

---

**THERMODYNAMICS  
– INTERACTION STUDIES  
– SOLIDS, LIQUIDS  
AND GASES**

---

Edited by **Juan Carlos Moreno-Piraján**

**INTECHWEB.ORG**

## **Thermodynamics – Interaction Studies – Solids, Liquids and Gases**

Edited by Juan Carlos Moreno-Piraján

### **Published by InTech**

Janeza Trdine 9, 51000 Rijeka, Croatia

### **Copyright © 2011 InTech**

All chapters are Open Access distributed under the Creative Commons Attribution 3.0 license, which permits to copy, distribute, transmit, and adapt the work in any medium, so long as the original work is properly cited. After this work has been published by InTech, authors have the right to republish it, in whole or part, in any publication of which they are the author, and to make other personal use of the work. Any republication, referencing or personal use of the work must explicitly identify the original source.

As for readers, this license allows users to download, copy and build upon published chapters even for commercial purposes, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

### **Notice**

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

**Publishing Process Manager** Viktorija Zgela

**Technical Editor** Teodora Smiljanic

**Cover Designer** Jan Hyrat

**Image Copyright** Zsolt, Biczó, 2010. Used under license from Shutterstock.com

First published September, 2011

Printed in Croatia

A free online edition of this book is available at [www.intechopen.com](http://www.intechopen.com)  
Additional hard copies can be obtained from [orders@intechweb.org](mailto:orders@intechweb.org)

Thermodynamics – Interaction Studies – Solids, Liquids and Gases,

Edited by Juan Carlos Moreno-Piraján

p. cm.

ISBN 978-953-307-563-1

**INTECH** OPEN ACCESS  
PUBLISHER

**INTECH** open

**free** online editions of InTech  
Books and Journals can be found at  
**[www.intechopen.com](http://www.intechopen.com)**



---

# Contents

---

## **Preface IX**

- Chapter 1 **Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design 1**  
Agnieszka K. Bronowska
- Chapter 2 **Atmospheric Thermodynamics 49**  
Francesco Cairo
- Chapter 3 **Thermodynamic Aspects of Precipitation Efficiency 73**  
Xinyong Shen and Xiaofan Li
- Chapter 4 **Comparison of the Thermodynamic Parameters Estimation for the Adsorption Process of the Metals from Liquid Phase on Activated Carbons 95**  
Svetlana Lyubchik, Andrey Lyubchik, Olena Lygina, Sergiy Lyubchik and Isabel Fonseca
- Chapter 5 **Thermodynamics of Nanoparticle Formation in Laser Ablation 123**  
Toshio Takiya, Min Han and Minoru Yaga
- Chapter 6 **Thermodynamics of the Oceanic General Circulation – Is the Abyssal Circulation a Heat Engine or a Mechanical Pump? 147**  
Shinya Shimokawa and Hisashi Ozawa
- Chapter 7 **Thermodynamic of the Interactions Between Gas-Solid and Solid-Liquid on Carbonaceous Materials 163**  
Vanessa García-Cuello, Diana Vargas-Delgadillo, Yesid Murillo-Acevedo, Melina Yara Cantillo-Castrillon, Paola Rodríguez-Estupiñán, Liliana Giraldo and Juan Carlos Moreno-Piraján
- Chapter 8 **Thermodynamics of Interfaces 201**  
Omid Moradi

- Chapter 9 **Exergy, the Potential Work** 251  
Mofid Gorji-Bandpy
- Chapter 10 **Dimensionless Parametric Analysis of Spark Ignited Free-Piston Linear Alternator** 271  
Jinlong Mao, Zhengxing Zuo and Huihua Feng
- Chapter 11 **Time Resolved Thermodynamics Associated with Diatomic Ligand Dissociation from Globins** 301  
Jaroslava Miksovská and Luisana Astudillo
- Chapter 12 **Some Applications of Thermodynamics for Ecological Systems** 319  
Eugene A. Silow, Andrey V. Mokry and Sven E. Jørgensen
- Chapter 13 **Statistical Thermodynamics of Material Transport in Non-Isothermal Mixtures** 343  
Semen Semenov and Martin Schimpf
- Chapter 14 **Thermodynamics of Surface Growth with Application to Bone Remodeling** 369  
Jean-François Ganghoffer
- Chapter 15 **Thermodynamic Aspects of CVD Crystallization of Refractory Metals and Their Alloys** 403  
Yu. V. Lakhotkin
- Chapter 16 **Effect of Stagnation Temperature on Supersonic Flow Parameters with Application for Air in Nozzles** 421  
Toufik Zebbiche
- Chapter 17 **Statistical Mechanics That Takes into Account Angular Momentum Conservation Law - Theory and Application** 445  
Illia Dubrovskiy
- Chapter 18 **The Role and the Status of Thermodynamics in Quantum Chemistry Calculations** 469  
Llored Jean-Pierre
- Chapter 19 **Thermodynamics of ABO<sub>3</sub>-Type Perovskite Surfaces** 491  
Eugene Heifets, Eugene A. Kotomin, Yuri A. Mastrikov, Sergej Piskunov and Joachim Maier
- Chapter 20 **Advances in Interfacial Adsorption Thermodynamics: Metastable-Equilibrium Adsorption (MEA) Theory** 519  
Gang Pan, Guangzhi He and Meiyi Zhang

- Chapter 21 **Towards the Authentic *Ab Initio* Thermodynamics** 543  
In Gee Kim
- Chapter 22 **Thermodynamics of the Phase Equilibria of Some Organic Compounds** 595  
Raisa Varushchenko and Anna Druzhinina
- Chapter 23 **Thermodynamics and Thermokinetics to Model Phase Transitions of Polymers over Extended Temperature and Pressure Ranges Under Various Hydrostatic Fluids** 641  
S  verine A.E. Boyer, Jean-Pierre E. Grolier,  
Hirohisa Yoshida, Jean-Marc Haudin and Jean-Loup Chenot
- Chapter 24 **Thermodynamics and Reaction Rates** 673  
Miloslav Pekař
- Chapter 25 **The Thermodynamics *in* Planck's Law** 695  
Constantinos Ragazas
- Chapter 26 **Statistical Thermodynamics** 717  
Anatol Malijevsk  y
- Chapter 27 **Thermodynamics Approach in the Adsorption of Heavy Metals** 737  
Mohammed A. Al-Anber
- Chapter 28 **Thermodynamics as a Tool for the Optimization of Drug Binding** 765  
Ruth Matesanz, Benet Pera and J. Fernando D  az
- Chapter 29 **On the Chlorination Thermodynamics** 785  
Brocchi E. A. and Navarro R. C. S.
- Chapter 30 **Thermodynamics of Reactions Among Al<sub>2</sub>O<sub>3</sub>, CaO, SiO<sub>2</sub> and Fe<sub>2</sub>O<sub>3</sub> During Roasting Processes** 825  
Zhongping Zhu, Tao Jiang, Guanghui Li,  
Yufeng Guo and Yongbin Yang
- Chapter 31 **Thermodynamic Perturbation Theory of Simple Liquids** 839  
Jean-Louis Bretonnet
- Chapter 32 **Probing Solution Thermodynamics by Microcalorimetry** 871  
Gregory M. K. Poon
- Chapter 33 **Thermodynamics of Metal Hydrides: Tailoring Reaction Enthalpies of Hydrogen Storage Materials** 891  
Martin Dornheim





---

## Preface

---

Thermodynamics is one of the most exciting branches of physical chemistry which has greatly contributed to the modern science. Since its inception, great minds have built their theories of thermodynamics. One should name those of Sadi Carnot, Clapeyron, Clausius, Maxwell, Boltzmann, Bernoulli, Leibniz etc. Josiah Willard Gibbs had perhaps the greatest scientific influence on the development of thermodynamics. His attention was for some time focused on the study of the Watt steam engine. Analysing the balance of the machine, Gibbs began to develop a method for calculating the variables involved in the processes of chemical equilibrium. He deduced the phase rule which determines the degrees of freedom of a physicochemical system based on the number of system components and the number of phases. He also identified a new state function of thermodynamic system, the so-called free energy or Gibbs energy ( $G$ ), which allows spontaneity and ensures a specific physicochemical process (such as a chemical reaction or a change of state) experienced by a system without interfering with the environment around it. The essential feature of thermodynamics and the difference between it and other branches of science is that it incorporates the concept of heat or thermal energy as an important part in the energy systems. The nature of heat was not always clear. Today we know that the random motion of molecules is the essence of heat. Some aspects of thermodynamics are so general and deep that they even deal with philosophical issues. These issues also deserve a deeper consideration, before tackling the technical details. The reason is a simple one - before one does anything, one must understand what they want.

In the past, historians considered thermodynamics as a science that is isolated, but in recent years scientists have incorporated more friendly approach to it and have demonstrated a wide range of applications of thermodynamics.

These four volumes of applied thermodynamics, gathered in an orderly manner, present a series of contributions by the finest scientists in the world and a wide range of applications of thermodynamics in various fields. These fields include the environmental science, mathematics, biology, fluid and the materials science. These four volumes of thermodynamics can be used in post-graduate courses for students and as reference books, since they are written in a language pleasing to the reader.

They can also serve as a reference material for researchers to whom the thermodynamics is one of the area of interest.

**Juan Carlos Moreno-Piraján**  
Department of Chemistry  
University of the Andes  
Colombia





# Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design

Agnieszka K. Bronowska  
*Heidelberg Institute for Theoretical Studies Heidelberg,  
Germany*

## 1. Introduction

Biologically relevant macromolecules, such as proteins, do not operate as static, isolated entities. On the contrary, they are involved in numerous interactions with other species, such as proteins, nucleic acid, membranes, small molecule ligands, and also, critically, solvent molecules. These interactions often display a remarkable degree of specificity and high affinity. Fundamentally, the biological processes rely on molecular organisation and recognition events. Binding between two interacting partners has both enthalpic ( $\Delta H$ ) and entropic ( $-T\Delta S$ ) components, which means the recognition event is associated with changes of both the structure and dynamics of each counterpart. Like any other spontaneous process, binding occurs only when it is associated with a negative Gibbs' free energy of binding ( $\Delta G$ ), which may have differing thermodynamic signatures, varying from enthalpy- to entropy-driven. Thus, the understanding of the forces driving the recognition and interaction require a detailed description of the binding thermodynamics, and a correlation of the thermodynamic parameters with the structures of interacting partners. Such an understanding of the nature of the recognition phenomena is of a great importance for medicinal chemistry and material research, since it enables truly rational structure-based molecular design.

This chapter is organised in the following way. The first part of it introduces general principles which govern macromolecular associations under equilibrium conditions: the free energy of binding and its enthalpic and entropic components, the contributions from both interacting partners, interaction energy of the association, and specific types of interactions – such as hydrogen bonding or van der Waals interactions, ligand and protein flexibility, and ultimately solvent effects (e.g. solute-solvent interactions, solvent reorganisation). The second part is dedicated to methods applied to assess particular contributions, experimental as well as computational. Specifically, there will be a focus on isothermal titrational calorimetry (ITC), solution nuclear magnetic resonance (NMR), and a discussion of computational approaches to the estimation of enthalpic and entropic contributions to the binding free energy. I will discuss the applicability of these methods, the approximations behind them, and their limitations. In the third part of this chapter, I will provide the reader several examples of ligand-protein interactions and focus on the forces driving the associations, which can be very different from case to case. Finally, I will address several practical aspects of assessing the thermodynamic parameters in molecular design, the

bottlenecks of methods employed in such process, and the directions for future development.

The information content provided by thermodynamic parameters is vast. It plays a prominent role in the elucidation of the molecular mechanism of the binding phenomenon, and – through the link to structural data – enables the establishment of the structure-activity relationships, which may eventually lead to rational design. However, the deconvolution of the thermodynamic data and particular contributions is not a straightforward process; in particular, assessing the entropic contributions is often very challenging.

Two groups of computational methods, which are particularly useful in assessment of the thermodynamics of molecular recognition events, will be discussed. One of them are methods based on molecular dynamics (MD) simulations, provide detailed insights into the nature of ligand-protein interactions by representing the interacting species as a conformational ensemble that follows the laws of statistical thermodynamics. As such, these are very valuable tools in the assessment of the dynamics of such complexes on short (typically, picosecond to tens of nanosecond, occasionally microsecond) time scales. I will give an overview of free energy perturbation (FEP) methods, thermodynamic integration (TI), and enhanced sampling techniques. The second group of computational methods relies on very accurate determinations of energies of the macromolecular systems studied, employing calculations based on approximate solutions of the Schrödinger equation. The spectrum of these quantum chemical (QM) methods applied to study ligand-protein interactions is vast, containing high-level ab initio calculations: from Hartree-Fock, through perturbational calculations, to coupled-clusters methods; DFT and methods based on it (including “frozen” DFT and SCC-DFTTB tight binding approaches); to semi-empirical Hamiltonians (such as AM1, PM3, PM6, just to mention the most popular ones) (Piela, 2007, Stewart, 2009). Computational schemes based the hybrid quantum mechanical –molecular mechanical (QM/MM) regimes will also be introduced. Due to the strong dependence of the molecular dynamics simulations on the applied force field, and due to the dependence of both MD simulations and QM calculations on the correct structure of the complex, validation of results obtained by these methodologies against experimental data is crucial.

Isothermal titration calorimetry (ITC) is one of the techniques commonly used in such validations. This technique allows for the direct measurement of all components of the Gibbs' equation simultaneously, at a given temperature, thus obtaining information on all the components of free binding energy during a single experiment. Yet since these are *de facto* global parameters, the decomposition of the factors driving the association, and investigation of the origin of force that drives the binding is usually of limited value. Nonetheless, the ITC remains the primary tool for description of the thermodynamics of ligand-protein binding (Perozzo *et al.*, 2004). In this chapter, I will give a brief overview of ITC and its applicability in the description of recognition events and to molecular design.

Another experimental technique, which has proven very useful in the experimental validation of computational results, is NMR relaxation. These measurements are extremely valuable, as they specifically investigate protein dynamics on the same time scales as MD simulations. As such, the results obtained can be directly compared with simulation outputs. In addition, the Lipari-Szabo model-free formalism (Lipari and Szabo, 1982) is relatively free of assumptions regarding the physical model describing the molecular motions. The only requirement is the internal dynamics being uncorrelated with the global tumbling of the system under investigation. The results of the Lipari-Szabo analysis, in the form of generalised order parameters ( $S_{LS}^2$ ), can be readily interpreted in terms of the

conformational entropy associated with the measured motions (Yang and Kay, 1996). It has been shown that for a wide range of motion models, the functional dependence of the conformational entropy on the order parameter is similar, suggesting that changes in order parameters can be related to changes in entropy in a model-independent manner. I will introduce the application of this model-free formalism to MD simulation, for the study of dynamical behaviour of ligand-protein complexes and the estimation of changes in the conformational entropy upon ligand-protein association. The MD simulations, performed on several proteins in complexes with their cognate ligands, indicate that the molecular ensembles provide a picture of the protein backbone dynamics that show a remarkably high degree of consistency with NMR relaxation data, regardless of the protein's size and structure (Schowalter and Brüschweiler, 2007).

In this chapter I will also address the enthalpy-entropy compensation phenomenon and the challenges it imposes on molecular design. The generality of this phenomenon have been a subject of debate for many years. Although this compensation is not a thermodynamic requirement as such (Ford, 2005, Sharp, 2001), it has been very frequently observed in protein-ligand interactions (Whitesides and Krishnamurthy, 2005). Briefly, stronger and more directed interactions are less entropically favourable, since the tight binding constricts molecular motions. The detailed mechanism of enthalpy-entropy compensation is, nonetheless, highly system-dependent, and this compensation does not obey a single functional form. An example of enthalpy-entropy compensation and its consequences to the design process will be provided.

A discussion of the thermodynamics of protein-ligand interactions would not be complete without commenting on dynamic allostery and cooperativity. The mechanism of allostery plays a prominent role in control of protein biological activity, and it is becoming accepted that protein conformational dynamics play an important role in allosteric function. Changes of protein flexibility upon ligand binding affect the entropic cost of binding at distant protein regions. Counter-intuitively, proteins can increase their conformational entropy upon ligand binding, thus reducing the entropic cost of the binding event (MacRaild *et al.*, 2007). I will discuss these phenomena, illustrating them through several examples of biologically-relevant protein-ligand interactions.

The overall aim of this chapter is to introduce the forces driving binding events, and to make the reader familiar with some general rules governing molecular recognition processes and equally to raise awareness of the limitations of these rules. Combining the structural information with equilibrium thermodynamic data does not yield an understanding of the binding energetics under non-equilibrium conditions, and global parameters, obtained during ITC experiments, do not enable us to assess the individual contributions to the binding free energy. Certain contributions, such as entropy, may behave in a strongly non-additive and highly correlated manner (Dill, 1997). This chapter will discuss the boundaries of rational molecular design guided by thermodynamic data.

## 2. Principles

### 2.1 Enthalpic and entropic components of free binding energy

A non-covalent association of two macromolecules is governed by general thermodynamics. Similarly to any other binding event (or – in a broader context – to any spontaneous process), it occurs only when it is coupled with a negative Gibbs' binding free energy (1), which is the sum of an enthalpic, and an entropic, terms:

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

where  $\Delta G$  is free binding energy,  $\Delta H$  is enthalpy,  $\Delta S$  entropy, and  $T$  is the temperature.

The enthalpic contribution to the free energy reflects the specificity and strength of the interactions between both partners. These include ionic, halogen, and hydrogen bonds, electrostatic (Coulomb) and van der Waals interactions, and polarisation of the interacting groups, among others. The simplest description of entropic contribution is that it is a measure of dynamics of the overall system. Changes in the binding entropy reflect loss of motion caused by changes in translational and rotational degrees of freedom of the interacting partners. On the other hand, changes in conformational entropy may be favourable and in some cases these may reduce the entropic cost of binding (MacRaild *et al.*, 2007). Solvation effects, such as solvent re-organisation, or the release of tightly bound water upon ligand binding can contribute significantly to the entropic term of the binding free energy.

The Gibbs equation can be also written as in equation (2):

$$\Delta G = -RT \ln K_d \quad (2)$$

where  $R$  is a gas constant,  $T$  is the temperature, and  $K_d$  is binding constant. This formulation emphasises the relationship between Gibbs energy and binding affinity. The ligand-protein association process can be represented in the form of a Born-Haber cycle. A typical cycle is showed in Figure 1. The 'intrinsic' free energy of binding between ligand  $L$  and protein  $P$  is represented by  $\Delta G_i$ , whereas the experimentally observable free energy of binding is represented by  $\Delta G_{obs}$ .

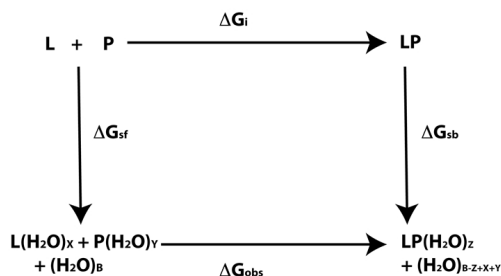


Fig. 1. An example of Born-Haber cycle for ligand-protein (LP) association. It relates the experimentally observed free energy of binding ( $\Delta G_{obs}$ ) with 'intrinsic' free energy of binding ( $\Delta G_i$ ) between ligand ( $L$ ) and protein ( $P$ ) and with solvation free energies of free interacting partners ( $\Delta G_{sf}$ ) and the resulting complex ( $\Delta G_{sb}$ ).  $X$ ,  $Y$ ,  $Z$ , and  $B$  refer to the number of water molecules involved in solvation of the unbound ligand ( $X$ ), unbound protein ( $Y$ ), ligand-protein complex ( $Z$ ), and to the bulk solvent ( $B$ ).

Two additional processes can be defined: the free energy of solvation of the free (unbound) interacting partners ( $\Delta G_{sf}$ ), and the free energy of solvation of the ligand-protein complex ( $\Delta G_{sb}$ ). Since the free energy is a state function, it is independent of the path leading from from one state of the system to another. Hence, the observable free energy of binding can be written as in equation (3):



$$\Delta G_{obs} = \Delta G_i + \Delta G_{sb} - \Delta G_{sf} \quad (3)$$

The equation above shows how the observable free energy of binding can be decomposed into the 'intrinsic' term, and the solvation contributions from the ligand-protein complex and unbound interactors. Similar decomposition can be done for the enthalpic and entropic terms separately, as these terms are also state functions.

Since the enthalpic and entropic contributions to the binding free energy depend on many system-specific properties (such as protonation states, binding of metal cations, changes in conformational entropy from one ligand to another in a way which is very difficult to predict, etc), the conclusion is that optimising the overall free energy remains the most viable approach to rational (structure-based) molecular design. Attempting to get an insight into individual components of the free energy requires re-thinking the whole concept of ligand-protein binding. This means regarding ligand-protein complexes as specifically interacting yet flexible ensembles of structures rather than rigid entities, and the role of solvation effects. The significant contribution of specific interactions and flexibility to the 'intrinsic' component of binding free energy, and solvation effects will be discussed next in this chapter.

## 2.2 Specific interactions

### 2.2.1 Electrostatic interactions

Electrostatic interactions, involved in ligand-protein binding events, can be roughly classified into three types; charge-charge, charge-dipole, and dipole-dipole. Typical charge-charge interactions are those between oppositely charged atoms, ligand functional groups, or protein side chains, such as positively charged (amine or imine groups, lysine, arginine, histidine) and negatively charged (carboxyl group, phosphate groups, glutamate side chain). An important contribution to the enthalpy change associated with a binding event arises from charge-dipole interactions, which are the interactions between ionised amino acid side chains and the dipole of the ligand moiety or water molecule. The dipole moments of the polar side chains of amino acid also affect their interaction with ligands.

### 2.2.2 Van der Waals interactions

Van der Waals interactions are very important for the structure and interactions of biological molecules. There are both attractive and repulsive van der Waals interactions that control binding events. Attractive van der Waals interactions involve two induced dipoles that arise from fluctuations in the charge densities that occur between adjacent uncharged atoms, which are not covalently bound. Repulsive van der Waals interactions occur when the distance between two involved atoms becomes very small, but no dipoles are induced. In the latter case, the repulsion is a result of the electron-electron repulsion that occurs in two partly-overlapping electron clouds.

Van der Waals interactions are very weak (0.1- 4 kJ/mol) compared to covalent bonds or electrostatic interactions. Yet the large number of these interactions that occur upon molecular recognition events makes their contribution to the total free energy significant.

Van der Waals interactions are usually treated as a simple sum of pairwise interatomic interactions (Wang *et al.*, 2004). Multi-atom VdW interactions are, in most cases, neglected. This follows the Axilrod-Teller theory, which predicts a dramatic (i.e. much stronger than for pairwise interactions) decrease of three-atom interactions with distance (Axilrod and Teller, 1943). Indeed, detailed calculations of single-atom liquids (Sadus, 1998) and solids

(Donchev, 2006) indicate that multi-body effects amount to only 5% of the total energy (Finkelstein, 2007). However, Finkelstein (2010) shows that those largely ignored multi-atom Van der Waals interactions may lead to significant changes in free energy in the presence of covalent bonds. Those changes can be comparable to those caused by the substitutions of one atom by another one in conventional pairwise Van der Waals interactions. Thus, the currently used force fields (applied in MD simulations) need to be revised.

### 2.2.3 Hydrogen bonds

Hydrogen bonds are non-covalent, attractive interactions between a hydrogen covalently bonded to some electronegative group (“donor”), and another electronegative atom, such as oxygen or nitrogen (“acceptor”). The hydrogen bond can be described as an electrostatic dipole-dipole interaction. However, it also has some features of covalent bonding: it is specific, directional, it produces interatomic distances shorter than sum of van der Waals radii, and usually it involves a limited number of interaction partners, which can be interpreted as a type of valence.

Proteins contain ample hydrogen bond donors and acceptors both in their backbone and in the side chains. The environment (aqueous solvent, protein-protein network, lipid bilayers) in which proteins of interest are immersed also contains numerous proton donors and acceptors – be it water molecule, interacting proteins, lipid headgroups, or DNA/RNA. Hydrogen bonding, therefore, occurs not only between ligand and protein and within the protein itself, but also within the surrounding medium.

Like all non-covalent interactions, hydrogen bonds are fairly weak: in biological conditions, the strength of hydrogen bonds varies between 5-30 kJ/mol (outside of biological systems, the strength of hydrogen bonds may vary from 2 kJ/mol to even 155 kJ/mol for HF<sub>2</sub><sup>-</sup>) (Emsley, 1980), which is weaker than ionic or covalent bonds. However, because of their relative weakness, they can be formed and broken rapidly during binding event, conformational changes, or protein folding. Thus, hydrogen bonds in biological systems may be switched on or off with energies that are within the range of thermal fluctuations. This is one of the prime factors that facilitates macromolecular association events, and biological activity. Another key factor is related to the strict geometric rules, followed by hydrogen bonds in biological systems. Namely, their orientations, lengths, and angular preferences, which make hydrogen bonding very specific. Due to these properties, the role of hydrogen bonds in governing specific interactions in biological recognition processes is absolutely crucial. Hydrogen bonds, both intra- and inter-molecular, are partly responsible for the secondary, tertiary, and quaternary structures of proteins, nucleic acids, and also some synthetic polymers. They play a pivotal role in molecular recognition events, and they tune the properties of the macromolecular system (e.g. mechanical strength, binding specificity). These geometric rules were among the first to be extracted from crystal structure databases (Bissantz *et al.*, 2010). While the preferred geometries of hydrogen bonds are easily defined, their contributions to binding free energy are system-specific (Davis and Teague, 1999, Williams and Ladbury, 2003). Hydrogen bonds always convey specificity to a recognition process but do not always add much binding free energy (Bissantz *et al.*, 2010).

Hydrogen bonds can vary quite considerably in their strength. Often, a stronger hydrogen bond implies higher penalty of desolvation, so the net free energy gain of a stronger hydrogen bond might be seriously compromised. However, such a picture is not always the case. Hydrogen bond strength, in the context of the free energy changes, should be carefully

examined, as it is likely to vary considerably from one ligand-protein system to another one (Barratt *et al.*, 2005, 2006).

Regarding weak hydrogen bonds, the most prominent donor is the CH group. These interactions, despite of their weakness, play an important role in stabilising appropriate conformations of ligand-protein complexes, for instance among the complexes between protein kinases and their inhibitors (Bissantz *et al.*, 2010). Protonated histidines can also act as strong CH donors (Chakrabarti and Bhattacharyya, 2007). Weak hydrogen bonds, their nature, and their role in ligand-protein interactions have been extensively reviewed by Panigrahi and Desiraju (2007).

#### 2.2.4 Halogen bonds and multipolar interactions

The concept of halogen bonds is similar to hydrogen bonds: both types of interactions involve relationships between an electron donor and electron acceptor. In hydrogen bonding, a hydrogen atom acts as the electron acceptor and forms a non-covalent bond by accepting electron density from an electronegative atom (“donor”). In halogen bonding, a halogen atom is the donor.

Despite of their prevalence in complexes between proteins and small organic inhibitors (many of them contain halogen atoms due to solubility and bioavailability) and their importance for medicinal chemistry, the significance of halogen bonds in biological context has been overlooked for a long time (Zhou *et al.*, 2010). For a number of years, halogen atoms were regarded as hydrophobic appendages, convenient – from the molecular design point of view – to fill apolar protein cavities. The nature of halogen interactions (such as directionality, sigma-holes) was not studied in detail and not regarded as very important. Indeed, halogen bonds are, in general, fairly weak interactions. On the other hand, in some cases they can compete with hydrogen bonds, thus should be considered in more details, given the importance of hydrogen bonds in ligand-protein interactions and given that many of synthesised small organic compounds contain halogen bonds in their structure (Bissantz *et al.*, 2010, Zhou *et al.*, 2010).

Halogens involved in halogen bonds are chlorine, bromine, iodine, and fluorine (not very often). All four halogens are capable of acting as donors (as proven by computational and experimental data) and follow the general trend:  $F < Cl < Br < I$ , with iodine normally forming the strongest bonds, as the strength increases with the size of the halogen atom. From the chemical point of view, the halogens, with the exception of fluorine, have unique electronic properties when bound to aryl or electron withdrawing alkyl groups. They show an anisotropy of electron density distribution with a positive area (so-called  $\sigma$ -hole) of electrostatic potential opposite the carbon-halogen bond (Clark *et al.*, 2007). The molecular origin of the  $\sigma$ -hole can be explained quantum chemically and the detailed description is provided in the work by Clark and coworkers (2007). Briefly, a patch of negative charge is formed around the central region of the bond between carbon and halogen atom, leaving the outermost region positive (hence the “hole”).

Available experimental data show the strong influence of halogen bonds on binding affinity. Replacement of hydrogen by halogen atom is often used by medicinal chemists in order to increase the affinity. Indeed, in a series of adenosine kinase inhibitors, a 200-fold affinity gain from hydrogen to iodine has been observed (Iltzsch *et al.*, 1995). Another spectacular, 300-fold affinity difference upon iodine substitution was observed in a series of HIV reverse-transcriptase inhibitors (Benjahad *et al.*, 2003). Unsurprisingly, substitution of hydrogen by iodine typically leads to the largest affinity gain, since the strength of the halogen bond increases with the size of halogen atom.

Halogen atoms can interact with the oxygen and with the carbon atoms of C=O groups, as well. The former attributes to the halogen bond formation, the latter is a hallmark of so-called orthogonal multipolar interactions. These interactions are formed by two dipolar functional groups, which are in a close distance from each other. Only recently it received attention in the field of medicinal chemistry and ligand-protein interactions (Paulini *et al.*, 2005), even though it has been described for a long time. This interaction is known to contribute to ligand-protein stabilisation (Fischer *et al.*, 2008), and it is particularly important in the context of halogen bonds (Bissantz *et al.*, 2010 and references therein). It is worth bearing in mind that in an orthogonal (perpendicular) orientation of two dipoles, the actual dipole contribution to interaction energy is zero. Thus, higher order electrostatic and dispersion terms must be responsible for this type of interaction. The disappearance of the dipole term may turn a repulsive electrostatic interaction into an attractive one. Because of its high electron density and low polarisability, fluorine's preference for dipolar interactions is more pronounced than for the other halogens (Bissantz *et al.*, 2010). Chlorine and other heavy halogens also form multipolar interactions with carbonyl groups, but they show a tendency for the C-X bond to be parallel rather than orthogonal to the amide plane, a consequence of the  $\sigma$ -hole (Bissantz *et al.*, 2010).

### 2.2.5 Hydrophobic interactions

The interactions between ligands and the hydrophobic side chains of proteins contribute significantly to the binding free energy. The hydrophobic residues mutually repel water and other polar groups and results in a net attraction of the non-polar groups of ligand. In addition, apolar and aromatic rings of tryptophan, phenylalanine, and tyrosine participate in "stacking" interactions with aromatic moieties of ligand. Many studies have demonstrated that the hydrophobic interactions, quantified by the amount of hydrophobic surface buried upon ligand binding, is the structural parameter correlating best with binding free energy (Bissantz *et al.*, 2010, Perozzo *et al.*, 2004). It holds well for very diverse sets of ligands (Boehm and Klebe, 1996) as well as for protein-protein interactions (Vallone *et al.*, 1998). It should be emphasised, though, that a considerable part of the affinity gain caused by hydrophobic interactions in hydrophobic binding pockets comes from sub-optimal solvation of the pocket in the unbound (apo) state.

Aromatic interactions, hydrophobic effect, and other solvent effects will be discussed further in the following parts of this chapter.

### 2.2.6 Interactions mediated by aromatic rings

Aromatic rings deserve special attention in the context of ligand-protein interactions. Interactions between ligands and protein aromatic side chains (Phe, Trp, and Tyr) are widespread in ligand-protein complexes (Bissantz *et al.*, 2010). The unique steric and electronic properties of these side chains, which give rise to large polarizabilities and quadrupole moments, result in preferred geometries upon interactions.

For interactions between two aromatic systems, two geometries are predominant: one, where two rings are parallel to each other, and the perpendicular, edge-to-face arrangement. High-accuracy *ab initio* CCSD(T) quantum chemical calculations of the dimerisation energy of benzene predict these two geometries to be isoenergetic (Hobza *et al.*, 1996), which agrees with experimental results qualitatively and quantitatively (Grover *et al.*, 1987, Krause *et al.*, 1991).

An introduction of heteroatoms into aromatic ring affects the ratio of both geometries. The preference to perpendicular interactions increases when the acidity of the interacting “side” atoms increases; this happens upon introduction of a strongly electron-withdrawing substituent in either ortho- or para-position. This was demonstrated by high-accuracy quantum chemical calculations by Sinnokrot and Sherrill (2004): The interaction between benzene as a donor and fluorobenzene as the acceptor, while both compounds were perpendicular to each other, was  $\sim 0.3$  kcal/mol weaker than that of the benzene dimer. With reverse of roles (fluorobenzene as the donor), the interaction became  $\sim 0.6$  kcal/mol stronger as compared to the benzene dimer.

For perpendicularly-oriented aromatic-aromatic interactions, studies on several model systems showed that aliphatic-aromatic interactions in the same orientation provide a favourable contribution to the free energy of the same magnitude as aromatic-aromatic interactions (Turk and Smithrud, 2001). For aliphatic-aromatic interactions, interaction energy becomes more favourable when acidity of the interacting CH unit of aliphatic counterpart increases. Study conducted by Tsuzuki *et al.* (2000) showed that ethane (sp<sup>3</sup> hybridisation of carbon atom, less acidic) is a worse binder of benzene than acetylene (sp hybridisation of carbon atom, more acidic), and the difference in dissociation energies between acetylene-benzene and ethane-benzene complexes is around 1 kcal/mol. In ligand-protein complexes, this type of interaction can be found in interactions between aromatic side chains and methyl groups. The strength of such interactions depends on the group to which the interacting methyl group is bound: the more electronegative the group, the more the preference towards perpendicular geometry of interacting methyl-aromatic side chain is pronounced (Bissantz *et al.*, 2010).  $\Pi - \pi$  interactions are also displayed by amide bonds of protein backbone (namely, their pi faces) and ion pairs - interactions between acidic (Asp, Glu) and basic (Lys, Arg) side chains.

Aromatic interactions are not limited to  $\pi - \pi$  interactions. Recently, the nature of favourable interactions between heavier halogens and aromatic rings has been studied, in particular in the context of halogen bonds. C-H - halogen interactions can be regarded as “very weak hydrogen bonds” (Desiraju, 2002).

### 2.3 Solvent effects, structural waters, and the bulk water

Any binding event displaces water molecules from the interaction interface or from the binding pocket, while simultaneously desolvating the ligand (or a part of it). Although most of those waters are disordered and loosely associated with protein structure, such displacement affects the whole solvation shell around the ligand-protein complex (Poornima and Dean, 1995b).

While the vast majority of those water molecules are mobile and easily displaceable, some are tightly bound to the protein structure. Tightly bound water molecules are often conserved across multiple crystal structures of ligand-protein complexes (Poornima and Dean, 1995c). Often, those water molecules play an important role in tuning the biological activity of the protein, as in the case of many enzymes (Langhorst *et al.*, 1999, Nagendra *et al.*, 1998, Poornima and Dean, 1995a). Those water molecules may be regarded as part of the protein structure. Ligand-protein interactions are often mediated by water molecules buried in the binding site and forming multiple hydrogen bonds with both binding partners (Poornima and Dean, 1995a-c). In other cases, those bound water molecules are released to

the bulk upon ligand binding. Such displacement may affect the thermodynamic signature of the binding event in a dramatic way. It is generally assumed that the release of a water molecule from a rigid environment should be entropically favorable. The upper limit of the entropy gained for transferring a water molecule from a protein to bulk solvent was estimated to be 2 kcal/mol at room temperature (Dunitz, 1994). This gain would be compensated by loss of enthalpy, so the total contribution to the free energy (as a sum of its enthalpic and entropic terms) of a single water molecule released from the protein to the bulk is difficult to guess. Moreover, in order to reach this 2 kcal/mol limit the water molecule would have to be fixed very rigidly while bound. This is often not the case, and it has been observed in numerous occasions that even very tightly bound, “structural” waters may retain a significant amount of residual mobility (Denisov *et al.*, 1997, Fischer and Verma, 1999, Matthews and Liu, 2009, Smith *et al.*, 2004).

“Structural” water molecules affect their surrounding not only via direct interactions (such as hydrogen-bonding network), but also by influencing the dynamical behaviour of their environment. Numerous cases have been reported when binding of the structural water affected protein flexibility (Fischer and Verma, 1999, Smith *et al.*, 2004). The direction of such influence cannot be predicted by simple rules, as it is heavily dependent on the details of the binding site – some protein become more dynamic upon water binding (Fischer and Verma, 1999), while other ones become more rigid (Mao *et al.*, 2000). Yet ignoring those water effects is likely to lead to substantial errors in the free energy predictions. The importance of the contributions of “structural” water molecules to binding events and its implications for drug design have been emphasised in a study by Michel *et al.* (Michel *et al.*, 2009).

The traditional, enthalpy-dominated view of ligand-protein association largely neglects solvation effects, which strongly affect the thermodynamic profile of a binding event. Recently it became clear that studying the hydration state of a protein binding pocket in the apo (unbound) state should be a routine procedure in rational drug design, as the role of solvation in tuning binding affinity is critical. Solvation costs are a plausible reason why some ligands, despite fitting into a binding site, fail during experimental tests as inhibitors. Young and coworkers showed that an optimised inhibitor of factor Xa turns virtually inactive when the isopropyl group interacting in the S4 pocket of factor Xa is substituted by hydrogen: The compound (PDB code 2J4I) is characterised by  $K_i$  of 1 nM. Replacing the isopropyl group by hydrogen reduces its affinity to 39  $\mu$ M. Substitution of this group by hydrogen, apart from reducing the number of favourable hydrophobic interactions, leads to unfavourable solvation of the binding pocket (Young *et al.*, 2007, and references therein).

Desolvation of the ligand itself may sometimes control the binding free energy. For highly hydrophilic ligands, the desolvation costs may be very high and make unfavourable contributions to the binding (Daranas *et al.* 2004, MacRaild *et al.*, 2007, Syme *et al.*, 2010). The calorimetric study of  $\beta$ -galactose derivatives binding to arabinose binding protein (ABP) showed dramatic differences in binding free energy between several deoxy derivatives (Daranas *et al.*, 2004). The most likely reason of 4-deoxygalactose failing to bind to ABP is the unfavourable desolvation cost (Bronowska and Homans, unpublished data).

Spectroscopic evidence shows that (1) water molecules in the first solvation shell (surrounding the hydrophobic solute) are more flexible than it was originally thought (Finney and Soper, 1994) and (2) hydrogen bonds at hydrophobic surfaces are weaker than it was assumed (Scatena *et al.*, 2001). In addition, the properties of the water molecules from first two solvation shells are very different from those of bulk water, as emerged from terahertz spectroscopy results (Ebbinghaus *et al.*, 2007, Heugen *et al.*, 2006).

## 2.4 Classical and non-classical hydrophobic effect

The concept of the classical hydrophobic effect relies on a hydrophobic solute disrupting the structure of bulk water. This decreases entropy due to ordering of water molecules around the hydrophobic entity. Such unfavourable effects can be minimised if solute molecules aggregate. Upon aggregation, water molecules form one larger “cage” surrounding the hydrophobic aggregate, and the surface area of such aggregate is smaller than the sum of surface areas of individual (non-aggregated) solutes. This makes the entropic contribution less unfavourable and hence makes the free energy more favourable (Homans, 2007).

If this mechanism was the sole driving force for a protein-ligand interaction, all binding events involving hydrophobic ligands would be entropy-driven. This is not the case. Several years ago, in the group of Steve Homans (University of Leeds), we studied the thermodynamics signature of ligand binding by the mouse major urinary protein (MUP). This protein is characterised by a strongly hydrophobic binding pocket and it binds a handful of very different hydrophobic ligands – long-chain alcohols and pyrazine derivatives, among others. Surprisingly, the ITC data showed that the binding was enthalpy-driven (Barratt *et al.*, 2005). This was combined with a negative change in heat capacity upon binding - a hallmark of the hydrophobic effect.

In order to elucidate the molecular origin of this unusual binding signature, we employed computational methods, such as molecular dynamics (MD) simulations. I will discuss the results in more details later in this chapter. The data showed that the key to this favorable enthalpy of binding of ligands to MUP seems to be the sub-optimal solvation of the binding pocket in apo (unbound) state: only a few water molecules remained there prior to ligand binding. The favourable enthalpic component was, thus, largely determined by ligand desolvation, with only a minor contribution from desolvation of the protein. Such complexation thermodynamics driven by enthalpic components have been referred to as the “non-classical hydrophobic effect”.

## 2.5 Enthalpy-entropy compensation, binding cooperativity, and protein flexibility

The enthalpic and entropic contributions are related. An increase in enthalpy by tighter binding may directly affect the entropy by the restriction of mobility of the interacting molecules (Dunitz, 1995). This phenomenon, referred to as enthalpy-entropy compensation, is widely observed, although its relevance is disputed (Ford, 2005). Such compensation, although frequently observed, is not a requirement: if it was, meaning that changes in  $\Delta H$  were always compensated by opposing changes in  $T\Delta S$ , optimisation of binding affinities would not be possible, which is clearly not the case.

In connection to the enthalpy-entropy compensation, ligand-protein interactions can be cooperative, which means the binding energy associated with them is different than the sum of the individual contributions to the binding free energies. Cooperativity provides a medium to transfer information, enhance or attenuate a response to changes in local concentration and regulate the overall signalling/reaction pathway. Its effects are either positive (synergistic) or negative (interfering), depending on whether the binding of the first ligand increases or decreases the affinity for subsequent ligands. Noncooperative (additive) binding does not affect the affinity for remaining ligands and the subsequent binding sites can be regarded as independent.

Cooperativity is often linked to pronounced conformational changes in the structure of the protein. It can be, in some cases, caused by structural tightening through the presence of additional interactions; inter-atomic distances become shorter and interaction becomes

enthalpically more favorable. Evidence for such a mechanism has been reported for many ligand-protein complexes; biotin-streptavidin being one of the most extensively studied (Williams *et al.*, 2003). In other cases, cooperativity can occur in the absence of any conformational changes of the protein, and be driven solely by changes in protein dynamics (Homans, 2005, Wand, 2001). Catabolite-activated protein (CAP) is a very good example of such dynamic allostery. CAP is a transcriptional activator that exists as a homodimer in solution, with each subunit comprising a ligand-binding domain at the N-terminal domain and a DNA-binding domain at the C-terminal domain (Harman, 2001). Two cyclic AMP (cAMP) molecules bind to CAP dimer, and this binding increases affinity of CAP for DNA (Harman, 2001). Binding of each cAMP molecule shows negative cooperativity, i.e. binding of the first cAMP molecule decreases affinity of binding of the second cAMP molecule to CAP. This is accompanied by absence of long-range structural changes. Thermodynamic analysis, performed by a combination of ITC and solution NMR, confirmed that the observed negative cooperativity was entirely driven by changes in protein entropy (Popovych *et al.*, 2009). Thus, it is more appropriate to describe the phenomenon of cooperativity in terms of thermodynamics rather than merely conformational changes (if any such changes can be observed), since it is fundamentally thermodynamic in its nature.

Examples above illustrate the importance of protein dynamics in binding events. Proteins tend to compensate the unfavourable entropic contribution to ligand binding by increasing their dynamics in regions distant from the ligand binding site (Evans and Bronowska, 2010, MacRaild *et al.*, 2007) Flexible binding sites may require more flexible ligand moieties than 'stiffer' ones. The traditional focus on the enthalpic term (direct and specific interactions) and dominance of the 'induced fit' model has led to an overly enthalpic view of the world that neglects protein flexibility. Such view of the ligand-protein binding events, although very intuitive, is flawed by neglect of entropic contributions and – as a consequence – an impairment to correct predictions of free binding energy. Although it is true that tighter interactions make binding more favourable, the thermodynamic signature of a “good” binder does not need to be dominated by an enthalpic term.

### 3. Methods

#### 3.1 Experimental methods

Many experimental techniques have been developed to study various aspects of ligand-protein thermodynamics. X-ray crystallography provides very valuable information about the enthalpic contribution (hydrogen and halogen bonds, electrostatic interactions, etc). Although it focuses on static structures of ligand-receptor complexes, it also yields some information on entropic contribution. B-factors (temperature factors), obtainable for heavy (non-hydrogen) atoms of the complex under investigation, are sensitive to the mean square displacements of atoms because of thermal motions, therefore they reflect on ligand-protein dynamics. However, B-factors do not distinguish time scales of the motions and their interpretation is not straightforward. X-ray (Makowski *et al.*, 2011) and neutron scattering (Frauenfelder and Mezei, 2010) also reflect on ligand-protein dynamics. The former one focuses on global changes in protein size and shape in a time-resolved manner, while the latter reports on motion amplitudes and time scales for positions of hydrogen atoms. Another technique useful in understanding protein dynamics both in unbound (apo) and bound (holo) forms is fluorescence spectroscopy (Weiss, 2000). Single molecule techniques



also offer an opportunity to measure contributions to binding events from interacting partners individually. Hydrogen-deuterium exchange mass spectrometry (HX-MS) and related methods, have been very successful in studying protein dynamics in large supramolecular complexes (Wales and Engen, 2006). The motion of the entire complex and individual contributors, and the dynamics of the binding events can be investigated by time-resolved HX-MS (Graf *et al.*, 2009). Another technique frequently used to study binding events is surface plasmon resonance (SPR), which allows for straightforward determination of equilibrium binding constants (Alves *et al.*, 2005). Terahertz spectroscopy is a relatively new technique, used primarily to probe solvation of macromolecules and their complexes (Ebbinghaus *et al.*, 2007). It is very sensitive to changes of the collective water network dynamics at the at the macromolecule-water interface. Terahertz absorption spectroscopy can also be used to probe collective modes in ligand-protein complexes (Xu *et al.*, 2006).

There are two groups of methods that deserve special attention in the context of thermodynamics of binding events and will be discussed more in details in the following part of this chapter. One of these is NMR spectroscopy, especially powerful for the study of ligand-protein dynamics, hence the entropic contribution to the binding free energy (Meyer and Peters, 2003). The other group contains calorimetric techniques, which are very important for the study of biological systems, their stability, and the thermodynamics of macromolecular interactions. Currently, two most popular techniques applied to investigate biological systems are differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC). The former quantifies the heat capacity and enthalpy of thermal denaturation, the latter measures the heat exchanged during macromolecular association. While DSC provides the way to estimate the stability of the system (protein, nucleic acid, ligand-protein complex, etc), ITC is an excellent tool to study the thermodynamics of binding events (Perozzo *et al.*, 2004). Since this chapter is dedicated to the thermodynamics of macromolecular associations, in the course of this chapter I will focus mainly on ITC and its applications to study biological systems.

### 3.1.1 Isothermal titration calorimetry (ITC)

ITC measures the heat evolved during macromolecular association events. In an ITC experiment, one binding partner (ligand) is titrated into a solution containing another binding partner (protein), and the extent of binding is determined by direct measurement of heat exchange (whether heat is being generated or absorbed upon the binding). ITC is the only experimental technique where the binding constant ( $K_d$ ), Gibbs free energy of binding ( $\Delta G$ ), enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ) can be determined in a single experiment (Perozzo *et al.*, 2004). ITC experiments performed at different temperatures are used to estimate the heat capacity change ( $\Delta C_p$ ) of the binding event (Perozzo *et al.*, 2004).

During last few decades, ITC has attracted interest of broader scientific community, as a powerful technique when applied in life sciences. Several practical designs emerged, but the greatest advances have happened during last 10 years. Development of sensitive, stable, and - last but not least - affordable calorimeters made calorimetry a very popular analytical procedure and ITC became the gold standard in estimations of macromolecular interactions. Given the ability of ITC to obtain a full thermodynamic description of the system studied, the technique has found widespread applicability in the study of biological systems. Apart from its versatility and simple experimental setup, ITC also has advantages over some other

techniques: the experiments can be performed in a physiologically relevant buffer, no surface effects have to be taken into account, and the interacting species do not require immobilisation or labelling.

ITC is also used for determination of binding affinity-independent reaction stoichiometry. The reaction stoichiometry is estimated from the titration equivalence point. Provided this, ITC is increasingly used in the analysis of macromolecular complexes involving multiple binding events (e.g. protein aggregation or the formation of multi-protein complexes). Systems that involve multiple binding events that occur at two or more interacting sites often demonstrate cooperativity, which is an important mechanism of regulation in biological systems (Brown, 2009).

Using ITC it is also possible to study protonation effects, in cases when protein-ligand binding is coupled to changes in the protonation state of the system. If the formation of the complex changes the protonation state of ligand as well as that of the protein (whether free or bound), proton transfer with the solvent occurs. As a result, the signal measured by ITC will contain the heat effect of protonation/ deprotonation, contributing to the overall heat of binding. Repeating the experiment at the same pH in buffers with different ionisation enthalpies but otherwise under the same conditions allows for the determination of the number of protons released/ accepted by buffer solution. From this, the intrinsic binding enthalpy corrected by protonation heats, can be established (4).

$$\Delta H_{obs} = \Delta H_{bind} + nH + \Delta H_{ion} \quad (4)$$

ITC can also provide information about solvation effects. If  $\Delta H$  is determined at a range of temperatures, the change in the constant pressure heat capacity ( $\Delta C_p$ ) for an interaction is given by the slope of the linear regression analysis of  $\Delta H_{obs}$  plotted vs. temperature. There is a strong correlation between  $\Delta C_p$  and the amount of desolvated (buried) surface area of a macromolecular complex. Thus, for the ligand-protein binding events,  $\Delta C_p$  is most often negative, when the complex is regarded as a reference state. Through this correlation, changes in  $\Delta C_p$  are measure of solvation state of the macromolecule and involvement of solvent effects in binding event (Perozzo *et al.*, 2004).

### 3.1.1.1 Experimental setup

In a typical ITC experiment, a solution of ligand is injected (titrated) into a solution of the protein, in small volumes, over the time. During that time, the changes in heat resulting from the interaction are monitored (Figure 2, upper panel). Each peak represents a heat change associated with the injection of a ligand sample into the protein solution inside the ITC reaction cell. Concentrations of both ligand and protein in their respective solutions are known. As the ligand-protein system reaches saturation, the heat changes diminish until only heats of dilution are observed. A binding curve is then obtained from a plot of the heats from each titration against the ratio of ligand and protein inside the ITC cell (Figure 2, lower panel). The binding curve is analysed with the appropriate binding model to determine the thermodynamic parameters.

ITC is a straightforward technique to accurately measure binding events with affinity range from mM up to high-nM. Problems occur when the ligand binds very tightly, in a single-digit nM and below. This is due to the titration curve becoming too steep to fit accurately. In such cases, the displacement experiments are commonly used. Such experimental setup consists of binding a low-affinity binding ligand first and then displacing it during titration

with a stronger binder of interest. However, this method requires precise knowledge of binding constants of those weak binders. The experimental setup of the displacement assay is often challenging, as there are several factors increasing the error of measurements.

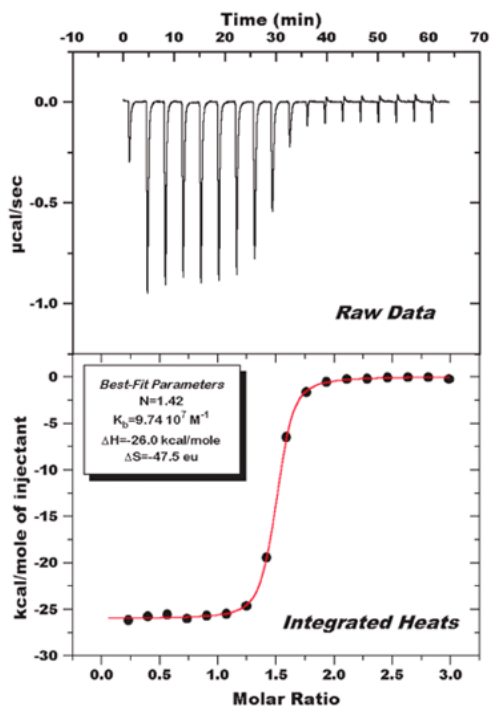


Fig. 2. An example of ITC data. Raw data, representing observed changes in heat resulting from interactions are shown in the upper panel. The resulting binding curve is displayed in the lower panel (from MicroCal materials <http://www.microcal.com/technology/itc.asp>).

It is worth remembering that ITC experiment not only measures the heat absorbed or released during binding reactions, but it also detects the total heat effect in the calorimetric cell upon titration of ligand. Thus, the experimental data contain contributions arising from non-specific effects, such as dilution of ligand and protein, mixing two solutions of slightly different compositions, temperature differences between the ITC cell and the titrating syringe, and so forth. In order to determine these contributions the control experiments need to be performed in order to extract the heat of ligand-protein complex formation.

### 3.1.1.2 Thermodynamic content of ITC data

The  $\Delta G$  determines the stability of any ligand-protein complex of interest, which makes it very useful for studies and predictions of structure-activity relationships. The conventional analysis of ITC data involves fitting an appropriate model (*i.e.* single- or two-site binding model) to the data, and obtaining the binding constant. Quite often, though, more sophisticated models (such as multiple interacting-site models) must be applied, if the behaviour of the system is more complex.

As mentioned earlier, observed overall  $\Delta G$  can be very similar regardless of the driving force, which can be very different from one case to another.  $\Delta G$  can be the same for an interaction with positive  $\Delta S$  and  $\Delta H$  (entropy-driven, binding signature dominated by the classical hydrophobic effect), an interaction with negative  $\Delta S$  and  $\Delta H$  (enthalpy-driven binding signature), or all sort of combinations of negative  $\Delta H$  and positive  $\Delta S$ . As described in the previous section, ligand-protein complexes tend to compensate for enthalpic and entropic contributions, making changes in  $\Delta G$  less sensitive to the molecular details of the interactions. Therefore, dissection of  $\Delta G$  into enthalpic and entropic contributions is of a fundamental importance for understanding of the binding energetics.

#### 3.1.1.2.1 Enthalpic contributions

The change in the enthalpy represents the changes in energy associated with specific, non-covalent interaction. However, such an interpretation is too simplistic to describe experimental  $\Delta H$  values, and the physical meaning of observed  $\Delta H$  seems to be more complex. The measured changes in enthalpy are the result of the formation and breaking of many individual bonds; it reflects the loss of protein-solvent hydrogen bonds and van der Waals interactions, the loss of ligand-solvent interactions, the formation of ligand-protein bonds, salt bridges and van der Waals contacts, the re-organisation of the intra-molecular hydrogen-bonding network of the protein, solvent reorganisation near the protein surface, conformational changes at the binding site due to the binding event, and many more. These individual components may produce either favourable or unfavourable contributions, depending on the system.

The treatment of each component individually is very challenging since the global heat effect of a particular interaction is a balance between the enthalpy of the ligand binding to the protein and to the solvent. Several approaches have been employed to investigate the energetics of individual bonds, including alanine scanning mutagenesis (Perozzo *et al.*, 2004 and references therein), and removal of particular hydrogen bonds at the binding site (Connelly *et al.*, 1994). However, these approaches suffer from the major bottleneck, resulting from the fact that a direct relation between the change in enthalpy and the removal of the corresponding specific interactions cannot be made *a priori*.

A large part of the observed  $\Delta H$  is due to a bulk hydration effect, as emerged from ITC studies carried out in water and deuterium (Connelly *et al.*, 1993). Frequently, water molecules are located at complex interfaces, improving the complementarity of the surfaces and extending hydrogen-bonding networks. This should contribute favourably to the enthalpy, but it may be offset by an unfavourable entropic contribution (Perozzo *et al.*, 2004). The role of interfacial water was studied by lowering water activity by adding osmolytes such as glycerol to the solution. It was found that complexes with a low degree of surface complementarity and no change in hydration are tolerant to osmotic pressure (Perozzo *et al.*, 2004, and references therein).

#### 3.1.1.2.2 Entropic contributions

$\Delta S$  may be calculated directly from  $\Delta G$  and  $\Delta H$ , according to the Gibbs' equation. Its physical representation is not straightforward. It is often related to the dynamics and flexibility of the system (Diehl *et al.*, 2010, Homans, 2007), sometimes dubbed as a 'measure of the system's disorder' (which is incorrect). It has been proposed that the  $\Delta S$  associated

with ligand-protein binding can, at a given temperature, be expressed as the sum of several contributing effects. The main one is related to solvent effects. The burial of water-accessible surface area upon binding event should result in release of confined or interfacial water molecules to the bulk. This should contribute favourably to the total entropy of interaction. A positive entropy change is usually a strong indication that water molecules have been released from the complex surface (Jelesarov and Bosshard, 1999). On the other hand, interfacial water remaining upon binding can also contribute positively to the total entropy of the interaction (Fischer and Verma, 1999).

Another important entropic contribution is related to the reduction of conformational (rotational and vibrational) degrees of freedom of protein side-chains. In addition to these, the ligand loses translational degrees of freedom upon binding. All these contribute unfavourably to the overall entropy of interaction. However, in some cases the protein increases the number of conformational degrees of freedom upon ligand binding, as observed by NMR and deduced from MD simulations (MacRaild *et al.*, 2007, Stoeckmann *et al.*, 2008). This is likely to happen in order to partly offset the unfavourable entropic contribution from ligand binding and thus to reduce the overall thermodynamic cost of that process.

#### 3.1.1.2.3 Enthalpy-entropy compensation

As mentioned in the previous section of this chapter, this phenomenon is described by the linear relationship between the change in enthalpy and the change in entropy. This means that favourable changes in binding enthalpy are compensated by opposite changes in binding entropy and vice versa, resulting in very small changes in overall free binding energy. Enthalpy-entropy compensation is an illustration of the 'motion opposes binding' rule, and it is believed to be a consequence of altering the weak inter-molecular interactions as well as being related to solvent effects. Since both  $\Delta H$  and  $\Delta S$  are connected to  $\Delta C_p$ , the correlation between enthalpy/entropy and heat capacity changes is clear.

Enthalpy-entropy compensation is a difficult problem to address in the context of rational molecular design. In such framework, the goal is to maximise the binding affinity of a complex of the designed compound and the protein target. The optimisation strategy requires simultaneous minimisation of both enthalpic and entropic penalties. However, reducing one of them usually means increasing the other.

### 3.1.2 Nuclear Magnetic Resonance (NMR) spectroscopy

Thermodynamics of biologically-relevant macromolecules and their complexes can be characterised by measurements using NMR spectroscopy. The basis of NMR spectroscopy is the non-zero nuclear magnetic moment of many elements, such as  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , or  $^{19}\text{F}$ . When put into an external static magnetic field ( $B$ ), the different nuclear spin states of these elements become quantised with energies proportional to their projections onto vector  $B$ . The energy differences are also proportional to the field strength and dependent on the chemical environment of the element, which makes NMR an ideal technique to study 3D structural and dynamical properties of the systems.

A variety of NMR methods have been introduced to study ligand-protein interactions. These methods include one-, two- and three-dimensional NMR experiments. Many studies, to date, proved the power of stable-isotope labelling and isotope-edited NMR in the investigation of ligand-protein interactions. Recent development of techniques allowed for

the study of ligand-induced conformational changes, investigating positions and dynamic behaviour of bound water molecules, and for quantification of conformational entropy. The steady-state heteronuclear Overhauser effects (NOEs) are very useful for structural analysis of three-dimensional structures of macromolecules in solution (Boehr *et al.*, 2006, Meyer and Peters, 2003). It is important to note that the NOE occurs through space, not through chemical bonds, which makes it applicable to characterise non-covalent binding events. When ligand binds the NOEs change dramatically, and transferred NOEs (trNOEs), relying on different tumbling times of free and bound interactors, can be observed.

Another NMR technique commonly used for identification of the ligand binding is chemical shift mapping (Meyer and Peters, 2003). Briefly, chemical shifts describe the dependence of nuclear magnetic energy levels on the electronic environment in the given macromolecule. Electron density, electronegativity, and aromaticity are among the factors affecting chemical shifts. Not surprisingly, binding event changes the chemical shifts of both interacting partners, particularly in the area of the association (e.g. protein binding pocket, protein-peptide interaction interface). Thus, changes in chemical shifts can be used to identify binding events and to describe the location of the binding.

Ligand-protein thermodynamics can be investigated using NMR relaxation analysis, which provides an insight into protein dynamics in the presence and the absence of ligand. These results can be integrated with thermodynamic data obtained from isothermal titration calorimetry (ITC) experiments and computational results (e.g. MD simulations). For proteins, the relaxation rates of backbone ( $^{15}\text{N}$ ) and side chains ( $^2\text{H}$  and  $^{13}\text{C}$ ), can be obtained. The time scales available to NMR ranges over 17 orders of magnitude, reflecting protein motions on timescales from picoseconds to milliseconds (Boehr *et al.*, 2006). This covers all the relevant motions of proteins and their complexes.

Backbone and side chain (methyl groups) NMR relaxation measurements revealed the role of protein dynamics in ligand binding and protein stability (Boehr *et al.*, 2006). Development of molecular biology techniques for incorporation of stable,  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes into expressed proteins allowed for design and application of modern multidimensional heteronuclear NMR techniques. As a consequence, the maximum size of the macromolecule studied using these techniques rose from about 10 kDa (when  $^1\text{H}$  homonuclear NMR is used) to 50 kDa and beyond (using  $^{13}\text{C}$  and  $^{15}\text{N}$  heteronuclear NMR with fractional  $^2\text{H}$  enrichment). Application of modern TROSY (transverse relaxation optimized spectroscopy) techniques further expanded the size limitations of NMR, reaching up to the 900 kDa (Fernandez and Wider, 2003).

While NMR methodologies are being developed to study ligand-protein complexes in solid state, special techniques have been developed specifically to study protein stability and folding (Baldus, 2006), or in-cell NMR (Burz *et al.*, 2006), providing complementary information to fluorescence studies in biological settings. In this chapter I will briefly discuss only application of relaxation analysis in solution for the study of ligand-protein thermodynamics, specifically intrinsic entropic contributions.

### 3.1.2.1 Slow and fast dynamics: from dynamics to entropy

Conformational changes that may be associated with ligand binding events generally occur on 'slow' (microsecond to millisecond) time scales and thus report on slower motions than protein backbone and side chain fluctuations (pico-to-nanoseconds). There is no straightforward relationship between 'slow' and 'fast' motions. Experiments on several ligand-enzyme systems have shown that binding events, which decrease the 'fast' motions, may increase, decrease, or not affect the 'slow' motions (Boehr *et al.*, 2006). This obviously has an effect on the overall entropy contribution, but this has not been fully explored.

NMR relaxation techniques have been used to study multiple time scale dynamics of ligand-protein complexes. Their results show that even though large conformational changes occur on 'slow' time scale, 'fast' (pico-to-nanosecond) protein motion plays important roles in all aspects of binding event. These are typically probed by measuring three relaxation rates: the longitudinal relaxation rate ( $R1$ ), the transverse relaxation rate ( $R2$ ), and the NOE. These relaxation rates are directly related to the spectral density function,  $J(\omega)$ . This function is proportional to the amplitude of the fluctuating magnetic field at the frequency  $\omega$ . Such fluctuating magnetic fields are caused by molecular motion in an external magnetic field, which is closely coupled to nuclear spin relaxation (Boehr *et al.*, 2006, and references therein). In early studies of ps-ns time scale protein dynamics, various models for protein internal motion were used to generate different spectral density functions that were then compared to the experimental data. Subsequently, Lipari and Szabo (Lipari and Szabo, 1996) generated a spectral density function (5) that is independent of any specific physical model of motion, which is shown in equation 5 and is referred to as model free formalism.

$$J(\omega) = \frac{S^2 \tau_m}{1 + \omega^2 \tau_m^2} + \frac{(1 - S^2) \tau}{1 + \omega^2 \tau^2} \quad (5)$$

For isotropic tumbling (ligand-protein complex tumbles in the water solution), where  $\tau_m$  is the correlation time for the overall rotational diffusion of the macromolecule,  $S^2$  is the order parameter, and  $\frac{1}{\tau} = \frac{1}{\tau_m} + \frac{1}{\tau_e}$ , where  $\tau_e$  is the time scale (ns) for the bond vector internal motions. An order parameter of 1 indicates complete restriction of internal motion, and  $S^2 = 0$  indicates unrestricted isotropic internal motion. It should be emphasised that  $S^2$  parameters have a straightforward physical interpretation. The simplest model relates  $S^2$  to 'diffusion in a cone' with semi-angle  $\theta$ , and is shown in Figure 3.

$$S^2 = \frac{\cos^2 \theta (1 + \cos \theta)^2}{4}$$

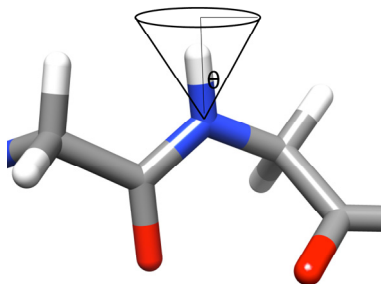


Fig. 3. Physical interpretation of  $S^2$  order parameters.  $S^2$  can be interpreted as a measure of a free rotation of a bond vector (here - N-H) in a cone. The semi-angle  $\theta$  is displayed. Smaller  $S^2$  parameters correspond to more flexible bond vectors.  $S^2 = 0$  means unrestricted rotation of the bond vector.

There were attempts to relate order parameters to structural characteristics of proteins and ligand-protein complexes. It was observed that amino acids with smaller side chains tend to show - intuitively - greater backbone flexibility than those with bulkier side chains (Goodman *et al.*, 2000). However, the variation of backbone amide  $S^2$  parameters is larger than the differences between the averages for different amino acid types. Backbone amide order parameters are also only weakly affected by secondary structure elements, with loops having only slightly smaller average  $S^2$  N-H values than helices or beta-turns (Kay *et al.*, 1989). Backbone  $S^2$  N-H values can be predicted from structures using a simple model that takes account of local contacts to the N-H and C=O atoms of each peptide group (Zhang and Brüschweiler, 2002).

A more sophisticated model for predicting dynamics from structure has recently been reported (tCONCOORD) (Seeliger *et al.*, 2007). tCONCOORD allows for a fast and efficient sampling of protein's conformational degrees of freedom based on geometrical restraints. Weak correlation between side chain order parameters and contact distance between the methyl carbon and neighboring atoms, with solvent exposure (Ming and Brüschweiler, 2004), and amino acid sequence conservation patterns (Mittermaier *et al.*, 2003) have been reported in literature. These results demonstrate that protein dynamics are strongly affected by the unique architecture of the protein as well as the environment. Thus, it cannot be readily predicted by the bioinformatic techniques, based on the primary/secondary structure analysis. Developing a fast and reliable method of assessment of protein dynamics is, nevertheless, crucial for predictions of ligand-protein interactions - as it will be shown in the course of this chapter, dynamics affects all stages of molecular recognition events.

Order parameters can be related to entropy through the relationship developed by Yang and Kay (1996). This formalism quantifies the conformational entropy associated with observable protein motions by means of a specific motion model. For a wide range of motion models, the functional dependence of entropy on the  $S^2$  parameter was demonstrated to be similar (Yang and Kay, 1996). This suggests that changes in  $S^2$  can be related to the conformational entropy change in a model-independent manner. This approach has many advantages: it is straightforward, relatively free of assumptions (the requirement is that the internal motions are uncorrelated with the global tumbling of the macromolecule), and applicable to both NMR experiments and theoretical approaches (MD simulations). Moreover, since  $S^2$  parameters are measured per bond vector, this approach enables site-specific reporting of any losses, gains, and redistributions of conformational entropy through different dynamic states of the ligand-protein complex.

However, the model-free formalism can give only a qualitative view of micro-to-millisecond time scale motions. Failure to correctly account for anisotropic molecular tumbling and the assumption that all motions are un-correlated seriously compromises the usefulness of this approach for studying dynamics associated with large conformational changes or concerted motions. Because of the time scales, alternative approaches must be implemented to study motions occurring at a millisecond time scale (e.g. R2 relaxation dispersion).

### 3.1.3 Combination of ITC and NMR

As described, ITC obtains free energy as the global parameter, thus, effects like ligand-induced conformational changes, domain-swapping, or protein oligomerisation, which contribute to the overall  $\Delta G$ , will not be resolved. In order to assess the role those factors



play in a binding event, a combination of ITC and other techniques (such as NMR) need to be used. A combination of ITC and NMR proves useful in studying cooperativity phenomena. Heteronuclear NMR spectroscopy is one of the few experimental techniques capable of measuring the occupancies of individual binding sites on proteins and therefore determining microscopic binding affinities. Coupling this site-specific data (e.g. chemical shift mapping and/or relaxation analysis data) with the macroscopic binding data from ITC allows a complete description of the binding properties of the system. A method of determining cooperativity using heteronuclear solution NMR spectroscopy has been described using an isotope-enriched two-dimensional heteronuclear single-quantum coherence experiment (2D HSQC) (Tochtrop *et al.*, 2002). The ligands are isotopically labelled (usually  $^1\text{H}$ ,  $^{15}\text{N}$ , or  $^{13}\text{C}$ ), while the receptor remains unlabelled. Spectra are acquired at different molar ratios and the peak volumes are integrated. Isotherms are generated by plotting the peak volume integration against molar ratio. The data is then fitted to site-specific binding models to obtain the thermodynamic parameters (Brown, 2009).

### 3.2 Computational approaches

Computational approaches to ligand-protein interaction studies have great potential and the development of various methods, briefly described in this chapter, have been truly outstanding. However, every method - computational, experimental alike - has its limitations and computational methods should not be used in a 'black box' manner; one should beware of the 'Garbage In Garbage Out' phenomenon. Yet it is evident that theoretical approaches have finally come to the stage that makes rational molecular design truly rational.

During a binding event, the ligand may bind in multiple orientations. The conformation of either of the interacting partners can change significantly upon association. The network of intramolecular interactions (e.g. hydrogen bonds, salt bridges) can dramatically change (breaking and/or creating new contacts), and new intermolecular interactions occur. Water molecules and ions can be expelled upon binding, or - on the contrary - bind more tightly. Finally, conformational or solvation entropic contributions may play significant role, affecting the free energy in a way which is difficult to predict.

Growing amount of calorimetric data available allows the investigation of the thermodynamic profiles for many ligand-protein complexes in detail. When structural data (crystal, NMR) are available as well - and often it is the case - it is very appealing to speculate about the link between the structure of the complex and the thermodynamics of the binding event. However, such speculations are challenging. It is important to bear in mind that both enthalpic and entropic contributions to the free energy terms obtained from ITC experiments are global parameters, containing a mixture of different contributions, which can have either equal or opposing signs and different magnitudes. This may lead to various thermodynamic signatures of a binding event. Moreover, 'structural' interpretation of intrinsic entropic contributions is notoriously difficult. Hence, the experimental thermodynamic data cannot be easily interpreted on the basis of structural information alone. Last but not least, the contribution from the solvation effects is difficult to get insight into, and although direct experimental estimations of solvation free energy have been attempted, these always require additional assumptions (Homans, 2007, Shimokhina *et al.*, 2006).

No doubt, a great advantage of theoretical approaches lies with gaining an insight about each of those contributions and their de-convolution. Binding events (ligand-protein

binding pose and the strength of their interactions) can be predicted by molecular docking, albeit intrinsic entropic contributions and solvation effects are usually ignored. Dynamic behaviour of proteins and ligands can be studied using extensive molecular dynamics (MD) simulation, which, combined with experimental NMR and ITC data, provide extremely valuable information on configurational entropy changes upon binding event, and hence about the intrinsic entropic contribution to the free energy. The global free energy changes can be studied by free energy perturbation (FEP) calculations, or related methods, such as thermodynamic integration (TI). Molecular docking methods allow for a quick assessment of enthalpic contributions, while solvent effects can be studied either by quantum chemical (QM) calculations (e.g. COSMO model), hybrid QM/MM schemes, or FEP-related methods. Theoretical approaches allow also for investigations of transient phenomena, e.g. short-living alternative conformers from an ensemble that contribute to the binding event but which cannot be readily observed. In a situation – which is not uncommon – when an experimental structure of the protein target or a part of it is missing (such as in cases of most G-protein-coupled receptors), computational approaches allow the generation of such structures (e.g. by homology modelling, threading, or *ab initio* predictions) and its use for predictions which can be validated experimentally, despite of the absence of protein structural data. Therefore, usage of theoretical methods is indispensable – not only for the interpretation of the existing experimental data, but also to direct and design new experiments.

Because of space limitations, only two theoretical methods, which are the most relevant for thermodynamics of molecular binding events, will be briefly discussed: MD-related methods (which includes MD simulations, FEP-like approaches, methods which use MD algorithms with enhanced sampling, and hybrid QM/MM schemes), and quantum chemical (QM) calculations. This division is not strict and many of these methods overlap, e.g. QM/MM methods use both MD simulations and QM calculations, and FEP-like methods have many flavours, including hybrid QM/MM-FEP.

### 3.2.1 Molecular Dynamics (MD) simulations

Molecular dynamics (MD) simulation consists of the numerical solution of the Newton's equations of motion of a system (e.g. protein, or a ligand-protein complex in water environment). The potential energy of the particle system is described by a function called force field (6). The potential energy of the system ( $U$ ) is described as a sum of energy terms for covalent bonds, angles, dihedral angles, a van der Waals non-bonded term, and a non-bonded electrostatic term (Cornell *et al.*, 1995). Since the kinetic energy is also taken into account, the system is able to move across the energy barriers on the potential energy surface, which implies substantial changes (e.g. conformational) during the simulation.

$$\begin{aligned}
 U = & \sum_{\text{bonds}} k_r (r - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k\phi [1 + \cos(n\phi + \phi_0)] \\
 & + \sum_{\text{atom } i} \sum_{j \neq i} 4\epsilon_{i,j} \left[ \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon_0 r_{i,j}}
 \end{aligned} \tag{6}$$

The principles of MD simulations, algorithms used, and different types of force fields applied (all-atom, united atom, coarse-grain, etc) have been described in many publications

(Klepeis *et al.*, 2009 and references therein). MD methods rely on quality of the force field (parameters, inclusion of non-additive effects, etc), description of solvent effects, adequate sampling, and quality of initial structures used for the simulations. The quality of the results relies also on the duration of the simulation. There are limits on the time scales at which the system of interest can be considered. Simulation runs are fairly short: typically nanoseconds to microseconds, rarely extending to milliseconds, if super-fast computers are employed. Since biological processes (ligand-protein binding, large conformational changes, etc) typically occur at micro-to-millisecond scales, one needs to assess whether or not a simulation has reached equilibrium before the averages calculated can be trusted. Furthermore, the averages obtained need to be subjected to a statistical analysis, to make an estimate of the errors.

MD methods have been widely employed to study ligand-protein binding phenomena, conformational changes, solvent effects, and to assess individual contributions to the binding free energy. These methods are particularly useful in assessing the conformational entropic contribution to the free energy. Information about ps-to-ns time scale molecular motions can be readily obtained from the MD simulation trajectory and analysed either through diagonalisation of the covariance matrix of displacements of atomic Cartesian coordinates - quasi-harmonic analysis, Schlitter's approach (7,8), analysed through principal component analysis (PCA), or quantified NMR-like via generalised order parameters. Entropy changes can be estimated from the MD trajectory through Yang and Kay's relationship (9). The order parameter analysis has the advantage of being able to calculate order parameters by-vector, thus providing site-specific information on flexibility and hence intrinsic entropic contribution. Computed parameters can be also directly compared to the experimental results of NMR relaxation analysis (Best and Vendruscolo, 2004). In last few years several studies proved the success of this methodology in estimating of entropic contributions to the binding thermodynamics.

$$S < S' = \frac{1}{2} k_B \ln \det \left[ \mathbf{1} + \frac{k_B T e^2}{\hbar^2} \mathbf{M} \sigma \right] \quad (7)$$

$$\sigma_{ij} = \left\langle (x_i - \langle x_i \rangle) (x_j - \langle x_j \rangle) \right\rangle \quad (8)$$

$$S^2 = \frac{3}{2} \left[ \langle x^2 \rangle^2 + \langle y^2 \rangle^2 + \langle z^2 \rangle^2 + 2 \langle xy \rangle^2 + 2 \langle yz \rangle^2 + 2 \langle yz \rangle^2 \right] - \frac{1}{2} \quad (9)$$

MD-based approaches used for binding thermodynamics calculations include free energy perturbation (FEP) (Foloppe and Hubbard, 2006), thermodynamic integration (TI) (Straatsma and Berendsen, 1988), lambda-dynamics simulations (Knight and Brooks, 2009), Molecular Mechanics-Poisson-Boltzmann Surface Area (MM-PBSA) (Gilson and Zhou, 2007), Linear Interaction Energy (LIE) (Gilson and Zhou, 2007), and hybrid quantum chemical/molecular mechanical (QM/MM) (Senn and Thiel, 2009) methods.

Free energy perturbation (FEP) is used to calculate free energy differences between two states from MD simulations. These two states can represent, for instance, unbound (apo) protein and a ligand-protein complex (holo), or two ligand-protein complexes with different ligands. In the framework of FEP, the difference in the free energy difference for two states is obtained from the Zwanzig equation (10).

$$\Delta G(A \rightarrow B) = G_B - G_A = -k_B T \ln \left\langle \exp \left( -\frac{E_B - E_A}{k_B T} \right) \right\rangle_A \quad (10)$$

where A and B represent two states (e.g. apo and holo protein), G is the difference between free energies of both states,  $k_B$  is the Boltzmann's constant, T is the temperature, and the triangular brackets denote an average over a simulation run for state A.

FEP calculations converge properly only when the difference between these two states is small enough; therefore it is usually necessary to divide a perturbation into a series of smaller 'steps', which are calculated independently.

Thermodynamic integration (TI) is a related method to calculate free energy differences. Since the free energy can be expressed by its relation to the canonical partition function, the free energy difference in two different states can be used to calculate the difference of potential energy. TI calculations are usually executed by designing a thermodynamic cycle (Figure 1), and integrating along the relevant path. The path can be either a real chemical process or an artificial change (e.g. substitution of a methyl group by hydrogen atom).

The MD methods, despite their numerous successes, suffer of two major bottlenecks. One is the results are critically dependent on the force field used, therefore requires caution when use of appropriate force field and parameters. Many modern force fields are parametrised on experimental NMR data, some are able to include - to some extent - non-additive effects (electronic polarisation). Application of QM/MM schemes allow the inclusion of quantum effects to some extent. Another bottleneck is related to the adequacy of sampling. It is known that - due to relatively short time scales investigated - only some subsets of potential conformational changes can be observed, and often the system gets 'stuck' in a minimum, which does not have to be the global one. This makes the results heavily biased towards the starting structure and is very likely to underestimate the degree of molecular motions observed in the system. Prolonging the simulation time helps to solve the sampling problem only to some extent, and significantly increases the computational cost of MD simulations. Thus, in order to overcome the sampling issue, various enhanced sampling techniques have been employed. One of such methods, frequently used, is replica exchange MD (REMD), which attempts to overcome the problem of multiple-minima by exchanging temperatures of several replicas of the system. These replicas are non-interacting with each other and they run at different temperatures. REMD is also called "parallel tempering" (Earl and Deem, 2005). Another approach used to improve sampling is to construct the bias potential and add it to the potential energy function of the system (force field). This group of methods, referred to as umbrella sampling methods, consist of metadynamics (Laio and Gervasio, 2008), conformational flooding, and accelerated dynamics (Lange et al., 2006). The core feature of metadynamics is the construction of so-called reference potential, which is one that is the most similar to the actual potential. That is, repulsive markers are placed in a coarse time line in a space that is spanned by a small number of relevant collective variables. These markers are then placed on top of the underlying free energy landscape in order to push the system to rapidly accumulate in the initial basin by discouraging it from revisiting points in configurational space. In this way, the system is allowed to escape the lowest transition state as soon as the growing biasing potential and the underlying free energy well exactly counterbalance each other, effectively allowing the simulation to escape free energy minima.

### 3.2.2 Quantum mechanical (QM) calculations

Ligand-protein interactions can be driven by quantum effects. These include charge transfer, halogen bonds, or polarisation. Stabilisation energy related to charge transfer can be several kcal/mol and force field-based schemes cannot describe this stabilisation correctly.

'Conventional' QM calculations, using HF, DFT, or semi-empirical methods provide a way to obtain the ground state energy of a ligand-protein system or a part of it. Most programs based on these are capable of studying molecular properties such as atomic charges, multipole moments, vibrational frequencies, and spectroscopic constants. In addition, there are methods allowing the study of excited-state processes, such as time-dependent DFT or restricted open-shell Kohn-Sham (ROKS) (Li and Liu, 2010).

The application area of QM methods is vast. QM calculations are used for charge derivation for molecular dynamics simulations, for description of direct interactions (hydrogen bonds, halogen bonds, aromatic stacking), for calculations of pKa, protonation, redox states, and for studying solvation effects, such as computing free solvation energies.

Derivation of accurate charges for a system being studied is an important step in preparation for MD simulation. Failure in charge representation will inevitably lead to incorrect results. Derivation of charges is done using QM calculations, usually in several steps, involving optimisation, electrostatic potential generation, and fitting charges into atoms. RESP methodology, based on charges derived from *ab initio* HF/6-31G\* level of theory has been for many years a standard in deriving charges for MD simulations (Bayly *et al.*, 1993, Cieplak *et al.*, 1995, Cornell *et al.*, 1993).

Charge distribution is also required for the calculation of the solvation properties using conductor-like screening model (COSMO) (Klamt and Schüürmann, 1993). COSMO, just like any other continuum solvent approach, approximates the solvent by a dielectric continuum, surrounding the solute molecules outside of a molecular cavity. In COSMO, the polarisation charges of the continuum, caused by the polarity of the solute molecule, is derived from a scaled-conductor approximation (hence the name). In this way, the charge distribution of the molecule, which can be obtained from the QM calculations, and the energy of the interaction between the solvent and the solute molecule can be determined.

QM calculations are also used for force field development, such as adding new parameters and incorporating non-additive effects. Several studies indicate that non-additive effects (e.g. electronic polarisation) significantly affect binding affinities of many ligands (Ji and Zhang, 2008, 2009 and references therein). Electrostatic interactions are critically dependent on charge distribution around both interacting species, and this distribution is heavily dependent on the conformation (geometry) of the complex. Description of hydrogen bonding is also affected by electronic polarisation – some hydrogen bonds, which are found broken during MD simulation using 'conventional' force fields are found to be stable, when non-additive force field is used (Ji and Zhang, 2009). Corrections for polarisation can be added to MD force fields in order to derive protein charges more accurately and provide a better description of electrostatic interactions. Protein polarisation is important for stabilisation of the native structures of proteins. MD simulations indicate that inclusion of polarisation effects not only improves the description of protein native structures, but also distinguishes native from decoy dynamically: the former are more stable than the latter under the polarised force fields. These observations provide strong evidence that inclusion of polarisation effects in calculations of ligand-protein interactions is likely to greatly improve accuracy of such calculations.

QM methods are also used in molecular docking. The application of QM methods to molecular docking was pioneered by Raha and Merz (2004), who developed a semi-empirical QM-based scoring function and studied ion-mediated ligand binding processes. Their conclusion was that quantum chemical description is required for metal-containing systems, mainly because of poorly-defined atom types of metal atoms in most of the force field parameters, which cannot describe the interactions between a small molecule ligand and a metal ion in the active site of the protein.

Applicability of QM methods to study ligand-protein system has been discussed in literature (Raha *et al.*, 2007, Stewart, 2007). As well as these successes, many examples of the failure of QM approaches have been demonstrated. However, it should be kept in mind that most of these studies were based on either DFT, or semi-empirical Hamiltonians, which do not describe van der Waals interactions and hydrogen bonding terms of ligand-protein interactions correctly. This is, indeed, a serious limitation of “fast”, hence more popular QM methods. A straightforward way to solve this problem is to add additional correction terms to the QM energy. It has been demonstrated that the addition of the dispersion energy and corrections for hydrogen bonds improved the performance of semi-empirical QM methods dramatically (Rezac *et al.*, 2009). The recently developed PM6-DH2 method (Fanfrlik *et al.*, 2010) yields, to date, the most accurate results for non-covalent interactions among the semi-empirical QM methods. For small non-covalent complexes, the results obtained were comparable to the high-level wave-function theory-based calculations within chemical accuracy (1 kcal/mol) (Rezac *et al.*, 2009).

Another major bottleneck of QM methods applied to studying ligand-protein interactions thermodynamics is the size of the system. DFT can handle up to 150 atoms, highly-accurate methods such as coupled-clusters can handle a few tens of atoms and require very fast computers and long computing times. This limitation of the size of the systems that can be studied seriously compromises its usage in the study of ligand-protein thermodynamics. For instance, the usage of the linear scaling algorithm MOZYME (Stewart, 2009) based on the localised orbitals allows size increases to systems as large as 18 000 atoms and above, which allows calculation of very large ligand-protein complexes. Due to these developments QM methods have become, therefore, very useful for fast and highly accurate predictions of ligand-protein interactions energetics.

An alternative approach is to use so-called divide-and-conquer (DIVCON) algorithm (Dixon and Merz, 1996, 1997). The principle is to divide a large system into many smaller subsystems, separately determine the electron density of each of these subsystems, and then to add the corresponding contributions from each subsystem in order to obtain the total electron density and the energy.

## 4. Examples

In the previous sections of this chapter, I briefly introduced the forces governing macromolecular associations and characterised methods commonly used to assess these contributions. Here, I will illustrate on several examples of 'real' ligand-protein systems and the way how their binding thermodynamics is studied.

### 4.1 Hydrophobic versus hydrophilic binding pocket: MUP and HBP

Both histamine-binding protein (HBP) and mouse major urinary protein (MUP) are members of lipocalin family of proteins, so their overall structures are similar (Figure 4). The

binding pocket of HBP contains a number of polar and charged residues, hence it is an example of a 'hydrophilic' binder. In contrast to HBP, the binding pocket of MUP is very 'hydrophobic'. Surprisingly, both HBP and MUP are characterised by similar overall entropy of ligand binding.

In our recent study (Syme *et al.*, 2010) we compared the driving forces for binding between these two proteins in terms of entropic contributions from ligand, protein, and solvent. We performed an extensive study combining x-ray crystallography, NMR spectroscopy, ITC, MD simulations, and QM calculations.

#### 4.1.1 Structures of HBP and MUP

The structure of MUP was solved both by x-ray crystallography and solution NMR (Barratt *et al.*, 2005, Kuser *et al.*, 2001, Timm *et al.*, 2001). Several ligand-MUP complexes were studied, including long-chain alcohols, pyrazine derivatives, and pheromones as ligands. Regardless of the chemical nature of ligand, the protein structures are very similar to each other: the desolvated ligand, which occupies the central, hydrophobic binding pocket, causes very few conformational changes (Figure 4, left panel).

The crystal structure of HBP complexed with histamine revealed two binding sites for the ligand: one of them possessing considerably higher affinity than the other (Figure 4, right panel) (Syme *et al.*, 2010). Therefore, in order to simplify the thermodynamic analysis of ligand binding, a mutant of HBP was designed. In this mutant, denoted as HBP-D24R, negatively charged aspartic acid D24 inside the "low" affinity site was replaced by larger and positively charged arginine. This abolished binding of ligand to the "low" affinity site.

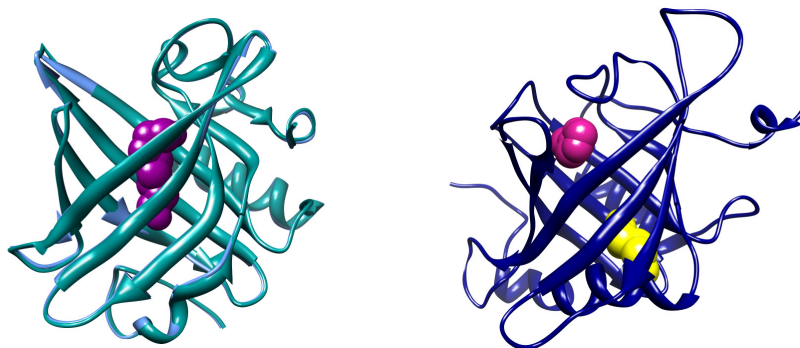


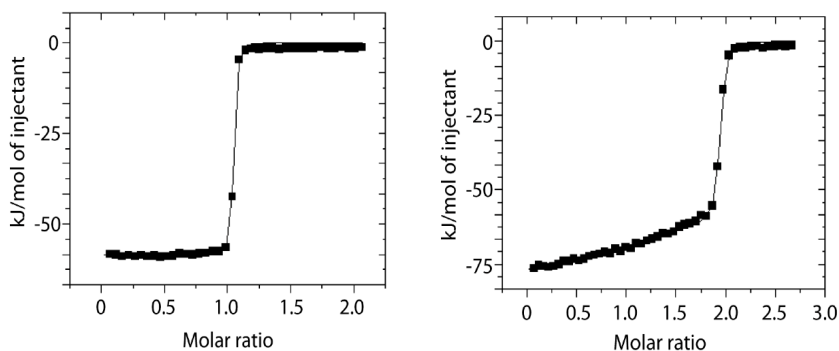
Fig. 4. Crystal structures of MUP (left panel) and HBP (right panel). Both MUP and HBP are displayed with their ligands bound: octanediol (purple), and histamine (pink). Ligands are represented as VDW spheres. For MUP, superimposed structures of apo (blue) and holo (dark cyan) protein are showed, in order to display a lack of major conformational changes associated with ligand binding. For HBP, the second (low affinity) binding site is showed and coloured yellow.

#### 4.1.2 Calorimetric studies of MUP

Given that the binding pocket of MUP is very hydrophobic, an entropy-driven binding signature might have been expected for ligand-MUP interactions. Surprisingly, global

thermodynamics data obtained for pyrazine ligands (Barratt *et al.* 2004) and alcohols (Barratt *et al.*, 2006) showed that binding is driven by favourable enthalpic contributions, rather than the classical hydrophobic effect. The only hydrogen bond that could be formed between a ligand and the protein binding site involved the hydroxyl group of tyrosine Y120. Barratt *et al.* (2004) reported that ITC measurements on the binding of isobutyl-methoxy-pyrazine (IBMP) to the Y120F (phenylalanine side chain lacks hydroxyl group) mutant showed slightly reduced enthalpy of binding compared to wild-type MUP, but the binding was nonetheless enthalpy-driven.

Binding of long-chain alcohols, such as n-octanol, n-nonanol, and 1,8 octan-diol was characterised by similar thermodynamic signature. Contrary to expectations, binding was enthalpy-driven (Barratt *et al.*, 2006). Each complex was characterised by a bridging water molecule between the hydroxyl group of Y120 and the hydroxyl group of ligand. The thermodynamic penalty to binding derived from the unfavourable desolvation of 1,8 octan-diol (+21 kJ/mol with respect to n-octanol, which came from an additional hydroxyl group facing a hydrophobic pocket) was partially offset by a favourable intrinsic contribution.



Ligand	Temp (K)	$\Delta H^\circ$ (kJ/mol)	$T\Delta S^\circ$ (kJ/mol)	$\Delta G^\circ$ (kJ/mol)	$K_d$ (nM)
HBP(D24R)	278	$-42.7^a \pm 0.9^b$	$6.5 \pm 1$	$-49.2 \pm 0.6$	$2.4 \pm 0.6$
	288	$-48.3 \pm 3.1$	$0.4 \pm 4.5$	$-48.8 \pm 1.1$	$2.9 \pm 1.3$
	298	$-58.3 \pm 1.2$	$-9.3 \pm 2.5$	$-49.1 \pm 1.4$	$2.5 \pm 1.4$
HBP(WT)	HA 298	$-70.9 \pm 7.1$	$-24.4 \pm 7.3$	$-46.5 \pm 1.3$	$7.0 \pm 3.6$
	LA 298	$-53.2 \pm 3.3$	$-12.0 \pm 3.6$	$-41.2 \pm 1.5$	$59.6 \pm 34.8$

Fig. 5. ITC data for obtained for HBP. Binding curves for HBP(D24R) mutant and wild-type BP are displayed in left and right panel, respectively. Bottom panel shows thermodynamic parameters for mutated and wild-type HBP, obtained from ITC measurements.



### 4.1.3 Calorimetric studies of HBP

Typical ITC isotherms for the binding of histamine to the HBP-D24R mutant (top panel) and wild-type HBP (middle panel), and the resulting thermodynamic parameters (bottom panel) are shown in Figure 5. Consistent with structural data, histamine bound to wild-type HBP with 2:1 stoichiometry (both binding sites occupied), while to mutant with 1:1 stoichiometry. In both cases, histamine bound with high affinity ( $K_d$  in nanomolar range), but the binding enthalpies and entropies were somewhat different, even though binding was largely enthalpy-driven in both cases. These differences in thermodynamic details are a manifestation of the enthalpy-entropy compensation, introduced in the previous sections of this chapter: wild-type HBP binds histamine with similar affinity than D24R mutant, but its enthalpic contribution is more favourable than that of D24R, at the expense of the entropic contribution, which is less favourable in wild-type HBP than in D24R mutant.

A major concern regarding the data shown in Figure 5 was the possibility of proton exchange (release or binding) during the binding event. In such case, the observed enthalpy change  $\Delta H_b$  would contain contributions from the ionisation of the buffer ( $\Delta H_{ion}$ ), as explained previously. To assess this effect, titrations between histamine and HBP-D24R mutant were performed in two different buffers, characterised by very different values of the ionisation enthalpy ( $\Delta H_{ion}$  47.45 kJ/mol and 3.6 kJ/mol, respectively). Thus, any contribution from proton exchange should be easily detectable on the basis of the differences in  $\Delta H_b$  for the histamine-HBP interaction. Obtained enthalpies (-58 and -61 kJ/mol, respectively) showed that there were no significant protonation effects associated with histamine-HBPD24R binding.

### 4.1.4 NMR relaxation measurements

In order to gain deeper insight into the entropic contribution to binding of ligands to HBP and MUP,  $^{15}\text{N}$  NMR relaxation measurements were employed to probe per-residue conformational entropies for backbone amides for the free (apo) protein and for the ligand-protein complexes. Backbone  $^{15}\text{N}$  longitudinal and transverse relaxation rates ( $R1 = 1/T1$  and  $R2 = 1/T2$ , respectively) were determined for the free protein and the complexes with ligands. Amide  $^{15}\text{N}$  and  $^1\text{H}$ - $^{15}\text{N}$  resonance assignments in the apo-HBP, apo-MUP, and the ligand-protein complexes were determined by use of conventional three-dimensional triple-resonance experiments.

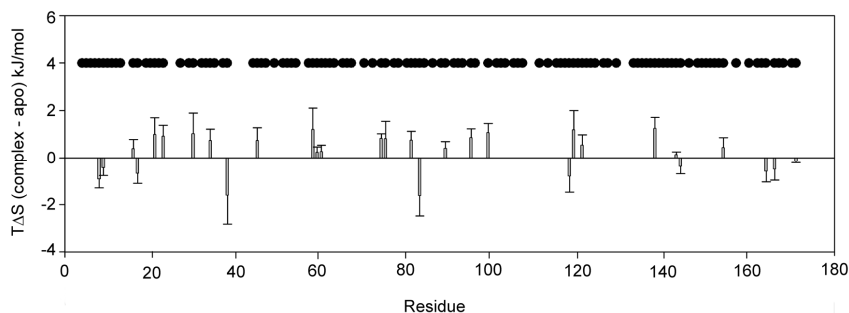


Fig. 6. Backbone amide entropy changes for HBP-D24R, quantified as differences between entropy of holo and apo protein for each protein residue. The plot shows histamine binding - induced changes in entropy assessed by  $^{15}\text{N}$  relaxation measurements.

For both HBP and MUP, both positive and negative changes in local backbone entropy were observed. Entropy changes were not restricted to the binding pocket but were dispersed over the protein (Figure 6). For HBP as well as MUP,  $T\Delta S$  summed over backbone amides was close to the error value, i.e., an overall change in backbone entropy that was not statistically different from zero (Syme *et al.*, 2010).

#### 4.1.5 Molecular dynamics (MD) simulations

MD simulations were used to examine the solvation of the binding pocket of MUP and HBP. The MUP binding pocket was found virtually devoid of water, even without any ligand bound (Barratt *et al.*, 2006). Even if water molecules were artificially “forced in” at the beginning of the MD simulation, they did not remain inside the pocket. This observation contributed to the explanation of the unexpected thermodynamic signature of the ligand binding to MUP, measured by ITC.

An average of five to six water molecules is observed in the binding pocket of HBP over the simulation time in the absence of ligand. Analysis of the diffusion and rotational correlation functions of these solvent molecules suggested that their dynamic behaviour was very similar to those of bulk water (Syme *et al.*, 2010), which suggests that the return of these water molecules to bulk solution on ligand binding would not offer any significant contribution to the binding entropy.

For HBP, side chain entropies were computed from the MD simulation. Moreover, as a test of the robustness of these simulations, backbone amide entropy changes were calculated from the MD simulation and compared to NMR data. The comparison was very favourable ( $16.4 \pm 1.0$  kJ/mol versus  $12.4 \pm 9.8$  kJ/mol), which raised confidence in the applied methodology. Contribution from the protein side chains was estimated at  $+17.4 \pm 1.8$  kJ/mol (Syme *et al.*, 2010). Taken together, these data strongly indicate an overall increase in entropy on ligand binding. This observation, although counter-intuitive, it is not without precedent in the literature (MacRaid *et al.*, 2007, Stoeckmann *et al.*, 2008).

MD simulations were also used to estimate the entropic contributions from the ligand. A contribution from the loss in vibrational degrees of freedom of histamine on binding to HBP was estimated using the Schlitter's method (Schlitter, 1993), leading to an unfavorable contribution of  $\approx -22 \pm 2.4$  kJ/mol.

In addition to this contribution, there was an assumption that internal degrees of freedom of the ligand are heavily constrained upon binding. The unfavorable contribution from the three relevant internal degrees of freedom of histamine amounted to  $\approx -12$  kJ/mol (Lundquist *et al.*, 2000). In addition, the entropic contribution from the loss of translational and rotational degrees of freedom of the ligand depends on the logarithm of the molecular mass, and on the basis of earlier work this represents an unfavorable contribution that can be estimated as  $\approx -25$  kJ/mol (Turnbull *et al.*, 2004).

#### 4.1.6 Solvation thermodynamics estimation

For MUP ligands, it was possible to measure their solvation free energies directly, using the water/vapor partitioning experiments (Shimokhina *et al.*, 2006). For histamine, experimental measurements could not be done due to non-volatility of histamine, and hence the ligand free solvation energies were calculated quantum-chemically, using the COSMO model (Klamt and Schüürmann, 1993). Prior to running calculations, the ligand was optimised using *ab initio* QM calculations (RFH/6-31G\* basis set). The optimised structure was

subjected to COSMO calculations at three different temperature settings (270, 300, and 330 K) in order to extract the enthalpic and entropic contributions to the free energy of solvation, using the finite-difference approach. The approximation used therein was based on the assumption that the heat capacity is constant over a certain range of temperatures near the target temperature,  $T$ . In the case of solute molecules solvated in water, this approximation usually holds near room temperature for temperature ranges (denoted as  $\Delta T$ ) as wide as 50 K. Using the finite-difference approximation, the entropy can be approximated at the target temperature as in equation (11):

$$\Delta S(T) = - \left( \frac{\Delta G(T + \Delta T) - \Delta G(T - \Delta T)}{2\Delta T} \right) \quad (11)$$

where  $\Delta S(T)$  denotes entropy at the target temperature,  $\Delta G(T)$  is the free energy of solvation energy, and  $\Delta T$  is the temperature difference. For the calculations presented here  $\Delta T$  was 30 K.

Obviously, it was very difficult to assess the accuracy of such calculations, but the experimental solvation thermodynamics of two related “fragments” of histamine have been reported (Cabani *et al.*, 1981). Comparison (Table 1) shows that the solvation free energies are reproduced very well, and the solvation entropies and enthalpies reasonably well (for *n*-propylamine) compared with experiment, which lends some confidence in the computed values for histamine.

Ligand	$\Delta G^0$ (kJ/mol)	$\Delta H^0$ (kJ/mol)	$T\Delta S^0$ (kJ/mol)
Histamine	-70.1	-137.2	-67.1
<i>n</i> -propylamine	-16.5	-65.5	-49
<i>n</i> -propylamine (exp)	-18.4	-55.8	-37.4
2-methyl imidazole	-43	-96.3	-53.3
2-methyl imidazole (exp)	-42.9	n.a.	n.a.

Table 1. Solvation parameters: free energy, enthalpy, and entropy at temperature 300 K for histamine and related ligands. All values are given in kJ/mol. Experimental values (exp) are taken from Cabani *et al.* (1981).

#### 4.1.7 Driving forces for ligand binding by MUP and HBP

Ligand binding to both HBP and MUP was strongly enthalpy-driven. The overall entropy of binding was the same within error for both MUP and HBP, yet the contributions from protein, ligand, and solvent are very different.

In the case of HBP, the dominant entropic contribution to binding arises from ligand desolvation, with a significant contribution from protein degrees of freedom. This contribution is favourable. However, the overall entropic contribution to binding is unfavourable, which indicates that the entropic contribution from desolvation of the protein binding pocket is strongly unfavourable.

In MUP, a favourable contribution to binding entropy is also derived from ligand desolvation, but is unable to overcome the unfavourable contribution from “freezing”

ligand degrees of freedom upon binding. The favourable entropic contribution from desolvation of the protein binding pocket that one would predict in a “classical” hydrophobic interaction is absent in MUP, since the occluded binding pocket is substantially desolvated prior to binding. This phenomenon has subsequently been observed in other proteins, such as streptavidin and HIV-protease receptors (Young *et al.*, 2007).

As suggested by MD simulations, there are around 6 water molecules (on average) occupying the binding site of HBP prior to ligand binding. As already stated, no significant change in entropic contribution should arise from displacing these waters upon ligand binding, since their dynamic behaviour inside the pocket is similar to the behaviour of the bulk water. However, four water molecules are sequestered in the binding pocket in the histamine-HBP complex. These waters are significantly more ordered than the bulk water (Syme *et al.*, 2010), which contributes negatively to the entropic term of binding free energy. This unfavourable entropic contribution can be estimated as about  $-30$  to  $-40$  kJ/mol, which is qualitatively consistent with the observed sign of  $\Delta C_p$  on binding.

In conclusion, in the case of HBP, we found favourable entropic contributions to binding from desolvation of the ligand. However, the overall entropy of binding was unfavourable due to a dominant unfavourable contribution arising from the loss of ligand degrees of freedom, together with the sequestration of solvent water molecules into the binding pocket in the complex. This can be contrasted with MUP, where desolvation of the protein binding pocket made a minor contribution to the overall entropy of binding given that the pocket is substantially desolvated prior to binding.

## 4.2 An integrated picture of galactose binding to Arabinose Binding Protein (ABP)

The arabinose binding protein is found in the periplasm of Gram-negative bacteria. It belongs to the family of proteins, which exert their biological function as components of osmotic shock-sensitive transport systems for sugars and amino acids (Vyas *et al.*, 1991). Besides its biological context, ABP is a very well-defined model system for structure-activity relationships in the hydrophilic ligand binding systems. As demonstrated by ITC, ABP interacts with its natural ligands, namely L-arabinose and D-galactose, and their deoxy derivatives (Daranas *et al.*, 2004). The interactions are enthalpy-driven. The galactose-ABP interactions served as a model for interactions between hydrophilic ligands and hydrophilic binding pockets. Such choice of the model system was made mainly because of the large unfavourable entropic contribution to binding, whose origin was difficult to understand in the framework of the current ligand-protein interaction paradigms. Such a large entropy change upon ligand binding is frequently observed in proteins interacting with carbohydrate ligands, and it contributes to the fact that these interactions are notoriously challenging for predictions and design. In order to address those issues, we employed a combination of solution NMR and MD simulations.

### 4.2.1 NMR measurements of ABP

An essential prerequisite to NMR studies of protein dynamics was the establishment of the  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  resonances that contributed to the investigated spectra. These assignments have been obtained for the complex of ABP with the ligand D-galactose using conventional triple-resonance assignment strategies (Daranas *et al.*, 2004). Assignments for the unbound (apo) ABP were determined from these results, using an approach that combined conventional triple-resonance assignment strategies and  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single

quantum coherence (HSQC) titrations of ABP with 1-deoxy-galactose, which is a fast-exchanging ligand (MacRaild *et al.*, 2007).

From these assignments, a comparison was made of the chemical shifts of the backbone amide resonances of ABP in the unbound state and in the complex (Figure 7). Large chemical shift changes were observed in the binding site area and in the region linking the two domains of ABP. This suggested that ligand binding might be associated with a substantial conformational change in ABP (Figure 8). Such conformational change (domain reorientation) upon binding is observed in other members of the periplasmic-binding protein family, and have been proposed for ABP on the basis of the results of small-angle X-ray scattering and theoretical studies (Mao *et al.*, 1982, Newcomer *et al.*, 1981). Smaller changes in chemical shift were also observed at sites distal to the binding site, which suggests that small conformational changes, resulting from protein dynamic behaviour, occur upon the binding event.

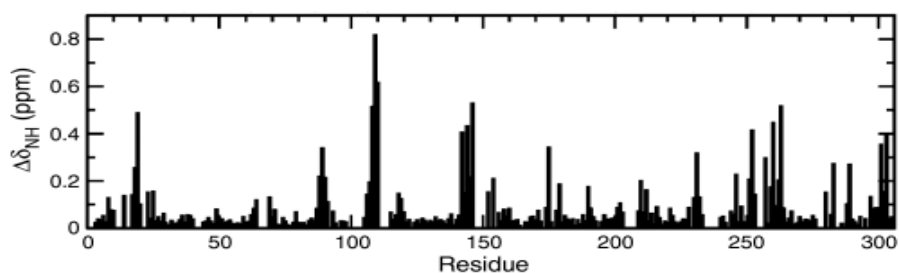


Fig. 7. Chemical shifts for galactose binding to ABP. Changes in backbone amide chemical shifts are plotted against protein residue number.

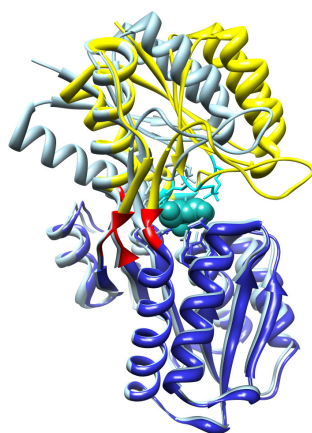


Fig. 8. Conformational changes between apo-ABP (light blue) and galactose-ABP complex. The apo protein is much more 'open' than the complex. Large conformational changes are observed in the hinge region (coloured red), and the reorientation of N-domain (yellow) and C-domain (dark blue) towards each other is quite pronounced. Bound galactose is coloured cyan and displayed as VDW spheres. Side chains of several residues involved in direct interactions with ABP are showed and coloured dark cyan.

Full sets of  $^{15}\text{N}$  relaxation measurements were made for ABP in its apo form and in complex with D-galactose, L-arabinose and D-fucose. The analysis of these measurements, assessed by means of Lipari-Szabo model-free approach (1982), allowed for extraction of the information on the extent of 'fast' (ps-ns time scale) motion of the protein backbone and  $^{15}\text{N}$ -containing side chains. Differences in Lipari-Szabo generalised order parameters ( $S^2$ ) between apo- and holo-ABP and were interpreted in terms of changes in dynamics accompanying the binding event. Thus, observed dynamic changes could be related to binding thermodynamics by means of the relationship between changes in order parameters derived from NMR relaxation and changes in conformational entropy. Because  $S^2$  parameters are specific to individual bond vectors, the described approach offered an unprecedented degree of structural resolution in thermodynamic analysis of protein function.

Surprisingly, generalised order parameters for apo ABP were, in general, larger than for the ABP-galactose complex. This suggests that 'fast' (pico- to nanosecond time scale) motions are more extensive in the ligand-protein complex than in the unbound protein.

#### 4.2.2 Molecular dynamics (MD) study

To confirm and further explore NMR relaxation result, we performed MD simulations of ABP in complex with galactose and its unbound state. In order to make a direct comparison between the observed order parameters and the simulations, backbone amide order parameters were calculated from the MD trajectory (MacRaild et al., 2007). A very good agreement between measured and calculated order parameters was observed, albeit with a small tendency for the simulation to underestimate the experimental values. The calculated changes in order parameter upon ligand binding reproduced the changes measured by NMR excellently, showing an approximately uniform decrease in order parameter (and, hence, increase of dynamics) upon galactose binding across the protein.

To understand the thermodynamic implications of the observed changes in dynamics, we employed the relationship between Lipari-Szabo order parameters ( $S^2$ ) and conformational entropy derived by Yang and Kay (1996).

In addition to generalised order parameters, we observed significant increase in backbone dynamics in the complex as compared with the apo protein by several other measures. RMS deviations from the average structure were significantly larger for the galactose-ABP complex than for apo-ABP. Fluctuations of backbone heavy-atom positions across the trajectory were generally larger in the complex than in the apo protein.

As well as validating the experimental results, the MD simulations revealed details of dynamic changes in regions that could not be measured experimentally. Importantly, the simulations revealed complex changes in the dynamics of the ABP binding site: counter-intuitively, several residues in the binding site showed increase in flexibility upon binding, which was consistent with the trend seen throughout the rest of the protein. Other binding site residues displayed a decrease in flexibility, more in keeping with the intuitive expectation that ligand binding will reduce the conformational freedom of binding site residues (Figure 9).

The total entropy alters due to changes in the pico- to nanosecond motion upon galactose binding, estimated from MD simulations and NMR results, was  $610(\pm 120)$  J/mol K, which gives  $T\Delta S = 188(\pm 37)$  kJ/mol at 308 K. Clearly, this latter value is overestimated, as the

assumption of un-correlated motion is not likely to hold for all residues of the protein. It is evident, however, that the entropy change associated with changes in 'fast' dynamics contributes favourably to the free energy of binding. As the result, it allows the reduction of the unfavourable entropic contribution associated with ligand binding.

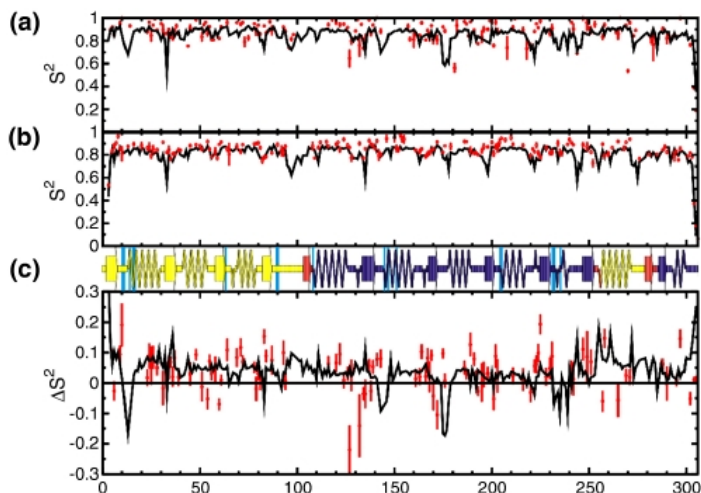


Fig. 9. Backbone amide order parameters for apo-ABP and galactose-ABP complex.  $S^2$  order parameters were obtained by NMR relaxation measurements (red dots) and MD simulations (black lines). Data obtained for apo-ABP are showed in panel (a), while results for galactose-ABP complex are displayed in panel (b). Panel (c) shows changes in order parameters induced by galactose binding to ABP, with protein secondary structure elements displayed above. The N-domain of ABP is coloured is yellow, the C-domain is coloured blue, and the hinge region is coloured red. Residues interacting with the ligand are coloured cyan. This colouring scheme is consistent with colouring in Figure 10. (figure taken from MacRaild *et al.*, 2007)

#### 4.2.3 Origins of entropic costs of binding

We investigated the origin of the large and unfavourable entropic contribution to the binding free energy of galactose-ABP, observed by ITC (Daranas *et al.*, 2004), in terms of the different contributions.

It is clear that formation of a ligand-protein complex will involve the loss of entropy associated with constraining the translational and rotational degrees of freedom of one binding partner with respect to the other. The magnitude of these unfavourable contributions to the ligand-protein interaction can be approximated. In the case of galactose-ABP interactions, we took an estimate of the loss of ligand translational and rotational entropy from the work by Turnbull *et al* (2004) and Lundquist and his coworkers (Lundquist and Toone, 2002), which gave at approximately 25 kJ/mol for the free energy penalty.

It is assumed that the bound ligand will experience a loss of entropy reflecting the loss of conformational flexibility of the ligand in solution. On the assumption that conformational degrees of freedom are substantially restricted upon ligand-protein binding, the entropic

penalty arising from loss in degrees of freedom of the galactose hydroxyl rotors alone is likely to be ~30 kJ/mol (Lundquist and Toone, 2002). Another contribution arises from the solvation effects. As calculated by means of QM/COSMO approach, desolvation energy of galactose is +87.6 kJ/mol (at 300 K), which is a significant unfavourable contribution (Bronowska and Homans, unpublished data).

It is known that the ABP binding site contains a significant number of tightly bound water molecules, which maintain the structure of the binding site and play a role in governing the specificity of ligand binding (Quiocho, 1993). Examination of the structures of ABP in complex with its ligands (galactose, fucose, and arabinose) revealed some 15 crystallographically resolved and structurally conserved water molecules within the binding site (MacRaild *et al.*, 2007). Dunitz (1994) estimated the maximal entropic cost of confinement of a single water molecule to 8 kJ/mol. Even if the cost of confinement of the water molecules in the binding site of ABP is lower than this maximal value, the overall cost of confining these water molecules in the binding site will be vast.

All these amount to unfavourable entropic contribution of galactose-ABP binding. However, the entropy of galactose-ABP interactions is much lower than the sum of these contributions: as measured by ITC the  $T\Delta S^0$  amounts to -61 kJ/mol at 308 K (Daranas *et al.*, 2004). Observed discrepancy can be explained in terms of favourable contribution of protein dynamics to the entropy of galactose-ABP interactions, which was observed by NMR measurements and MD simulations.

### 4.3 Enthalpy-entropy compensation revisited: bovine carbonic anhydrase II

In studies on the thermodynamics of ligand-protein interactions, it is usually assumed that the bound ligand is fixed in the binding site. However, there is little direct experimental evidence for this assumption, and in the case of binding of *p*-substituted benzenesulfonamide inhibitors to bovine carbonic anhydrase II (BCA II), the observed thermodynamic binding signature assessed by ITC measurements leads indirectly to the conclusion that the bound ligands retain a considerable degree of flexibility (Krishnamurthy *et al.*, 2006).

BCAII and its binding to a large panel of ligands was reported in literature as a classic example of enthalpy-entropy compensation. Whitesides and his coworkers studied these interactions for a series of *p*-(glycine)*n*-substituted benzenesulfonamides (where *n* = 1-5) and found almost perfect enthalpy-entropy compensation across the series: as demonstrated by ITC measurements, the enthalpy of binding became less favorable and the entropy more favorable with increasing chain length. Changes in heat capacity were independent of chain length, which indicated that the observed changes in binding thermodynamic signatures across the series cannot be explained on the basis of the classical hydrophobic effect. In addition, strong evidence exists that these thermodynamic signatures are not driven by solvent effects (Syme *et al.*, 2010). To explain the observed data, a model was proposed, which assumed the increased mobility of ligand upon the chain length growth. Such increased flexibility of the bound ligand (favourable entropic contribution) would be compensated by a decreased number of direct ligand-protein contacts (unfavourable enthalpic contribution).

Such an increase in mobility of the ligand can be readily probed by <sup>15</sup>N NMR relaxation measurements and computational studies. Thus, we investigated these using a combination of the solution NMR and MD simulations. Two series of ligands were studied: The first



(series 1) consisted of six ligands with different chain lengths ( $n = 1-6$ ), isotopically  $^{15}\text{N}$ -labeled at the terminal amide, whereas the second (series 2) comprised six ligands with the same chain length ( $n = 6$ ), but isotopically  $^{15}\text{N}$ -labeled at a single amide at each position  $n$ . The ligands bound at the BCAII binding site are shown in Figure 10.

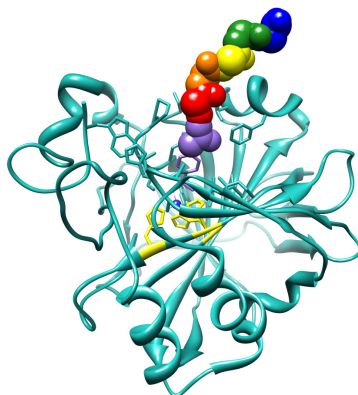


Fig. 10. The binding of investigated benzenesulfonamides to BCAII. The protein backbone is coloured cyan, with three histidines (yellow) coordinating zinc atom (dark blue dot) displayed. The ligand is showed, with glycine side chains represented as VDW spheres and coloured as follows:  $n=1$  – purple,  $n=2$  – red,  $n=3$  – orange,  $n=4$  – bright yellow,  $n=5$  – dark green, and  $n=6$  – blue.

Contrary to expectation, we found that the observed thermodynamic binding signature could not be explained by residual ligand dynamics in the bound state, but rather results from the indirect influence of ligand chain length on protein dynamics.

Chemical shift changes on ligand binding were monitored by simple one-dimensional  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectra. It was possible, since each ligand was selectively  $^{15}\text{N}$ -labeled. The results (Figure 11) indicated that residues proximal to the aromatic moiety of ligand show substantial changes in chemical shift, whereas residues more distant to the aromatic moiety show much smaller changes. The former is indicative of interactions with the protein, while the latter suggests less substantive interactions with the protein by these more distant residues.

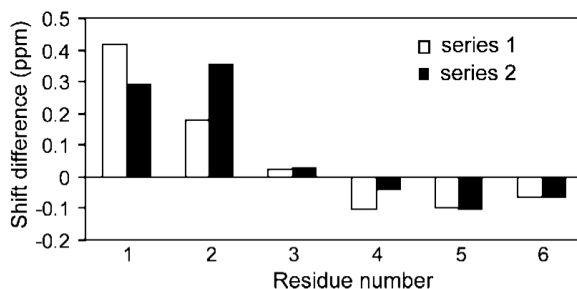


Fig. 11. Observed chemical shift differences between free ligands and ligand-BCAII complexes, for both series of ligands, plotted against the number of glycine residues ( $n$  from 1 to 6) in the side chain (Stoekmann *et al.*, 2008).

In order to obtain a more detailed picture of ligand dynamics in the bound state,  $^{15}\text{N}$  NMR relaxation data ( $R1$ ,  $R2$ , NOE) were measured for each series of ligands. The results obtained showed that the ligand chain became more dynamic as  $n$  increased. However, there were substantial differences between series 1 and 2 ligands: series 1 ligands were much more dynamic for small  $n$  values, while in case of series 2 ligands, the three residues nearest the aromatic ring adopted slow dynamic motions, whereas the three residues distal to the aromatic ring adopted faster dynamics similar to series 1 ligands.

The relaxation data were analysed using the Lipari-Szabo model-free approach (1982) and the 'fast' dynamics was quantified by means of generalised order parameters. The entropic contributions arising from these 'fast' motions were assessed from  $S^2$  parameters using the relation derived by Yang and Kay (1996).

The order parameters of all ligand Gly residues in both series were much smaller than those of backbone residues within protein, indicating a comparatively high mobility of the ligand chain. The first two residues of the ligand were relatively immobile adjacent to the aromatic ring and are engaged in direct interactions with the protein, whereas the last four Gly residues were significantly more mobile, with  $S^2$  values indicating motions being unrestricted by BCaII protein. From  $S^2$  values, the entropic contribution to ligand-BCaII binding arising from each Gly residue, could be obtained. These data are shown in Figure 12.

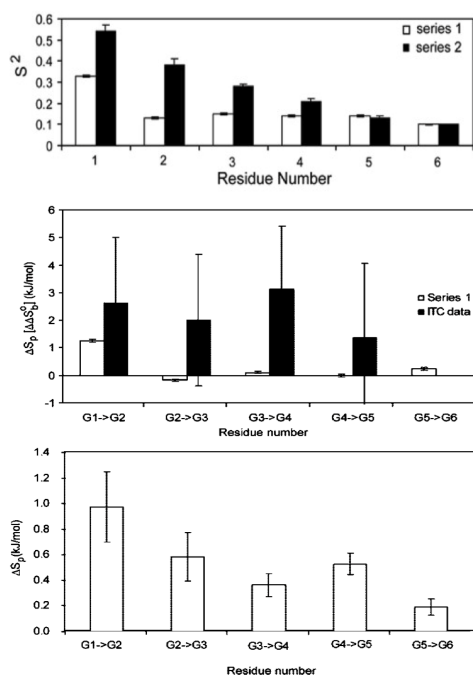


Fig. 12. Top panel: order parameters for backbone amide (N-H) bond vectors for ligand series 1 (white) and 2 (black), obtained from NMR relaxation measurements. Middle panel: Derived entropy differences for series 1 ligands (white) compared to entropies of binding obtained from ITC measurements (black) (Krishnamurthy *et al.*, 2006). Bottom panel: Entropy differences for series 2 ligands (per N-H bond vector), obtained by NMR.

Comparing these results with ITC data by Krishnamurthy *et al.* (2006), it is clear that a poor correlation exists between the change in ligand conformational entropy determined from NMR relaxation studies and the entropies of binding derived from ITC (Figure 14, middle panel). It indicates that a model based on increased dynamics of the ligand in the bound state is not a plausible explanation for the observed thermodynamic binding data. This is not entirely unexpected since the ITC values are global parameters, which include contributions not only from the ligand, but from protein and solvent as well. However, the role of solvation is unlikely to be the driving one in the case of ligand-BCAII binding – for three reasons. First,  $\Delta C_p$  values for the interaction determined by ITC are independent of Gly chain length (Stoeckmann *et al.*, 2008). Second, these values are fairly small: around 80 J/mol/K. Finally, ligands are not fully desolvated upon the binding event: more distal residues extend beyond the binding pocket and they interact with water molecules. The observed increase in entropy with respect to the ligand chain length is approximately linear, which argues against a significant solvation contribution.

It was hoped that assessment of the protein contribution would shed light on the observed binding signature. To achieve this, MD simulations of both series of ligands in complexes with BCAII were performed (Stoeckmann *et al.*, 2008). In order to validate the methodology, generalised order parameters for ligand amide vectors were calculated from the trajectory and compared to NMR data. These MD trajectories were then used to probe the influence of ligand binding on protein dynamics. Specifically,  $S^2$  values for backbone amide bond vectors, side chain terminal heavy-atom bond vectors, and corresponding conformational entropies were calculated for each complex with series 1 ligands.

The results obtained showed that the aromatic moiety became correspondingly more rigid with respect to series 1 ligand chain length. This was consistent with the NMR data showing that addition of successive glycine residues decreased the dynamics of the preceding units. Moreover, we observed the trend of increased dynamics of protein residue side chains with respect to ligand chain length (Table 2). This counter-intuitive observation that ligand binding increases protein dynamics has been observed in a number of ligand-protein systems, including ABP, which was described in the previous section of this chapter.

Residues	Gly2-Gly1	Gly3-Gly2	Gly4-Gly3	Gly5-Gly4	Gly6-Gly5
Biding site	4.37 ± 1.1	5.28 ± 1.2	4.33 ± 1.0	3.11 ± 1.0	6.04 ± 1.3
Whole protein	14.9 ± 1.7	4.6 ± 1.8	5.5 ± 2.2	9.9 ± 2.4	8.4 ± 2.5

Table 2. Differences in per-residue entropies quantified as TDS (in kJ/mol at temperature 300 K) for residues in the binding pocket of BCAII as well as for the whole BCAII protein. Displayed differences are result of changing side chain length of the ligand (Gly<sub>n</sub>-Gly<sub>n-1</sub>).

Summarising, our results suggest that the enthalpy-entropy compensation observed for binding of ArGly<sub>n</sub>O- ligands to BCA II derives principally from an increase in protein dynamics, rather than ligand dynamics, with respect to the ligand chain length. Krishnamurthy and his coworkers showed that enthalpy-entropy compensation was observed for a range of BCAII ligands, whose structurally distinct chain types gave similar thermodynamic signatures (Krishnamurthy *et al.*, 2006). This suggests that a common process is underway that is unlikely to be related to specific interactions between the chain

and the protein. In our study, we demonstrated an increase in protein dynamics upon binding longer-chained ligands. This observation provides an explanation for the enthalpy-entropy compensation across these structurally distinct ligands.

## 5. Conclusions

The notion of the binding event being the result of shape complementarity between ligand and protein binding site (key-and-lock model) has been a paradigm in the description of binding events and molecular recognition phenomena for a long time. The recent discovery of the important role played by protein dynamics and solvent effects, as well as the enthalpy-entropy compensation phenomenon, challenged this concept, and demanded the thorough examination of entropic contributions and solvent effects. Assessment of all these contributions to the thermodynamics of ligand-protein binding is a challenging task. Although understanding the role of each contribution and methods allowing for a complete dissection of thermodynamic contributions are tasks far from being completed, significant progress has been made in recent years. For instance, development of high-resolution heteronuclear NMR methods allowed for assessment of the contribution from protein degrees of freedom to the intrinsic entropy of binding. The usefulness of such approach has been demonstrated in the course of this chapter on several ligand-protein examples. In addition, progresses in the development of MD-related methodologies and advanced force fields enabled the application of the NMR-derived formalism on relevant time scales and the assessment of the intrinsic entropic contributions solely using computational methods. Development of QM methods allows the study of larger and larger systems, while advances in ITC calorimetry allow the use of very small amounts of reagents for a single experiment.

Despite this progress, much remains to be done. The enthalpy-entropy compensation phenomenon seems to be widespread among ligand-protein systems. It seems universal: binding restricts motions, while motions oppose tight confinement. However, our current knowledge about intrinsic protein dynamics is still insufficient to allow us to predict this phenomenon and hence to exploit it for the purposes of rational molecular design. Another challenge lies within the quantification of solvation contributions. There seem to be conflicting data regarding the contributions from confined water molecules. Their influence on binding can be favourable or unfavourable, enthalpy- or entropy- driven. Bound water molecules can be released upon ligand binding or – on the contrary – bind tighter (Poornima CS and Dean, 1995a-c). Their presence can make the protein structure more rigid (Mao *et al.*, 2000), or more flexible (Fischer and Verma, 1999). Finally, protein binding sites can be fully solvated prior to binding, or fully desolvated (Barratt *et al.*, 2006, Syme *et al.*, 2010). The only common feature that seems to exist is that the contribution of the solvation effects to the ligand-protein binding thermodynamics can be – and often is – significant.

Last but not least, intrinsic entropic contributions are notoriously difficult to quantify. A handful of experimental and theoretical methods can be employed to quantify these contributions, as have been described. However, all of these methods have their limitations, and one should be aware of these and of the assumptions that are being made. Theoretical results should be treated with caution, experimental data likewise, as they are based on many approximations and heavily dependent on the conditions applied. Care must be taken not to over-extrapolate data, and not fall the victim to confirmation bias.

Fundamentally, in order to predict free energy of binding accurately, it would be necessary to go beyond predicting a single 'dominant' conformation of the ligand-protein complex. It

should be emphasised that the overall shape of the free energy landscape controls the binding free energy. This shape is affected by the depth and width of the local minima, and the height and breadth of the energy barriers. The factors that shape that landscape include intrinsic entropic contributions of both interacting partners, ligand poses, protein conformations, solvent effects, and protonation states. Computational and experimental approaches combined together can provide insight into this crucial but otherwise hidden landscape, which is pivotal not only to understand the origin of each contribution and its role in the binding event, but which can allow a truly rational molecular design.

## 6. Acknowledgements

I would like to thank my collaborators and coauthors of my publications: Steve Homans, Chris MacRaild, Arnout Kalverda, Liz Barratt, Bruce Turnbull, Antonio Hernandez Daranas, Neil Syme, Cairiona Dennis, Dave Evans, Natalia Shimokhina, Pavel Hobza, Jindra Fanfrlik, Honza Rezac, Honza Konvalinka, Jiri Vondrasek, Jiri Cerny, Henning Stoeckmann, Stuart Warriner, Rebecca Wade, and Frauke Gräter. I also would like to thank for the financial support: BBSRC (United Kingdom), DAAD (Germany), DFG (Germany), Heidelberg Institute for Theoretical Sciences, and University of Heidelberg, Germany.

## 7. References

- Alves, I.D.; Park, C.K.; Hruby, V.J. (2005). Plasmon resonance methods in GPCR signaling and other membrane events. *Curr Protein Pept Sci.* 6(4), pp. 293-312.
- Axilrod, B.M.; & Teller, E. (1943). Interaction of the van der Waals' type between three atoms, *J Chem Phys.* 1943;11, pp.299-300. doi: 10.1063/1.1723844.
- Baldus, M. (2006). Molecular interactions investigated by multi-dimensional solid-state NMR, *Curr Opin Struct Biol.* 2006 Oct;16(5), pp. 618-23.
- Barratt, E.; Bingham, R.J.; Warner, D.J.; Laughton, C.A.; Phillips, S.E.; Homans, S.W. (2005). Van der Waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water, *J Am Chem Soc.* 2005 Aug 24;127(33), pp. 11827-34.
- Barratt, E.; Bronowska, A.K.; Vondrásek, J.; Cerný, J.; Bingham, R.; Phillips, S.; Homans, S.W. (2006). Thermodynamic penalty arising from burial of a ligand polar group within a hydrophobic pocket of a protein receptor., *J Mol Biol.* 2006 Oct 6;362(5), pp. 994-1003.
- Bayly, C.I.; Cieplak, P.; Cornell, W.D.; Kollman, P.A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges the RESP model. *J. Phys. Chem.* 1993;97, pp. 10269-10280.
- Benjahad, A.; Guillemont, J.; Andries, K.; Nguyen, C.H.; Grierson, D.S. (2003). 3-Iodo-4-phenoxy pyridinones (IOPY's), a new family of highly potent non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* 2003, 13, pp. 4309-4312.
- Best, R.B.; & Vendruscolo, M. (2004). Determination of protein structures consistent with NMR order parameters, *J Am Chem Soc.* 2004 Jul 7;126(26), pp. 8090-1.
- Bissantz, C.; Kuhn, B.; & Stahl, M. (2010). A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* 2010;53(14), pp. 5061-5084. doi: 10.1021/jm100112j.

- Boehm, H.J.; & Klebe, G. (1996). What can we learn from molecular recognition in protein–ligand complexes for the design of new drugs?. *Angew. Chem., Int. Ed.* 1996, 35, pp. 2588–2614.
- Boehr, D.D.; Dyson, H.J.; & Wright, P.E. (2009). An NMR perspective on enzyme dynamics. *Chem Rev*, 2006; 106, pp. 3055–3079.
- Brown, A. (2009). Analysis of Cooperativity by Isothermal Titration Calorimetry, *Int. J. Mol. Sci.* 2009, 10, pp. 3457 – 3477. doi:10.3390/ijms10083457
- Burz, D.S.; Dutta, K.; Cowburn, D.; Shekhtman, A. (2006). In-cell NMR for protein-protein interactions (STINT-NMR), *Nat Protoc.* 2006;1(1), pp. 146–52.
- Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. J. (1981) *Solution Chem.* 1981, 10, pp. 563–595.
- Chakrabarti, P.; & Bhattacharyya, R. (2007). Geometry of nonbonded interactions involving planar groups in proteins. *Prog. Biophys. Mol. Biol.* 2007, 95, pp. 83–137.
- Cieplak, P.; Cornell, W.D.; Bayly, C.; Kollman, P.A. (1995). Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* 1995;16, pp. 1357–1377.
- Clark, T.; Hennemann, M.; Murray, J.S.; Politzer, P. (2007). Halogen bonding: the sigma hole. *J. Mol. Model.* 2007, 13, pp. 291–296.
- Connelly, P.R.; Thomson, J.A.; Fitzgibbon, M.J.; Bruzzese, F.J. (1993). Probing hydration contributions to the thermodynamics of ligand binding by proteins. Enthalpy and heat capacity changes of tacrolimus and rapamycin binding to FK506 binding protein in D<sub>2</sub>O and H<sub>2</sub>O, *Biochemistry.* 1993 Jun 1;32(21), pp. 5583–90.
- Connelly, P.R.; Aldape, R.A.; Bruzzese, F.J.; Chambers, S.P.; Fitzgibbon, M.J.; Fleming, M.A.; Itoh, S.; Livingston, D.J.; Navia, M.A.; Thomson, J.A. et al. (1994). Enthalpy of hydrogen bond formation in a protein-ligand binding reaction, *Proc Natl Acad Sci U S A.* 1994 Mar 1;91(5), pp. 1964–8.
- Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Kollman, P.A. (1993). Application of RESP charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation. *J. Am. Chem. Soc.* 1993;115, pp. 9620–9631.
- Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M. Jr.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc.* 117, pp. 5179–5197.
- Daranas, A.H.; Shimizu, H.; & Homans, S.W. (2004). Thermodynamics of binding of D-galactose and deoxy derivatives thereof to the L-arabinose-binding protein, *J Am Chem Soc.* 2004 Sep 29;126(38), pp. 11870–6.
- Davis, A.M.; & Teague, S.J. (1999). Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew. Chem., Int. Ed.* 1999, 38, pp. 736–749.
- Denisov, V.P.; Venu, K.; Peters, J.; Hoerlein, H.D.; Halle, B. (1997) Orientational disorder and entropy of water in protein cavities. *J. Phys. Chem. B* 1997, 101, pp. 9380–9389.
- Desiraju, G.R. (2002) Hydrogen bridges in crystal engineering: interactions without borders. *Acc. Chem. Res.* 2002, 35, pp. 565–573.
- Diehl, C.; Engström, O.; Delaine, T.; Håkansson, M.; Genheden, S.; Modig, K.; Leffler, H.; Ryde, U.; Nilsson, U.J.; Akke, M. (2010). Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3, *J Am Chem Soc.* 2010 Oct 20;132(41), pp. 14577–89.
- Dill, K.A. (1997). Additivity principles in biochemistry, *J. Biol. Chem.*, 272, pp. 701–704.

- Dixon, S.L. & K.M. Merz, Jr. (1996). Semiempirical Molecular Orbital Calculations with Linear System Size Scaling. *J. Chem. Phys.*, 1996. 104: pp. 6643-6649.
- Dixon, S.L. and K.M. Merz, Jr. (1997). Fast, Accurate Semiempirical Molecular Orbital Calculations for Macromolecules. *J. Chem. Phys.*, 1997. 107(3): pp. 879-893.
- Donchev, A.G. (2006). Many-body effects of dispersion interaction. *J Chem Phys.* 2006, pp. 125:074713. doi: 10.1063/1.2337283.
- Dunitz, J.D. (1994). The entropic cost of bound water in crystals and biomolecules. *Science* 1994, 264, p. 670.
- Dunitz, J.D. (1995). Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions., *Chem Biol*, 1995;2(11), pp. 709-12.
- Earl, D.J.; & Deem, M.W. (2005). Parallel tempering: theory, applications, and new perspectives, *Phys Chem Chem Phys.* 2005 Dec 7;7(23), pp. 3910-6
- Ebbinghaus, S.; Kim, S.J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D.M.; Havenith, M. (2007). An extended dynamical hydration shell around proteins. *Proc Natl Acad Sci U S A.* 2007 Dec 26;104(52), pp- 20749-52.
- Emsley, J. (1980). Very Strong Hydrogen Bonds. *Chem Soc Rev*, 1980;9(1), pp. 91-124.
- Evans, D.A.; & Bronowska, A.K. (2010). Implications of fast-time scale dynamics of human DNA/RNA cytosine methyltransferases (DNMTs) for protein function, *Theoretica Chimica Acta*, Volume 125, Numbers 3-6, pp. 407- 418, doi: 10.1007/s00214-009-0681-2
- Fanfrlík, J.; Bronowska, A.K.; Rezáč, J.; Prenosil, O.; Konvalinka, J.; Hobza P. (2010). JA reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H- bonding: HIV-1 protease with 22 ligands, *Phys Chem B.* 2010 Oct 7;114(39), pp. 12666-78.
- Fernandez, C.; & Wider, G. (2003). TROSY in NMR studies of the structure and function of large biological macromolecules, *Curr Opin Struct Biol.* 2003 Oct;13(5), pp. 570-80.
- Finkelstein, A.V. (2007). Average and extreme multi-atom Van der Waals interactions: strong coupling of multi-atom Van der Waals interactions with covalent bonding. *Chem Cent J.* 2007, pp. 1:21. doi: 10.1186/1752-153X-1-21.
- Finney, J.L.; & Soper, A.K.(1994). Solvent structure and perturbations in solutions of chemical and biological importance. *Chem. Soc. Rev.* 1994, 23, pp. 1-10.
- Fischer, S.; & Verma, C.S. (1999). Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 1999, 96, pp. 9613-9615.
- Fischer, F.R; Wood, P.A.; Allen, F.H.; Diederich, F. (2008). Orthogonal dipolar interactions between amide carbonyl groups, *Proc Natl Acad Sci U S A.* 2008 Nov 11;105(45), pp. 17290-4.
- Foloppe, N.; & Hubbard, R. (2006). Towards predictive ligand design with free-energy based computational methods?, *Curr Med Chem.* 2006;13(29), pp. 3583-608.
- Ford, D.M. (2005). Enthalpy-entropy compensation is not a general feature of weak association. *J. Am. Chem. Soc.* 2005, 127, pp. 16167-16170.
- Frauenfelder, H.; & Mezei, F. (2010). Neutron scattering and protein dynamics, *Acta Crystallogr D Biol Crystallogr.* 2010 Nov;66(Pt 11), pp. 1229-31.
- Gilson, M.K.; & Zhou, H.X. (2007). Calculation of protein-ligand binding affinities, *Annu Rev Biophys Biomol Struct.* 2007;36, pp. 21-42.

- Goodman, J.L.; Pagel, M.D.; & Stone, M.J. (2000). Relationships between protein structure and dynamics from a database of NMR-derived backbone order parameters, *J Mol Biol.* Jan 28;295(4), pp. 963-78.
- Graf, C.; Stankiewicz, M.; Kramer, G.; Mayer, M.P. (2009). Spatially and kinetically resolved changes in the conformational dynamics of the Hsp90 chaperone machine, *EMBO J.* 2009 Mar 4;28(5), pp. 602-13.
- Grover, J.R.; Walters, E.A.; & Hui, E.T. (1987). Dissociation energies of the benzene dimer and dimer cation. *J. Phys. Chem.* 1987, 91, pp. 3233–3237.
- Harman, J.G. (2001). Allosteric regulation of the cAMP receptor protein, *Biochim Biophys Acta.* 2001 May 5;1547(1), pp. 1-17.
- Heugen, U.; Schwaab, G.; Bründermann, E.; Heyden, M.; Yu, X.; Leitner, D.M.; Havenith, M. (2006). Solute-induced retardation of water dynamics probed directly by terahertz spectroscopy, *Proc Natl Acad Sci U S A.* 2006 Aug 15, 103(33), pp. 12301-6.
- Hobza, P.; Selzle, H.L.; & Schlag, E.W. (1996). Potential Energy Surface for the Benzene Dimer. Results of ab initio CCSD(T) Calculations Show Two Nearly Isoenergetic Structures: T-Shaped and Parallel Displaced, *J Phys Chem.* 1996;100(48), pp. 18790-94.
- Homans, S.W. (2005). Probing the binding entropy of ligand-protein interactions by NMR, *ChemBiochem.* 2005 Sep;6(9), pp. 1585-91.
- Homans, S.W. (2007). Dynamics and thermodynamics of ligand-protein interactions, *Top Curr Chem* 2007, 272, pp. 51-82.
- Iltzsch, M.H.; Uber, S.S.; Tankersley, K.O.; el Kouni, M.H. (1995). Structure–activity relationship for the binding of nucleoside ligands to adenosine kinase from *Toxoplasma gondii*. *Biochem. Pharmacol.* 1995, 49, pp. 1501–1512.
- Jelesarov, I.; & Bosshard, H.R. (1999). Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition, *J Mol Recognition*, 12(1), pp.3-18.
- Ji, C.G.; & Zhang, J.Z. (2009). Electronic polarization is important in stabilizing the native structures of proteins, *J Phys Chem B.* 2009 Dec 10;113(49), pp. 16059-64.
- Ji, C.G.; Mei, Y.; & Zhang, J.Z. (2008). Developing polarized protein-specific charges for protein dynamics: MD free energy calculation of pKa shifts for Asp26/Asp20 in thioredoxin, *Biophys J.* 2008 Aug;95(3), pp. 1080-1088.
- Klamt, A.; & Schüürmann, G. (1993). COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient, *J.Chem.Soc.Perkin Trans. 2*, p. 799-805.
- Klepeis, J.L.; Lindorff-Larsen, K.; Dror, R.O.; Shaw, D:E. (2009). Long-timescale molecular dynamics simulations of protein structure and function, *Curr Opin Struct Biol.* 2009 Apr;19(2), pp. 120-7.
- Knight, J.L.; & Brooks, C.L. III. (2009). Lambda-Dynamics free energy simulation methods. *Journal of Computational Chemistry*, 2009, pp. 1692-1700.
- Krause, H.; Ernstberger, B.; Neusser, H.J. (1991). Binding energies of small benzene clusters. *Chem. Phys. Lett.* 1991, 184, pp. 411–417.
- Krishnamurthy, V. M.; Bohall, B. R.; Semetey, V.; Whitesides, G. M. (2006). The paradoxical thermodynamic basis for the interaction of ethylene glycol, glycine, and sarcosine chains with bovine carbonic anhydrase II: an unexpected manifestation of enthalpy/entropy compensation. *J. Am. Chem. Soc.* 2006, 128, pp. 5802–5812.



- Kuser, P.R.; Franzoni, L.; Ferrari, E.; Spisni, A.; Polikarpov, I. (2001). The X-ray structure of a recombinant major urinary protein at 1.75 Å resolution. A comparative study of X-ray and NMR-derived structures, *Acta Crystallogr D Biol Crystallogr.* 2001 Dec;57(Pt 12), pp. 1863-9.
- Laio, A.; & Gervasio, F.L. (2008). Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science, *Rep. Prog. Phys.* 71, p. 126601. doi: 10.1088/0034-4885/71/12/126601
- Lange, O.F.; Schäfer, L.V.; & Grubmüller, H. (2006). Flooding in GROMACS: accelerated barrier crossings in molecular dynamics, *J Comput Chem.* 2006 Nov 15;27(14), pp. 1693-702.
- Langhorst, U.; Loris, R.; Denisov, V.P.; Doumen, J.; Roose, P.; Maes, D.; Halle, B.; Steyaert, J. (1999). Dissection of the structural and functional role of a conserved hydration site in RNase T1. *Protein Sci.* 1999 Apr;8(4), pp. 722-30.
- Li, Z.; & Liu, W. (2010). Spin-adapted open-shell random phase approximation and time-dependent density functional theory. I. Theory, *J Chem Phys.* 2010 Aug 14;133(6), pp. 064106.
- Lipari, G.; & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J. Am. Chem. Soc.*, 104, pp. 4546-4559.
- Lu, Y. et al. (2009) Halogen bonding--a novel interaction for rational drug design? *J Med Chem* 52 (9), pp. 2854-2862.
- Lundquist, J.J.; & Toone, E.J. (2002). The cluster glycoside effect, *Chem. Rev.* 2002;102, pp. 555-578.
- MacRaid, C.A.; Daranas, A.H.; Bronowska, A.; Homans, S.W. (2007). Global changes in local protein dynamics reduce the entropic cost of carbohydrate binding in the arabinose-binding protein, *J. Mol. Biol.*, 2007, 368(3), pp. 822- 832.
- Makowski, L.; Gore, D.; Mandava, S.; Minh, D.; Park, S.; Rodi, D.J.; Fischetti RF. (2011). X-ray solution scattering studies of the structural diversity intrinsic to protein ensembles, *Biopolymers.* 2011 Apr 1. doi: 10.1002/bip.21631
- Mao, Y.; Ratner, M.A.; & Jarrold, M.F. (2000). One water molecule stiffens a protein. *J. Am. Chem. Soc.* 2000, 122, pp. 2950-2951.
- Mao, B.; Pear, M.R.; McCammon, J.A.; Quioco F.A. (1982). Hinge-bending in L-arabinose-binding protein. The "Venus's- flytrap" model. *J. Biol. Chem.* 1982;257, pp. 1131-1133.
- Matthews, B.W.; & Liu, L. (2009). A review about nothing: are apolar cavities in proteins really empty?. *Protein Sci.* 2009, 18, pp. 494-502.
- Meyer, B.; & Peters, T. (2003). NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors, *Angew Chem. Int. Ed.* 2003, 42(8), pp. 864-90.
- Michel, J.; Tirado-Rives, J.; & Jorgensen, W. L. (2009). Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J. Am. Chem. Soc.* 2009, 131, pp. 15403-15411.
- Ming, D.; & Brüschweiler, R. (2004). Prediction of methyl-side chain dynamics in proteins, *J Biomol NMR.* 2004 Jul;29(3), pp. 363-8.
- Mittermaier, A.; Davidson, A.R.; & Kay, L.E. (2003). Correlation between <sup>2</sup>H NMR side-chain order parameters and sequence conservation in globular proteins, *J Am Chem Soc.* 2003 Jul 30;125(30), pp. 9004-5.

- Nagendra, H.G.; Sukumar, N.; & Vijayan, M. (1998). Role of water in plasticity, stability, and action of proteins: the crystal structures of lysozyme at very low levels of hydration, *Proteins*. 1998 Aug 1;32(2), pp. 229-40.
- Newcomer, M.E.; Lewis, B.A.; & Quioco, F.A. (1981). The radius of gyration of L-arabinose-binding protein decreases upon binding of ligand. *J. Biol. Chem.* 1981;256, pp. 13218-13222.
- Panigrahi, S.K.; & Desiraju, G.R.(2007). Strong and weak hydrogen bonds in the protein–ligand interface. *Proteins* 2007, 67, pp. 128–141.
- Paulini, R.; Müller, K.; & Diederich, F. (2005). Orthogonal multipolar interactions in structural chemistry and biology. *Angew. Chem., Int. Ed.* 2005, 44, pp. 1788–1805.
- Perozzo, R.; Folkers, G.; & Scapozza, L. (2004). Thermodynamics of protein-ligand interactions: history, presence, and future aspects, *J. Recept. Signal Transduct. Res.*, 2004, 24(1-2), pp. 1-52.
- Piela, L. (2007). *Ideas of Quantum Chemistry*, Elsevier, ISBN: 9780444522276.
- Poornima, C.S.; & Dean, P.M. (1995). Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions., *J Comput Aided Mol Des.* 1995 Dec;9(6), pp. 500-12.
- Poornima, C.S.; & Dean, P.M. (1995). Hydration in drug design. 2. Influence of local site surface shape on water binding. *J. Comput.-Aided Mol. Des.* 1995, 9, pp. 513–520.
- Poornima, C.S.; & Dean, P.M. (1995). Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins., *J Comput Aided Mol Des.* 1995 Dec;9(6), pp. 521-31.
- Popovych, N.; Sun, S., Ebright, R.H.; Kalodimos, C.G. (2006). Dynamically driven protein allostery, *Nat Struct Mol Biol.* 2006 Sep;13(9), pp. 831-8.
- Quioco, F.A. (1993). Probing the atomic interactions between proteins and carbohydrates. *Biochem. Soc. Trans.* 1993;21, pp. 442–448.
- Raha, K.; & Merz, K.M. Jr. (2004). A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding, *J Am Chem Soc.* 2004 Feb 4;126(4), pp. 1020-1.
- Raha, K.; Peters, M.B.; Wang, B.; Yu, N.; Wollacott, A.M.; Westerhoff, L.M.; Merz, K.M. Jr. (2007). The role of quantum mechanics in structure-based drug design, *Drug Discov Today.* 2007 Sep;12(17-18), pp. 725-31.
- Rezac, J.; Fanfrlik, J.; Salahub, D.; Hobza, P. (2009). Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes, *J. Chem. Theory Comp.* 2009, 5, pp. 1749-1760.
- Ringer, A.L.; Senenko, A.; & Sherrill, C.D. (2007). Models of S/pi interactions in protein structures: Comparison of the H<sub>2</sub>S–benzene complex with PDB data. *Protein Sci.* 2007, 16, pp. 2216–2223.
- Sadus, R.J. (1998). Exact calculation of the effect of three-body Axilrod-Teller interactions on vapour-liquid phase coexistence. *Fluid Phase Equilibria.* 1998;144, pp. 351–360. doi: 10.1016/S0378-3812(97)00279-3.
- Scatena, L.F.; Brown, M.G.; & Richmond, G.L. (2001). Water at hydrophobic surfaces: weak hydrogen bonding and strong orientation effects. *Science* 2001, 292, pp. 908–912.
- Schlitter, J. (1993). Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* 1993, 215, pp. 617–621.

- Seeliger, D.; Haas, J.; & de Groot, B.L. (2007). Geometry-based sampling of conformational transitions in proteins, *Structure*, 2007;15(11), pp. 1482-92.
- Senn, H.M.; & Thiel, W. (2009). QM/MM methods for biomolecular systems, *Angew Chem Int Ed Engl*. 2009;48(7), pp. 1198-229.
- Sharp, K. (2001). Entropy-enthalpy compensation: fact or artifact?. *Protein Sci*. 2001, 10, pp. 661-667.
- Shimokhina, N.; Bronowska, A.K.; & Homans, S.W. (2006). Contribution of Ligand Desolvation to Binding Thermodynamics in a Ligand-Protein Interaction, *Angew Chem Int Ed Engl*. 2006 Sep 25;45(38), pp. 6374-6.
- Showalter, S.A.; & Brüschweiler, R. (2007). Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field, *J. Chem. Theory Comput.*, 2007, 3, pp. 961-975.
- Sinnokrot, M.O.; & Sherrill, C.D. (2004). Substituent effects in pi-pi interactions: sandwich and T-shaped configurations. *J. Am. Chem. Soc.* 2004, 126, pp. 7690-7697.
- Smith, J.C.; Merzel, F.; Bondar, A.N.; Tournier, A.; Fischer S. (2004). Structure, dynamics and reactions of protein hydration water., *Philos Trans R Soc Lond B Biol Sci*. 2004 August 29; 359(1448), pp. 1181-1190.
- Stewart, J.J. (2009). Application of the PM6 method to modeling proteins, *J Mol Model*. 2009 Jul;15(7), pp. 765-805
- Stoekmann, H.; Bronowska, A.K.; Syme, N.R.; Thompson, G.S.; Kalverda, A.P.; Warriner, S.L.; Homans, S.W. (2008). Residual ligand entropy in the binding of p-substituted benzenesulfonamide ligands to bovine carbonic anhydrase II, *J Am Chem Soc*. 2008 Sep 17;130(37), pp. 12420-6.
- Straatsma, T.P.; & Berendsen, H.J.C. (1988). Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations, *J. Chem. Phys.* 89, pp. 5876- 5886.
- Timm, D.E.; Baker, L.J.; Mueller, H.; Zidek, L.; Novotny, M.V. (2001). Structural basis of pheromone binding to mouse major urinary protein (MUP-I), *Protein Sci*. 2001 May;10(5), pp. 997-1004.
- Tochtrop, G.P.; Richter, K.; Tang, C.; Toner, J.J.; Covey, D.F.; Cistola, D.P. (2002). Energetics by NMR: site-specific binding in a positively cooperative system. *Proc. Natl. Acad. Sci. U.S.A.* 2002, 99, pp. 1847-1852.
- Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanaba, K. (2000). Origin of the attraction and directionality of the NH/pi interaction: comparison with OH/pi and CH/pi interactions. *J. Am. Chem. Soc.* 2000, 122, pp. 11450- 11458.
- Turk, J.A.; & Smithrud, D.B. (2001). Synthesis and physical properties of protein core mimetics, *J. Org. Chem.* 2001, 66, pp. 8328-8335.
- Turnbull, W. B.; Precious, B. L.; & Homans, S. W. (2004). Dissecting the cholera toxin-ganglioside GM1 interaction by isothermal titration calorimetry, *J. Am. Chem. Soc.* 2004, 126, pp. 1047-1054.
- Vallone, B.; Miele, A.E.; Vecchini, P.; Chiancone, E.; Brunori, M. (1998). Free energy of burying hydrophobic residues in the interface between protein subunits. *Proc. Natl. Acad. Sci. U.S.A.* 1998, 95, pp. 6103-6107.
- Vyas, N.K.; Vyas, M.N.; & Quijcho, F.A. (1991). Comparison of the periplasmic receptors for L-arabinose, D-glucose/D- galactose, and D-ribose. *Structural and Functional Similarity*, *J Biol Chem*. 1991 Mar 15;266(8), pp. 5226-37.

- Wand, A.J. (2001). Dynamic activation of protein function: a view emerging from NMR spectroscopy, *Nat Struct Biol.* 2001 Nov;8(11), pp. 926-31.
- Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. (2004). Development and testing of a general amber force field. *J Comp Chem.* 2004;25, pp. 1157–1174. doi: 10.1002/jcc.20035.
- Weiss, S. (2000). Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy, *Nat Struct Biol.* 2000 Sep;7(9), pp. 724-9.
- Whitesides, G.M.; & Krishnamurthy, V.M. (2005). Designing ligands to bind proteins, *Q. Rev. Biophys.*, 2005, 38, pp. 385- 395.
- Williams, M.A.; & Ladbury, J.E. (2003). Hydrogen bonds in protein–ligand complexes. *Methods Princ. Med. Chem.* 2003, 19, pp. 137–161.
- Williams, D.H.; Stephens, E.; & Zhou, M. (2003). Ligand binding energy and catalytic efficiency from improved packing within receptors and enzymes, *J Mol Biol.* 2003 May 30;329(2), pp. 389-99.
- Xu, J.; Plaxco, K.W.; & Allen, J.S. (2006). Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy, *Protein Sci.* 2006 May; 15(5), pp. 1175–1181.
- Yang, D.; & Kay, L.E. (1996). Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding, *J. Mol. Biol.*, 1996, 263(2), pp. 369-82.
- Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. (2007). Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, pp. 808–813.
- Zauhar, R.J.; Colbert, C.L.; Morgan, R.S.; Welsh, W.J. (2000). Evidence for a strong sulfur–aromatic interaction derived from crystallographic data. *Biopolymers* 2000, 53, pp. 233–248.
- Zhou, P.; Tian, F.; Zou, J.; Shang, Z. (2010). Rediscovery of halogen bonds in protein–ligand complexes. *Mini Rev Med Chem.* 2010;10(4), pp. 309-314.

# Atmospheric Thermodynamics

Francesco Cairo

*Consiglio Nazionale delle Ricerche – Istituto di  
Scienze dell'Atmosfera e del Clima  
Italy*

## 1. Introduction

Thermodynamics deals with the transformations of the energy in a system and between the system and its environment. Hence, it is involved in every atmospheric process, from the large scale general circulation to the local transfer of radiative, sensible and latent heat between the surface and the atmosphere and the microphysical processes producing clouds and aerosol. Thus the topic is much too broad to find an exhaustive treatment within the limits of a book chapter, whose main goal will be limited to give a broad overview of the implications of thermodynamics in the atmospheric science and introduce some of its jargon. The basic thermodynamic principles will not be reviewed here, while emphasis will be placed on some topics that will find application to the interpretation of fundamental atmospheric processes. An overview of the composition of air will be given, together with an outline of its stratification in terms of temperature and water vapour profile. The ideal gas law will be introduced, together with the concept of hydrostatic stability, temperature lapse rate, scale height, and hydrostatic equation. The concept of an air parcel and its enthalpy and free energy will be defined, together with the potential temperature concept that will be related to the static stability of the atmosphere and connected to the Brunt-Vaisala frequency.

Water phase changes play a pivotal role in the atmosphere and special attention will be placed on these transformations. The concept of vapour pressure will be introduced together with the Clausius-Clapeyron equation and moisture parameters will be defined. Adiabatic transformation for the unsaturated and saturated case will be discussed with the help of some aerological diagrams of common practice in Meteorology and the notion of neutral buoyancy and free convection will be introduced and considered referring to an exemplificative atmospheric sounding. There, the Convective Inhibition and Convective Available Potential Energy will be introduced and examined. The last subchapter is devoted to a brief overview of warm and cold clouds formation processes, with the aim to stimulate the interest of reader toward more specialized texts, as some of those listed in the conclusion and in the bibliography.

## 2. Dry air thermodynamics and stability

We know from experience that pressure, volume and temperature of any homogeneous substance are connected by an *equation of state*. These physical variables, for all gases over a

wide range of conditions in the so called *perfect gas approximation*, are connected by an equation of the form:

$$pV=mRT \quad (1)$$

where  $p$  is pressure (Pa),  $V$  is volume ( $\text{m}^3$ ),  $m$  is mass (kg),  $T$  is temperature (K) and  $R$  is the *specific gas constant*, whose value depends on the gas. If we express the amount of substance in terms of number of moles  $n=m/M$  where  $M$  is the gas molecular weight, we can rewrite (1) as:

$$pV=nR^*T \quad (2)$$

where  $R^*$  is the *universal gas constant*, whose value is  $8.3143 \text{ J mol}^{-1} \text{ K}^{-1}$ . In the kinetic theory of gases, the perfect gas is modelled as a collection of rigid spheres randomly moving and bouncing between each other, with no common interaction apart from these mutual shocks. This lack of reciprocal interaction leads to derive the internal energy of the gas, that is the sum of all the kinetic energies of the rigid spheres, as proportional to its temperature. A second consequence is that for a mixture of different gases we can define, for each component  $i$ , a partial pressure  $p_i$  as the pressure that it would have if it was alone, at the same temperature and occupying the same volume. Similarly we can define the partial volume  $V_i$  as that occupied by the same mass at the same pressure and temperature, holding *Dalton's law* for a mixture of gases  $i$ :

$$p=\sum p_i \quad (3)$$

Where for each gas it holds:

$$p_i V=n_i R^* T \quad (4)$$

We can still make use of (1) for a mixture of gases, provided we compute a specific gas constant  $R$  as:

$$\bar{R} = \frac{\sum m_i R_i}{m} \quad (5)$$

The atmosphere is composed by a mixture of gases, water substance in any of its three physical states and solid or liquid suspended particles (aerosol). The main components of dry atmospheric air are listed in Table 1.

Gas	Molar fraction	Mass fraction	Specific gas constant ( $\text{J Kg}^{-1} \text{ K}^{-1}$ )
Nitrogen (N <sub>2</sub> )	0.7809	0.7552	296.80
Oxygen (O <sub>2</sub> )	0.2095	0.2315	259.83
Argon (Ar)	0.0093	0.0128	208.13
Carbon dioxide (CO <sub>2</sub> )	0.0003	0.0005	188.92

Table 1. Main component of dry atmospheric air.

The composition of air is constant up to about 100 km, while higher up molecular diffusion dominates over turbulent mixing, and the percentage of lighter gases increases with height. For the pivotal role water substance plays in weather and climate, and for the extreme variability of its presence in the atmosphere, with abundances ranging from few percents to

millionths, it is preferable to treat it separately from other air components, and consider the atmosphere as a mixture of dry gases and water. In order to use a state equation of the form (1) for moist air, we express a specific gas constant  $R_d$  by considering in (5) all gases but water, and use in the state equation a *virtual temperature*  $T_v$  defined as the temperature that dry air must have in order to have the same density of moist air at the same pressure. It can be shown that

$$T_v = \frac{T}{1 - \frac{e}{p} \left(1 - \frac{M_w}{M_d}\right)} \quad (6)$$

Where  $M_w$  and  $M_d$  are respectively the water and dry air molecular weights.  $T_v$  takes into account the smaller density of moist air, and so is always greater than the actual temperature, although often only by few degrees.

## 2.1 Stratification

The atmosphere is under the action of a gravitational field, so at any given level the downward force per unit area is due to the weight of all the air above. Although the air is permanently in motion, we can often assume that the upward force acting on a slab of air at any level, equals the downward gravitational force. This *hydrostatic balance* approximation is valid under all but the most extreme meteorological conditions, since the vertical acceleration of air parcels is generally much smaller than the gravitational one. Consider an horizontal slab of air between  $z$  and  $z + \delta z$ , of unit horizontal surface. If  $\rho$  is the air density at  $z$ , the downward force acting on this slab due to gravity is  $g\rho\delta z$ . Let  $p$  be the pressure at  $z$ , and  $p + \delta p$  the pressure at  $z + \delta z$ . We consider  $\delta p$  as negative, since we know that pressure decreases with height. The hydrostatic balance of forces along the vertical leads to:

$$-\delta p = g\rho\delta z \quad (7)$$

Hence, in the limit of infinitesimal thickness, the *hypsoetric equation* holds:

$$-\frac{\partial p}{\partial z} = -g\rho \quad (8)$$

leading to:

$$p(z) = \int_z^\infty g\rho dz \quad (9)$$

As we know that  $p(\infty)=0$ , (9) can be integrated if the air density profile is known.

Two useful concepts in atmospheric thermodynamic are the geopotential  $\Phi$ , an exact differential defined as the work done against the gravitational field to raise 1 kg from 0 to  $z$ , where the 0 level is often taken at sea level and, to set the constant of integration,  $\Phi(0)=0$ , and the *geopotential height*  $Z = \Phi/g_0$ , where  $g_0$  is a mean gravitational acceleration taken as 9,81 m/s.

We can rewrite (9) as:

$$Z(z) = \frac{1}{g_0} \int_z^\infty g dz \quad (10)$$

Values of  $z$  and  $Z$  often differ by not more than some tens of metres.

We can make use of (1) and of the definition of virtual temperature to rewrite (10) and formally integrate it between two levels to formally obtain the geopotential thickness of a layer, as:

$$\Delta Z = \frac{R_d}{g_0} \int_{p_1}^{p_2} T_v \frac{dp}{p} \quad (11)$$

The above equations can be integrated if we know the virtual temperature  $T_v$  as a function of pressure, and many limiting cases can be envisaged, as those of constant vertical temperature gradient. A very simplified case is for an isothermal atmosphere at a temperature  $T_v=T_0$ , when the integration of (11) gives:

$$\Delta Z = \frac{R_d T_0}{g_0} \ln \left( \frac{p_1}{p_2} \right) = H \ln \left( \frac{p_1}{p_2} \right) \quad (12)$$

In an isothermal atmosphere the pressure decreases exponentially with an e-folding scale given by the *scale height*  $H$  which, at an average atmospheric temperature of 255 K, corresponds roughly to 7.5 km. Of course, atmospheric temperature is by no means constant: within the lowest 10-20 km it decreases with a *lapse rate* of about 7 K km<sup>-1</sup>, highly variable depending on latitude, altitude and season. This region of decreasing temperature with height is termed *troposphere*, (from the Greek “turning/changing sphere”) and is capped by a region extending from its boundary, termed *tropopause*, up to 50 km, where the temperature is increasing with height due to solar UV absorption by ozone, that heats up the air. This region is particularly stable and is termed *stratosphere* (“layered sphere”). Higher above in the *mesosphere* (“middle sphere”) from 50 km to 80-90 km, the temperature falls off again. The last region of the atmosphere, named *thermosphere*, sees the temperature rise again with altitude to 500-2000K up to an isothermal layer several hundreds of km distant from the ground, that finally merges with the interplanetary space where molecular collisions are rare and temperature is difficult to define. Fig. 1 reports the atmospheric temperature, pressure and density profiles. Although the atmosphere is far from isothermal, still the decrease of pressure and density are close to be exponential. The atmospheric temperature profile depends on vertical mixing, heat transport and radiative processes.

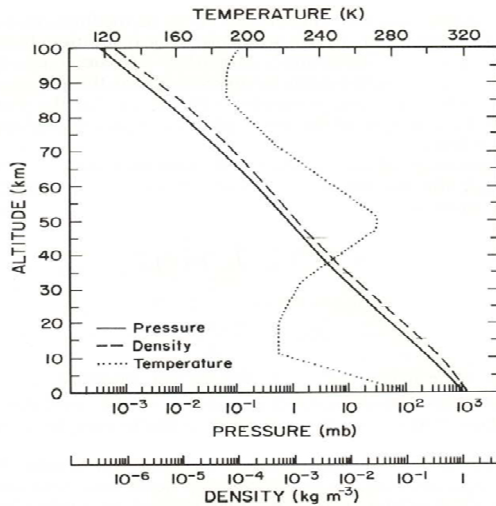


Fig. 1. Temperature (dotted line), pressure (dashed line) and air density (solid line) for a standard atmosphere.



## 2.2 Thermodynamic of dry air

A system is *open* if it can exchange matter with its surroundings, *closed* otherwise. In atmospheric thermodynamics, the concept of “air parcel” is often used. It is a good approximation to consider the air parcel as a closed system, since significant mass exchanges between airmasses happen predominantly in the few hundreds of metres close to the surface, the so-called *planetary boundary layer* where mixing is enhanced, and can be neglected elsewhere. An air parcel can exchange energy with its surrounding by work of expansion or contraction, or by exchanging heat. An *isolated* system is unable to exchange energy in the form of heat or work with its surroundings, or with any other system. The first principle of thermodynamics states that the *internal energy*  $U$  of a closed system, the kinetic and potential energy of its components, is a state variable, depending only on the present state of the system, and not by its past. If a system evolves without exchanging any heat with its surroundings, it is said to perform an *adiabatic* transformation. An air parcel can exchange heat with its surroundings through diffusion or thermal conduction or radiative heating or cooling; moreover, evaporation or condensation of water and subsequent removal of the condensate promote an exchange of latent heat. It is clear that processes which are not adiabatic ultimately lead the atmospheric behaviours. However, for timescales of motion shorter than one day, and disregarding cloud processes, it is often a good approximation to treat air motion as adiabatic.

### 2.2.1 Potential temperature

For adiabatic processes, the first law of thermodynamics, written in two alternative forms:

$$c_v dT + p dv = \delta q \quad (13)$$

$$c_p dT - v dp = \delta q \quad (14)$$

holds for  $\delta q = 0$ , where  $c_p$  and  $c_v$  are respectively the specific heats at constant pressure and constant volume,  $p$  and  $v$  are the specific pressure and volume, and  $\delta q$  is the heat exchanged with the surroundings. Integrating (13) and (14) and making use of the ideal gas state equation, we get the Poisson's equations:

$$T v^{\gamma-1} = \text{constant} \quad (15)$$

$$T p^{-\kappa} = \text{constant} \quad (16)$$

$$p v^{\gamma} = \text{constant} \quad (17)$$

where  $\gamma = c_p/c_v = 1.4$  and  $\kappa = (\gamma-1)/\gamma = R/c_p \approx 0.286$ , using a result of the kinetic theory for diatomic gases. We can use (16) to define a new state variable that is conserved during an adiabatic process, the *potential temperature*  $\theta$ , which is the temperature the air parcel would attain if compressed, or expanded, adiabatically to a reference pressure  $p_0$ , taken for convention as 1000 hPa.

$$\theta = T \left( \frac{p_0}{p} \right)^{\kappa} \quad (18)$$

Since the time scale of heat transfers, away from the planetary boundary layer and from clouds is several days, and the timescale needed for an air parcel to adjust to environmental pressure changes is much shorter,  $\theta$  can be considered conserved along the air motion for one week or more. The distribution of  $\theta$  in the atmosphere is determined by the pressure



### 2.2.2 Entropy and potential temperature

The second law of the thermodynamics allows for the introduction of another state variable, the *entropy*  $s$ , defined in terms of a quantity  $\delta q/T$  which is not in general an exact differential, but is so for a reversible process, that is a process proceeding through states of the system which are always in equilibrium with the environment. Under such cases we may pose  $ds = (\delta q/T)_{rev}$ . For the generic process, the heat absorbed by the system is always lower than what can be absorbed in the reversible case, since a part of heat is lost to the environment. Hence, a statement of the second law of thermodynamics is:

$$ds \geq \frac{\delta q}{T} \quad (23)$$

If we introduce (22) in (23), we note how such expression, connecting potential temperature to entropy, would contain only state variables. Hence equality must hold and we get:

$$d \ln \theta = \frac{ds}{c_p} \quad (24)$$

That directly relates changes in potential temperature with changes in entropy. We stress the fact that in general an adiabatic process does not imply a conservation of entropy. A classical textbook example is the adiabatic free expansion of a gas. However, in atmospheric processes, adiabaticity not only implies the absence of heat exchange through the boundaries of the system, but also absence of heat exchanges between parts of the system itself (Landau et al., 1980), that is, no turbulent mixing, which is the principal source of irreversibility. Hence, in the atmosphere, an adiabatic process always conserves entropy.

### 2.3 Stability

The vertical gradient of potential temperature determines the stratification of the air. Let us differentiate (18) with respect to  $z$ :

$$\frac{\partial \ln \theta}{\partial z} = \frac{\partial \ln T}{\partial z} + \frac{R}{c_p} \left( \frac{\partial p_0}{\partial z} - \frac{\partial p}{\partial z} \right) \quad (25)$$

By computing the differential of the logarithm, and applying (1) and (8), we get:

$$\frac{T}{\theta} \frac{\partial \theta}{\partial z} = \frac{\partial T}{\partial z} + \frac{g}{c_p} \quad (26)$$

If  $\Gamma = -(\partial T/\partial z)$  is the environment lapse rate, we get:

$$\Gamma = \Gamma_d - \frac{T}{\theta} \frac{\partial \theta}{\partial z} \quad (27)$$

Now, consider a vertical displacement  $\delta z$  of an air parcel of mass  $m$  and let  $\rho$  and  $T$  be the density and temperature of the parcel, and  $\rho'$  and  $T'$  the density and temperature of the surrounding. The restoring force acting on the parcel per unit mass will be:

$$f_z = -\frac{\rho' - \rho}{\rho'} g \quad (28)$$

That, by using (1), can be rewritten as:

$$f_z = -\frac{T - T'}{T} g \quad (29)$$

We can replace  $(T-T')$  with  $(\Gamma_d - \Gamma) \delta z$  if we acknowledge the fact that the air parcel moves adiabatically in an environment of lapse rate  $\Gamma$ . The second order equation of motion (29) can be solved in  $\delta z$  and describes buoyancy oscillations with period  $2\pi/N$  where  $N$  is the Brunt-Vaisala frequency:

$$N = \left[ \frac{g}{T} (\Gamma_d - \Gamma) \right]^{1/2} = \left[ \frac{g}{c_p} \frac{\partial \theta}{\partial z} \right]^{1/2} \quad (30)$$

It is clear from (30) that if the environment lapse rate is smaller than the adiabatic one, or equivalently if the potential temperature vertical gradient is positive,  $N$  will be real and an air parcel will oscillate around an equilibrium: if displaced upward, the air parcel will find itself colder, hence heavier than the environment and will tend to fall back to its original place; a similar reasoning applies to downward displacements. If the environment lapse rate is greater than the adiabatic one, or equivalently if the potential temperature vertical gradient is negative,  $N$  will be imaginary so the upward moving air parcel will be lighter than the surrounding and will experience a net buoyancy force upward. The condition for atmospheric stability can be inspected by looking at the vertical gradient of the potential temperature: if  $\theta$  increases with height, the atmosphere is stable and vertical motion is discouraged, if  $\theta$  decreases with height, vertical motion occurs. For average tropospheric conditions,  $N \approx 10^{-2} \text{ s}^{-1}$  and the period of oscillation is some tens of minutes. For the more stable stratosphere,  $N \approx 10^{-1} \text{ s}^{-1}$  and the period of oscillation is some minutes. This greater stability of the stratosphere acts as a sort of damper for the weather disturbances, which are confined in the troposphere.

### 3. Moist air thermodynamics

The conditions of the terrestrial atmosphere are such that water can be present under its three forms, so in general an air parcel may contain two gas phases, dry air (d) and water vapour (v), one liquid phase (l) and one ice phase (i). This is a heterogeneous system where, in principle, each phase can be treated as an homogeneous subsystem open to exchanges with the other systems. However, the whole system should be in thermodynamical equilibrium with the environment, and thermodynamical and chemical equilibrium should hold between each subsystem, the latter condition implying that no conversion of mass should occur between phases. In the case of water in its vapour and liquid phase, the chemical equilibrium imply that the vapour phases attains a saturation vapour pressure  $e_s$  at which the rate of evaporation equals the rate of condensation and no net exchange of mass between phases occurs.

The concept of chemical equilibrium leads us to recall one of the thermodynamical potentials, the *Gibbs function*, defined in terms of the *enthalpy* of the system. We remind the definition of enthalpy of a system of unit mass:

$$h = u + pv \quad (31)$$

Where  $u$  is its specific internal energy,  $v$  its specific volume and  $p$  its pressure in equilibrium with the environment. We can think of  $h$  as a measure of the total energy of the system. It includes both the internal energy required to create the system, and the amount of energy required to make room for it in the environment, establishing its volume and balancing its pressure against the environmental one. Note that this additional energy is not stored in the system, but rather in its environment.

The First law of thermodynamics can be set in a form where  $h$  is explicited as:

$$\delta q = dh - vdp \quad (32)$$

And, making use of (14) we can set:

$$dh = c_p dT \quad (33)$$

By combining (32), (33) and (8), and incorporating the definition of geopotential  $\Phi$  we get:

$$\delta q = d(h + \Phi) \quad (34)$$

Which states that an air parcel moving adiabatically in an hydrostatic atmosphere conserves the sum of its enthalpy and geopotential.

The specific Gibbs free energy is defined as:

$$g = h - Ts = u + pv - Ts \quad (35)$$

It represents the energy available for conversion into work under an isothermal-isobaric process. Hence the criterion for thermodynamical equilibrium for a system at constant pressure and temperature is that  $g$  attains a minimum.

For an heterogeneous system where multiple phases coexist, for the  $k$ -th species we define its *chemical potential*  $\mu_k$  as the partial molar Gibbs function, and the equilibrium condition states that the chemical potentials of all the species should be equal. The proof is straightforward: consider a system where  $n_v$  moles of vapour ( $v$ ) and  $n_l$  moles of liquid water ( $l$ ) coexist at pressure  $e$  and temperature  $T$ , and let  $G = n_v \mu_v + n_l \mu_l$  be the Gibbs function of the system. We know that for a virtual displacement from an equilibrium condition,  $dG > 0$  must hold for any arbitrary  $dn_v$  (which must be equal to  $-dn_l$ , whether its positive or negative) hence, its coefficient must vanish and  $\mu_v = \mu_l$ .

Note that if evaporation occurs, the vapour pressure  $e$  changes by  $de$  at constant temperature, and  $d\mu_v = v_v de$ ,  $d\mu_l = v_l de$  where  $v_v$  and  $v_l$  are the volume occupied by a single molecule in the vapour and the liquid phase. Since  $v_v \gg v_l$  we may pose  $d(\mu_v - \mu_l) = v_v de$  and, using the state gas equation for a single molecule,  $d(\mu_v - \mu_l) = (kT/e) de$ . In the equilibrium,  $\mu_v = \mu_l$  and  $e = e_s$  while in general:

$$(\mu_v - \mu_l) = kT \ln \left( \frac{e}{e_s} \right) \quad (36)$$

holds. We will make use of this relationship we we will discuss the formation of clouds.

### 3.1 Saturation vapour pressure

The value of  $e_s$  strongly depends on temperature and increases rapidly with it. The celebrated Clausius -Clapeyron equation describes the changes of saturated water pressure above a plane surface of liquid water. It can be derived by considering a liquid in equilibrium with its saturated vapour undergoing a Carnot cycle (Fermi, 1956). We here simply state the result as:

$$\frac{de_s}{dT} = \frac{L_v}{T\alpha} \quad (37)$$

Retrieved under the assumption that the specific volume of the vapour phase is much greater than that of the liquid phase.  $L_v$  is the latent heat, that is the heat required to convert

a unit mass of substance from the liquid to the vapour phase without changing its temperature. The latent heat itself depends on temperature – at 1013 hPa and 0°C is  $2.5 \cdot 10^6$  J kg<sup>-1</sup>, - hence a number of numerical approximations to (37) have been derived. The World Meteorological Organization bases its recommendation on a paper by Goff (1957):

$$\begin{aligned} \text{Log}_{10} e_s = & 10.79574 (1 - 273.16 / T) - 5.02800 \text{Log}_{10}(T / 273.16) + \\ & + 1.50475 \cdot 10^{-4} (1 - 10(-8.2969 * (T / 273.16 - 1))) + 0.42873 \cdot 10^{-} \\ & -3 (10(+4.76955 * (1 - 273.16 / T)) - 1) + 0.78614 \end{aligned} \quad (38)$$

Where  $T$  is expressed in K and  $e_s$  in hPa. Other formulations are used, based on direct measurements of vapour pressures and theoretical calculation to extrapolate the formulae down to low  $T$  values (Murray, 1967; Bolton, 1980; Hyland and Wexler, 1983; Sonntag, 1994; Murphy and Koop, 2005) uncertainties at low temperatures become increasingly large and the relative deviations within these formulations are of 6% at -60°C and of 9% at -70°.

An equation similar to (37) can be derived for the vapour pressure of water over ice  $e_{si}$ . In such a case,  $L_v$  is the latent heat required to convert a unit mass of water substance from ice to vapour phase without changing its temperature. A number of numerical approximations holds, as the Goff-Gratch equation, considered the reference equation for the vapor pressure over ice over a region of -100°C to 0°C:

$$\begin{aligned} \text{Log}_{10} e_{si} = & -9.09718 (273.16 / T - 1) - 3.56654 \text{Log}_{10}(273.16 / T) + \\ & + 0.876793 (1 - T / 273.16) + \text{Log}_{10}(6.1071) \end{aligned} \quad (39)$$

with  $T$  in K and  $e_{si}$  in hPa. Other equations have also been widely used (Murray, 1967; Hyland and Wexler, 1983; Marti and Mauersberger, 1993; Murphy and Koop, 2005).

Water evaporates more readily than ice, that is  $e_s > e_{si}$  everywhere (the difference is maxima around -20°C), so if liquid water and ice coexists below 0°C, the ice phase will grow at the expense of the liquid water.

### 3.2 Water vapour in the atmosphere

A number of moisture parameters can be formulated to express the amount of water vapour in the atmosphere. The *mixing ratio*  $r$  is the ratio of the mass of the water vapour  $m_v$ , to the mass of dry air  $m_d$ ,  $r = m_v / m_d$  and is expressed in g/kg<sup>-1</sup> or, for very small concentrations as those encountered in the stratosphere, in parts per million in volume (ppmv). At the surface, it typically ranges from 30-40 g/kg<sup>-1</sup> at the tropics to less than 5 g/kg<sup>-1</sup> at the poles; it decreases approximately exponentially with height with a scale height of 3-4 km, to attain its minimum value at the tropopause, driest at the tropics where it can get as low as a few ppmv. If we consider the ratio of  $m_v$  to the total mass of air, we get the *specific humidity*  $q$  as  $q = m_v / (m_v + m_d) = r / (1 + r)$ . The *relative humidity*  $RH$  compares the water vapour pressure in an air parcel with the maximum water vapour it may sustain in equilibrium at that temperature, that is  $RH = 100 e / e_s$  (expressed in percentages). The dew point temperature  $T_d$  is the temperature at which an air parcel with a water vapour pressure  $e$  should be brought isobarically in order to become saturated with respect to a plane surface of water. A similar definition holds for the frost point temperature  $T_f$ , when the saturation is considered with respect to a plane surface of ice.

The *wet-bulb temperature*  $T_w$  is defined operationally as the temperature a thermometer would attain if its glass bulb is covered with a moist cloth. In such a case the thermometer is

cooled upon evaporation until the surrounding air is saturated: the heat required to evaporate water is supplied by the surrounding air that is cooled. An evaporating droplet will be at the wet-bulb temperature. It should be noted that if the surrounding air is initially unsaturated, the process adds water to the air close to the thermometer, to become saturated, hence it increases its mixing ratio  $r$  and in general  $T \geq T_w \geq T_{dr}$ , the equality holds when the ambient air is already initially saturated.

### 3.3 Thermodynamics of the vertical motion

The saturation mixing ratio depends exponentially on temperature. Hence, due to the decrease of ambient temperature with height, the saturation mixing ratio sharply decreases with height as well.

Therefore the water pressure of an ascending moist parcel, despite the decrease of its temperature at the dry adiabatic lapse rate, sooner or later will reach its saturation value at a level named *lifting condensation level* (LCL), above which further lifting may produce condensation and release of latent heat. This internal heating slows the rate of cooling of the air parcel upon further lifting.

If the condensed water stays in the parcel, and heat transfer with the environment is negligible, the process can be considered reversible - that is, the heat internally added by condensation could be subtracted by evaporation if the parcel starts descending - hence the behaviour can still be considered adiabatic and we will term it a *saturated adiabatic process*. If otherwise the condensate is removed, as instance by sedimentation or precipitation, the process cannot be considered strictly adiabatic. However, the amount of heat at play in the condensation process is often negligible compared to the internal energy of the air parcel and the process can still be considered well approximated by a saturated adiabat, although it should be more properly termed a *pseudoadiabatic process*.

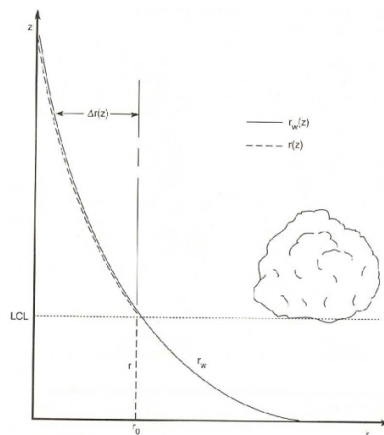


Fig. 3. Vertical profiles of mixing ratio  $r$  and saturated mixing ratio  $r_s$  for an ascending air parcel below and above the lifting condensation level. (source: Salby M. L., Fundamentals of Atmospheric Physics, Academic Press, New York.)

#### 3.3.1 Pseudoadiabatic lapse rate

If within an air parcel of unit mass, water vapour condenses at a saturation mixing ratio  $r_s$ , the amount of latent heat released during the process will be  $-L_w dr_s$ . This can be put into (34) to get:

$$-L_w dr_s = c_p dT + g dz \quad (40)$$

Dividing by  $c_p dz$  and rearranging terms, we get the expression of the *saturated adiabatic lapse rate*  $\Gamma_s$ :

$$\Gamma_s = -\frac{dT}{dz} = \frac{\Gamma_d}{\left(1 + \left(\frac{L_w}{c_p}\right)\left(\frac{dr_s}{dz}\right)\right)} \quad (41)$$

Whose value depends on pressure and temperature and which is always smaller than  $\Gamma_d$ , as should be expected since a saturated air parcel, since condensation releases latent heat, cools more slowly upon lifting.

### 3.3.2 Equivalent potential temperature

If we pose  $\delta q = -L_w dr_s$  in (22) we get:

$$\frac{d\theta}{\theta} = -\frac{L_w dr_s}{c_p T} \simeq -d\left(\frac{L_w r_s}{c_p T}\right) \quad (42)$$

The approximate equality holds since  $dT/T \ll dr_s/r_s$  and  $L_w/c_p$  is approximately independent of T. So (41) can be integrated to yield:

$$\theta_e = \theta \exp\left(\frac{L_w r_s}{c_p T}\right) \quad (43)$$

That defines the *equivalent potential temperature*  $\theta_e$  (Bolton, 1990) which is constant along a pseudoadiabatic process, since during the condensation the reduction of  $r_s$  and the increase of  $\theta$  act to compensate each other.

### 3.4 Stability for saturated air

We have seen for the case of dry air that if the environment lapse rate is smaller than the adiabatic one, the atmosphere is stable: a restoring force exist for infinitesimal displacement of an air parcel. The presence of moisture and the possibility of latent heat release upon condensation complicates the description of stability.

If the air is saturated, it will cool upon lifting at the smaller saturated lapse rate  $\Gamma_s$  so in an environment of lapse rate  $\Gamma$ , for the saturated air parcel the cases  $\Gamma < \Gamma_s$ ,  $\Gamma = \Gamma_s$ ,  $\Gamma > \Gamma_s$  discriminates the absolutely stable, neutral and unstable conditions respectively. An interesting case occurs when the environmental lapse rate lies between the dry adiabatic and the saturated adiabatic, that is  $\Gamma_s < \Gamma < \Gamma_d$ . In such a case, a moist unsaturated air parcel can be lifted high enough to become saturated, since the decrease in its temperature due to adiabatic cooling is offset by the faster decrease in water vapour saturation pressure, and starts condensation at the LCL. Upon further lifting, the air parcel eventually get warmer than its environment at a level termed *Level of Free Convection (LFC)* above which it will develop a positive buoyancy fuelled by the continuous release of latent heat due to condensation, as long as there is vapour to condense. This situation of *conditional instability* is most common in the atmosphere, especially in the Tropics, where a forced finite uplifting of moist air may eventually lead to spontaneous convection. Let us refer to figure 4 and follow such process more closely. In the figure, which is one of the meteograms discussed later in the chapter, pressure decreases vertically, while lines of constant temperature are tilted 45° rightward, temperature decreasing going up and to the left.



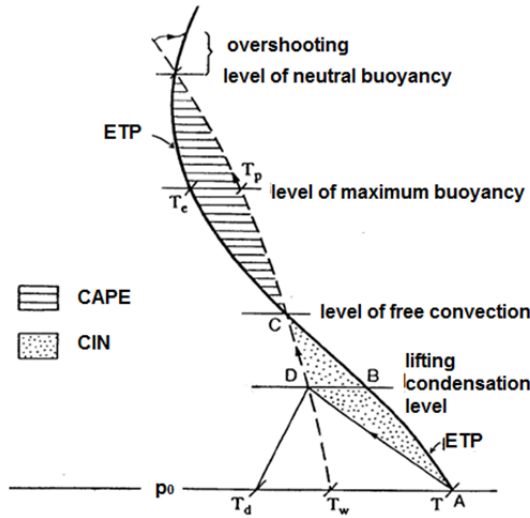


Fig. 4. Thick solid line represent the environment temperature profile. Thin solid line represent the temperature of an ascending parcel initially at point A. Dotted area represent CIN, shaded area represent CAPE.

The thick solid line represent the environment temperature profile. A moist air parcel initially at rest at point A is lifted and cools at the adiabatic lapse rate  $\Gamma_d$  along the thin solid line until it eventually get saturated at the Lifting Condensation Level at point D. During this lifting, it gets colder than the environment. Upon further lifting, it cools at a slower rate at the pseudoadiabatic lapse rate  $\Gamma_s$  along the thin dashed line until it reaches the Level of Free Convection at point C, where it attains the temperature of the environment. If it gets beyond that point, it will be warmer, hence lighter than the environment and will experience a positive buoyancy force. This buoyancy will sustain the ascent of the air parcel until all vapour condenses or until its temperature crosses again the profile of environmental temperature at the *Level of Neutral Buoyancy (LNB)*. Actually, since the air parcel gets there with a positive vertical velocity, this level may be surpassed and the air parcel may overshoot into a region where it experiences negative buoyancy, to eventually get mixed there or splash back to the LNB. In practice, entrainment of environmental air into the ascending air parcel often occurs, mitigates the buoyant forces, and the parcel generally reaches below the LNB.

If we neglect such entrainment effects and consider the motion as adiabatic, the buoyancy force is conservative and we can define a potential. Let  $\rho$  and  $\rho'$  be respectively the environment and air parcel density. From Archimede's principle, the buoyancy force on a unit mass parcel can be expressed as in (29), and the increment of potential energy for a displacement  $\delta z$  will then be, by using (1) and (8):

$$dP = f_b = \left(\frac{T'-T}{T}\right) g \delta z = R(T' - T) d \log p \tag{44}$$

Which can be integrated from a reference level  $p_0$  to give:

$$dP(p) = -R \int_{p_0}^p (T' - T) d \log p = -RA(p) \tag{45}$$

Referring to fig. 4,  $A(p)$  represent the shaded area between the environment and the air parcel temperature profiles. An air parcel initially in A is bound inside a “potential energy well” whose depth is proportional to the dotted area, and that is termed *Convective Inhibition (CIN)*. If forcedly raised to the level of free convection, it can ascent freely, with an available potential energy given by the shaded area, termed *CAPE (Convective Available Potential Energy)*.

In absence of entrainment and frictional effects, all this potential energy will be converted into kinetic energy, which will be maximum at the level of neutral buoyancy. CIN and CAPE are measured in J/Kg and are indices of the atmospheric instability. The CAPE is the maximum energy which can be released during the ascent of a parcel from its free buoyant level to the top of the cloud. It measures the intensity of deep convection, the greater the CAPE, the more vigorous the convection. Thunderstorms require large CAPE of more than 1000 Jkg<sup>-1</sup>.

CIN measures the amount of energy required to overcome the negatively buoyant energy the environment exerts on the air parcel, the smaller, the more unstable the atmosphere, and the easier to develop convection. So, in general, convection develops when CIN is small and CAPE is large. We want to stress that some CIN is needed to build-up enough CAPE to eventually fuel the convection, and some mechanical forcing is needed to overcome CIN. This can be provided by cold front approaching, flow over obstacles, sea breeze.

CAPE is weaker for maritime than for continental tropical convection, but the onset of convection is easier in the maritime case due to smaller CIN.

We have neglected entrainment of environment air, and detrainment from the air parcel, which generally tend to slow down convection. However, the parcels reaching the highest altitude are generally coming from the region below the cloud without being too much diluted.

Convectively generated clouds are not the only type of clouds. Low level stratiform clouds and high altitude cirrus are a large part of cloud cover and play an important role in the Earth radiative budget. However convection is responsible of the strongest precipitations, especially in the Tropics, and hence of most of atmospheric heating by latent heat transfer.

So far we have discussed the stability behaviour for a single air parcel. There may be the case that although the air parcel is stable within its layer, the layer as a whole may be destabilized if lifted. Such case happen when a strong vertical stratification of water vapour is present, so that the lower levels of the layer are much moister than the upper ones. If the layer is lifted, its lower levels will reach saturation before the uppermost ones, and start cooling at the slower pseudoadiabatic rate, while the upper layers will still cool at the faster adiabatic rate. Hence, the top part of the layer cools much more rapidly of the bottom part and the lapse rate of the layer becomes unstable. This *potential (or convective) instability* is frequently encountered in the lower levels in the Tropics, where there is a strong water vapour vertical gradient.

It can be shown that condition for a layer to be potentially unstable is that its equivalent potential temperature  $\theta_e$  decreases within the layer.

### 3.5 Tephigrams

To represent the vertical structure of the atmosphere and interpret its state, a number of diagrams is commonly used. The most common are *emagrams*, *Stüve diagrams*, *skew T- log p diagrams*, and *tephigrams*.

An *emagram* is basically a  $T$ - $z$  plot where the vertical axis is  $\log p$  instead of height  $z$ . But since  $\log p$  is linearly related to height in a dry, isothermal atmosphere, the vertical coordinate is basically the geometric height.

In the *Stüve diagram* the vertical coordinate is  $p^{(R/c_p)}$  and the horizontal coordinate is  $T$ : with this axes choice, the dry adiabats are straight lines.

A *skew  $T$ - $\log p$  diagram*, like the emagram, has  $\log p$  as vertical coordinate, but the isotherms are slanted. *Tephigrams* look very similar to skew  $T$  diagrams if rotated by  $45^\circ$ , have  $T$  as horizontal and  $\log \theta$  as vertical coordinates so that isotherms are vertical and the isentropes horizontal (hence tephigram, a contraction of  $T$  and  $\Phi$ , where  $\Phi = c_p \log \theta$  stands for the entropy). Often, tephigrams are rotated by  $45^\circ$  so that the vertical axis corresponds to the vertical in the atmosphere.

A tephigram is shown in figure 5: straight lines are isotherms (slope up and to the right) and isentropes (up and to the left), isobars (lines of constant  $p$ ) are quasi-horizontal lines, the dashed lines sloping up and to the right are constant mixing ratio in  $\text{g/kg}$ , while the curved solid bold lines sloping up and to the left are saturated adiabats.

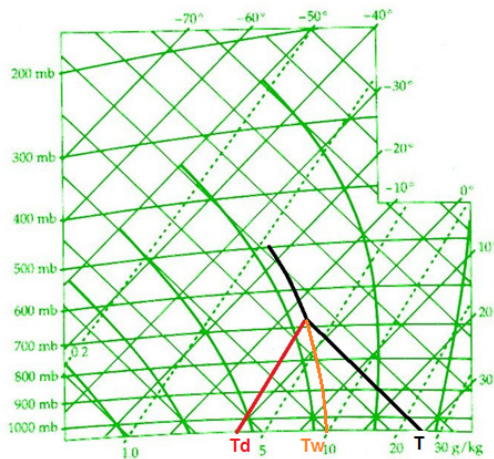


Fig. 5. A tephigram. Starting from the surface, the red line depicts the evolution of the Dew Point temperature, the black line depicts the evolution of the air parcel temperature, upon uplifting. The two lines intersect at the LCL. The orange line depicts the saturated adiabat crossing the LCL point, that defines the wet bulb temperature at the ground pressure surface.

Two lines are commonly plotted on a tephigram – the temperature and dew point, so the state of an air parcel at a given pressure is defined by its temperature  $T$  and  $T_d$ , that is its water vapour content. We note that the knowledge of these parameters allows to retrieve all the other humidity parameters: from the dew point and pressure we get the humidity mixing ratio  $w$ ; from the temperature and pressure we get the saturated mixing ratio  $w_s$ , and relative humidity may be derived from  $100 \cdot w/w_s$ , when  $w$  and  $w_s$  are measured at the same pressure.

When the air parcel is lifted, its temperature  $T$  follows the dry adiabatic lapse rate and its dew point  $T_d$  its constant vapour mixing ratio line - since the mixing ratio is conserved in

unsaturated air - until the two meet at the LCL where condensation may start to happen. Further lifting follows the Saturated Adiabatic Lapse Rate. In Figure 5 we see an air parcel initially at ground level, with a temperature of  $30^\circ$  and a Dew Point temperature of  $0^\circ$  (which as we can see by inspecting the diagram, corresponds to a mixing ratio of approx. 4 g/kg at ground level) is lifted adiabatically to 700 mb which is its LCL where the air parcel temperature following the dry adiabats meets the air parcel dew point temperature following the line of constant mixing ratio. Above 700 mb, the air parcel temperature follows the pseudoadiabat. Figure 5 clearly depicts the *Normand's rule*: The dry adiabatic through the temperature, the mixing ratio line through the dew point, and the saturated adiabatic through the wet bulb temperature, meet at the LCL. In fact, the saturated adiabat that crosses the LCL is the same that intersect the surface isobar exactly at the wet bulb temperature, that is the temperature a wetted thermometer placed at the surface would attain by evaporating - at constant pressure - its water inside its environment until it gets saturated.

Figure 6 reports two different temperature soundings: the black dotted line is the dew point profile and is common to the two soundings, while the black solid line is an early morning sounding, where we can see the effect of the nocturnal radiative cooling as a temperature inversion in the lowermost layer of the atmosphere, between 1000 and 960 hPa. The state of the atmosphere is such that an air parcel at the surface has to be forcedly lifted to 940 hPa to attain saturation at the LCL, and forcedly lifted to 600 hPa before gaining enough latent heat of condensation to become warmer than the environment and positively buoyant at the LFB. The temperature of such air parcel is shown as a grey solid line in the graph.

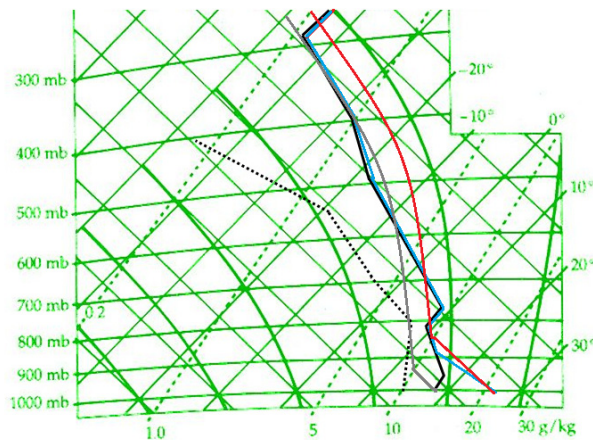


Fig. 6. A tephigram showing with the black and blue lines two different temperature soundings, and with the grey and red lines two different temperature histories of an air parcel initially at ground level, upon lifting. The dotted line is the common  $T_d$  profile of the two soundings.

The blue solid line is an afternoon sounding, when the surface has been radiatively heated by the sun. An air parcel lifted from the ground will follow the red solid line, and find itself immediately warmer than its environment and gaining positive buoyancy, further increased by the release of latent heat starting at the LCL at 850 hPa. Notice however that a

second inversion layer is present in the temperature sounding between 800 hPa and 750 hPa, such that the air parcel becomes colder than the environment, hence negatively buoyant between 800 hPa and 700 hPa. If forcedly uplifted beyond this stable layer, it again attains a positive buoyancy up to above 300 hPa.

As the tephigram is a graph of temperature against entropy, an area computed from these variables has dimensions of energy. The area between the air parcel path is then linked to the CIN and the CAPE. Referring to the early morning sounding, the area between the black and the grey line between the surface and 600 hPa is the CIN, the area between 600 hPa and 400 hPa is the CAPE.

#### 4. The generation of clouds

Clouds play a pivotal role in the Earth system, since they are the main actors of the atmospheric branch of the water cycle, promote vertical redistribution of energy by latent heat capture and release and strongly influence the atmospheric radiative budget.

Clouds may form when the air becomes supersaturated, as it can happen upon lifting as explained above, but also by other processes, as isobaric radiative cooling like in the formation of *radiative fogs*, or by mixing of warm moist air with cold dry air, like in the generation of airplane contrails and *steam fogs* above lakes.

*Cumulus* or *cumulonimbus* are classical examples of convective clouds, often precipitating, formed by reaching the saturation condition with the mechanism outlined hereabove.

Other types of clouds are *alto-cumulus* which contain liquid droplets between 2000 and 6000m in mid-latitudes and cluster into compact herds. They are often, during summer, precursors of late afternoon and evening developments of deep convection.

*Cirrus* are high altitude clouds composed of ice, rarely opaque. They form above 6000m in mid-latitudes and often promise a warm front approaching. Such clouds are common in the Tropics, formed as remains of anvils or by in situ condensation of rising air, up to the tropopause. *Nimbo-stratus* are very opaque low clouds of undefined base, associated with persistent precipitations and snow. *Strato-cumulus* are composed by water droplets, opaque or very opaque, with a cloud base below 2000m, often associated with weak precipitations.

*Stratus* are low clouds with small opacity, undefined base under 2000m that can even reach the ground, forming fog. Images of different types of clouds can be found on the Internet (see, as instance, <http://cimss.ssec.wisc.edu/satmet/gallery/gallery.html>).

In the following subchapters, a brief outline will be given on how clouds form in a saturated environment. The level of understanding of water cloud formation is quite advanced, while it is not so for ice clouds, and for glaciation processes in water clouds.

##### 4.1 Nucleation of droplets

We could think that the more straightforward way to form a cloud droplet would be by condensation in a saturated environment, when some water molecules collide by chance to form a cluster that will further grow to a droplet by picking up more and more molecules from the vapour phase. This process is termed *homogeneous nucleation*. The survival and further growth of the droplet in its environment will depend on whether the Gibbs free energy of the droplet and its surrounding will decrease upon further growth. We note that,

by creating a droplet, work is done not only as expansion work, but also to form the interface between the droplet and its environment, associated with the *surface tension* at the surface of the droplet of area  $A$ . This originates from the cohesive forces among the liquid molecules. In the interior of the droplet, each molecule is equally pulled in every direction by neighbouring molecules, resulting in a null net force. The molecules at the surface do not have other molecules on all sides of them and therefore are only pulled inwards, as if a force acted on interface toward the interior of the droplet. This creates a sort of pressure directed inward, against which work must be exerted to allow further expansion. This effect forces liquid surfaces to contract to the minimal area.

Let  $\sigma$  be the energy required to form a droplet of unit surface; then, for the heterogeneous system droplet-surroundings we may write, for an infinitesimal change of the droplet:

$$dG = -SdT + Vdp + (\mu_v - \mu_l)dm_v + \sigma dA \quad (46)$$

We note that  $dm_v = -dm_l = -n_l dV$  where  $n_l$  is the number density of molecules inside the droplet. Considering an isothermal-isobaric process, we came to the conclusion that the formation of a droplet of radius  $r$  results in a change of Gibbs free given by:

$$\Delta G = 4\pi r^2 \sigma - \frac{4}{3}\pi r^3 n_l K T \ln\left(\frac{e}{e_s}\right) \quad (47)$$

Where we have used (36). Clearly, droplet formation is thermodynamically unfavoured for  $e < e_s$ , as should be expected. If  $e > e_s$ , we are in supersaturated conditions, and the second term can counterbalance the first to give a negative  $\Delta G$ .

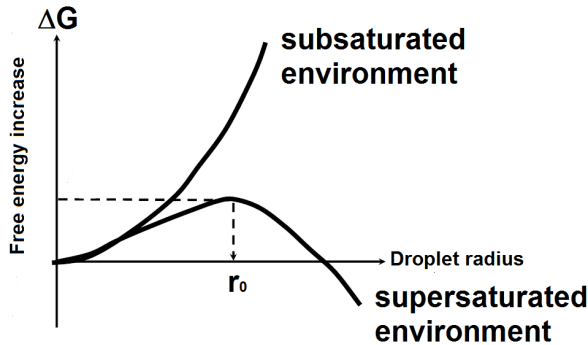


Fig. 7. Variation of Gibbs free energy of a pure water droplet formed by homogeneous nucleation, in a subsaturated (upper curve) and a supersaturated (lower curve) environment, as a function of the droplet radius. The critical radius  $r_0$  is shown.

Figure 7 shows two curves of  $\Delta G$  as a function of the droplet radius  $r$ , for a subsaturated and supersaturated environment. It is clear that below saturation every increase of the droplet radius will lead to an increase of the free energy of the system, hence is thermodynamically unfavourable and droplets will tend to evaporate. In the supersaturated case, on the contrary, a critical value of the radius exists, such that droplets that grows by casual collision among molecules beyond that value, will continue to grow: they are said to get *activated*. The expression for such *critical radius* is given by the *Kelvin's formula*:

$$r_0 = \frac{2\sigma}{n_l K T \ln\left(\frac{e}{e_s}\right)} \quad (48)$$

The greater  $e$  with respect to  $e_s$ , that is the degree of supersaturation, the smaller the radius beyond which droplets become activated.

It can be shown from (48) that a droplet with a radius as small as 0.01  $\mu\text{m}$  would require a supersaturation of 12% for getting activated. However, air is seldom more than a few percent supersaturated, and the homogeneous nucleation process is thus unable to explain the generation of clouds. Another process should be invoked: the *heterogeneous nucleation*. This process exploits the ubiquitous presence in the atmosphere of particles of various nature (Kaufman et al., 2002), some of which are soluble (hygroscopic) or wettable (hydrophilic) and are called *Cloud Condensation Nuclei (CCN)*. Water may form a thin film on wettable particles, and if their dimension is beyond the critical radius, they form the nucleus of a droplet that may grow in size. Soluble particles, like sodium chloride originating from sea spray, in presence of moisture absorb water and dissolve into it, forming a droplet of solution. The saturation vapour pressure over a solution is smaller than over pure water, and the fractional reduction is given by *Raoult's law*:

$$f = \frac{e'}{e} \quad (49)$$

Where  $e$  is the vapour pressure over pure water, and  $e'$  is the vapour pressure over a solution containing a *mole fraction*  $f$  (number of water moles divided by the total number of moles) of pure water.

Let us consider a droplet of radius  $r$  that contains a mass  $m$  of a substance of molecular weight  $M_s$  dissolved into  $i$  ions per molecule, such that the effective number of moles in the solution is  $im/M_s$ . The number of water moles will be  $((4/3)\pi r^3 \rho - m)/M_w$  where  $\rho$  and  $M_w$  are the water density and molecular weight respectively. The water mole fraction  $f$  is:

$$f = \frac{\frac{(4/3)\pi r^3 \rho - m}{M_w}}{\frac{(4/3)\pi r^3 \rho - m}{M_w} + \frac{im}{M_s}} = \left(1 + \frac{imM_w}{M_s((4/3)\pi r^3 \rho - m)}\right)^{-1} \quad (50)$$

Eq. (49) and (50) allows us to express the reduced value  $e'$  of the saturation vapour pressure for a droplet of solution. Using this result into (48) we can compute the saturation vapour pressure in equilibrium with a droplet of solution of radius  $r$ :

$$\frac{e'}{e_s} = \exp\left(\frac{2\sigma}{n_l K T r}\right) \left(1 + \frac{imM_w}{M_s((4/3)\pi r^3 \rho - m)}\right)^{-1} \quad (51)$$

The plot of supersaturation  $e'/e_s - 1$  for two different values of  $m$  is shown in fig. 8, and is named *Köhler curve*.

Figure 8 clearly shows how the amount of supersaturation needed to sustain a droplet of solution of radius  $r$  is much lower than what needed for a droplet of pure water, and it decreases with the increase of solute concentration. Consider an environment supersaturation of 0.2%. A droplet originated from condensation on a sphere of sodium chloride of diameter 0.1  $\mu\text{m}$  can grow indefinitely along the blue curve, since the peak of the curve is below the environment supersaturation; such droplet is *activated*. A droplet originated from a smaller grain of sodium chloride of 0.05  $\mu\text{m}$  diameter will grow until

when the supersaturation adjacent to it is equal to the environmental: attained that maximum radius, the droplet stops its grow and is in stable equilibrium with the environment. Such *haze dropled* is said to be *unactivated*.

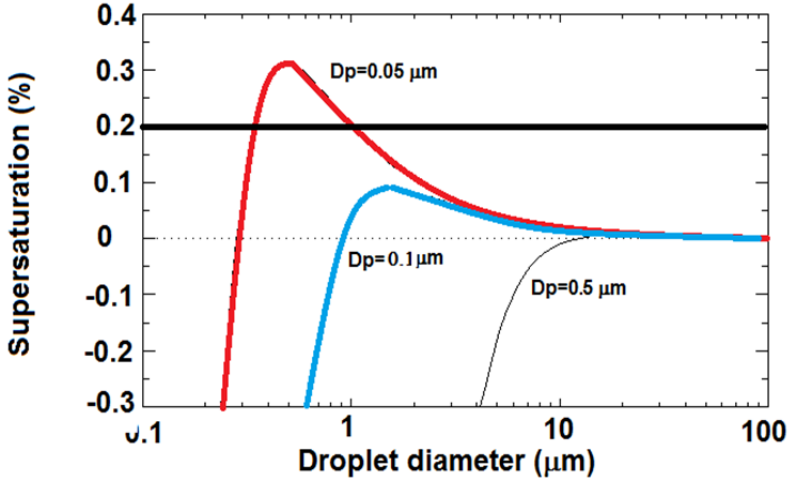


Fig. 8. Kohler curves showing how the critical diameter and supersaturation are dependent upon the amount of solute. It is assumed here that the solute is a perfect sphere of sodium chloride (source: [http://en.wikipedia.org/wiki/Köhler\\_theory](http://en.wikipedia.org/wiki/Köhler_theory)).

#### 4.2 Condensation

The droplet that is able to pass over the peak of the Köhler curve will continue to grow by condensation. Let us consider a droplet of radius  $r$  at time  $t$ , in a supersaturated environment whose water vapour density far from the droplet is  $\rho_v(\infty)$ , while the vapour density in proximity of the droplet is  $\rho_v(r)$ . The droplet mass  $M$  will grow at the rate of mass flux across a sphere of arbitrary radius centred on the droplet. Let  $D$  be the diffusion coefficient, that is the amount of water vapour diffusing across a unit area through a unit concentration gradient in unit time, and  $\rho_v(x)$  the water vapour density at a distance  $x > r$  from the droplet. We will have:

$$\frac{dM}{dt} = 4\pi x^2 D \frac{d\rho_v(x)}{dx} \quad (52)$$

Since in steady conditions of mass flow this equation is independent of  $x$ , we can integrate it for  $x$  between  $r$  and  $\infty$  to get:

$$\frac{dM}{dt} \int_r^\infty \frac{dx}{x^2} = \int_{\rho_v(r)}^{\rho_v(\infty)} d\rho_v(x) \quad (53)$$

Or, expliciting  $M$  as  $(4/3)\pi r^3 \rho_l$ :

$$\frac{dr}{dt} = \frac{D}{r \rho_l} (\rho_v(\infty) - \rho_v(r)) = \frac{D \rho_v(\infty)}{r \rho_l e(\infty)} (e(\infty) - e(r)) \quad (54)$$



Where we have used the ideal gas equation for water vapour. We should think of  $e(r)$  as given by  $e'$  in (49), but in fact we can approximate it with the saturation vapour pressure over a plane surface  $e_s$ , and pose  $(e(\infty)-e(r))/e(\infty)$  roughly equal to the supersaturation  $S=(e(\infty)-e_s)/e_s$  to come to:

$$\frac{dr}{dt} r = \frac{D\rho_v(\infty)}{\rho_l} S \quad (55)$$

This equation shows that the radius growth is inversely proportional to the radius itself, so that the rate of growth will tend to slow down with time. In fact, condensation alone is too slow to eventually produce rain droplets, and a different process should be invoked to create droplet with radius greater than few tens of micrometers.

### 4.3 Collision and coalescence

The droplet of density  $\rho_l$  and volume  $V$  is suspended in air of density  $\rho$  so that under the effect of the gravitational field, three forces are acting on it: the gravity exerting a downward force  $\rho_l Vg$ , the upward Archimede's buoyancy  $\rho V$  and the *drag force* that for a sphere, assumes the form of the *Stokes' drag*  $6\pi\eta rv$  where  $\eta$  is the viscosity of the air and  $v$  is the steady state terminal fall speed of the droplet. In steady state, by equating those forces and assuming the droplet density much greater than the air, we get an expression for the terminal fall speed:

$$v = \frac{2}{9} \frac{\rho_l g r^2}{\eta} \quad (56)$$

Such speed increases with the droplet dimension, so that bigger droplets will eventually collide with the smaller ones, and may entrench them with a *collection efficiency*  $E$  depending on their radius and other environmental parameters, as for instance the presence of electric fields. The rate of increase of the radius  $r_1$  of a spherical collector drop due to collision with water droplets in a cloud of *liquid water content*  $w_l$ , that is is the mass density of liquid water in the cloud, is given by:

$$\frac{dr}{dt} = \frac{(v_1-v_2)w_l E}{4\rho_l} \cong \frac{v_1 w_l E}{4\rho_l} \cong \quad (57)$$

Since  $v_1$  increases with  $r_1$ , the process tends to speed up until the collector drops became a rain drop and eventually pass through the cloud base, or split up to reinitiate the process.

### 4.4 Nucleation of ice particles

A cloud above  $0^\circ$  is said a *warm cloud* and is entirely composed of water droplets. Water droplet can still exists in *cold clouds* below  $0^\circ$ , although in an unstable state, and are termed *supecooled*. If a cold cloud contains both water droplets and ice, is said *mixed cloud*; if it contains only ice, it is said *glaciated*.

For a droplet to freeze, a number of water molecules inside it should come together and form an *ice embryo* that, if exceeds a critical size, would produce a decrease of the Gibbs free energy of the system upon further growing, much alike the homogeneous condensation from the vapour phase to form a droplet. This glaciations process is termed *homogeneous freezing*, and below roughly  $-37^\circ\text{C}$  is virtually certain to occur. Above that temperature, the

critical dimensions of the ice embryo are several micrometers, and such process is not favoured. However, the droplet can contain impurities, and some of them may promote collection of water droplets into an ice-like structure to form a ice-like embryo with dimension already beyond the critical size for glaciations. Such particles are termed *ice nuclei* and the process they start is termed *heterogeneous freezing*. Such process can start not only within the droplet, but also upon contact of the ice nucleus with the surface of the droplet (*contact nucleation*) or directly by deposition of ice on it from the water vapour phase (*deposition nucleation*). Good candidates to act as ice nuclei are those particle with molecular structure close to the hexagonal ice crystallography. Some soil particles, some organics and even some bacteria are effective nucleators, but only one out of  $10^3$ - $10^5$  atmospheric particles can act as an ice nucleus. Nevertheless ice particles are present in clouds in concentrations which are orders of magnitude greater than the presence of ice nuclei. Hence, ice multiplication processes must be at play, like breaking of ice particles upon collision, to create ice splinterings that enhance the number of ice particles.

#### 4.5 Growth of ice particles

Ice particles can grow from the vapour phase as in the case of water droplets. In a mixed phase cloud below  $0^\circ\text{C}$ , a much greater supersaturation is reached with respect to ice that can reach several percents, than with respect to water, which hardly exceed 1%. Hence ice particles grows faster than droplets and, since this deplete the vapour phase around them, it may happen that around a growing ice particle, water droplets evaporate. Ice can form in a variety of shapes, whose basic habits are determined by the temperature at which they grow. Another process of growth in a mixed cloud is by *riming*, that is by collision with supercooled droplets that freeze onto the ice particle. Such process is responsible of the formation of hailstones.

A process effective in cold clouds is the *aggregation* of ice particles between themselves, when they have different shapes and/or dimension, hence different fall speeds.

#### 5. Conclusion

A brief overview of some topic of relevance in atmospheric thermodynamic has been provided, but much had to remain out of the limits of this introduction, so the interested reader is encouraged to further readings. For what concerns moist thermodynamics and convection, the reader can refer to chapters in introductory atmospheric science textbooks like the classical Wallace and Hobbs (2006), or Salby (1996). At a higher level of deepening the classical reference is Iribarne and Godson (1973). For the reader who seeks a more theoretical approach, Zdunkowski and Bott (2004) is a good challenge. Convection is thoughtfully treated in Emmanuel (1994) while a sound review is given in the article of Stevens (2005). For what concerns the microphysics of clouds, the reference book is Pruppacher and Klett (1996). A number of seminal journal articles dealing with the thermodynamics of the general circulation of the atmosphere can be cited: Goody (2003), Pauluis and Held (2002), Renno and Ingersoll (2008), Pauluis et al. (2008) and references therein. Finally, we would like to suggest the Bohren (2001) delightful book, for which a scientific or mathematical background is not required, that explores topics in meteorology and basic physics relevant to the atmosphere.

## 6. References

- Bohren, C. F., (2001), *Clouds in a Glass of Beer: Simple Experiments in Atmospheric Physics*, John Wiley & Sons, Inc., New York.
- Bolton, M.D., (1980), The computation of equivalent potential temperature, *Mon. Wea. Rev.*, 108, 1046-1053.
- Emanuel, K., (1984), *Atmospheric Convection*, Oxford Univ. Press, New York.
- Fermi, E., (1956), *Thermodynamics*, Dover Publications, London.
- Goff, J. A., (1957), Saturation pressure of water on the new Kelvin temperature scale, *Transactions of the American society of heating and ventilating engineers*, pp. 347-354, meeting of the American Society of Heating and Ventilating Engineers, Murray Bay, Quebec, Canada, 1957.
- Goody, R. (2003), On the mechanical efficiency of deep, tropical convection, *J. Atmos. Sci.*, 50, 2287-2832.
- Hyland, R. W. & A. Wexler A., (1983), Formulations for the Thermodynamic Properties of the saturated Phases of H<sub>2</sub>O from 173.15 K to 473.15 K, *ASHRAE Trans.*, 89(2A), 500-519.
- Iribarne J. V. & Godson W. L., (1981), *Atmospheric Thermodynamics*, Springer, London.
- Kaufman Y. J., Tanrè D. & O. Boucher, (2002), A satellite view of aerosol in the climate system, *Nature*, 419, 215-223.
- Landau L. D. & Lifshitz E. M., (1980), *Statistical Physics*, Plenum Press, New York.
- Marti, J. & Mauersberger K., (1993), A survey and new measurements of ice vapor pressure at temperatures between 170 and 250 K, *Geophys. Res. Lett.* , 20, 363-366.
- Murphy, D. M. & Koop T., (2005), Review of the vapour pressures of ice and supercooled water for atmospheric applications, *Quart. J. Royal Met. Soc.*, 131, 1539-1565.
- Murray, F. W., (1967), On the computation of saturation vapor pressure, *J. Appl. Meteorol.*, 6, 203-204, 1967.
- Pauluis, O; & Held, I.M. (2002). Entropy budget of an atmosphere in radiative-convective equilibrium. Part I: Maximum work and frictional dissipation, *J. Atmos. Sci.*, 59, 140-149.
- Pauluis, O., Czaja A. & Korty R. (2008). The global atmospheric circulation on moist isentropes, *Science*, 321, 1075-1078.
- Pruppacher H. D. & Klett, J. D., (1996), *Microphysics of clouds and precipitation*, Springer, London.
- Renno, N. & Ingersoll, A. (1996). Natural convection as a heat engine: A theory for CAPE, *J. Atmos. Sci.*, 53, 572-585.
- Salby M. L., (1996), *Fundamentals of Atmospheric Physics*, Academic Press, New York.
- Sonntag, D., (1994), Advancements in the field of hygrometry, *Meteorol. Z., N. F.*, 3, 51-66.
- Stevens, B., (2005), Atmospheric moist convection, *Annu. Rev. Earth. Planet. Sci.*, 33, 605-643.
- Wallace J.M & Hobbs P.V., (2006), *Atmospheric Science: An Introductory Survey*, Academic Press, New York.

Zdunkowski W. & Bott A., (2004), *Thermodynamics of the Atmosphere: A Course in Theoretical Meteorology*, Cambridge University Press, Cambridge.

# Thermodynamic Aspects of Precipitation Efficiency

Xinyong Shen<sup>1</sup> and Xiaofan Li<sup>2</sup>

<sup>1</sup>*Key Laboratory of Meteorological Disaster of Ministry of Education  
Nanjing University of Information Science and Technology*

<sup>2</sup>*NOAA/NESDIS/Center for Satellite Applications and Research*

<sup>1</sup>*China*

<sup>2</sup>*USA*

## 1. Introduction

Precipitation efficiency is one of important meteorological parameters and has been widely used in operational precipitation forecasts (e.g., Doswell et al., 1996). Precipitation efficiency has been defined as the ratio of precipitation rate to the sum of all precipitation sources from water vapor budget (e.g., Auer and Marwitz, 1968; Heymsfield and Schotz, 1985; Chong and Hauser, 1989; Dowell et al., 1996; Ferrier et al., 1996; Li et al., 2002; Sui et al., 2005) after Braham (1952) calculated precipitation efficiency with the inflow of water vapor into the storm through cloud base as the rainfall source more than half century ago. Sui et al. (2007) found that the estimate of precipitation efficiency with water vapor process data can be more than 100% or negative because some rainfall sources are excluded or some rainfall sinks are included. They defined precipitation efficiency through the inclusion of all rainfall sources and the exclusion of all rainfall sinks from surface rainfall budget derived by Gao et al. (2005), which fixed precipitation efficiency to the normal range of 0-100%.

In additional to water vapor processes, thermal processes also play important roles in the development of rainfall since precipitation is determined by environmental thermodynamic conditions via cloud microphysical processes. The water vapor convergence and heat divergence and its forced vapor condensation and depositions in the precipitation systems could be major sources for precipitation while these water vapor and cloud processes could give some feedback to the environment. Gao et al. (2005) derived a water vapor related surface rainfall budget through the combination of cloud budget with water vapor budget. Gao and Li (2010) derived a thermally related surface rainfall budget through the combination of cloud budget with heat budget. In this chapter, precipitation efficiency is defined from the thermally related surface rainfall budget (*PEH*) and is calculated using the data from the two-dimensional (2D) cloud-resolving model simulations of a pre-summer torrential rainfall event over southern China in June 2008 (Wang et al., 2010; Shen et al., 2011a, 2011b) and is compared with the precipitation efficiency defined from water vapor related surface rainfall budget (Sui et al., 2007) to study the efficiency in thermodynamic aspect of the pre-summer heavy rainfall system.

The impacts of ice clouds on the development of convective systems have been intensively studied through the analysis of cloud-resolving model simulations (e.g., Yoshizaki, 1986;

Nicholls, 1987; Fovell and Ogura, 1988; Tao and Simpson, 1989; McCumber et al., 1991; Tao et al., 1991; Liu et al., 1997; Grabowski et al., 1999; Wu et al., 1999; Li et al., 1999; Grabowski and Moncrieff, 2001; Wu, 2002; Grabowski, 2003; Gao et al., 2006; Ping et al., 2007). Wang et al. (2010) studied microphysical and radiative effects of ice clouds on a pre-summer heavy rainfall event over southern China during 3-8 June 2008 through the analysis of sensitivity experiments and found that microphysical and radiative effects of ice clouds play equally important roles in the pre-summer heavy rainfall event. The total exclusion of ice microphysics decreased model domain mean surface rain rate primarily through the weakened convective rainfall caused by the exclusion of radiative effects of ice clouds in the onset phase and through the weakened stratiform rainfall caused by the exclusion of ice microphysical effects in the development and mature phases, whereas it increased the mean rain rate through the enhanced convective rainfall caused by the exclusion of ice microphysical effects in the decay phase. Thus, effects of ice clouds on precipitation efficiencies are examined through the analysis of the pre-summer heavy rainfall event in this chapter. Precipitation efficiency is defined in section 2. Pre-summer heavy rainfall event, model, and sensitivity experiments are described in section 3. The control experiment is discussed in section 4. Radiative and microphysical effects of ice clouds on precipitation efficiency and associated rainfall processes are respectively examined in sections 5 and 6. The conclusions are given in section 7.

## 2. Definitions of precipitation efficiency

The budgets for specific humidity ( $q_v$ ), temperature ( $T$ ), and cloud hydrometeor mixing ratio ( $q_l$ ) in the 2D cloud resolving model used in this study can be written as

$$\frac{\partial q_v}{\partial t} = -\frac{\partial(u'q_v')}{\partial x} - \bar{u}^o \frac{\partial q_v'}{\partial x} - \bar{w}^o \frac{\partial q_v'}{\partial z} - w' \frac{\partial \bar{q}_v}{\partial z} - \frac{1}{\rho} \frac{\partial}{\partial z} (\rho w' q_v') - S_{qv} - \bar{u}^o \frac{\partial \bar{q}_v}{\partial x} - \bar{w}^o \frac{\partial \bar{q}_v}{\partial z} \quad (1a)$$

$$\frac{\partial T}{\partial t} = -\frac{\partial}{\partial x} (\bar{u}^o + u') T' - \pi \bar{u}^o \frac{\partial \bar{\theta}^o}{\partial x} - \pi \bar{w}^o \frac{\partial}{\partial z} (\bar{\theta} + \theta') - \pi w' \frac{\partial \bar{\theta}}{\partial z} - \frac{\pi}{\bar{\rho}} \frac{\partial}{\partial z} (\bar{\rho} w' \theta') + \frac{Q_{ch}}{c_p} + \frac{Q_R}{c_p} \quad (1b)$$

$$\frac{\partial q_l}{\partial t} = -\frac{\partial(uq_l)}{\partial x} - \frac{1}{\rho} \frac{\partial}{\partial z} (\rho w q_l) + \frac{1}{\rho} \frac{\partial}{\partial z} (\rho (w_{Tr} q_r + w_{Ts} q_s + w_{Tg} q_g)) + S_{qv} \quad (1c)$$

where

$$S_{qv} = P_{CND} + P_{DEP} + P_{SDEP} + P_{GDEP} - P_{REVP} - P_{MLTS} - P_{MLTG} \quad (2a)$$

$$Q_{ch} = L_v S_{qv} + L_f P_{18} \quad (2b)$$

$\theta$  is potential temperature;  $u$  and  $w$  are zonal and vertical components of wind, respectively;  $\rho$  is air density that is a function of height;  $c_p$  is the specific heat of dry air at constant pressure;  $L_v$ ,  $L_s$ , and  $L_f$  are latent heat of vaporization, sublimation, and fusion at  $T_o=0^\circ\text{C}$ , respectively,  $L_s=L_v+L_f$ ;  $T_{ov}=-35^\circ\text{C}$ ; and cloud microphysical processes in (2) can be found in Gao and Li (2008).  $Q_R$  is the radiative heating rate due to the convergence of net flux of solar and IR radiative fluxes.  $w_{Tr}$ ,  $w_{Ts}$ , and  $w_{Tg}$  in (1c) are terminal velocities for raindrops, snow,

and graupel, respectively; overbar denotes a model domain mean; prime is a perturbation from model domain mean; and superscript  $^{\circ}$  is an imposed observed value. The comparison between (1) and (2) shows that the net condensation term ( $S_{qv}$ ) links water vapor, heat, and cloud budgets.

$$\begin{aligned}
 P_{18} = & P_{DEP} + P_{SDEP} + P_{GDEP} - P_{MLTS} - P_{MLTG} \\
 & + P_{SACW}(T < T_o) + P_{SFW}(T < T_o) + P_{GACW}(T < T_o) \\
 & + P_{IACR}(T < T_o) + P_{GACR}(T < T_o) + P_{SACR}(T < T_o) \\
 & + P_{GFR}(T < T_o) - P_{RACS}(T > T_o) - P_{SMLT}(T > T_o) \\
 & - P_{GMLT}(T > T_o) + P_{IHOM}(T < T_{oo}) - P_{IMLT}(T > T_o) \\
 & + P_{IDW}(T_{oo} < T < T_o),
 \end{aligned} \tag{2c}$$

Following Gao et al. (2005) and Sui and Li (2005), the cloud budget (1c) and water vapor budget (1a) are mass integrated and their budgets can be, respectively, written as

$$P_S - Q_{CM} = Q_{WVS} = Q_{WVOUT} + Q_{WVIN} \tag{3}$$

$$Q_{WVT} + Q_{WVF} + Q_{WVE} = Q_{WVS} \tag{4}$$

where

$$P_S = P_r + P_s + P_g \tag{5a}$$

$$P_r = \overline{\rho w_T q_r} |_{z=0}, \tag{5b}$$

$$P_s = \overline{\rho w_T q_s} |_{z=0}, \tag{5c}$$

$$P_g = \overline{\rho w_T q_g} |_{z=0}, \tag{5d}$$

$$Q_{CM} = -\frac{\partial [q_l]}{\partial t} - [u \frac{\partial q_l}{\partial x}] - [w \frac{\partial q_l}{\partial z}] \tag{5e}$$

$$Q_{WVOUT} = [P_{CND}] + [P_{DEP}] + [P_{SDEP}] + [P_{GDEP}] \tag{5f}$$

$$Q_{WVIN} = -[P_{REVP}] - [P_{MLTG}] - [P_{MLTS}] \tag{5g}$$

$$Q_{WVT} = -\frac{\partial [q_v]}{\partial t} \tag{5h}$$

$$Q_{WVF} = -[\bar{u}^{\circ} \frac{\partial \bar{q}_v}{\partial x}] - [\bar{w}^{\circ} \frac{\partial \bar{q}_v}{\partial z}] - [\frac{\partial (\bar{u}' q_v')}{\partial x}] - [\bar{u}^{\circ} \frac{\partial q_v'}{\partial x}] - [\bar{w}^{\circ} \frac{\partial q_v'}{\partial z}] - [w' \frac{\partial \bar{q}_v'}{\partial z}] \tag{5i}$$

$$Q_{WVE} = E_s \tag{5j}$$

Here,  $P_s$  is precipitation rate, and in the tropics,  $P_s=0$  and  $P_g=0$ ,  $P_s=P_r$ ;  $E_s$  is surface evaporation;  $[0]=\int_{z_b}^{z_t} \bar{\rho} dz$ ,  $z_t$  and  $z_b$  are the heights of the top and bottom of the model atmosphere, respectively.

The heat budget (1b) is mass integrated and can be written as

$$S_{HT} + S_{HF} + S_{HS} + S_{LHLF} + S_{RAD} = Q_{WVS} \quad (6)$$

where

$$S_{HT} = \frac{c_p}{L_v} \frac{\partial [T]}{\partial t} \quad (7a)$$

$$S_{HF} = \frac{c_p}{L_v} \left[ \frac{\partial}{\partial x} (\bar{u}^o + u') T' + \pi \bar{u}^o \frac{\partial \bar{\theta}^o}{\partial x} + \pi \bar{w}^o \frac{\partial}{\partial z} (\bar{\theta} + \theta') + \pi w' \frac{\partial \bar{\theta}}{\partial z} \right] \quad (7b)$$

$$S_{HS} = -\frac{c_p}{L_v} H_s \quad (7c)$$

$$S_{LHLF} = -\frac{L_f}{L_v} < P_{18} > \quad (7d)$$

$$S_{RAD} = -\frac{1}{L_v} < Q_R > \quad (7e)$$

$H_s$  is surface sensible heat flux.

The equations (3), (4), and (6) indicate that the surface rain rate ( $P_s$ ) is associated with favorable environmental water vapor and thermal conditions through cloud microphysical processes ( $Q_{WVOUT}+Q_{WVIN}$ ). Following Gao and Li (2010), the cloud budget (3) and water vapor budget (4) are combined by eliminating  $Q_{WVOUT}+Q_{WVIN}$  to derive water vapor related surface rainfall equation ( $P_{SWV}$ ),

$$P_{SWV} = Q_{WVT} + Q_{WVF} + Q_{WVE} + Q_{CM} \quad (8a)$$

In (8a), the surface rain rate ( $P_{SWV}$ ) is associated with local atmospheric drying ( $Q_{WVT} > 0$ )/moistening ( $Q_{WVT} < 0$ ), water vapor convergence ( $Q_{WVF} > 0$ )/divergence ( $Q_{WVF} < 0$ ), surface evaporation ( $Q_{WVE}$ ), and decrease of local hydrometeor concentration/hydrometeor convergence ( $Q_{CM} > 0$ ) or increase of local hydrometeor concentration/hydrometeor divergence ( $Q_{CM} < 0$ ). Similarly, the cloud budget (3) and heat budget (6) are combined by eliminating  $Q_{WVOUT}+Q_{WVIN}$  to derive thermally related surface rainfall equation ( $P_{SH}$ ),

$$P_{SH} = S_{HT} + S_{HF} + S_{HS} + S_{LHLF} + S_{RAD} + Q_{CM}. \quad (8b)$$

In (8b), the surface rain rate ( $P_{SH}$ ) is related to local atmospheric warming ( $S_{HT} > 0$ )/cooling ( $S_{HT} < 0$ ), heat divergence ( $S_{HF} > 0$ )/convergence ( $S_{HF} < 0$ ), surface sensible heat ( $S_{HS}$ ), latent heat due to ice-related processes ( $S_{LHLF}$ ), radiative cooling ( $S_{RAD} > 0$ )/heating ( $S_{RAD} < 0$ ), and



decrease of local hydrometeor concentration/hydrometeor convergence ( $Q_{CM} > 0$ ) or increase of local hydrometeor concentration/hydrometeor divergence ( $Q_{CM} < 0$ ).  $P_{SWV} = P_{SH} = P_S$ .

From (8), precipitation efficiencies can be respectively defined as

$$PEWV = \frac{P_S}{\sum_{i=1}^4 H(Q_i)Q_i} \quad (9a)$$

$$PEH = \frac{P_S}{\sum_{i=1}^5 H(S_i)S_i + H(Q_{CM})Q_{CM}} \quad (9b)$$

where  $Q_i=(Q_{WVT}, Q_{WVF}, Q_{WVE}, Q_{CM})$ ;  $S_i=(S_{HT}, S_{HF}, S_{HS}, S_{LHLF}, S_{RAD})$ ;  $H$  is the Heaviside function,  $H(F)=1$  when  $F>0$ , and  $H(F)=0$  when  $F \leq 0$ . Large-scale heat precipitation efficiency ( $PEH$ ) is first introduced in this study, whereas large-scale water vapor precipitation efficiency ( $PEWV$ ) is exactly same to  $LSPE2$  defined by Sui et al. (2007).

### 3. Pre-summer rainfall case, model, and experiments

The pre-summer rainy season is the major rainy season over southern China, in which the rainfall starts in early April and reaches its peak in June (Ding, 1994). Although the rainfall is a major water resource in annual water budget, the torrential rainfall could occur during the pre-summer rainfall season and can lead to tremendous property damage and fatalities. In 1998, for instance, the torrential rainfall resulted in over 30 billion USD in damage and over 100 fatalities. Thus, many observational analyses and numerical modeling have been contributed to understanding of physical processes responsible for the development of pre-summer torrential rainfall (e.g., Krishnamurti et al., 1976; Tao and Ding, 1981; Wang and Li, 1982; Ding and Murakami, 1994; Simmonds et al., 1999). Recently, Wang et al. (2010) and Shen et al. (2011a, 2011b) conducted a series of sensitivity experiments of the pre-summer torrential rainfall occurred in the early June 2008 using 2D cloud-resolving model and studied effects of vertical wind shear, radiation, and ice clouds on the development of torrential rainfall. They found that these effects on torrential rainfall are stronger during the decay phase than during the onset and mature phases. During the decay phase of convection on 7 June 2008, the increase in model domain mean surface rain rate resulting from the exclusion of vertical wind shear is associated with the slowdown in the decrease of perturbation kinetic energy due to the exclusion of barotropic conversion from mean kinetic energy to perturbation kinetic energy. The increase in domain-mean rain rate resulting from the exclusion of cloud radiative effects is related to the enhancement of condensation and associated latent heat release as a result of strengthened radiative cooling. The increase in the mean surface rain rate is mainly associated with the increase in convective rainfall, which is in turn related to the local atmospheric change from moistening to drying. The increase in mean rain rate caused by the exclusion of ice clouds results from the increases in the mean net condensation and mean latent heat release caused by the strengthened mean radiative cooling associated with the removal of radiative effects of ice clouds. The increase

in mean rain rate caused by the removal of radiative effects of water clouds corresponds to the increase in the mean net condensation.

The pre-summer torrential rainfall event studied by Wang et al. (2010) and Shen et al. (2011a, 2011b) will be revisited to examine the thermodynamic aspects of precipitation efficiency and effects of ice clouds on precipitation efficiency. The cloud-resolving model (Soong and Ogura, 1980; Soong and Tao, 1980; Tao and Simpson, 1993) used in modeling the pre-summer torrential rainfall event in Wang et al. (2010) is the 2D version of the model (Sui et al., 1994, 1998) that was modified by Li et al. (1999). The model is forced by imposed large-scale vertical velocity and zonal wind and horizontal temperature and water vapor advections, which produces reasonable simulation through the adjustment of the mean thermodynamic stability distribution by vertical advection (Li et al., 1999). The modifications by Li et al. (1999) include: (1) the radius of ice crystal is increased from  $50\mu\text{m}$  (Hsie et al., 1980) to  $100\mu\text{m}$  (Krueger et al., 1995) in the calculation of growth of snow by the deposition and riming of cloud water, which yields a significant increase in cloud ice; (2) the mass of a natural ice nucleus is replaced by an average mass of an ice nucleus in the calculation of the growth of ice clouds due to the position of cloud water; (3) the specified cloud single scattering albedo and asymmetry factor are replaced by those varied with cloud and environmental thermodynamic conditions. Detailed descriptions of the model can be found in Gao and Li (2008). Briefly, the model includes prognostic equations for potential temperature and specific humidity, prognostic equations for hydrometeor mixing ratios of cloud water, raindrops, cloud ice, snow, and graupel, and perturbation equations for zonal wind and vertical velocity. The model uses the cloud microphysical parameterization schemes (Lin et al., 1983; Rutledge and Hobbs, 1983, 1984; Tao et al., 1989; Krueger et al., 1995) and solar and thermal infrared radiation parameterization schemes (Chou et al., 1991, 1998; Chou and Suarez, 1994). The model uses cyclic lateral boundaries, and a horizontal domain of 768 km with 33 vertical levels, and its horizontal and temporal resolutions are 1.5 km and 12 s, respectively.

The data from Global Data Assimilation System (GDAS) developed by the National Centers for Environmental Prediction (NCEP), National Oceanic and Atmospheric Administration (NOAA), USA are used to calculate the forcing data for the model over a longitudinally oriented rectangular area of  $108\text{--}116^\circ\text{E}$ ,  $21\text{--}22^\circ\text{N}$  over coastal areas along southern Guangdong and Guangxi Provinces and the surrounding northern South China Sea. The horizontal and temporal resolutions for NCEP/GDAS products are  $1^\circ\times 1^\circ$  and 6 hourly, respectively. The model is imposed by large-scale vertical velocity, zonal wind (Fig. 1), and horizontal temperature and water vapor advections (not shown) averaged over  $108\text{--}116^\circ\text{E}$ ,  $21\text{--}22^\circ\text{N}$ . The model is integrated from 0200 Local Standard Time (LST) 3 June to 0200 LST 8 June 2008 during the pre-summer heavy rainfall. The surface temperature and specific humidity from NCEP/GDAS averaged over the model domain are uniformly imposed on each model grid to calculate surface sensible heat flux and evaporation flux. The 6-hourly zonally-uniform large-scale forcing data are linearly interpolated into 12-s data, which are uniformly imposed zonally over model domain at each time step. The imposed large-scale vertical velocity shows the gradual increase of upward motions from 3 June to 6 June. The maximum upward motion of  $18\text{ cm s}^{-1}$  occurred around 9 km in the late morning of 6 June. The upward motions decreased dramatically on 7 June. The lower-tropospheric westerly winds of  $4\text{--}12\text{ m s}^{-1}$  were maintained during the rainfall event.

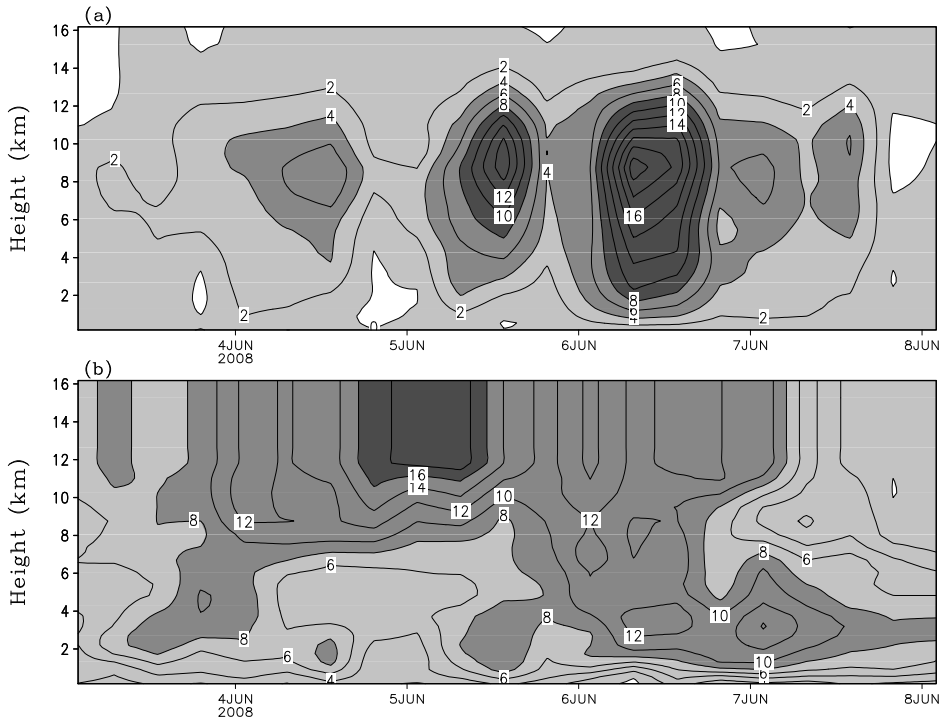


Fig. 1. Temporal and vertical distribution of (a) vertical velocity ( $\text{cm s}^{-1}$ ) and (b) zonal wind ( $\text{m s}^{-1}$ ) from 0200 LST 3 June – 0200 LST 8 June 2008. The data are averaged in a rectangular box of  $108\text{--}116^\circ\text{E}$ ,  $21\text{--}22^\circ\text{N}$  from NCEP/GDAS data. Ascending motion in (a) and westerly wind in (b) are shaded.

In the control experiment (C), the model is integrated with the initial vertical profiles of temperature and specific humidity from NCEP/GDAS at 0200 LST 3 June 2008. The model is integrated with the initial conditions and constant large-scale forcing at 0200 LST 3 June for 6 hours during the model spin-up period and the 6-hour model data are not used for analysis. The comparison in surface rain rate between the simulation and rain gauge observation averaged from 17 stations over southern Guangdong and Guangxi reveals a fair agreement with a gradual increase from 3–6 June and a rapid decrease from 6–7 June (Fig. 2). Their RMS difference ( $0.97 \text{ mm h}^{-1}$ ) is significantly smaller than the standard deviations of simulated ( $1.22 \text{ mm h}^{-1}$ ) and observed ( $1.26 \text{ mm h}^{-1}$ ) rain rates. The differences in surface rain rate between the simulation and observation can reach  $2 \text{ mm h}^{-1}$ , as seen in the previous studies (e.g., Li et al., 1999; Xu et al., 2007; Wang et al., 2009). The differences may partially be from the comparison of small hourly local sampling of rain gauge observations over 35% of model domain over land and no rain gauge observations over 65% of model domain over ocean with large model domain averages of model data in the control experiment with imposed 6-hourly large-scale forcing. The convection may be affected by land-ocean contrast and orography; these effects are included in the large-scale forcing imposed in the model.

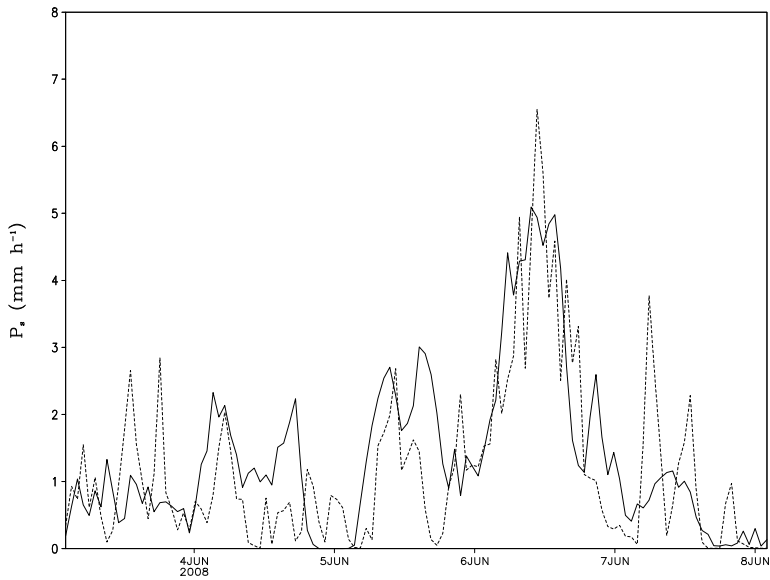


Fig. 2. Time series of surface rain rates ( $\text{mm h}^{-1}$ ) simulated in the control experiment (solid) and from rain gauge observation (dash).

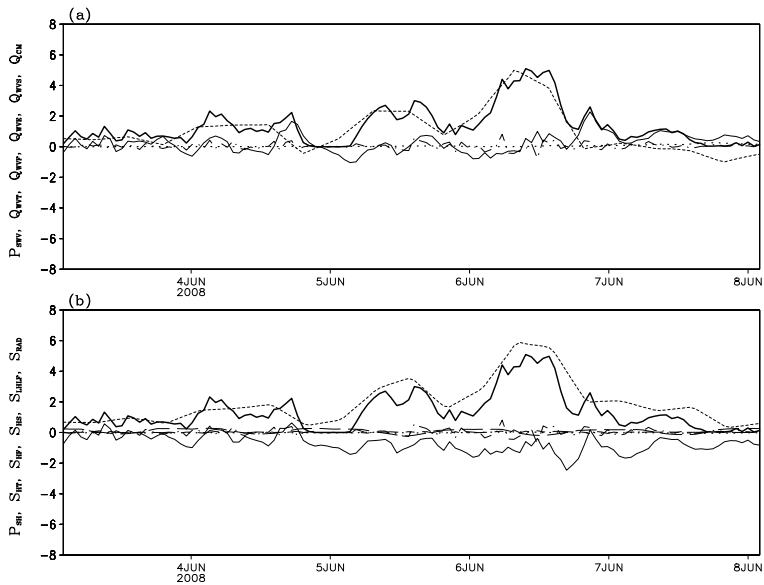


Fig. 3. Time series of model domain means of (a)  $P_{SWV}$  (dark solid),  $Q_{WVT}$  (light solid),  $Q_{WVF}$  (short dash),  $Q_{WVE}$  (dot),  $Q_{CM}$  (dot dash), and (b)  $P_{SH}$  (dark solid),  $S_{HT}$  (light solid),  $S_{HF}$  (short dash),  $S_{HS}$  (dot),  $S_{LHLF}$  (long short dash),  $S_{RAD}$  (long dash), and  $Q_{CM}$  (dot dash) in C. Unit is  $\text{mm h}^{-1}$ .

To investigate effects of ice clouds on precipitation efficiency, two sensitivity experiments are examined in this study. Experiment CNIR is identical to C except that the mixing ratios of ice hydrometeor are set to zero in the calculation of radiation. Experiment CNIR is compared with C to study radiative effects of ice clouds on rainfall responses to the large-scale forcing. Experiment CNIM is identical to C except in CNIM ice clouds (the ice hydrometeor mixing ratio and associated microphysical processes) are excluded. The comparison between CNIM and CNIR reveals impacts of the removal of microphysical efficient of ice clouds on rainfall responses to the large-scale forcing in the absence of radiative effects of ice clouds. The hourly model simulation data are used in the following discussions of this study.

#### 4. The control experiment: C

Model domain mean surface rain rate starts on 3 June 2008 with the magnitude of about  $1 \text{ mm h}^{-1}$  (Fig. 3), which corresponds to the weak upward motions with a maximum of  $2 \text{ cm s}^{-1}$  at 6-8 km (Fig. 1a). The rain rate increases to  $2 \text{ mm h}^{-1}$  as the upward motions increase up to over  $6 \text{ cm s}^{-1}$  on 4 June. When the upward motions weaken in the evening of 4 June and a weak downward motion occurs near the surface, the mean rainfall vanishes. As upward motions pick their strengths on 5 June, the mean rain rate intensifies (over  $2 \text{ mm h}^{-1}$ ). The mean rainfall reaches its peak on 6 June (over  $4 \text{ mm h}^{-1}$ ) as the upward motions have a maximum of over  $20 \text{ cm s}^{-1}$ . The upward motions rapidly weaken on 7 June, which leads to the significant reduction in the mean rainfall. Thus, four days (4, 5, 6, and 7 June) are defined as the onset, development, mature, and decay phases of the rainfall event, respectively. During 3-6 June, the mean rainfall is mainly associated with the mean water vapor convergence ( $Q_{WVF}>0$ ) in water vapor related surface rainfall budget and the mean heat divergence ( $S_{HF}>0$ ) in thermally related surface rainfall budget. Local atmospheric drying ( $Q_{WVT}>0$ ) and moistening ( $Q_{WVT}<0$ ) occur while the mean local atmospheric cooling ( $S_{HT}<0$ ) prevails. The mean hydrometeor loss/convergence ( $Q_{CM}$ ) has small hourly fluctuations. The mean radiative heating during the daytime and mean radiative cooling during the nighttime have the much smaller magnitudes than the mean heat divergence and the mean local heat change do in thermally related surface rainfall budget. On 7 June, the mean water vapor related surface rainfall budget shows that the rainfall is associated with local atmospheric drying while water vapor divergence prevails. The mean thermally related surface rainfall budget reveals that the rainfall is related to heat divergence while the heat divergence cools local atmosphere.

The calculation of precipitation efficiency using model domain mean model simulation data shows that  $PE_{WV}$  generally is higher than  $P_{WH}$  (Fig. 4a) because the rainfall source from the mean water vapor convergence in water vapor related surface rainfall budget is weaker than the rainfall source from the mean heat divergence in thermally related surface rainfall budget (Fig. 3). This suggests that the precipitation system is more efficient in the consumption of rainfall source from water vapor than in the consumption of the rainfall source from heat. The root-mean-squared (RMS) difference between  $PE_{WV}$  and  $PE_H$  is 24.4%. Both  $PE_{WV}$  and  $PE_H$  generally increase as surface rain rate increases (Fig. 5a). This indicates that the precipitation system generally is more efficient for high surface rain rates than for low surface rain rates. The ranges of  $PE_{WV}$  and  $PE_H$  are smaller when surface rain rate is higher than  $3 \text{ mm h}^{-1}$  (70-100%) than when surface rain rate is lower than  $3 \text{ mm h}^{-1}$  (0-100%).

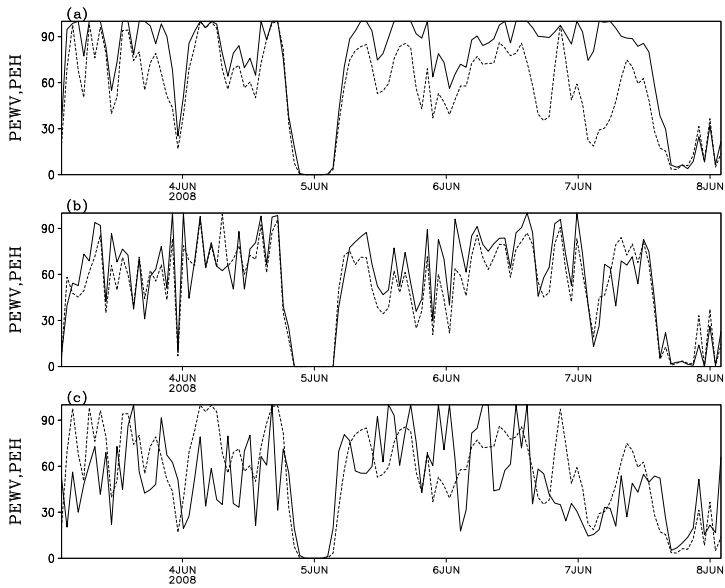


Fig. 4. Time series of  $PEWV$  (solid) and  $PEH$  (dash) calculated using (a) model domain mean data and data from (b) convective and (c) raining stratiform regions in C. Unit is %.

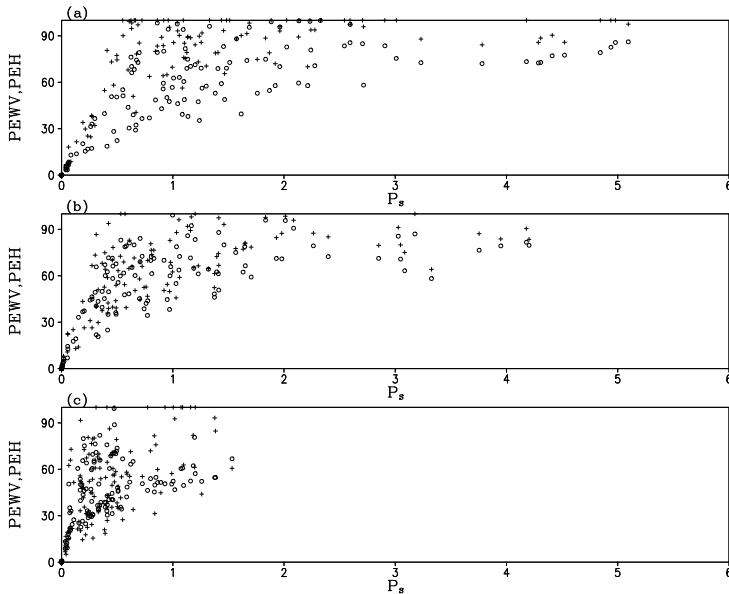


Fig. 5.  $PEWV$  versus  $P_s$  (cross) and  $PEH$  versus  $P_s$  (open circle) calculated using (a) model domain mean data and data from (b) convective and (c) raining stratiform regions in C. Units are % for  $PEWV$  and  $PEH$  and  $\text{mm h}^{-1}$  for  $P_s$ .

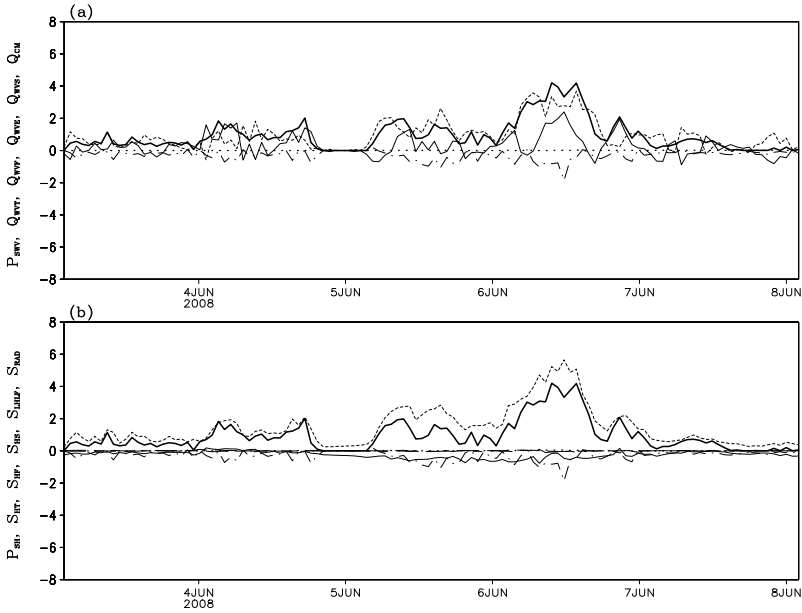


Fig. 6. Time series of contributions of model domain mean of (a)  $P_{SWV}$  (dark solid),  $Q_{WVT}$  (light solid),  $Q_{WVF}$  (short dash),  $Q_{WVE}$  (dot),  $Q_{CM}$  (dot dash), and (b)  $P_{SH}$  (dark solid),  $S_{HT}$  (light solid),  $S_{HF}$  (short dash),  $S_{HS}$  (dot),  $S_{LHLF}$  (long short dash),  $S_{RAD}$  (long dash), and  $Q_{CM}$  (dot dash) from convective regions in C. Unit is  $\text{mm h}^{-1}$ .

Model domain mean surface rainfall consists of convective and stratiform rainfall. Convective rainfall differs from stratiform rainfall in four ways. First, convective rain rates are higher than stratiform rain rates. Second, convective rainfall is associated with stronger horizontal reflectivity gradients than stratiform rainfall. Third, upward motions associated with convective rainfall are much stronger than those associated with stratiform rainfall. Fourth, the accretion of cloud water by raindrops via collisions in strong updraft cores and the vapor deposition on ice particles are primary microphysical processes that are responsible for the development of convective and stratiform rainfall, respectively (Houghton 1968). The convective-stratiform rainfall partitioning scheme used in this study is developed by Tao and Simpson (1993) and modified by Sui et al. (1994). This scheme partitions each vertical column containing clouds in 2-D  $x$ - $z$  framework into convective or stratiform based on the following criterion. Model grid point is identified as convective if it has a rain rate twice as large as the average taken over the surrounding four grid points, the one grid point on either side of this grid point, and any grid point with a rain rate of  $20 \text{ mm h}^{-1}$  or more. All non-convective cloudy points are considered as stratiform. In addition, grid points over stratiform regions are further checked and identified as convective when following conditions are met. Over raining stratiform regions, cloud water below the melting level is greater than  $0.5 \text{ g kg}^{-1}$  or the maximum updraft above  $600 \text{ hPa}$  exceeds  $5 \text{ m s}^{-1}$ , or over non-raining stratiform regions, cloud water mixing ratio of  $0.025 \text{ g kg}^{-1}$  or more exists or the maximum updraft exceeds  $5 \text{ m s}^{-1}$  below the melting level.

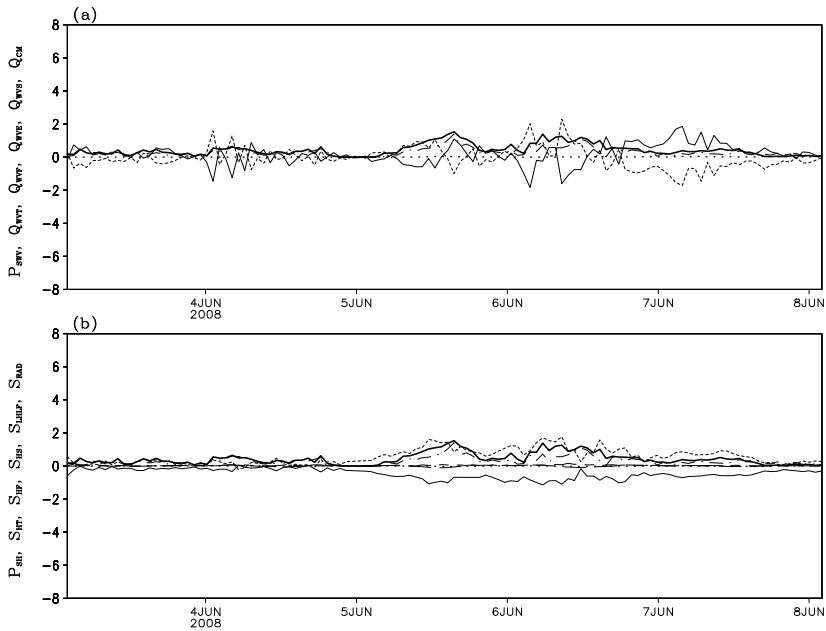


Fig. 7. Time series of contributions of model domain mean of (a)  $P_{SWV}$  (dark solid),  $Q_{WVT}$  (light solid),  $Q_{WVF}$  (short dash),  $Q_{WVE}$  (dot),  $Q_{CM}$  (dot dash), and (b)  $P_{SH}$  (dark solid),  $S_{HT}$  (light solid),  $S_{HF}$  (short dash),  $S_{HS}$  (dot),  $S_{LHLF}$  (long short dash),  $S_{RAD}$  (long dash), and  $Q_{CM}$  (dot dash) from raining stratiform regions in C. Unit is  $\text{mm h}^{-1}$ .

Convective rain rate (Fig. 6) is much higher than stratiform rain rate (Fig. 7) and mainly accounts for model domain mean surface rain rate (Fig. 3). Over convective regions, rainfall is associated with water vapor convergence in water vapor related surface rainfall budget (Fig. 6a) and heat divergence in thermally related surface rainfall budget (Fig. 6b).  $Q_{CM}$  is negative over convective regions whereas  $Q_{CM}$  is positive over raining stratiform regions (Fig. 8), which indicates the transport of hydrometeor concentration from convective regions ( $Q_{CM} < 0$ ) to raining stratiform regions ( $Q_{CM} > 0$ ). The hydrometeor transport is associated with the local atmospheric drying ( $Q_{WVT} > 0$ ) over convective regions because  $Q_{WVT}$  and  $Q_{CM}$  have similar magnitudes but opposite signs. The water vapor convergence, local atmospheric drying, and heat divergence are the rainfall sources whereas the local atmospheric cooling and the transport of hydrometeor concentration from convective regions to raining stratiform regions are the rainfall sinks.  $PE_{WV}$  generally is higher than  $PE_H$  (Fig. 4b) because the rainfall source from water vapor convergence in water vapor related surface rainfall budget is weaker than the rainfall source from heat divergence in thermally related surface rainfall budget (Fig. 6).  $PE_{WV}$  is lower than  $PE_H$  on 4 and 7 June because the rainfall source from water vapor convergence is stronger than the rainfall source from heat divergence. The RMS difference between  $PE_{WV}$  and  $PE_H$  is 12.5%. Both  $P_{WEV}$  and  $PE_H$  increase as convective rainfall intensifies (Fig. 5b). Precipitation efficiency ranges from 60% to 90% as convective rain rate is higher than  $2 \text{ mm h}^{-1}$ , whereas it ranges from 0 to 100% as convective rain rate is lower than  $2 \text{ mm h}^{-1}$ .



Over raining stratiform regions, rainfall is primarily associated with the transport of hydrometeor concentration from convective regions to raining stratiform regions because water vapor divergence dries local atmosphere on 3 and 7 June and water vapor convergence moistens local atmosphere on 4-6 June and heat divergence cools local atmosphere (Fig. 7). *PWH* is generally higher than *PEWV* on 3, 4, and 7 June whereas it is generally lower than *PEWV* on 5-6 June (Fig. 4c). The RMS difference between *PEWV* and *PEH* is 23.8%, which largely accounts for the RMS difference in the model domain mean calculations. The ranges of precipitation efficiencies for the surface rain rate of lower than  $1 \text{ mm h}^{-1}$  (0-100%) are larger than those for the surface rain rate of higher than  $1 \text{ mm h}^{-1}$  (45-100%).

**5. Radiative effects of ice clouds: CNIR versus C**

The calculations of model domain mean simulation data show that *PEWV* is insensitive to radiative effects of ice clouds on 4 June, whereas the exclusion of radiative effects of ice clouds decreases *PEH* (Table 1). The removal of radiative effects of ice clouds increases *PEWV*, but it barely affects *PEH* on 5 June. The elimination of radiative effects of ice clouds decreases *PEWV* and *PEH* on 6 June. The exclusion of radiative effects of ice clouds increases *PEWV* but it decreases *PEH* on 7 June. On 4 June, the water vapor related surface rainfall budgets reveal that all rainfall processes contribute to rain rate in C and CNIR (Table 2), which leads to 100% of *PEWV* in the two experiments. The thermally related surface rainfall budgets show that rainfall is associated with heat divergence and radiative cooling in the two experiments (Table 3). Thus, local atmospheric cooling ( $S_{HT} < 0$ ) makes *PEH* less than 100% in the two experiments.

(a)	Model domain mean			Convective regions			Raining stratiform regions		
	C	CNIR	CNIM	C	CNIR	CNIM	C	CNIR	CNIM
4 June	100.0	100.0	91.5	77.8	68.8	83.0	82.3	66.7	58.3
5 June	92.0	98.4	81.0	71.5	70.3	80.6	100.0	90.8	100.0
6 June	99.8	96.3	91.3	81.9	80.5	87.3	100.0	82.2	100.0
7 June	62.8	64.6	60.4	55.0	67.2	77.2	28.1	36.8	51.5

(b)	Model domain mean			Convective regions			Raining stratiform regions		
	C	CNIR	CNIM	C	CNIR	CNIM	C	CNIR	CNIM
4 June	77.0	72.0	68.5	73.5	64.1	77.0	65.2	67.5	47.0
5 June	63.3	63.8	57.0	52.1	51.5	56.5	47.3	47.1	34.4
6 June	71.9	69.3	68.7	70.6	69.4	77.6	49.1	43.7	34.3
7 June	36.7	35.1	34.2	48.6	49.2	49.4	33.4	28.2	29.1

Table 1. (a) *PEWV* and (b) *PEH* calculated data averaged daily and over model domain, convective regions, and raining stratiform regions in C, CNIR, and CNIM. Unit is %.

		Model domain mean			Convective regions			Raining stratiform regions			
		C	CNIR	CNIM	C	CNIR	CNIM	C	CNIR	CNIM	
4	June	$P_{SWV}$	1.16	1.07	0.92	0.86	0.73	0.80	0.30	0.34	0.13
		$Q_{WVT}$	0.22	0.19	0.16	0.52	0.46	0.31	-0.06	-0.17	-0.09
		$Q_{WVF}$	0.87	0.86	0.84	0.57	0.60	0.64	0.09	0.18	0.13
		$Q_{WVE}$	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
		$Q_{CM}$	0.05	0.01	-0.09	-0.24	-0.33	-0.16	0.28	0.32	0.09
5	June	$P_{SWV}$	1.54	1.59	1.27	0.92	0.99	0.93	0.62	0.59	0.35
		$Q_{WVT}$	-0.13	-0.03	-0.17	0.10	0.35	0.21	0.00	-0.06	0.10
		$Q_{WVF}$	1.59	1.53	1.55	1.17	1.06	0.93	0.20	0.19	0.13
		$Q_{WVE}$	0.03	0.04	0.02	0.00	0.01	0.00	0.01	0.02	0.00
		$Q_{CM}$	0.05	0.04	-0.13	-0.37	-0.42	-0.22	0.41	0.44	0.12
6	June	$P_{SWV}$	2.95	2.84	2.69	2.21	2.21	2.36	0.74	0.64	0.32
		$Q_{WVT}$	0.34	0.41	0.31	0.77	0.92	0.87	0.03	-0.14	0.17
		$Q_{WVF}$	2.57	2.51	2.62	1.92	1.80	1.83	0.23	0.38	0.03
		$Q_{WVE}$	0.04	0.04	0.02	0.01	0.01	0.00	0.01	0.02	0.00
		$Q_{CM}$	-0.01	-0.11	-0.26	-0.49	-0.53	-0.34	0.47	0.38	0.12
7	June	$P_{SWV}$	0.54	0.51	0.46	0.29	0.33	0.32	0.25	0.19	0.15
		$Q_{WVT}$	0.64	0.59	0.67	-0.09	0.08	0.08	0.65	0.30	0.19
		$Q_{WVF}$	-0.32	-0.28	-0.23	0.52	0.39	0.32	-0.63	-0.32	-0.14
		$Q_{WVE}$	0.16	0.17	0.10	0.01	0.01	0.00	0.02	0.04	0.01
		$Q_{CM}$	0.05	0.03	-0.08	-0.15	-0.16	-0.09	0.20	0.17	0.09

Table 2. Water vapor related surface rainfall budget ( $P_{SWV}$ ,  $Q_{WVT}$ ,  $Q_{WVF}$ ,  $Q_{WVE}$ , and  $Q_{CM}$ ) averaged daily and over model domain, convective regions, and raining stratiform regions in C, CNIR, and CNIM. Unit is  $\text{mm h}^{-1}$ .

The removal of radiative effects of ice clouds decreases  $PEH$  from C to CNIR through the enhanced local atmospheric cooling associated with the enhanced radiative cooling because of similar heat divergence in the two experiments. On 5 June, the elimination of radiative effects of ice clouds increases  $PEWV$  through the reduced local atmospheric moistening ( $Q_{WVT}<0$ ) associated with the decreased water vapor convergence ( $Q_{WVF}>0$ ). The two experiments have similar  $PEH$  because of similar heat related rainfall processes. On 6 June, the exclusion of radiative effects of ice clouds decreases  $PEWV$  and  $PEH$  from C to CNIR through the enhanced hydrometeor gain ( $Q_{CM}>0$ ) because water vapor and heat processes have similar contributions to rainfall ( $Q_{WVT}+Q_{WVF}+Q_{WVE} \approx S_{HT}+S_{HF}+S_{HS}+S_{LHLF}+S_{RAD}$ ) in the two experiments. On 7 June, the removal of radiative effects of ice clouds increases  $PEWV$  through the decreased water vapor divergence, whereas it decreases  $PEH$  through the enhanced local atmospheric cooling associated with the enhanced radiative cooling.

Over convective regions, the exclusion of radiative effects of ice clouds decreases  $PEWV$  and  $PEH$  through the intensified transport of hydrometeor concentration from convective regions to raining stratiform regions ( $Q_{CM}<0$ ) because water vapor and heat processes have similar contributions to rainfall in the two experiments on 4 June. Similar magnitudes of rainfall sources associated with water vapor, heat, and cloud processes lead to similar

*PEWV* and *PEH* on 5 and 6 June. On 7 June, the exclusion of radiative effects of ice clouds increases *PEWV* through the local atmospheric change from moistening in C to drying in CNIR associated with the decrease in water vapor convergence while the two experiments have similar transport rates of hydrometeor concentration from convective regions to raining stratiform regions. The two experiments have similar *PEH* because of similar thermal processes. Note that radiative cooling is negligibly small over convective regions. Over raining stratiform regions, the removal of radiative effects of ice clouds decreases *PEWV* through the enhanced local atmospheric moistening on 4 June. Because the increase in rainfall from C to CNIR is similar to the increase in rainfall source that is from the heat divergence and the transport of hydrometeor concentration from convective regions to raining stratiform regions, the increase in stratiform rainfall from C to CNIR leads to the increase in *PEH*. On 5 June, the elimination of radiative effects of ice clouds decreases *PEWV* because the local water vapor is barely changed in C and the local atmospheric moistening occurs in CNIR. The two experiments have similar *PEH* due to similar rainfall sources from thermal and cloud processes. On 6 June, the exclusion of radiative effects of ice clouds leads to the decreases in *PEWV* through the local atmospheric change from drying in C to moistening in CNIR and reduces *PEH* through the enhanced local atmospheric cooling. On 7 June, the removal of radiative effects of ice clouds increases *PEWV* through the slowdown in water vapor divergence. Because the decrease in rainfall is similar to the decrease in the rainfall source from heat divergence and transport of hydrometeor concentration from convective regions to raining stratiform regions as a result of similar local atmospheric cooling rate in the two experiments, the decrease in stratiform rainfall from C to CNIR leads to the decrease in *PEH*.

## 6. Microphysical effects of ice clouds: CNIM versus CNIR

The calculations of model domain mean simulation data show the decreases in *PEWV* and *PEH* from CNIR to CNIM during the life span of pre-summer heavy rainfall event (Figs. 9a-10a). On 4 June, the exclusion of microphysical effects of ice clouds decreases *PEWV* and *PEH* through the hydrometeor change from loss in CNIR to gain in CNIM and the weakened local atmospheric cooling (Figs. 11a-17a). On 5 June, the decrease in *PEWV* is associated with the intensification in local atmospheric moistening and the hydrometeor change from loss in CNIR to gain in CNIM. The reduction in *PEH* is related to the hydrometeor change from loss in CNIR to gain in CNIM. On 6 June, the decrease in *PEWV* corresponds to the strengthened hydrometeor gain. *PEH* is barely changed in the two experiments because of similar rates of rainfall source from thermal processes. On 7 June, the decreases in *PEWV* and *PEH* result from the hydrometeor change from loss in CNIR to gain in CNIM although water vapor divergence and local atmospheric cooling are weakened.

Over convective regions, *PEWV* and *PEH* are increased from CNIR to CNIM during the life span of pre-summer heavy rainfall event (Figs. 9b-10b). The exclusion of microphysical effects of ice clouds increases *PEWV* and *PEH* through the weakened transport of hydrometeor concentration from convective regions to raining stratiform regions during 4-6 June (Figs. 11b-17b). The decrease in local atmospheric cooling also contributes to the increases in *PEWV* and *PEH* on 6 June. On 7 June, the removal of microphysical effects of ice clouds increases *PEWV* through the weakened transport of hydrometeor concentration from

convective regions to raining stratiform regions, whereas it barely changes  $PEH$  because of the offset between the weakened transport of hydrometeor concentration from convective regions to raining stratiform regions and the enhanced local atmospheric cooling (Figs. 14b and 15b).

		Model domain mean			Convective regions			Raining stratiform regions		
		C	CNIR	CNIM	C	CNIR	CNIM	C	CNIR	CNIM
4 June	$P_{SH}$	1.16	1.07	0.92	0.86	0.73	0.80	0.30	0.34	0.13
	$S_{HT}$	-0.34	-0.41	-0.33	-0.05	-0.06	-0.08	-0.16	-0.16	-0.14
	$S_{HF}$	1.30	1.28	1.24	1.15	1.11	1.02	0.15	0.14	0.18
	$S_{HS}$	-0.01	-0.01	-0.01	0.01	0.00	0.00	0.00	0.00	0.00
	$S_{LHLF}$	0.02	0.00	0.00	-0.01	-0.02	0.00	0.03	0.02	0.00
	$S_{RAD}$	0.13	0.18	0.12	0.01	0.02	0.01	0.00	0.02	0.00
	$Q_{CM}$	0.05	0.01	-0.09	-0.24	-0.33	-0.16	0.28	0.32	0.09
5 June	$P_{SH}$	1.54	1.59	1.27	0.92	0.99	0.93	0.62	0.59	0.35
	$S_{HT}$	-0.88	-0.88	-0.82	-0.46	-0.48	-0.49	-0.68	-0.66	-0.65
	$S_{HF}$	2.30	2.28	2.14	1.76	1.91	1.63	0.88	0.74	0.89
	$S_{HS}$	-0.01	-0.02	-0.01	0.00	0.01	0.00	-0.01	-0.01	0.00
	$S_{LHLF}$	0.01	0.01	0.00	-0.01	-0.03	0.00	0.03	0.04	0.00
	$S_{RAD}$	0.07	0.16	0.09	0.00	0.01	0.01	-0.01	0.03	0.00
	$Q_{CM}$	0.05	0.04	-0.13	-0.37	-0.42	-0.22	0.41	0.44	0.12
6 June	$P_{SH}$	2.95	2.84	2.69	2.21	2.21	2.36	0.74	0.64	0.32
	$S_{HT}$	-1.15	-1.14	-0.97	-0.40	-0.40	-0.34	-0.76	-0.81	-0.62
	$S_{HF}$	4.00	3.92	3.81	3.10	3.15	3.02	0.98	1.00	0.82
	$S_{HS}$	0.00	-0.01	0.01	0.02	0.01	0.01	-0.01	-0.01	0.00
	$S_{LHLF}$	0.03	0.03	0.00	-0.03	-0.04	0.00	0.06	0.06	0.00
	$S_{RAD}$	0.06	0.16	0.10	0.00	0.01	0.01	0.00	0.02	0.01
	$Q_{CM}$	-0.01	-0.11	-0.26	-0.49	-0.53	-0.34	0.47	0.38	0.12
7 June	$P_{SH}$	0.54	0.51	0.46	0.29	0.33	0.32	0.25	0.19	0.15
	$S_{HT}$	-0.90	-0.92	-0.81	-0.15	-0.17	-0.23	-0.48	-0.47	-0.36
	$S_{HF}$	1.30	1.26	1.29	0.59	0.66	0.64	0.51	0.47	0.42
	$S_{HS}$	-0.03	-0.03	-0.01	0.00	0.00	0.00	-0.01	-0.01	0.00
	$S_{LHLF}$	0.02	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.01
	$S_{RAD}$	0.09	0.15	0.06	0.00	0.00	0.00	0.00	0.01	0.00
	$Q_{CM}$	0.05	0.03	-0.08	-0.15	-0.16	-0.09	0.20	0.17	0.09

Table 3. Thermally related surface rainfall budget ( $P_{SWV}$ ,  $S_{HT}$ ,  $S_{HF}$ ,  $S_{HS}$ ,  $S_{RAD}$ , and  $Q_{CM}$ ) averaged daily and over model domain, convective regions, and raining stratiform regions in C, CNIR, and CNIM. Unit is  $\text{mm h}^{-1}$ .

Over raining stratiform regions, the elimination of microphysical effects of ice clouds decreases  $PEWV$  through the weakened local atmospheric moistening and reduces  $PEH$  through the weakened local atmospheric cooling on 4 June. During 5-6 June, the exclusion of microphysical effects of ice clouds increases  $PEWV$  because all rainfall processes favors rainfall in CNIM but the local atmospheric moistening reduces rainfall in CNIR. Although

the removal of microphysical effects of ice clouds barely impacts local atmospheric cooling on 5 June and it decreases local atmospheric cooling on 6 June, the decreases in stratiform rainfall are associated with the slowdown in transport of hydrometeor concentration from convective regions to raining stratiform regions. As a result, the decreases in stratiform rainfall lead to the decreases in  $PEH$  from CNIR to CNIM. On 7 June, the elimination of microphysical effects of ice clouds increases  $PEWV$  through the weakened water vapor divergence and increases  $PEH$  through the weakened local atmospheric cooling.

## 7. Conclusions

Precipitation efficiency can be well defined through diagnostic surface rainfall budgets. From thermally related surface rainfall budget, precipitation efficiency associated with heat processes ( $PEH$ ) is first defined in this study as the ratio of surface rain rate and the rainfall source from heat and cloud budgets. Precipitation efficiency associated with water vapor processes ( $PEWV$ ) was defined by Sui et al. (2007) as the ratio of surface rain rate to the rainfall source from water vapor and cloud budgets. In this study, both precipitation efficiencies and their responses to effects of ice clouds are investigated through an analysis of sensitivity cloud-resolving modeling data of a pre-summer heavy rainfall event over southern China during June 2008. The major results include:

- The calculations of model domain mean simulation data show that  $PEH$  is lower than  $PEWV$  because heat divergence contributes more to surface rainfall than water vapor convergence does. Precipitation efficiencies are lower during the decay phase than during the development of rainfall.  $PEH$  is generally lower than  $PEWV$  over convective regions, whereas it is generally higher than  $PEWV$  over raining stratiform regions. Precipitation efficiencies increase as surface rain rate increases.
- $PEWV$  has different responses to radiative effects of ice clouds during the different stages of the rainfall event. The exclusion of Microphysical effects of ice clouds generally decreases  $PEWV$  in the calculations of model domain mean simulation data, whereas it generally increases  $PEWV$  over raining regions.
- The exclusion of radiative effects of ice clouds generally decreases  $PEH$ . The removal of microphysical effects of ice clouds generally decreases  $PEH$  except that it increases  $PEH$  over convective regions.
- Effects of ice clouds on precipitation efficiencies can be explained by the analysis of surface rainfall budgets. The changes in  $PEWV$  are mainly associated with the changes in local atmospheric moistening and transport of hydrometeor concentration from convective regions to raining stratiform regions during the life span of pre-summer heavy rainfall event and the change in water vapor divergence on 7 June. The changes in  $PEH$  are mainly related to the changes in local atmospheric cooling and radiative cooling and transport of hydrometeor concentration from convective regions to raining stratiform regions during the life span of pre-summer heavy rainfall event.

## 8. Acknowledgment

The authors thank W.-K. Tao at NASA/GSFC for his cloud resolving model, and Dr. N. Sun at I. M. Systems Group, Inc. for technical assistance to access NCEP/GDAS data. This study

is supported by the National Key Basic Research and Development Project of China under Grant No. 2011CB403405, the National Natural Science Foundation of China under Grant No. 41075039, the Chinese Special Scientific Research Project for Public Interest under Grant No. GYHY200806009, and the Qinglan Project of Jiangsu Province of China under Grant No. 2009.

## 9. References

- Auer, A. H. Jr. & Marwitz, J. D. (1968) Estimates of air and moisture flux into hailstorms on the High Plains. *Journal of Applied Meteorology*, Vol.6, No.2, (April 1968), pp. 196-198, ISSN 0021-8952.
- Braham, R. R. Jr. (1952) The water and energy budgets of the thunderstorm and their relation to thunderstorm development. *Journal of Meteorology*, Vol.9, No.4, (August 1952), pp. 227-242, ISSN 0095-9634.
- Chong, M. & Hauser, D. (1989) A tropical squall line observed during the CORT 81 experiment in West Africa. Part II: Water budget. *Monthly Weather Review*, Vol.117, No.4, (April 1989), pp. 728-744, ISSN 0027-0644.
- Chou, M.-D. & Suarez, M. J. (1994) An efficient thermal infrared radiation parameterization for use in general circulation model. NASA Tech. Memo. 104606, Vol.3, pp. 1-85, Available from NASA/Goddard Space Flight Center, Code 913, Greenbelt, MD 20771.
- Chou, M.-D.; Kratz, D. P. & Ridgway, W. (1991) Infrared radiation parameterization in numerical climate models. *Journal of Climate*, Vol.4, No.4, (April 1991), pp. 424-437, ISSN 0894-8755.
- Chou, M.-D.; Suarez, M. J.; Ho, C.-H.; Yan, M. M.-H. & Lee K.-T., (1998) Parameterizations for cloud overlapping and shortwave single scattering properties for use in general circulation and cloud ensemble models. *Journal of Climate*, Vol.11, No.2, (February 1998), pp. 201-214, ISSN 0894-8755.
- Ding, Y. (1994) *Monsoon over China*, Springer, ISBN 978-90-481-4161-6.
- Ding Y. & Murakami, M. (1994) *Asia monsoon*, Meteorological Press, ISBN 7-5029-1709-8, Beijing, China.
- Doswell, C. A., III; Brooks, H. E. & Maddox, R. A. (1996) Flash flood forecasting: An ingredients-based methodology. *Weather Forecasting*, Vol.11, No.4, (December 1996), pp. 560-581, ISSN 0882-8156.
- Ferrier, B. S.; Simpson, J. & Tao, W.-K. (1996) Factors responsible for precipitation efficiencies in midlatitude and tropical squall simulations. *Monthly Weather Review*, Vol.124, No.10, (October 1996), pp. 2100-2125, ISSN 0027-0644.
- Fovell, R. G. & Ogura, Y. (1988) Numerical simulation of a midlatitude squall line in two dimensions. *Journal of the Atmospheric Sciences*, Vol.45, No.24, (December 1988), pp. 3846-3879, ISSN 0022-4928.
- Gao, S. & Li, X. (2008) *Cloud-resolving modeling of convective processes*. Springer, ISBN 978-1-4020-8275-7, Dordrecht.
- Gao, S. & Li, X. (2010) Precipitation equations and their applications to the analysis of diurnal variation of tropical oceanic rainfall. *Journal of Geophysical Research*, Vol.115, No.D08204, (April 2010), ISSN 0148-0227.
- Gao, S.; Cui, X.; Zhu, Y. & Li, X. (2005) Surface rainfall processes as simulated in a cloud resolving model. *J. Geophys. Res.*, Vol.110, No.D10202, (May 2005), ISSN 0148-0227.

- Gao, S.; Ran, L. & Li, X. (2006) Impacts of ice microphysics on rainfall and thermodynamic processes in the tropical deep convective regime: A 2D cloud-resolving modeling study. *Monthly Weather Review*, Vol.134, No.10, (October 2006), pp. 3015-3024, ISSN 0027-0644.
- Grabowski, W. W. (2003) Impact of ice microphysics on multiscale organization of tropical convection in two-dimensional cloud-resolving simulations. *Quarterly Journal of the Royal Meteorological Society*, Vol. 129, No.587 (February 2003), pp. 67-81, ISSN 0035-9009.
- Grabowski, W. W. & Moncrieff, M. M. (2001) Large-scale organization of tropical convection in two-dimensional explicit numerical simulations. *Quarterly Journal of the Royal Meteorological Society*, Vol. 127, No.572, (February 2001), pp. 445-468, ISSN 0035-9009.
- Grabowski, W. W., X. Wu, and M. W. Moncrieff, 1999: Cloud-resolving model of tropical cloud systems during Phase III of GATE. Part III: Effects of cloud microphysics. *Journal of the Atmospheric Sciences*, Vol.56, No.14, (July 1999), pp. 2384-2402, ISSN 0022-4928.
- Heymsfield, G. M. & Schotz, S. (1985) Structure and evolution of a severe squall line over Oklahoma. *Monthly Weather Review*, Vol.113, No.9, (September 1985), pp. 1563-1589, ISSN 0027-0644.
- Houghton, H. G. (1968) On precipitation mechanisms and their artificial modification. *Journal of Applied Meteorology*, Vol.7, No.5, (October 1968), pp. 851-859, ISSN 0021-8952.
- Hsie, E. Y.; Farley, R. D. & Orville, H. D. (1980) Numerical simulation of ice phase convective cloud seeding. *Journal of Applied Meteorology*, Vol.29, No.8, (August 1980), pp. 950-977, ISSN 0021-8952.
- Krishnamurti, T. N.; Molinari, J. & Pan, H.-L. (1976) Numerical simulation of the Somali Jet, *Journal of the Atmospheric Sciences*, Vol.33, No.12, (December 1976), pp. 2350-2362, ISSN 0022-4928.
- Krueger, S. K.; Fu, Q.; Liou, K.-N. & Chin, H.-N. S. (1995) Improvement of an ice-phase microphysics parameterization for use in numerical simulations of tropical convection, *Journal of Applied Meteorology*, Vol.34, No.1, (January 1995), pp. 281-287, ISSN 0021-8952.
- Li, X.; Sui, C.-H.; Lau, K.-M. & Chou, M.-D. (1999) Large-scale forcing and cloud-radiation interaction in the tropical deep convective regime. *Journal of the Atmospheric Sciences*, Vol.56, No.17, (September 1999), pp. 3028-3042, ISSN 0022-4928.
- Li, X.; Sui, C.-H. & Lau, K.-M. (2002) Precipitation efficiency in the tropical deep convective regime: A 2-D cloud resolving modeling study. *Journal of Meteorological Society of Japan*, Vol.80, No.2, (April 2002), pp. 205-212, ISSN 0026-1165.
- Lin, Y.-L.; Farley, R. D. & Orville, H. D. (1983) Bulk parameterization of the snow field in a cloud model, *Journal of Applied Meteorology and climatology*, Vol.22, No.6, (June 1983), pp. 1065-1092, ISSN 0733-3021.
- Liu, C.; Moncrieff, M. W. & Zipser, E. J. (1997): Dynamic influence of microphysics in tropical squall lines: A numerical study. *Monthly Weather Review*, Vol.125, No.9, (September 1997), pp. 2193-2210, ISSN 0027-0644.
- McCumber, M.; Tao, W.-K.; Simpson, J.; Penc, R. & Soong, S.-T. (1991) Comparison of ice - phase microphysical parameterization schemes using numerical simulations of

- tropical convection. *Journal of Applied Meteorology and Climatology*, Vol.30, No.7, (July 1991), pp. 985-1004, ISSN 0894-8763.
- Nicholls, M. E. (1987) A comparison of the results of a two-dimensional numerical simulation of a tropical squall line with observations. *Monthly Weather Review*, Vol.115, No.12, (December 1987), pp. 3055-3077, ISSN 0027-0644.
- Ping, F.; Luo, Z. & Li, X. (2007) Microphysical and radiative effects of ice microphysics on tropical equilibrium states: A two-dimensional cloud-resolving modeling study. *Monthly Weather Review*, Vol.135, No.7, (July 2007), pp. 2794-2802, ISSN 0027-0644.
- Rutledge, S. A. & Hobbs, P. V. (1983) The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. Part VIII: A model for the "seeder-feeder" process in warm-frontal rainbands, *Journal of the Atmospheric Sciences*, Vol.40, No.5, (May 1983), pp. 1185-1206, ISSN 0022-4928.
- Rutledge, S. A. & Hobbs, P. V. (1984) The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. Part XII: A diagnostic modeling study of precipitation development in narrow cold-frontal rainbands, *Journal of the Atmospheric Sciences*, Vol.41, No.20., (October 1984), 2949-2972, ISSN 0022-4928.
- Shen, X.; Wang, Y.; Zhang, N. & Li, X. (2010) Precipitation and cloud statistics in the deep tropical convective regime. *Journal of Geophysical Research*, Vol.115, No.D24205, ISSN 0747-7309.
- Shen, X.; Wang, Y. & Li, X. (2011) Radiative effects of water clouds on rainfall responses to the large-scale forcing during pre-summer heavy rainfall over southern China. *Atmospheric Research*, Vol.99, No.1, (January 2011), pp.120-128, ISSN0169-8095.
- Shen, X.; Wang, Y. & Li, X. (2011) Effects of vertical wind shear and cloud radiative processes on responses of rainfall to the large-scale forcing during pre-summer heavy rainfall over southern China. *Quarterly Journal of the Royal Meteorological Society*, Vol.137, No.654, (February 2011), pp. 236-249, ISSN 0035-9009.
- Simmonds, I.; Bi, D. & Hope, P. (1999) Atmospheric water vapor flux and its association with rainfall over China in summer. *Journal of Climate*, Vol.12, No.5, (May 1999), pp. 1353-1367, ISSN 0894-8755.
- Soong, S. T. & Ogura, Y. (1980) Response of tradewind cumuli to large-scale processes. *Journal of the Atmospheric Sciences*, Vol.37, No.9, (September 1980), pp. 2035-2050, ISSN 0022-4928.
- Soong, S. T., and W. K. Tao, 1980: Response of deep tropical cumulus clouds to mesoscale processes. *Journal of the Atmospheric Sciences*, Vol.37, No.9, (September 1980), pp. 2016-2034, ISSN 0022-4928.
- Sui, C.-H. & Li, X. (2005) A tendency of cloud ratio associated with the development of tropical water and ice clouds. *Terrestrial Atmospheric and Oceanic Sciences*, Vol.16, No.4, (June 2005), pp. 419-434, ISSN 1017-0839.
- Sui, C.-H.; Lau, K.-M.; Tao, W.-K. & Simpson, J. (1994) The tropical water and energy cycles in a cumulus ensemble model. Part I: Equilibrium climate. *Journal of the Atmospheric Sciences*, Vol. 51, No.5, (March 1994), pp. 711-728, ISSN 0022-4928.
- Sui, C.-H.; Li, X. & Lau, K.-M. (1998) Radiative-convective processes in simulated diurnal variations of tropical oceanic convection, *Journal of the Atmospheric Sciences*, Vol.55, No.13, (July 1998), pp. 2345-2359, ISSN 0022-4928.



- Sui, C.-H.; Li, X.; Yang, M.-J. & Huang, H.-L. (2005) Estimation of oceanic precipitation efficiency in cloud models. *Journal of the Atmospheric Sciences*, Vol.62, No.12, (December 2005), pp. 4358-4370, ISSN 0022-4928.
- Sui, C.-H.; Li, X. & Yang, M.-J. (2007) On the definition of precipitation efficiency. *Journal of the Atmospheric Sciences*, Vol. 64, No.12, (December 2007), pp 4506-4513, ISSN 0022-4928.
- Tao, S. & Ding, Y. (1981) Observational evidence of the influence of the Qinghai-Xizang (Tibet) plateau on the occurrence of heavy rain and severe convective storms in China. *Bulletin of American Meteorological Society*, Vol.62, No.1, (January 1981), pp. 23-30, ISSN 0003-0007.
- Tao, W.-K. & Simpson, J. (1989) Modeling study of a tropical squall-type convective line, *Journal of the Atmospheric Sciences*, Vol.46, No.1, (January 1989), pp. 177-202, ISSN 0022-4928.
- Tao, W.-K. & Simpson, J. (1993) The Goddard Cumulus Ensemble model. Part I: Model description. *Terrestrial Atmospheric and Oceanic Sciences*, Vol.4, No. 1, (March 1993), pp. 35-72, ISSN 1017-0839.
- Tao, W.-K; Simpson J. & McCumber, M. (1989) An ice-water saturation adjustment, *Monthly Weather Review*, Vol.117, No.1, (January 1989), pp. 231-235, ISSN 0027-0644.
- Tao, W.-K.; Simpson, J. & Soong, S.-T. (1991) Numerical simulation of a subtropical squall line over the Taiwan Strait. *Monthly Weather Review*, Vol.119, No.11, (November 1991), pp. 2699-2723, ISSN 0027-0644.
- Tao, W.-K.; Simpson, J.; Sui, C.-H.; Ferrier, B.; Lang, S.; Scala, J.; Chou M.-D. & Pickering, K. (1993) Heating, moisture and water budgets of tropical and midlatitude squall lines: Comparisons and sensitivity to longwave radiation. *Journal of the Atmospheric Sciences*, Vol.50, No.5, (March 1993), pp. 673-690, ISSN 0022-4928.
- Wang, D.; Li, X.; Tao, W.-K.; Liu, Y. & Zhou, H. (2009) Torrential rainfall processes associated with a landfall of severe tropical storm Bilis (2006): A two-dimensional cloud-resolving modeling study. *Atmospheric Research*, Vol.91, No.1, (January 2009), pp. 94-104 ISSN 0169-8095.
- Wang, J. & Li, M. (1982) Cross equator flow from Australia and summer monsoon circulations and precipitation over China. *Sci. Atmos. Sin*, Vol.6, No.1, (January 1982), pp. 1-10, ISSN 1006-9895.
- Wang, Y.; Shen, X. & Li, X. (2010) Microphysical and radiative effects of ice clouds on responses of rainfall to the large-scale forcing during pre-summer heavy rainfall over southern China. *Atmospheric Research*, Vol.97, No.1-2, (July 2010), pp. 35-46, ISSN 0169-8095.
- Wu, X. (2002) Effects of ice microphysics on tropical radiative-convective-oceanic quasi-equilibrium states. *Journal of the Atmospheric Sciences*, Vol.59, No.11, (June 2002), 1885-1897, ISSN 0022-4928.
- Wu, X.; Hall, W. D.; Grabowski, W. W.; Moncrieff, M. W.; Collins, W. D. & Kiehl, J. T. (1999) Long-term behavior of cloud systems in TOGA COARE and their interactions with radiative and surface processes. Part II: Effects of ice microphysics on cloud-radiation interaction. *Journal of the Atmospheric Sciences*, Vol.56, No.18, (September 1999), pp. 3177-3195, ISSN 0022-4928.
- Xu, X.; Xu, F. & Li, B. (2007) A cloud-resolving modeling study of a torrential rainfall event over China. *Journal of Geophysical Research*, Vol.112, No.D17204, ISSN 0148-0227.

Yoshizaki, M. (1986) Numerical simulations of tropical squall-line clusters: Two-dimensional model. *Journal of Meteorological Society of Japan*, Vol.64, No.4, (August 1986), pp. 469-491, ISSN 0026-1165.

# Comparison of the Thermodynamic Parameters Estimation for the Adsorption Process of the Metals from Liquid Phase on Activated Carbons

Svetlana Lyubchik, Andrey Lyubchik, Olena Lygina,  
Sergiy Lyubchik and Isabel Fonseca  
*REQUIMTE, Faculdade Ciência e Tecnologia, Universidade Nova de Lisboa  
Quinta de Torre, Campus da Caparica, 2829-516 Caparica  
Portugal*

## 1. Introduction

Over the past decades investigation of the adsorption process on activated carbons has confirmed their great potential for industrial wastewater purification from toxic and heavy metals. This chapter is focused on the adsorption of Cr (III) in high-capacity solid adsorbents such as activated carbons. There are abundant publications on heavy metal adsorption on activated carbons with different oxygen functionalities covering wide-range conditions (solution pH, ionic strength, initial sorbate concentrations, carbon loading and etc. (Brigatti et al., 2000; Carrott et al., 1997; Li et al., 2011; Lyubchik et al., 2008; Tikhonova et al., 2008; Kołodzyńska, 2010; Anirudhan & Radhakrishnan, 2011). Although much has been accomplished in this area, less attention has been given to the kinetics, thermodynamics and temperature dependence of the adsorption process, which is still under continuing debates (Ramesh et al., 2007; Myers, 2004). The principal problem in interpretation of solution adsorption studies lies in the relatively low comparability of the data obtained by different research groups. These are due to the differences in the nature of the carbons, conditions of the adsorption processes and the chosen methodology of the metals adsorption analysis. Furthermore, the adsorption from the solution is much more complex than that from the gas phase.

In general, the molecules attachment to the solid surface by adsorption is a broad subject (Myers, 2004). Therefore, only complex investigation of the metal ions/carbon surfaces interaction at the aqueous-solid interface can help to understand the metals adsorption mechanism, which is an important point in optimization of the conditions of their removal by activated carbons (Anirudhan & Radhakrishnan, 2008; Argun et al., 2007; Aydin & Aksoy, 2009; Ramesh et al., 2007; Liu et al., 2004). Particularly, thermodynamics has the remarkable ability to connect seemingly unrelated properties (Myers, 2004). The most important application of thermodynamics is the calculation of equilibrium between phases of the adsorption process profile. The basis for thermodynamic calculations is the adsorption isotherm, which gives the amount of the metals adsorbed in the porous structure as a function of the amount at equilibrium in the solutions. Whether the adsorption isotherm has been experimentally determined, the data points must be fitted with analytical equations for interpolation, extrapolation, and for the calculation of thermodynamic properties by numerical integration or differentiation (Myers, 2004; Ruthven, 1984).

It has to be noted, that the thermodynamics applies only to equilibrium adsorption isotherms. The equilibrium of heavy metals adsorption on activated carbons is still in its infancy due to the complexity of operating mechanisms of metal ions binding to carbon with ion exchange, complexation, and surface adsorption as the prevalent ones (Brown et al., 2000). Furthermore, these processes are strongly affected by the pH of the aqueous solution (Liu et al., 2004; Chen and Lin, 2001; Brigatti et al., 2000). The influence of pH is generally attributed to the variation, with pH, in the relative distribution of the metal and carbon surface species, in their charge and proton balance (Csobán et al., 1998; Kratochvil and Volesky, 1998). Therefore, the equilibrium constants of each type of the species on each type of the activated sites are very important for the controlling of metals ions capture by activated carbons (Carrott et al., 1997; Chen & Lin, 2001).

Another area of the debates is an optimum contact time to reach the adsorption equilibrium and, once again, regardless of the solution pHs, the differences in metal ions speciation, adsorbents charge and potential, complicate the overall process and make a comparison of the results of a metals capture by activated carbons difficult. The majority of studies on the sorption kinetics have revealed a two-step behaviour of the adsorption systems (Brigatti et al., 2000; Csobán et al., 1998; Raji et al., 1998) with fast initial uptake and much slower gradual uptake afterwards, which might take days even months (et al., 2000; Csobán et al., 1998; Raji et al., 1998; Kumar et al., 2000; Ajmal et al., 2001; Lakatos et al., 2002; Chakir et al., 2002; Leist et al., 2000; Csobán & Joó, 1999). Some of the authors reported the optimum contact time of minutes (Kumar et al., 2000; Ajmal et al., 2001), whereas, at the other extreme, that of hundred hours (Brigatti et al., 2000; Lakatos et al., 2002) for equilibrium to be attained; and the average values reported for the heavy metal binding were of 1–5 hours (Csobán et al., 1998; Raji et al., 1998; Chakir et al., 2002; Leist et al., 2000; Csobán and Joó, 1999). It has been also stressed that adsorption thermodynamics is drastically affected by the equilibrium pH of the solutions. Regardless of the equilibrium pH, adsorption of the heavy metals by a single adsorbent could be completed in a quite different contact time (Carrott et al., 1997; Lalvani et al., 1998; Farias et al., 2002; Perez-Candela et al., 1995). Taking into account that equilibration of metal ions uptake by activated carbons depends on the equilibrium pH, authors agreed (Lyubchik et al., 2003) with the statement (Carrott et al., 1997) that it would be appropriate to express adsorption results in terms of the final solution pH. However, this practice is not widely used by the investigators.

Due to the prolonged time is needed to accomplish thermodynamic equilibrium conditions, the adsorption experiments are often carried out under pseudo-equilibrium condition, when the actual time is chosen either to accomplish the rapid adsorption step or, rather arbitrary, to ensure that the saturation level of the carbon is reached (Kumar et al., 2000). However, once again, the adsorption models are all valid only and, therefore, applicable only to complete equilibration.

The study presented herein is part of the work aimed the exploration of the mechanism of Cr (III) adsorption on activated carbons associated with varying of surface oxygen functionality and porous texture. The mechanism of chromium adsorption was investigated through a series of equilibrium and kinetic experiments under varying pH, temperature, initial chromium concentration, carbon loading for wide-ranging carbons of different surface properties (i.e. texture and surface groups) (Lyubchik et al., 2004; Lyubchik et al., 2005; Lyubchik et al., 2008); and particular objective of the current study is evaluation of the thermodynamics (entropy, enthalpy, free energy) parameters of the adsorption process in the system “Cr (III) – activated carbon”.

Thermodynamics were evaluated through a series of the equilibrium experiments under varying temperature, initial chromium concentration, carbon loading for two sets of the commercial activated carbons and their oxidised by post-chemical treatment forms with different texture and surface functionality. This approach served the dual purpose: i) gained deep insight into various carbon's structural characteristics and their effect on thermodynamics of the Cr (III) adsorption; and ii) gained insight, which often very difficult or impossible to obtain by other mean, into equilibrium of the Cr (III) adsorption on activated carbon. The thermodynamics parameters were evaluated using both the thermodynamic equilibrium constants and the Langmuir, Freundlich and BET constants. The obtained data on thermodynamic parameters were compared, when it was possible.

## 2. Experimental

### 2.1 Materials

Two commercially available activated charcoals GR MERCK 2518 and GAC Norit 1240 Plus (A- 10128) were chosen as adsorbents. The activated carbons were used as supplied (parent carbons) and after their oxidative post treatments. Chemical treatment aimed at introduction of the surface oxygen functional groups on the carbon surface. In some conditions, the chemical treatments also changed the carbons porous texture.

#### 2.1.1 Surface modification

Commercial activated charcoals GR MERCK 2518 and GAC Norit 1240 Plus (A- 10128) have been subjected to the post-chemical treatment with 1 M nitric acid at boiling temperature during 6 h. The oxidized materials, were subsequently washed with distilled water until neutral media, and dried in an oven at 110 °C for 24 h.

#### 2.1.2 Surface characterization

The textural characterization of the carbon samples was based on nitrogen adsorption isotherms at 77K. These experiments were carried out with Surface Area & Porosimetry Analyzer, Micromeritics ASAP 2010 apparatus. Prior to the adsorption testing, the samples were outgassing at 240 °C for 24 h under a pressure of  $10^{-3}$  Pa. The apparent surface areas were determined from the adsorption isotherms using the BET equation; the Dubinin-Raduskhevich and B.J.H. methods were applied respectively to determine the micro- and mesopores volume. The oxidation treatment resulted in reduction of the apparent surface area with mesopores formation (Table 1).

The carbon's point zero charge ( $pH_{PZC}$  values) were obtained by acid-base titration (Sontheimer, 1988).  $pH_{PZC}$  decreases when the carbon surface is treated with nitric acid (Table 1). The parent carbons and their oxidized forms were characterized by elemental and proximate analyses using an Automatic CHNS-O Elemental Analyzer and a Flash EATM 1112 (Table 2). The oxygen content significantly increases when the carbon surface is treated with nitric acid.

The carbon surface was also characterized by temperature-programmed desorption with a Micromeritics TPD/TPR 2900 equipment. A quartz microreactor was connected to a mass spectrometer set up (Fisons MD800) for continuous analysis of gases evolved in a MID (multiple ion detection) mode. Surface oxygen groups on carbon materials decomposed

upon heating by releasing CO and CO<sub>2</sub> at different temperatures (Table 3). The assignment of the TPD peaks to the specific surface groups was based on the data published in the literature (Figueiredo, 1999). Thus, a CO<sub>2</sub> peak results from decomposition of the carboxylic acid groups at low temperatures (below 400 °C), or lactones at high temperatures (650 °C); carboxylic anhydrous decompose as CO and CO<sub>2</sub> at the same temperature (around 650 °C). Ether (700 °C), phenol (600-700 °C) and carbonyls/quinones (700-980 °C) decompose as CO. The treatment by nitric acid resulted in an increase in carboxylic acids and anhydrous carboxylic, lactones and phenol groups.

Carbons	S <sub>BET</sub> , (m <sup>2</sup> /g)	V <sub>total</sub> , (cm <sup>3</sup> /g)	V <sub>micro</sub> , (cm <sup>3</sup> /g)	S <sub>meso</sub> , (m <sup>2</sup> /g)	S <sub>micro</sub> , (m <sup>2</sup> /g)	pH <sub>PzC</sub>
Merck_initial	755	0.33	0.31	41	714	7.02
Merck_1 M HNO <sub>3</sub>	1017	0.59	0.55	40	977	3.41
Norit_initial	770	0.40	0.32	41	729	6.92
Norit_1 M HNO <sub>3</sub>	945	0.43	0.41	72	873	4.41

Table 1. Textural and surface characteristics of the studied activated carbons.

Carbons	Proximate analysis (wt %)			Elemental analysis (wt %)			
	Moisture	Volatile	Ash	C	H	N	O
Norit_initial	3.9	6.7	2.8	95.2	0.40	0.48	3.90
Norit_1M HNO <sub>3</sub>	1.8	7.9	2.0	87.9	0.60	2.60	8.90
Merck_initial	2.0	9.1	3.2	92.8	0.25	0.40	6.50
Merck_1M HNO <sub>3</sub>	1.7	12.8	2.0	86.3	0.30	0.54	12.80

Table 2. Proximate and elemental analyses of the studied activated carbons

Carbons	Oxygen evolved, (g/100g)		
	CO <sub>2</sub>	CO	CO/CO <sub>2</sub>
Norit_initial	0.49	1.18	2.41
Norit_1M HNO <sub>3</sub>	3.18	5.94	1.86
Merck_initial	0.44	1.15	2.61
Merck_1M HNO <sub>3</sub>	3.05	18.7	6.22

Table 3. Surface oxygen functionality of the studied activated carbons

All chemicals used were of an analytical grade. Salt Cr<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>OH<sub>2</sub>, which is used in the tanning industry, was used as a source of trivalent chromium. Metal standard was prepared by dissolution of Cr (III) salt in pure water, which was first deionized and then doubly distilled. The initial pH of the resulting Cr (III) solution was 3.2. The chromium solution was always freshly prepared and used within a day in order to avoid its aging.

## 2.2 Adsorption process analysis

### 2.2.1 Batch experiments

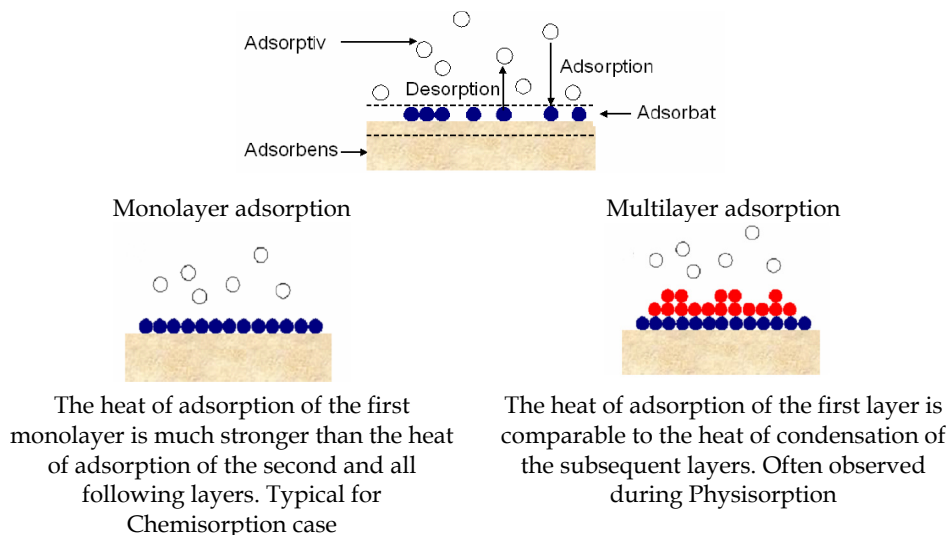
Batch laboratory techniques were utilized to study the equilibrium of Cr (III) adsorption on Norit and Merck activated carbons. The adsorption isotherms were obtained at four different temperatures: 22, 30, 40 and 50 °C. All adsorption isotherms were determined at initial pH of the resulting Cr (III) solution i.e. 3.2, without adding any buffer to control the pH to prevent introduction of any new electrolyte into the systems.

The batch tests were conducted by loading a desirable amount of sorbent to the 250 ml Erlenmeyer flasks containing the Cr(III) solution of fixed (at 200 ppm, which is 10 times lower than the initial concentration present in the tannery wastewater) concentration. Each of the 10 samples used for one experiment consisted of a known carbon dosage from a range 1.2 – 20 g/l in 25 ml of Cr(III) 200 ppm solution, which were shaking on a gyratory shaker at 180 rev/min for 1-7 days (depending on the temperature of the experiment). Each experiment was performed for both initial and post-treated with peroxide, 1 M and acid forms of Norit and Merck carbons, thus generated a total of  $10 \times 2 \times 2 = 40$  samples for each experimental temperature. Furthermore, in some cases, for the batch tests the conditions were changed for fixed carbon loading at 4.8 g/l, whereas Cr(III) concentration were varied from 50 to 2000 ppm. Experiments were duplicated for quality control. The standard deviation of the adsorption parameters was under 1.5 %.

At the end of the experiments, the adsorbent was removed by filtration through membrane filters with a pore size of 0.45  $\mu\text{m}$ . The chromium equilibrium concentration was measured spectrophotometrically, using UV-Visible GBC 918 spectrometer, at fixed wavelength  $\lambda = 420$  nm according to the standard procedure.

### 2.3 Supporting theory

In a typical adsorption process, species/materials in gaseous or liquid form (the adsorptive) become attached to a solid or liquid surface (the adsorbent) and form the adsorbate [Scheme 1], (Christmann, 2010).



Scheme 1. Presentation of the typical adsorption process (after Christmann, 2010)

Since the adsorptive and the adsorbent often undergo a chemical reactions, the chemical and physical properties of the adsorbate is not always just the sum of the individual properties of the adsorptive and the adsorbent, and often represents a phase with new properties (Christmann, 2010).

When the adsorbent and adsorptive are contacted long enough, the equilibrium is established between the amount of adsorptive adsorbed on the carbon surface (the adsorbate) and the amount of adsorptive in the solution. The equilibrium relationship is described by isotherms. Therefore, the adsorption isotherm for the metal adsorption is the relation between the specific amount adsorbed ( $q_{eq}$ , expressed in (mmol) of the adsorbate per (g) of the solid adsorbent) and the equilibrium concentrations of the adsorptive in liquid phase ( $C_{eq}$ , in expressed in (mmol) of the adsorptive per (l) of the solution), when amount adsorbed is equals  $q_{eq}$ .

Chemical equilibrium between adsorbate and adsorptive leads to a constant surface concentration ( $\Gamma$ ) in [mmol/m<sup>2</sup>]. Constant ( $\Gamma$ ) is maintained when the fluxes of adsorbing and desorbing particles are equal, thus the initial adsorptive concentration and temperature dependence of the liquid-solid phase equilibrium are considered (Christmann, 2010).

A common procedure is to equate the chemical potentials and their derivatives of the phases involved. Note: the chemical potential ( $\mu$ ) is the derivative of the Gibbs energy ( $dG$ ) with respect to the mole number ( $n_i$ ) in question (Christmann, 2010), which is for the adsorption process from the liquid phase is the equilibrium concentrations of the adsorptive in liquid phase ( $C_{eq}$ ), when amount adsorbed on the carbon surface is equals ( $q_{eq}$ ) [1]:

$$\mu_i = \left[ \frac{dG}{dn_i} \right]_{p,T, other\ mole\ numbers\ (C_{eq})} \quad (1)$$

The decisive quantities when studying the adsorption process are the heat of adsorption and its coverage dependence to lateral particle-particle interactions, as well as the kind and number of binding states (Christmann, 2010). The most relevant thermodynamic variable to describe the heat effects during the adsorption process is the differential isosteric heat of adsorption ( $\Delta H_x$ , kJ mol<sup>-1</sup>), that represents the energy difference between the state of the system before and after the adsorption of a differential amount of adsorbate on the adsorbent surface (Christmann, 2010). The physical basis is the Clausius-Clapeyron equation [2]:

$$\frac{1}{(C_{eq})} \left[ \frac{d(C_{eq})}{dT} \right]_{\Gamma} = \left[ \frac{d \ln(C_{eq})}{d\left(\frac{1}{T}\right)} \right]_{\Gamma} = -\frac{\Delta H_x}{R} \quad (2)$$

Knowledge of the heats of sorption is very important for the characterization and optimization of an adsorption process. The magnitude of ( $\Delta H_x$ ) value gives information about the adsorption mechanism as chemical ion-exchange or physical sorption: for physical adsorption, ( $\Delta H_x$ ) should be below 80 kJmol<sup>-1</sup> and for chemical adsorption it ranges between 80 and 400 kJmol<sup>-1</sup> (Saha & Chowdhury, 2011). It also gives some indication about the adsorbent surface heterogeneity.

*Langmuir Isotherm:* A model assumes monolayer coverage and constant binding energy between surface and adsorbate [3]:



$$q_{eq} = \frac{K_L \times q_{max} C_{eq}}{1 + K_L C_{eq}} \quad (3)$$

where  $q_{max}$  is the maximum adsorption capacity (monolayer coverage), i.e. mmol of the adsorbate per (g) of adsorbent;

$K_L$  is the constant of Langmuir isotherm if the enthalpy of adsorption is independent of coverage.

The constant  $K_L$  depends on (i) the relative stabilities of the adsorbate and adsorptive species involved, (ii) on the temperature of the system, and (iii) on the initial concentration of the metal ions in the solution. Factors (ii) and (iii) exert opposite effects on the concentration of adsorbed species: the surface coverage may be increased by raising the initial metal concentration in the solution but will be reduced if the surface temperature is raised (Christmann, 2010).

If the desorption energy is equal to the energy of adsorption, then the first-order processes has been assumed both for the adsorption and the desorption reaction. Whether the deviation exists, the second-order processes should be considered, when adsorption/desorption reactions involving rate-limiting dissociation. From the initial slope of a log - log plot of a Langmuir adsorption isotherm the order of adsorption can be easily determined: if a slope is of 1, that is 1<sup>st</sup> order adsorption; if a slope is of 0.5, that is 2<sup>nd</sup> order adsorption process (Christmann, 2010).

*BET (Brunauer, Emmett and Teller) Isotherm:* This is a more general, multi-layer model.

It assumes that a Langmuir isotherm applies to each layer and that no transmigration occurs between layers. It also assumes that there is equal energy of adsorption for each layer except for the first layer [4]:

$$q_{eq} = \frac{K_{BET} \times q_{max} C_{eq}}{(C_{init} - C_{eq}) \times [1 + (K_{BET} - 1) \times (C_{eq} / C_{init})]} \quad (4)$$

where  $C_{init}$  is saturation (solubility limit) concentration of the metal ions (in mmol/l) and  $K_{BET}$  is a parameter related to the binding intensity for all layers;

Two limiting cases can be distinguished: (i) when  $C_{eq} \ll C_{init}$  and  $K_{BET} \gg 1$  BET isotherm approaches Langmuir isotherm ( $K_L = K_{BET} / C_{init}$ ); (ii) when the constant  $K_{BET} \gg 1$ , the heat of adsorption of the very first monolayer is large compared to the condensation enthalpy; and adsorption into the second layer only occurs once the first layer is completely filled. Conversely, if  $K_{BET}$  is small, then a multilayer adsorption already occurs while the first layer is still incomplete (Christmann, 2010). In general, as solubility of solute increases the extent of adsorption decreases.

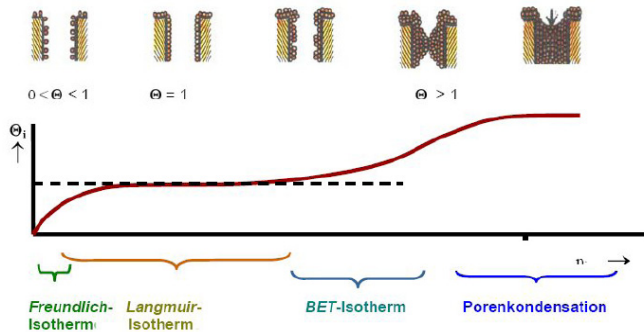
This is known as the "Lundelius' Rule". Solute-solid surface binding competes with solute-solvent attraction. Factors which affect solubility include molecular size (high MW- low solubility), ionization (solubility is minimum when compounds are uncharged), polarity (as polarity increases get higher solubility because water is a polar solvent).

*Freundlich Isotherm:* For the special case of heterogeneous surface energies in which the energy term ( $K_F$ ) varies as a function of surface coverage the Freundlich model are used [5]:

$$q_{eq} = K_F \times C_{eq}^{1/n} \quad (5)$$

where  $K_F$  and  $1/n$  are Freundlich constants related to adsorption capacity and adsorption efficiency, respectively.

To determine which model (Scheme 2) to use to describe the adsorption isotherms for particular adsorbate/adsorbent systems, the experimental data were analyzed using model's linearization.



Scheme 2. Models presentation of the adsorption process (after Christmann 2010), where symbol ( $\theta$ ) is the fraction of the surface sites occupied.

## 2.4 Theoretical calculations

### 2.4.1 Isotherms analysis

The results of Cr (III) adsorbed on activated carbons were quantified by mass balance. To test the system at equilibrium, the following parameters were used: adsorption capacity of the carbon ( $q_{eq}$ ) expressed in terms of metal amount adsorbed on the unitary sorbent mass (mmol/g), i.e.  $[Cr\ III]_{uptake}$ ; and sorption efficiency of the system ( $R\%$ ) indicated from the percentage of removed metal ions relative to the initial amount, i.e.  $[Cr_{Rem}]$ , %. These parameters have been calculated as indicated below [6, 7]:

$$q_{eq} = \frac{(C_{init} - C_{eq})}{m} \quad (6)$$

$$R\% = \frac{(C_{init} - C_{eq})}{C_{init}} 100 \quad (7)$$

where  $C_{init}$  and  $C_{eq}$  are, respectively, the initial and equilibrium concentrations of metal ions in solution (mmol/l) and  $m$  is the carbon dosage (g/l).

The data for the uptake of Cr (III) at different temperatures has been processed in accordance with the linearised form of the Freundlich [8], Langmuir [9] and BET [10] isotherm equations.

For the Freundlich isotherm the log-log version was used [8]:

$$\log q_{eq} = \log K_F + 1/n \log C_{eq} \quad (8)$$

The Langmuir model linearization (a plot of  $1/q_{eq}$  vs  $1/C_{eq}$ ) was expected to give a straight line with intercept of  $1/q_{max}$  [9]:

$$\frac{1}{q_{eq}} = \frac{1}{K_L q_{max}} \frac{1}{C_{eq}} + \frac{1}{q_{max}} \quad (9)$$

The BET model linearization equation [10] was used:

$$\frac{C_{eq}}{(C_{init} - C_{eq})q_{eq}} = \frac{K_{BET} - 1}{K_{BET}q_{max}} \frac{C_{eq}}{C_{init}} + \frac{1}{K_{BET}q_{max}} \quad (10)$$

For a successful determination of a BET model the limiting case of  $K_{BET} \gg 1$  is required. In this case, a plot of  $\left[ \frac{C_{eq}}{(C_{init} - C_{eq})q_{eq}} \right]$  vs  $\frac{C_{eq}}{C_{init}}$  yields a straight line with positive slope and intercept from which the constant ( $K_{BET}$ ) and the monolayer sorption capacity ( $q_{max}$ ) can be obtain.

### 2.4.2 Thermodynamic parameters

Thermodynamic parameters such as change in Gibb's free energy  $\Delta G^0$ , enthalpy  $\Delta H^0$  and entropy  $\Delta S^0$  were determined using the following equation [11]:

$$K_d = \frac{q_{eq}}{C_{eq}} \quad (11)$$

where  $K_d$  is the apparent equilibrium constant,  $q_{eq}$  (or  $[Cr III]_{uptake}$ ); is the amount of metal adsorbed on the unitary sorbent mass (mmol/g) at equilibrium and  $C_{eq}$  (or  $[Cr III]_{eq}$ ) equilibrium concentrations of metal ions in solution (mmol/l), when amount adsorbed is equals  $q_{eq}$ ;

$\frac{q_{eq}}{C_{eq}}$  - relationship depends on the type of the adsorption that occurs, i.e. multi-layer, chemical, physical adsorption, etc.

The thermodynamic equilibrium constants ( $K_d$ ) of the Cr III adsorption on studied activated carbons were calculated by the method suggested by (Khan and Singh, 1987) from the intercept of the plots of  $\ln(q_{eq}/C_{eq})$  vs.  $q_{eq}$

Then, the standard free energy change  $\Delta G^0$ , enthalpy change  $\Delta H^0$  and entropy change  $\Delta S^0$  were calculated from the Van't-Hoff equation [12].

$$\Delta G^0 = -RT \ln K_d, \quad (12)$$

where  $K_d$  is the apparent equilibrium constant;  $T$  is the temperature in Kelvin and  $R$  is the gas constant (8.314 Jmol<sup>-1</sup>K<sup>-1</sup>):

The slope and intercept of the Van't-Hoff plot [13] of  $\ln K_d$  vs.  $1/T$  were used to determine the values of  $\Delta H^0$  and  $\Delta S^0$ ,

$$\ln K_d = \left( \frac{-\Delta H^0}{R} \right) \frac{1}{T} + \frac{\Delta S^0}{R} \quad (13)$$

Then, the influence of the temperature on the system entropy was evaluated using the equations [14]

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 \quad (14)$$

The thermodynamic parameters of the adsorption were also calculated by using the Langmuir constant ( $K_L$ ), Freundlich constants ( $K_F$ ) and the BET constant ( $K_{BET}$ ) for the

equations [12–14] instead of ( $K_d$ ). The obtained data on thermodynamic parameters were compared, when it was possible.

The differential isosteric heat of adsorption ( $\Delta H_x$ ) at constant surface coverage was calculated using the Clausius-Clapeyron equation [15]:

$$\frac{d \ln(C_{eq})}{dT} = - \frac{\Delta H_x}{RT^2} \quad (15)$$

Integration gives the following equation [16]:

$$\ln(C_{eq}) = - \left[ \frac{\Delta H_x}{R} \right] \frac{1}{T} + K \quad (16)$$

where  $K$  is a constant.

The differential isosteric heat of adsorption was calculated from the slope of the plot of  $\ln(C_{eq})$  vs  $1/T$  and was used for an indication of the adsorbent surface heterogeneity. For this purpose, the equilibrium concentration ( $C_{eq}$ ) at constant amount of adsorbate adsorbed was obtained from the adsorption isotherm data at different temperatures according to (Saha & Chowdhury, 2011).

### 3. Results and discussion

#### 3.1 Adsorption isotherms

The equilibrium measurements focused on the determination of the adsorption isotherms. Figures 1–4 show the relationship between the amounts of chromium adsorbed per unit mass of carbon, i.e.  $[\text{Cr(III)}]_{\text{uptake}}$  in mmol/g, and its equilibrium concentration in the solution, i.e.  $[\text{Cr(III)}]_{\text{eq}}$  in mmol/l, at the temperatures of 22, 30, 40 and 50 °C. The carbon adsorption capacity improved with temperature and gets the maximum at 40 °C in the case of the oxidized Norit and Merck carbons and slightly improved with temperature in the case of the parent Norit and Merck activated carbons. The isotherms showed two different shapes. There are isotherms of type III (Fig. 1, 2) for the oxidized samples and of type IV (Fig. 3, 4) for the parent Norit and Merck carbons. Therefore in all cases, the adsorption of the polar molecules (like Cr III solution) on unpolar surface (like the studied activated carbons) is characterized by initially rather repulsive interactions leading to a reduced uptake (Fig. 1, 2), while the increasing presence of adsorbate molecules facilitate the ongoing adsorption leading to isotherms of type III. Furthermore, the porous adsorbents are used and additional capillary condensation effects appeared leading to isotherms of type IV (Fig. 3, 4).

Batch adsorption thermodynamics was described by the three classic empirical models of Freundlich (Eq. 8), Langmuir (Eq. 9) and BET (Eq.10). Regression analysis of the linearised isotherms of Freundlich ( $\log q_{eq}$  vs  $\log C_{eq}$ ) and Langmuir ( $1/q_{eq}$  vs  $1/C_{eq}$ ) and

$\left( \left[ \frac{C_{eq}}{(C_{init} - C_{eq})q_{eq}} \right] \text{ vs } \frac{C_{eq}}{C_{init}} \right)$  using the slope and the intercept of the obtained straight line

gave the sorption constants ( $K_F$ ,  $1/n$  and  $K_L$ ,  $K_{BET}$ ,  $q_{max}$ ). The related parameters for the fitting of Freundlich, Langmuir and BET equations and correlation coefficients ( $R^2$ ) at different temperatures are summarized in Tables 4.

Based on the results, we can conclude that the Freundlich model appeared to be the most “universal” to describe the equilibrium conditions for all studied activated carbons over the

entire range of temperatures, when the Langmuir and BET models were appropriate for one or another of the adsorption systems only.

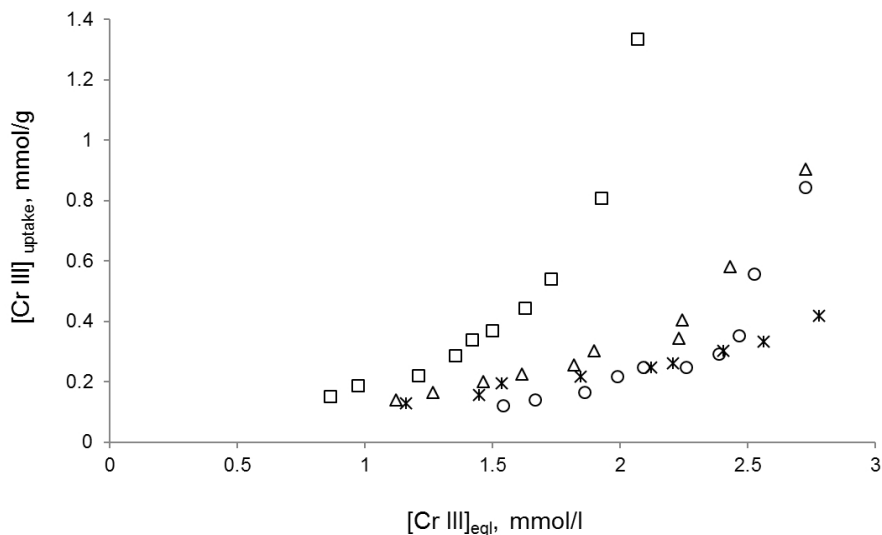


Fig. 1. Isotherms of the Cr (III) adsorption on modified by 1M HNO<sub>3</sub> Norit activated carbon at different temperatures: (\*) - 22; (O) - 30; (□) - 40 and (Δ) - 50 °C.

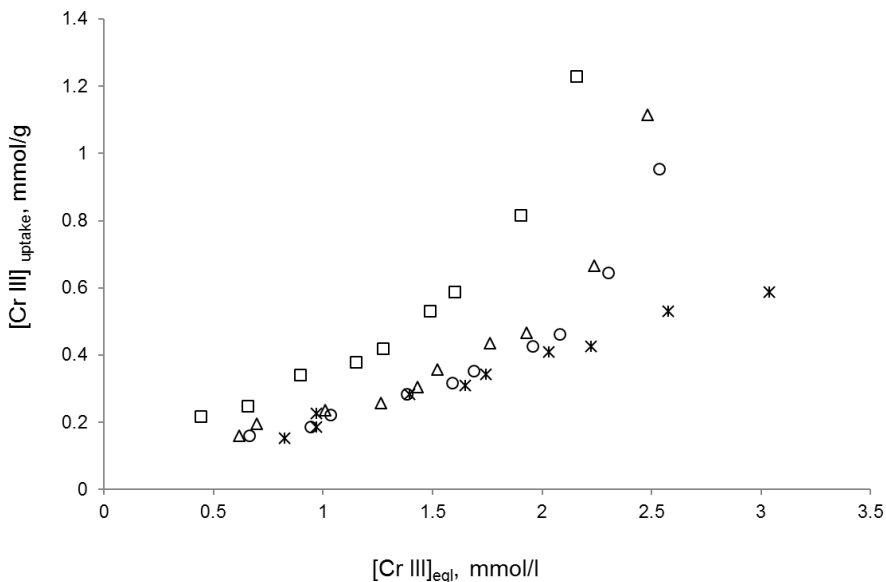


Fig. 2. Isotherms of the Cr (III) adsorption on modified by 1M HNO<sub>3</sub> Merck activated carbon at different temperatures: (\*) - 22; (O) - 30; (□) - 40 and (Δ) - 50 °C.

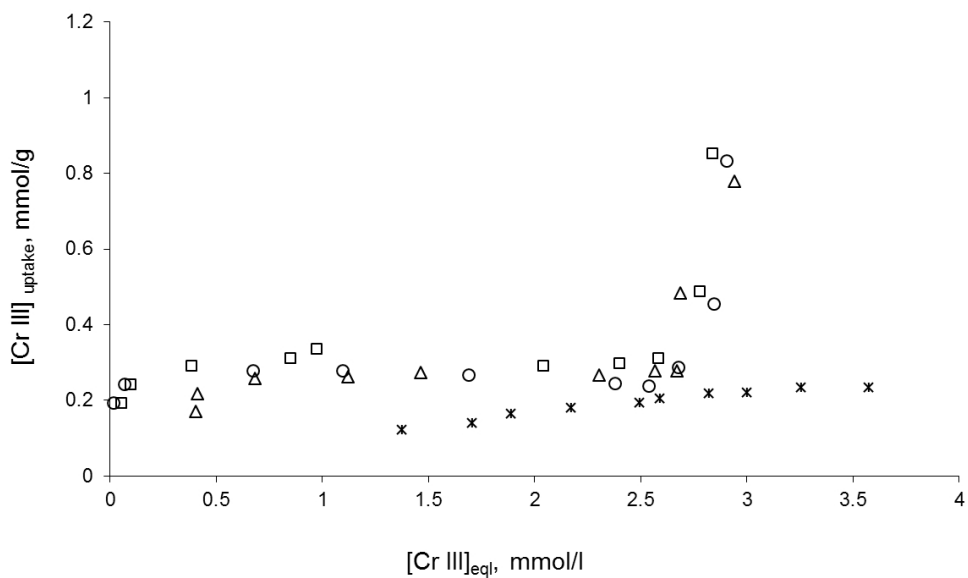


Fig. 3. Isotherms of the Cr (III) adsorption on initial Merck activated carbon at different temperatures: (\*) - 22; (O) - 30; (□) - 40 and (△) - 50 °C.

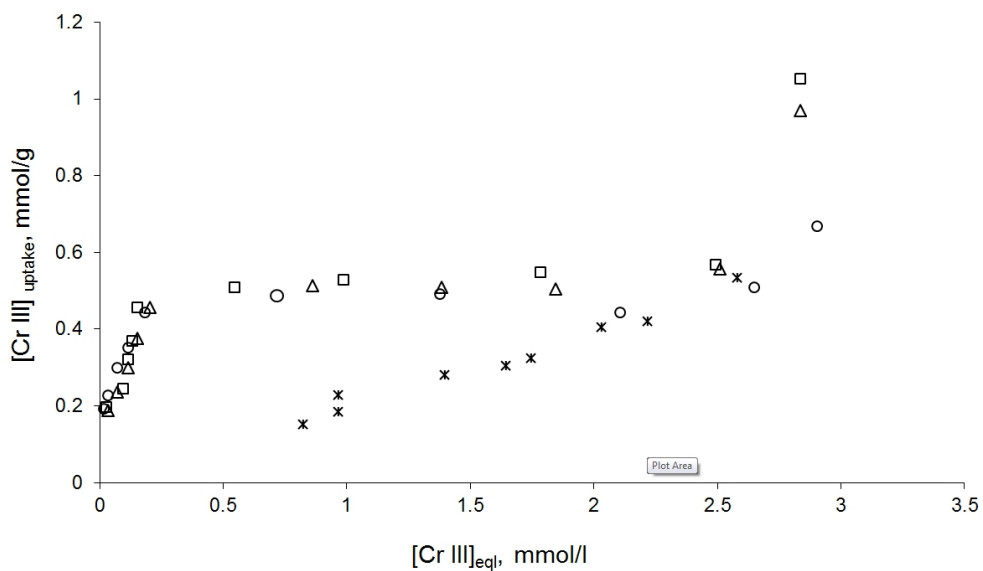


Fig. 4. Isotherms of the Cr(III) adsorption on initial Norit activated carbon at different temperatures: (\*) - 22; (O) - 30; (□) - 40 and (△) - 50 °C.

T, °C	Langmuir constants			Freundlich constants			BET constants			Equilibrium constants		
	R <sup>2</sup>	q <sub>max</sub> , mmol/g	K <sub>L</sub> , l/mmol	R <sup>2</sup>	K <sub>F</sub> , mol/g	1/n	R <sup>2</sup>	q <sub>max</sub> , mmol/g	K <sub>BET</sub>	R <sup>2</sup>	K <sub>d</sub>	
Fixed [Cr III] = 200 ppm, pH3.2												
Merck												
22	Initial	0.7671	0.1290	75.9837	0.9795	10.6608	0.7836	0.9641	0.1197	9.1498	0.0945	4.701
30	Initial	0.7921	0.2617	138.7455	0.5253	3.9487	0.0508	0.9608	0.1092	3.4681	0.1976	6.688
40	Initial	0.7711	0.3332	23.7812	0.6158	3.2485	0.1351	0.9443	0.1164	-16.8765	0.2040	5.445
50	Initial	0.7730	0.3027	4.2890	0.6773	4.2274	0.1748	0.8825	0.0997	-9.3369	0.1677	4.754
22	1M HNO <sub>3</sub>	0.9877	-3.7564	-0.0466	0.9898	5.4386	1.0756	0.9214	0.3525	3.1462	0.9630	3.9361
30	1M HNO <sub>3</sub>	0.9606	2.3453	0.1021	0.9595	4.7632	1.2235	0.6241	0.3164	2.5019	0.9717	4.7063
40	1M HNO <sub>3</sub>	0.9042	2.1961	0.2201	0.9671	2.5448	1.0175	0.5632	0.5651	2.6392	0.9636	5.6350
50	1M HNO <sub>3</sub>	0.9403	2.2412	0.1245	0.9680	4.1034	0.9795	0.8566	0.2990	3.8750	0.9745	5.2799
Norit												
22	Initial	0.9728	0.3509	22.0336	0.6793	3.7895	0.1017	0.9436	0.1931	9.7116	0.1412	3.5450
30	Initial	0.9411	0.4684	31.7875	0.8272	3.7895	0.1820	0.9973	0.4087	176.2481	0.2345	5.0420
40	Initial	0.8679	0.5344	15.4698	0.8058	2.1710	0.2360	0.9899	0.4127	130.3293	0.1845	3.9250
50	Initial	0.9576	0.5419	12.7623	0.8327	1.9333	0.2320	0.9854	0.4020	148.4132	0.0945	4.6290
22	1M HNO <sub>3</sub>	0.9728	-1.0185	-0.0954	0.9644	0.1022	1.2550	0.9641	0.3525	9.5353	0.7945	3.1000
30	1M HNO <sub>3</sub>	0.9688	-0.1399	-0.2946	0.9701	28.9194	2.6228	0.3015	0.2937	3.2066	0.9727	4.3925
40	1M HNO <sub>3</sub>	0.9810	-0.3438	-0.3443	0.9677	9.7227	1.6942	0.7065	0.2134	2.7445	0.9672	5.0415
50	1M HNO <sub>3</sub>	0.9827	-0.4389	-0.2106	0.9588	9.4387	1.6454	0.7735	0.1910	2.7281	0.9860	4.6223
Fixed [Carbon] = 4 g/l, pH3.2												
Merck												
22	Initial	0.9915	0.1159	84.5720	0.9752	1.1620	0.0615	0.9670	0.0689	10.4093	0.0667	3.1676
22	1M HNO <sub>3</sub>	0.9661	1.0690	0.3985	0.9868	3.1705	0.8362	0.9746	0.4179	2.3340	0.9701	5.2972
Norit												
22	Initial	0.9716	0.2756	28.0537	0.9792	3.2751	0.3748	0.9786	0.1157	9.5029	0.1740	3.2031
22	1M HNO <sub>3</sub>	0.9851	0.5617	0.8277	0.9891	0.2496	1.5384	0.9817	0.1720	8.0431	0.9758	4.7848

Table 4. Parameters of the Cr(III) adsorption on studied activated carbons at different temperatures

The Langmuir model was applicable ( $R^2$  ca. 0.96) for the parent Norit carbon, which has low apparent surface area and poor surface oxygen functionality (Tabl. 1, 3), thus indicating strong specific interaction between the surface and the adsorbate and confirmed the monolayer formation on the carbon surface. The lower values of the correlation coefficients ( $R^2$  ca. 0.76) for the parent Merck carbon indicated less strong fitting of the experimental data, most

probably due to less developed porous structure of this carbon. Large values of the Langmuir constant ( $K_L$ ) of ca. 75-140 (which are relative to the adsorption energy) implied a strong bonding on a finite number of binding sites. Langmuir constants (Table 4) slightly increased with temperature increase indicating an endothermic process of the Cr (III) adsorption on studied activated carbons. This observation could be attributed to the increasing an interaction between adsorbent and adsorbate at higher temperatures for the endothermic reactions (Kapoor & Viraraghavan, 1997). There were unfavourable data correlations (the negative values of  $q_{max}$  and  $K_L$ ) for the Langmuir model application (Tabl. 4). It can be seen that the Langmuir model did not fit the adsorption run for the Norit oxidized sample, while it fitted it for the Merck oxidized carbon. Although the Langmuir isotherm model does not correspond to the ion-exchange phenomena, in the present study it was used for oxidized forms of carbon to evaluate their sorption capacity ( $q_{max}$ ). According to the obtained results the oxidized Merck carbon possessed the highest adsorbate uptake (c.f.  $q_{max}$  data, Tabl. 4).

A more general BET (Brunauer, Emmett and Teller) multi-layer model was also used to establish an appropriate correlation of the equilibrium data for the studied carbons. The model assumes the application of the Langmuir isotherm to each layer and no transmigration between layers. It also assumes equal adsorption energy for each layer except the first. It was shown, that in all cases, when Langmuir model failed, the BET model fitted the adsorption runs with better correlations, and an opposite, when Langmuir model better correlated the equilibrium data, BET model was less applicable (c.f. the related parameters for the fitting of Langmuir and BET equations for parent Merck and oxidized Norit, Tabl. 4). Still, in some cases, BET isotherm could not fit the experimental data well (as pointed by the low correlation values) or not even suitable for the adsorption equilibrium expression (for instance, negative values of  $K_{BET}$  Tabl. 4). From the obtained data, three limiting cases are distinguished: (i) when  $C_{eq} \ll C_{init}$  and  $K_{BET} \gg 1$ , BET isotherm approaches Langmuir isotherm ( $K_L = K_{BET}/C_{init}$ ), it was the case of the parent Norit carbon; and (ii) when the constant  $K_{BET} \gg 1$ , the heat of adsorption of the very first monolayer is large compared to the condensation enthalpy and adsorption into the second layer only occurs once the first layer is completely filled, these were the cases of the Cr (III) adsorption by oxidized Merck and Norit carbons; (iii) when  $K_{BET}$  is small, which was the case of the parent Merck carbon, then a multilayer adsorption already occurs while the first layer is still incomplete. In the last case that is most probably connected to the less developed porous structure of the parent Merck.

Based on the obtained results (Tabl. 4), the Freundlich model appeared to be the most "universal" to describe the equilibrium conditions in all studied adsorption systems over the entire range of temperatures. The linear relationships ( $R^2 \sim 0.95-0.99$ ) were observed among the plotted parameters at different temperatures for oxidized samples indicating the applicability of the Freundlich equation. The Cr (III) isotherms showed Freundlich characteristics with a slope of  $\sim 1$  in a log-log representation for the oxidized Merck and Norit activated carbons. These values were in the range of  $\sim 0.2$  for the parent Merck and Norit carbons; and  $1/n$  was found to be more than 2.6 in the case of oxidized Norit carbon. Larger value of  $n$  (smaller value of  $1/n$ ) implies stronger interaction between adsorbent and adsorbate [39]. It is known that the values of  $0.1 < (1/n) < 1.0$  shows that adsorption of Cr (III) is favorable (Mckay et al., 1982) and the magnitude of  $(1/n)$  of to 1 indicates linear adsorption leading to identical adsorption energies for all (Weber & Morris, 1963). Freundlich constants ( $K_F$ ) related to adsorption capacity. In average, these values were in a range of (2-9) and decreased by rising the temperature for all studied carbons.



While Langmuir and BET isotherms indicate the homogeneity of the adsorbent surface and uniform energies of the adsorption, the Freundlich type isotherm hints towards transmigration of sorbate in the plane of the surface and its heterogeneity. Therefore, the surface of studied activated carbons could be made up of small heterogeneous adsorption patches which are very much similar to each other in respect of adsorption phenomenon. Since here the Norit and Merck activated carbons were used as supplied and after post-chemical oxidative treatment, Cr(III) uptake on initial carbons, i.e. those without surface functionality, taken place mainly due to physisorption and increased with the increase in temperature. For oxidized samples total adsorption increases with the temperature until certain temperature, and further temperature rising led to the reversal adsorption capacity when the total adsorption decreases with the temperature. The cross over appears at 40°C. This can be explained by the fact that for carbon reached by surface functionality there is more than one mechanism of chromium sorption: along with the normal physisorption the chemisorption of chromium on the active sites takes place leading to increased adsorption via surface exchange reactions, then with the rise in temperature, i.e.  $T > 40\text{ }^{\circ}\text{C}$ , the ionic exchange is no longer the main mechanism of sorption.

### 3.2 Adsorption thermodynamics

The adsorption process involves a solid phase (adsorbent) and a liquid phase containing a dissolved species (adsorptive) to be adsorbed (adsorbate). The affinity of the adsorbent for the adsorbate determines its distribution between the solid and liquid phases. When the sorption equilibrium is established, the adsorbate immobilized in the solid sorbent is in equilibrium with the residual concentration of adsorptive remaining in the liquid phase.

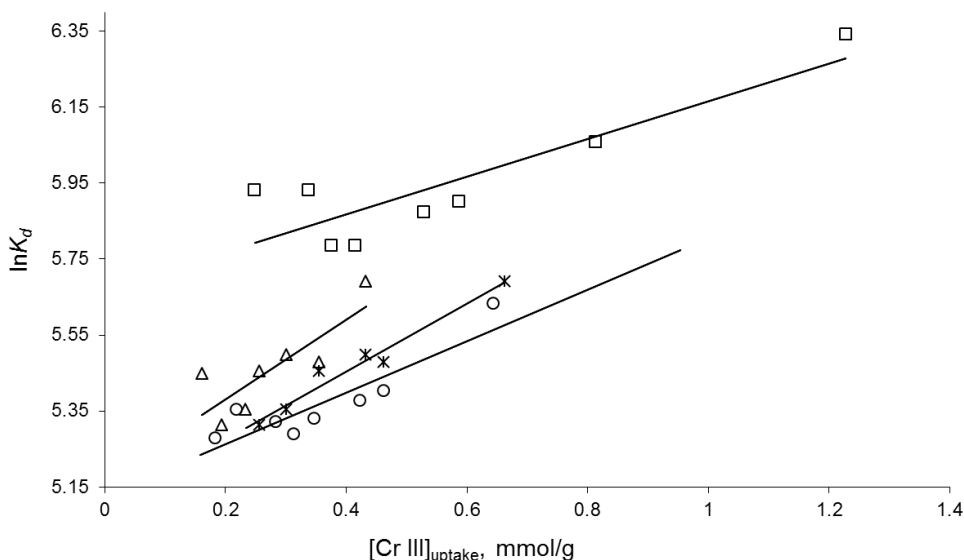


Fig. 5. Plots of  $\ln [Cr III]_{\text{uptake}}/[Cr III]_{\text{eq}}$  vs.  $[Cr III]_{\text{uptake}}$  for the Cr(III) adsorption on modified by 1M  $HNO_3$  Merck activated carbon at (\*) - 22; (O) - 30; (□) - 40 and (△) - 50 °C.

The value for the apparent equilibrium constant ( $K_d$ ) of the adsorption process of the Cr (III) in aqueous solution on studied activated carbons were calculated with respect to temperature using the method of [Khan and Singh] by plotting  $\ln (q_{eq}/C_{eq})$  vs.  $q_{eq}$  and extrapolating to zero  $q_{eq}$  (Fig. 5, 6) and presented in Table. 4. In general,  $K_d$  values increased with temperature in the following range of the studied activated carbons: Merck\_initial < Norit\_initial < Norit\_treated by 1M HNO<sub>3</sub> < Merck\_treated by 1M HNO<sub>3</sub> (Tabl. 4.). However, it should be noted that in the case of the parent Norit and Merck activated carbons, the experimental data did not serve well for the apparent equilibrium constants calculation (as pointed by the low correlation values ( $R^2$ ) on Fig. 7).

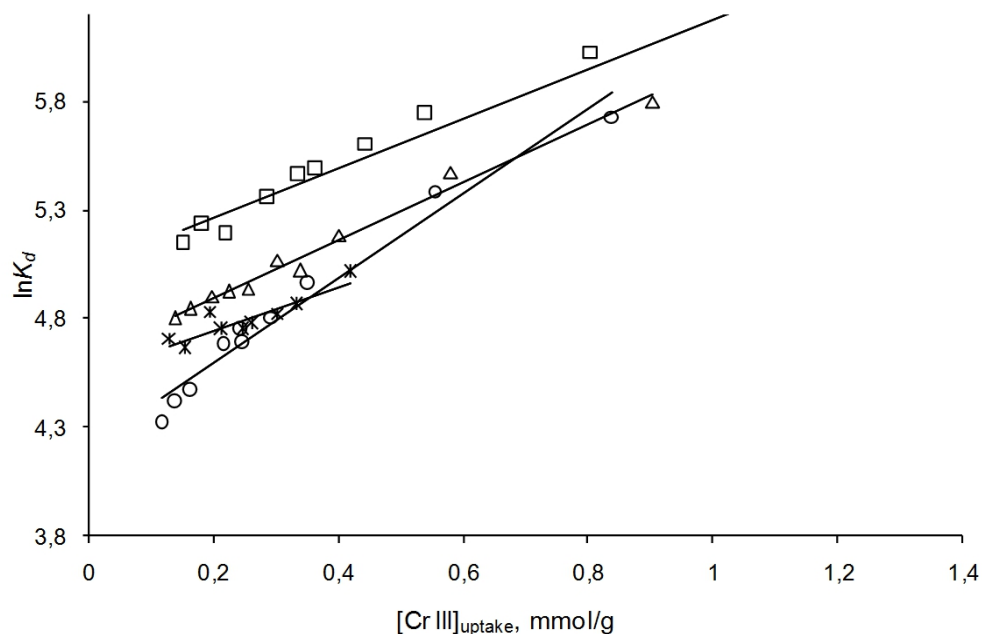


Fig. 6. Plots of  $\ln [Cr(III)]_{uptake}/[Cr(III)]_{eq}$  vs.  $[Cr(III)]_{uptake}$  for the Cr(III) adsorption on modified by 1M HNO<sub>3</sub> Norit activated carbon at (\*) - 22; (O) - 30; (□) - 40 and (Δ) - 50 °C.

As-depicted irregular pattern of linearised forms of  $[\ln (q_{eq}/C_{eq})$  vs.  $q_{eq}]$ , (Fig. 7) are likely to be caused by less developed porous structure of the parent materials and their poor surface functionality, thus low adsorption and, consequently, by the pseudo-equilibrium conditions in the systems with parent activated Norit and Merck carbons.

Thermodynamic parameters for the adsorption were calculated from the variations of the thermodynamic equilibrium constant ( $K_d$ ) by plotting of  $\ln K_d$  vs.  $1/T$ . Then the slope and intercept of the lines are used to determine the values of  $\Delta H^0$  and the equations (13) and (14) were applied to calculate the standard free energy change  $\Delta G^0$  and entropy change  $\Delta S^0$  with the temperature (Table 5).

Based on the results obtained using the thermodynamic equilibrium constant ( $K_d$ ) some tentative conclusions can be given. The free energy of the process at all temperatures was

negative and decreased with the rise in temperature (Fig. 9 (II) and 10 (II)), which indicates that the process is spontaneous in nature is more favourable at higher temperatures. The entropy change ( $\Delta S^0$ ) values were positive, that indicates a high randomness at the solid/liquid phase with some structural changes in the adsorbate and the adsorbent (Saha, 2011). This could be possible because the mobility of adsorbate ions/molecules in the solution increase with increase in temperature and that the affinity of adsorbate on the adsorbent is higher at high temperatures (Saha, 2011). The positive values of  $\Delta H^0$  indicate the endothermic nature of the adsorption process, which fact was evidenced by the increase in the adsorption capacity with temperature (Tabl. 5). The magnitude of  $\Delta H^0$  may also give an idea about the type of sorption. As far as physical adsorption is usually exothermic process and the heat evolved is of 2.1–20.9 kJ mol<sup>-1</sup> (Saha 2011); while the heats of chemisorption is in a range of 80–200 kJ mol<sup>-1</sup> (Saha 2011), and the enthalpy changes for ion-exchange reactions are usually smaller than 8.4 kJ/mol (Nakajima & Sakaguchi, 1993), it is appears that sorption of Cr(III) on studied activated carbons is rather complex reaction. It has to be pointed out, that owing to different operating mechanisms for the Cr (III) adsorption on studied samples, given the  $K_d$  values are not vary linear with the temperature (see Fig. 8 (IV) and the regression coefficients in Tabl. 5) and hence applying of the van't Hoff type equation for the computation of the thermodynamic parameters for the adsorption on the studied carbons is not fully correct, especially in a case of parent carbons (see Fig. 9 (IV) and 10 (IV)).

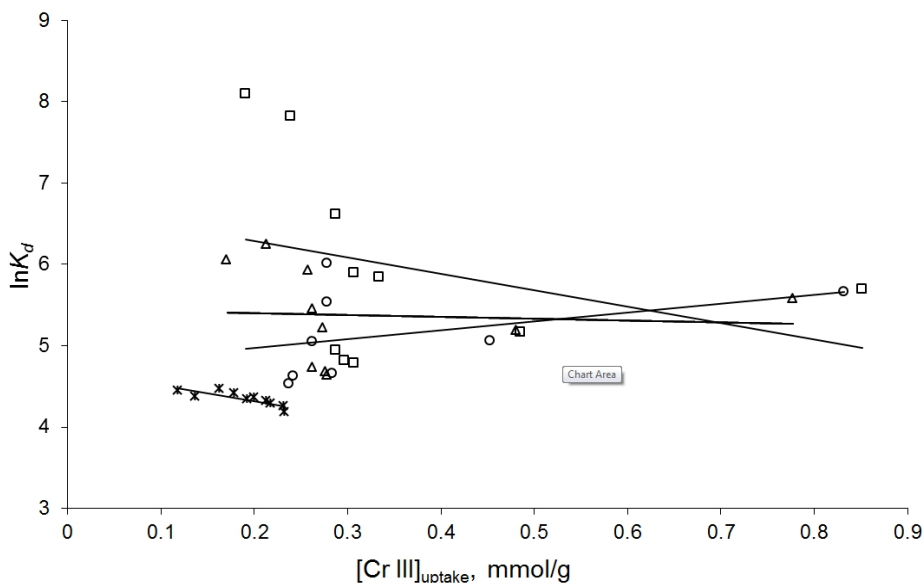


Fig. 7. Plots of  $\ln [Cr(III)]_{uptake} / [Cr(III)]_{eq}$  vs.  $[Cr(III)]_{uptake}$  for the Cr(III) adsorption by parent Merck activated carbon at (\*) – 22; (O) – 30; (□) – 40 and (Δ) – 50 °C.

On the other hand, Langmuir, Freundlich and BET constants showed similar variation with temperature (Fig. 8 (I), (II) and (III)), and hence were also used to calculate the thermodynamic parameters (compare the  $R^2$  for different calculations, Table 5).

T. °C	For Langmuir isotherms				For Freundlich isotherms				For BET isotherms				For Thermodynamic equilibrium constants			
	G kJ mol <sup>-1</sup>	R <sup>2</sup>	S kJ mol <sup>-1</sup>	H kJ mol <sup>-1</sup>	G kJ mol <sup>-1</sup>	R <sup>2</sup>	S kJ mol <sup>-1</sup>	H kJ mol <sup>-1</sup>	G kJ mol <sup>-1</sup>	R <sup>2</sup>	S kJ mol <sup>-1</sup>	H kJ mol <sup>-1</sup>	G kJ mol <sup>-1</sup>	R <sup>2</sup>	S kJ mol <sup>-1</sup>	H kJ mol <sup>-1</sup>
Merck																
22	Initial	10.6544	-	-0.2649	-	-5.8225	-0.0631	-	-5.4464	-	-0.0714	-	-3.8080	-	0.0069	-
30	Initial	12.4649	0.7514	-0.2519	-88.8023	-3.4706	-0.0692	-24.4253	-3.1426	0.999	-0.0771	-26.5045	-4.8021	0.0476	0.0100	-1.7635
40	Initial	-8.2721	-	-0.2573	-	-3.0756	-0.0682	-	-	-	-	-	-4.4239	-	0.0085	-
50	Initial	-3.9223	-	-0.2628	-	-3.8834	-0.0636	-	-	-	-	-	-4.1996	-	0.0075	-
22	1M HNO <sub>3</sub>	-	-	-	-	-4.1666	-0.0281	-	-2.8200	-	0.0297	-	-3.3711	-	0.0415	-
30	1M HNO <sub>3</sub>	5.7662	0.4910	0.0095	8.6466	-3.9445	-0.0281	-12.4485	-2.3174	0.3768	0.0273	5.9464	-3.9141	0.0422	0.0422	8.8801
40	1M HNO <sub>3</sub>	3.9513	-	0.0150	-	-2.4383	-0.0320	-	-2.5333	-	0.0297	-	-4.5134	0.8734	0.0428	-
50	1M HNO <sub>3</sub>	5.6124	-	0.0094	-	-3.8032	-0.0268	-	-3.6489	-	0.0271	-	-4.4823	-	0.0414	-
Norit																
22	Initial	-7.7178	-	-0.0466	-	-3.2777	-0.0436	-	-5.5931	-	0.2376	-	-3.1136	-	0.0253	-
30	Initial	-8.7411	0.4508	-0.0420	-21.4743	-1.9589	-0.0468	-16.1303	-13.0695	0.6100	0.2560	64.4903	-4.0882	0.4316	0.0279	4.3579
40	Initial	-7.1497	-	-0.0458	-	-1.7209	-0.0460	-	-12.7129	-	0.2467	-	-3.5694	-	0.0253	-
50	Initial	-6.8598	-	-0.0452	-	-1.9399	-0.0439	-	-13.4691	-	0.2414	-	-4.1279	-	0.0263	-
22	1M HNO <sub>3</sub>	-	-	-	-	-	-	-	-5.4464	-	-0.0876	-	-2.7836	-	0.0470	-
30	1M HNO <sub>3</sub>	-	-	-	-	-8.5022	-0.1237	-45.9976	-2.9445	0.6414	-0.0936	-31.2994	-3.7397	0.7670	0.0489	11.0731
40	1M HNO <sub>3</sub>	-	-	-	-	-5.9373	-0.1280	-	-2.6355	-	-0.0916	-	-4.2229	-	0.0489	-
50	1M HNO <sub>3</sub>	-	-	-	-	-6.0471	-0.1237	-	-2.7035	-	-0.0885	-	-4.1239	-	0.0470	-

Table 5. Thermodynamic parameters of the Cr III adsorption on studied activated carbons at different temperatures

According to the calculation using ( $K_L$ ), ( $K_F$ ) and ( $K_{BET}$ ) constants (Tabl. 6), the free energy of the processes at all temperatures was negative and increased with the temperature rise (Fig. 9 (I), (II), (III) and Fig. 10 (I), (II), (III)), which indicates spontaneous in nature adsorption processes. While, an increase in the negative value of  $\Delta G^0$  with temperature indicates that the adsorption process is more favorable at low temperatures indicating the typical tendency for physical adsorption mechanism.

The overall process on oxidized carbons seems to be endothermic; whereas that on initial Norit and Merck activated carbons is more evident being exothermic, the negative values of  $\Delta H^0$  in the last case indicate that the product is energetically stable (Tabl. 6). Had the physisorption been the only adsorption process, the enthalpy of the system should have been exothermic. The result suggests that Cr (III) sorption on initial activated carbons is either physical adsorption nor simple ion-exchange reactions, whereas it on oxidized carbons is much more complicated process. Probably, the transport of metal ions through the particle solution interface into the porous carbon texture followed by the adsorption on the available surface sites are both responsible for the Cr (III) uptake.

The negative  $\Delta S^0$  value shows a greater order of reaction during the adsorption on initial activated carbons that could be due to fixation of Cr (III) to the adsorption sites resulting in a decrease in the degree of freedom of the systems. In some cases of oxidized Merck carbon the entropy at all the temperatures positive and is slightly decreases with the temperature with an exception for 40°C. It means that with the temperature the ion-exchange and the replacement reactions have taken place resulted in creation of the steric hindrances (Helfferich, 1962) which is reflected in the increased values for entropy of the system, but at 50°C, these processes are completed and the system has returned to a stable form. Thus it can be concluded that physisorption occurs at a room temperature, ion-exchange and the replacement reactions start with the rise in the temperature and they became less important at  $T > 40^\circ\text{C}$ .

Based on adsorption in-behind physical meaning, some general conclusions can be drawn. When the activated carbon is rich by surface oxygen functionality and has well developed porous structure, including mesopores, the evaluation of the thermodynamic parameters can be well presented by all of ( $K_d$ ), ( $K_L$ ), ( $K_F$ ) and ( $K_{BET}$ ) constants. When similar, but more microporous carbon is used, the thermodynamic parameters is better to present by ( $K_d$ ), ( $K_F$ ) and ( $K_{BET}$ ) constants. However, when the carbon has less developed structure and surface functionality, thermodynamic parameters is better to evaluate based on ( $K_L$ ) and ( $K_F$ ) constants. As a robust equation, Freundlich isotherm fits nearly all experimental adsorption data, and is especially excellent for highly heterogeneous carbons. Therefore ( $K_F$ ) constants can be used for the comparison of the calculated thermodynamic parameters for different activated carbons. However, predictive conclusions can be hardly drawn from systems operating at different conditions and proper analysis will require relevant model as one of the vital basis.

### 3.3 Isostatic heat of the adsorption

The equilibrium concentration  $[\text{Cr III}]_{\text{eq}}$  of the adsorptive in the solution at a constant  $[\text{Cr III}]_{\text{uptake}}$  was obtained from the adsorption data at different temperatures (Fig. 1 - 4). Then isosteric heat of the adsorption ( $\Delta H_s$ ) was obtained from the slope of the plots of  $\ln[\text{Cr III}]_{\text{eq}}$  versus  $1/T$  (Fig. 11, 12) and was plotted against the adsorbate concentration at the adsorbent surface  $[\text{Cr III}]_{\text{eq}}$ , as shown in Fig. 13.

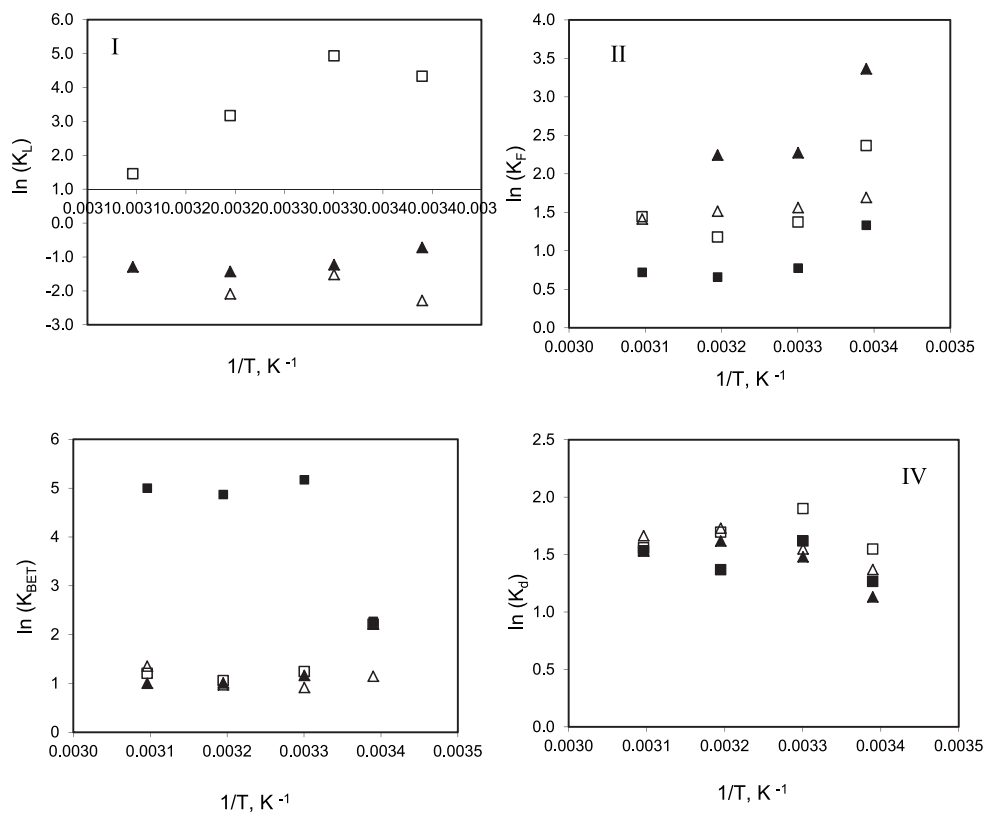


Fig. 8. Plots of Langmuir ( $K_F$ ); Freundlich ( $K_F$ ), BET ( $K_{BET}$ ) and thermodynamic equilibrium constants ( $K_d$ ) vs temperature for the adsorption of Cr(III) on parent Norit (■) and Merck (□) and modified by 1M  $HNO_3$  Norit (▲) and Merck (△) activated carbons.

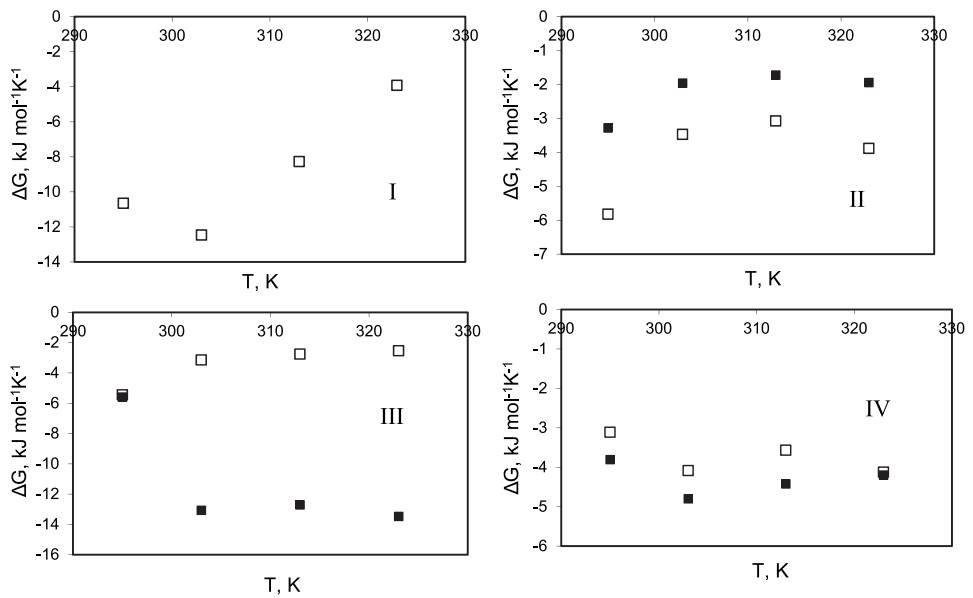


Fig. 9. Plot of Gibb's free energy change ( $\Delta G_0$ ) vs temperature, calculated on Langmuir (I); Freundlich (II), BET (III) and thermodynamic equilibrium (IV) constants for Cr(III) adsorption on parent Norit (■) and Merck (□) activated carbons

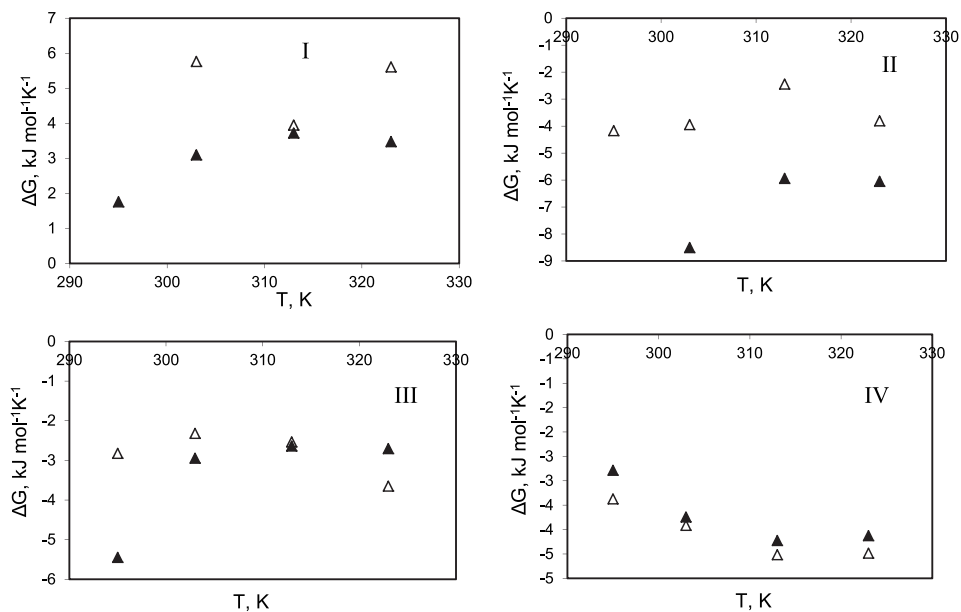


Fig. 10. Plot of Gibb's free energy change ( $\Delta G^0$ ) vs temperature, calculated on Langmuir (I); Freundlich (II), BET (III) and thermodynamic equilibrium(IV) constants for Cr(III) adsorption on modified by by 1M  $\text{HNO}_3$  Norit (▲) and Merck ( $\Delta$ ) activated carbons

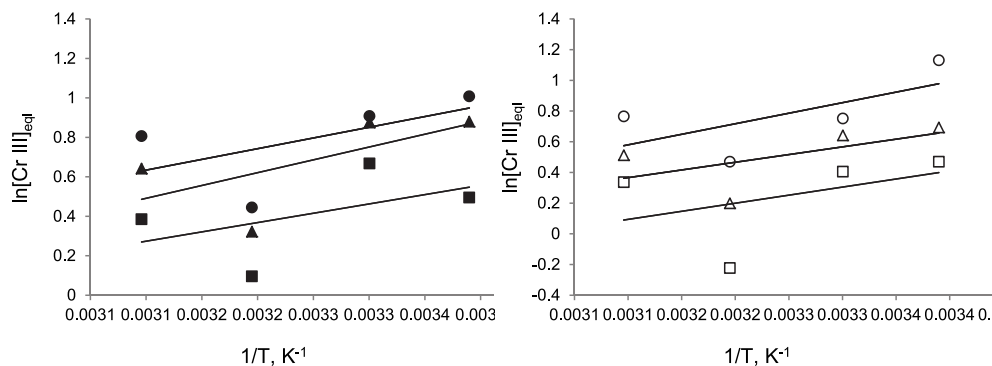


Fig. 11. Plot of  $\ln[\text{Cr III}]_{\text{eq}}$  vs  $1/T, \text{K}^{-1}$ , calculated for the modified activated carbons 1M  $\text{HNO}_3$  Norit : at  $[\text{Cr III}]_{\text{uptake}}$  (●) - 0.4; (■) - 0.3; (▲) - 0.2 mmol/g; and 1M  $\text{HNO}_3$  Merck: at  $[\text{Cr III}]_{\text{eq}}$  (○) - 0.6; (□) - 0.4 and ( $\Delta$ ) - 0.3 mmol/g.



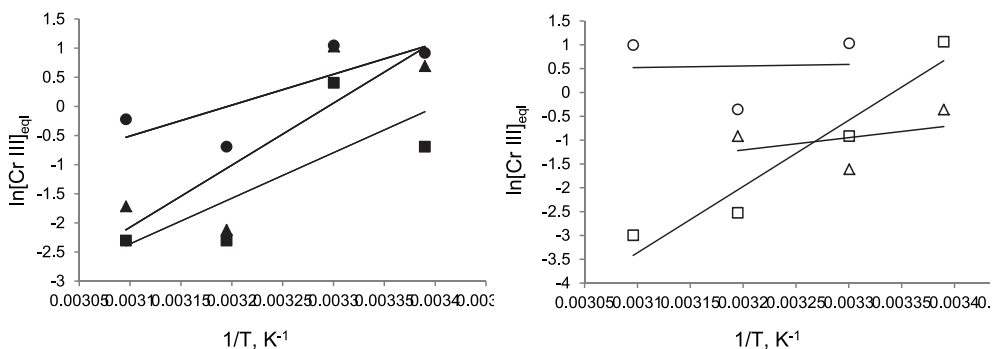


Fig. 12. Plot of  $\ln[Cr III]_{eq}$  vs  $1/T, K^{-1}$ , calculated for the parent Norit : at  $[Cr III]_{uptake}$  (●) - 0.5; (■) - 0.4; (▲) - 0.26 mmol/g; and parent Merck: at  $[Cr III]_{eq}$  (○) - 0.3; (□) - 0.26 and (△) - 0.22 mmol/g.

The plots revealed that  $(\Delta H_x)$  is dependent on the loading of the sorbate, indicating that the adsorption sites are energetically heterogeneous towards Cr III adsorption. For oxidized by 1M  $HNO_3$  Norit and 1M  $HNO_3$  Merck activated carbons (Fig. 13), the isosteric heat of adsorption steadily increased with an increase in the surface coverage, suggesting the occurrence of positive lateral interactions between adsorbate molecules on the carbon surface (Do 1998). In contrary, for the parent Norit and Merck activated carbons (Fig. 13), the  $(\Delta H_x)$  is very high at low coverage and decreases sharply with an increase in  $[Cr III]_{uptake}$ . It has been suggested that the high  $(\Delta H_x)$  values at low surface coverage are due to the existence of highly active sites on the carbon surface. The adsorbent-adsorbate interaction takes place initially at lower surface coverage resulting in high heats of adsorption. Then, increasing in the surface coverage gives rise to lower heats of the adsorption (Christmann, 2010). The magnitude of the  $(\Delta H_x)$  values ranged in 10-140  $kJ mol^{-1}$  revealed that the adsorption mechanism for the studied activated carbons is complex and can be attributed to the combined chemical-physical adsorption processes.

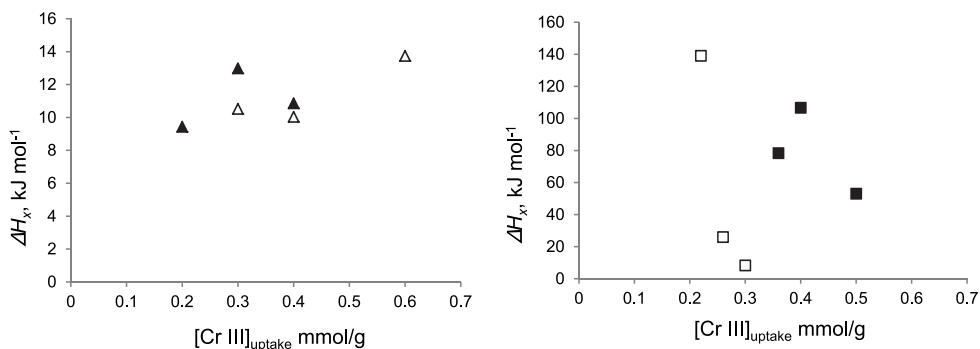


Fig. 13. Plot of isosteric heating  $(\Delta H_x)$  as a function of the amount adsorbed of the parent Norit (■) and Merck (□) activated carbons and their oxidized by 1M  $HNO_3$  Norit (▲) and 1M  $HNO_3$  Merck (△) forms.

### 3.4 General remarks

It should be stressed, however, that the interpretation of the results presented here is tentative. According to our previous investigation on the equilibrium for the studied systems at different pHs and at a room temperature there are both slow and fast Cr(III) uptakes by Norit and Merck carbons (Lyubchik, 2005). The actual time to reach equilibrium is strongly depended on the initial and equilibrium pH of the solution, as well as on the surface functionality and material texture, and was varied between 0.5 and 3 months for different carbons at different pHs. The process did not appear to achieve equilibrium over the time interval used for the batch experiment of ca. 0.5-1 month, especially for the carbons reached by surface functionality (i.e. those modified by nitric acid), as well as for the all systems at moderated acidic pH values, i.e. pH 2 and 3.2. Thus, for the Norit and Merck carbons treated by 1 M HNO<sub>3</sub> the chromium removal increased from 40–50 % to 55–65 % as the contact time is increased from 0.5 to 3 months at pH 3.2. At pH 3.2 the carbon's surface might have different affinities to the different species of chromium existing in the solution. Under real equilibrium conditions our results showed that studied Merck activated carbons adsorb Cr (III) from the aqueous solution more effective then corresponded Norit samples. It is related to the microporous texture of Norit carbons that could be inaccessible for large enough Cr (III) cations (due to their surrounded layers of adsorbed water).

This finding points out that the chosen current conditions for batch experiment at different temperatures could be out of the equilibrium conditions for the studied systems. Therefore current analysis of the thermodynamic parameters should be corrected taking into account the behaviors of the systems in complete equilibrium state.

## 4. Conclusion

The adsorption isotherms are crucial to optimize the adsorbents usage; therefore, establishment of the most appropriate correlation of an equilibrium data is essential. Experimental data on adsorption process from liquid phase on activated carbon are usually fitted to several isotherms, where Langmuir and Freundlich models are the most reported in literature. To determine which model to use to describe the adsorption isotherms the experimental data were analyzed using linearised forms of three, the widespread-used, Langmuir, Freundlich and BET models for varied activated carbons.

As a robust equation, Freundlich isotherm fitted nearly all experimental adsorption data, and was especially excellent for highly heterogeneous adsorbents, like post-treated by HNO<sub>3</sub> Merck and Norit activated carbons. It was shown, that in all cases, when Langmuir model fall-short to represent the equilibrium data, the BET model fitted the adsorption runs with better correlations, and an opposite, when Langmuir model better correlated the equilibrium data, BET model was less applicable. In some cases, chosen models were not able to fit the experimental data well or were not even suitable for the equilibrium data expression. As-depicted irregular pattern of experimental data and applied linearised models are likely to be caused by the complex nature of the studied activated carbons. Different adsorption behavior is related to the varied porous structure, nature and amount of surface functional groups, as well as to the different operating mechanism of the Cr (III) with temperatures rising.

The thermodynamics parameters were evaluated using both the thermodynamic equilibrium constants and the Langmuir, Freundlich and BET constants. The obtained data were compared, when it was possible. Based on adsorption in- behind physical meaning general conclusions were drawn. However, it should be stressed, that the interpretation of the results presented here is tentative. The principal drawback of adsorption studies in a liquid phase is associated with the relatively low precision of the measurements and the long equilibration time that is requires. These factors imply that an extensive experimental effort is needed to obtain reliable adsorption data in sufficient quantity to allow evaluated the process thermodynamics. Therefore, the adsorption experiments are carried out either under pseudo-equilibrium condition when the actual time is chosen to accomplish the rapid adsorption step or under equilibrium condition when the contact time is chosen rather arbitrary to ensure that the saturation level of the carbon is reached. While, the adsorption models are all valid only and, therefore, applicable only to completed equilibration.

## 5. References

- Anirudhan, T. & Radhakrishnan, P. (2011). Adsorptive Removal and Recovery of U(VI), Cu(II), Zn(II), and Co(II) from Water and Industry Effluents. *Bioremediation Journal*, Vol.15, No1, pp. 39-56
- Anirudhan, T. & Radhakrishnan, P. (2008). Thermodynamics and Kinetics of Adsorption of Cu(II) from Aqueous Solutions onto a New Cation Exchanger Derived from Tamarind Fruit Shell. *Journal of Chemical Thermodynamics*, Vol.40, pp. 702-709. ISSN: 0021- 9614.
- Ajmal, M.; Rao, R.; Ahmad, R.; Ahmad, J. & Rao, L. (2001). Removal and Recovery of Heavy Metals from Electroplating Wastewater by using Kyanite as an Adsorbent. *Journal of Hazardous Materials*, Vol.87, pp. 127-137
- Araújo, M. & Teixeira, J. (1997). Trivalent Chromium Sorption on Alginate Beads. *Int. Biodeterioration and Biodegradation* Vol.40, pp. 63-74
- Argun, M.; Dursun, S.; Ozdemir, C. & Karatas, M. (2007). Heavy Metal Adsorption by Modified Oak Sawdust: Thermodynamics and Kinetics. *Journal of Hazardous Materials*, Vol.141, pp. 77-85. ISSN: 0304-3894.
- Aydin, Y.A. & Aksoy, N.D. (2009). Adsorption of Chromium on Chitosan: Optimization, Kinetics and Thermodynamics. *Chemical Engineering Journal*, Vol. 151, pp. 188-194. ISSN: 1385-8947.
- Bailey, S.; Olin, T.; Bricka R. & Adrian D. (1999). A Review of Potentially Low-cost Sorbents for Heavy Metals. *Water Research*, Vol.33, No.11, pp. 2469-2479
- Brigatti, M.; Franchini, G.; Lugli, C.; Medici, L.; Poppi L. & Turci, E. (2000). Interaction between Aqueous Chromium Solutions and Layer Silicates. *Applied Geochemistry*, Vol.15, pp. 1307-1316
- Brown, P.; Gill, S. & Allen, S. (2000). Metal Removal from Wastewater using Peat. *Water Research*, Vol.34, No.10, pp. 3907-3916
- Carrott, P.; Ribeiro Carrott, M.; Nabais, J. & Prates Ramalho, J. (1997). Influence of Surface Ionization on the Adsorption of Aqueous Zinc Species by Activated Carbons, *Carbon*, Vol.35, pp. 403-410

- Chakir, A.; Bessiere, J.; Kacemi, K. & Marouf, B. (2002). A Comparative Study of the Removal of Trivalent Chromium from Aqueous Solutions by Bentonite and Expanded Perlite. *Journal of Hazardous Material*, Vol.95, No1-2, pp. 29-46
- Chen, J. & Lin, M. (2001). Surface Change and Metal Ion Adsorption on H-type Activated Carbon: Experimental Observation and Modelling Simulation by the Surface Complex Formation Approach. *Carbon*, Vol.39, pp. 1491-1504
- Christmann, K. (2010). Adsorption. Lecture Series 2010/2011: "Modern Methods in Heterogeneous Catalysis Research", Institut für Chemie und Biochemie, Freie Universität, Berlin, Available from [http://www.fhirlin.mpg.de/acnew/departament/pages/teaching/pages/teaching\\_wintersemester\\_2010\\_2011/klaus\\_christmann\\_adsorption\\_101105.pdf](http://www.fhirlin.mpg.de/acnew/departament/pages/teaching/pages/teaching_wintersemester_2010_2011/klaus_christmann_adsorption_101105.pdf)
- Csobán, K.; Párkányi-Berka, M.; Joó P. & Behra, Ph. (1998). Sorption Experiments of Cr(III) onto Silica. *Colloids and Surfaces A: Physicochem. Eng. Aspects*, Vol.141, pp. 347-364
- Csobán, K. & Joó, P. (1999). Sorption of Cr (III) on Silica and Aluminium Oxide: Experiments and Modelling. *Colloids and Surfaces A: Physicochem. Eng. Aspects*, Vol.151, pp. 97-112
- Do, D.D. (1998). *Adsorption Analysis: Equilibria and Kinetics*. Series on Chemical Engineering, Vol. 2. Imperial College Press, London.
- Farias, R.; Souza, A.; Nunes, L. & Cardoso, V. (2002). Adsorption of Cr(III) and Fe(III) on an Inorganic-Organic Silicon Hybrid Surface. *Colloids and Surfaces A: Physicochem. Eng. Aspects*, Vol.211, No.2-3, pp. 295-298
- Ferro-García, M.; Rivera-Utrilla, J.; Bautista-Toledo, I. & Moreno-Castilla, C. (1998). Adsorption of Humic Substances on Activated Carbon from Aqueous Solutions and Their Effect on the Removal of Cr(III) Ions. *Langmuir*, Vol.14, pp. 1880-1886
- Figueiredo J.; Pereira M.; Freitas M. & Orfão J. (1999). Modification of the surface chemistry of activated carbons. *Carbon*, Vol.37, pp. 1379-1389
- Helfferich, F. (1962). *Ion Exchange*, McGraw-Hill, New York, 624 pp.
- Indraswati, N. & Ismadjia, S. (2009). Equilibrium and Kinetic Studies in Adsorption of Heavy Metals using Biosorbent: A
- Kapoor, A. & Viraraghavan, T. (1997). Nitrate Removal from Drinking Water. *Journal of Environmental Engineering*, Vol.123, No.4, pp. 370-379
- Khan, A. & Singh, R. (1987). Adsorption Thermodynamics of Carbofuran on Sn(IV) a Rsenosilicate in H<sup>+</sup>, Na<sup>+</sup> and Ca<sup>+</sup> Forms. *Colloids and Surfaces*, Vol.24, pp. 33-42.
- Kołodyńska, D. (2010). The Effects of the Treatment Conditions on Metal Ions Removal in the Presence of Complexing Agents of a New Generation. *Desalination*, Vol.263, No1-3, pp. 159-169
- Kratochvil, D. & Volesky, B. (1998) Advances in the Biosorption of Heavy Metals, *Trends in Biotechnology*, Vol.16, pp. 291-300
- Kumar, A.; Rao, N. & Kaul, S. (2000). Alkali-treated Straw and Insoluble Straw Xanthate as Low Cost Adsorbents for Heavy Metal Removal - Preparation, Characterization and Application. *Bioresource Technology*, Vol.71, pp. 133-142
- Lakatos, J.; Brown S. & Snape, C. (2002). Coals as Sorbents for the Removal and Reduction of Hexavalent Chromium from Aqueous Waste Streams. *Fuel*, Vol.81, pp. 691-698

- Lalvani, S.; Wiltowski, T.; Hübner, Weston, A. & Mandich, N. (1998). Removal of Hexavalent Chromium and Metal Cations by a Selective and Novel Carbon Adsorbent. *Carbon*, Vol.39, pp. 1219-1226
- Leist, M.; Casey R. & Caridi, D. (2000). The Management of Arsenic Wastes: Problems and Prospects. *Journal of Hazardous Materials*, Vol.76, pp. 125-138
- Li, Y.; Xia, B.; Zhao, Q.; Liu, F.; Zhang, P.; Du, Q.; Wang, D.; Li, D.; Wang, Z. & Xia, Y. (2011). Removal of Copper Ions from Aqueous Solution by Calcium Alginate Immobilized Kaolin. *Journal of Environmental Sciences*, Vol.23, No3, pp. 404-411
- Liu, X.; Goodfellow, M. & Zheng, O. (2004). Correlation of pH Dependant Equilibrium Isotherms of Heavy Metal Biosorption with a Modified Freundlich Model *Environmental Technology*, Vol.25, No12, pp. 1341-1348
- Lyubchik, S.; Melo, R., Palma, C. & Fonseca I. (2003). Adsorption Equilibrium in the System "Cr (III) - Activated Carbon, In: *NATO Science Series IV. Earth and Environmental Science, V.24, Role of Interfaces in Environmental Protection*, S. Barany (Ed.), 355-377, Kluwer Academic Publishers, ISSN 1568-1238, Dordrecht, The Netherlands
- Lyubchik, S.; Lyubchik, A.; Lygina, E.; Lyubchik, S.I., Makarova, T.; Vital, J.; Rego, A. & Fonseca, I. (2008). Simultaneous Removal of 3d Transition Metals from Multi-Component Solutions by Activated Carbons from Co-mingled Wastes. *Separation and Purification Technology*, Vol.60, No3, pp. 264-271
- Lyubchik, S.; Perepichka, I.; Galushko, O.; Lyubchik, A.; Lygina, E. & Fonseca, I. (2005). Optimization of the Conditions for the Cr (III) Adsorption on Activated Carbon. *Adsorption*, Vol.11, No5-6, pp. 581-593
- Mckay, G.; Blair, H. & Garden, J. Adsorption of dyes on chitin. I. Equilibrium studies. *Journal of Applied Polymer Science*, Vol.27, pp. 3043-3057
- Myers, L. (2004). Thermodynamics of Adsorption, In: *Chemical Thermodynamics for Industry*, T.M. Letcher, (Ed.), 243-252, Royal Society of Chemistry, ISBN 0-85404-591-0, Cambridge, UK
- Nakajima, A. & Sakaguchi, T. (1993). Uptake and recovery of gold by immobilized persimmon tannin. *Journal of Chemical Technology and Biotechnology*, Vol.57, No4, pp. 321-326
- Perez-Candela, M.; Martin-Martinez, J. & Torregrosa-Macia, R. (1995). Chromium (VI) Removal with Activated Carbons. *Water Research*, Vol.29, pp. 2174-2180
- Raji, C. & Anirudhan, T. (1998). Batch Cr(VI) Removal by Polyacrylamide-grafted Sawdust: Kinetics and Thermodynamics. *Water Research*, Vol.32, pp. 3772-3780
- Ramesh, A.; Lee, D. & Wong, J. (2005). Thermodynamic Parameters for Adsorption Equilibrium of Heavy Metals and Dyes from Wastewater with Low-cost Adsorbents. *Journal of Colloid and Interface Science*, Vol.291, pp. 588-592
- Ruthven, D. (1984). *Principles of Adsorption and Adsorption Processes*, John Wiley & Sons, New York
- Saha, P. & Chowdhury S. (2011). *Insight into Adsorption Thermodynamics*, InTech, ISBN:978-953-307-544-0, Vienna, Austria
- Sontheimer H.; Crittenden J. & Summers R. (1988). *Activated Carbon for Water Treatment*. DVGW-Forschungsstelle am Engler-Bunte-Institut der Universität Karlsruhe (TH), AWWA Research Foundation, Karlsruhe, 126 pp.

- Tikhonova, L.; Goba, V.; Kovtun, M.; Tarasenko, Yu.; Khavryuchenko, V.; Lyubchik, S. & Boiko, A. (2008). Sorption of Metal Ions from Multicomponent Aqueous Solutions by Activated Carbons Produced from Waste. *Russian Journal of Applied Chemistry*, Vol.81, No8, pp. 1348-1355
- Weber, W. & Morris, J. (1963). Kinetics of adsorption on carbon from solutions. *J. Sanit. Eng. Div. Am. Soc. Civ. Eng.* Vol.89, pp. 31-60

# Thermodynamics of Nanoparticle Formation in Laser Ablation

Toshio Takiya<sup>1</sup>, Min Han<sup>2</sup> and Minoru Yaga<sup>3</sup>

<sup>1</sup>*Hitachi Zosen Corporation*

<sup>2</sup>*Nanjing University*

<sup>3</sup>*University of the Ryukyus*

<sup>1,3</sup>*Japan*

<sup>2</sup>*China*

## 1. Introduction

Nanometer-sized particles, or nanoparticles, are smaller than conventional solid-state materials and possess great potential for new, useful properties due to peculiar quantum effects (Roco, M. C., 1998). Highly functional devices synthesized from nanoparticles have been studied for use in various fields, such as semiconductors (Liqiang, J., 2003; Lu, M., 2006), photocatalysis (Liqiang, J., 2004), secondary batteries (Ito, S., 2005; Kim, K., 2009, 2010), superconductors (Strickland, N. M., 2008), and bonding substances (Ide, E., 2005). In the present chapter, we discuss the thermodynamics related to nanoparticle formation.

Cooling processes of expanding vapor evaporated from a solid surface, such as gas evaporation, arc discharge, sputtering, pulsed microplasma and pulsed laser ablation (PLA), have been applied as a method of nanoparticle formation in the gaseous phase (Wegner, K., 2006). The PLA method, under reduced atmospheric pressure, has been found to be especially promising since it provides the following capabilities (Chrissey, D. B., 1994): (i) ablation of target material regardless of melting point due to the high intensity and focused laser beam pulse, (ii) flexibility in choice of atmospheric gaseous species and pressure, (iii) ease of production of the non-equilibrium state of the high-pressure field due to the formation of shock waves, (iv) ability to obtain many different structured materials, from thin films to micrometer-sized particles, by controlling vapor association and condensation, and (v) ease of synthesis of nano-compounds of non-stoichiometric composition by preparing target materials with desired compositional ratios. The PLA method has been widely used for nanoparticle formation because the formed nanoparticles have diameters smaller than 10 nm with low size dispersion and can be formed as basic materials for highly functional devices via effective utilization of these capabilities (Li, S., 1998; Li, Q., 1999; Patrone, L., 1999, 2000; Wu, H. P., 2000; Suzuki, N., 2001; Inada, M., 2003; Seto, T., 2006).

To understand the process of nanoparticle formation by the PLA method, two perspectives are necessary: (i) the thermodynamics of the microscopic processes associated with the nucleation and growth of nanoparticles, and (ii) the thermodynamics of the macroscopic processes associated with the laser irradiated surface of the target supplying the raw

gaseous materials, combined with the surrounding atmosphere, to provide adequate conditions for nucleation and subsequent growth.

Due to its importance in both academia and industry, the chemical thermodynamics of nanoparticle formation in the gaseous phase have been studied extensively (Finney, E. E., 2008). Two processes are important in these studies: (i) homogeneous nucleation, whereby vapors generated in the PLA process reach super-saturation and undergo rapid phase change, and (ii) growth, during which the nanoparticles continue to grow by capturing surrounding atoms and nuclei in the vapor. The size and generation rate of critical nuclei are important factors for understanding the homogeneous nucleation process. To evaluate the generation rate of critical nuclei, we need to know the partition function of each size of nuclei. If an assembled mass of each size of nuclei can be regarded as a perfect gas, then the partition functions can be calculated using statistical thermodynamic methods. However, because it is generally difficult to directly calculate the nucleus partition function and incorporate the calculated results into continuous fluid dynamics equations, what has been used in practice is the so-called surface free energy model, in which the Gibbs free energy of the nanoparticles is represented by the chemical potential and surface free energy of the bulk materials. In contrast, a kinetic theory has been used for treating the mutual interference following nucleation, such as nanoparticle condensation, evaporation, aggregation, coalescence, and collapse, in the nanoparticle growth process.

Since statistical thermodynamics is a valid approach for understanding the mechanisms of nanoparticle formation, microscopic studies have increased aggressively in recent years. In the case of using a deposition process of nanoparticles for thin-film fabrication for industrial use, however, it is necessary to optimize the process by regulating the whole flow field of nanoparticle formation. In cases in which several vapors (plumes) generated during laser ablation are identified as a continuous fluid, macroscopic studies are needed using, for example, continuous fluid dynamics with a classical nucleation model.

Some studies have evaluated the thermodynamics and fluid dynamics that are involved in nanoparticle formation by using tools such as numerical analysis with an evaporation model, a blast wave model, and a plasma model. However, the shock waves generated in the early stage of PLA result in extensive reflection and diffraction which increasingly complicate clarification of the nanoparticle formation process. Up to now, no attempt to introduce shock wave generation and reflection into the plume dynamics has been reported in relation to nanoparticle formation. We note in particular that thermodynamic confinement could occur at the points of interference between the shock wave and the plume, and that nanoparticles with uniform thermodynamic state variables subsequently could be formed in the confinement region, thus making such a system a new type of nanoparticle generator.

In Section 2 of the present chapter, we review the thermodynamics and fluid dynamics of nanoparticle formation during PLA. After providing analytical methods and models of 1D flow calculation in Section 3, we present the calculation results for laser-irradiated material surfaces, sudden evaporation from the surfaces, Knudsen layer formation, plume progression, and shock wave generation, propagation, and reflection. Extensive 2D flow calculation results (without nanoparticle formation) are presented in Section 4 to explore the flow patterns inside the new type of nanoparticle generator. The experimental results for the various nanoparticles formed by the generator are presented in Section 5. Finally, conclusions are given in Section 6.



## 2. Thermodynamics of nanoparticle formation

### 2.1 Nucleation and growth of nanoparticles

In nanoparticle formation, the following stages must be considered: (i) homogeneous nucleation, where vapor atoms produced by laser ablation have been supersaturated, and (ii) particle growth, where the critical nuclei are growing, capturing atoms on their surfaces, and making the transition into large particles.

At the first stage of homogeneous nucleation, the nucleation rate and the size of critical nuclei are important factors. The nucleation rate,  $I$ , is the number of nuclei that are created per unit volume per unit time. To evaluate the nucleation rate, the number density of nanoparticles at equilibrium is needed. In the present case, it is assumed that the nanoparticles are grown only in the capture of a single molecule without causing other nuclei to collapse. That is, when a nanoparticle consisting of  $i$  atoms is indicated by  $A_i$  (hereinafter,  $i$ -particle), the reaction process related to the nanoparticle formation is expressed as follows:



If the molecular partition functions of the various sizes of nanoparticle are derived by statistical mechanical procedure, the equilibrium constants for each equation are known. As a result, the number density of the nanoparticles at equilibrium can be inferred assuming ideal gas behavior. Namely, the equilibrium constant  $K_{i-1,i}$  between ( $i-1$ )-particle and  $i$ -particle is

$$K_{i-1,i} = \frac{Q_i}{Q_{i-1}Q_1} \exp\left(\frac{D_{i-1,i}}{kT}\right)
 \tag{2}$$

Here,  $Q_i$  is the  $i$ -particle partition function,  $D_{i-1,i}$  is the dissociation energy of one atom for the  $i$ -particle,  $k$  is the Boltzmann constant, and  $T$  is the temperature of the system. In general, to explicitly calculate the Gibbs' free energy change from the molecular partition function of nanoparticles and to incorporate these into a continuous fluid dynamics equation are extremely difficult. Therefore, the so-called surface free energy model, where Gibbs' free energy change is represented by the surface tension and chemical potential of bulk materials, can be adopted. Furthermore, when assuming a steady reaction process for nanoparticle formation, the critical nucleation rate,  $I$ , is represented as (Volmer, M., 1939)

$$I = \frac{n^2 c v_c}{4r_*} \sqrt{\frac{3W_*}{\pi kT}} \exp\left(-\frac{W_*}{kT}\right)
 \tag{3}$$

where  $n$  is the number density of species in the vapor,  $c$  is the average relative speed between nanoparticles and atomic vapor,  $v_c$  is the volume per atom in the vapor,  $r_*$  is the radius of the critical nuclei,  $W_*$  is the energy of formation for critical nuclei,  $k$  is Boltzmann constant, and  $T$  is the temperature of the system. The exponential term appeared in the above formula seems to be an essential factor for thermodynamic considerations in

nucleation process. The supersaturation,  $S$ , which is implicitly included in the variable  $W$ , is a dominant factor which significantly affects nucleation rate,  $I$ .

Once the Gibbs' free energy change,  $G$ , is known, the critical nucleus radius,  $r_*$ , can be easily obtained. For this, an assumption of capillary phenomena (capillarity assumption) is used as a condition for mechanical equilibrium of the particles and the extreme value at  $dG=0$  may be considered. When the surface tension of the nanoparticle is depicted by  $\sigma$ , the radius of critical nucleus is

$$r_* = \frac{2\sigma v_c}{kT \ln S} \quad (4)$$

Here, as in the case of nucleation rate, the degree of supersaturation,  $S$ , is what determines the size of the critical nucleus.

Next, it was assumed for convenience that the nanoparticle growth first occurred after its nucleus reached the critical nucleus size. In other words, the Gibbs' free energy of nanoparticle formation begins to decrease after it reaches maximum value at the critical nucleus size. At this time, the number of atomic vapor species condensing per unit area of particle surface per unit time,  $\beta$ , can be determined using the number density,  $N_r$ , of the species in the atomic vapor near the surface of a nanoparticle possessing radius,  $r$ , and assuming the equilibrium Maxwell-Boltzman distribution,

$$\beta = \xi N_r \sqrt{\frac{kT}{2\pi m}} \quad (5)$$

Here,  $\xi$  is the condensation coefficient, which represents the ratio of the number of condensing atoms to colliding atoms, and  $m$  is the mass of the vapor species. When the vapor species are in equilibrium with the nanoparticles, the number density is represented by  $N_{r,eq}$  and the number of atoms evaporating,  $\alpha$ , from the nanoparticle surface per unit time and area is given by

$$\alpha = \xi N_{r,eq} \sqrt{\frac{kT}{2\pi m}} \quad (6)$$

Therefore, the growth rate of the nanoparticle radius is

$$\frac{dr}{dt} = (\beta - \alpha) v_c \quad (7)$$

In this equation, the variable  $a$  is the equilibrium value corresponding to the temperature of the nanoparticle, while the kinetic parameters of the surrounding vapors, which affect significantly the variable  $\beta$ , are dominant.

As mentioned above, when the two processes of nanoparticle nucleation and growth are considered, each parameter governing the processes is different. That is, the degree of supersaturation dominates as a non-equilibrium thermodynamic parameter for nucleation, while the state variables related to the surrounding vapors are important as molecular kinetic parameters for particle growth. Thus, separating the nucleation and growth processes in time by using the difference, could hypothetically lead to the formation of nanoparticles of uniform size.

## 2.2 Thermal analysis and Knudsen layer analysis

In the view of gas dynamics, the PLA process can be classified into (i) evaporation of the target material and (ii) hydrodynamic expansion of the ablated plume into the ambient gas. We make the approximation herein of a pure thermal evaporation process and neglect the interaction between the evaporated plume and the incident laser beam. For the fairly short laser pulses (~10 ns) that are typical for PLA experiments, it is reasonable to consider the above two processes as adjacent stages. The energy of the laser irradiation is spent heating, melting, and evaporating the target material. The surface temperature of the target can be computed using the heat flow equation (Houle, F. A., 1998). For very high laser fluences, the surface temperature approaches the maximum rapidly during the initial few nanoseconds of the pulse. The evaporation process becomes important when the surface temperature of target approaches the melting point. With the laser fluence and pulse duration we considered, thermally activated surface vaporization can reasonably be used to describe the evaporation due to pulsed laser irradiation of the target. The saturated vapor pressure,  $p_v$ , in equilibrium at the target surface can be calculated using the Clausius-Clapeyron equation from the surface temperature,  $T_s$ . The flux of vapor atoms leaving the surface can be written as

$$J = \frac{\eta p_v}{\sqrt{2\pi k T_s m}} \quad (8)$$

where  $\eta$  ( $\approx 1$ ) denotes the sticking coefficient of surface atoms and  $m$  is the atomic mass of the vapor atom. The total number of ablated atoms is an integration of  $J$  over time and surface area.

To obtain the initial condition for vapor expansion problem, we can perform a Knudsen layer analysis to get the idealized states of the gas just leaving the Knudsen layer (Knight, C. J., 1979). The local density,  $n_0$ , mean velocity,  $u_0$ , and temperature,  $T_0$ , of the vapor just outside the Knudsen layer can be calculated from the jump conditions and may be deduced very simply using

$$\frac{T_0}{T_s} = \left[ \sqrt{1 + \pi \left(\frac{g}{8}\right)^2} - \sqrt{\pi} \frac{g}{8} \right]^2 \quad (9)$$

$$\frac{n_0}{n_s} = \sqrt{\frac{T_s}{T_0}} \left[ \left( g^2 + \frac{1}{2} \right) e^{g^2} \operatorname{erfc}(g) - \frac{g}{\sqrt{\pi}} \right] + \frac{1}{2} \frac{T_s}{T_0} \left[ 1 - \sqrt{\pi} g e^{g^2} \operatorname{erfc}(g) \right] \quad (10)$$

$$u_0 = \sqrt{\kappa \frac{k T_0}{m}} \quad (11)$$

where  $n_s$  is the saturated vapor density at the target surface  $g$  is a function of Mach number and  $\kappa$  is the adiabatic index. The idealized states just beyond the Knudsen layer are calculated by using the above equations (Han, M., 2002).

## 3. One dimensional flow problems

### 3.1 Fluid dynamics of laser ablated plume

Since the processes described above for nanoparticle formation arise in the high temperature plume generated by laser ablation, it is important to know the thermodynamic state of the

species in the plume. The one-dimensional unsteady Euler compressible fluid equation can be obtained using the numerical scheme in order to solve the thermodynamic state of the plume species, as well as to understand the nanoparticle nucleation and growth. Discretization of the system equation was driven by a finite volume method in which a total variation diminishing (TVD) scheme for capturing the shock wave was adopted as a numerical viscosity term. In the present study, because the time evolution of the plume and shock wave interference need to be considered, a three-order precision Runge-Kutta scheme was used as the accurate time calculation.

The conservation equations of mass, momentum, and energy, which describe the behavior of the laser plume in an ambient gas, are as follows (Shapiro, A. H., 1953),

$$\frac{\partial \mathbf{Q}}{\partial t} + \frac{\partial \mathbf{E}}{\partial x} = \mathbf{W} \quad (12)$$

where

$$\mathbf{Q} = [\rho_v \quad \rho_g \quad \rho_m u \quad e \quad C_1 \quad C_2 \quad C_3 \quad C_4]^T \quad (13)$$

$$\mathbf{E} = [\rho_v u \quad \rho_g u \quad p + \rho_m u^2 \quad (e + p)u \quad C_1 u \quad C_2 u \quad C_3 u \quad C_4 u]^T \quad (14)$$

$$\mathbf{W} = [-\dot{\rho}_c \quad 0 \quad 0 \quad \lambda \dot{\rho}_c \quad I \quad \dot{r} C_1 \quad 2\dot{r} C_3 \quad 4\pi \dot{r} C_3]^T \quad (15)$$

Here,  $x$  and  $t$  are distance and time, respectively, and the variables  $\rho$ ,  $u$ ,  $p$  and  $e$  are the density, velocity, pressure, and the total energy per unit volume, respectively. The sub-indices for the vapor, the ambient gas, and the gas mixture are expressed respectively as  $v$ ,  $g$  and  $m$ . Moreover,  $\lambda$  is latent heat for the bulk material of the nanoparticle. In addition, the dotted variables  $\dot{\rho}$  and  $\dot{r}$  represent the time derivative related to the density and the radius of nanoparticle, respectively.  $C_1$ ,  $C_2$ ,  $C_3$ , and  $C_4$  are transient intermediate variables; among these, the last variable,  $C_4$ , also represents the nanoparticle density,  $\rho_c$ .

### 3.2 Calculation model for 1D flow

Figure 1 shows a numerical calculation model of nanoparticle formation during laser ablation. The one-dimensional computational domain, also called the confined space in the present study, is surrounded by a solid wall on the left and a laser target on the right (Takiya, T., 2007, 2010). The confined space is initially filled with ambient gas. The figure represents the initial state of the flow field immediately after laser irradiation. The target surface is melted by laser irradiation and then saturated vapor of high temperature and pressure is present near the surface. Outside it, the Knudsen layer, the non-equilibrium thermodynamic region where Maxwell-Boltzmann velocity distribution is not at equilibrium, appears. Following the Knudsen layer is the initial plume expansion, which is the equilibrium thermodynamic process. In this case, the high temperature and high pressure vapor, which is assumed to be in thermodynamic equilibrium, is on the outer side of the Knudsen layer and is given as the initial conditions for a shock tube problem. In the calculation, the high temperature and high pressure vapor is suddenly expanded, and a

plume is formed forward. With the expansion of the plume, the ambient gas that originally filled the space is pushed away to the right and towards the solid wall.

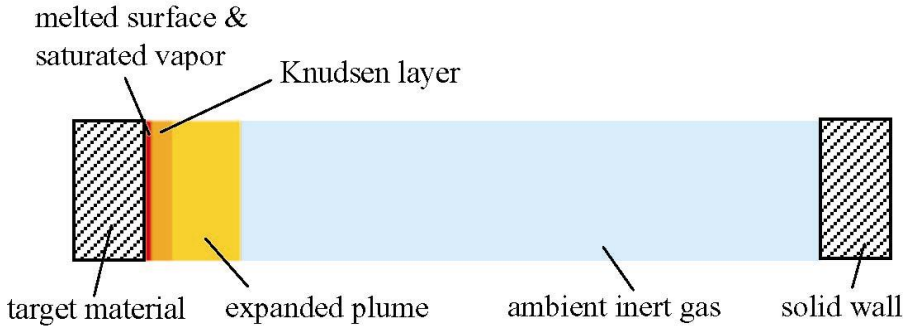


Fig. 1. Calculation model for 1D flow

### 3.3 Physical values and conditions

In this calculation, Si was selected as the target for laser ablation. Physical properties of Si used in the calculations are shown in Table 1 (Weast, R. C., 1965; Touloukian, Y. S., 1967; AIST Home Page, 2006).

As parameters in the simulation, the atmospheric gas pressure,  $P_{\text{atm}}$ , and target-wall distance,  $L_{\text{TS}}$ , may be varied, but conditions of  $P_{\text{atm}} = 100$  Pa and  $L_{\text{TS}} = 20$  mm were most commonly used in the present study. To examine the confinement effect on the nanoparticle formation, however, parametric numerical experiments for  $L_{\text{TS}} = 20, 40, 60, 80,$  and  $200$  mm were also conducted.

Thermal conductivity at 300K	156 [W/(m·K)]
Density	2330 [kg/m <sup>3</sup> ]
Thermal capacity at 298K	712 [J/(kg·K)]
Melting enthalpy	1.804 [kJ/g]
Vaporization enthalpy at 1550K	15.67 [kJ/g]
Melting point	1683 [K]
Boiling point	2628 [K]

Table 1. Physical values of Si

The parameters for laser irradiation of the target, the surface, and the vapor conditions are shown in Table 2. Here, the Laser energy is the energy per single laser pulse, the Laser fluence is the energy density of laser beam having a diameter of 1 mm, the Surface temperature is the temperature of the target surface resulting from the thermal analysis, and the Vapor temperature and Vapor density at the Knudsen layer are the conditions resulting from the Knudsen layer analysis.

Laser energy	20 mJ
Laser fluence	25 mJ / mm <sup>2</sup>
Surface temperature	6100 K
Temperature at Knudsen layer	4050 K
Vapor density at Knudsen layer	1.54 kg / m <sup>3</sup>

Table 2. Parameters for laser irradiation

### 3.4 Typical results for 1D flow

To substantially separate the nucleation and the growth of nanoparticles and facilitate the formation of uniform-sized nanoparticles, the behavior of the shock wave incidentally generated by laser ablation was investigated.

Nanoparticle evaporation is generally thought to be due to an increase in temperature during the passage of shock waves. Therefore, comparatively weak shock waves, which occur in soft laser ablation, were used to promote nanoparticle growth without the evaporation. When soft laser ablation in the confined space was studied, the shock wave and plume were generated, followed by the collision of the reflected shock wave into the plume front. For verification of these processes, a simulation was also carried out with the one-dimensional compressible fluid equations.

A typical flow profile in the calculation showing the change in densities of the Si vapor, helium gas, and nanoparticles between the target surface and the solid wall are shown in Figure 2. Figure 2(a) indicates these densities in the early stages following laser ablation. In general, the silicon vapor atoms in the plume generated by laser ablation are in the electronically excited state by the high energy of the laser. In the plume front, an emission has been observed with de-excitation based on collisions between the vapor atoms and helium gas. Pushing away helium gas by expansion, the plume gradually increases the density in the front region by reaction. Because the ablation laser pulse is limited to a very short time duration, the plume cannot continue to push away helium gas. The clustering of atomic vapors can thus be promoted in the compressed region of plume due to an increase in supersaturation. In front of the plume, it is clearly shown that a shock wave is formed and propagated in helium gas. A transition is observed wherein the plume propagation speed is greater than the speed of the shock wave (Figures 2(b) to 2(d)). On the other hand, while the peak height of plume density progressively decreases, the spatial density of the nanoparticles continues to increase. The shock wave crashes into the right side wall and reflects to the left (Figures 2(e) and 2(f)). In addition, the peak position of nanoparticle density is slightly shifted from the peak position of vapor density. The shock wave is strengthened by reflection to the right side wall, followed by collision with the plume (Figure 2(g)). Figure 2(h) shows the state just after the collision between the reflected shock wave and the plume. The shock wave penetrates into the plume, enhanced the plume density, and thus slightly pushes it back to the left (Figure 2(i)). When the shock wave has completely passed through the plume, the spatial density of nanoparticles effectively increases (Figure 2(j)).

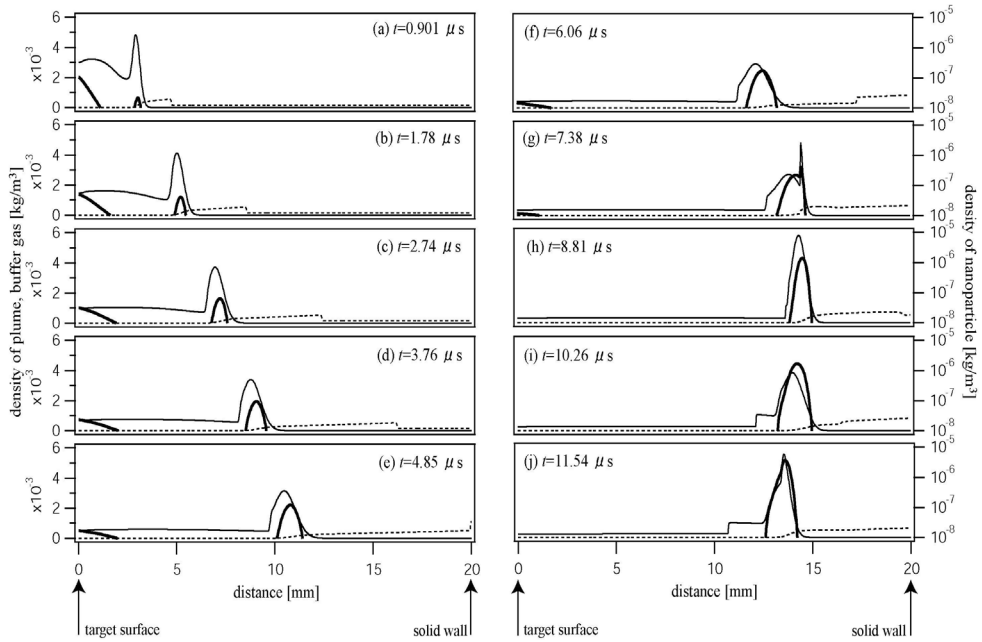


Fig. 2. Typical flow field calculated using the methods and conditions presented in Section 3.3.

### 3.5 Nucleation and growth

Using the same conditions as discussed in the previous section, more detail on the time variation of the state variables is presented in this section.

Figure 3 (a) shows the time variation of the total mass of nanoparticles in the confined space. The horizontal axis is the elapsed time from laser irradiation. This axis is logarithmic to facilitate simultaneous description of the multiple phenomena occurring over several different time scales. The mass of nanoparticles increases between  $0.001 \mu\text{s}$  and  $0.1 \mu\text{s}$  (Figure 3(a)). After  $0.1 \mu\text{s}$ , the mass becomes constant and begins to rise again at  $10 \mu\text{s}$ . The time of the second mass increase is consistent with the moment at which the reflected shock wave collides with the plume. The time variation of the spatially averaged nucleation rate in the confined space is shown in Figure 3(b). The nucleation rate reaches a maximum value at  $0.01 \mu\text{s}$ . The integrated value of nucleation also increases rapidly in the early stages and then becomes constant (Figure 3(c)), which means that the nucleation phenomenon is completed very early on.

The variation of nanoparticle size, which corresponds to the spatially averaged number of atoms constructing the nanoparticle, is shown in Figure 3(d). Since the nanoparticle size starts to increase at  $10 \mu\text{s}$ , when the reflected shock wave arrives at the plume front, it substantially determines the final nanoparticle size, which indicates that the growth of the nanoparticles is facilitated by the effect of the reflected shock wave. Because the nucleation is completed at a very early stage, as already shown, the calculated results also show that nanoparticle growth can be clearly separated from the nucleation process.

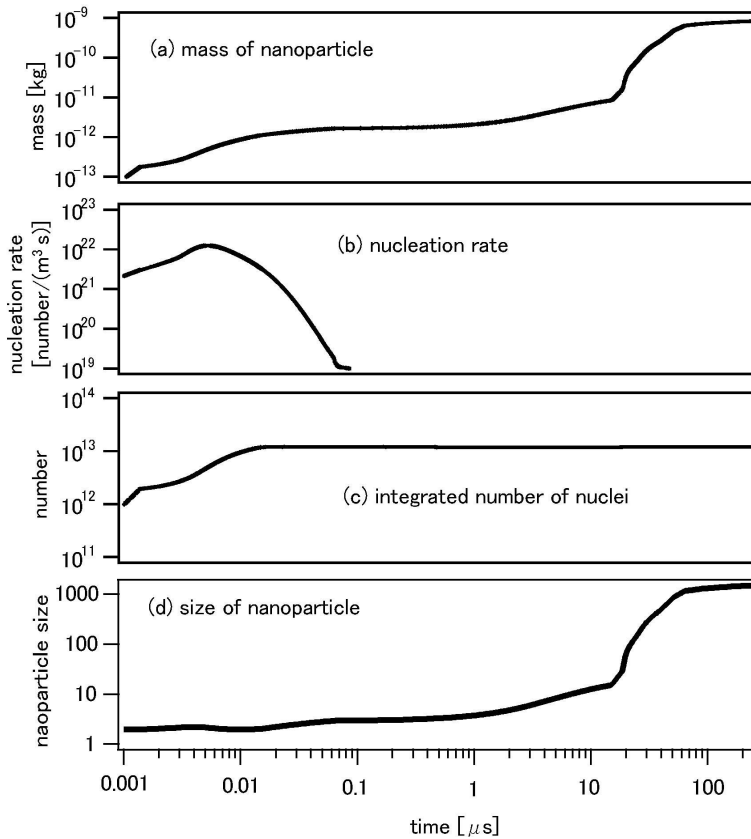


Fig. 3. Time variation of nucleation and growth of nanoparticles

### 3.6 Influence of confinement

The change in nanoparticle size over time was also examined; nanoparticle size increased when the shock wave hit the plume front. Before examining this process further, however, the typical nanoparticle size, as well as the locations of the plume front and the shock wave, must be clearly defined.

There is a definite relationship between the size and spatial density of nanoparticles. The nanoparticle size generally has a distribution, which is especially large in the region of the plume front. The width of the nanoparticle density distribution is smaller than the spread in nanoparticle size and has a sharper distribution profile. The peak positions of the two distributions are almost identical. This means that the maximum nanoparticle size is placed at the location where the nanoparticle density is also at a maximum. Therefore, the typical nanoparticle size in the space can be regarded as the maximum nanoparticle size.

The location of the shock wave propagating through the ambient gas is defined as the maximum value of the derivative for the change in gas density. On the other hand, the plume front is defined as the compression region in the atomic vapor, which comes into contact with the atmospheric gas and high-density area.



The time variation of the nanoparticle size and the positions of the shock wave and the plume, which were defined above, are shown in Figure 4. The left vertical axis is the nanoparticle radius, the right vertical axis is the position in the calculation region, and the horizontal axis is the elapsed time from laser irradiation. The dashed line, thick solid line, and the shaded area represent the nanoparticle size, the position of the shock wave, and the plume front, respectively. The shock waves are propagated backward and forward in the space by reflecting on the target surface and the opposed wall. The width of the shaded area, which represents the plume front, gradually broadens. In addition, the first,  $T_{C1}$ , and second,  $T_{C2}$ , times when the shock wave interferes with the plume front are shown. This interference can be seen as opportunities to enhance the growth rate of nanoparticles. The slope of the dashed line in Figure 4 represents the nanoparticle growth rate, which changes from 17.5 to 52.0  $\mu\text{m/s}$  at  $T_{C1}$ , and from 16.0 to 34.2  $\mu\text{m/s}$  at  $T_{C2}$ . Referring back to Eq. (7), the growth rate of the nanoparticles was determined by a kinetic balance between the condensation rate of nanoparticles, which is based on a macroscopic collision cross-section of the ambient vapors, and the evaporation rate of nanoparticles corresponding to the nanoparticle temperature. Therefore, the fact that the nanoparticle growth rate increases when the shock wave and plume collide means that the shock wave effectively increases the macroscopic collision cross-section.

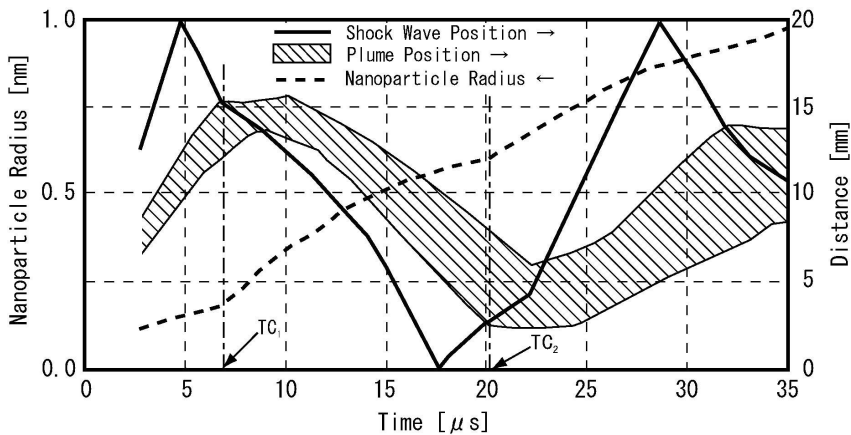


Fig. 4. Increase of nanoparticle radius from interference between shock wave and plume

To investigate the effect of the distance between the target surface and the solid wall on the rate of nanoparticle growth enhanced by the shock wave passage, the numerical simulation was performed under the following conditions:  $L_{TS} = 20, 40, 60, 80,$  and  $200$  mm. The calculated results for the increase of nanoparticle radius are indicated in Figure 5 against the elapsed time from laser irradiation. Nanoparticle growth was promoted by the passage of the shock wave under all of these conditions. The nanoparticle radius,  $r$ , increased with time and eventually reaches a constant value. A balance between the evaporation rate and condensation rate is reached at the maximum radius, and the growth rate of nanoparticles asymptotically approaches zero. When the radius of the nanoparticle is compared among the various distances between the target surface and the solid wall, the shorter  $L_{TS}$  resulted in a larger value of  $r$ . Therefore, larger nanoparticles can be obtained with smaller distances because there are more opportunities for the shock waves to pass through the plume front before the condensation rate balances the evaporation rate.

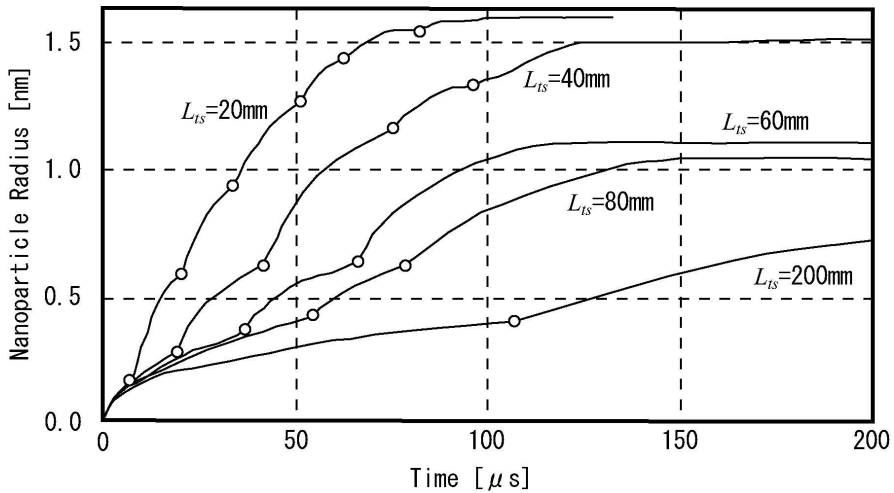


Fig. 5. Influence of distance between target and wall on nanoparticle growth. The circles indicate the arrival time of the shock wave.

## 4. Two dimensional flow problem

### 4.1 Calculation model for 2D flow

On the basis of the results described in the previous section, we have proposed a new method for the direct generation of monodisperse nanoparticles. This method makes use of interaction phenomena between the plume and shock wave arising in an ellipsoidal cell following laser ablation in ambient gas. The method is based on the hypothesis that monodisperse nanoparticles are instantaneously formed inside a narrow region constructed from a diffusion mixture of vapor and ambient gas during the interaction between a plume and shock wave. Such a region forms at one focal point of the ellipsoidal cell, while the plume is ejected from the other focal point with laser irradiation being accompanied by shock waves. Here, the ellipsoidal cell is used as an experimental device based on this principle to obtain uniformly sized nanoparticles, which does not require an additional size classifier like a differential mobility analyzer (Camata, R. P., 1996), and therefore is expected to show high efficiency (Iwata, Y., 2002).

The basic idea of the proposed device, illustrated in Figure 6, is as follows: the target material is exposed to a high-power pulsed laser; the ablated vapor suddenly expands due to high temperature; the expansion results in a propagating shock wave (Figure 6(a)); the vapor is fed by the ablation process for a period of the exposure of the pulse laser; the plume propagates toward the cell exit (Figure 6(b)); the ablation stops after a short duration, while the shock wave and the plume continue to propagate and start to interact (Figures 6(b) and 6(c)); and after the complex interaction between them, the monodispersed nanoparticles are produced and extracted through the cell exit (Figure 6(d)). During the interaction, it is important for the nanoparticles to grow to a certain size.

To investigate the effect of this new model, 2D calculations were performed. For the governing equations, we have chosen the axisymmetric, two-dimensional, compressible Navier-Stokes equations, because the experiments showed that the laser-ablated plume travels straight toward the cell exit with no distortion. The equations are solved by a finite

volume method using the MUSCL-type total variation diminishing (TVD) scheme with a curvilinear generalized coordinate (Yaga, M., 2005, 2008; Fukuoka, F., 2008)

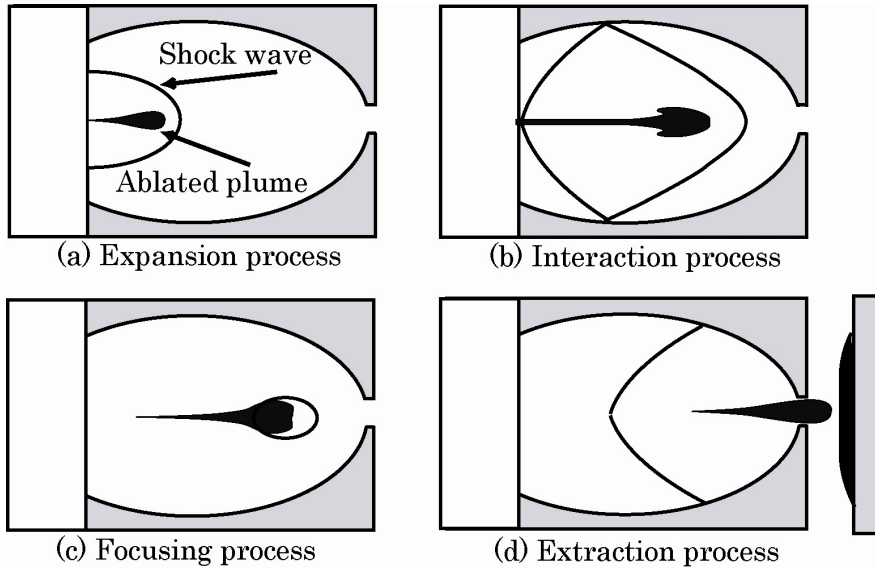


Fig. 6. Behavior of plume and shock wave in an ellipsoidal cell.

#### 4.2 Boundary and initial conditions

The contours of the wall are calculated by the following equation:

$$\sqrt{\frac{x^2}{a^2} + \frac{y^2}{b^2}} = 1 \quad (16)$$

where,  $a$  and  $b$  are constants with a relation of  $a = b(1 + \sqrt{5})/2$ .

For the boundary conditions, non-slip conditions are applied to the cell wall, except for the position of the plume ejection. Outgoing flow conditions are applied to the boundaries outside the cell. The position of the plume ejection is set at one of the focal points of the ellipsoidal cell. The sudden ejection generates a traveling shock wave which converges at the other focal point. The cell exit, through which the flow passes during the propagations of the shock and pressure waves, connects the inside and outside of the cell. During the focusing process of the propagating shock wave, the interaction between the converging shock wave and plume plays an important role in the growing nanoparticle size. An ejected jet of gas is shut off after a certain period so that the calculation can be used for a basic reference for PLA techniques. Then, the ejected gas is considered to be a plume traveling toward the exit of the cell on the right side wall. It is clear that many parameters are involved in this process. We have chosen the three main parameters to be the Mach number, jet duration, and diameter of the exit hole, because, in related experiments, the controllable

parameters are the laser power, duration of the laser pulse, and diameter of the cell exit hole. We therefore assume that the experimental parameters related to the laser power and pulse correspond to the jet Mach number and jet duration, respectively. Hence, we have selected the above three parameters to be tested, fixing all the other parameters. The states of the gas inside and outside of the cell are initially at rest, that is, the ambient properties such as the pressure, density, temperature are uniform over the whole calculation region.

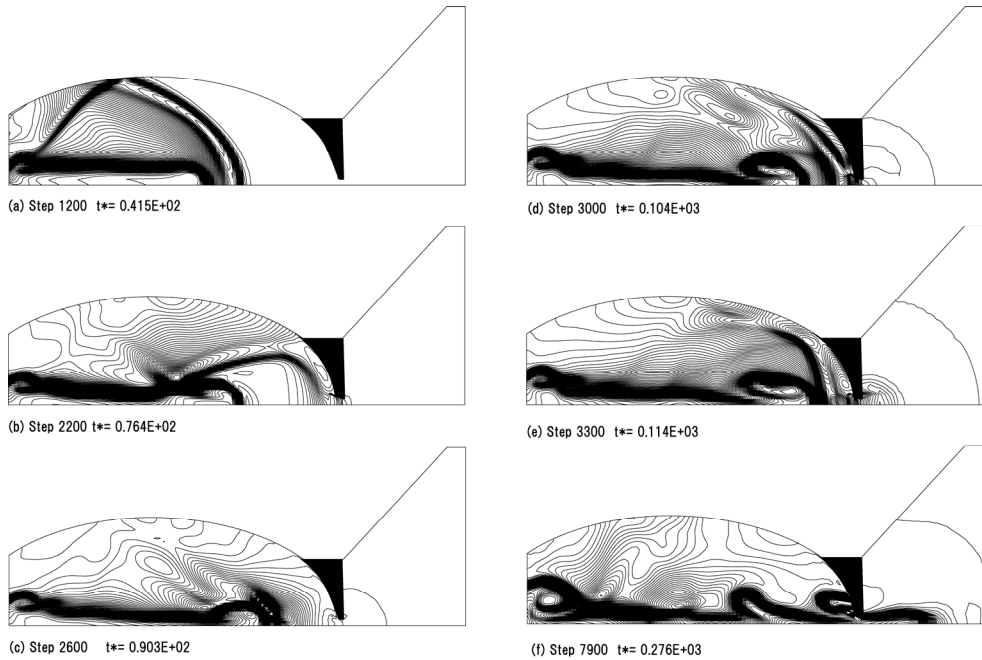


Fig. 7. Time variations of density contours in the ellipsoidal cell.

### 4.3 Shock wave behavior and interaction with plume

Figures 7(a) to (f) show the calculated density fields using certain parameters. As illustrated in the previous section, a shock wave was generated by the sudden expansion of the ejected plume in the ellipsoidal cell. Together with the plume jet, the shock wave propagated towards the right wall of the cell. The plume has decelerated while the shock wave continues to move towards the exit hole (Figure 7(a)). The distance between the plume front and the shock wave increased. The propagating shock wave was reflected from the upper wall of the cell and changed direction toward the focal point. The propagating shock wave arrived at the exit hole and was reflected from the cell wall (Figure 7(b)). The shock wave behind the plume started to interact with the plume front. Figure 7(c) shows the moment when the shock wave was focused at the focal point of the cell and, at the same time, the plume front was located at almost the same point. The plume seems to be blocked by the converging shock wave. Namely, it was recognized that a confinement of the plume was occurred at the time. The transmitted shock wave through the exit hole is so weak that the density contour fields outside of the cell cannot be seen. After the shock wave has

converged at the focal point, it began to spread out again. Most of the spreading shock wave, except for the left traveling shock wave, impinged on the wall again (Figure 7(d)). The shock wave around the exit hole causes weak compression waves to be transmitted into the cell through the hole. After the spreading shock wave was reflected from the wall again (Figure 7(e)), it started to converge at the focal point, where the plume is already present. Then, the shock wave, having been through two reflections, is strong enough to deform the plume and flatten the vortices. The deformed plume front still has enough momentum to transmit through the exit hole (Figure 7(f)). However, part of the plume is left in the cell due to the small diameter of the exit hole. Figure 7(f) suggests that the exit diameter is an important factor in evaluating how much of the plume can get through the exit hole. By choosing the suitable size of exit hole, we can efficiently extract the plume, which once converged at the focal point of the ellipsoidal cell. If these suitable conditions are applied to the actual laser ablation process in the ellipsoidal cell, the laser ablated plume can be confined by a converging shock wave followed by a generation of monodispersed nanoparticles.

## 5. Experimental results

### 5.1 Experimental equipment

The main part of the experimental setup, shown in Figure 8, is composed of the generation and deposition chamber. The laser beam is introduced into the generation chamber. The laser plume is confined by the ellipsoidal cell and uniform-sized nanoparticles are generated. An ambient gas is supplied into the ellipsoidal cell and the cell is filled with the gas. The deposition chamber is connected to the generation chamber through a skimmer, where the nanoparticles are extracted by the ambient gas flow. A substrate is placed in a vertical position for the deposition of the nanoparticles.

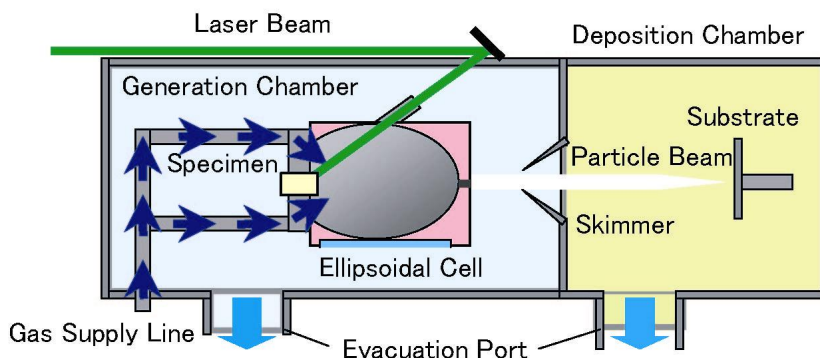


Fig. 8. Schematic diagram of the apparatus for the nanoparticle formation using Pulsed Laser Ablation

When the high-power pulsed laser is directed onto a solid sample in the ellipsoidal cell filled with the ambient gas, the plume is ejected perpendicular to the sample surface. At that time, a shock wave is also generated, driven by the plume expansion, and then propagates in the ambient gas. By placing the sample surface on the focal point of the ellipsoidal cell, the shock wave is reflected on the inner wall of the cell and focused onto the other focal point.

The plume then collides with the ambient gas, which is considerably denser because of the shock wave focusing in the vicinity of this focal point. A mixture region, where the thermodynamic states are uniform, is formed in the boundary between the ambient gas of high density and the plume front and where the mono-disperse nanoparticles are formed.

## 5.2 Typical examples

Experiments were carried out with an ellipsoidal cell having an exit hole diameter of 2 mm, a long axis of 30 mm, and a short axis of 19 mm. An Nd:YAG laser with second harmonic generator ( $\lambda = 532$  nm) was used, and the sample was oxygen-free copper. Helium was chosen as the ambient gas, and the cell pressures were 100, 500, and 1000 Pa. The energy of the laser pulse was 256 mJ, with a pulse duration of 8 ns. The laser irradiated area was measured to be 3.88 mm<sup>2</sup> from observation of a laser trace on the sample surface.

An image of the copper nanoparticles after 30 laser pulses obtained with transmission electron microscopy (TEM) is shown in Figure 9. Although some grain growth due to aggregation of nanoparticles after generation is recognized in the case of gas pressure 1000 Pa, monodisperse nanoparticles less than 10 nm are easily obtainable by properly controlling the ambient gas pressure.

Nanoparticle size distributions analyzed using TEM with a 0.5  $\mu\text{m}$  square field of vision are shown in Figure 10. Using gas pressures of 100 and 500 Pa, nanoparticles with average diameters below 10 nm were obtained. In both cases, the particle size distribution can be approximated using a lognormal distribution function. The geometric standard deviation,  $\sigma$ , ranges from 1.09 to 1.12. Furthermore, it was confirmed in other experiments that the standard deviation can be further reduced by reducing the diameter of the exit hole of the ellipsoidal cell.

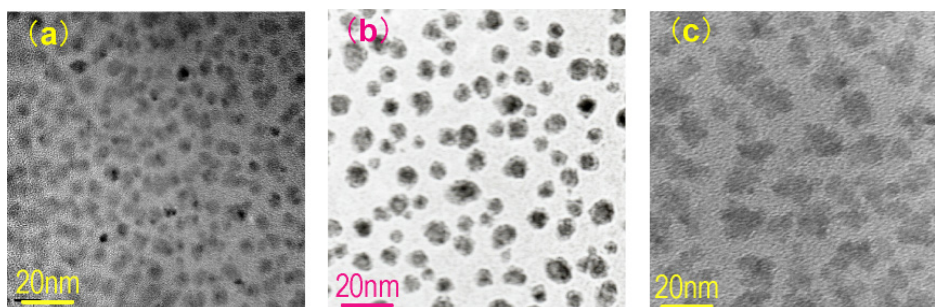


Fig. 9. TEM images of Cu nanoparticles formed under (a)100, (b)500 and (c)1000 Pa of helium gas pressure.

An electron diffraction pattern and corresponding TEM image of copper nanoparticles is shown in Figure 11. By comparing the diffraction pattern of the copper nanoparticles with that of only the carbon film on which the nanoparticles were collected, we confirmed that the copper nanoparticles are as crystalline. Debye Scherrer rings are observed in the electron diffraction pattern where most of Laue spots are very small, suggesting that the crystallized nanoparticles are facing various directions with respect to the nanoparticle crystal axis.

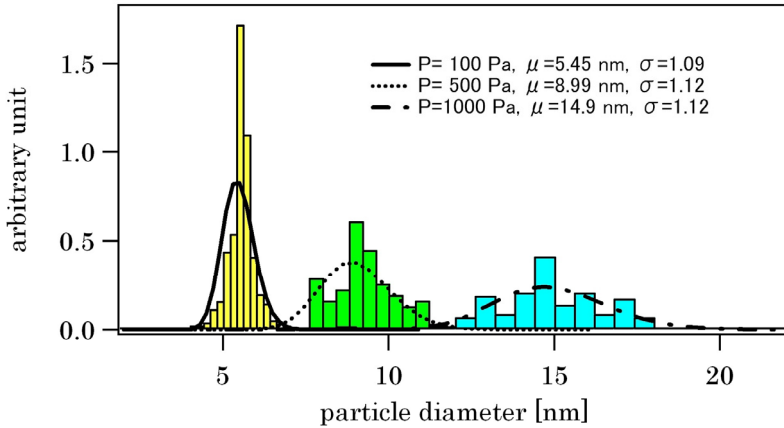


Fig. 10. Histograms showing the diameters and the logarithmic normal distributions of the nanoparticles.

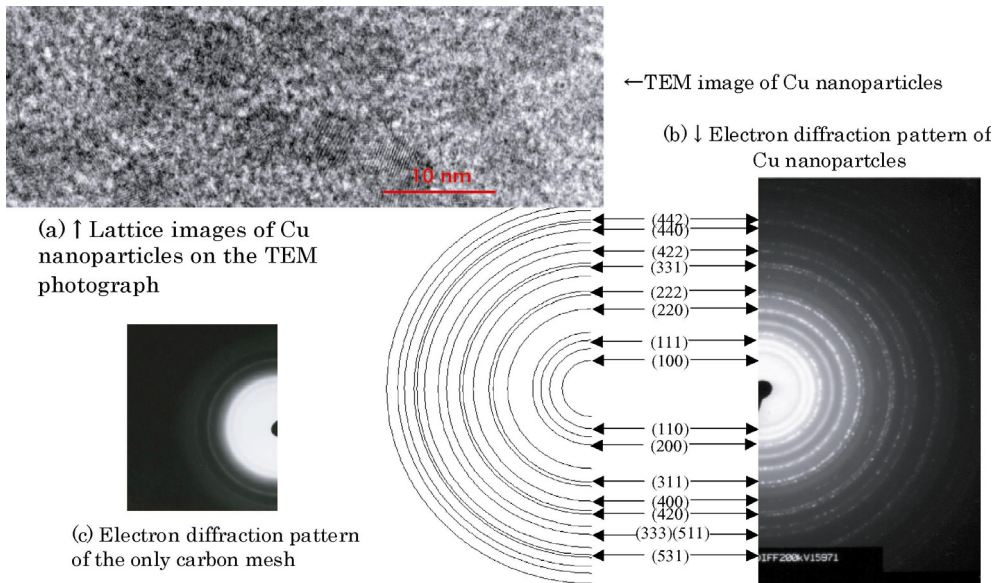


Fig. 11. TEM images and corresponding the electron diffraction patterns of Cu nanoparticles.

### 5.3 Confinement effects

As already stated, in order to produce monodispersed nanoparticles, it is important to confine the plume within the converging shock waves. To confirm the importance of

confining the plume at the focal point of the ellipsoidal cell, further nanoparticle formation experiments were carried out.

Figure 12 is a schematic diagram of the apparatus with an ellipsoidal cell. The laser spot is intentionally shifted by a distance,  $x$ , from the central axis of the ellipsoidal cell, while the target surface is also intentionally inclined by an angle,  $\theta$ , against a plane perpendicular to the central axis. Figure 13 shows some of the results for nanoparticles produced as a result of changing these parameters. The experimental results shown in Figure 13(a), which are obtained under the conditions  $x = 0.0$  mm and  $\theta = 0.0^\circ$ , represent monodispersed nanoparticles. When the target surface has no inclination but the laser spot is shifted  $x = 2$

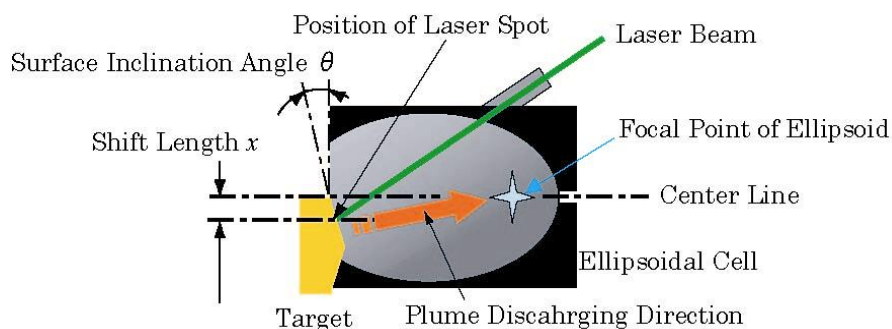


Fig. 12. Schematic of experiment demonstrating the importance of confinement

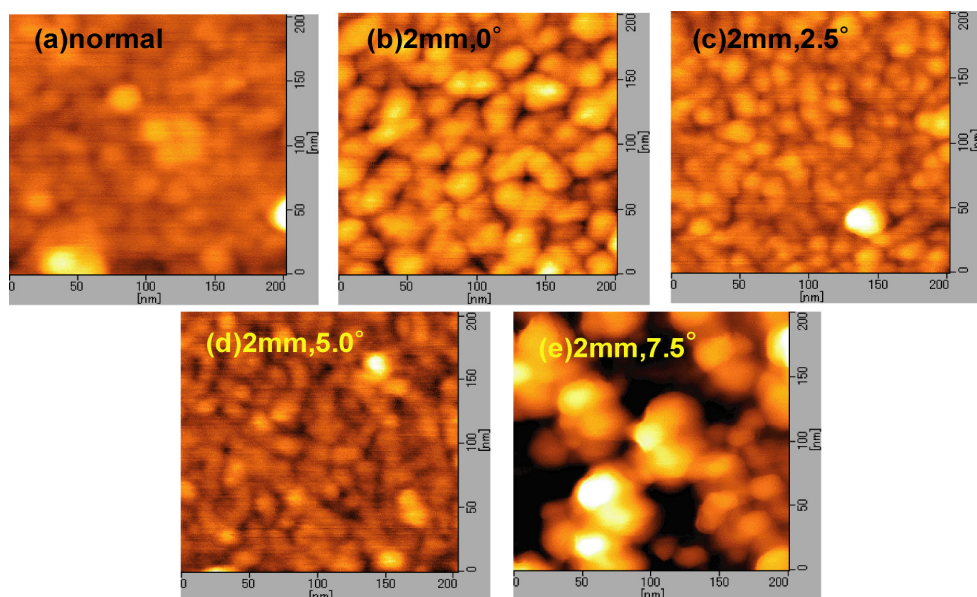


Fig. 13. Influence of shock wave confinement on deposited nanoparticles morphology in the ellipsoidal cell (field of view:200×200nm)



mm, as shown in Figure 13(b), some aggregation is observed. The result in Figure 13(c), where  $x = 2.0$  mm and  $\theta = 2.5^\circ$ , shows the appearance of fine nanoparticles, similar to the normal case (Figure 13(a)). The mainly small and uniformly sized nanoparticles shown in Figure 13(d) formed under conditions of  $x = 2.0$  mm and  $\theta = 5.0^\circ$ . In contrast, when  $x = 2$  mm,  $\theta = 7.5^\circ$ , secondary particles were generated by nanoparticle aggregation (Figure 13(e)). Although the position of the laser spot is shifted and also the density of laser energy is slightly changed (Figures 13(c) and 13(d)) relative to the normal case (Figure 13(a)), the sizes of the resulting nanoparticles were found to be finely dispersed, similar to the normal case. The confinement effect of the plume by the converging shock wave plays a role in these cases, because the plume ejection is approximately directed to the focal point of the ellipsoidal cell. The result of Figure 13(e) indicates that the residence time of nanoparticles in the ellipsoidal cell increased due to circulation by a vortex flow resulting from the shifted direction of the plume ejection relative to the focal point.

#### 5.4 Low temperature sintering

As mentioned above, nanoparticle size was found to be monodispersed in the ellipsoidal cell under appropriate conditions. We will now discuss a case in which the monodispersed nanoparticles were sintered under low-temperature conditions. This low-temperature sintering procedure could serve as a metal bonding technique.

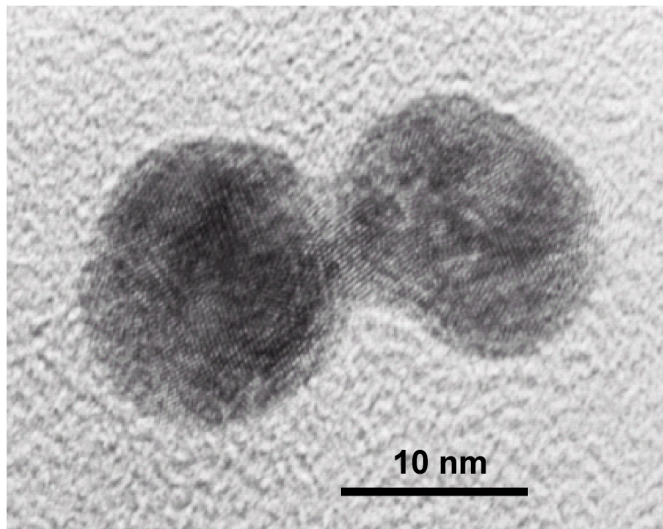


Fig. 14. Two gold nanoparticles forming a neck and binding to each other.

The bonding of metal is an important process for the construction of fine mechanical parts and heat sinks. Conventional bonding methods such as diffusion bonding, melted alloy bonding, hot isostatic pressing and silver brazing cause thermal stress at the interface between two metals because of differences in thermal expansion between the bonded parts. This thermal stress in turn causes warping of the bonded material. Therefore, low-temperature metal bonding is desired to overcome these problems. Since the melting point of metals decreases with decreasing particle size, metal nanoparticle paste has been used as

a low-temperature bonding material. However, the bonding strength of nanoparticle paste is relatively low. Since the sintering of monodispersed nanoparticles has been observed to effectively bond metals, it is important to elucidate this sintering phenomenon in order to optimize the strength of the metal bonding.

The TEM image in Figure 14 shows two gold nanoparticles bonding to each other. In crystallized metallic nanoparticles, bonding between the nanoparticles starts to form even at room temperature if the crystal orientations of the two particles are coincident at the interfaces as shown.

Even if the crystal orientations do not match, it is possible for nanoparticles to bond to each other by using a low-temperature sintering effect which lowers the melting point of the material making up the nanoparticles. In the sintering phenomena of two particles at a certain high temperature, melting, vaporization and diffusion locally occurring in the particle surface result in a fusion at the narrowest neck portion of the contact area between the two particles.

It is well known that the melting point of a substance decreases with decreasing the particle size of materials. The decrement of the melting point,  $\Delta T$ , for a nanoparticle of diameter  $d$  is expressed as follows (Ragone, D. V, 1996):

$$\Delta T = -\frac{4V_s\gamma_{l-s}T_m}{\Delta H_m} \frac{1}{d} \quad (17)$$

where,  $V_s$  is the volume per mole,  $\Delta H_m$  is the melting enthalpy per mole,  $\gamma_{l-s}$  is the interface tension between the liquid and solid phase, and  $\Delta T_m$  is the melting point for the bulk material. If we assume that the material is copper,  $\Delta T$  is about 160 K for a copper nanoparticle having a diameter of 10 nm. We also assume that the interface tension,  $\gamma_{l-s}$  is half the value of bulk surface tension.

The decrease in the melting point results in a decrease in the sintering temperature and strengthens the diffusion bonding at relatively low temperatures. In general, diffusion bonding is enhanced by the sintering process, in which atomic transport occurs between the small bumps on the material surface. By irradiating nanoparticles onto the surface of the materials before bonding, the number of effective small bumps greatly increases.

In some experiments, the aggregation of the nanoparticles was found to be the smallest when the helium background gas pressure was suitable for the dispersion conditions. AFM images of nanoparticles formed under these conditions by the PLA method show that the size of the nanoparticles ranges from 10 nm to several tens of nm. Annealing at comparatively low temperature was performed on nanoparticles formed under these conditions. Figure 15(a) shows an AFM image of nanoparticles before annealing, and Figures 15(b), 15(c), and 15(d) show them after annealing at 473 K, 573 K and 673 K, respectively. As can be seen from the images, nanoparticle size increased with annealing temperature.

According to sintering process theory, the final diameter of a nanoparticle,  $d_f$ , is dependent on the annealing temperature. Particle growth rate can be expressed using the surface area of a nanoparticle by (Koch, W. 1990):

$$\frac{da}{dt} \propto -\frac{1}{\tau}(a - a_f) \quad (18)$$

where  $t$  is the time,  $\tau$  is the characteristic time of particle growth by sintering,  $a$  is the surface area, and  $a_f$  the value of the surface area at a final size. The particle growth rate is dependent on  $\tau$ , which is determined by two main types of the diffusion: lattice diffusion and the grain boundary diffusion. The characteristic time of the lattice diffusion,  $\tau_l$ , is proportionate to the third power of the particle diameter,  $d$ , and temperature,  $T$ , and it is inversely proportional to the surface energy,  $\gamma$ , and the diffusion constant,  $D$ . Therefore,  $\tau_l$  is expressed as (Greer, J. R., 2007)

$$\tau_l \propto \frac{kTd^3}{\gamma D} = \frac{kTd^3}{\gamma D_0} \exp\left(\frac{\varepsilon}{kT}\right) \quad (19)$$

where  $k$  is the Boltzman constant,  $D_0$  is the vibrational constant, and  $\varepsilon$  the activation energy for diffusion. If  $\tau$  used in Eq.(18) is known, the final diameter,  $d_f$ , can be estimated from the correlation between the diameter and annealing time.

As shown in Eq. (19), the characteristic time  $\tau_l$  seems to increase proportionally with temperature, but  $\tau_l$  actually decreases with increasing temperature due to the large contribution of temperature in the exponential term of the equation. However, the characteristic time  $\tau_b$  for grain boundary diffusion is always shorter than  $\tau_l$  under low-temperature conditions. As a result, if  $\tau_b$  is used as the value of  $\tau$  in Eq.(18), the final particle size  $d_f$  can be estimated by measuring the particle sizes at specified time intervals.

Since a large  $\tau$  value corresponds to an unfavorable degree of the sintering, it is necessary to reduce the value of  $\tau$  in order to enhance the sintering process. It can be deduced from Eq. (19) that it is effective to not only increase temperature but also to decrease the diameter of the nanoparticles. From the viewpoint of low-temperature bonding, however, it is preferable to keep the temperature as low as possible and to decrease the size of the nanoparticles before annealing.

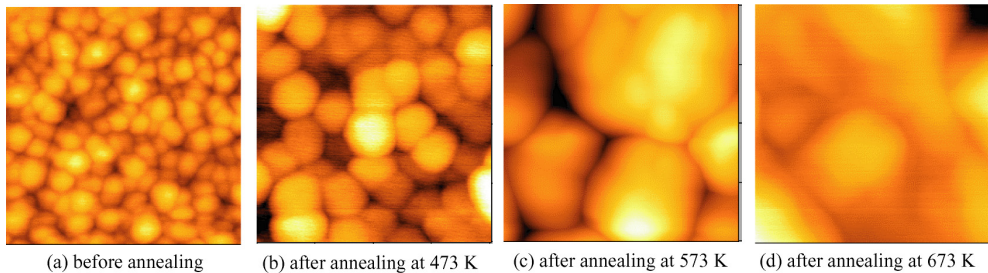


Fig. 15. Nanoparticle sintering at various temperatures (field of view:200×200nm).

## 6. Summary

In this chapter, several topics on the thermodynamics of nanoparticles formation under laser ablation were explored.

Firstly, thermodynamics related to some general aspects of nanoparticle formation in the gas phase and the principles behind of pulsed laser ablation (PLA) was explained. We divided the problem into the following parts for simplicity: (i) nanoparticle nucleation and growth, (ii) melting and evaporation by laser irradiation, and (iii) Knudsen layer formation. All these considerations were then used to build a model of nanoparticle formation into fluid dynamics equations.

Secondly, fluid dynamics concerning nanoparticle formation in a high speed flow was developed. Interactions between the shock waves and plume, generation of nuclei, and growth of nanoparticles could all be treated with a single calculation. We conducted one-dimensional calculations with the equation, and found conditions wherein the timing of the nucleation and growth processes could be separated based on interactions between the shock wave and plume. The existence of certain conditions for nanoparticle formation in the narrow region between the plume and the buffer gas were confirmed from the numerical results. In addition, reflected shock waves substantially contribute to the growth of nanoparticles by increasing particle radius, but do not contribute to the increase of nanoparticle numbers by promoting nucleation.

A new model of nanoparticle generator, employing an ellipsoidal cell, was then formulated based on the results of the one-dimensional calculations. To evaluate the performance of the cell, axi-symmetric two-dimensional calculations were conducted using Navier-Stokes equations without nanoparticle formation. The behavior of shock wave and plume became clear with the use of density contour maps. The reflection and conversion of shock waves, the interaction between shock wave and plume, and ejection of gas through the cell exit were clearly illustrated.

The ellipsoidal cell was manufactured and PLA process was experimentally carried out in the cell. Cu nanoparticles formed in the experiment were typically of uniform size, under 10 nm in diameter, and had a narrow size distribution, with a standard deviation around 1.1 for the lognormal distribution. The narrow distribution of nanoparticle size possibly originated from the effect of ellipsoidal cell, because the fine, uniform nano-sized particles could not be obtained unless the direction of plume ejection was coincident with the focal point of the ellipsoidal cell. Such uniformly sized nanoparticles are important for practical use as indicated by the following example.

Finally, the thermodynamics of nanoparticle sintering was explored, in particular the transition of nanoparticle appearance with changes in temperature, as well as the possibility of low temperature bonding. Since the melting point of nanoparticles sensitively depends on size, it is important to prepare uniformly sized nanoparticles for bonding at low temperatures.

## 7. References

- AIST Home Page, Research Information Database, Network Database System for Thermophysical Property Data, (2006), [http://riodb.ibase.aist.go.jp/TPDB/DBGVsupport/detail/silicon\\_en.html](http://riodb.ibase.aist.go.jp/TPDB/DBGVsupport/detail/silicon_en.html).
- Camata, R. P., Atwater, H. A., Vahala, K. J. and Flagan, R. C. (1996), Size classification of silicon nanocrystals, *Appl. Phys. Lett.* 68 (22), 3162-3164.
- Chrissey, D.B. and Hubler G.K. (Eds.) (1994), *Pulsed Laser Deposition of Thin Films*, Wiley-Interscience, New York.
- Finney, E. E. and Finke, R. G. (2008), Nanocluster nucleation and growth kinetic and mechanistic studies: A review emphasizing transition-metal nanoclusters, *Journal of Colloid and Interface Science* 317, 351-374.
- Fukuoka, H., Yaga, M. and Takiya, T. (2008), Study of Interaction between Unsteady Supersonic Jet and Shock Waves in Elliptical Cell, *Journal of Fluid Science and Technology*, 3-7, 881-891.

- Greer, J. R. and Street, R. A. (2007), Thermal cure effects on electrical performance of nanoparticle silver inks, *Acta Mater.* 55, 6345-6349.
- Han, M., Gong, Y. Zhou, J. Yin, C. Song, F. Muto, M. Takiya T. and Iwata, Y. (2002), Plume dynamics during film and nanoparticles deposition by pulsed laser ablation, *Phys. Lett.*, A302, 182-189.
- Houle F. A. and Hinsberg, W. D. (1998), Stochastic simulation of heat flow with application to laser-solid interactions, *Appl. Phys.*, A66, 143-151.
- Ide, E., Angata, S., Hirose, A. and Kobayashi, K. (2005), Metal-metal bonding process using Ag metallo-organic nanoparticles, *Acta Materialia* 53, 2385-2393.
- Inada, M., Nakagawa, H., Umezu, I. and Sugimura, A. (2003), Effects of hydrogenation on photoluminescence of Si nanoparticles formed by pulsed laser ablation, *Materials Science and Engineering B101*, 283-285.
- Ito, S., Nakaoka, K., Kawamura, M., Ui, K., Fujimoto, K. and Koura, N. (2005), Lithium battery having a large capacity using  $\text{Fe}_3\text{O}_4$  as a cathode material, *Journal of Power Sources* 146, 319-322.
- Iwata, Y. Kishida, M. Muto, M. Yu, S. Sawada, T. Fukuda, A. Takiya, T. Komura A. and Nakajima, K. (2002), Narrow size-distributed silicon cluster beam generated using a spatiotemporal confined cluster source, *Chem. Phys. Lett.*, 358, 36-42.
- Kim, K., Park, J. H., Doo, S. G., Nam, J. D. and Kim, T. (2009), Generation of size and structure controlled Si nanoparticles using pulse plasma for energy devices, *Thin Solid Films* 517, 4184-4187.
- Kim, K., Park, J. H., Doo, S. G. and Kim, T. (2010), Effect of oxidation on Li-ion secondary battery with non-stoichiometric silicon oxide ( $\text{SiO}_x$ ) nanoparticles generated in cold plasma, *Thin Solid Films* 518, 6547-6549.
- Knight, C. J. (1979), Theoretical Modeling of Rapid Surface Vaporization with Back Pressure, *AIAA J.*, 17-5, 519-523.
- Koch W. and Friedlander S. K. (1990) · The effect of particle coalescence on the surface area of a coagulation aerosol, *Journal of Colloid and Interface Science*, 140-2, 419.
- Li, S. and El-Shall, M. S. (1998), Synthesis of nanoparticles by reactive laser vaporization: silicon nanocrystals in polymers and properties of gallium and tungsten oxides, *Applied Surface Science* 127-129, 330-338.
- Li, Q., Sasaki, T., Koshizaki, N. (1999), Pressure dependence of the morphology and size of cobalt (II,III) oxide nanoparticles prepared by pulsed-laser ablation, *Appl. Phys. A* 69, 115-118.
- Liqiang, J., Xiaojun, S., Jing, S., Weimin, C., Zili, X., Yaoguo, D. and Honggang, F. (2003), Review of surface photovoltage spectra of nanosized semiconductor and its applications in heterogeneous photocatalysis, *Solar Energy Materials & Solar Cells* 79, 133-151.
- Liqiang, J., Baiqi, W., Baifu, X., Shudan, L., Keying, S., Weimin, C. and Honggang, F. (2004), Investigations on the surface modification of ZnO nanoparticle photocatalyst by depositing Pd, *Journal of Solid State Chemistry* 177, 4221-4227.
- Lu, M., Gong, H., Song, T., Wang, J. P., Zhang, H. W. and Zhou, T. J. (2006), Nanoparticle composites: FePt with wide-band-gap semiconductor, *Journal of Magnetism and Magnetic Materials* 303, 323-328.

- Patrone, L., Nelson, D., Safarov, V.I., Giorgio, S., Sentis, M. and Marine, W. (1999), Synthesis and properties of Si and Ge nanoclusters produced by pulsed laser ablation, *Appl. Phys. A* 69 [Suppl.], S217-S221.
- Patrone, L., Nelson, D., Safarov, V. I., Sentis, M. and Marine, W. (2000), Photoluminescence of silicon nanoclusters with reduced size dispersion produced by laser ablation, *Journal of Applied Physics* Vol.87, No.8, 3829-3837.
- Ragone · D. V. (1996), *Chemical physics of materials II*, Maruzen, (Translated into Japanese).
- Roco, M. C. (1998), Reviews of national research programs in nanoparticle and nanotechnology research in the U.S.A., *J. Aerosol Sci.* Vol. 29, No. 5/6, pp. 749-760.
- Seto, T., Koga, K., Takano, F., Akinaga, H., Orii, T., Hirasawa, M. and Murayama, M. (2006), Synthesis of magnetic CoPt/SiO<sub>2</sub> nano-composite by pulsed laser ablation, *Journal of Photochemistry and Photobiology A: Chemistry* 182, 342-345.
- Shapiro, A. H. (1953) *The Dynamics and Thermodynamics of COMPRESSIBLE FLUID FLOW*, Ronald Press, New York.
- Strickland, N. M., Long, N. J., Talantsev, E. F., Hoefakker, P., Xia, J. A., Rupich, M. W., Zhang, W., Li, X., Kodendath, T. and Huang, Y. (2008), Nanoparticle additions for enhanced flux pinning in YBCO HTS films, *Current Applied Physics* 8, 372-375.
- Suzuki, N., Makino, T., Yamada, Y. and Yoshida, T. (2001), Monodispersed, nonagglomerated silicon nanocrystallites, *Applied Physics Letters*, Vol.78, No.14 , 2043-2045.
- Takiya, T., Umezu, I., Yaga, M. and Han, M. (2007), Nanoparticle Formation in the Expansion Process of a Laser Ablated Plume, *J. Phys. Conf. Ser.* 59, 445.
- Takiya, T., Fukuda, N., Inoue, N., Han, M., Yaga, M. and Iwata, Y. (2010), Dynamics of the Shock Wave Accompanied by Nanoparticle Formation in the PLA Processes, *Adv. Studies Theor. Phys.*, Vol. 4, no.7, 305 - 316.
- Touloukian, Y. S., editor. (1967), *Thermophysical Properties of High Temperature Solid Materials*, The Macmillan Co., New York.
- Volmer, M. (1939), *Kinetik der Phasenbildung*, T. Steinkopff, Dresden, Leipzig.
- Weast, R. C., editor. (1965), *Handbook of Chemistry and Physics* 46th edition, The Chemical Rubber Co., Ohio.
- Wegner, K., Piseri, P., Tafreshi H. V. and Milani, P. (2006), Cluster beam deposition: a tool for nanoscale science and technology, *J. Phys. D: Appl. Phys.* 39, R439-R459.
- Wu, H. P., Okano, A. and Takayanagi, K. (2000), Photoluminescence properties of size-selected Si nanocluster films prepared by laser ablation, *Appl. Phys. A* 71, 643-646.
- Yaga, M., Takiya T. and Iwata, Y. (2005), Numerical study of unsteady compressible flow driven by supersonic jet injected into ellipsoidal cell with small exit hole, *Shock waves*, 14-5/6, 403-411.
- Yaga, M., Fukuoka, H., Iwata, Y. and Takiya, T. (2008), Behavior of Shock Waves Formed by Unsteady Supersonic Jet Injected into Cell, *Journal of Thermal Science*, 17-1, pp.50-55.

# Thermodynamics of the Oceanic General Circulation – Is the Abyssal Circulation a Heat Engine or a Mechanical Pump?

Shinya Shimokawa<sup>1</sup> and Hisashi Ozawa<sup>2</sup>

*<sup>1</sup>National Research Institute for Earth  
Science and Disaster Prevention*

*<sup>2</sup>Hiroshima University  
Japan*

## 1. Introduction

The oceanic general circulation has been investigated mainly from a dynamic perspective. Nevertheless, some important contributions to the field have been made also from a thermodynamic viewpoint. This chapter presents description of the thermodynamics of the oceanic general circulation. Particularly, we examine entropy production of the oceanic general circulation and discuss its relation to a thermodynamic postulate of a steady closed circulation such as the oceanic general circulation: Sandström's theorem. Also in this section, we refer to another important thermodynamic postulate of an open non-equilibrium system such as the oceanic general circulation: the principle of Maximum Entropy Production.

### 1.1 Outline of oceanic general circulation

Oceanic general circulation is the largest current in the world ocean, making a circuit from the surface to the bottom over a few thousand years. The present oceanic general circulation, briefly speaking, is a series of flows, in which seawater sinks from restricted surface regions in high latitudes of the Atlantic Ocean to the deep bottom ocean. It later comes to broad surface regions of the Pacific Ocean, and returns to the Atlantic Ocean through the surface of the Indian Ocean (see Fig. 1). The atmosphere affects the daily weather, whereas the ocean affects the long-term climate because of its larger heat capacity. Therefore, it is important for our life to elucidate the oceanic general circulation.

The causes generating the oceanic general circulation are momentum flux by wind stress at the sea surface and density flux by heating, cooling, precipitation, and evaporation through the sea surface, except for tides. In general, the oceanic general circulation is explained as consisting of surface (wind-driven) circulation attributable to the momentum flux and abyssal (thermohaline) circulation caused by the density flux. However, the distinction between them is not simple because diapycnal mixing, which is important for abyssal circulation, depends largely on wind, as described in the next sub-section. Moreover, diapycnal mixing depends also on tides.

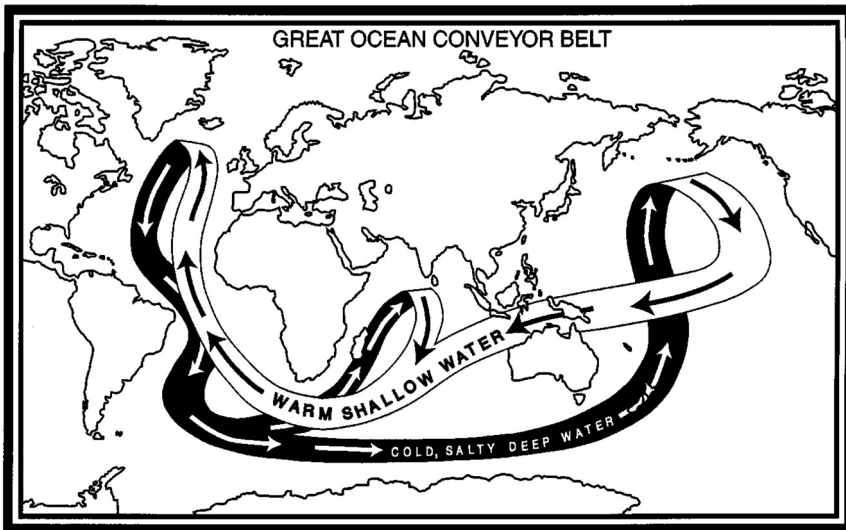


Fig. 1. Illustration of oceanic general circulation (Broecker, 1987).

### 1.2 Energy sources of abyssal circulation

Sustained abyssal circulation is a manifestation of conversion of potential energy to kinetic energy within the system. Production of potential energy is mainly the result of diapycnal mixing in the ocean interior, geothermal heating through the ocean floor, and the meridional distribution of precipitation, evaporation, and runoff (e.g., Gade & Gustafsson, 2004).

Diapycnal mixing results from turbulent diffusion by wind and tides. The most reasonable mechanism to transfer energy from the surface to the deeper layer is regarded as breaking and wave-wave interaction of internal waves generated by wind and tides (e.g., Muller & Briscoe, 2000). The wind and tidal dissipation quantities have been estimated respectively as about 1 TW (Wunsch, 1998) and 1 TW (Egbert & Ray, 2000). Using these estimates and  $R_f = 0.15$  (Osborn, 1980) as the flux Richardson number,  $\gamma = R_f / (1 - R_f) = 0.18$  as the ratio of potential energy to available energy, and  $S = 3.6 \times 10^{14} \text{ m}^2$  as the total surface area of the ocean, the production of potential energy caused by diapycnal mixing has been estimated as about  $1.0 \times 10^{-3} \text{ W m}^{-2}$  ( $= 2\text{TW} / (3.6 \times 10^{14} \text{ m}^2) \times 0.18$ ).

Geothermal heating through the ocean floor causes a temperature increase and a thermal expansion in seawater, and generates potential energy. Production of potential energy caused by geothermal heating has been estimated as about 0.11 (Gade & Gustafsson, 2004) - 0.14 (Huang, 1999)  $\times 10^{-3} \text{ W m}^{-2}$ .

Precipitation (evaporation) is a flux of mass to (from) the sea surface and consequently a flux of potential energy. On average, the warm (cold) tropics with high (low) sea level are regions of evaporation (precipitation). These therefore tend to reduce the potential energy. The value integrated for the entire ocean shows a net loss of potential energy. Loss of potential energy attributable to precipitation, evaporation, and runoff has been estimated as less than 0.02 (Gade & Gustafsson, 2004) - 0.03 (Huang, 1998)  $\times 10^{-3} \text{ W m}^{-2}$ . These contributions can be negligible.



In addition, there can be work done on the ocean by surface heating and cooling. Heating (cooling) causes an expansion (contraction) with a net rise (fall) in the centre of mass and an increase (decrease) in potential energy. The exact estimate of the effect is difficult, but it will be small compared to the effect of the wind forcing. The best recent estimate of work done on the ocean by surface heating and cooling is zero (Wunsch & Ferrari, 2004).

### 1.3 "Missing mixing" problem

Munk (1966) estimated that the magnitude of diapycnal mixing to drive and maintain abyssal circulation is about  $K \approx 10^{-4} \text{ m}^2 \text{ s}^{-1}$ . He reached that figure by fitting of vertical profiles of tracers with one-dimensional vertical balance equation of advection and diffusion as

$$K \frac{d^2 T}{dz^2} = w \frac{dT}{dz}, \quad (1)$$

where  $K$  is a diapycnal mixing coefficient,  $T$  denotes a tracer variable such as temperature, salinity and radioactive tracers,  $z$  signifies a vertical coordinate, and  $w$  represents the upwelling velocity. The estimated value has been regarded as reasonable because the total upwelling of deep water estimated using the above  $K$  is consistent with the total sinking of deep water estimated by observations in the sinking area.

However, some direct observations of turbulence (Gregg, 1989) and dye diffusion (Ledwell et al., 1993) in the deep ocean indicate a diapycnal mixing of only  $K \approx 10^{-5} \text{ m}^2 \text{ s}^{-1}$ . Moreover, this is consistent with mixing estimated from the energy cascade in an internal wave spectrum (called "background") (McComas & Mullar, 1981). This difference of  $K$  is designated as the "missing mixing" problem.

On the other hand, recent observations of turbulence show larger diapycnal mixing of  $K \geq 10^{-4} \text{ m}^2 \text{ s}^{-1}$  (Ledwell et al., 2000; Polizin et al., 1997), although such observations are limited to areas near places with large topographic changes such as seamounts (called "hot spots"), where internal waves are strongly generated as sources of diapycnal mixing. Munk & Wunsch (1998) reported that the value averaged over the entire ocean including "background" and "hot spots" can be about  $K \approx 10^{-4} \text{ m}^2 \text{ s}^{-1}$ , which remains controversial.

### 1.4 Abyssal circulation as a heat engine or a mechanical pump

Traditionally, the abyssal circulation has been treated as a heat engine (or a buoyancy process) driven by an equatorial hot source and polar cold sources. Broecker & Denton (1990) reported that abrupt changes in the ocean's overturning causes the ocean's heat loss, which might engender large swings in high-latitude climate, such as that occurring during the ice age. They also suggested a descriptive image of abyssal circulation: a conveyor-belt (see Fig. 1). Peixoto & Oort (1992) investigated the atmosphere-ocean system as a heat engine using the concept of available potential energy developed by Lorenz (1955).

Toggweiler (1994) reported that the abyssal formation in the North Atlantic is induced by upwelling because of strong surface wind stress in the Antarctic circumpolar current (a mechanical pump or a mechanical process). This mechanism is inferred from the "missing mixing" problem, as stated in section 1.3. If "background" diapycnal mixing for maintaining abyssal circulation is weaker than Munk's estimate, then another new mechanism to pump

up water from the deep layer to the surface is needed, provided that sinking can occur in the cold saline (i.e. dense) region of the North Atlantic. Drake Passage is located in the region of westerly wind band where water upwells from below to feed the diverging surface flow. Because net poleward flow above the ridges is prohibited (there is no east-west side wall to sustain an east-west pressure gradient in the Antarctic circumpolar current region), the upwelled water must come from below the ridges, i.e., from depths below 1500–2000 m. In addition, very little mixing energy is necessary to upwell water because of weak stratification near Antarctica.

### 1.5 Sandström theorem

Related to a closed steady circulation such as abyssal circulation, there is an important thermodynamic postulate: Sandström's theorem (Sandström, 1908, 1916)<sup>1</sup>.

Sandström considered the system moving as a cycle of the heat engine with the following four stages (see Fig. 2).

1. Expansion by diabatic heating under constant pressure
2. Adiabatic change (expansion or contraction) from the heating source to the cooling source
3. Contraction by diabatic cooling under constant pressure
4. Adiabatic change (contraction or expansion) from the cooling source to the heating source

When the system moves anti-clockwise (expansion in stage 2 and contraction in stage 4), i.e., the heating source ( $d\alpha > 0$ ;  $\alpha$  is a specific volume that is equal to the volume divided by the mass) is located at the high-pressure side and the cooling source ( $d\alpha < 0$ ) is located at the low-pressure side (Fig. 2a;  $P_{\text{heating}} > P_{\text{cooling}}$ ), the work done by the system is positive:

$$\oint P d\alpha > 0. \quad (2)$$

In contrast, when the system moves clockwise (contraction in stage 2 and expansion in stage 4), i.e., the cooling source is located at the high-pressure side and the heating source is located at the low-pressure side (Fig. 2b;  $P_{\text{heating}} < P_{\text{cooling}}$ ). Therefore, the work done by the system is negative:

$$\oint P d\alpha < 0. \quad (3)$$

Consequently, Sandström suggested that a closed steady circulation can only be maintained in the ocean if the heating source is located at a higher pressure (i.e. a lower level) than the cooling source.

Regarding the atmosphere, the heating source is located at the ground surface and the cooling source is located at the upper levels because the atmosphere is almost transparent to shortwave radiation of the sun, which heats the ground surface directly. Then heat is transferred from the heated surface by vertical convection. Therefore, the atmosphere can be regarded as a heat engine.

---

<sup>1</sup> An English translation of Sandström (1906) is available as an appendix in Kuhlbrodt (2008), but the Sandström papers are written in German, and are not easy to obtain. Other explanations of Sandström's theorem can be found in some textbooks of oceanic and atmospheric sciences: Defunt (1961), Houghton (2002), and Huang (2010).

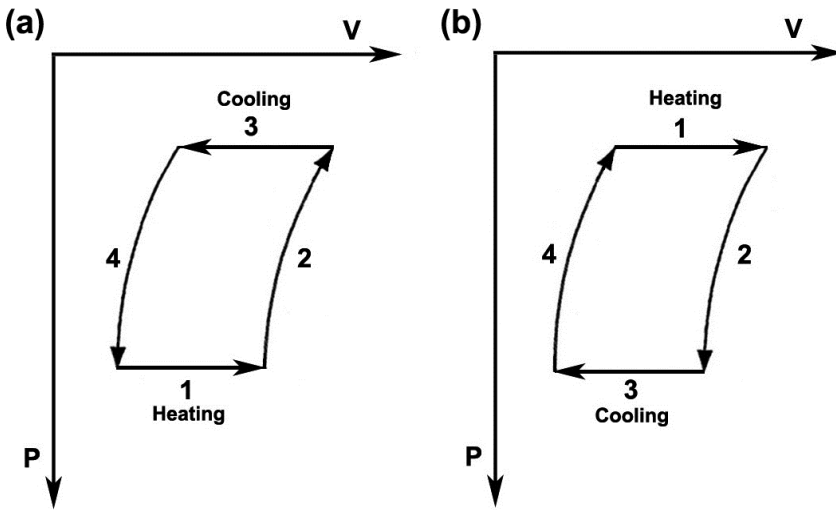


Fig. 2. Heat engines of two types discussed by Sandström (1916): (a) anti-clockwise and (b) clockwise.

### 1.6 Principle of maximum entropy production and oceanic general circulation

In this sub-section, we briefly explain another important thermodynamic postulate of stability of a nonlinear non-equilibrium system such as the oceanic general circulation, the principle of the maximum Entropy Production and consider the stability of oceanic general circulation from a global perspective because local processes of generation and dissipation of kinetic energy in a turbulent medium remain unknown.

The ocean system can be regarded as an open non-equilibrium system connected with surrounding systems mainly via heat and salt fluxes. The surrounding systems consist of the atmosphere, the Sun and space. Because of the curvature of the Earth's surface and the inclination of its rotation axis relative to the Sun, net gains of heat and salt are found in the equatorial region; net losses of heat and salt are apparent in polar regions. The heat and salt fluxes bring about an inhomogeneous distribution of temperature and salinity in the ocean system. This inhomogeneity produces the circulation, which in turn reduces the inhomogeneity. In this respect, the formation of the circulation can be regarded as a process leading to final equilibrium of the whole system: the ocean system and its surroundings. In this process, the rate of approach to equilibrium, i.e., the rate of entropy production by the oceanic circulation, is an important factor.

Related to the rate of entropy production in an open non-equilibrium system, Sawada (1981) reported that such a system tends to follow a path of evolution with a maximum rate of entropy production among manifold dynamically possible paths. This postulate has been called the principle of Maximum Entropy Production (MEP), which has been confirmed as valid for mean states of various nonlinear fluid systems, e.g., the global climate system of the Earth (Ozawa & Ohmura, 1997; Paltridge, 1975, 1978), those of other planets (Lorenz et al., 2001), the oceanic general circulation including both surface and abyssal circulations (Shimokawa, 2002; Shimokawa & Ozawa, 2001, 2002, 2007), and thermal convection and shear turbulence (Ozawa et al., 2001). Therefore, it would seem that MEP can stand for a

universal principle for time evolution of non-equilibrium systems (see reviews of Kleidon and Lorenz, 2005; Lorenz, 2003; Martyushev & Seleznev, 2006; Ozawa et al., 2003; Whitfield, 2005). However, although some attempts have been made to seek a theoretical framework of MEP (e.g., Dewar, 2003, 2005), we remain uncertain about its physical meaning.

### 1.7 Main contents of this chapter

As described above, the problem of whether the abyssal circulation is a heat engine or mechanical pump and how it is related to the Sandström theorem are important for better understanding of the oceanic general circulation. In the following sections, we discuss the problem referring to the results of numerical simulations of the oceanic general circulation. In section 2, a numerical model and method are described. In section 3, a calculation method of entropy production rate in the model is explained. In section 4, details of entropy production in the model are described. In section 5, by referring to the results, the problem of whether the abyssal circulation is a heat engine or mechanical pump and how it is related to the Sandström theorem is discussed.

## 2. Numerical model and method

The numerical model used for this study is the Geophysical Fluid Dynamics Laboratory's Modular Ocean Model (Pacanowski, 1996). The model equations consist of Navier-Stokes equations subject to the Boussinesq, hydrostatic, and rigid-lid approximations along with a nonlinear equation of state that couples two active variables, temperature and salinity, to the fluid velocity. A convective adjustment scheme is used to represent the vertical mixing process. Horizontal and vertical diffusivity coefficients are, respectively,  $10^3 \text{ m}^2 \text{ s}^{-1}$  and  $10^{-4} \text{ m}^2 \text{ s}^{-1}$ . The time-step of the integration is 5400 s.

The model domain is a rectangular basin of  $72^\circ$  longitude by  $140^\circ$  latitude with a cyclic path, representing an idealized Atlantic Ocean (Fig. 3(a)). The southern hemisphere includes an Antarctic Circumpolar Current passage from  $48^\circ\text{S}$  to  $68^\circ\text{S}$ . The horizontal grid spacing is 4 degrees. The ocean depth is 4500 m with 12 vertical levels (Shimokawa & Ozawa, 2001). All boundary conditions for wind stress, temperature and salinity are arranged as symmetric about the equator (Figs. 3(b), 3(c), and 3(d)). The wind stress is assumed to be zonal (eastward or westward direction, Fig. 3(b)). A restoring boundary condition is applied: The surface temperature and salinity are relaxed to their prescribed values (Figs. 3(c) and 3(d)), with a relaxation time scale of 20 days over a mixed layer depth of 25 m. The corresponding fluxes of heat and salt are used to calculate  $F_h$  and  $F_s$  at the surface. The initial temperature distribution is described as a function of depth and latitude. The initial salinity is assumed to be constant (34.9‰). The initial velocity field is set to zero. Numerical simulation is conducted for a spin-up period of 5000 years.

Figure 4 shows a zonally integrated meridional stream function at years 100, 1000, 2000, 3000, 4000, and 5000, after starting the calculations. At year 100, the circulation pattern is almost symmetric about the equator. The sinking cell in the southern hemisphere does not develop further because of the existence of the Antarctic Circumpolar Current. In contrast, the sinking cell in the northern hemisphere develops into deeper layers, and the circulation pattern becomes asymmetric about the equator. The oceanic circulation becomes statistically steady after year 4000. Temperature variations are shown to be less than 0.1 K after year 4000. In the steady state, the northern deep-water sinking cell is accompanied by an Antarctic bottom-water sinking cell and by a northern intrusion cell from the south. The flow pattern is apparently a basic one in the idealised Atlantic Ocean.

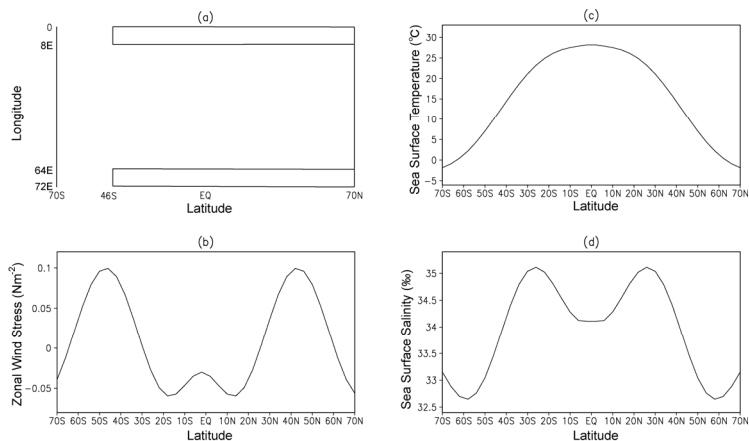


Fig. 3. (a) Model domain, and forcing fields of the model as functions of latitude, (b) forced zonal wind stress ( $\text{N m}^{-2}$ ) defined as positive eastward, (c) prescribed sea surface temperature ( $^{\circ}\text{C}$ ), and (d) prescribed sea surface salinity ( $\text{‰}$ ).

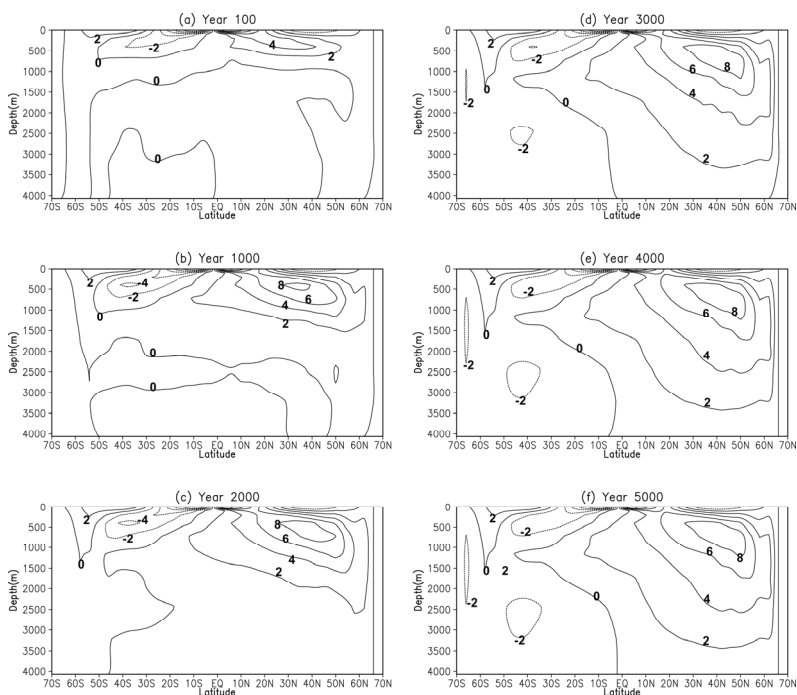


Fig. 4. The zonally integrated meridional stream function at years (a) 100, (b) 1000, (d) 2000, (e) 3000, (d) 4000, and (e) 5000 after starting the numerical calculations. The contour line interval is  $2 \text{ SV}$  ( $10^6 \text{ m}^3 \text{ s}^{-1}$ ). The circulation pattern reached a statistically steady-state after year 4000.

### 3. Entropy production rate calculation

According to Shimokawa & Ozawa (2001) and Shimokawa (2002), the entropy increase rate for the ocean system is calculable as

$$\begin{aligned} \frac{dS}{dt} = & \int \frac{1}{T} \left[ \frac{\partial(\rho c T)}{\partial t} + \text{div}(\rho c T v) + p \text{div}(v) \right] dV + \int \frac{F_h}{T} dA \\ & - ak \int \left[ \frac{\partial C}{\partial t} + \text{div}(C v) \right] \ln C dV - ak \int F_s \ln C dA \end{aligned} \quad (4)$$

where  $\rho$  stands for the density,  $c$  denotes the specific heat at constant volume,  $T$  signifies the temperature,  $a = 2$  is van't Hoff's factor representing the dissociation effect of salt into separate ions ( $\text{Na}^+$  and  $\text{Cl}^-$ ),  $k$  is the Boltzmann's constant,  $C$  is the number concentration of salt per unit volume of seawater,  $F_h$  and  $F_s$  are the heat and salt fluxes per unit surface area respectively, defined as positive outward, and  $dV$  and  $dA$  are the small volume and surface elements, respectively.

If we can assume that the seawater is incompressible ( $\text{div } v = 0$ ) and that the volumetric heat capacity is constant ( $\rho c = \text{const.}$ ), then the divergence terms in (4) disappear. In this case, we obtain

$$\frac{dS}{dt} = \int \frac{\rho c}{T} \frac{\partial T}{\partial t} dV + \int \frac{F_h}{T} dA - ak \int \frac{\partial C}{\partial t} \ln C dV - ak \int F_s \ln C dA \quad (5)$$

The first two terms in the right-hand side represent the entropy production rate attributable to heat transport in the ocean. The next two terms represent that attributable to the salt transport. The first and third terms vanish when the system is in a steady state because the temperature and the salinity are virtually constant ( $\partial T / \partial t = \partial C / \partial t = 0$ ). In the steady state, entropy produced by the irreversible transports of heat and salt is discharged completely into the surrounding system through the boundary fluxes of heat and salt, as expressed by the second and fourth terms in equation (5).

The general expression (4) can be rewritten in a different form. A mathematical transformation (Shimokawa and Ozawa, 2001) can show that

$$\frac{dS}{dt} = \int F_h \cdot \text{grad} \left( \frac{1}{T} \right) dV + \int \frac{\Phi}{T} dV - ak \int \frac{F_s \cdot \text{grad}(C)}{C} dV, \quad (6)$$

where  $F_h$  and  $F_s$  respectively represent the flux densities of heat and salt (vector in three-dimensional space) and  $\Phi$  is the dissipation function, representing the rate of dissipation of kinetic energy into heat by viscosity per unit volume of the fluid. The first term on the right-hand side is the entropy production rate by thermal dissipation (heat conduction). The second term is that by viscous dissipation; the third term is that by molecular diffusion of salt ions. Empirically, heat is known to flow from hot to cold via thermal conduction, and the dissipation function is always non-negative ( $\Phi \geq 0$ ) because the kinetic energy is always dissipated into heat by viscosity. Molecular diffusion is also known to take place from high to low concentration (salinity). Therefore, the sum should also be positive. This is a consequence of the Second Law of Thermodynamics.

#### 4. Results – details of entropy production in the model

We describe here the details of entropy production in the model from the final state of the spin-up experiment (Fig. 4(f)). Because entropy production due to the salt transport is negligible (Shimokawa and Ozawa, 2001), local entropy production can be estimated from the first term in equation (6) as

$$A = \frac{\rho C}{T^2} (A_x + A_y + A_z), A_x = D_h \left( \frac{dT}{dx} \right)^2, A_y = D_h \left( \frac{dT}{dy} \right)^2, A_z = D_v \left( \frac{dT}{dz} \right)^2, \quad (7)$$

where  $D_h$  denotes horizontal diffusivity of  $10^3 \text{ m}^2 \text{ s}^{-1}$ ,  $D_v$  stands for vertical diffusivity of  $10^{-4} \text{ m}^2 \text{ s}^{-1}$  (see section 2), and other notation is the same as that used earlier in the text. It is assumed here that  $F_h = -k \text{ grad}(T) = -\rho c D_E \text{ grad}(T)$ , where  $k = \rho c D_E$  signifies thermal conductivity and where  $D_E$  represents the eddy diffusivity ( $D_h$  or  $D_v$ ). Figure 5 shows zonal, depth and zonal-depth averages of each term in equation (7). The quantities not multiplied by  $dV$  represent the values at the site, and the quantities multiplied by  $dV$  represent the values including the effect of layer thickness.

It is apparent from the zonal average of  $A$  (Fig. 5(a)) that entropy production is large in shallow-intermediate layers at low latitudes. This is apparent also in the zonal-depth average of  $A \times dV$  (Fig. 5(c)). However, it is apparent from the depth average of  $A \times dV$  (Fig. 5(b)) that entropy production is large at the western boundaries at mid-latitudes and at low latitudes. Consequently, entropy production is greatest at the western boundaries at mid-latitudes as the depth average, but it is highest at low latitudes as the depth-zonal average. It is apparent as the figures show of  $A_x$ ,  $A_y$  and  $A_z$  (Figs. 5(d), (g) and (j)) that  $A_x$  is large in shallow layers at mid-latitudes,  $A_y$  is large in shallow-intermediate layers at high latitudes, and that  $A_z$  is large in shallow-intermediate layers at low latitudes. It is also apparent that as the figures show of  $A_x \times dV$ ,  $A_y \times dV$  and  $A_z \times dV$  (Figs. 5(e), 5(f), 5(h), 5(i), 5(k) and 5(l)) that  $A_x \times dV$  is large at the western boundaries at mid-latitudes,  $A_y \times dV$  is large at high latitudes, and  $A_z \times dV$  is large at low latitudes. Additionally, it is apparent that the values of  $A_z$  ( $A_z \times dV$ ) is the largest, and those of  $A_x$  ( $A_x \times dV$ ) are smaller than those of  $A_y$  ( $A_y \times dV$ ) and  $A_z$  ( $A_z \times dV$ ).

Consequently, there are three regions with large entropy production: shallow-intermediate layers at low latitudes, shallow layers at the western boundaries at mid-latitudes, and shallow-intermediate layers at high latitudes. It can be assumed that the contribution of shallow-intermediate layers at low latitudes results from the equatorial current system. That of western boundaries at mid-latitudes results from the western boundary currents such as Kuroshio, and that of intermediate layers at high latitudes results from the meridional circulation of the global ocean. It is apparent that high dissipation regions at low latitudes expand into the intermediate layer in the zonal averages of  $A \times dV$  and  $A_z \times dV$ . These features appear to indicate that equatorial undercurrents and intermediate currents in the equatorial current system are very deep and strong currents which can not be seen at other latitudes (Colling, 2001). It is also apparent that high dissipation regions at high latitudes in the northern hemisphere intrude into the intermediate layer in the zonal averages of  $A \times dV$  and  $A_y \times dV$ , and the peak of northern hemisphere is larger than that of southern hemisphere in the zonal-depth averages of  $A$  and  $A_y$ . These features appear to represent the characteristics of the circulation with northern sinking (Fig. 4(f)).

Strictly speaking, we should consider dissipation in a mixed layer and dissipation by convective adjustment for entropy production in the model. Dissipation in a mixed layer can be estimated from the first term in (6) as

$$B = \frac{\rho C (T_r - T_s)}{T^2 \Delta t_r}, \quad (8)$$

where  $T_r$  signifies restoring temperature (Fig. 3(c)),  $T_s$  is the sea surface temperature in the model, and  $\Delta t_r$  stands for the relaxation time of 20 days (see section 2). It is assumed here that  $F_h = -k \text{grad}(T) = -\rho c D_M \text{grad}(T)$ , where  $k = \rho c D_M$  is thermal conductivity,  $D_M = \Delta z_r^2 / \Delta t_r$  represents diffusivity in the mixed layer, and  $\Delta z_r$  is the mixed layer thickness of 25 m (see section 2). The estimated value of  $B$  is lower than that of  $A$  by three or four orders: it is negligible. Dissipation by convective adjustment can be estimated from the first term in (5) such that

$$C = \frac{\rho C (T_b - T_a)}{T_b \Delta t}, \quad (9)$$

where  $T_b$  is the temperature before convective adjustment,  $T_a$  is the temperature after convective adjustment, and  $\Delta t$  is the time step of 5400 s (see section 2). In fact,  $T_b$  is identical to  $T_a$  at the site where convective adjustment has not occurred. The value of  $C$  is negligible because the effect of convective adjustment is small in the steady state.

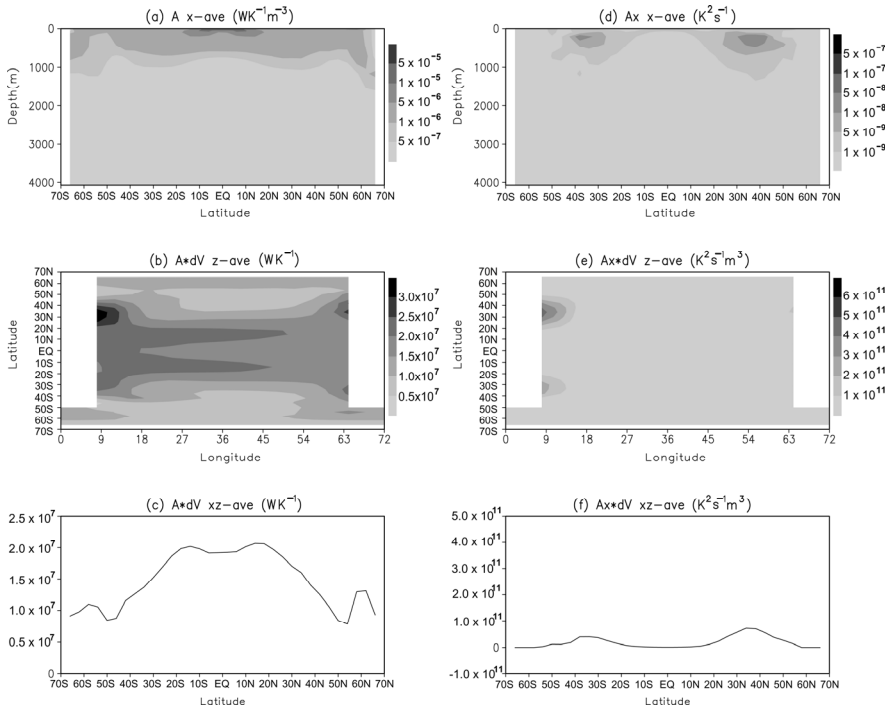


Fig. 5. Entropy production in the model.



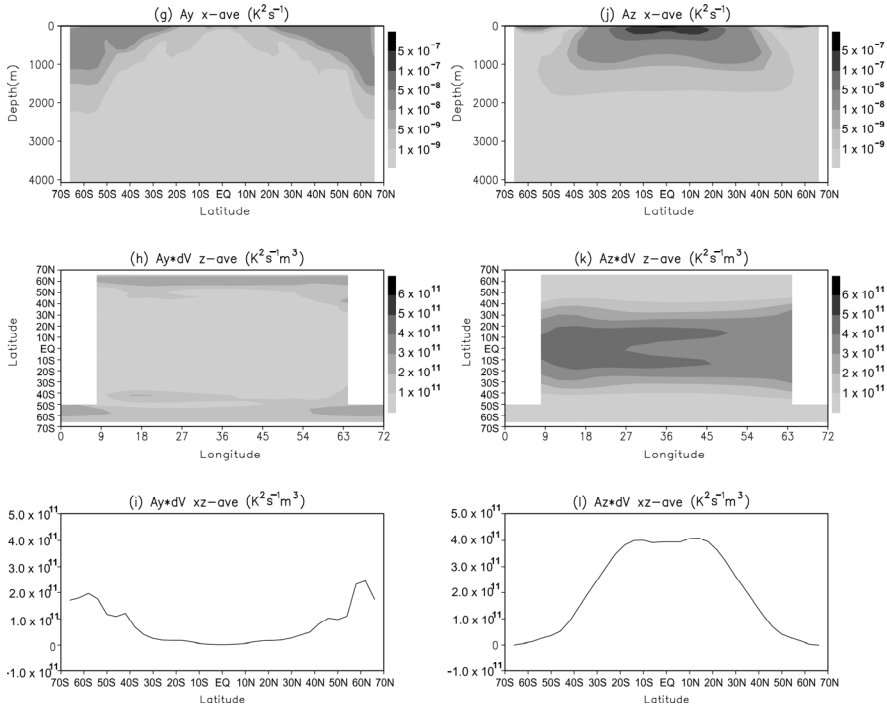


Fig. 5. (continued)

(a) zonal average of  $A$ , (b) depth average of  $A \times dV$ , (c) zonal-depth average of  $A \times dV$ , (d) zonal average of  $A_x$ , (e) depth average of  $A_x \times dV$ , (f) zonal-depth average of  $A_x \times dV$ , (g) zonal average of  $A_y$ , (h) depth average of  $A_y \times dV$ , (i) zonal-depth average of  $A_y \times dV$ , (j) zonal average of  $A_z$ , (k) depth average of  $A_z \times dV$ , (l) zonal-depth average of  $A_z \times dV$ . The unit for  $A$  is  $W K^{-1} m^{-3}$ . The unit for  $A \times dV$  is  $W K^{-1}$ . The unit for  $A_x$ ,  $A_y$ , and  $A_z$  is  $K^2 s^{-1}$ . The unit for  $A_x \times dV$ ,  $A_y \times dV$ , and  $A_z \times dV$  is  $K^2 s^{-1} m^3$ . The contour interval is indicated at the right side of each figure.

## 5. Discussion – Sandström theorem and abyssal circulation

As stated in section 1.5, Sandström suggested that a closed steady circulation can only be maintained in the ocean if the heating source is located at a higher pressure (i.e. a lower level) than that of the cooling source. Therefore, he suggested that the oceanic circulation is not a heat engine.

Huang (1999) showed using an idealized tube model and scaling analysis that when the heating source is at a level that is higher than the cooling source such as the real ocean, the circulation is mixing controlled, and in the contrary case, the circulation is friction-controlled. He also suggested that, within realistic parameter regimes, the circulation requires external sources of mechanical energy to support mixing to maintain basic stratification. Consequently, oceanic circulation is only a heat conveyer, not a heat engine.

Yamagata (1996) reported that the oceanic circulation can be driven steadily as a heat engine only with great difficulty, considering the fact that the efficiency as a heat engine of the

oceanic circulation calculated heating and cooling sources at the sea surface is very low, in addition to a view of Sandström's theorem. He therefore concluded that the oceanic circulation might not be driven steadily as a heat engine, but that it shows closed circulation by transferral to mechanically driven (e.g. wind-driven) flow on the way: the oceanic circulation might be sustained with a mixture of the buoyancy process and mechanical process.

However, these arguments are based on the assumption that the heating source is located only at the sea surface. If a diabatic heating because of turbulent diffusion takes place in the ocean interior (and the cooling source is placed at the sea surface), then Sandström's theorem is not violated. The important quantity in this respect is diapycnal diffusion, as stated in section 1, which corresponds to  $A_z$  in our model. As stated in section 4,  $A_z$  in our model showed high entropy production attributable to turbulent diapycnal diffusion down to 1000 m in the whole equatorial region (<30 deg). By contrast, the diapycnal diffusion at high latitude is very small and is confined to the surface in Fig. 5(j). Although there also exists dissipation caused by convective adjustment in the polar region, it can be negligible as the regional average: the region of adiabatic heating at low latitudes extends into the deeper layer (i.e. a higher pressure), but the region of adiabatic cooling at high latitudes is confined to the surface (i.e. a lower pressure). These results support the inference described above. In addition, the real ocean is also affected by dynamic interaction among tides, topography, and the resultant diabatic heating, which has not been considered in our model.

Moreover, the inference is supported by some experimental studies that the circulation is possible if external heating and cooling are placed at the same level (Park & Whitehead, 1999), or even if external heating is placed at a higher level than external cooling (Coman et al. 2006). Coman et al. (2006) reported that heat diffusion (whether by molecular conduction or turbulent mixing) allows heat to enter and leave the fluid at the boundary and causes the heating to be distributed throughout at least the depth of the boundary layer. Warmed water ascends towards the surface after having warmed and expanded at higher pressures than the surface pressure. Positive work is available from the heating and cooling cycle, even when the heating source is above the cooling source. Therefore, they concluded that Sandström theorem cannot be used to discount the formation of a deep convective overturning in the oceans by the meridional gradient of surface temperature or buoyancy forcing suggested by Jeffreys (1925). In addition, the driving force of the circulation in these experiments is only internal diabatic heating by molecular conduction or turbulent diffusion: the real ocean includes stronger diabatic heating due to external forcing of wind and tide, as explained in sections 1.2 and 1.3. In the equatorial region, the flow structure consisting of equatorial undercurrents and intermediate currents is organized such that forced mixing by wind stress at the surface accelerates turbulent heat transfer into the deeper layer. However, in the polar regions, forced mixing by wind stress at the surface does not reach the deeper layer, and adiabatic cooling is confined to the surface. For that reason, seawater expands at the high-pressure intermediate layer in the equatorial region because of heating and contracts at the low-pressure surface in the polar regions because of cooling. Consequently, mechanical work outside (i.e. kinetic energy) is generated and the circulation is maintained. The above inference will be strengthened in consideration of the real ocean.

Using numerical simulations, Hughes & Griffiths (2006) showed that by including effects of turbulent entrainment into sinking regions, the model convective flow requires much less energy than Munk's prediction. Results obtained using their model indicate that the ocean

overturning is feasibly a convective one. Therefore, they suggested that there might be no need to search for “missing mixing.” As stated in section 1.4, the idea of the ocean as “mechanical pump” was the idea derived to solve the “missing mixing” problem: the “mechanical pump” was introduced as another new mechanism of diapycnal mixing to maintain abyssal circulation. If their conclusion is correct in the real ocean, then the assumption of a “mechanical pump” (i.e. “missing mixing”) is not necessary. Small “background” diapycnal mixing might be sufficient to maintain abyssal circulation.

It is possible that the idea of the ocean as a “heat engine” is not fully contradicted by the idea of the ocean as a “mechanical pump”: it can be considered that a circulation driven as a “heat engine” is strengthened by a pump-up flow driven as a “mechanical pump”. In a sense, the idea of a mixture of buoyancy processes and mechanical processes by Yamagata (1996) might be right on target.

As stated in section 1.3, although recent observations of turbulence show large diapycnal mixing, such observations are limited to a few locations. It is not clear how much is the value of diapycnal mixing averaged in the entire ocean. Although global mapping of diapycnal diffusivity based on expendable current profiler surveys has been tried (Hibiya et al., 2006), the observed places remain limited. To verify the thermodynamic structure of the oceanic general circulation suggested in this chapter, the entire structure of adiabatic heating and cooling should be resolved. Particularly, observations of the following are recommended: 1) the structure of turbulent heat transfer into the intermediate layer because of forced mixing by wind stress at the surface and the resultant adiabatic heating in the equatorial region, 2) the process of adiabatic cooling confined to the surface and the subsequent concentrated sinking in the polar regions. In addition, direct observations of sinking and upwelling, not inferred from other observations, are important because the inferred value might include the effects of assumptions and errors. The observation of sinking is difficult because of severe climates in polar winter, with the worst conditions occurring when the sinking occurs. Moreover, observation of the upwelling itself is extremely difficult because of the low velocity. Future challenges must include technical improvements of observational instruments.

## 6. Conclusion

This chapter presented discussion of the problem of whether the abyssal circulation is a heat engine or a mechanical pump. We also discussed how it is related to the Sandström theorem, referring to results of numerical simulations of the oceanic general circulation. The results obtained using our model show high-entropy production due to turbulent diapycnal diffusion down to 1000 m in the entire equatorial region (<30 deg). By contrast, diapycnal diffusion at high latitude is very small and is confined to the surface: the region of adiabatic heating at low latitudes extends into the deeper layer (i.e. a higher pressure), but the region of adiabatic cooling at high latitudes is confined to the surface (i.e. lower pressure). In this case, Sandström’s theorem is not violated. In the equatorial region, the flow structure consisting of equatorial undercurrents and intermediate currents is organized such that forced mixing by wind stress at the surface accelerates turbulent heat transfer into the deeper layer. However, in polar regions, forced mixing by wind stress at the surface does not reach the deeper layer, and adiabatic cooling is confined to the surface. Consequently, seawater expands at a high-pressure intermediate layer in the equatorial region because of

heating and contracts at a low-pressure surface in polar regions because of cooling. Therefore, mechanical work outside (i.e. kinetic energy) is generated and the circulation is maintained. The results suggest that abyssal circulation can be regarded as a heat engine, which does not contradict Sandström's theorem.

## 7. Acknowledgments

This research was supported by the National Research Institute for Earth Science and Disaster Prevention, and by Hiroshima University.

## 8. References

- Broecker, W. S., & G. H. Denton (1990). What Drives Glacial Cycles?, *Scientific American*, pp. 43–50
- Broecker, W. S. (1987). The Largest Chill, *Natural History*, Vol. 97, No. 2, pp. 74–82
- Colling, A. (Ed.) (The Open University) (2001). *Ocean Circulation*, Butterworth–Heinemann, ISBN 978-0-7506-5278-0, Oxford
- Coman, M. A.; R. W. Griffiths & G. O. Hughes (2006). Sandström's Experiments Revisited, *J. Mar. Res.*, Vol. 64, pp. 783–796
- Defant, A. (1961). *Physical Oceanography*, Pergamon Press, London
- Dewar, R. C. (2003). Information Theory Explanation of the Fluctuation Theorem, Maximum Entropy Production and Self-organized Criticality in Non-equilibrium Stationary States, *J. Phys. A Math. Gen.*, Vol. 36, pp. 631–641
- Dewar, R. C. (2005). Maximum Entropy Production and the Fluctuation Theorem, *J. Phys. A Math. Gen.*, Vol. 38, pp. L371–L381
- Egbert, G. D., & R. D. Ray (2000). Significant Dissipation of Tidal Energy in the Deep Ocean Inferred from Satellite Altimeter data, *Nature*, Vol. 403, pp. 775–778
- Gade, H., & K. Gustafson (2004). Application of Classical Thermodynamics Principles to the Study of Oceanic Overturning Circulation, *Tellus*, Vol. 56A, pp. 371–386
- Gregg, M. C. (1989). Scaling Turbulent Dissipation in the Thermocline, *J. Geophys. Res.*, Vol. 94, pp. 9686–9698
- Hibiya, T.; M. Nagasawa & Y. Niwa (2006). Global Mapping of Diapycnal Diffusivity in the Deep Ocean Based on the Results of Expendable Current Profiler (XCP) surveys, *Geophys. Res. Lett.*, Vol. 33, L03611, doi:10.1029/2005GL025218
- Houghton, J. (2002). *The Physics of Atmosphere* (3rd ed.), Cambridge University Press, ISBN 978-0-521-01122-8, Cambridge
- Huang, R. X. (1998). On Available Potential Energy in a Boussinesq Ocean, *J. Phys. Oceanogr.*, Vol. 28, pp. 669–678
- Huang, R. X. (1999). Mixing and Energetics of the Oceanic Thermohaline Circulation, *J. Phys. Oceanogr.*, Vol. 29, pp. 727–746
- Huang, R. X. (2010). *Ocean Circulation*, Cambridge University Press, ISBN 978-0-521-85228-9, Cambridge
- Huges, G. O., & R. W. Griffiths (2006). A Simple Convective Model of the Global Overturning Circulation, Including Effects of Entrainment into Sinking Regions, *Ocean Modelling*, Vol. 12, pp. 46–79
- Jeffreys, H. (1925). On fluid motions produced by differences of temperature and humidity. *Q. J. R. Meteorol. Soc.*, Vol. 51, pp. 347–356

- Kleidon, A., & R. D. Lorenz (Eds.) (2005). *Non-Equilibrium Thermodynamics and the Production of Entropy: Life, Earth and Beyond*, Springer, Berlin
- Kuhlbrodt, T. (2008). On Sandström's Inferences from his Tank Experiments: a hundred years later, *Tellus*, Vol. A60, pp. 819–835
- Ledwell, J. R.; A. J. Watson & C. S. Law (1993). Evidence for Slow Mixing across the Pycnocline from An Open-ocean Tracer-release Experiment, *Nature*, Vol. 364, pp. 231–246
- Ledwell, J. R.; E. T. Montgomery; K. L. Polzin; L. C. St. Laurent; R. W. Schmitt & J. M. Toole (2000). Evidence for Enhanced mixing over Rough Topography in the Abyssal Ocean, *Nature*, Vol. 403, pp. 179–182
- Lorenz, E. N. (1955). Available Potential Energy and the Maintenance of the General Circulation, *Tellus*, Vol. 7, pp. 157–167
- Lorenz, R. D.; J. I. Lunine; P. G. Withers & C. P. McKay (2001). Titan, Mars, and Earth: Entropy Production by Latitudinal Heat Transport, *Geophys. Res. Lett.*, Vol. 28, pp. 415–418
- Lorenz, R. D. (2003). Full Steam Ahead, *Science*, Vol. 299, pp. 837–838
- Martyushev, L. M., & V. D. Seleznev (2006). Maximum Entropy Production Principle in Physics, Chemistry and Biology, *Phys. Rep.*, Vol. 426, pp. 1–45
- McComas, C. H. (1981). The Dynamic Balance of Internal Waves, *J. Phys. Oceanogr.*, Vol. 11, pp. 970–986
- Munk, W. H. (1966). Abyssal Recipes, *Deep-Sea Res.*, Vol. 13, pp. 707–730
- Munk, W. H., & C. Wunsch (1998). The Moon and Mixing: Abyssal Recipes II, *Deep-Sea Res.*, Vol. 45, 1977–2010
- Müller, P., & M. Briscoe (2000). Diapycnal Mixing and Internal Waves, *Oceanography*, Vol. 13, pp. 98–103
- Osborn, T. R. (1980). Estimates of the Local Rate of Vertical Diffusion from Dissipation Measurements, *J. Phys. Oceanogr.*, Vol. 10, pp. 83–104
- Ozawa, H., & A. Ohmura (1997). Thermodynamics of a Global-mean State of the Atmosphere – A State of Maximum Entropy Increase, *J. Clim.*, Vol. 10, pp. 441–445
- Ozawa, H.; S. Shimokawa & H. Sakuma (2001). Thermodynamics of Fluid Turbulence: A Unified Approach to the maximum Transport Properties, *Phys. Rev.*, Vol. E64, doi:10.1103/Phys. Rev. E. 64.026303
- Ozawa, H.; A. Ohmura; R. D. Lorenz & T. Pujol (2003). The Second Law of Thermodynamics and the Global Climate System: A Review of the maximum Entropy Production Principle, *Rev. Geophys.*, Vol. 41, doi:10.1029/2002RG000113
- Pacanowski, R. C. (Ed.) (1996). MOM2 Documentation, User's Guide and Reference Manual (GFDL Ocean Technical Report 3), Geophysical Fluid Dynamics Laboratory, Available from <http://www.gfdl.noaa.gov/ocean-model>
- Paltridge, G. W. (1975). Global Dynamics and Climate – A System of Minimum Entropy Exchange, *Q. J. R. Meteorol. Soc.*, Vol. 101, pp. 475–484
- Paltridge, G. W. (1978). The Steady-State Format of Global Climate, *Q. J. R. Meteorol. Soc.*, Vol. 104, pp. 927–945
- Park, Y. G., & J. A. Whitehead (1999). Rotating Convection Driven by Differential Bottom Heating, *J. Phys. Oceanogr.*, Vol. 29, pp. 1208–1220
- Peixoto, J. P., & A. H. Oort (1992). *Physics of Climate*, Amer. Inst. Phys., ISBN 0-88318-712-4, New York

- Polzin, K. L.; J. M. Toole; J. R. Ledwell & R. W. Schmitt (1997). Spatial Variability of Turbulent Mixing in the Abyssal Ocean, *Science*, Vol. 276, pp. 93–96
- Sandström, J. W. (1908). Dynamische Versuche mit Meerwasser, *Annalen der Hydrographie und Maritimen Meteorologie*, Vol. 36, pp. 6–23
- Sandström, J. W. (1916). Meteorologische Studien im Schwedischen Hochgebirge, *Göteborgs Kungl. Vetenskaps- och Vitterhetssamhälles Handlingar*, Vol. 17, pp. 1–48
- Sawada, Y. (1981). A Thermodynamic Variational Principle in Nonlinear Non-equilibrium Phenomena, *Prog. Theor. Phys.*, Vol. 66, pp. 68–76
- Shimokawa, S., & H. Ozawa (2001). On the Thermodynamics of the Oceanic General Circulation: Entropy Increase Rate of an Open Dissipative System and its Surroundings, *Tellus*, Vol. A53, pp. 266–277
- Shimokawa, S., & H. Ozawa (2002). On the Thermodynamics of the Oceanic General Circulation: Irreversible Transition to a State with Higher Rate of Entropy Production, *Q. J. R. Meteorol. Soc.*, Vol. 128, pp. 2115–2128
- Shimokawa, S. (2002). Thermodynamics of the Oceanic General Circulation: Entropy Increase Rate of a Fluid System (PhD thesis), The University of Tokyo, Tokyo
- Shimokawa, S., & H. Ozawa (2007). Thermodynamics of Irreversible Transitions in the Oceanic General Circulation, *Geophys. Res. Lett.*, Vol. 34, L12606, doi:10.1029/2007GL030208
- Toggweiler, J. R. (1994). The Ocean's Overturning Circulation, *Physics Today*, Vol. 47, pp. 45–50
- Yamagata, T. (1996). The Ocean Determining Decadal and Centurial Climate Variability (in Japanese), In *Kikou Hendouron (Climate Variability)*, A. Sumi (Ed.), 69–101, Iwanami Shoten, ISBN 4-00-010731-3, Tokyo
- Whitfield, J. (2005). Order Out of Chaos, *Nature*, Vol. 436, pp. 905–907
- Wunsch, C. (1998). The Work Done by the Wind on the Oceanic General Circulation, *J. Phys. Oceanogr.*, Vol. 28, pp. 2332–2340
- Wunsch, C., & R. Ferrari (2004). Vertical Mixing, Energy, and the General Circulation of the Oceans, *Ann. Rev. Fluid Mech.*, Vol. 36, pp. 281–314

# Thermodynamic of the Interactions Between Gas-Solid and Solid-Liquid on Carbonaceous Materials

Vanessa García-Cuello<sup>1</sup>, Diana Vargas-Delgadillo<sup>1</sup>,  
Yesid Murillo-Acevedo<sup>1</sup>, Melina Yara Cantillo-Castrillon<sup>1</sup>, Paola  
Rodríguez-Estupiñán<sup>1</sup>, Liliana Giraldo<sup>1</sup> and Juan Carlos Moreno-Piraján<sup>2</sup>

<sup>1</sup>*Facultad de Ciencias, Departamento de Química, Universidad Nacional de Colombia*

<sup>2</sup>*Facultad de Ciencias, Departamento de Química, Grupo de Investigación en Sólidos Porosos y Calorimetría, Universidad de Los Andes Colombia*

## 1. Introduction

For decades the man has had to face one of the major problems resulting from technological development and global population growth, environmental pollution, which has impacted on the different systems of life. The impacts of technological progress attained by man, have necessitated the establishment of international rules and regulations that set limits and establish a balance between development and the effects caused by the same (**Rodríguez 2003, Callister 2007, Rodriguez-Reinoso, 2007**). For this reason, we have launched various alternative solutions to environmental problems, including the synthesis and use of porous materials from organic waste or waste products with high carbon content, has been successful mainly in catalysis, adsorption and gas separation.

Activated carbon is a material that consists of microcrystals elementary hexagonal planes which are not well targeted, but displaced relative to each other and overlapping each other, so they have a high percentage of highly disordered structure. In fact there are hexagonal folding sheets with spaces of varying size (usually less than 2 nm) which make up the porosity of the material (**Marsh & Rodriguez-Reinoso, 2006**). These characteristics confer an exceptionally high surface area and good absorbent properties can be exploited in different areas. The production of activated carbon is linked to the purification of products and environmental protection. To the extent that the demands of purity of products require more sophisticated processes and emissions standards become more stringent, the activated carbon evolves, the production of the classic styles granular and powder have been joined by other like fibers, fabrics, monoliths among others (**Blanco et al., 2000**). Forms of activated carbon that are known and marketed, recent studies have shown that the monoliths exhibit characteristics that differentiate them from conventional ways, including the following highlights: allow the passage of gases with a very drop small, have a high geometric surface per unit weight / volume, the gas flow is very uniform, with easy handling, resistance to friction, reduce the constraints generated by phenomena of internal diffusion and mass transfer, these properties the have become used as support materials or adsorbents that favor direct adsorption process in the gas phase (**Nakagawa et al., 2007**).

In addition to their interesting adsorptive properties, during the process of obtaining activated carbons it is possible to modify and / or design their properties through a treatment of pre-or post-synthesis in order to obtain materials. These materials can be intended for more specific uses taking into consideration that the adsorption capacity of activated carbon depends on its textural characteristics and also on its surface chemistry (**Dias et al. 2007, Petit et al. 2010**). The surface chemistry of activated carbon is determined by the presence of atoms different from carbon atoms in graphene layers also known as heteroatoms, the most common are: oxygen, hydrogen and nitrogen. These heteroatoms produce a wide variety of surface groups, these groups can be acidic such as: carboxylic, lactone, anhydride and phenolic, while the basic character is attributed to groups such as quinones, pyrones besides of the delocalized  $\pi$  electrons in graphene layers or with a neutral character like carbonyl and ether (**Montes-Moran et al. 2004, Rivera-Utrilla & Sanchez-Polo. 2003**); surface groups determine parameters such as acidity and total basicity, surface charge and hydrophobicity, this will depend mainly on the type and concentration of groups on the surface of activated carbon.

Surface chemistry of activated carbon can be modified mainly by oxidation treatments and controlled thermal treatments that promote the formation and / or selective removal of different groups, as reported by Figueiredo and Pereira 2010. The role of surface chemistry in adsorption of gases, metal ions and organic molecules, is mainly due to specific interactions between surface groups and the different species.

In general, any organic material with relatively high carbon content is capable of being transformed into activated carbon. The activated carbon produced by manufacturing can come from wood and forest residues or other biomass (peat, lignite and other coals) as well as various polymers and natural or synthetic fibers. The factors that must be taken into account when choosing a suitable precursor are good availability and low cost, low mineral content and the resulting carbon possesses good mechanical properties and adsorption capacity. Wood waste, coconut shells, nuts, seeds and fruits along with some coals and petroleum coke are the most commonly used precursors (**Giraldo et al. 2008**). The complexity of the structure of activated carbon, can be characterized textural, chemical and energy by using different techniques among which include: physical adsorption of gases or vapors, scanning electron microscopy (SEM), electron microscopy transmission (TEM), infrared spectroscopy (FT-IR), X-ray diffraction, temperature programmed desorption (TPD), mercury porosimetry, immersion calorimetry and adsorption rates of methylene blue adsorption, density, size, moisture, ash, abrasion resistance and other parameters applicable to the specifications of coal (**Marsh & Rodriguez-Reinoso, 2006**).

The most common use is given to activated carbon adsorption processes of pollutants in liquid and gas phase. Adsorption is a process by which atoms, ions or molecules are trapped or retained on the surface of a material, which is different from absorption, which involves volume. This process involves an adsorbent, which in this case is activated carbon and adsorbate, which is the contaminant retained

The adsorption is exothermic, which can be classified as chemical or physical, depending on the magnitude of the interactions established between the retained material and the adsorbent surface. The amount of material that accumulates depends on the dynamic equilibrium is reached between the rate at which the material is adsorbed to the surface and the speed at which it evaporates or releases, in turn, they usually depend importantly on temperature.



Adsorption is a simple and attractive method for removing contaminants in aqueous phase, due to its high efficiency and easy handling. Given that the adsorption phenomenon is clearly exothermic, it is possible to characterize all the interactions inherent in the process of adsorption. For this purpose we describe the phenomenon of adsorption in terms of energy: If the interaction is solid-gas, may be the differential molar enthalpy of adsorption from the molar internal energy of the adsorbed phase and energy  $U_a$  adsorbate molar  $U_g$ , then, the change in internal energy as a result of adsorption can be represented as:

$$\frac{\Delta U_a}{n_a} = \bar{U}_a - \bar{U}_g = \Delta \bar{U}_a \quad (1)$$

In which,  $\frac{\Delta U_a}{n_a}$  is the molar energy change of the system, caused by the transfer of n moles of adsorbate from the gas phase to the adsorbed state.

This change in molar internal energy of the system is the integral molar energy of adsorption,  $\Delta U_a$ , which depends on interactions adsorbate-adsorbate and adsorbate-adsorbent.  $U_a$  is the average molar internal energy, for all the adsorbed molecules.

When an infinitesimal amount of moles of adsorbate,  $dn_a$ , is transferred to the solid surface from the gas phase at constant volume, the change in internal energy of the system will:

$$\Delta \bar{U}_a = \bar{U}_a - \bar{U}_g \quad (2)$$

Where  $\Delta \bar{U}_a$  is the differential molar energy of adsorption, and is defined as:

$$\bar{U}_a = \left( \frac{\partial U_a}{\partial n_a} \right)_{T,A} \quad (3)$$

Similar to what was expressed for energy,

$$\Delta H_a = H_a - H_g \quad (4)$$

Where  $H_a$  and  $H_g$  are the molar enthalpies of the adsorbate adsorbed state and gas phase. Likewise, the differential molar adsorption enthalpy is:

$$\Delta \bar{H}_a = \bar{H}_a - \bar{H}_g \quad (5)$$

In which:

$$\bar{H}_a = \left( \frac{\partial H_a}{\partial n_a} \right)_{T,P} \quad (6)$$

With the theoretical basis described above, it is possible to know then the extent of heat evolved in the adsorption process, known as heat of adsorption and is defined as a thermodynamic function, which can be used to characterize the surface of a solid .

There are two ways of expressing the heat of adsorption:

*Integral heat of adsorption*, amount of heat,  $Q$ , released when one gram of solid adsorbent adsorbs  $x$  grams. Its units are Joules / gram of adsorbent.

*Differential heat of adsorption*,  $\Delta H$ , amount of heat generated when  $x$  moles of adsorbate adsorbed. Its units are Joules / mole of adsorbate.

The differential heat of adsorption is also known as isosteric heat of adsorption and can be determined by using calorimetry or by adsorption isotherms, measured at two or more temperatures. (Bansal et al. 1988, Moreno-Piraján & Giraldo, 2007).

On the other hand, it is necessary to know the pore structure of the solid. Thus, for microporous solids without external surface (Kraus 1955),  $\Delta H_i$  immersion enthalpy is related to the isosteric heat of adsorption  $q^{isost}$  by the expression:

$$-\Delta H_i(T) = \int_0^1 q^{isost.}(T; \Theta) d\Theta - \Delta H_{vap.}(T) \quad (7)$$

Where:

$$\Theta = \text{degree of volume filling } W/W_o$$

The negative sign is because  $q^{isost}$  convention is positive, while  $\Delta H_i$  is negative.

Dubinin (Moreno-Piraján & Giraldo 2007, Dubinin 1975) define of net Heat of adsorption ( $q^{net}$ ) as:

$$q^{net} = q^{isost.} - \Delta H_{vap.}$$

when  $\Delta H_{vap}$  is vaporization enthalpy.

For the general case of  $n$ , the above equation becomes:

$$q^{net} = E_0 \left[ \left( \ln \frac{n_a^o}{n_a} \right)^{\frac{1}{n}} + \left( \frac{\alpha T}{n} \right) \left( \ln \frac{n_a^o}{n_a} \right)^{\left( \frac{1}{n} \right) - 1} \right] \quad (8)$$

Donde,

$\alpha$  = Thermal expansion coefficient of adsorbate

$n_a$  = Amount adsorbed at relative pressure  $P/P_o$

$n_a^o$  = Amount adsorbed at relative pressure  $P/P_o$

$E_o$  = Characteristic energy for vapor adsorption

$\Theta$  = Micropore filling degree

For activated carbons with  $n=2$

$$q^{net} = \beta E_o \left[ \left( \ln \frac{1}{\Theta} \right)^{\frac{1}{2}} + \left( \frac{\alpha T}{2} \right) \left( \ln \frac{1}{\Theta} \right)^{\left( \frac{1}{2} \right) - 1} \right] \quad (9)$$

$$-\Delta H_i = \int_0^1 q^{net} d\Theta$$

By replacing in equation (9):

$$-\Delta H_i = \beta E_0 \int_0^1 \left( \ln \frac{1}{\Theta} \right)^{\frac{1}{2}} d\Theta + \frac{\alpha T \beta E_0}{2} \int_0^1 \left( \ln \frac{1}{\Theta} \right)^{-\frac{1}{2}} d\Theta \quad (10)$$

The solution for the integrals of this type, is a function gamma ( $\Gamma(x)$ )

$$\int_0^1 \ln \left( \frac{1}{x} \right)^p dx = \Gamma(p+1) \quad (11)$$

$$\Gamma(p+1) = \begin{cases} \text{para } p = \frac{1}{2} : \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2} \\ \text{para } p = -\frac{1}{2} : \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \end{cases} \quad (12)$$

Thus, equation (10) can be written as:

$$\Delta H_i = -\frac{\beta E_0 \sqrt{\pi} (1 + \alpha T)}{2} \quad (13)$$

Which is valid for the micropores filling, in activated carbons at temperature T (**Moreno-Piraján & Giraldo 2007**).

For a coal with a micropore volume  $W_0$  and an adsorbate with a molar volume  $V_m$ , Stoekli and colleagues established a relationship between the enthalpy of different types of activated carbons in various body fluids and the parameters obtained by adsorption of vapors of these liquids on the same solid. In the equation of Stoekli and Kraehenbuehl (**Moreno-Piraján & Giraldo 2007, Stoekli & Kraehenbuehl 1989**):

$$\Delta H_i = -\frac{\beta E_0 W_0 \sqrt{\pi} (1 + \alpha T)}{2V_m} \quad (14)$$

Which corresponds to the immersion enthalpy  $\Delta H_i$ , en J/g.

Where  $\beta$  is the adsorbate affinity coefficient,  $E_0$  is the characteristic free energy of adsorption of vapor of reference,  $W_0$  is the total volume of micropores of the solid, is the coefficient of thermal expansion at temperature T, and  $V_m$  is the volume molar.

The immersion in a nonpolar solvent, which is known molecular dimensions, and generate a thermal effect only by filling the micropores of the solid, relates the thermal effect of the adsorbent surface as follows

$$\Delta H_{\text{exp.}} = \Delta H_i + h_i S_{\text{ext.}} \quad (15)$$

where  $h_i$  is specific enthalpy in J/m<sup>2</sup>.

As the external surface is defined as:

$$S_{\text{ext.}} = \frac{\Delta H_{i\text{exp.}}}{h_i} - \frac{\Delta H_i}{h_i} \quad (16)$$

So, the external surface may be known as:

$$S_{ext.} = A_{TOTAL} - A_{MICROPO.} \quad (17)$$

For a solid contribution also presents the mesopores and macropores, the total area is expressed as:

$$A_{TOTAL} = A_{MICROP.} + A_{MESOP.} + A_{MACROP.} + S_{abierta.} \quad (18)$$

Where,

$A_{TOTAL}$ : Total area

$A_{MICROP.}$ : Microporous area

$A_{MESOP.}$ : Mesoporous area

$A_{MACROP.}$ : Macroporous area

$S_{ABIERTA}$ : Open area non-porous

When activated carbon presents neither external surface nor development of meso and macropores, the experimental enthalpy will:

$$\Delta H_{exp} = \Delta H_i \quad (19)$$

determinate by equation (15).

Until here, have been described the theoretical foundations of solid-gas interactions and solid-liquid, which can calculate the magnitude of the thermal effect generated and the type of interaction between the adsorbate and the adsorbent, which is proportional to the affinity between them, and allows for relationships that give an approximation of the surface area of adsorbent.

Now, referring to such calorimetric techniques, such as immersion calorimetry and adsorption for measuring the thermal effect of the interaction of a solid with a liquid or gas, which leads to determine the thermodynamic variables in different interfaces: solid-liquid, solid-gas and liquid-liquid, which provide valuable information about the behavior of the solid. Such is the importance of these techniques, which in recent years, several studies are focused on the design, construction and development of immersion calorimeters adsorption.

#### *Calorimetric technique*

Calorimetry is a technique used to measure a large number of processes that generate heat in their development, and which can determine the enthalpy changes  $\Delta H$  and combination with other techniques to obtain the free energy change  $\Delta G$  and entropy changes  $\Delta S$ . Furthermore, this technique has been used to study physical and chemical events that occur on the surface of a solid. (Moreno 1996).

However, there is a team that is capable of performing all the above measures at a time, so different equipment should be used according to the type of study that you want to perform and the amount of heat generated in the system. For this reason, there are combustion calorimeters, adsorption and immersion, which basically consist of:

Calorimetric cell, which should be designed with a capacity that is associated with the type of sensors require the calorimeter and additionally must have regard to the material of construction depending on the system to be studied. In this cell are often attached a heating sample holder.

Electrical calibration system, which consists of an electrical circuit which provides power to the calorimeter to determine the constant of the system under study. Electrical work

supplied can be calculated if one knows the potential (Eh) through the heating resistor (Rh), the current (I) and heating time (t).

$$W_{elect} = Eh * I * t \quad (20)$$

Thermometric system for measuring thermal effect, which consists of different types of sensors, which can be proportional to the temperature or property connected with the transfer of heat.

According to the system you want to measure, you must use a specific calorimetric system. Below is a brief description of immersion calorimetry and sorption

*Immersion calorimetry, measurement of solid-liquid interactions.*

For many years, immersion microcalorimetry has been a useful technique for the characterization of powders and porous solids like activated carbons and oxides (Hemminger & Höhne 1984). Technique involves immersing a known quantity of a solid in a specific liquid, and measure the heat generated due to wet the solid, liquid immersion.

In the absence of complex effects such as filling of micropores, is usually taken as a first approximation, the energy due to the immersion of a solid degassed  $\Delta_{im}U^o$ , which is proportional to the solid surface, A, according to Equation 21:

$$\Delta_{im}U^o = A \cdot \Delta_{im}u^{i,o} \quad (21)$$

in which the energy of immersion per unit area,  $\Delta_{im}u^{i,o}$  is characteristic of the nature of solid-liquid system.

When  $\Delta_{im}U$ , is known for a given solid-liquid system, the adsorbent surface (A) can be evaluated. When the surface of the sample of adsorbent is less than 1 m<sup>2</sup>, generates heat due to immersion, which is easily measured by calorimetric procedures and therefore the immersion microcalorimetry can be used to evaluate the specific surface of adsorbent (Rouquerol et al. 1999).

Immersion calorimetry is a useful technique to assess the total area and size distribution of micropores of a microporous carbon (Denoyel et al.1993), assuming that the energy of the dip is proportional to the area available for liquid immersion to any size and shape of the pores. In addition, it is assumed, from the point of view of energy per unit external surface area of solid has the same behavior (Rouquerol et al. 1999, Hemminger & Höhne 1984).

The Figure 1 shows the immersion calorimetric heat conduction unit. To experimentally measure the immersion heat, the adsorbent is immersed in the liquid which is to determine the interaction. You can use a microcalorimeter heat conduction, which is expected to be reached thermal equilibrium between all components of the calorimetric system: the cell containing the immersion liquid, the vial containing the solid under study, a heating pad for perform system calibration, temperature sensors should be arranged around the cell containing the immersion fluid and the surroundings. To achieve this, the entire system must be completely insulated from temperature fluctuations. Once thermal equilibrium is reached, it is the breaking of the ampoule to allow liquid to come into contact and the adsorbent, it ends with an electrical calibration. Throughout the experiment, recorded the potential generated by the sensors, should have the thermal effect sensor thermocouples or thermopiles and evaluates the area under the curve of the signal generated in response to solid-liquid interaction.

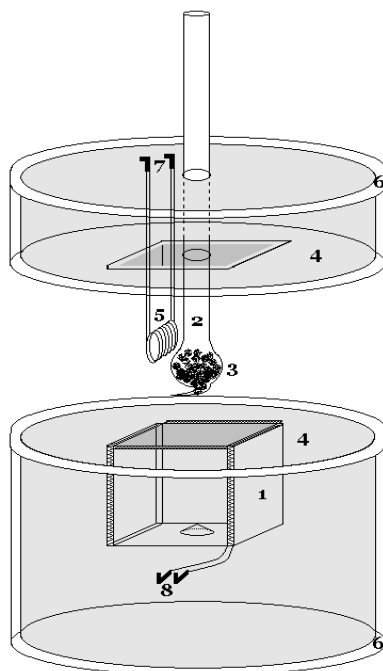


Fig. 1. Calorimeter immersion scheme Tian type. (1)Sensors System; (2) Sample cell; (3) Sample; (4) Heat Sink; (5) Heat resistance for calibration; (6) Insulation jacket; (7) Output of resistance to power supply; (8) Output of sensors system to interface multimeter.

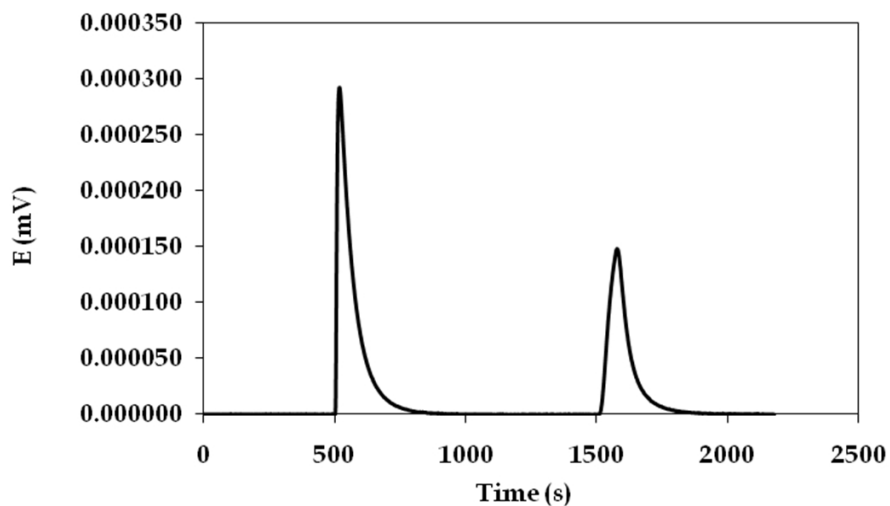


Fig. 2. Thermogram obtained for the immersion of an activated carbon pellet ore (CAP), in benzene

Figure 2 shows a typical thermogram obtained for the immersion of an activated carbon pellet ore (CAP), in benzene. It can be seen in the range of 0 to 500 seconds, the baseline obtained, which illustrates the heat balance and low noise level in the calorimetric signal. Table 1 shows the values of surface properties obtained by immersion calorimetry, for this same sample, two samples obtained by the modification of the CAP.

Sample	$E_o$ kJ/mol	$W_o$ cm <sup>3</sup> /g	$S_{BET}$ m <sup>2</sup> /g
CAP	7.47	0.43	1248
CAPRED	6.48	0.38	1089
CAPN65	7.47	0.43	1253

Table 1. Surface properties obtained for three activated carbons by gas adsorption

The sample CAPRED is a modification of CAP, obtained by heating the same until 1373 K, under nitrogen, for 3h. CAPN65 sample is a sample obtained by modification of CAP through the impregnation of CAP with 65% HNO<sub>3</sub> and heating it to 473 K, for 2 hours.

As shown in Table 1, the modification with HNO<sub>3</sub> and 65% did not produce a significant change in the surface properties of the sample. This behavior is attributed to the low temperature at which it made the change, which did not affect the porous structure of the solid. The modification to 1373K nitrogen affected the pore structure of the solid, reducing the volume of micropores and consequently, the surface area there of in Figure 3 shows the isotherms of nitrogen at 77 K for these three samples.

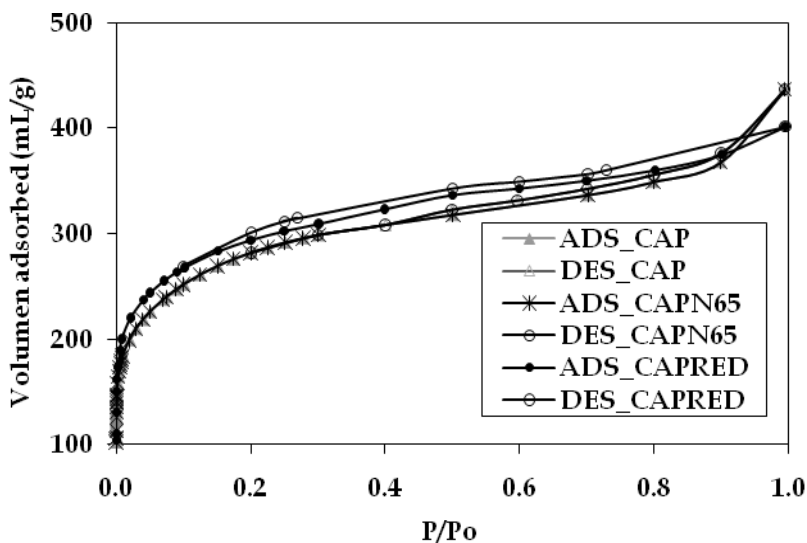


Fig. 3. Nitrogen adsorption isotherms at 77 K, for the three carbons under study

The isotherm can be observed further that the sample has CAPRED mesoporosity development, so it appears the hysteresis loop in it. CAP and CAPN65 isotherms are

virtually identical, confirming that the modification with 65% HNO<sub>3</sub> shows no effect on the texture of activated carbon.

Table 2 shows the surface properties obtained by immersion calorimetry for these three samples

Sample	S <sub>ext</sub> m <sup>2</sup> /g	ΔH <sub>imm</sub> J/g	ΔH <sub>exp</sub> J/g	A <sub>MICROP</sub> m <sup>2</sup> /g	A <sub>total</sub> m <sup>2</sup> /g
CAP	37	-44	-48	1219	1256
CAPRED	55	-33	-39	1065	1120
CAPN65	64	-34	-41	1219	1283

Table 2. Surface properties obtained for three activated carbons by immersion calorimetry in benzene

The results for ΔH<sub>imm</sub>, from equation (14), and ΔH<sub>exp</sub> are the experimental results. The external surface area S<sub>ext</sub>, of micropores A<sub>MICROP</sub>, and the total area A<sub>total</sub>, were obtained from the equations (16), (17) and (18) respectively.

Table 3 shows the parameters used for calculations of surface properties obtained by immersion calorimetry into benzene.

α	B	V <sub>m</sub>
1,24E-03	1	88,9

Table 3. Physical characteristics of benzene.

Figure 4 shows a relationship between the areas obtained by gas adsorption and that obtained by immersion calorimetry.

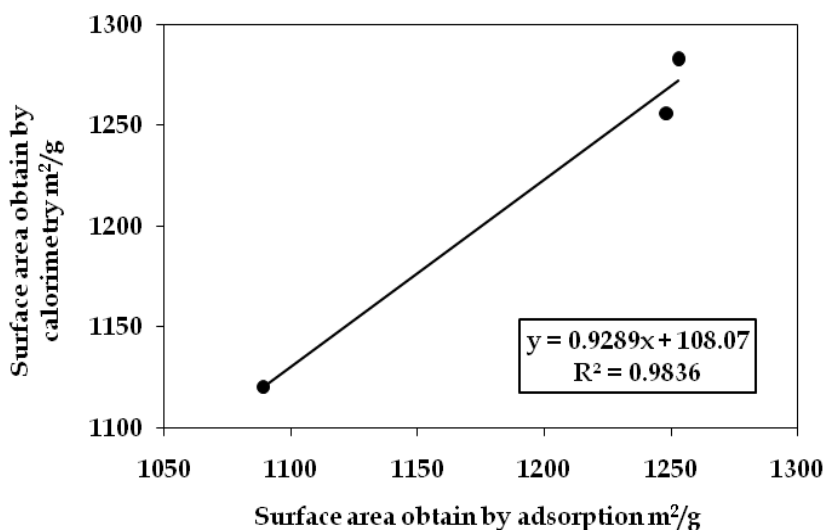


Fig. 4. Relationship between the total area obtained by adsorption calorimetry and nitrogen adsorption.



From these results we can see good correlation between the results obtained by the two methods compared, which shows a correlation coefficient of 0.9836, confirming that immersion calorimetry is a characterization parameter for solid-liquid interactions. You could make a more exhaustive with probe molecules of different sizes to benzene, since the pore size distribution can affect the calorimetric data (Molina-Sabio et al. 2008).

*Adsorption calorimetry, measurement of solid-gas interactions.*

There are several reasons to determine the heat of adsorption to characterize the surface energy of materials (Rouquerol et al. 1999), provide basic data for development of new theories of equilibrium and kinetics of adsorption (Zimmermann & Keller 2003), design and plants improve separation processes by adsorption and desorption, PSA, VSA, TSA and their combinations (Ruthven 1984, Yang 1997).

Adsorption calorimetry in combination with other physical or chemical properties to describe the properties of a solid surface (Garcia-Cuello et al. 2009, Llewellyn & Maurin 2005, Garcia-Cuello et al. 2008, Moreno & Giraldo 2005).

To experimentally measure the heat of adsorption, calorimetric unit is used as shown in Figure 5.

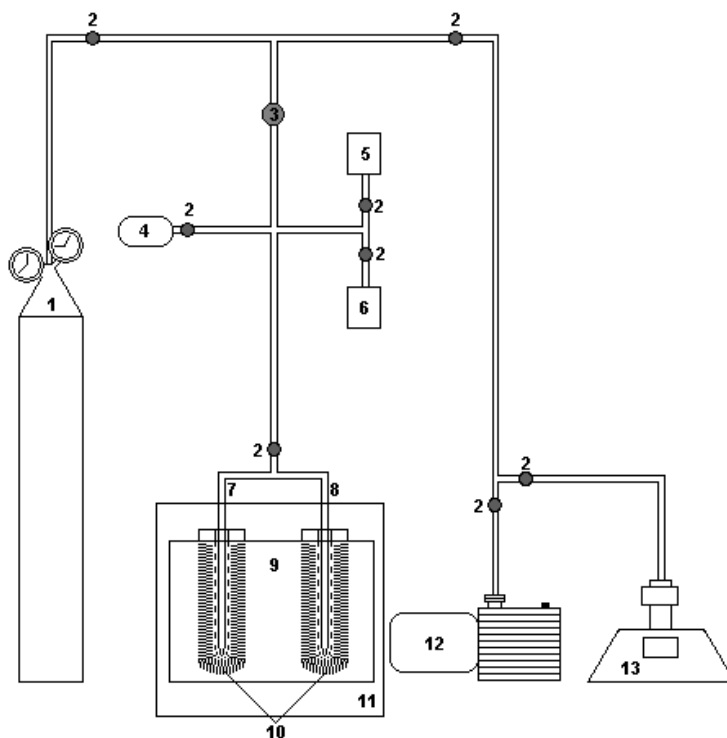


Fig. 5. Adsorption calorimeter scheme. (1) Adsorbate, (2) precision valves, (3) needle valve, (4) Volume calibration, (5) pressure transducer 1 to 1000mbar, (6) pressure transducer 10<sup>-4</sup> to 1 mbar , (7) measuring cell, (8) reference cell, (9) Calorimeter adsorption (10) thermopile sensors in 3D layout type, (11) thermostat, (12) Rotary Vacuum Pump, (13) Pump ultra high vacuum

The heats of adsorption measured at a temperature of liquefaction of the adsorbate, in the case of nitrogen at 77 K and 273 K. CO<sub>2</sub> For this, use a thermostat bath at that temperature. Make contact with the solid adsorbate successive small doses. This allows measure the evolution of the interaction energy compared to coverage. Before start the calorimetric measurements. To start the measurements in the microcalorimeter, initially must be empty throughout the adsorption system, including the solid sample under study, using a vacuum system that achieves at least 10<sup>-3</sup> Torr. When the system reaches the expected vacuum level, are the respective gas injection, waiting time for a balance between system components and are simultaneously recorded volumes of gas adsorbed and the heat evolved at each injection. Developed to sense heat, temperature sensors are used thermopile type, with appropriate sensitivity to detect heat from 10 to 100 J / g. Pressure readings are made using a pressure sensor with adequate sensitivity and precision must be known in the injection volume. The differential molar adsorption energy can be obtained by equation (3), and evaluating the area under the curve obtained in the experiment, which is the signal generated by the thermopile due to solid-gas interaction which is proportional to the adsorption energy (Garcia-Cuello et al. 2008, Garcia-Cuello et al. 2009).

*Preparation, characterization, modification and use of carbonaceous Materials*

Preparation, characterization, modification and use of carbonaceous materials like activated carbon in different presentation such as: granulate, powder, pelettes, char, monoliths, among other, it has been object investigation during many years. Next are presented some results of investigations developed in the by the authors about these porous solids and their employment in the adsorption of pollutants in liquid and gas phase.

*Bone char in the adsorption of derivates phenolics*

The bovine bone char (BBC) have received attention by industry of treatment waste water; due to its advantages in front of others adsorbents between these are found: low cost and adsorbent versatility for wide variety pollutants (Deyder et al., 2005). The BBC was prepared in the following way: The bones were cleaned from meat and fat and cut by saw to pieces of approximate size 4-10 cm. Subsequently, bones were washed with tap water for several times. The bones were then transferred to the oven for drying at 353 K. After 24 h, the dried bones were crushed and milled into different particle sizes in the range of 2-3 mm. These particles are burned in an inert atmosphere. This process was carried out in a tubular fixed bed reactor from room temperature to 1073 K for 2 h at a heating rate of 3 K min<sup>-1</sup> and a flow of N<sub>2</sub> 80 cm<sup>3</sup> min<sup>-1</sup>.

The adsorption from solution depends on the chemical and physical characteristics of the solid as surface area, porosity and surface chemistry, see Table 4. The study about this process has shown dependence with the solution characteristics as pH, ionic strength and temperature (Moreno-Castilla & López-Ramos M.V., 2007). These factors have influence in the adsorption mechanism and in consequence, the magnitude in that the system - (solid-liquid) - liberates heat.

S <sub>BET</sub> (m <sup>2</sup> / g)	157
Pore Volume (cm <sup>3</sup> /g)	0.14
Pore Size (nm)	3.0
Acid Sites (meq/g)	0.23
Basic Sites (meq / g)	0.42
PZC	8.5

Table 4. Physical and chemical characteristics of the BBC

The chemical properties of the adsorbent depends the surface concentration of acid and basic sites, but these are in pH function of solution because the charge on the surface depends of this property. In this study was used 2,4-Dinitrophenol (DNP) a organic compounds commonly used for tincture manufacturing, wood preservatives, explosives, substances for insects control and other chemical products (Su-Hsia & Ruey-Shin, 2009, Tae Young et al., 2001) that in aqueous solution can be found as ionic or nonionic species Figure 6.

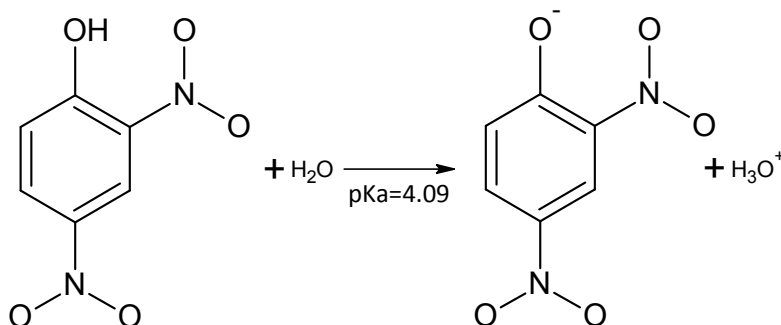


Fig. 6. Species of DNP in aqueous solution.

The adsorption isotherm represents the thermodynamic equilibrium between the adsorbed solute and the solute in solution, the obtained equilibrium data which are used to assess the ability of adsorbent to adsorb a particular molecule.

Figure 7 shows the influence of concentration on the adsorption of DNP on CHB, where the mass of solute adsorbed onto the adsorbent continues to increase when raising the concentration of solute in equilibrium and is not asymptotic at high concentrations.

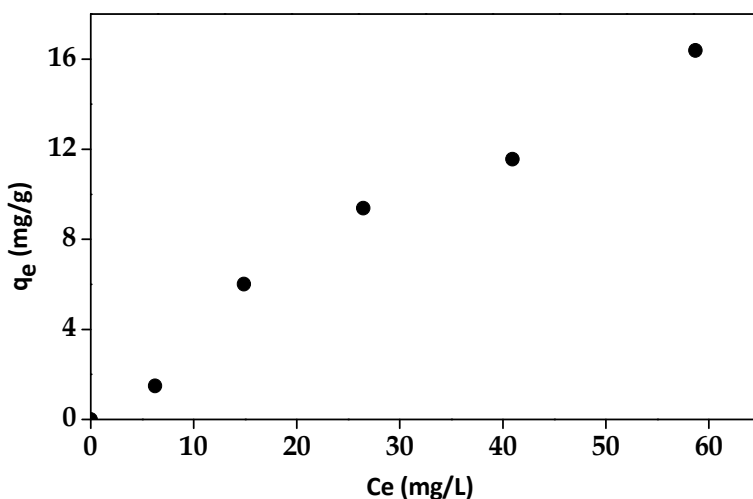


Fig. 7. Adsorption isotherm of DNP on bone char

In the literature on liquid phase adsorption has been reported different mathematical models to represent the adsorption isotherms, the most used are the Langmuir and Freundlich model. The first assumes: (i) uniform adsorption energies on the surface, (ii) no interaction between adsorbed molecules (III) adsorption occurs at specific sites. Meanwhile, the second (I) assumes that the adsorbent surface is energetically heterogeneous, (ii) that increasing the concentration of adsorbate, increases the amount adsorbed on the surface (Oke et al., 2008, Moreno et al., 2010).

These models are represented mathematically as shown in table 5:

Isotherm	Equation	Lineal Form	Graphic
Langmuir	$q_e = \frac{q_m * b * C_e}{1 + b * C_e}$	$\frac{1}{q_e} = \frac{1}{b * q_m} \frac{1}{C_e} + \frac{1}{q_m}$	$\frac{1}{q_e}$ vs. $\frac{1}{C_e}$
Freundlich	$q_e = k_f * C_e^{1/n}$	$Ln q_e = Ln k_f + \frac{1}{n} Ln C_e$	$Ln q_e$ vs. $Ln C_e$

Table 5. Mathematics models of Langmuir and Freundlich.

where  $q_e$  is the amount adsorbed at  $C_e$  (mg/L), concentration of DNP at equilibrium,  $b$  (L/mg), and  $q_0$  (mg/g) are the Langmuir constants related to the energy of adsorption and maximum capacity, respectively;  $k_f$  ( $mg^{1-1/n} l^{1/n} g^{-1}$ ) and  $1/n$  are the Freundlich constants related to the adsorption capacity and intensity, respectively; and  $q_e$  (mg/g) is the mass of DNP adsorbed per mass of adsorbent.

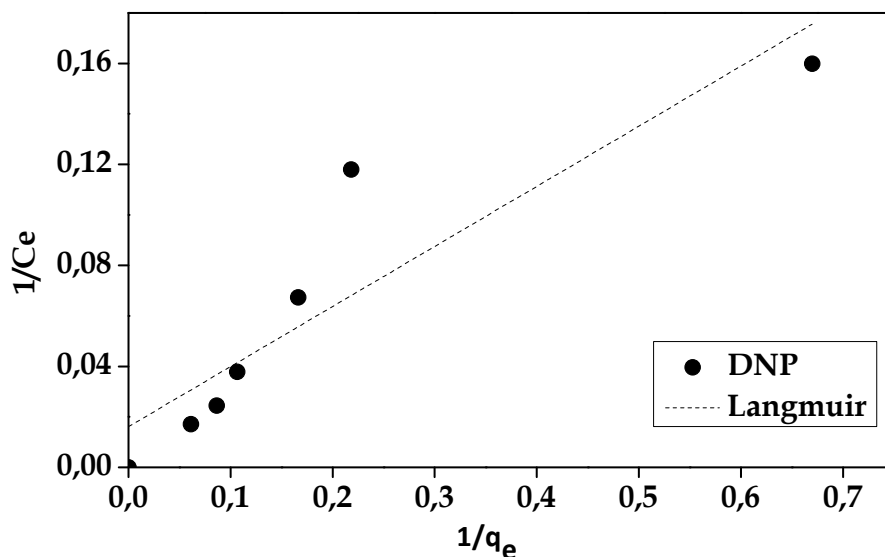


Fig. 8. Adsorption isotherm Langmuir model

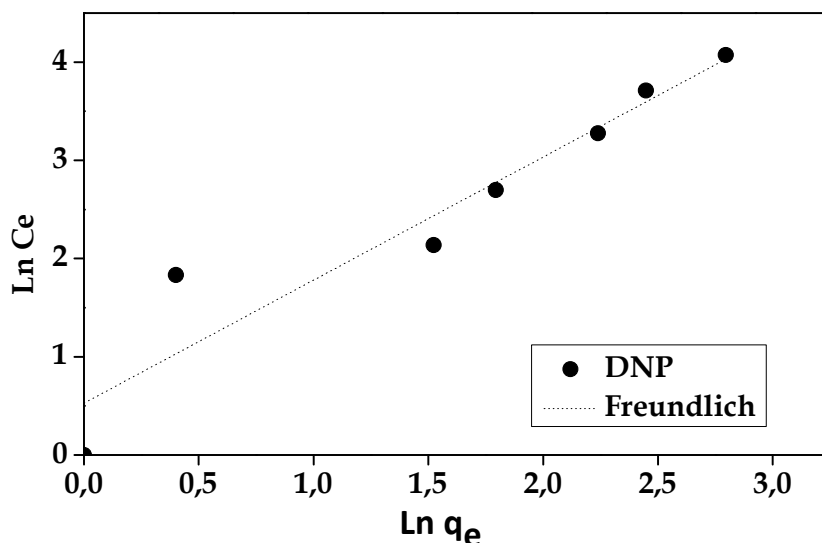


Fig. 9. Adsorption isotherm Freundlich model

Isotherm	Parameter Values
Langmuir	$q_m = 61.96$ $b = 0.068$ $R^2 = 0.7969$
Freundlich	$K_F = 0.593$ $n = 0.798$ $R^2 = 0.8907$

Table 6. Isotherm parameters for Langmuir and Freundlich models.

Correlating the experimental data of adsorption of DNP on BBC with both models, Figure 8 and 9 shows the typical behavior of the Freundlich isotherm, which contrasts with the parameters and correlation coefficients, see Table 6. This model describes the surface of the adsorbent is energetically heterogeneous and includes the lateral interactions between adsorbate molecules. In this type of liquid-solid systems, it is important understand that when a model fits the experimental data does not support the adsorption mechanism occurs under the principles of the model. Although these data are adjusted by mathematical methods - statistics to calculate the parameters given, these methods do not consider the interactions between adsorbate and surface active sites.

Depending on the thermodynamic conditions of the system, heat is produced when a solid comes into contact with the solution; this intensity is determined by immersion enthalpy. It is set for a specific amount of a solid and measured by a technique known as immersion

calorimetry (Blanco et al., 2008). When make this type of measure, where contact between a solid and a solution is involved, there are different interactions that contribute to the total amount of heat produced. Among these are interactions between water and the groups on the solid's surface, the filling of pores and adsorption on the surface. Furthermore, there are also adsorption of and interactions with the solute; these depend on the characteristics of the solution (Moreno-Piraján et al., 2007).

The values of the enthalpies of immersion were evaluated from the thermograms, where the heat generated by the process of adsorption is proportional to the area under the curve of the peak generated by the thermal effect. Figure 10 shows the typical thermograms for the immersion of BBC in DNP solutions of 10 and 30 mg/L.

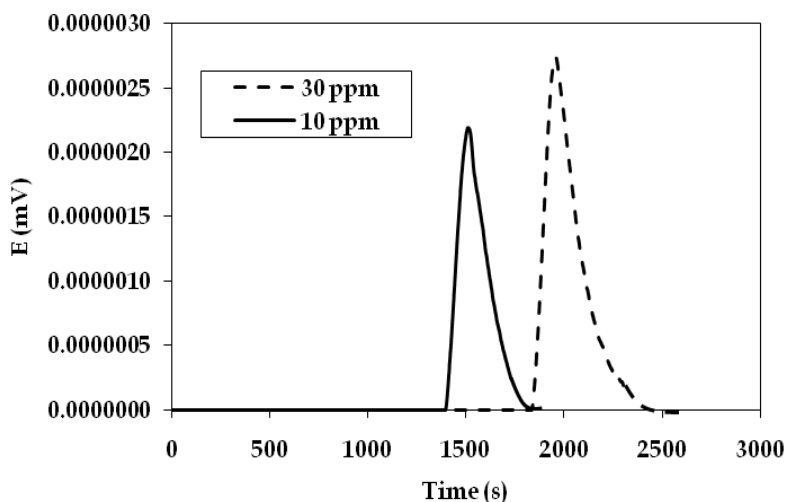


Fig. 10. Thermograms of BBC immersion in a solution of DNP at concentrations of 10 and 30 mg/L at 298 K.

Figure 11 shows the (a) interactions between bone char and DNP in solutions at different concentrations and the (2) interactions with the adsorbate char was obtained subtracting the effect of char-water interactions.

As can be seen in the Figure 11, at low concentrations (10-30 mg/L) there was a greater interaction between the BBC and the adsorbate (DNP); however, as the concentration increased (50-100 mg/L) there was a decrease in enthalpy, i.e. weaker interactions between the adsorbent and the adsorbate.

When relating the enthalpies as a function of adsorbed amount of DNP can be seen that the enthalpy is directly proportional to the percent of retention, this behavior is due to the main morphological characteristic of the material is its heterogeneity, therefore the heat generated is different because that the adsorbate has occupied the most active sites than the immediately occupy.

The differential free energy of adsorption that occurs in the time interval, in which it is carry a calorimetry measure, is determined relating the kinetics of the process to this time interval. Where  $t_{inicial}$  is the time in that started the immersion solid-liquid and  $t_{final}$  is the time in that ended the calorimetric measurement. The free energy difference as a thermodynamic

parameter is the fundamental criterion of spontaneity (Smiciklas et al., 2008), and may be calculated considering the initial concentration (Co) and the concentration in the equilibrium (Ce) to tfinal, by equation (22).

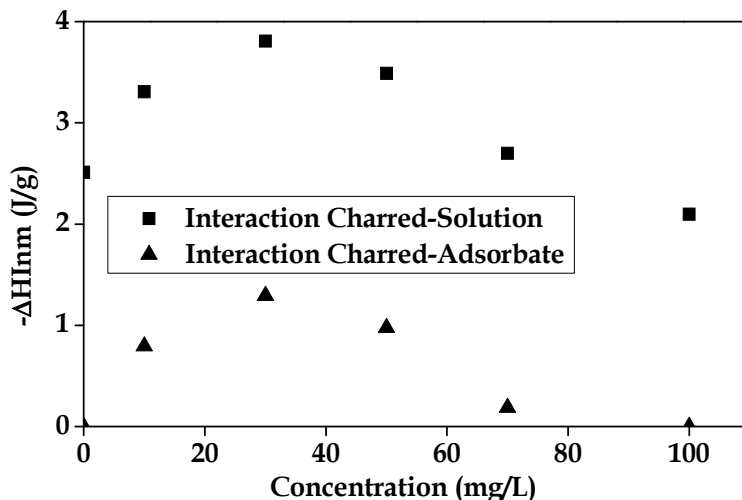


Fig. 11. Enthalpies of immersion of BBC on DNP to different concentration.

$$\Delta G_{li}^{ff} = -RT \ln \left( \frac{C_o}{C_e} \right) \quad (22)$$

Where  $\Delta G$  (kJ/mol) is differential free energy change; R is universal gas constant, and T (K) is absolute temperature. Reaction occurs spontaneously if G is a negative quantity. From the above equation, the differential change Gibbs free energy for the adsorption process of DNP on BBC to tfinal (293 K) is -113.0 kJ/mol, this negative value indicate that the adsorption of DNP is thermodynamically feasible.

In the specific case of the solution 30 mg/L which has a differential  $\Delta H_{imm} = -56.10$  kJ/(g\*mol) and substituting the parameters known in equation (23) determine the differential entropy of the process is equivalent to 580 J/(mol K) this positive value suggests that the organization of the adsorbate in the solid-liquid interface and coincide with value obtained for free energy.

$$\Delta S_{li}^{ff} = \frac{\Delta H_{imm} - \Delta G}{T} \quad (23)$$

#### Granular activated carbon for adsorption of nickel

The samples used for nickel removal were obtained from a commercial granular activated carbon made from coconut shell GAC, which was oxidized with GACoxN 6M nitric acid, two parts of this sample were treated one at 723 K and another 1023 K under nitrogen atmosphere, GACoxN723 and GACoxN1023, and a final sample obtained by heating the sample at 1173 K GAC, GAC1173, these samples were characterized by N<sub>2</sub> physisorption at -

196 ° C and their surface chemistry by Boehm and determining the point of zero charge, in addition, an immersion calorimetry was conducted in different liquids, such as benzene, carbon tetrachloride and water.

Sample	Área <sub>BET</sub> m <sup>2</sup> /g	V <sub>o</sub> cm <sup>3</sup> /g	Carboxylic µmol/g	Lactonic µmol/g	Phenolic µmol/g	Acidity Total µmol/g	Basicity Total µmol/g	pzc
GAC	842	0.34	72.2	40.5	85.0	198	90.5	5.4
GACoxN	816	0.35	267	52.4	73.7	393	48.6	3.4
GACoxN723	903	0.32	95.3	60.2	112	268	103	7.9
GACoxN1023	935	0.37	2.36	10.2	47.9	60.5	266	8.2
GAC1173	876	0.35	0.00	11.5	34.9	46.4	278	8.9

Table 7. Physical and chemical parameters of samples

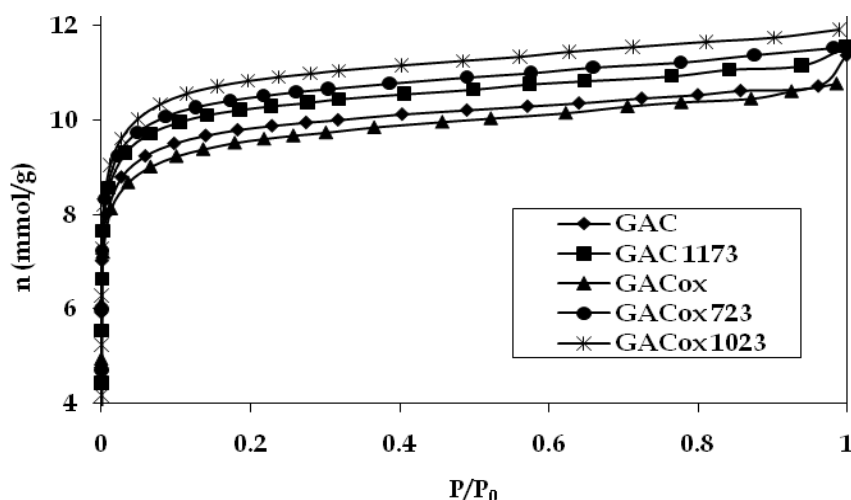


Fig. 12. N<sub>2</sub> adsorption isotherms at 77 K for different samples

The isotherms of nitrogen obtained for each sample are shown in Figure 1. These are classified as type I adsorption isotherms, where a knee at low relative pressures is evidenced, characteristic of microporous solids in accordance with data obtained after applying the Dubinin-Raduskevich equation. It is important to note that the oxidation process caused a decrease in surface area, this is explained by considering that the oxidation with nitric acid promotes the formation of surface oxygenated groups at the edges of the openings of the pores, these groups are mainly carboxylic and carbonyl (Dias *et al.* 2007; Daud & Houshamnd. 2010; Yin *et al.* 2007) besides producing the collapse of certain porous structures (Radovic *et al.* 2000; Yin *et al.* 2007; Silvestre-Alvero *et al.* 2009), additionally, a surface area increase can be observed even in relation to the sample treated with a higher temperature. This is a result of selective removal of surface groups formed in the oxidation processes, which break down into carbon monoxide and carbon dioxide. In other words, with heat treatment more carbon atoms are lost promoting surface area increase in the solid.



On the other hand, we evaluated the changes in surface chemistry of each sample, taking into consideration the important role of surface chemistry on the removal of dissolved metals in aqueous solutions. Table 7 shows the results of the amount of surface groups of each of the samples obtained through Boehm titration. It is observed that the content of acid groups increased by the oxidation treatment, favoring mainly the formation of carboxylic groups, such as reported in other studies (Gao *et al.* 2009). Additionally heat treatment changed the number of groups according to their different thermal stabilities, so, in general it is considered that at low temperatures (about 700 K) and in an inert atmosphere carboxylic groups decompose; in the range of 1000 K lactone groups, carboxylic anhydrides, phenol and ether decomposition is favored; and in higher temperatures up to 1200 K quinone and pyrone groups decompose. On the other hand the values of zero point of charge are consistent with changes in surface chemistry of each sample according to the treatment applied. (Chingombe *et al.* 2005; Szymański *et al.* 2002; Figueiredo *et al.* 1999; Figueiredo & Pereira. 2010)

As for the characterization of samples obtained by immersion calorimetry, it is important to note that the enthalpies of immersion allow to evaluate the type of interactions that occur between the solid and the wetting liquid, considering that: if there are no specific interactions between the molecules of the wetting liquid and the solid surface, the immersion enthalpy corresponds to the accessible area of the molecule of the liquid, and if on the contrary, there are specific interactions as in the case of some samples immersed in water, the immersion enthalpy would indicate the hydrophobic or hydrophilic character of the surface of the sample. (Stoekli *et al.* 2001; Szymański *et al.* 2002)

Table 8 shows the results obtained by calculating the enthalpies of immersion in benzene, carbon tetrachloride and water. As for the results using molecules with bipolar moments equal to zero it was observed that: the enthalpies of immersion changed according to changes of surface area as shown in Figure 2 and for the oxidized sample, which has a lower surface area, the enthalpy of immersion is less than for the original sample and even for the heat-treated ones in according to what it was discussed in the analysis of nitrogen adsorption isotherms, the same trend was observed for the enthalpies of immersion in carbon tetrachloride, although values were lower in these enthalpies of immersion, this is basically due to the difference in the size of the molecules of each liquid, which for benzene is 0.37 nm and for carbon tetrachloride is 0.66 nm, in other words, the carbon tetrachloride molecule has diffusion restrictions, therefore the interactions involved correspond only to pores in which the molecule does not have this restrictions, this situation does not occur with benzene that is smaller.

Sample	$\Delta H_{ImmC_6H_6}$ (J/g)	$\Delta H_{ImmH_2O}$ (J/g)	$\Delta H_{ImmCCl_4}$ (J/g)
GAC	-106.4	-49.65	-75.05
GACoxN	-94.98	-66.59	-85.87
GACoxN723	-107.9	-53.32	-50.76
GACoxN1023	-128.8	-37.39	-57.01
GAC1173	-145.1	-32.39	-94.29

Table 8. Enthalpies of immersion in Benzene, Carbon Tetrachloride and water.

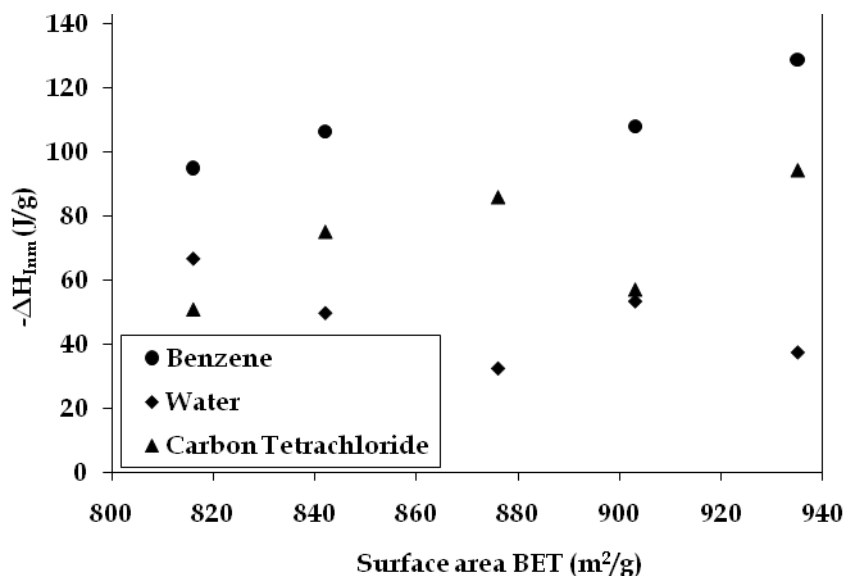


Fig. 13. Enthalpies of immersion in Benzene, Carbon Tetrachloride and water in terms of BET area

On the other hand, the difference in the enthalpies of immersion in water of different samples indicates the change in surface chemistry (Giraldo & Moreno-Piraján. 2008; López-Ramón *et al.* 2000), as a result of the different treatments that samples underwent, that is, the development or removal of surface groups on the surface of the solid, thus, a greater amount of oxygenated surface groups as in GACox's case which leads to a bigger enthalpy of immersion, as a consequence of the interactions established between the polar molecule as is the water molecule and oxygen surface groups developed in the sample, which is consistent with the chemical characterization, these groups were mostly acid type, specifically carboxyl groups. It is also observed that in thermally treated samples decreased enthalpies of immersion in water due to the selective decomposition of the groups present on the surface and therefore a decrease in specific interactions with the water molecule. Additionally, it is possible to conclude that the interactions of water does not occur exclusively with surface groups of the different samples because the sample CAG1173 in which one would expect to have a minimum amount of oxygenated surface groups, also has an calorimetric effect attributed to interactions dispersive type and non-specific type. As for the hydrophobic character of the surface is found that this decrease with the oxidation process and gradually increases with the heat treatments, being higher in the sample treated at 1173 K. Figures 14 and 15 shows the typical thermograms, obtained in the immersion of a solid in the different liquids used. These two figures were chosen because you can see the difference in magnitude of the peaks corresponding to each liquid for to the most oxidized sample (GACoxN) and sample treated to highest temperature in an atmosphere inert (GAC1173).

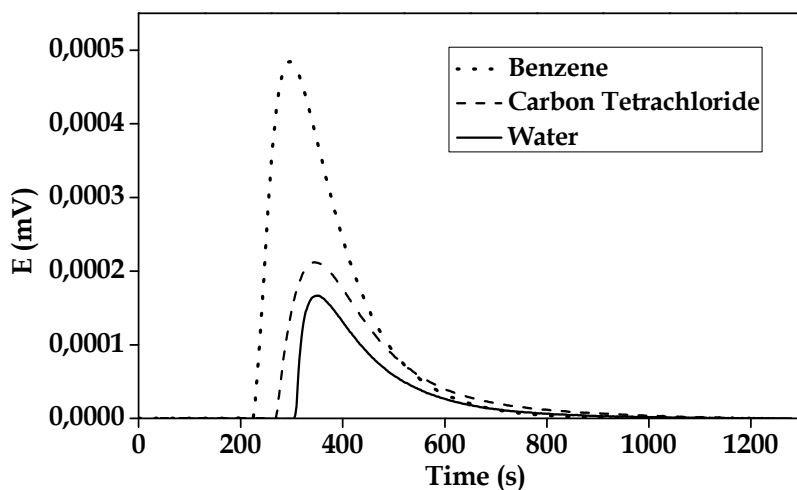


Fig. 14. Thermogram of Immersion Calorimetry in Benzene, Carbon Tetrachloride and water of GAC1173 sample

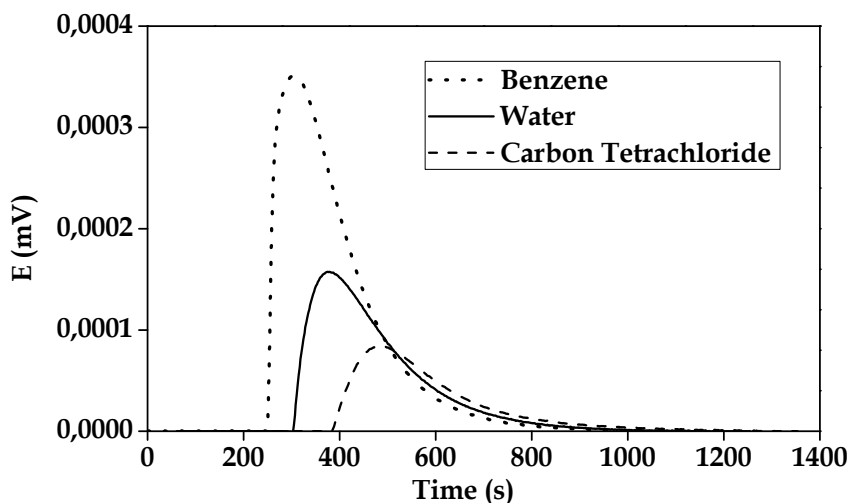


Fig. 15. Thermogram of Immersion Calorimetry in Benzene, Carbon Tetrachloride and water of GACoxN1173 sample

Finally, the samples were used for the removal of nickel from aqueous solution, for this, 0,500 g of each sample were put in contact with 50ml of the nickel solution of concentrations from 100 to 500 mg /L, initial pH of the mixture was adjusted to 6, taking into account that in this pH is nickel is found as Ni (II). The experimental data obtained in the adsorption process were adjusted to the Redlich-Peterson model and are shown in Figure 16.

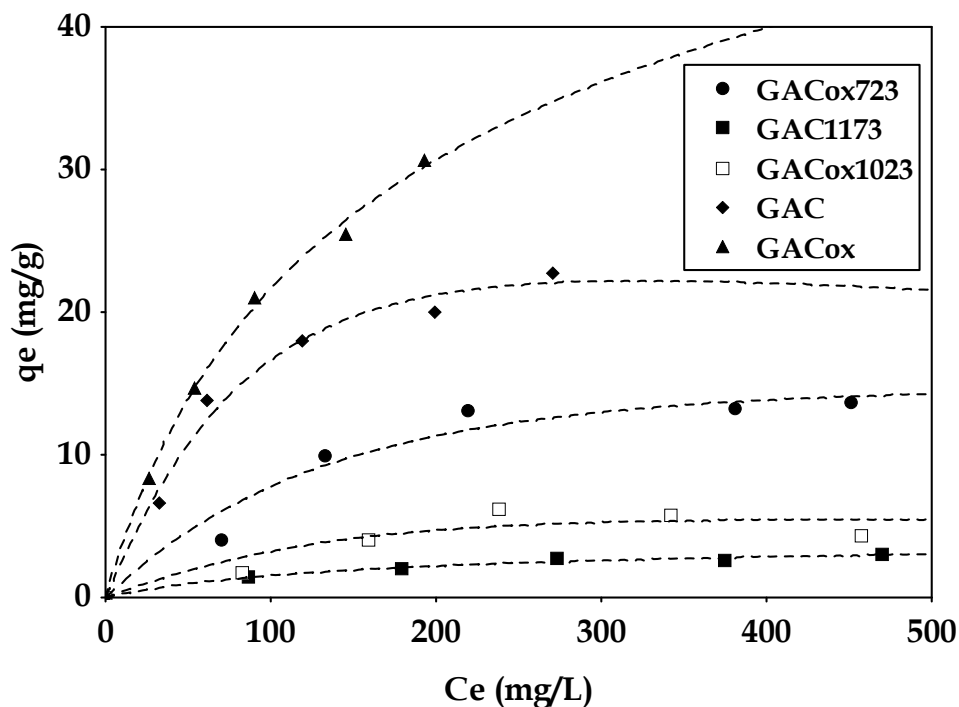


Fig. 16. Adsorption isotherm of Nickel on different samples fit the Redlich-Peterson model.

The importance of the role of oxygenated groups on the activated carbon surface in the adsorption process of ions from aqueous solution has been highlighted by many authors (Puziy *et al.* 2002). It is generally considered that the removal of an ion is mainly attributed to the interaction of surface groups and the ion, through various mechanisms, such as: formation of metal complexes like COOH-M and / or donor-acceptor reactions of electrons (Petit *et al.* 2010; Moreno-Castilla *et al.* 2010), that is, by establishing specific interactions, therefore, these mechanisms are favored when the solid undergoes an oxidation process as in the case of the sample GACoxN, which has a higher adsorption capacity with respect to other ones, this capacity decreased in heat-treated samples after the process oxidation ratifying the importance of the presence of oxygenated surface groups, although it is important to note that the adsorption capacity of GAC1173 sample is lower, it is also suggested to contemplate within the mechanisms of adsorption interactions that are not only specific but also of the dispersive type, to a lesser extent but to complement the adsorption process.

#### *Activate carbon for the adsorption of phenol*

Among various industrial waste to generate contamination there are tires, this waste is a difficult material to degrade and handle due to its physicochemical composition, generating a problem of global nature. Therefore alternatives different have been proposed for reuse, among these are energy production through incineration, combustion, and pyrolysis

processes. (Nadem et al. 2001) Another alternative that being studied at present is the production of activated carbon from this waste, there by creating a double benefit for the environment.

A study was conducted about to granular activated carbon adsorbents prepared from tires. To this end, the tires were cut into pieces with a size of 10 mm thick, two samples were treated with phosphoric acid at 20 and 40% p/p (TCP20 and TCP40) and other samples were treated with potassium hydroxide to the same concentrations (TCK20 And TCK40), then underwent to a carbonization process in a horizontal furnace at 1123 K for 2 hours. In this way is prepare by physical activation with CO<sub>2</sub>, samples were subjected to a pyrolysis process with N<sub>2</sub> at 923 K, and then activation with CO<sub>2</sub> at two temperatures 1123 K and 1223 K (TCCO<sub>2</sub>-1123 and TCCO<sub>2</sub>-1223) during 2 hours. All samples were characterized by N<sub>2</sub> adsorption at 77 K and immersion calorimetry in benzene. Some of the results obtain are compiled in Table 9.

Samples	SBET (m <sup>2</sup> /g)	Vo DR (cm <sup>3</sup> /g)	Eo (KJ/mol)	-ΔH <sub>imm</sub> C <sub>6</sub> H <sub>6</sub> (J/g)
TCP20	71.17	0.018	13.59	5.670
TCP40	52.85	0.013	13.62	3.120
TCK20	149.2	0.068	19.95	35.55
TCK40	157.3	0.070	19.63	26.17
TCCO <sub>2</sub> -850	25.44	0.009	14.23	11.48
TCCO <sub>2</sub> -950	58.12	0.020	14.92	21.36

Table 9. Characteristics of the samples obtained by gas adsorption and immersion calorimetry

Figure 17 shows the adsorption isotherms for three coals of the series prepared. the TCK40 sample presents a type I isotherm according to IUPAC classification, feature of microporous carbons, the samples and TCCO<sub>2</sub>-1223 and TCP20 have type II isotherms characteristic of mesoporous carbon, where adsorption occurs in open surface with the formation of multilayers. (Martín-Martínez 1988)

Immersion calorimetry as mentioned along this chapter, allow complement the characterization of porous materials. Figure 18 shows the thermograms obtained for the immersion of the samples in benzene, which is a liquid of wet to assess the area accessible to the molecule. It is observed that the highest enthalpies are obtained for samples prepared with sodium hydroxide which is consistent with the surface areas of these samples. By contrast the samples activated with phosphoric acid have low values compared with those activated with CO<sub>2</sub>, probably due to the presence of phosphorus compounds in activated carbon, which prevents the access of the benzene molecule. (Marsh & Rodríguez-Reinoso. 2007)

From Dubinin Radushkevich equation was calculated pore volume and the characteristic energy for the samples.

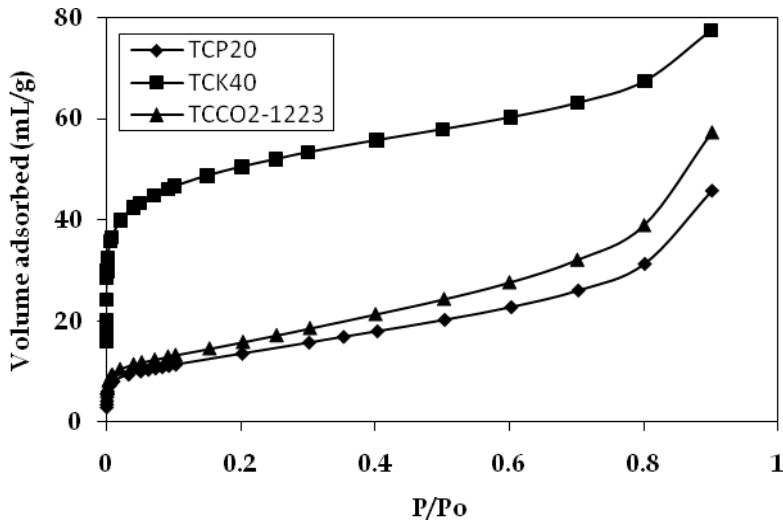


Fig. 17. Isotherms of  $N_2$  of the samples TCP20, TCK40 y TCCO2-1223

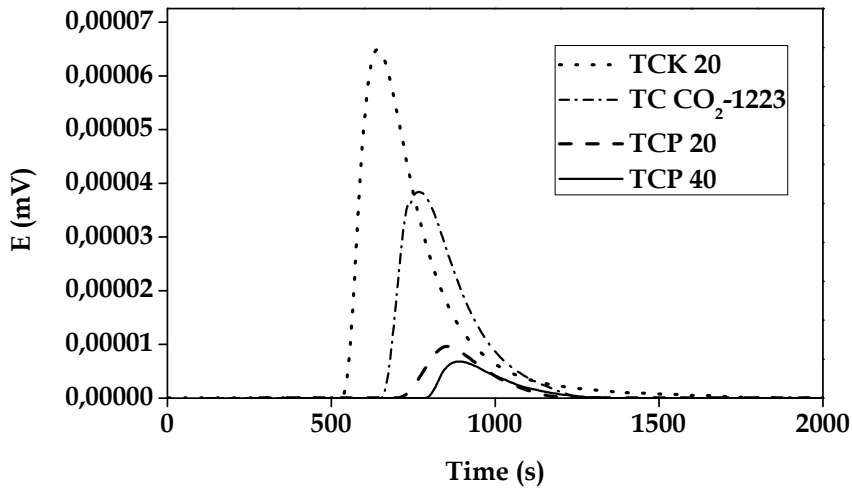


Fig. 18. Thermograms obtained for the samples.

$$\ln V = \ln V_o - D \left( \ln \frac{P_o}{P} \right)^2 \quad (24)$$

Where  $V$  is the volume adsorbed at certain pressures,  $P/P_o$  is the partial pressure,  $V_o$  is the micropore volume and  $D$  is a constant. **Bansal et al. 1988**

Figure 19 shows the graphs of DR for the samples TCK40 and TCK20 where you can see that a deviation from linearity near the saturation pressure, explaining that he has a multilayer adsorption and capillary condensation in mesopores (Martín-Martínez 1988) consistent with the isotherms of N<sub>2</sub> obtained which present a mesoporosity of activated carbons.

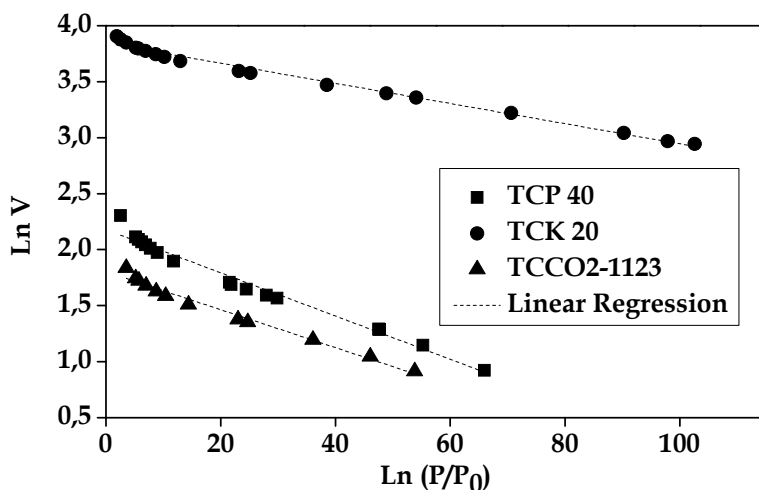


Fig. 19. The graphs of DR for the samples. Organized as follows way TCP40-TCK20-TCCO2-1123.

From of the constant D to be the slope of the graph permit to calculate the characteristic energy of adsorption given by the following equation:

$$\ln V = \ln V_o - D \left( \ln \frac{P_o}{P} \right)^2 \quad (24)$$

$$E_o = \frac{RT}{\sqrt{D}\beta} \quad (25)$$

Where R is the gas constant, 80414J/mol, T is the critical temperature of liquide nitrogen, 77K and  $\beta$  is the affinity coefficient of nitrogen, 0.34.

Stoeckli and Krahenbuehl were the first to correlate the enthalpy of benzene with the microporous parameters, in this work is realized this correlation but for mesoporous carbons. Figure 20 relates the characteristic energy of the nitrogen molecule with the enthalpy of immersion of the benzene molecule. In the activated carbons with carbon dioxide and potassium hydroxide gives a higher characteristic energy to a higher enthalpy of immersion, unlike those activated with phosphoric acid, which has a decrease in enthalpy with increasing energy feature, this behavior shows proportionality existing between the enthalpy of benzene and energy characteristic of N<sub>2</sub>, despite being two different methods to perform. Samples with higher BET surface area have a higher enthalpy of immersion in benzene which is the expected behavior because it has a greater surface arranged to interact with benzene (Silvestre-Albero et al.2004).

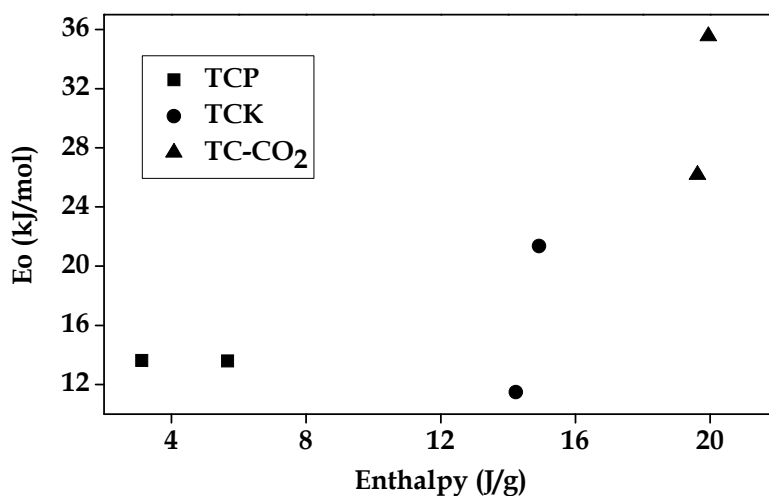


Fig. 20. Relación entre la entalpía de inmersión y la energía característica del nitrógeno.

*Activated Carbon monoliths for CO<sub>2</sub> adsorption*

Taking into account the interest they have taken the activated carbon monoliths in recent years, and its potential use in gas adsorption, is being developed in the research group work which seeks to make a contribution to knowledge of chemistry of solid adsorbents through the preparation, characterization and functionalization of carbon materials granular and monolithic type contribute to the study of the process of adsorption / gas capture a high environmental interest such as carbon dioxide. The CO<sub>2</sub> adsorption, has been studied as a way to retain the gas and check for interactions and conditions that govern the process to be more efficient and better use, the problem with CO<sub>2</sub> is not simple reason that alternatives are sought treatment with new materials, which will open avenues and possibilities according as knowledge of processes such as adsorption is broader.

As a preliminary approach to the preparation of carbonaceous materials of potential interest in the CO<sub>2</sub> adsorption, has been carried out the preparation of activated carbon monoliths disk type take taking advantage of two materials source lignocellulosic s generated as waste in large quantities in Colombia; coconut shell (samples COD) and African palm stone (samples CUD), the endocarps of these precursors are impregnated with H<sub>3</sub>PO<sub>4</sub> solutions at different concentrations for a period of 2 hours at 358K, then take a uniaxial press, where the shaping is done by pressing at 423 K, structures are then carbonized in a horizontal furnace at a linear heating rate of 1 Kmin<sup>-1</sup> to a temperature of 723K remaining there 2 hours. Finally, the monoliths obtained are washed with hot distilled water until neutral pH to remove any traces of chemical agent used in the impregnation (Rodríguez-Reinoso et al., 2004, Vargas et al., 2010).

Subsequently, textural, chemical and energy characterization of monoliths is performed to establish their behavior. The adsorption isotherms of N<sub>2</sub> at 77K and CO<sub>2</sub> at 273 K are determined, the experimental data fit the Langmuir model, and further immersion calorimetry in benzene are performed (0.37 nm) to establish energy correlations.



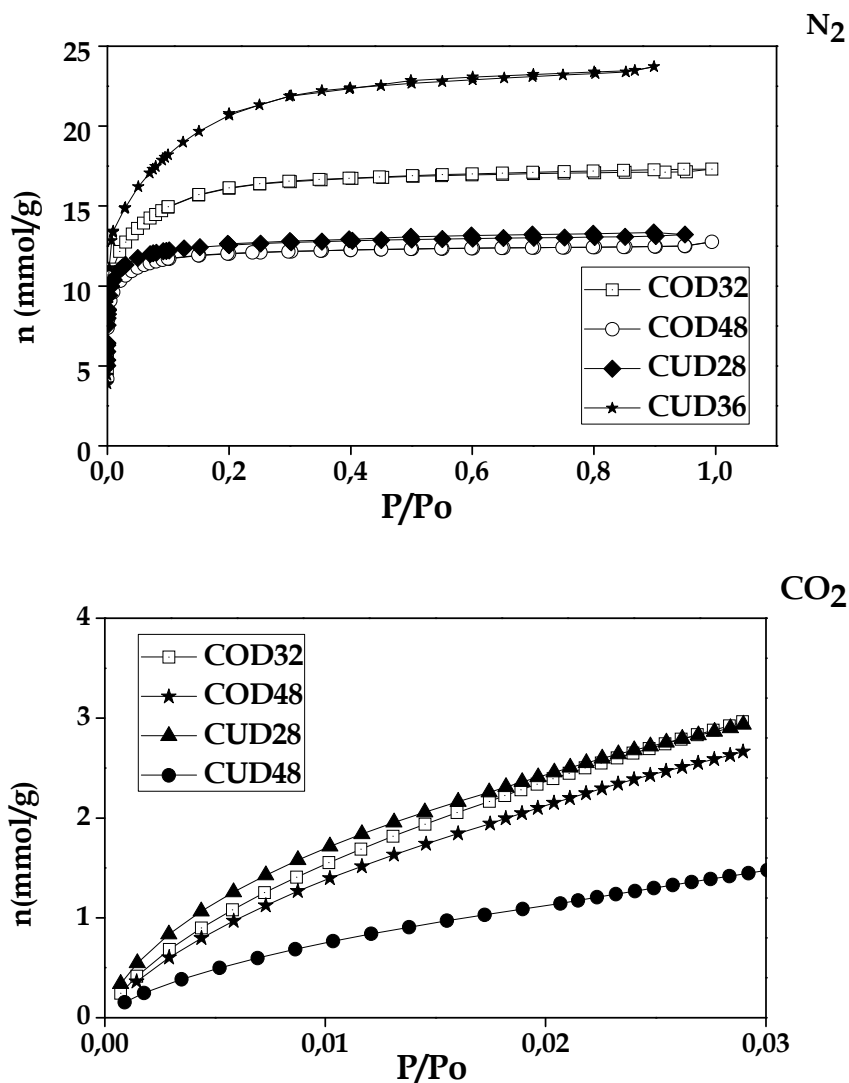


Fig. 21. Nitrogen adsorption isotherms at 77K and CO<sub>2</sub> at 273K for the monoliths with high and low adsorption capacity, with each precursor.

Some of nitrogen and carbon dioxide isotherms obtained for disks are shown in Figure 22, is evidence the obtaining of microporous solids fact is justified by the form type I isotherms, these solids have a surface area between 975 and 1711 m<sup>2</sup>g<sup>-1</sup> and n<sub>0</sub> between 11.49 and 18.02 mmol, experimental results indicated that the monoliths prepared from African palm stone have higher adsorption capacity and therefore a larger surface area, further shows that the change in the concentration of H<sub>3</sub>PO<sub>4</sub> produces a greater effect on the textural characteristics of samples CUD compared with the COD.

The obtained carbon monoliths were tested as potential adsorbents for CO<sub>2</sub> finding a retention capacity between 88-164 mgCO<sub>2</sub>g<sup>-1</sup> at 273K and atmospheric pressure, in Figure 22 to observe the isotherms of the samples with higher and lower CO<sub>2</sub> adsorption capacity in each series, the monoliths with a better performance in the retention of this gas were COD32 and CUD28.

The table 10 compiles the characteristics of the carbon monoliths prepared, show the data obtained for the interaction of three molecules of interest in the characterization of materials. Additionally, adsorption data were used for the calculation of three parameters:  $n_{oDR}$ ,  $n_{mL}$ ,  $K_L$  which are measures of the adsorption capacity.

Sample	N <sub>2</sub>		CO <sub>2</sub>				C <sub>6</sub> H <sub>6</sub>	
	S <sub>BET</sub> (m <sup>2</sup> /g)	n <sub>o</sub>	n <sub>o</sub>	n <sub>m</sub>	K	E <sub>o</sub> (KJ/mol)	-ΔH <sub>imm</sub> (J/g)	E <sub>o</sub> (KJ/mol)
COD28	1270	14.19	4.88	6.95	0.029	16.01	130	20.90
COD32	1320	13.86	5.10	6.64	0.031	16.87	147	24.03
COD36	1318	14.15	4.91	6.56	0.035	16.80	132	21.33
COD48	975	11.49	4.75	4.75	0.055	18.58	112	22.43
CUD28	1013	12.12	4.93	5.36	0.054	19.12	123	21.47
CUD32	1397	13.35	4.38	6.87	0.028	16.76	130	21.12
CUD36	1711	18.02	2.92	4.53	0.027	16.85	120	14.80
CUD48	1706	18.65	2.36	3.99	0.025	17.63	96	11.48

Table 10. Characteristics of carbon monoliths.

Figure 22 shows the relationship between the number of moles of the monolayer determined by two different models,  $n_m$  by the Langmuir model and  $n_o$  calculated from Dubinin Raduskevich, shows that the data are a tendency for both precursors although they are calculated from models with different considerations. There are two points that fall outside the general trend CUD28 and COD32 samples, which despite having the highest value of  $n_o$  in each series not have the highest  $n_m$ .

The Dubinin Raduskevich equation is use to determinate, the characteristic adsorption energies of N<sub>2</sub> and CO<sub>2</sub> (E<sub>o</sub>) for each samples, likewise by the Stoeckli y Krahenbüehl equation (equation 14) was determined benzene (E<sub>o</sub>), in Figure 23 shows the relationship between the characteristic energies determined by two different characterization techniques and found two trends in the data which shows the heterogeneity of carbonaceous surfaces of the prepared samples. The characteristic energy of CO<sub>2</sub> adsorption, is lower in almost all the monoliths compared to E<sub>o</sub> of immersion in benzene, this is consistent considering that due to the size of the CO<sub>2</sub> molecule 0.33 nm, this can be accessed easily to narrow pores,

while benzene has a size of 0.37 nm for slit-shape pores and 0.56 nm for cylindrical restricts its accessibility and generates an increase in  $E_o$ . In Figure 19a shows that the COD samples show a trend, except COD32 which again leaves the general behavior, this can be attributed to the monolith has a narrow micropores limits the interaction with the benzene molecule, generating a higher  $E_o$ .

In the case of samples CUD48 and CUD36 which present a larger surface area, there is a greater more  $CO_2$   $E_o$  compared to benzene  $E_o$ , in these samples increased the concentration of chemical agent degrades carbonaceous matrix producing a widening pore that provides access to benzene and leads to a decrease in  $E_o$ .

Figure 24 relates the characteristic adsorption energy in benzene with the immersion enthalpy in this molecule, can be observed for most samples an increase of the immersion enthalpy with the characteristic energy of the process, which is consistent since the characteristic energy is a measure of the magnitude of the interaction between the solid and the adsorbate is ratified with the increase of enthalpy value.

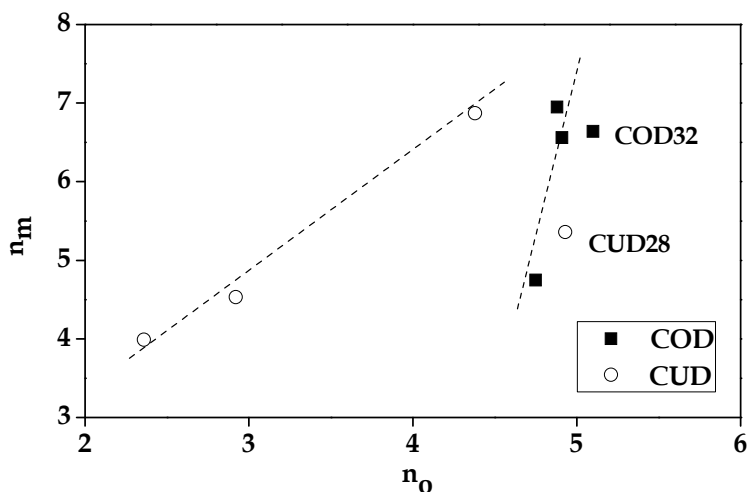


Fig. 22. Relationship between  $n_m$  and  $n_o$  samples of each series.

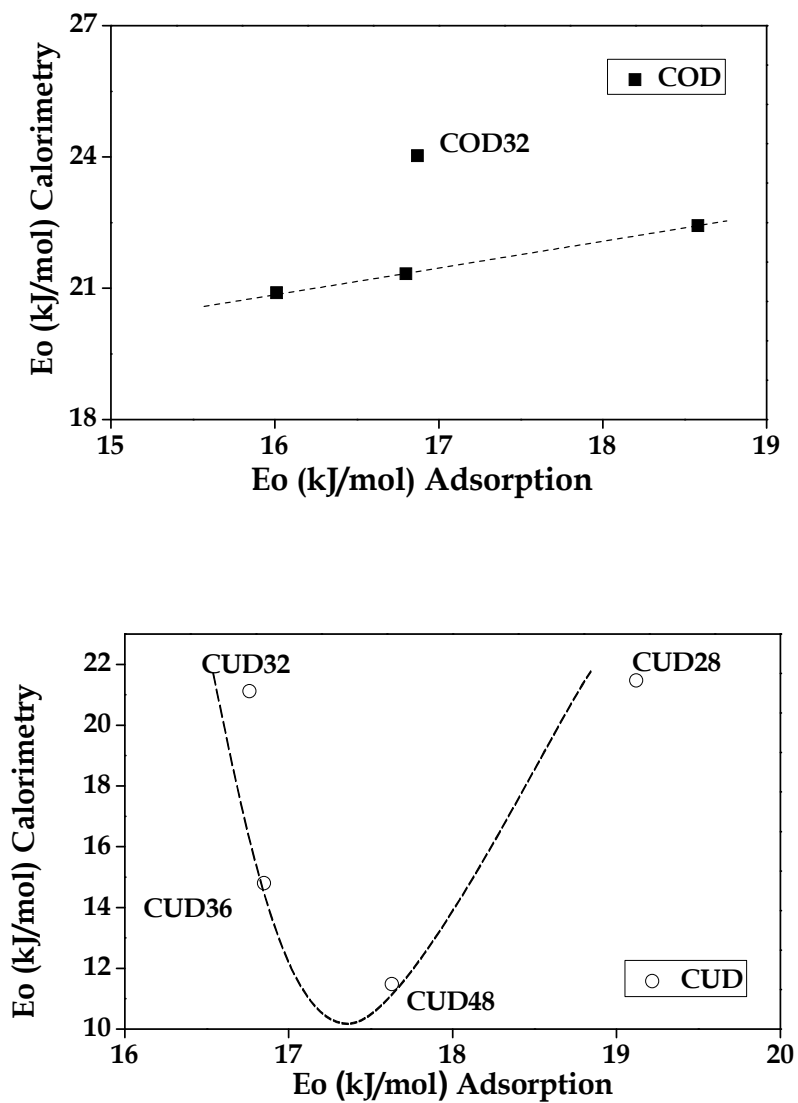


Fig. 23. Relationship between the characteristic immersion energy of benzene and the characteristic adsorption energy of  $\text{CO}_2$ .

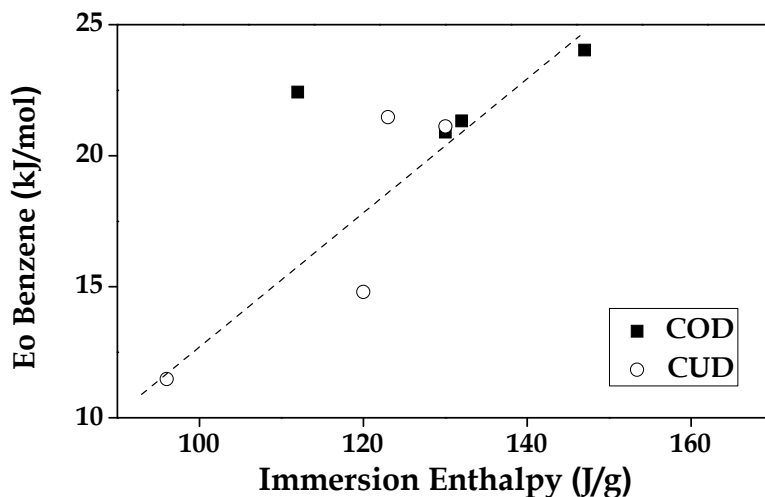
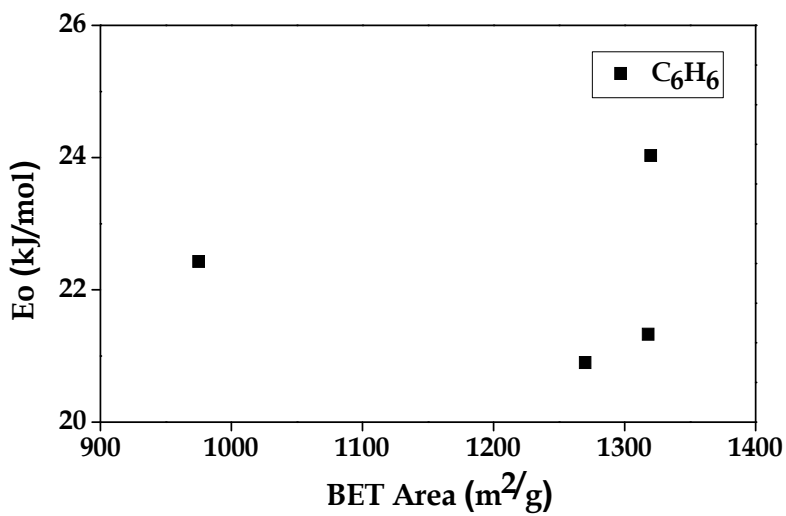
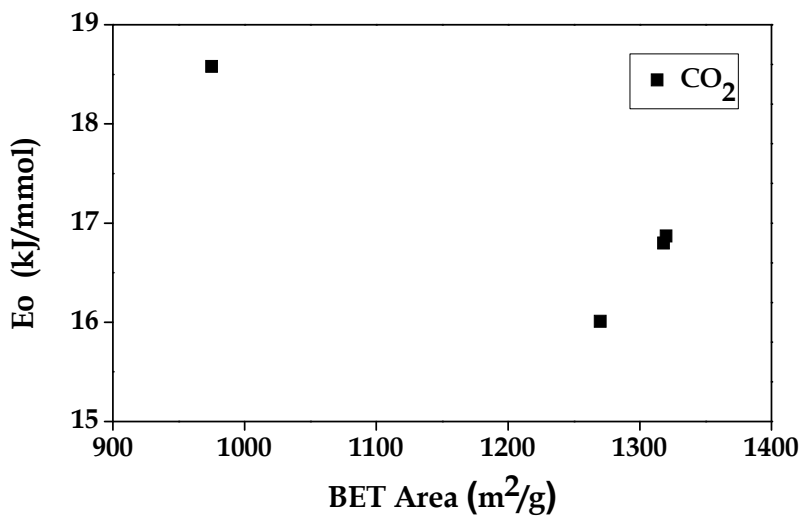


Fig. 24. Relationship between the characteristic adsorption energy in benzene and the immersion enthalpy.

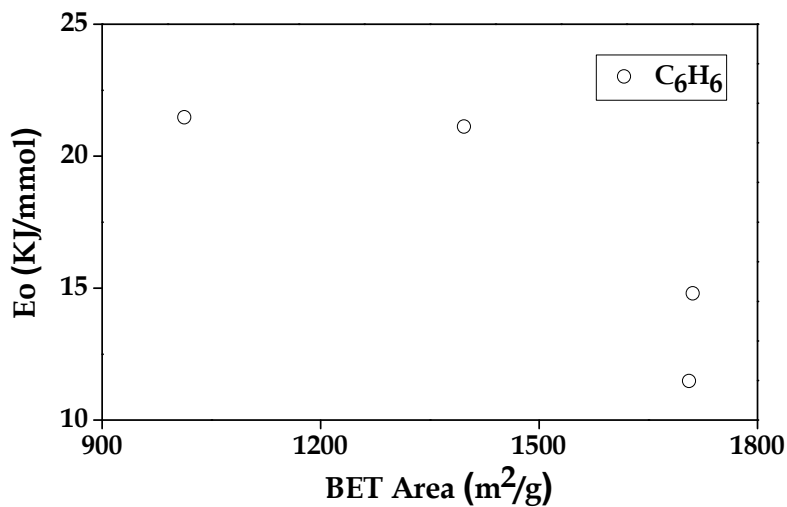
Additionally, establishing correlations between energetic parameters determined by different models and textural characteristics, figure 25 a) and b) show the relationship between the characteristic energy and BET area of the COD samples, different behaviors can be observed for each molecule, in the case characteristic adsorption energy of benzene shows a decrease with increasing area of the discs for samples COD28, COD48, but there was an increase in the COD36 and COD32 samples with higher values for surface area. To CUD, as shown in Figure 25 c) and d) in the case of benzene adsorption, for all samples shows a decrease in  $E_o$ . The characteristic adsorption energy carbon dioxide molecule shows a decrease with increasing the BET area, for COD32, COD36 there is a slight increase in  $E_o$  attributed to these samples have more narrow micropores that can be seen in the value of  $n_0$ ,  $CO_2$ . A similar trend shows the CUD discs; the decrease in the characteristic energy with increasing surface area of the monoliths is related to the increased amount of mesopores in the material, since the adsorption energy decreases with increasing pore size (Stoekli et al., 1989).



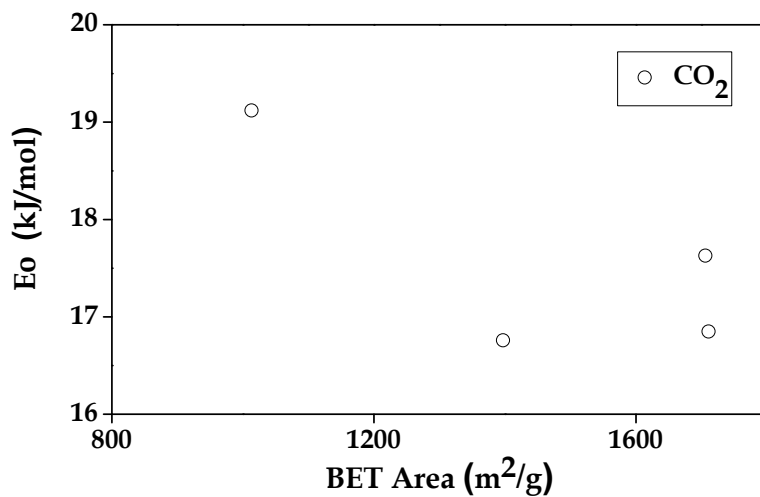
a)



b)



c)



d)

Fig. 25. Relationship between the characteristic energy and BET area of the series. a,b) COD. c,d) CUD.

## 2. References

- Aydin, Gokhan., Karakurt, Izzet., Aydiner, Kerim. (2010). Evaluation of geologic storage options of CO<sub>2</sub>: Applicability, cost, storage capacity and safety. *Energ Policy*, Vol. 38, No. 9, pp. 5072-5080.
- Bansal R. C., Goyal M. (2005). Activated carbon adsorption. Taylor & Francis Group. ISBN: 0-8247-5344-5 New York, United States. pp.487 .
- Bansal, R.C., Donnet, J. B., Stoeckli, H. F. (1988). Active Carbon, Marcel Dekker, ISBN 92-828-5537-6, New York, United States.
- Blanco, D., Giraldo, L., Moreno, J. C. (2008). Resorcinol adsorption from aqueous solution on activated carbon. Relation adsorption isotherm and immersion enthalpy, *Revista Colombiana de Química*, Vol. 37, No. 2, pp. 203-216.
- Blanco, J., Ávila, P., Martín, M.P. (2000). Honeycomb monoliths of activated carbons for effluent gas purification. *Micropor. Mesopor. Mat.* Vol. 37, No. 1-2, pp. 201-08.
- Callister, W. D. (2007). Introducción a la ciencia e ingeniería de los materiales, Editorial Reverté, Vol. 1, Tercera edición, ISBN 978-84-291-7252-2, Barcelona, España.
- Carlson, Annelie. (2003). Energy systems and the climate dilemma Reflecting the impact on CO<sub>2</sub> emissions by reconstructing regional energy systems, *Energ Policy*, Vol.31, No. 10, pp. 951-959.
- Chingombe, P., Saha, B., Wakeman, R.J. (2005) Surface modification and characterisation of a coal-based activated carbon. *Carbon*. Vol.43, No. 15, pp 3132-3143.
- Daud, W., Houshamnd, A.H. (2010) Textural characteristics, surface chemistry and oxidation of activated carbon. *Journal of Natural Gas Chemistry*. Vol. 19, No.3, pp 267-279
- Denoyel, R., Fernandez-Colinas, J., Grillet, Y., Rouquerol, J. (1993). Assessment of the surface area and microporosity of activated charcoals from immersion calorimetry and nitrogen adsorption data. *Langmuir*, Vol. 9, No. 2, pp.515-518.
- Deydier, E., Guilet, R., Sarda, S., Sharrock, P. (2005). Physical and chemical characterisation of crude meat and bone meal combustion residue: “waste or raw material?”, *J. Hazard. Mater.*, Vol. 121, pp. 141-148.
- Dias, J.M., Alvim-Ferraza, M.C.M., Almeida, M.F., Rivera-Utrilla, J., Sánchez-Polo, M. (2007) Waste materials for activated carbon preparation and its use in aqueous-phase treatment: A review. *Journal of Environmental Management*. Vol. 85, No. 4, pp. 833-846
- Dubin, M. M. (1975). Progress in Surface and Membrane Science, Academic Press, Edited by Cadenhead D. A., Vol. 9, ISBN 0-12-571809-8, New York, United States.
- Figueiredo, J.L., Pereira, M.F.R. (2010) The role of surface chemistry in catalysis with carbons *Catalysis Today* Vol. 150, No. 1-2, 26 pp 2-7
- Figueiredo, J.L., Pereira, M.F.R., Freitas, M.M.A., Órfão, J.J.M. (1999). Modification of the surface chemistry of activated carbons. *Carbon*, Vol. 37, No. 9, pp. 1379-1389
- Gao, Z., Bandosz, T.J., Zhao, Z., Hand, M., Qiu, J. (2009) Investigation of factors affecting adsorption of transition metals on oxidized carbon nanotubes. *Journal of Hazardous Materials*. Vol.167, No. 1-3, pp 357-365.
- García-Cuello, V., Moreno-Piraján, J. C., Giraldo-Gutiérrez, L., Sapag, K., Zgrablich, G. (2008). Determination of Differential Enthalpy and Isotherm by Adsorption Calorimetry. *R. Lett. Phys. Chem.*, 4 pp.



- García-Cuello, V., Moreno-Piraján, J. C., Giraldo-Gutiérrez, L., Sapag, K., Zgrablich, G. (2009). A new microcalorimeter of adsorption for the determination of differential enthalpies. *Micropor. Mesopor. Mat.* Vol. 120 No. 1, pp. 239-245.
- Giraldo, L., García-Cuello, V., Moreno-Piraján, J. C. (2008). Caracterización superficial en fase gas y líquida de carbones activados. *Rev.Ing.Universidad de los Andes*, Vol. 27, No. 1, pp. 7-16.
- Giraldo, L., Moreno, J.C. (2008) Pb(II) and Cr(VI) adsorption from aqueous solution on activated carbons obtained from sugar cane husk and sawdust. *Journal of Analytical and Applied Pyrolysis*, Vol. 81, No. 2, pp 278-284.
- Hemminger, W. & Höhne, G. (1984). *Calorimetry: Fundamentals and Practice*, Verlag Chemie, ISBN 2-87352-027-2, Florida, United States.
- Jiménez, O. (2000). Preparación de carbón activado a partir de distintos tipos de material vegetal colombiano de desechos con posible aplicación en la remoción de fenol en agua residuales, Trabajo de Grado. Departamento de Química. Universidad Nacional de Colombia. Bogotá, Colombia.
- Kraus, G. (1955). The heat of immersion of carbon black in water, methanol and n-hexane. *J. phys. Chem.* Vol. 59, No. 4, pp. 343-345.
- Llewellyn, P. L. & Maurin, G. (2005). Gas adsorption microcalorimetry and modelling to characterize zeolites and related materials, *C. R. Chim.*, Vol. 8 No. 1, pp. 283-302.
- López-Ramón, M.V., Stoeckli, F., Moreno-Castilla, C., Carrasco-Marín, F. (2000) Specific and non-specific interactions of water molecules with carbon surfaces from immersion calorimetry. *Carbon*. Vol. 38, No. 6, pp.825-829.
- Marsh, H., Rodríguez-Reinoso, F. (2007) *Encyclopedia of Materials: Science and Technology*. Activated Carbon. 542 p.
- Marsh, Harry., Rodríguez-Reinoso, Francisco. (2006). Activated Carbon (Origins). En: *Activated carbon*, Elsevier Science Ltd ISBN: 0080444636, pp. 13-81.
- Martin Matinez, J.M. (1988). Adsorción física de gases y vapores por carbones . Universidad de Alicante. 465 p. ISBN: 84-86809-33-9. Edita: Secretariado de Publicaciones de la Universidad de Alicante. Alicante, España.
- Menéndez, J. A. (1998). On the use of calorimetric techniques for the characterization of carbons: A brief review. *Thermochim. Acta*, Vol. 312, No. 1, pp. 79 - 86.
- Molina-Sabio, M., Nakagawa, Y., Rodríguez-Reinoso, F. (2008). Possible errors in microporosity in chemically activated carbon deduced from immersion calorimetry. *Carbon*, Vol. 46, No. 1, pp. 329-334.
- Montes-Morán, M.A., Suárez, D., Menéndez, J.A., Fuente, E.(2004) On the nature of basic sites on carbon surfaces: an overview. *Carbon*. Vol. 42, No. 7, pp.1219-1225.
- Moreno, J. C. & Giraldo, L. (2005). Setups for simultaneous measurement of isotherms and adsorption heats. *Rev. Sci. Instrum.* Vol. 76, No. 1, pp. 1-8.
- Moreno, J. C. (1996). Microcalorímetro de conducción de calor: diseño, construcción, calibración y su aplicación al estudio de sólidos porosos. Tesis doctoral, Universidad Nacional de Colombia, Bogotá, Colombia.
- Moreno, J.C., Gómez, R., Giraldo, L. (2010). Removal of Mn, Fe, Ni and Cu Ions from Wastewater Using Cow Bone Charcoal, *Materials*, Vol. 3, pp. 452-466.

- Moreno-Castilla, C., López-Ramos, M.V. (2007). Adsorción de compuestos orgánicos disueltos en agua sobre carbones activados. En: *Sólidos Porosos: preparación, caracterización y aplicaciones*, Ediciones Uniandes, Editor académico, Juan Carlos Moreno-Piraján, ISBN 978-958-695-276-7, Bogotá, Colombia, pp. 281-299.
- Moreno-Castilla, C., Álvarez-Merino, M.A., Pastrana-Martínez, L.M., López-Ramón, M.V. (2010) Adsorption mechanisms of metal cations from water on an oxidized carbon surface. *Journal of Colloid and Interface Science*. Vol. 345, No. 2, pp. 461-466.
- Moreno-Piraján, J. C. & Giraldo, L. (2007). Instrumentación calorimétrica aplicada a la determinación de entalpías de inmersión de sólidos porosos. En: *Sólidos Porosos: preparación, caracterización y aplicaciones*, Ediciones Uniandes, Editor académico, Juan Carlos Moreno-Piraján, ISBN 978-958-695-276-7, Bogotá, Colombia, pp. 281-299.
- Moreno-Piraján, J.C., Navarrete, L., Giraldo, L., García, V. (2007). Adsorption of Phenol and 3 Chlorophenol on Activated Carbons using Immersion Calorimetry, *Información Tecnológica*, Vol. 18 , No. 3, pp. 71-80.
- Nadhemi, K. H., Xiao, D. C., Mohammed, M. F. (2001). Adsorption kinetics for the removal of chromium(VI) from aqueous solution by adsorbents derived from used tyres and sawdust. *Chemical Engineering Journal*, Vol 84, pp. 95-105.
- Nakagawa, Y., Molina-Sabio, M., Rodríguez-Reinoso, F. (2007). Modification of the porous structure along the preparation of activated carbon monoliths with  $H_3PO_4$  and  $ZnCl_2$ . *Micropor. Mesopor. Mat.* Vol. 103, No. 1-3, pp. 29-34.
- Oke, I.A., Olarinoye, N.O., Adewusi, S.R.A. (2008). Adsorption kinetics for arsenic removal from aqueous solutions by untreated powdered eggshell. *Adsorption*, Vol. 14, pp. 73-83.
- Petit, C., Peterson, G.W., Mahle, J., Bandosz, T.J. (2010) The effect of oxidation on the surface chemistry of sulfur-containing carbons and their arsine adsorption capacity. *Carbon*. Vol.48, No. 6, pp. 1779-1787.
- Puziy, A.M., Poddubnaya, O.I., Martínez-Alonso, A., Suárez-García, F., Tascón, J.M.D. (2002) Synthetic carbons activated with phosphoric acid: I. Surface chemistry and ion binding properties. *Carbon*. Vol.40, No.9, pp.1493-1505.
- Radovic, L.R., Moreno-Castilla, C., Rivera-Utrilla, J. (2000) Carbon Materials as Adsorbents in Aqueous Solutions. In *Chemistry and Physics of Carbon*. A Serie of Advances. Ed Marcel Dekker New York. p 293-297
- Rivera-Utrilla, J., Sánchez-Polo, M. (2003) Adsorption of Cr(III) on ozonised activated carbon. Importance of  $\pi-\pi$  cation interactions. *Water Research*. Vol. 37, No. 14, pp. 3335-3340.
- Rodríguez, F. J. (2003). Procesos de potabilización del agua e influencia del tratamiento de ozonización, Ediciones Díaz de Santos S.A, ISBN 978-84-7978-587-1, Madrid, España.
- Rodríguez-Reinoso, F. (2007). El carbón activado como adsorbente universal. En: *Sólidos Porosos: preparación, caracterización y aplicaciones*, Ediciones Uniandes, Editor académico, Juan Carlos Moreno Piraján, ISBN 978-958-695-276-7, Bogotá, Colombia, pp.1-42.

- Rodríguez-Reinoso, F., Almansa, C., Molina-Sabio, M. (2004). Adsorption of methane into ZnCl<sub>2</sub> activated carbon derived disc. *Micropor. Mesopor. Mat*, Vol. 76, No. 1-3, pp. 185-191.
- Rouquerol, F., Rouquerol, J., Sing, K. S. W. (1999). Adsorption powders and porous solids Principles, methods and applications, Academic Press, 1st edition, ISBN 0125989202, San Diego, London.
- Ruthven, D. M. (1984). Principles of Adsorption and Adsorption Processes, Wiley - Interscience, ISBN 0471866067, New York, United States.
- Silvestre-Albero A., Silvestre-Albero J., Sepúlveda-Escribano A., Rodríguez-Reinoso, F. (2009) Ethanol removal using activated carbon: Effect of porous structure and surface chemistry. *Microporous and Mesoporous Materials*. Vol.120, No. 1-2, pp. 62-68.
- Silvestre-Albero, J., Gómez de Salazar, C., Sepulveda-Escribano, A., Rodríguez-Reinoso, F. (2001). Characterization of microporous solids by immersion calorimetry. *Colloid and Surfaces A: Physicochem and Eng Aspect*, Vol. 187-188, pp.151-165.
- Smicklas, I., Dimovic, S., Sljivic, M., Plecas, I. (2008). The batch study of Sr<sup>2+</sup> sorption by bone char, *J. Environ. Sci. Health, Part A: Toxic/Hazard. Subst. Environ. Eng.*, Vol. 43, No. 2, pp. 210-217.
- Stoeckli, F., Ballerini, L., De Bernardini, S. (1989). On the evolution of micropore widths and areas in the course of activation. *Carbon*, Vol. 27, No. 3, pp. 501-502.
- Stoeckli, F., Moreno-Castilla, C., Carrasco-Marín, F., Lopez-Ramón, M.V. (2001) Distribution of surface oxygen complexes on activated carbons from immersion calorimetry, titration and temperature programmed desorption techniques. *Carbon*. Vol 39, No. 14, pp.2235-2237.
- Stoeckli, H. F. & Kraehenbühl F. (1989). Recent developments in the Dubinin equation. *Carbon*, Vol. 27, No.1, pp. 125-128
- Su-Hsia, L. & Ruey-Shin, J. (2009). Adsorption of phenol and its derivatives from water using synthetic resins and low-cost natural adsorbents: A review. *J. Environ. Manag.*, Vol. 90, No. 3, pp. 1336-1349.
- Szymański, G., Biniak, S., Rychlicki . G. (2002) Carbon surface polarity from immersion calorimetry. *Fuel processing Technology*. Vol.79, pp. 217-223
- Szymański, G., Karpiński, Z., Biniak S., Świątkowski, A. (2002) The effect of the gradual thermal decomposition of surface oxygen species on the chemical and catalytic properties of oxidized activated carbon. *Carbon*. Vol.40, No.14, pp.2627-2639
- Tae Young, K., Seung Jai, K., Sung Young, C. (2001). Effect of pH on Adsorption of 2,4-Dinitrophenol onto an Activated Carbon. *Korean J. Chem. Eng.*, Vol.18, No. 5, pp. 755-760.
- Vargas, D.P., Giraldo, L., Moreno-Piraján, J.C. (2010). Preparation and Characterization of Activated Carbon Monoliths with Potential Application as Phenol Adsorbents., *E-J. Chem*. No. 7, pp. 531-539.
- Yang, R. T. (1997). Gas Separation by Adsorption Processes, Imperial College Press, Butterworth, ISBN 9780471297413, London, UK.
- Yin, C.Y., Aroua, M.K., Daud, W. (2007) Review of modifications of activated carbon for enhancing contaminant uptakes from aqueous solutions *Separation and Purification Technology*. Vol 52, No.3, pp.403-415.

Zimmermann, W. & Keller, J. U. (2003). A new calorimeter for simultaneous measurement of isotherms and heats of adsorption. *Thermochim. Acta*, Vol. 403, No. 1, pp. 31 – 41.

# Thermodynamics of Interfaces

Omid Moradi

*Shahre-Qods Branch, Islamic Azad University,  
Iran*

## 1. Introduction

Thermodynamics is the branch of science that is concerned with the principles of energy transformation in *macroscopic* systems. Macroscopic properties of matter arise from the behavior of a very large number of molecules. Thermodynamics is based upon experiment and observation, summarized and generalized in the *Laws of Thermodynamics*. These laws are not derivable from any other principle: they are in fact improvable and therefore can be regarded as assumptions only; nevertheless their validity is accepted because exceptions have never been reported. These laws do not involve any postulates about atomic and molecular structure but are founded upon observation about the universe as it is, in terms of instrumental measurements. In order to represent the state of a gas or a liquid or a solid system, input data of average quantities such as temperature ( $T$ ), pressure ( $P$ ), volume ( $V$ ), and concentration ( $c$ ) are used. These averages reduce the enormous number of variables that one needs to start a discussion on the positions and momentums of billions of molecules. We use the thermodynamic variables to describe the state of a system, by forming a *state function*:

$$P=f(V, T, n) \quad (1)$$

This simply shows that there is a physical relationship between different quantities that one can measure in a gas system, so that gas pressure can be expressed as a function of gas volume, temperature and number of moles,  $n$ . In general, some relationships come from the specific properties of a material and some follow from physical laws that are independent of the material (such as the laws of thermodynamics). There are two different kinds of thermodynamic variables: *intensive variables* (those that do not depend on the size and amount of the system, like temperature, pressure, density, electrostatic potential, electric field, magnetic field and molar properties) and *extensive variables* (those that scale linearly with the size and amount of the system, like mass, volume, number of molecules, internal energy, enthalpy and entropy). Extensive variables are additive whereas intensive variables are not (Adamson, A.W. and Gast, A.P. 1997).

In thermodynamic terms, the object of a study is called the system, and the remainder of the universe, the surroundings. Amounts of the order of a mole of matter are typical in a system under consideration, although thermodynamics may remain applicable for considerably smaller quantities. The imaginary envelope, which encloses the system and separates it from its surroundings, is called the boundary of the system. This boundary may serve either to isolate the system from its surroundings, or to provide for interaction in specific ways

between the system and surroundings. In practice, if a reactor is used to carry out a chemical reaction, the walls of the reactor that are in contact with the thermo stated liquid medium around the reactor may be assumed to be the surroundings of the experimental system. For particles such as colloids, the medium in which they are immersed may act as the surroundings, provided nothing beyond this medium influences the particle. An isolated system is defined as a system to or from which there is no transport of matter and energy. When a system is isolated, it cannot be affected by its surroundings. The universe is assumed to be an isolated system. Nevertheless, changes may occur within the system that are detectable using measuring instruments such as thermometers, pressure gauges etc. However, such changes cannot continue indefinitely, and the system must eventually reach a final static condition of internal equilibrium. If a system is not isolated, its boundaries may permit exchange of matter or energy or both with its surroundings. A closed system is one for which only energy transfer is permitted, but no transfer of mass takes place across the boundaries, and the total mass of the system is constant. As an example, a gas confined in an impermeable cylinder under an impermeable piston is a closed system. For a closed system, this interacts with its surroundings; a final static condition may be reached such that the system is not only internally at equilibrium but also in external equilibrium with its surroundings. A system is in equilibrium if no further spontaneous changes take place at constant surroundings. Out of equilibrium, a system is under a certain stress, it is not relaxed, and it tends to equilibrate. However, in equilibrium, the system is fully relaxed. If a system is in equilibrium with its surroundings, its macroscopic properties are fixed, and the system can be defined as a given thermodynamic state. It should be noted that a thermodynamic state is completely different from a molecular state because only after the precise spatial distributions and velocities of all molecules present in a system are known can we define a molecular state of this system. An extremely large number of molecular states correspond to one thermodynamic state, and the application of statistical thermodynamics can form the link between them (Lyklema, J. 2005), (Dabrowski A., 2001).

## 2. Energy, work and heat

### 2.1 The first law of thermodynamics

Generally, when a system passes through a process it exchanges energy  $U$  with its environment. The energy change in the system  $\Delta U$  may result from performing work  $w$  on the system or letting the system perform work, and from exchanging heat  $q$  between the system and the environment

$$\Delta U = q + w \quad (2)$$

The heat and the work supplied to a system are withdrawn from the environment, such that, according to the first law of thermodynamics

$$\Delta U_{\text{system}} + \Delta U_{\text{environment}} = 0 \quad (3)$$

The First Law of thermodynamics states that the energy content of the universe (or any other isolated system) is constant. In other words, energy can neither be created nor annihilated. It implies the impossibility of designing a perpetuum mobile, a machine that performs work without the input of energy from the environment. The First Law also implies that for a system passing from initial state 1 to final state 2 the energy change

$\Delta_{1 \rightarrow 2} U$  does not depend on the path taken to go from 1 to 2. A direct consequence of that conclusion is that  $U$  is a function of state: when the macroscopic state of a system is fully specified with respect to composition, temperature, pressure, and so on (the so-called state variables), its energy is fixed. This is not the case for the exchanged heat and work. These quantities do depend on the path of the process. For an infinitesimal small change of the energy of the system

$$\Delta U = \partial q + \partial w \quad (4)$$

For  $w$  and, hence,  $\partial w$ , various types of work may be considered, such as mechanical work resulting from compression or expansion

of the system, electrical work, interfacial work associated with expanding or reducing the interfacial area between two phases, and chemical work due to the exchange of matter between system and environment. All types of work are expressed as  $\int X dY$ , where  $X$  and  $Y$  are state variables.  $X$  is an intensive property (independent of the extension of the system) and  $Y$  the corresponding extensive property (it scales with the extension of the system). Examples of such combinations of intensive and extensive properties are pressure  $p$  and volume  $V$ , interfacial tension  $\gamma$  and interfacial area  $A$ , electric potential  $\Psi$  and electric charge  $Q$ , the chemical potential  $\mu_i$  of component  $i$ , and the number of moles  $n_i$  of component  $i$ . As a rule,  $X$  varies with  $Y$  but for an infinitesimal small change of  $Y$ ,  $X$  is approximately constant. Hence, we may write

$$dU = \partial q - p dV + \gamma dA + \psi dQ + \sum_i \mu_i dn_i \quad (5)$$

The terms of type  $X dY$  in Eq. above represent mechanical (volume), interfacial, electric, and chemical works, respectively.  $\sum_i$  implies summation over all components in the system. It is obvious that for homogeneous systems the  $\gamma dA$  term is not relevant.

## 2.2 The second law of thermodynamics: entropy

According to the First Law of thermodynamics the energy content of the universe is constant. It follows that any change in the energy of a system is accompanied by an equal, but opposite, change in the energy of the environment. At first sight, this law of energy conservation seems to present good news: if the total amount of energy is kept constant why then should we be frugal in using it? The bad news is that all processes always go in a certain direction, a direction in which the energy that is available for performing work continuously decreases.

Entropy,  $S$ , is the central notion in the Second Law. The entropy of a system is a measure of the number of ways the energy can be stored in that system. In view of the foregoing, any spontaneous process goes along with an entropy increase in the universe that is,  $\Delta S > 0$ . If as a result of a process the entropy of a system decreases, the entropy of the environment must increase in order to satisfy the requirement  $\Delta S > 0$  (Levine, I.N., 1990).

Based on statistical mechanics, the entropy of a system, at constant  $U$  and  $V$  can be expressed by Boltzmann's law

$$S_{u,v} = k_B \ln w \quad (6)$$

where  $w$  is the number of states accessible to the system and  $k_B$  is Boltzmann's constant. For a given state  $w$  is fixed and, hence, so is  $S$ . It follows that  $S$  is a function of state. It furthermore follows that  $S$  is an extensive property: for a system comprising two subsystems (a and b)  $w = w_a + w_b$  and therefore, because of,  $S = S_a + S_b$ . The entropy change in a system undergoing a process  $1 \rightarrow 2$  is thermodynamically formulated in terms of the heat  $\delta q$  taken up by that system and the temperature  $T$  at which the heat uptake occurs (Sraelachvili, J. 1991):

$$\Delta S \geq \int_1^2 \frac{\delta q}{T} \quad (7)$$

Because the temperature may change during the heat transfer is written in differential form (Pitzer, K.S. and Brewer L. 1961).

### 2.3 Reversible processes

In contrast to the entropy, heat is not a function of state. For the heat change it matters whether a process  $1 \rightarrow 2$  is carried out reversibly or irreversibly. For a reversible process, that is, a process in which the system is always fully relaxed

$$\Delta S = \int_1^2 \frac{\delta q_{rev}}{T} \quad (8)$$

Infinitesimal small changes imply infinitesimal small deviations from equilibrium and, therefore, reversibility. The term  $\delta q$  in (5) may then be replaced by  $TdS$ , which gives

$$dU = Tds - pdV + \gamma dA + \psi dQ + \sum_i \mu_i dn_i \quad (9)$$

where all terms of the right-hand side are now of the form  $XdY$ . Equation (9) allows the intensive variables  $X$  to be expressed as differential quotients, such as, for instance,

$$\gamma = \left( \frac{\partial U}{\partial A} \right)_{s, v, Q, n_i, s} \quad (10)$$

where the subscripts indicate the properties that are kept constant. In other words, the interfacial tension equals the energy increment of the system resulting from the reversible extension of the interface by one unit area under the conditions of constant entropy, volume, electric charge, and composition. The required conditions make this definition very impractical, if not in operational. If the interface is extended it is very difficult to keep variables such as entropy and volume constant.

The other intensive variables may be expressed similarly as the change in energy per unit extensive property, under the appropriate conditions (Tempkin M. I. and Pyzhev V., 1940).

### 2.4 Maxwell relations

At equilibrium, implying that the intensive variables are constant throughout the system, (9) may be integrated, which yields

$$U = TdS - pV + \gamma A + \psi Q + \sum_i \mu_i n_i \quad (11)$$



To avoid impractical conditions when expressing intensive variables as differential quotients as, for example auxiliary functions are introduced. These are the enthalpy  $H$ , defined as

$$H \equiv U + pV \quad (12)$$

the Helmholtz energy

$$A \equiv U - TS \quad (13)$$

and the Gibbs energy

$$G \equiv U + pV - TS = H - TS = A + pV \quad (14)$$

Since  $U$  is a function of state, and  $p$ ,  $V$ ,  $T$ , and  $S$  are state variables,  $H$ ,  $A$ , and  $G$  are also functions of state. The corresponding differentials are

$$dH = TdS + Vdp + \gamma dA + \psi dQ + \sum_i \mu_i dn_i \quad (15)$$

$$dA = -SdT - pdV + \gamma dA + \psi dQ + \sum_i \mu_i dn_i \quad (16)$$

$$dG = -SdT + Vdp + \gamma dA + \psi dQ + \sum_i \mu_i dn_i \quad (17)$$

Expressing  $\gamma$ ,  $\Psi$ , or  $\mu_i$  as a differential quotient requires constancy of  $S$  and  $V$ ,  $S$  and  $p$ ,  $T$  and  $V$ , and  $T$  and  $p$ , when using the differentials  $dU$ ,  $dH$ ,  $dA$ , and  $dG$ , respectively. In most cases the conditions of constant  $T$  and  $V$  or constant  $T$  and  $p$  are most practical. It is noted that for heating or cooling a system at constant  $p$ , the heat exchange between the system and its environment is equal to the enthalpy exchange. Hence, for the heat capacity, at constant  $p$ ,

$$C_p = \left(\frac{\partial q}{\partial T}\right)_p = \left(\frac{dH}{dT}\right)_p \quad (18)$$

In general, for a function of state  $f$  that is completely determined by variables  $x$  and  $y$ ,  $df = A dx + B dy$ . Cross-differentiation in  $df$  gives  $(\partial A / \partial y)_x = (\partial B / \partial x)_y$ , known as a Maxwell relation. Similarly, cross-differentiation in  $dU$ ,  $dH$ ,  $dA$ , and  $dG$  yields a wide variety of Maxwell relations between differential quotients. For instance, by cross-differentiation in  $dG$  we find, (Lyklema, L. 1991), (Pitzer, K.S. and Brewer L. 1961).

$$\left(\frac{\partial \gamma}{\partial T}\right)_{p,A,Q,n_i,S} = -\left(\frac{\partial S}{\partial A}\right)_{T,p,Q,n_i,S} \quad (19)$$

## 2.5 Molar properties and partial molar properties

Molar properties, indicated by a lowercase symbol, are defined as an extensive property  $Y$  per mole of material:  $y = Y/n$ . Since they are expressed per mole, molar quantities are intensive.

For a single component system  $Y$  is a function of  $T$ ;  $p$ ; . . . ;  $n$ . Many extensive quantities vary linearly with  $n$ , but for some (e.g., the entropy) the variation with  $n$  is not proportional. In the latter case  $y$  is still a function of  $n$ . In a two-, three- or multi-component system (i.e., a mixture), the contribution of each component to the functions of state, say, the energy of the

system cannot be assigned unambiguously. This is because the energy of the system is not simply the sum of the energies of the constituting components but includes the interaction energies between the components as well. It is impossible to specify which part of the total interaction energy belongs to component  $i$ . For that reason partial molar quantities  $y_i$  are introduced. They are defined as the change in the extensive quantity  $Y$  pertaining to the whole system due to the addition of one mole of  $n_i$  under otherwise constant conditions. Because by adding component  $i$  the composition of the mixture and, hence, the interactions between the components are affected,  $y_i$  is defined as the differential quotient (Prausnitz, J.M., and et al. 1999).

$$y_i \equiv \left( \frac{\partial Y}{\partial n_i} \right)_{T,p,\dots,n_j \neq i} \quad (20)$$

The partial molar quantities are operational; that is, they can be measured. Now  $Y_{T,p,\dots,n_i,s}$  can be obtained as  $\sum_i n_i y_i$ . A partial molar quantity often encountered is the partial molar Gibbs energy (Aveyard, R. and Haydon, D.A., 1973),

$$g_i \equiv \left( \frac{\partial G}{\partial n_i} \right)_{T,p,\dots,n_j \neq i} \quad (21)$$

According to (17)

$$g_i \equiv \left( \frac{\partial G}{\partial n_i} \right)_{T,p,\dots,n_j \neq i} = \mu_i \quad (22)$$

that is, at constant  $T, p, \dots, n_j \neq i$ , the chemical potential of component  $i$  in a mixture equals its partial molar Gibbs energy.

By cross-differentiation in (17) the temperature- and pressure-dependence of  $\mu_i$  can be derived as

$$\left( \frac{\partial \mu_i}{\partial T} \right)_{p,\dots,n_i,s} = - \left( \frac{\partial S}{\partial n_i} \right)_{T,p,\dots,n_j \neq i} = -s_i \quad (23)$$

With

$$\mu_i = g_i \equiv H_i - Ts_i \quad (24)$$

it can be deduced that

$$\left( \frac{\partial(\mu_i / T)}{\partial T} \right)_{p,\dots,n_i,s} = - \frac{H}{T^2} \quad (25)$$

The pressure-dependence of  $\mu_i$  is also obtained from (17):

$$\left( \frac{\partial \mu_i}{\partial p} \right)_{T,\dots,n_i,s} = \left( \frac{\partial V}{\partial n_i} \right)_{T,p,\dots,n_j \neq i} \equiv v_i \quad (26)$$

For an ideal gas

$$v_i = \frac{RT}{p_i} \quad (27)$$

in which R is the universal gas constant. Combining (26) and (27) gives, after integration, an expression for  $\mu_i(p_i)$  in an ideal gas

$$\mu_i = \mu_i^0 + RT \ln p_i \quad (28)$$

where  $\mu_i^0$  is an integration constant that is independent of  $p_i$ ;  $\mu_i^0 \equiv \mu_i(p_i = 1)$ , its value depending on the units in which  $p_i$  is expressed. Similarly, without giving the derivation here, it is mentioned that for the chemical potential of component i in an ideal solution

$$\mu_i = \mu_i^0 + RT \ln c_i \quad (29)$$

where  $c_i$  is the concentration of i in the solution. In more general terms, for an ideal mixture

$$\mu_i = \mu_i^0 + RT \ln X_i \quad (30)$$

where  $X_i$  is the mole fraction of i in the mixture, defined as  $X_i \equiv n_i / \sum_i n_i$ . Note that the value of  $\mu_i^0$  is the one obtained for  $\mu_i$  by extrapolating to  $X_i=1$  assuming ideality of the mixture. This value deviates from the real value of  $\mu_i$  for pure i, because in the case of pure i the "mixture" is as far as possible from ideal. As said  $\mu_i$  and  $\mu_i^0$  are defined per unit  $X_i$ ,  $c_i$ , and  $p_i$ , respectively, and their values are therefore independent of the configurations of i in the mixture. They do depend on the interactions between i and the other components and therefore on the types of substances in the mixture. Because  $X_i$ ,  $c_i$ , and  $p_i$  are expressed in different units, the values for  $\mu_i$  and  $\mu_i^0$  differ (Keller J.U., 2005)

The  $RT \ln X_i$  term in Eq. (30) or, for that matter, the  $RT \ln c_i$  and  $RT \ln p_i$  terms in (28) and (29) do not contain any variable pertaining to the types of substances in the mixture. Hence, these terms are generic. Interpretation of the  $RT \ln X_i$  term follows from

$$\left( \frac{\partial \mu_i}{\partial T} \right)_{p_i, n_i, S} = \frac{\partial \mu_i^0}{\partial T} + R \ln X_i \quad (31)$$

which, because of (23), gives

$$s_i = s_i^0 - R \ln X_i \quad (32)$$

The partial molar entropy of i is composed of a part  $s_i^0$ , which is independent of the configurations of i in the mixture but dependent on the interactions of i with the other components, and a part  $R \ln X_i$ , which takes into account the possible configurations of i. It follows that the  $RT \ln X_i$  (or  $RT \ln c_i$  or  $RT \ln p_i$ ) term in the expressions for  $\mu_i$  stems from the configurational possibilities as well.

### 3. Basic thermodynamics of interfaces

For an open system of variable surface area, the Gibbs free energy must depend on composition, temperature,  $T$ , pressure,  $p$ , and the total surface area,  $A$ :

$$G = G(T, p, A, n_1, n_2, \dots, n_k) \quad (33)$$

From this function it follows that:

$$dG = \left( \frac{\partial G}{\partial T} \right)_{p, n_i} dT + \left( \frac{\partial G}{\partial p} \right)_{p, n_j} dp + \left( \frac{\partial G}{\partial A} \right)_{T, p, n_i} dA + \sum_{i=1}^{i=k} \left( \frac{\partial G}{\partial n_i} \right)_{T, p, n_j} dn_i \quad (34)$$

The first two partial differentials refer to constant composition, so we may use the general definitions:

$$G = H - TS = U + PV - TS \quad (35)$$

To obtain

$$S = - \left( \frac{\partial G}{\partial T} \right)_{p, n_j} \quad (36)$$

and

$$V = \left( \frac{\partial G}{\partial p} \right)_{p, n_i} \quad (37)$$

Insertion of these relations into (35) gives us the fundamental result

$$dG = -SdT + Vdp + \gamma dA + \sum_{i=1}^{i=k} \mu_i dn_i \quad (38)$$

where the chemical potential  $\mu_i$  is defined as:

$$\mu_i = \left( \frac{\partial G}{\partial n_i} \right)_{T, p, n_j} \quad (39)$$

and the surface energy  $\gamma$  as:

$$\gamma = \left( \frac{\partial G}{\partial A} \right)_{T, p, n_j} \quad (40)$$

The chemical potential is defined as the increase in free energy of a system on adding an infinitesimal amount of a component (per unit number of molecules of that component added) when  $T$ ,  $p$  and the composition of all other components are held constant. Clearly, from this definition, if a component 'i' in phase A has a higher chemical potential than in phase B (that is,  $\mu_i^A > \mu_i^B$ ) then the total free energy will be lowered if molecules are transferred from phase A to B and this will occur in a spontaneous process until the

chemical potentials equalize, at equilibrium. It is easy to see from this why the chemical potential is so useful in mixtures and solutions in matter transfer (open) processes (Norde, W., 2003). This is especially clear when it is understood that  $m_i$  is a simple function of concentration, that is:

$$\mu_i = \mu_i^0 + kT \ln C_i \quad (41)$$

for dilute mixtures, where  $m_i^0$  is the standard chemical potential of component 'i', usually 1 M for solutes and 1 atm for gas mixtures. This equation is based on the entropy associated with a component in a mixture and is at the heart of why we generally plot measurable changes in any particular solution property against the log of the solute concentration, rather than using a linear scale. Generally, only substantial changes in concentration or pressure produce significant changes in the properties of the mixture. (For example, consider the use of the pH scale.) (Koopal L.K., and et al. 1994).

### 3.1 Thermodynamics for closed systems

The First Law of Thermodynamics is the law of conservation of energy; it simply requires that the total quantity of energy be the same both before and after the conversion. In other words, the total energy of any system and its surroundings is conserved. It does not place any restriction on the conversion of energy from one form to another. The interchange of heat and work is also considered in this first law. In principle, the internal energy of any system can be changed, by heating or doing work on the system. The First Law of Thermodynamics requires that for a closed (but not isolated) system, the energy changes of the system be exactly compensated by energy changes in the surroundings. Energy can be exchanged between such a system and its surroundings in two forms: heat and work. Heat and work have the same units (joule, J) and they are ways of transferring energy from one entity to another. A quantity of heat,  $Q$ , represents an amount of energy in transit between a system and its surroundings, and is not a property of the system. Heat flows from higher to lower temperature systems. Work,  $W$ , is the energy in transit between a system and its surroundings, resulting from the displacement of external force acting on the system. Like heat, a quantity of work represents an amount of energy and is not a property of the system. Temperature is a property of a system while heat and work refer to a process. It is important to realize the difference between temperature, heat capacity and heat: temperature,  $T$ , is a property which is equal when heat is no longer conducted between bodies in thermal contact and can be determined with suitable instruments (thermometers) having a reference system depending on a material property (for example, mercury thermometers show the density differences of liquid mercury metal with temperature in a capillary column in order to visualize and measure the change of temperature). Suppose any closed system (thus having a constant mass) undergoes a process by which it passes from an initial state to a final state. If the only interaction with its surroundings is in the form of transfers of heat,  $Q$ , and work,  $W$ , then only the internal energy,  $U$ , can be changed, and the First Law of Thermodynamics is expressed mathematically as (Lyklema, J. ;2005 & Keller J.U.;2005)

$$\Delta U = U_{final} - U_{initial} = Q + W \quad (42)$$

where  $Q$  and  $W$  are quantities inclusive of sign so that when the heat transfers from the system or work is done by the system, we use negative values in Equation (11). Processes

where heat should be given to the system (or absorbed by the system) ( $Q > 0$ ) are called endothermic and processes where heat is taken from the system (or released from the system) ( $Q < 0$ ) are called exothermic. The total work performed on the system is  $W$ . There are many different ways that energy can be stored in a body by doing work on it: volumetrically by compressing it; elastically by straining it; electrostatically by charging it; by polarizing it in an electric field  $E$ ; by magnetizing it in a magnetic field  $H$ ; and chemically, by changing its composition to increase its chemical potential. In interface science, the formation of a new surface area is also another form of doing work. Each example is a different type of work – they all have the form that the (differential) work performed is the change in some extensive variable of the system multiplied by an intensive variable. In thermodynamics, the most studied work type is pressure–volume work,  $W_{PV}$ , on gases performed by compressing or expanding the gas confined in a cylinder under a piston. All other work types can be categorized by a single term, *non-pressure–volume work*,  $W_{non-PV}$ . Then,  $W$  is expressed as the sum of the pressure–volume work,  $W_{PV}$ , and the non pressure–volume work,  $W_{non-PV}$ , when many types of work are operative in a process (Miladinovic N., Weatherley L.R. 2008).

Equation (11) states that the internal energy,  $\Delta U$  depends only on the initial and final states and in no way on the path followed between them. In this form, heat can be defined as *the work-free transfer of internal energy from one system to another*. Equation (11) applies both to *reversible* and *irreversible* processes. A *reversible process* is an infinitely slow process during which departure from equilibrium is always infinitesimally small. In addition, such processes can be reversed at any moment by infinitesimal changes in the surroundings (in external conditions) causing it to retrace the initial path in the opposite direction. A reversible process proceeds so that the system is never displaced more than differentially from an equilibrium state. An *irreversible process* is a process where the departure from equilibrium cannot be reversed by changes in the surroundings. For a differential change, Equation (11) is often used in the differential form (Scatchard, G. 1976), (Zeldowitsch J., 1934):

$$dU = \delta W + \delta Q \quad (43)$$

for reversible processes involving infinitesimal changes only. The internal energy,  $U$  is a function of the measurable quantities of the system such as temperature, volume, and pressure, which are all state functions like internal energy itself. The differential  $dU$  is an exact differential similar to  $dT$ ,  $dV$ , and  $dP$ ; so we can always integrate

$$\left[ \int_1^2 f(U) dU \right] \text{ expression.}$$

### 3.2 Derivation of the gibbs adsorption isotherm

Let us consider the interface between two phases, say between a liquid and a vapor, where a solute (i) is dissolved in the liquid phase. The real concentration gradient of solute near the interface may look like Figure 10.1. When the solute increases in concentration near the surface (e.g. a surfactant) there must be a surface excess of solute  $n_i^s$ , compared with the bulk value continued right up to the interface. We can define a surface excess concentration (in units of moles per unit area) as:

$$\Gamma_i = \frac{n_i^\delta}{A} \tag{44}$$

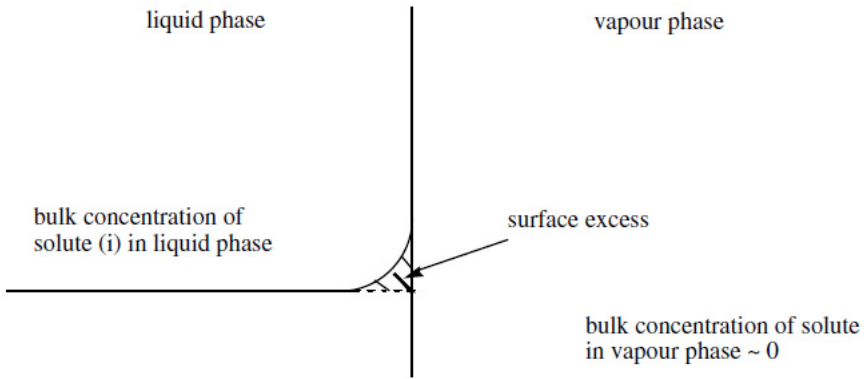


Fig. 1. Diagram of the variation in solute concentration at an interface between two phases.

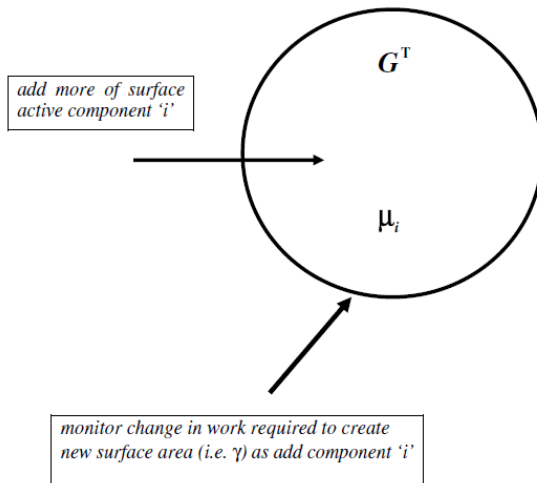


Fig. 2. Diagrammatic illustration of the change in surface energy caused by the addition of a solute.

where  $A$  is the interfacial area (note that  $\Gamma_i$  may be either positive or negative). Let us now examine the effect of adsorption on the interfacial energy ( $\gamma$ ). If a solute 'i' is positively adsorbed with a surface density of  $\Gamma_i$ , we would expect the surface energy to decrease on increasing the bulk concentration of this component (and vice versa). This situation is illustrated in Figure 10.2, where the total free energy of the system  $G^T$  and  $\mu_i$  are both increased by addition of component i but because this component is favourably adsorbed at the surface (only relative to the solvent, since both have a higher energy state at the surface),

the work required to create new surface (i.e.  $\gamma$ ) is reduced. Thus, although the total free energy of the system increases with the creation of new surface, this process is made easier as the chemical potential of the selectively adsorbed component increases (i.e. with concentration). This reduction in surface energy must be directly related to the change in chemical potential of the solute and to the amount adsorbed and is therefore given by the simple relationship (Zeldowitsch J., 1934):

$$d\gamma = -\Gamma_i d\mu_i \quad (45)$$

or, for the case of several components,

$$d\gamma = -\sum_i \Gamma_i d\mu_i \quad (46)$$

The change in  $\mu_i$  is caused by the change in bulk solute concentration. This is the Gibbs surface tension equation. Basically, these equations describe the fact that increasing the chemical potential of the adsorbing species reduces the energy required to produce new surface (i.e.  $\gamma$ ). This, of course, is the principal action of surfactants, which will be discussed in more detail in a later section. Using this result let us now consider a solution of two components

$$d\gamma = -\Gamma_1 d\mu_1 - \Gamma_2 d\mu_2 \quad (47)$$

and hence the adsorption excess for one of the components is given by

$$\Gamma_1 = -\left(\frac{\partial\gamma}{\partial\mu_1}\right)_{T,\mu_2} \quad (48)$$

Thus, in principle, we could determine the adsorption excess of one of the components from surface tension measurements, if we could vary  $\mu_1$  independently of  $\mu_2$ . But the latter appears not to be possible, because the chemical potentials are dependent on the concentration of each component. However, for dilute solutions the change in  $\mu$  for the solvent is negligible compared with that of the solute. Hence, the change for the solvent can be ignored and we obtain the simple result that

$$d\gamma = -\Gamma_1 d\mu_1 \quad (49)$$

Now, since  $\mu_1 = \mu_2 + RT \ln c_1$ , differentiation with respect to  $c_1$  gives

$$\left(\frac{\partial\mu_1}{\partial c_1}\right)_T = RT \left(\frac{\partial \ln c_1}{\partial c_1}\right)_T = \frac{RT}{c_1} \quad (50)$$

Then substitution in (49) leads to the result:

$$\Gamma_1 = -\frac{1}{RT} \left(\frac{\partial\gamma}{\partial \ln c_1}\right)_T = \frac{c_1}{RT} \left(\frac{\partial\gamma}{\partial c_1}\right)_T \quad (51)$$

This is the important Gibbs adsorption isotherm. (Note that for concentrated solutions the activity should be used in this equation.) An experimental measurement of  $\gamma$  over a range of concentrations allows us to plot  $\gamma$  against  $\ln c_1$  and hence obtain  $\Gamma_1$ , the adsorption density at



the surface. The validity of this fundamental equation of adsorption has been proven by comparison with direct adsorption measurements. The method is best applied to liquid/vapor and liquid/liquid interfaces, where surface energies can easily be measured. However, care must be taken to allow equilibrium adsorption of the solute (which may be slow) during measurement.

Finally, it should be noted that (51) was derived for the case of a single adsorbing solute (e.g. a non-ionic surfactant). However, for ionic surfactants such as CTAB, two species (CTA<sup>+</sup> and Br<sup>-</sup>) adsorb at the interface. In this case the equation becomes (Murrell, J.N. and Jenkins, A.D. 1994), (Ng J.C.Y., and et al. 2002):

$$\Gamma_1 = -\frac{1}{2RT} \left( \frac{\partial \gamma}{\partial \ln c_1} \right)_{TT} \quad (52)$$

because the bulk chemical potentials of both ions change with concentration of the surfactant.

#### 4. Fundamentals of pure component adsorption equilibrium

Adsorption equilibria information is the most important piece of information in understanding an adsorption process. No matter how many components are present in the system, the adsorption equilibria of pure components are the essential ingredient for the understanding of how many those components can be accommodated by a solid adsorbent. With this information, it can be used in the study of adsorption kinetics of a single component, adsorption equilibria of multicomponent systems, and then adsorption kinetics of multicomponent systems. In this section, we present the fundamentals of pure component equilibria. Various fundamental equations are shown, and to start with the proceeding we will present the most basic theory in adsorption: the Langmuir theory (1918). This theory allows us to understand the monolayer surface adsorption on an ideal surface. By an ideal surface here, we mean that the energy fluctuation on this surface is periodic and the magnitude of this fluctuation is larger than the thermal energy of a molecule ( $kT$ ), and hence the troughs of the energy fluctuation are acting as the adsorption sites. If the distance between the two neighboring troughs is much larger than the diameter of the adsorbate molecule, the adsorption process is called localised and each adsorbate molecule will occupy one site. Also, the depth of all troughs of the ideal surface are the same, that is the adsorption heat released upon adsorption on each site is the same no matter what the loading is. After the Langmuir theory, we will present the Gibbs thermodynamics approach. This approach treats the adsorbed phase as a single entity, and Gibbs adapted the classical thermodynamics of the bulk phase and applied it to the adsorbed phase. In doing this the concept of volume in the bulk phase is replaced by the area, and the pressure is replaced by the so-called spreading pressure. By assuming some forms of thermal equation of state relating the number of mole of adsorbate, the area and the spreading pressure (analogue of equations of state in the gas phase) and using them in the Gibbs equation, a number of fundamental equations can be derived, such as the linear isotherm, etc (Mohan D., Pittman Jr C.U. 2006).

Following the Gibbs approach, we will show the vacancy solution theory developed by Suwanayuen and Danner in 1980. Basically in this approach the system is assumed to consist of two solutions. One is the gas phase and the other is the adsorbed phase. The

difference between these two phases is the density. One is denser than the other. In the context of this theory, the vacancy solution is composed of adsorbate and vacancies. The latter is an imaginary entity defined as a vacuum space which can be regarded as the solvent of the system. Next, we will discuss one of the recent equations introduced by Nitta and his co-workers. This theory based on statistical thermodynamics has some features similar to the Langmuir theory, and it encompasses the Langmuir equation as a special case. Basically it assumes a localized monolayer adsorption with the allowance that one adsorbate molecule can occupy more than one adsorption site. Interaction among adsorbed molecules is also allowed for in their theory. As a special case, when the number of adsorption sites occupied by one adsorbate molecule is one, their theory is reduced to the Fowler-Guggenheim equation, and further if there is no adsorbate-adsorbate interaction this will reduce to the Langmuir equation. Another model of Nitta and co-workers allowing for the mobility of adsorbed molecules is also presented in this section.

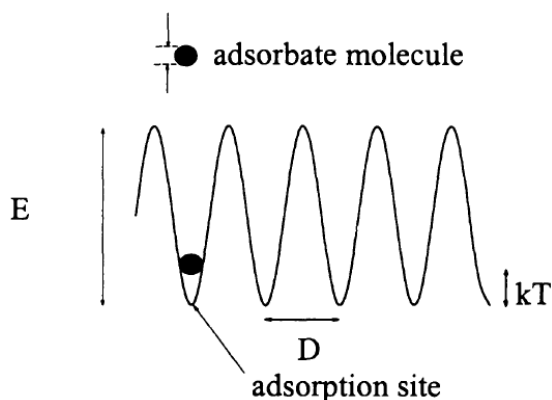


Fig. 3. Surface energy fluctuations

#### 4.1 Langmuir equation

Langmuir (1918) was the first to propose a coherent theory of adsorption onto a flat surface based on a kinetic viewpoint, that is there is a continual process of bombardment of molecules onto the surface and a corresponding evaporation (desorption) of molecules from the surface to maintain zero rate of accumulation at the surface at equilibrium. The assumptions of the Langmuir model are:

##### 4.1.1 Surface is homogeneous, that is adsorption energy is constant over all sites

##### 4.1.2 Adsorption on surface is localized, that is adsorbed atoms or molecules are adsorbed at definite, localized sites

##### 4.1.3 Each site can accommodate only one molecule or atom

The Langmuir theory is based on a kinetic principle, that is the rate of adsorption (which is the striking rate at the surface multiplied by a sticking coefficient, sometimes called the accommodation coefficient) is equal to the rate of desorption from the surface. The rate of striking the surface, in mole per unit time and unit area, obtained from the kinetic theory of gas is:

$$R_s = \frac{P}{\sqrt{2\pi MR_g T}} \quad (53)$$

To give the reader a feel about the magnitude of this bombardment rate of molecule, we tabulate below this rate at three pressures

<b>P (Torr)</b>	<b><math>R_s</math> (molecules/cm<sup>2</sup>/sec)</b>
760	$3 \times 10^{23}$
1	$4 \times 10^{20}$
$10^{-3}$	$4 \times 10^{17}$

Table 1. Magnitude of bombardment rate of molecule

This shows a massive amount of collision between gaseous molecules and the surface even at a pressure of  $10^{-3}$  Torr. A fraction of gas molecules striking the surface will condense and is held by the surface force until these adsorbed molecules evaporate again (see Figure 10.4). Langmuir (1918) quoted that there is good experimental evidence that this fraction is unity, but for a real surface which is usually far from ideal this fraction could be much less than unity. Allowing for the sticking coefficient  $\alpha$  (which accounts for non perfect sticking), the rate of adsorption in mole adsorbed per unit bare surface area per unit time is:

$$R_a = \frac{\alpha P}{\sqrt{2\pi MR_g T}} \quad (54)$$

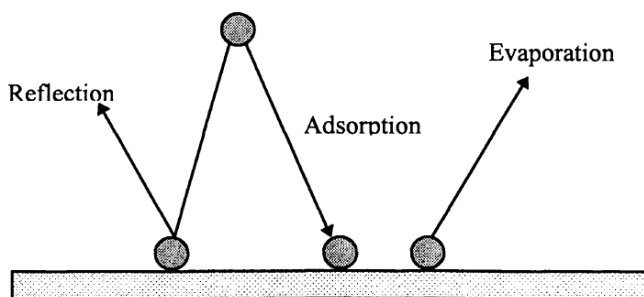


Fig. 4. Schematic diagram of Langmuir adsorption mechanism on a flat surface

This is the rate of adsorption on a bare surface. On an occupied surface, when a molecule strikes the portion already occupied with adsorbed species, it will evaporate very quickly, just like a reflection from a mirror. Therefore, the rate of adsorption on an occupied surface is equal to the rate given by eq. (54) multiplied by the fraction of empty sites, that is:

$$R_a = \frac{\alpha P}{\sqrt{2\pi MR_g T}} (1 - \theta) \quad (55)$$

where  $\theta$  is the fractional coverage. Here  $R_a$  is the number of moles adsorbed per unit area (including covered and uncovered areas) per unit time. The rate of desorption from the surface is equal to the rate, which corresponds to fully covered surface ( $k_d$ ), multiplied by the fractional coverage, that is:

$$R_d = k_d \theta = k_{d\infty} \exp\left(\frac{-E_d}{R_s T}\right) \theta \quad (56)$$

where  $E_d$  is the activation energy for desorption, which is equal to the heat of adsorption for physically adsorbed species since there is no energy barrier for physical adsorption. The parameter  $k_{d\infty}$  is the rate constant for desorption at infinite temperature. The inverse of this parameter is denoted as

$$\tau_{d\infty} = \frac{1}{k_{d\infty}} \quad (57)$$

The average residence time of adsorption is defined as:

$$\tau_a = \tau_{d\infty} e^{E_d/R_s T} \quad (58)$$

This means that the deeper is the potential energy well higher  $E_d$  the longer is the average residence time for adsorption. For physical adsorption, this surface residence time is typically ranging between  $10^{-13}$  to  $10^{-9}$  sec, while for chemisorption this residence time has a very wide range, ranging from  $10^{-6}$  (for weak chemisorption) to about  $10^9$  for systems such as CO chemisorbed on Ni. Due to the Arrhenius dependence on temperature this average surface residence time changes rapidly with temperature, for example a residence time of  $10^9$  at 300K is reduced to only 2 sec at 500K for a system having a desorption energy of 120 kJ/mole. Equating the rates of adsorption and desorption (Equations. 55 and 56), we obtain the following famous Langmuir isotherm written in terms of fractional loading:

$$\theta = \frac{bP}{1 + bP} \quad (59)$$

where

$$b = \frac{\alpha \exp(Q / R_s T)}{k_{d\infty} \sqrt{2\pi M R_s T}} = b_\infty \exp(Q / R_s T) \quad (60)$$

Here  $Q$  is the heat of adsorption and is equal to the activation energy for desorption,  $E_d$ . The parameter  $b$  is called the affinity constant or Langmuir constant. It is a measure of how strong an adsorbate molecule is attracted onto a surface. The pre exponential factor  $b_\infty$  of the affinity constant is:

$$b_\infty = \frac{\alpha}{k_{d\infty} \sqrt{2\pi M R_s T}} \quad (61)$$

which is inversely proportional to the square root of the molecular weight. When  $P$  is in Torr, the magnitude of  $b_\infty$  for nitrogen is given by Hobson (1965) as:

$$b_{\infty} = 5.682 \times 10^{-5} (MT)^{-1/2} \text{Torr}^{-1} \quad (62)$$

The isotherm equation (59) reduces to the Henry law isotherm when the pressure is very low ( $bP \ll 1$ ), that is the amount adsorbed increases linearly with pressure, a constraint demanded by statistical thermodynamics. When pressure is sufficiently high, the amount adsorbed reaches the saturation capacity, corresponding to a complete coverage of all adsorption sites with adsorbate molecules, this is called monolayer coverage,  $\theta \rightarrow 1$ . The behavior of the Langmuir isotherm ( $\Theta$  versus  $P$ ) is shown in Fig. 5.

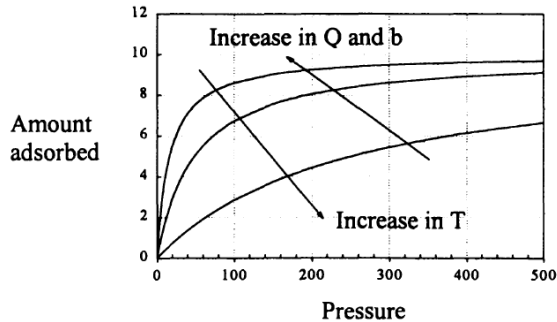


Fig. 5. Behavior of the Langmuir equation

When the affinity constant  $b$  is larger, the surface is covered more with adsorbate molecule as a result of the stronger affinity of adsorbate molecule towards the surface. Similarly, when the heat of adsorption  $Q$  increases, the adsorbed amount increases due to the higher energy barrier that adsorbed molecules have to overcome to evaporate back to the gas phase. Increase in the temperature will decrease the amount adsorbed at a given pressure. This is due to the greater energy acquired by the adsorbed molecule to evaporate. The isotherm equation (59) written in the form of fractional loading is not useful for the data correlation as isotherm data are usually collated in the form of amount adsorbed versus pressure. We now let  $C_{\mu}$  be the amount adsorbed in mole per unit mass or volume<sup>1</sup>, and  $C_{\mu s}$  be the maximum adsorbed concentration corresponding to complete monolayer coverage, then the Langmuir equation written in terms of the amount adsorbed useful for data correlation is:

$$C_{\mu} = C_{\mu s} \frac{b(T)P}{1 + b(T)P} \quad (63)$$

where

$$b(T) = b_{\infty} \exp(Q / R_s T) \quad (64)$$

Here we use the subscript  $\mu$  to denote the adsorbed phase, and this will be applied throughout this text. For example,  $C_{\mu}$  is the concentration of the adsorbed phase, and  $D_{\mu}$  is

<sup>1</sup> This volume is taken as the particle volume minus the void volume where molecules are present in free form.

the diffusion coefficient of the adsorbed phase,  $V_\mu$  is the volume of the adsorbed phase, etc. The temperature dependence of the affinity constant (e.g. 60) is  $T^{-1/2} \exp(Q/R_g T)$ . This affinity constant decreases with temperature because the heat of adsorption is positive, that is adsorption is an exothermic process. Since the free energy must decrease for the adsorption to occur and the entropy change is negative because of the decrease in the degree of freedom, therefore

$$\Delta H = \Delta G + T\Delta S < 0 \quad (65)$$

The negativity of the enthalpy change means that heat is released from the adsorption process. The Langmuir equation can also be derived from the statistical thermodynamics, based on the lattice statistics.

#### 4.2 Isostatic heat of adsorption

One of the basic quantities in adsorption studies is the isosteric heat, which is the ratio of the infinitesimal change in the adsorbate enthalpy to the infinitesimal change in the amount adsorbed. The information of heat released is important in the kinetic studies because when heat is released due to adsorption the released energy is partly absorbed by the solid adsorbent and partly dissipated to the surrounding. The portion absorbed by the solid increases the particle temperature and it is this rise in temperature that slows down the adsorption kinetics because the mass uptake is controlled by the rate of cooling of the particle in the later course of adsorption. Hence the knowledge of this isosteric heat is essential in the study of adsorption kinetics. The isosteric heat may or may not vary with loading. It is calculated from the following thermodynamic van't Hoff equation:

$$\frac{\Delta H}{R_g T^2} = - \left( \frac{\partial \ln P}{\partial T} \right)_{C_\mu} \quad (66)$$

For Langmuir isotherm of the form given in eq. (63), we take the total differentiation of that equation and substitute the result into the above van't Hoff equation to get:

$$\frac{\Delta H}{R_g T^2} = \frac{Q}{R_g T^2} + \delta(1 + bP) \quad (66)$$

in which we have allowed for the maximum adsorbed concentration ( $C_{\mu s}$ ) to vary with temperature and that dependence is assumed to take the form:

$$\frac{1}{C_{\mu s}} \frac{dC_{\mu s}}{dT} = -\delta \quad (67)$$

Since  $(1+bP) = 1/(1-\theta)$ , eq. (66) will become:

$$-\Delta H = Q + \frac{\delta R_g T^2}{1-\theta} \quad (68)$$

The negativity of the enthalpy change indicates that the adsorption process is an exothermic process. If the maximum adsorbed concentration,  $C^s$ , is a function of temperature and it decreases with temperature, the isosteric heat will increase with the loading due to the second term in the RHS of eq. (68). For the isosteric heat to take a finite value at high

coverage (that is  $\Theta \rightarrow 1$ ) the parameter  $\delta$  (thermal expansion coefficient of the saturation concentration) must be zero. This is to say that the saturation capacity is independent of temperature, and as a result the heat of adsorption is a constant, independent of loading.

**4.3 Isotherms based on the gibbs approach**

The last section dealt with the basic Langmuir theory, one of the earliest theories in the literature to describe adsorption equilibria. One should note that the Langmuir approach is kinetic by nature. Adsorption equilibria can be described quite readily by the thermodynamic approach. What to follow in this section is the approach due to Gibbs. More details can be found in Yang (1987) and Rudzinski and Everett (1992).

In the bulk  $\alpha$ -phase containing N components (Fig. 6), the following variables are specified: the temperature  $T^\alpha$ , the volume  $V^\alpha$  and the numbers of moles of all species  $n_i^\alpha$  (for  $i = 1, 2, \dots, N$ ). The upper script is used to denote the phase. With these variables, the total differential Helmholtz free energy is:

$$dA = -S^\alpha dT^\alpha - P^\alpha dV^\alpha + \sum_{i=1}^N \mu_i^\alpha dn_i^\alpha \tag{69}$$

where  $S^\alpha$  is the entropy of the  $\alpha$  phase,  $P^\alpha$  is the pressure of that phase,  $n^\alpha$  is the number of molecule of the species  $i$ , and  $(\mu_i^\alpha)$  is its chemical potential.

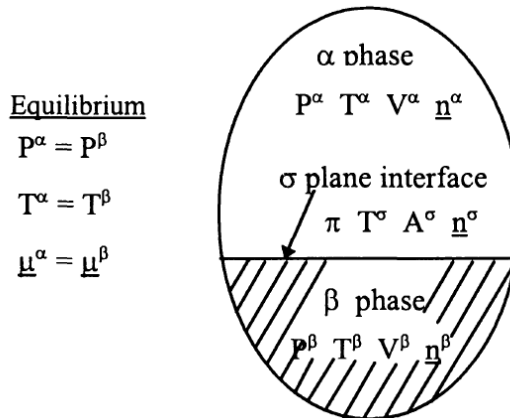


Fig. 6. Equilibrium between the phases  $\alpha$  and  $\beta$  separated by a plane interface a  $\delta$

Similarly, for the  $\beta$  -phase, we can write a similar equation for the differential Helmholtz free energy:

$$dA = -S^\beta dT^\beta - P^\beta dV^\beta + \sum_{i=1}^N \mu_i^\beta dn_i^\beta \tag{70}$$

If equilibrium exists between the two phases with a plane interface (Fig. 6), we have:

$$T^\alpha = T^\beta; P^\alpha = P^\beta; \mu_i^\alpha = \mu_i^\beta \tag{71}$$

that is equality in temperature, pressure and chemical potential is necessary and sufficient for equilibrium for a plane interface.

$$A^\alpha = -P V^\alpha + \sum_{i=1}^N \mu_i^\alpha n_i^\alpha \quad (72)$$

Differentiating eq. (72) and subtracting the result from eq. (69) will give the following Gibbs-Duhem equation:

$$-V^\alpha dP - S^\alpha dT + \sum_{i=1}^N \mu_i^\alpha n_i^\alpha = 0 \quad (73)$$

for the bulk  $\alpha$  phase. As a special case of constant temperature and pressure, the Gibbs-Duhem's relation is reduced to:

$$\sum_{i=1}^N \mu_i^\alpha n_i^\alpha = 0 \quad (74)$$

Similarly, the Gibbs-Duhem equation for the ( $\beta$ -phase at constant temperature and pressure is:

$$\sum_{i=1}^N \mu_i^\beta n_i^\beta = 0 \quad (75)$$

## 5. Thermodynamics of the surface phase

We now can develop a similar thermodynamic treatment for the surface phase  $\delta$ , which is the interface between the phases  $\alpha$  and  $\beta$ , and is in equilibrium with these two phases. When the adsorbed phase is treated as a two dimensional surface, fundamental equations in classical thermodynamics can still be applied. Applying the same procedure to surface free energy, we will obtain the Gibbs adsorption equation. This is done as follows. The total differentiation of the surface free energy takes the form similar to eq. (69) with  $P^\alpha dV^\beta$  being replaced by  $\pi dA$  :

$$dA^\sigma = -S^\sigma dT - \pi dA + \sum_{i=1}^N \mu_i dn_i^\sigma \quad (76)$$

where the surface chemical potentials  $\mu_i$  have the same values as those of the two joining phases,  $\pi$  is the spreading pressure, playing the same role as pressure in the bulk phase. Integrating eq. (76) with constant T,  $\pi$  and  $\mu_i$  yields:

$$A^\sigma = -\pi dA + \sum_{i=1}^N \mu_i n_i^\sigma \quad (77)$$

which is an analogue of eq. (72). Differentiation of this equation yields:

$$dA^\sigma = -\pi dA - Ad\pi + \sum_{i=1}^N \mu_i dn_i^\sigma + \sum_{i=1}^N n_i^\sigma d\mu_i \quad (78)$$

Subtracting eq. (72) from eq. (70), we have the Gibbs equation for a planar surface:



$$-S^\sigma dT - Ad\pi + \sum_{i=1}^N n_i^\sigma d\mu_i = 0 \quad (79)$$

Adsorption equilibria experiments are usually carried out at constant temperature, therefore the Gibbs adsorption isotherm equation is:

$$-Ad\pi + \sum_{i=1}^N n_i^\sigma d\mu_i = 0 \quad (80)$$

For pure component systems ( $N = 1$ ), we have:

$$-Ad\pi + nd\mu = 0 \quad (81)$$

where we have dropped the superscript  $\sigma$  for clarity. At equilibrium, the chemical potential of the adsorbed phase is equal to that of the gas phase, which is assumed to be ideal, i.e.

$$\mu = \mu_g = \mu_g^0 + R_g T \ln P \quad (82)$$

Substituting eq.(82) into eq.(81), the following Gibbs isotherm equation is derived:

$$\left( \frac{d\pi}{d \ln P} \right)_T = \frac{n}{A} R_g T \quad (83)$$

This equation is the fundamental equation relating gas pressure, spreading pressure and amount adsorbed. It is very useful in that if the equation of state relating the spreading pressure and the number of mole on the adsorbed phase is provided, the isotherm expressed as the number of mole adsorbed in terms of pressure can be obtained (Atkins, P.W., 1998).

### 5.1 Linear isotherm

For an ideal surface at infinite dilution, the equation of state relating the spreading pressure and the number of mole on the surface has the following form:

$$\pi A = nR_g T \quad (84)$$

an analogue of the ideal gas law (i.e. diluted systems), that is the spreading pressure is linear with the number of molecules on a surface of area  $A$ . Substituting this equation of state into the Gibbs equation, we get:

$$\pi = \left( \frac{d\pi}{d \ln P} \right)_T \quad (85)$$

Integrating this equation at constant  $T$ , we obtain  $n = C(T)P$ , where  $C(T)$  is some function of temperature. This equation means that at equilibrium the spreading pressure in the adsorbed phase is linearly proportional to the pressure in the gas phase. The spreading pressure is not, however, useful in the correlation of adsorption equilibrium data. To relate the amount adsorbed in the adsorbed phase in terms of the gas phase pressure, we use the equation of state to finally get:

$$\frac{n}{A} = K(T)P \quad (86)$$

where

$$K(T) = \frac{C(T)}{R_g T} \quad (87)$$

The parameter  $K(T)$  is called the Henry constant. The isotherm obtained for the diluted system is a linear isotherm, as one would anticipate from such condition of infinite dilution.

## 5.2 Volmer isotherm

We have seen in the last section that when the system is dilute (that is the equation of state follows eq. 84), the isotherm is linear because each adsorbed molecule acts independently from other adsorbed molecules. Now let us consider the case where we allow for the finite size of adsorbed molecules. The equation of state for a surface takes the following form:

$$\pi(A - A_0) = nR_g T \quad (88)$$

where  $A_0$  is the minimum area occupied by  $n$  molecules. The Gibbs equation (83) can be written in terms of the area per unit molecule as follows:

$$\left( \frac{\partial \pi}{\partial \ln P} \right)_T = \frac{R_g T}{\delta} \quad (89)$$

where the variable  $\delta$  is the area per unit molecule of adsorbate

$$\delta = \frac{A}{n} \quad (90)$$

Integrating equation (89) at constant temperature, we have:

$$\ln P = \frac{1}{R_g T} \int \delta d\pi \quad (91)$$

We rewrite the equation of state in terms of the new variable  $\delta$  and get:

$$\pi(\delta - \delta_0) = R_g T \quad (92)$$

Substituting the spreading pressure from the equation of state into the integral form of the Gibbs equation (91), we get:

$$\ln P = - \int \frac{\delta d\delta}{(\delta - \delta_0)^2} \quad (93)$$

But the fractional loading is simply the minimum area occupied by  $n$  molecules divided by the area occupied by the same number of molecules, that is

$$\theta = \frac{A_0}{A} = \frac{(A_0/n)}{(A/n)} = \frac{\delta_0}{\delta} \quad (94)$$

Written in terms of the fractional loading,  $\Theta$ , eq. (93) becomes:

$$\ln P = \int \frac{d\theta}{\theta(1-\theta)^2} \quad (95)$$

Carrying out the integration, we finally get the following equation:

$$b(T)P = \frac{\theta}{1-\theta} \exp\left(\frac{\theta}{1-\theta}\right) \quad (96)$$

where the affinity constant  $b(T)$  is a function of temperature, which can take the following form:

$$b(T) = b_{\infty} \exp\left(\frac{Q}{R_s T}\right) \quad (97)$$

Eq. (97) is known as the Volmer equation, a fundamental equation to describe the adsorption on surfaces where the mobility of adsorbed molecules is allowed, but no interaction is allowed among the adsorbed molecules. The factor  $\exp(\Theta/(1-\Theta))$  in eq. (96) accounts for the mobility of the adsorbate molecules. If we arrange eq. (96) as follows:

$$\frac{\theta}{1-\theta} = b \cdot \exp\left(-\frac{\theta}{1-\theta}\right) P \quad (98)$$

the Volmer equation is similar to the Langmuir isotherm equation with the apparent affinity as

$$b_{app} = b \cdot \exp\left(-\frac{\theta}{1-\theta}\right) \quad (99)$$

The difference between the Volmer equation and the Langmuir equation is that while the affinity constant remains constant in the case of Langmuir mechanism, the "apparent" affinity constant in the case of Volmer mechanism decreases with loading. This means that the rate of increase in loading with pressure is much lower in the case of Volmer compared to that in the case of Langmuir.

### 5.3 Hill-deboer isotherm

It is now seen that the Gibbs isotherm equation (83) is very general, and with any proper choice of the equation of state describing the surface phase an isotherm equation relating the amount on the surface and the gas phase pressure can be obtained as we have shown in the last two examples. The next logical choice for the equation of state of the adsorbate is an equation which allows for the co-volume term and the attractive force term. In this theme the following van der Waals equation can be used:

$$\left(\pi + \frac{a}{\delta^2}\right)(\delta - \delta_0) = R_s T \quad (100)$$

With this equation of state, the isotherm equation obtained is:

$$bP = \frac{\theta}{1-\theta} \exp\left(\frac{\theta}{1-\theta}\right) \exp(-c\theta) \quad (101)$$

where

$$bP = b_{\infty} \exp\left(\frac{Q}{R_g T}\right), c = \frac{2a}{R_g T \delta_0} = \frac{zw}{R_g T} \quad (102)$$

where  $z$  is the coordination number (usually taken as 4 or 6 depending on the packing of molecules), and  $w$  is the interaction energy between adsorbed molecules. A positive  $w$  means attraction between adsorbed species and a negative value means repulsion that is the apparent affinity is increased with loading when there is attraction between adsorbed species, and it is decreased with loading when there is repulsion among the adsorbed species. The equation as given in eq. (101) is known as the Hill-de Boer equation, which describes the case where we have mobile adsorption and lateral interaction among adsorbed molecules. When there is no interaction between adsorbed molecules (that is  $w = 0$ ), this Hill-de Boer equation will reduce to the Volmer equation obtained. The first exponential term in the RHS of eq. (101) describes the mobility of adsorbed molecules, and when this term is removed we will have the case of localized adsorption with lateral interaction among adsorbed molecules, that is:

$$bP = \frac{\theta}{1-\theta} \exp(-c\theta) \quad (103)$$

This equation is known in the literature as the Fowler-Guggenheim equation, or the quasi approximation isotherm. This equation can also be derived from the statistical thermodynamics. Due to the lateral interaction term  $\exp(-c\theta)$ , the Fowler-Guggenheim equation and the Hill-de Boer equation exhibit a very interesting behavior. This behavior is the two dimensional condensation when the lateral interaction between adsorbed molecules is sufficiently strong (Adam, N.K., 1968).

#### 5.4 Harkins-jura isotherm

We have addressed the various adsorption isotherm equations derived from the Gibbs fundamental equation. Those equations (Volmer, Fowler-Guggenheim and Hill de Boer) are for monolayer coverage situation. The Gibbs equation, however, can be used to derive equations which are applicable in multilayer adsorption as well. Here we show such application to derive the Harkins-Jura equation for multilayer adsorption. Analogous to monolayer films on liquids, Harkins and Jura (1943) proposed the following equation of state:

$$\pi = b - a\delta \quad (104)$$

where  $a$  and  $b$  are constants. Substituting this equation of state into the Gibbs equation (67) yields the following adsorption equation:

$$\ln\left(\frac{P}{P_0}\right) = B - \frac{C}{V^2} \quad (105)$$

which involves only measurable quantities. Here  $P_0$  is the vapor pressure. This equation can describe isotherm of type II shown in Figure 10.7. The classification of types of isotherm will be discussed in detail in next section. But for the purpose of discussion of the Harkins-Jura equation, we explain type II briefly here. Type II isotherm is the type which exhibits a similar behavior to Langmuir isotherm when the pressure is low, and when the pressure is further increased the amount adsorbed will increase in an exponential fashion. Rearranging the Harkins-Jura equation (105) into the form of adsorbed amount versus the reduced pressure, we have:

$$V = \frac{\sqrt{C/B}}{\sqrt{1 + \frac{1}{B} \ln(1/x)}} \quad (106)$$

where  $x$  is the reduced pressure ( $x=P/P_0$ ), We see that when the pressure approaches the vapor pressure, the adsorbed amount reaches a maximum concentration given below:

$$\lim V = V_{\max} = \sqrt{C/B} \quad (107)$$

Thus, the Harkins-Jura isotherm equation can be written as

$$\frac{V}{V_{\max}} = \frac{1}{\sqrt{1 + \frac{1}{B} \ln(1/x)}} \quad (108)$$

from which we can see that the only parameter which controls the degree of curvature of the isotherm is the parameter  $B$ .

### 5.5 Characteristics of isotherm

To investigate the degree of curvature of the Harkins-Jura equation (108), we study its second derivative:

$$\frac{d^2}{dx^2} \left( \frac{V}{V_{\max}} \right) = \frac{1}{2Bx^2} \frac{\left( \frac{3}{2B} - \left[ 1 + \frac{1}{B} \ln(1/x) \right] \right)}{\left[ 1 + \frac{1}{B} \ln(1/x) \right]^{5/2}} \quad (109)$$

To find the inflexion point, we set the second derivative to zero and obtain the reduced pressure at which the isotherm curve has an inflexion point

$$x_{\text{inf.}} = \exp\left[-\left(\frac{3}{2} - B\right)\right] \quad (110)$$

For the Harkins-Jura equation to describe the Type II isotherm, it must have an inflexion point occurring at the reduced pressure between 0 and 1, that is the restriction on the parameter  $B$  between  $3/2 > B > 0$ . The restriction of positive  $B$  is due to the fact that if  $B$  is negative, eq. (105) does not always give a real solution. With the restriction on  $B$  as shown in eq. (109), the minimum reduced pressure at which the inflexion point occurs is (by putting  $B$  to zero in eq. 110):

$$x_{\text{inf.}} = \exp\left[-\left(\frac{3}{2}\right)\right] \approx 0.22 \quad (111)$$

Fig. 7 shows typical plots of the Harkins-Jura equation.

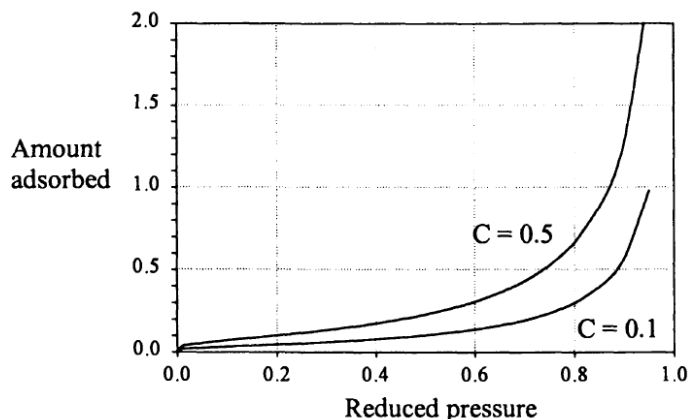


Fig. 7. Plots of the Harkins-Jura equation versus the reduced pressure with  $B = 0.01$

Jura and Harkins claimed that this is the simplest equation found so far for describing adsorption from sub-monolayer to multilayer regions, and it is valid over more than twice the pressure range of any two-constant adsorption isotherms. They showed that for  $\text{TiO}_2$  in the form of anatase, their isotherm agrees with the data at both lower and higher values of pressure than the commonly used BET equation. Harkins and Jura (1943) have shown that a plot of  $\ln(P/P_0)$  versus  $1/v^2$  would yield a straight line with a slope of  $C$ . The square root of this constant is proportional to the surface area of the solid. They gave the following formula:

$$S_g = 4.06\sqrt{C} \quad (112)$$

where  $v$  is the gas volume at STP adsorbed per unit g, and  $S$  has the unit of  $\text{m}^2/\text{g}$ . They also suggested that if the plot of  $\ln(P/P_0)$  versus  $1/v^2$  exhibits two straight lines, the one at lower pressure range should be chosen for the area calculation as this is the one in which there exists a transition from a monolayer to a polylayer.

### 5.6 Other isotherms from gibbs equation

We see that many isotherm equations (linear, Volmer, Hill-deBoer, Harkins-Jura) can be derived from the generic Gibbs equation. Other equations of state relating the spreading pressure to the surface concentration can also be used, and thence isotherm equations can be obtained. The following table (Table 2) lists some of the fundamental isotherm equations from a number of equations of state (Ross and Olivier, 1964; Adamson, 1984).

Since there are many fundamental equations which can be derived from various equations of state, we will limit ourselves to a few basic equations such as the Henry law equation, the Volmer, the Fowler-Guggenheim, and the Hill-de Boer equation. Usage of more complex fundamental equations other than those just mentioned needs justification for doing so.

Equation of state	Isotherm	Name
$\pi\delta = R_s T$	$bP = \theta$	Henry law
$\pi\delta = R_s T \ln\left(\frac{\delta}{\delta - \delta_0}\right)$	$bP = \frac{\theta}{1 - \theta}$	Langmuir
$\pi(\delta - \delta_0) = R_s T$	$bP = \frac{\theta}{1 - \theta} \exp\left(\frac{\theta}{1 - \theta}\right)$	Volmer
$\pi\delta = R_s T \frac{\delta}{\delta_0} \ln\left(\frac{\delta}{\delta - \delta_0}\right) - \frac{cw}{2} \frac{\delta_0}{\delta}$	$bP = \frac{\theta}{1 - \theta} \exp\left(-\frac{cw\theta}{RT}\right)$	Fowler-Guggenheim
$\left(\pi + \frac{a}{\delta^2}\right)(\delta - \delta_0) = R_s T$	$bP = \frac{\theta}{1 - \theta} \exp\left(\frac{\theta}{1 - \theta}\right) \exp(-c\theta)$	Hill-deBoer
$\left(\pi + \frac{a}{\delta^3}\right)(\delta - \delta_0) = R_s T$	$bP = \frac{\theta}{1 - \theta} \exp\left(\frac{\theta}{1 - \theta}\right) \exp(-\theta^2)$	-
$\left(\pi + \frac{a}{\delta^2}\right)\left(\delta - \frac{\delta_0}{\delta}\right) = R_s T$	$bP = \sqrt{\frac{\theta}{1 - \theta}} \exp\left(\frac{\theta}{1 - \theta}\right) \exp(-c\theta)$	-

Table 2. Isotherm Equations derived from the Gibbs Equation

### 5.7 Equation 2D of state adsorption isotherm

Numbers of fundamental approaches have been taken to derive the necessary adsorption isotherm. If the adsorbed fluid is assumed to behave like a two dimensional non-ideal fluid, then the Equation of State developed for three dimensional fluids can be applied to two dimensional fluids with a proper change of variables. The 2D- equation of state (2D.EOS) adsorption isotherm equations are not popularly used in the description of data, but they have an advantage of easily extending to multicomponent mixtures by using a proper mixing rule for the adsorption parameters. For 3D fluids, the following 3 parameter EOS equation is popularly used (Dubinin M. M., Radushkevich L.V. 1947)

$$\left(p + \frac{a}{v^2 + \alpha bv + \beta b^2}\right)(v - b) = R_s T \quad (113)$$

where p is the pressure, v is the volume per unit mole, a and b are parameters of the fluid and  $\alpha$  and  $\beta$  represent numerical values. Different values obtained of  $\alpha$  and  $\beta$ , give different forms of equation of state. For example, when  $\alpha = \beta = 0$ , we recover the famous van der Waals equation. Written in terms of molar density d (mole/volume), the 3D-EOS will become:

$$\left(p + \frac{ad^2}{1 + \alpha bd + \beta b^2 d^2}\right)(1 - bd) = dR_s T \quad (114)$$

Adopting the above form, we can write the following equation for the 2D-EOS as follows:

$$\left(\pi + \frac{a_s \delta^2}{1 + \alpha b_s \delta + \beta b_s^2 \delta^2}\right)(1 - b_s \delta) = \delta R_s T \quad (115)$$

where  $n$  is the spreading pressure,  $a$  is the surface density (mole/area) and the parameters  $\alpha$  and  $\beta$  are the 2D analogs of  $a$  and  $b$  of the 3D-EOS. Written in terms of the surface concentration (mole/mass), the above equation becomes:

$$\left( A\pi + \frac{a_s w^2}{1 + \alpha b_s w + \beta b_s^2 w^2} \right) (1 - b_s w) = w R_s T \quad (116)$$

where  $A$  is the specific area (m<sup>2</sup>/g). To provide an EOS to properly fit the experimental data, Zhou et al. (1994) suggested the following form containing one additional parameter

$$\left( A\pi + \frac{a_s w^2}{1 + \alpha b_s w + \beta b_s^2 w^2} \right) (1 - (b_s w)^m) = w R_s T \quad (117)$$

This general equation reduces to special equations when the parameters  $\alpha$ ,  $\beta$  and  $m$  take some specific values. The following table shows various special cases deduced from the above equation.

$\alpha$	$\beta$	$m$	EOS
0	0	1	van der Waals
0	0	1/2	Eyring
1	0	1	Soave-Redlich-Kwong
2	-1	1	Peng-Robinson

Table 3. The various special cases deduced from equations

To fit many experimental data, Zhou et al. (1994) have found that  $m$  has to be less than 1/2. They suggested a value of 1/3 for  $m$  to reduce the number of parameters in the 2D-EOS equation (117). At equilibrium, the chemical potential of the adsorbed phase is the same as that of the gas phase that is

$$\mu_a = \mu_g = \mu_g^0 + R_s T \ln P \quad (118)$$

The chemical potential of the adsorbed phase is related to the spreading pressure according to the Gibbs thermodynamics equation rewritten here for clarity:

$$-A d\pi + n d\mu = 0 \quad (119)$$

Thus

$$\left( \frac{d\pi}{d \ln P} \right) = \frac{n}{A} R_s T = \delta R_s T \quad (119)$$

But the spreading pressure is a function of  $a$  as governed by the equation of state (115). We write

$$d \ln P = \frac{1}{R_s T} \frac{1}{\delta} \left( \frac{\partial \pi}{\partial \delta} \right)_T d\delta \quad (120)$$



Integrating the above equation, we get

$$\int^p d \ln P = \frac{1}{R_s T} \int^s \frac{1}{\delta} \left( \frac{\partial \pi}{\partial \delta} \right)_T d\delta \quad (121)$$

Eq. (121) is the adsorption isotherm equation relating the surface density  $a$  (mole/m<sup>2</sup>) in terms of the gas phase pressure. The applicability of this isotherm equation rests on the ability of the 2D-EOS (eq. 121) to describe the state of the adsorbed molecule. Discussions on the usage of the above equation in the fitting of experimental data are discussed in Zhou et al. (1994).

## 6. Empirical isotherm equations

In this section, we present a number of popularly used isotherm equations. We start first with the earliest empirical equation proposed by Freundlich, and then Sips equation which is an extension of the Freundlich equation, modified such that the amount adsorbed in the Sips equation has a finite limit at sufficiently high pressure (or fluid concentration). We then present the two equations which are commonly used to describe well many data of hydrocarbons, carbon oxides on activated carbon and zeolite: Toth and Unilan equations. A recent proposed equation by Keller et al. (1996), which has a form similar to that of Toth, is also discussed. Next, we describe the Dubinin equation for describing micropore filling, which is popular in fitting data of many microporous solids. Finally we present the relatively less used equations in physical adsorption, Jovanovich and Tempkin, the latter of which is more popular in the description of chemisorption systems (Erbil, H.Y.,1997).

### 6.1 Freundlich equation

The Freundlich equation is one of the earliest empirical equations used to describe equilibria data. The name of this isotherm is due to the fact that it was used extensively by Freundlich (1932) although it was used by many other researchers. This equation takes the following form:

$$C_\mu = KP^{1/n} \quad (122)$$

where  $C_\mu$  is the concentration of the adsorbed species, and  $K$  and  $n$  are generally temperature dependent. The parameter  $n$  is usually greater than unity. The larger is this value; the adsorption isotherm becomes more nonlinear as its behavior deviates further away from the linear isotherm. To show the behavior of the amount adsorbed versus pressure (or concentration) we plot  $(C_\mu/C_{\mu 0})$  versus  $(P/P_0)$  as shown in Figure 10.8, that is

$$\frac{C_\mu}{C_{\mu 0}} = \left( \frac{P}{P_0} \right)^{1/n} \quad (123)$$

where  $P_0$  is some reference pressure and  $C_{\mu 0}$  is the adsorbed concentration at that reference pressure,  $C_{\mu 0} = KP_0^{1/n}$

We see from Figure 10.8 that the larger is the value of  $n$ , the more nonlinear is the adsorption isotherm, and as  $n$  is getting larger than about 10 the adsorption isotherm is approaching a so-called rectangular isotherm (or irreversible isotherm). The term "irreversible isotherm" is normally used because the pressure (or concentration) needs to go

down to an extremely low value before adsorbate molecules would desorb from the surface. The Freundlich equation is very popularly used in the description of adsorption of organics from aqueous streams onto activated carbon. It is also applicable in gas phase systems having heterogeneous surfaces, provided the range of pressure is not too wide as this isotherm equation does not have a proper Henry law behavior at low pressure, and it does not have a finite limit when pressure is sufficiently high. Therefore, it is generally valid in the narrow range of the adsorption data. Parameters of the Freundlich equation can be found by plotting  $\log_{10}(C_M)$  versus  $\log_{10}(P)$

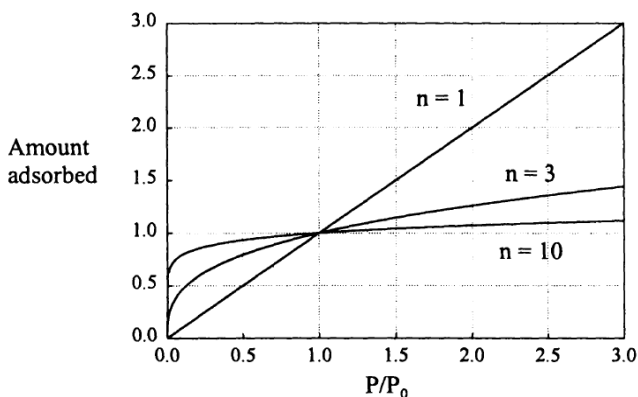


Fig. 8. Plots of the Freundlich isotherm versus  $P/P_0$

$$\log_{10}(C_{\mu}) = \log_{10} K + \frac{1}{n} \log_{10} P \quad (124)$$

which yields a straight line with a slope of  $(1/n)$  and an intercept of  $\log_{10}(K)$ .

### 6.1.1 Temperature dependence of K and n

The parameters  $K$  and  $n$  of the Freundlich equation (122) are dependent on temperature. Their dependence on temperature is complex, and one should not extrapolate them outside their range of validity. The system of CO adsorption on charcoal has temperature-dependent  $n$  such that its inverse is proportional to temperature. This exponent was found to approach unity as the temperature increases. This, however, is taken as a specific trend rather than a general rule. To derive the temperature dependence of  $K$  and  $n$ , we resort to an approach developed by Urano et al. (1981). They assumed that a solid surface is composed of sites having a distribution in surface adsorption potential, which is defined as:

$$A' = R_g T \ln \left( \frac{P_0}{P} \right) \quad (125)$$

The adsorption potential  $A'$  is the work (energy) required to bring molecules in the gas phase of pressure  $P$  to a condensed state of vapor pressure  $P_0$ . This means that sites associated with this potential  $A$  will have a potential to condense molecules from the gas phase of pressure  $P$  If the adsorption potential of the gas

$$A = R_g T \ln\left(\frac{P_0}{P}\right) \tag{126}$$

is less than the adsorption potential  $A'$  of a site, then that site will be occupied by an adsorbate molecule. On the other hand, if the gas phase adsorption potential is greater, then the site will be unoccupied (Fig. 9). Therefore, if the surface has a distribution of surface adsorption potential  $F(A')$  with  $F(A')dA'$  being the amount adsorbed having adsorption potential between  $A'$  and  $A'+dA'$ , the adsorption isotherm equation is simply:

$$C_\mu = \int_A^\infty F(A')dA' \tag{127}$$

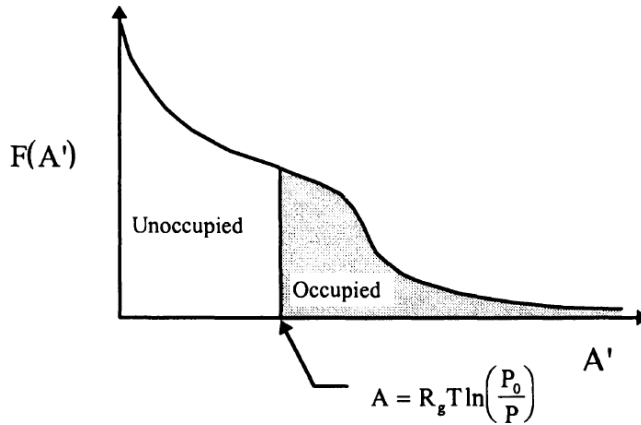


Fig. 9. Distribution of surface adsorption potential

If the density function  $F(A')$  takes the form of decaying exponential function

$$F(A) = \delta \cdot \exp(-A / A_0) \tag{128}$$

where  $A_0$  is the characteristic adsorption potential, the above integral can be integrated to give the form of the Freundlich equation:

$$C_\mu = KP^{1/n} \tag{129}$$

where the parameter  $K$  and the exponent  $(1/n)$  are related to the distribution parameters  $\delta$ ,  $A_0$ , and the vapor pressure and temperature as follows:

$$K = (\delta A_0) P_0^{-R_g T/A_0} \tag{130}$$

$$\frac{1}{n} = \frac{R_g T}{A_0} \tag{131}$$

The parameter  $n$  for most practical systems is greater than unity; thus eq. (131) suggests that the characteristic adsorption energy of surface is greater than the molar thermal energy  $R_g T$ .

Provided that the parameters  $\beta$  and  $A_0$  of the distribution function are constant, the parameter  $1/n$  is a linear function of temperature, that is  $nRT$  is a constant, as experimentally observed for adsorption of CO in charcoal for the high temperature range (Rudzinski and Everett, 1992). To find the temperature dependence of the parameter  $K$ , we need to know the temperature dependence of the vapor pressure, which is assumed to follow the Clapeyron equation:

$$\ln P_0 = \alpha - \frac{\beta}{T} \quad (132)$$

Taking the logarithm of  $K$  in eq. (131) and using the Clapeyron equation (132), we get the following equation for the temperature dependence of  $\ln K$ :

$$\ln K = \left[ \ln(\delta A_0) + \frac{\beta R_g}{A_0} \right] - \frac{\alpha R_g T}{A_0} \quad (133)$$

This equation states that the logarithm of  $K$  is a linear function of temperature, and it decreases with temperature. Thus the functional form to describe the temperature dependence of  $K$  is

$$K = K_0 \exp\left(-\frac{\alpha R_g T}{A_0}\right) \quad (134)$$

and hence the explicit temperature dependence form of the Freundlich equation is:

$$C_\mu = K_0 \exp\left(-\frac{\alpha R_g T}{A_0}\right) P^{R_g T / A_0} \quad (135)$$

Since  $\ln C_\mu$  and  $1/n$  are linear in terms of temperature, we can eliminate the temperature and obtain the following relationship between  $\ln K$  and  $n$ :

$$\ln K = \left[ \ln(\delta A_0) + \frac{\beta R_g}{A_0} \right] - \frac{\alpha}{n} \quad (136)$$

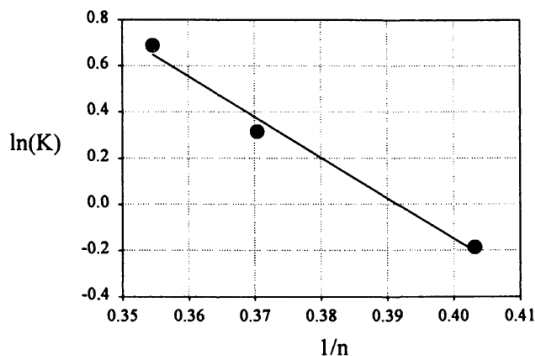


Fig. 10. Plot of  $\ln(K)$  versus  $1/n$  for propane adsorption on activated carbon

suggesting that the two parameters  $K$  and  $n$  in the Freundlich equation are not independent. Huang and Cho (1989) have collated a number of experimental data and have observed the linear dependence of  $\ln(K)$  and  $(1/n)$  on temperature. We should, however, be careful about using this as a general rule for extrapolation as the temperature is sufficiently high, the isotherm will become linear, that is  $n = 1$ , meaning that  $1/n$  no longer follows the linear temperature dependence as suggested by eq. (131). Thus, eq. (136) has its narrow range of validity, and must be used with extreme care. Using the propane data on activated carbon, we show in Figure 10 that  $\ln K$  and  $1/n$  are linearly related to each other, as suggested by eq.(136).

**6.2 Heat of adsorption**

Knowing  $K$  and  $n$  as a function of temperature, we can use the van't Hoff equation

$$\Delta H = -R_s T^2 \left( \frac{\partial \ln P}{\partial T} \right)_{C_\mu} \tag{137}$$

to determine the isosteric heat of adsorption. The result is (Huang and Cho, 1989)

$$\Delta H = - \left[ \ln(\sigma A_0) + \frac{R_s \beta}{A_0} \right] A_0 + A_0 \ln C_\mu \tag{138}$$

Thus, the isosteric heat is a linear function of the logarithm of the adsorbed amount.

**6.3 Sips equation (langmuir-freundlich)**

Recognizing the problem of the continuing increase in the adsorbed amount with an increase in pressure (concentration) in the Freundlich equation, Sips (1948) proposed an equation similar in form to the Freundlich equation, but it has a finite limit when the pressure is sufficiently high.

$$C_\mu = C_{\mu s} \frac{(bP)^{1/n}}{1 + (bP)^{1/n}} \tag{139}$$

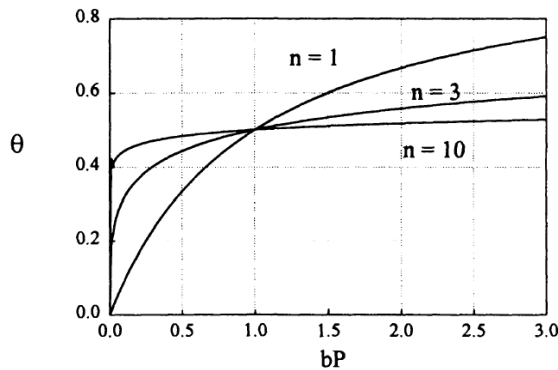


Fig. 11. Plots of the Sips equation versus  $bP$

In form this equation resembles that of Langmuir equation. The difference between this equation and the Langmuir equation is the additional parameter "n" in the Sips equation. If this parameter n is unity, we recover the Langmuir equation applicable for ideal surfaces. Hence the parameter n could be regarded as the parameter characterizing the system heterogeneity. The system heterogeneity could stem from the solid or the adsorbate or a combination of both. The parameter n is usually greater than unity, and therefore the larger is this parameter the more heterogeneous is the system. Figure 11 shows the behavior of the Sips equation with n being the varying parameter. Its behavior is the same as that of the Freundlich equation except that the Sips equation possesses a finite saturation limit when the pressure is sufficiently high. However, it still shares the same disadvantage with the Freundlich isotherm in that neither of them have the right behavior at low pressure, that is they don't give the correct Henry law limit. The isotherm equation (139) is sometimes called the Langmuir-Freundlich equation in the literature because it has the combined form of Langmuir and Freundlich equations.

To show the good utility of this empirical equation in fitting data, we take the same adsorption data of propane onto activated carbon used earlier in the testing of the Freundlich equation. The following Figure (Figure 10.12) shows the degree of good fit between the Sips equation and the data. The fit is excellent and it is fairly widely used to describe data of many hydrocarbons on activated carbon with good success. For each temperature, the fitting between the Sips equation and experimental data is carried out with MatLab nonlinear optimization outline, and the optimal parameters from the fit are tabulated in the following table. A code ISOFIT1 provided with this book is used for this optimization, and students are encouraged to use this code to exercise on their own adsorption data.

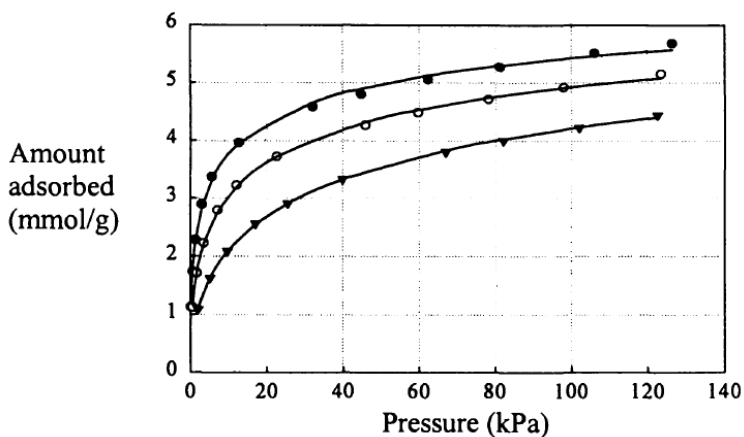


Fig. 12. Fitting of the propane/activated carbon data with the Sips equation (symbol -data; line:fitted equation)

The optimal parameters from the fitting of the Sips equation with the experimental data are tabulated in Table 4.

T (K)	C <sub>μs</sub> (mmole/g)	b (kPa <sup>-1</sup> )	n (-)
283	7.339	0.1107	2.306
303	7.232	0.04986	2.117
333	7.583	0.01545	1.956

Table 4. Optimal parameters for the Sips equation in fitting propane data on activated carbon

The parameter n is greater than unity, suggesting some degree of heterogeneity of this propane/ activated carbon system. The larger is this parameter, the higher is the degree of heterogeneity. However, this information does not point to what is the source of the heterogeneity, whether it be the solid structural property, the solid energetically property or the sorbet property. We note from the above table that the parameter n decreases with temperature, suggesting that the system is "apparently" less heterogeneous as temperature increases.

### 6.3.1 The temperature dependence of the sips equation

For useful description of adsorption equilibrium data at various temperatures, it is important to have the temperature dependence form of an isotherm equation. The temperature dependence of the Sips equation

$$C_{\mu} = C_{\mu s} \frac{(bP)^{1/n}}{1 + (bP)^{1/n}} \quad (140)$$

for the affinity constant b and the exponent n may take the following form:

$$b = b_{\infty} \exp\left(\frac{Q}{R_s T}\right) = b_0 \exp\left[\frac{Q}{R_s T_0} \left(\frac{T_0}{T}\right) - 1\right] \quad (141)$$

$$\frac{1}{n} = \frac{1}{n_0} + \alpha \left(1 - \frac{T_0}{T}\right) \quad (142)$$

Here  $b_{\infty}$  is the adsorption affinity constant at infinite temperature,  $b_0$  is that at some reference temperature  $T_0$  is the parameter n at the same reference temperature and  $\alpha$  is a constant parameter. The temperature dependence of the affinity constant b is taken from the of the Langmuir equation. Unlike Q in the Langmuir equation, where it is the isosteric heat, invariant with the surface loading, the parameter Q in the Sips equation is only the measure of the adsorption heat. The temperature-dependent form of the exponent n is empirical and such form in eq. (142) is chosen because of its simplicity. The saturation capacity can be either taken as constant or it can take the following temperature dependence:

$$C_{\mu s} = C_{\mu s,0} \exp\left[x\left(1 - \frac{T}{T_0}\right)\right] \quad (143)$$

Here  $C_{\mu s,0}$  is the saturation capacity at the reference temperature  $T_0$ , and x is a constant parameter. This choice of this temperature-dependent form is arbitrary. This temperature

dependence form of the Sips equation (142) can be used to fit adsorption equilibrium data of various temperatures simultaneously to yield the parameter  $b_0$ ,  $C_{\mu s,0}$ ,  $Q/RT_0$ , ratio and  $\alpha$ .

#### 6.4 Toth equation

The previous two equations have their limitations. The Freundlich equation is not valid at low and high end of the pressure range, and the Sips equation is not valid at the low end as they both do not possess the correct Henry law type behavior. One of the empirical equations that is popularly used and satisfies the two end limits is the Toth equation. This equation describes well many systems with sub-monolayer coverage, and it has the following form:

$$C_{\mu} = C_{\mu s} \frac{bP}{[1 + (bP)^t]^{1/t}} \quad (144)$$

Here  $t$  is a parameter which is usually less than unity. The parameters  $b$  and  $t$  are specific for adsorbate-adsorbent pairs. When  $t = 1$ , the Toth isotherm reduces to the famous Langmuir equation; hence like the Sips equation the parameter  $t$  is said to characterize the system heterogeneity. If it is deviated further away from unity, the system is said to be more heterogeneous. The effect of the Toth parameter  $t$  is shown in Figure 10-13, where we plot the fractional loading ( $C_{\mu}/C_{\mu s}$ ) versus  $bP$  with  $t$  as the varying parameter. Again we note that the more the parameter  $t$  deviates from unity, the more heterogeneous is the system. The Toth equation has correct limits when  $P$  approaches either zero or infinity.

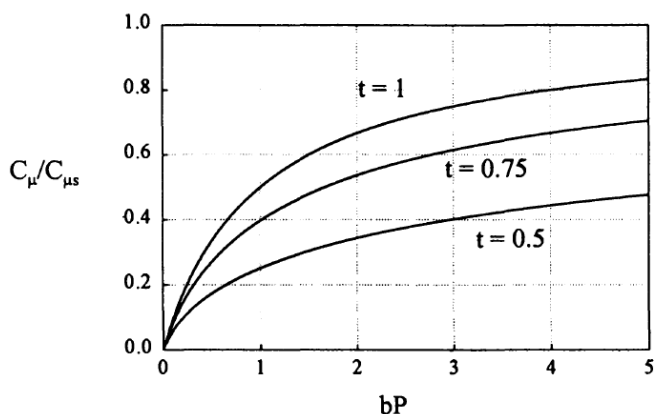


Fig. 13. Plot of the fractional loading versus  $bP$  for the Toth equation

Being the three-parameter model, the Toth equation can describe well many adsorption data. We apply this isotherm equation to fit the isotherm data of propane on activated carbon. The extracted optimal parameters are:  $C_{\mu s}=33.56$  mmole/g,  $b=0.069$  (kPa)<sup>-1</sup>,  $t=0.233$ . The parameter  $t$  takes a value of 0.233 (well deviated from unity) indicates a strong degree of heterogeneity of the system. Several hundred sets of data for hydrocarbons on Nuxit-al charcoal obtained by Szepesy and Illes (Valenzuela and Myers, 1989) can be described well by this equation. Because of its simplicity in form and its correct behavior at low and high



pressures, the Toth equation is recommended as the first choice of isotherm equation for fitting data of many adsorbates such as hydrocarbons, carbon oxides, hydrogen sulfide, and alcohols on activated carbon as well as zeolites. Sips equation presented in the last section is also recommended but when the behavior in the Henry law region is needed, the Toth equation is the better choice.

#### 6.4.1 Temperature dependence of the toth equation

Like the other equations described so far, the temperature dependence of equilibrium parameters in the Toth equation is required for the purpose of extrapolation or interpolation of equilibrium at other temperatures as well as the purpose of calculating isosteric heat. The parameters  $b$  and  $t$  are temperature dependent, with the parameter  $b$  taking the usual form of the adsorption affinity that is

$$b = b_{\infty} \exp\left(\frac{Q}{R_s T}\right) = b_0 \exp\left[\frac{Q}{R_s T_0} \left(\frac{T_0}{T} - 1\right)\right] \quad (145)$$

where  $b_{\infty}$  is the affinity at infinite temperature,  $b_0$  is that at some reference temperature  $T_0$  and  $Q$  is a measure of the heat of adsorption. The parameter  $t$  and the maximum adsorption capacity can take the following empirical functional form of temperature dependence

$$t = t_0 + \alpha \left(1 - \frac{T_0}{T}\right) \quad (146)$$

$$C_{\mu s} = C_{\mu s,0} \exp\left[\alpha \left(1 - \frac{T_0}{T}\right)\right] \quad (147)$$

The temperature dependence of the parameter  $t$  does not have any sound theoretical footing; however, we would expect that as the temperature increases this parameter will approach unity.

#### 6.5 Keller, staudt and toth's equation

Keller and his co-workers (1996) proposed a new isotherm equation, which is very similar in form to the original Toth equation. The differences between their equation and that of Toth are that:

- the exponent  $\alpha$  is a function of pressure instead of constant as in the case of Toth
- the saturation capacities of different species are different

The form of Keller et al.'s equation is:

$$C_{\mu} = C_{\mu s} \alpha_m \frac{bP}{[1 + (bP)^{\alpha}]^{1/\alpha}} \quad (148)$$

$$\alpha = \frac{1 + \alpha_m \beta P}{1 + \beta P} \quad (149)$$

where the parameter  $\alpha_m$  takes the following equation:

$$\frac{\alpha_m}{\alpha_m^*} = \left( \frac{r}{r^*} \right)^{-D} \quad (150)$$

Here  $r$  is the molecular radius, and  $D$  is the fractal dimension of sorbent surface. The saturation parameter  $C_{\mu s}$ , the affinity constant  $b$ , and the parameter  $\beta$  have the following temperature dependence):

$$C_{\mu s} = C_{\mu s,0} \exp\left[\chi\left(1 - \frac{T}{T_0}\right)\right] \quad (151)$$

$$b = b_0 \exp\left[\frac{Q_1}{R_g T_0} \left(\frac{T_0}{T} - 1\right)\right] \quad (152)$$

$$\beta = \beta_0 \exp\left[\frac{Q_2}{R_g T_0} \left(\frac{T_0}{T} - 1\right)\right] \quad (153)$$

Here the subscript 0 denotes for properties at some reference temperature  $T_0$ . The Keller et al.'s equation contains more parameters than the empirical equations discussed so far. Fitting the Keller et equation with the isotherm data of propane on activated carbon at three temperatures 283, 303 and 333 K, we found the fit is reasonably good, comparable to the good fit observed with Sips and Toth equations. The optimally fitted parameters are:

$C_{\mu s,0}$	16.08 mmole/g
$b_0$	0.9814 (kPa) <sup>-1</sup>
$\beta_0$	3.225 (kPa) <sup>-1</sup>
$\alpha_m$	0.4438
$Q_1/RT_0$	10.94
$Q_2/RT_0$	-0.2863
$\chi$	0.0002476

Table 5. The parameters for Keller, Staudt and Toth's Equation

### 6.6 Dubinin-radushkevich equation

The empirical equations dealt with so far, Freundlich, Sips, Toth, Unilan and Keller et al., are applicable to supercritical as well as subcritical vapors. In this section we present briefly a semi-empirical equation which was developed originally by Dubinin and his co-workers for sub critical vapors in microporous solids, where the adsorption process follows a pore filling mechanism. Hobson and co-workers and Earnshaw and Hobson (1968) analysed the data of argon on Corning glass in terms of the Polanyi potential theory. They proposed an equation relating the amount adsorbed in equivalent liquid volume ( $V$ ) to the adsorption potential

$$A = R_g T \ln\left(\frac{P_0}{P}\right) \quad (154)$$

where  $P_0$  is the vapor pressure. The premise of their derivation is the functional form  $V(A)$  which is independent of temperature. They chose the following functional form:

$$\ln V = \ln V_0 BA^2 \tag{155}$$

where the logarithm of the amount adsorbed is linearly proportional to the square of the adsorption potential. Eq. (155) is known as the Dubinin-Radushkevich (DR) equation. Writing this equation explicitly in terms of pressure, we have:

$$V = V_0 \exp \left[ -\frac{1}{(\beta E_0)^2} \left( R_s T \ln \frac{P}{P_0} \right)^2 \right] \tag{156}$$

where  $E_0$  is called the solid characteristic energy towards a reference adsorbate. Benzene has been used widely as the reference adsorbate. The parameter  $\beta$  is a constant which is a function of the adsorptive only. It has been found by Dubinin and Timofeev (1946) that this parameter is proportional to the liquid molar volume. Fig. 14 shows plots of the DR equation versus the reduced pressure with  $E/R_gT$  as the varying parameter (Foo K.Y., Hameed B.H., 2009).

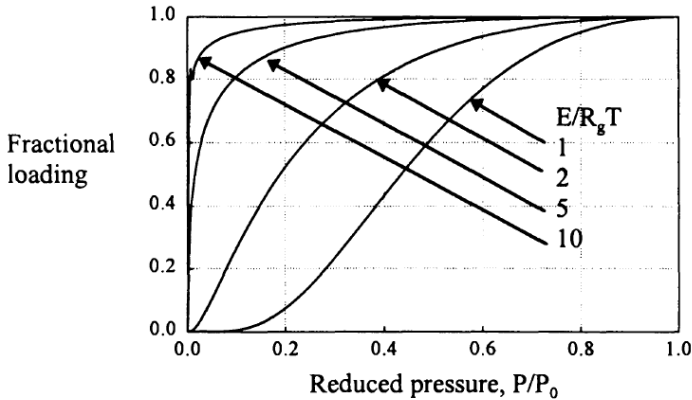


Fig. 14. Plots of the DR equation versus the reduced pressure

We see that as the characteristic energy increases the adsorption is stronger as the solid has stronger energy of interaction with adsorbate. One observation in that equation is that the slope of the adsorption isotherm at zero loading is not finite, a violation of the thermodynamic requirement Eq. (156) when written in terms of amount adsorbed (mole/g) is:

$$C_\mu = C_{\mu s} \exp \left[ -\frac{1}{(\beta E_0)^2} \left( R_s T \ln \frac{P}{P_0} \right)^2 \right] \tag{157}$$

Where the maximum adsorption capacity is:

$$C_{\mu S} = \frac{W_0}{V_M(T)} \quad (158)$$

The parameter  $W_0$  is the micropore volume and  $V_M$  is the liquid molar volume. Here we have assumed that the state of adsorbed molecule in micropores behaves like liquid. Dubinin-Radushkevich equation (157) is very widely used to describe adsorption isotherm of sub-critical vapors in microporous solids such as activated carbon and zeolite. One debatable point in such equation is the assumption of liquid-like adsorbed phase as one could argue that due to the small confinement of micropore adsorbed molecules experience stronger interaction forces with the micropore walls, the state of adsorbed molecule could be between liquid and solid. The best utility of the Dubinin-Radushkevich equation lies in the fact that the temperature dependence of such equation is manifested in the adsorption potential  $A$ , defined as in eq. (154), that is if one plots adsorption data of different temperatures as the logarithm of the amount adsorbed versus the square of adsorption potential, all the data should lie on the same curve, which is known as the characteristic curve. The slope of such curve is the inverse of the square of the characteristic energy  $E = \beta E_0$ . To show the utility of the DR equation, we fit eq. (157) to the adsorption data of benzene on activated carbon at three different temperatures, 283, 303 and 333 K. The data are tabulated in Table 10.6 and presented graphically in Figure 10.15.

283 K		303 K		333 K	
P (kPa)	$C_{\mu}$ (mmole/g)	P (kPa)	$C_{\mu}$ (mmole/g)	P (kPa)	$C_{\mu}$ (mmole/g)
0.0133	1.6510	0.0001	0.4231	0.0010	0.4231
0.0933	3.2470	0.0002	0.8462	0.0267	0.8450
0.2932	3.8750	0.0133	1.1110	0.0533	1.1090
0.6798	4.2560	0.0267	1.4060	0.0933	1.4030
1.5590	4.5270	0.0666	1.9540	0.2532	1.9460
2.6520	4.6600	0.0933	2.1660	0.3732	2.1520
4.2920	4.8060	0.1599	2.5090	0.6531	2.4870
6.3580	4.9480	0.3466	2.9730	1.3330	2.9290
8.7440	5.0480	0.6931	3.4310	2.6120	3.3470
10.0200	5.0840	1.2800	3.7610	4.3590	3.6260
		2.8260	4.1490	7.6640	3.9380
		3.9320	4.2770	9.4770	4.0370
		6.6380	4.4410	11.5600	4.2340
		8.5570	4.5370		
		10.410	4.5880		

Table 6. Adsorption data of benzene on activated carbon

The vapor pressure and the liquid molar volume of benzene are given in the following table.

T (K)	$P_0$ (kPa)	$v_M$ (cc/mmole)
303	16.3	0.0900
333	52.6	0.0935
363	150	0.0970

Table 7. Vapor pressure and liquid molar volume of benzene

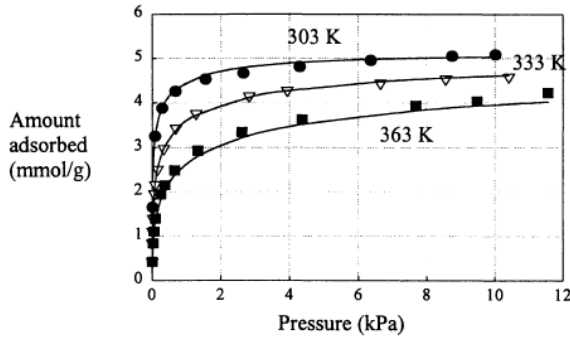


Fig. 15. Fitting the benzene/ activated carbon data with the DR equation

By fitting the equilibria data of all three temperatures simultaneously using the ISOFIT1 program, we obtain the following optimally fitted parameters:  $W_0 = 0.45 \text{ cc/g}$ ,  $E = 20,000 \text{ Joule/mole}$ . Even though only one value of the characteristic energy was used in the fitting of the three temperature data, the fit is very good as shown in Fig. 15, demonstrating the good utility of this equation in describing data of sub-critical vapors in microporous solids.

**6.7 Jovanovich equation**

Of lesser use in physical adsorption is the Jovanovich equation. It is applicable to mobile and localized adsorption (Hazlitt et al, 1979). Although it is not as popular as the other empirical equations proposed so far, it is nevertheless a useful empirical equation:

$$1 - \theta = \exp \left[ -a \left( \frac{P}{P_0} \right) \right] \tag{159}$$

or written in terms of the amount adsorbed:

$$C_\mu = C_{\mu s} [1 - e^{-bP}] \tag{160}$$

where

$$b = b_\infty \exp(Q / R_g T) \tag{161}$$

At low loading, the above equation will become  $C_\mu \approx (C_{\mu s} b)P = HP$ . Thus, this equation reduces to the Henry's law at low pressure. At high pressure, it reaches the saturation limit. The Jovanovich equation has a slower approach toward the saturation than that of the Langmuir equation.

**6.8 Temkin equation**

Another empirical equation is the Temkin equation proposed originally by Slygin and Frumkin (1935) to describe adsorption of hydrogen on platinum electrodes in acidic solutions (chemisorption systems). The equation is (Rudzinski and Everett, 1992):

$$v(P) = C \ln(c.P) \tag{162}$$

where  $C$  and  $c$  are constants specific to the adsorbate-adsorbent pairs. Under some conditions, the Temkin isotherm can be shown to be a special case of the Unilan equation (162).

### 6.9 BET<sup>2</sup> isotherm

All the empirical equations dealt with are for adsorption with "monolayer" coverage, with the exception of the Freundlich isotherm, which does not have a finite saturation capacity and the DR equation, which is applicable for micropore volume filling. In the adsorption of sub-critical adsorbate, molecules first adsorb onto the solid surface as a layering process, and when the pressure is sufficiently high (about 0.1 of the relative pressure) multiple layers are formed. Brunauer, Emmett and Teller are the first to develop a theory to account for this multilayer adsorption, and the range of validity of this theory is approximately between 0.05 and 0.35 times the vapor pressure. In this section we will discuss this important theory and its various versions modified by a number of workers since the publication of the BET theory in 1938. Despite the many versions, the BET equation still remains the most important equation for the characterization of mesoporous solids, mainly due to its simplicity. The BET theory was first developed by Brunauer et al. (1938) for a *flat* surface (no curvature) and there is *no limit* in the number of layers which can be accommodated on the surface. This theory made use of the same assumptions as those used in the Langmuir theory, that is the surface is energetically homogeneous (adsorption energy does not change with the progress of adsorption in the same layer) and there is no interaction among adsorbed molecules. Let  $S_0$ ,  $S_1$ ,  $S_2$  and  $S_n$  be the surface areas covered by no layer, one layer, two layers and  $n$  layers of adsorbate molecules, respectively (Fig. 16).

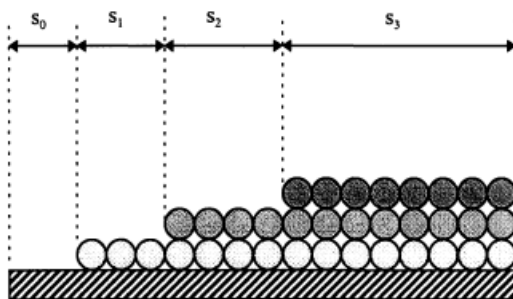


Fig. 16. Multiple layering in BET theory

The concept of kinetics of adsorption and desorption proposed by Langmuir is applied to this multiple layering process, that is the rate of adsorption on any layer is equal to the rate of desorption from that layer. For the first layer, the rates of adsorption onto the free surface and desorption from the first layer are equal to each other:

$$a_1 P s_0 = b_1 s_1 \exp\left(\frac{-E_1}{R_g T}\right) \quad (163)$$

<sup>2</sup> Brunauer, Emmett and Teller

where  $a_1$ ,  $b_1$  and  $E_1$  are constant, independent of the amount adsorbed. Here  $E_1$  is the interaction energy between the solid and molecule of the first layer, which is expected to be higher than the heat of vaporization. Similarly, the rate of adsorption onto the first layer must be the same as the rate of evaporation from the second layer, that is:

$$a_2 P s_0 = b_2 s_2 \exp\left(\frac{-E_2}{R_s T}\right) \tag{164}$$

The same form of equation then can be applied to the next layer, and in general for the  $i$ -th layer, we can write

$$a_i P s_{i-1} = b_i s_i \exp\left(\frac{-E_i}{R_s T}\right) \tag{165}$$

The total area of the solid is the sum of all individual areas, that is

$$S = \sum_{i=0}^{\infty} s_i \tag{166}$$

Therefore, the volume of gas adsorbed on surface covering by one layer of molecules is the fraction occupied by one layer of molecules multiplied by the monolayer coverage  $V_m$ :

$$V_1 = V_m \left(\frac{s_1}{S}\right) \tag{166}$$

The volume of gas adsorbed on the section of the surface which has two layers of molecules is:

$$V_2 = V_m \left(\frac{2s_2}{S}\right) \tag{167}$$

The factor of 2 in the above equation is because there are two layers of molecules occupying a surface area of  $s_2$  (Fig. 16). Similarly, the volume of gas adsorbed on the section of the surface having " $i$ " layers is:

$$V_i = V_m \left(\frac{i s_i}{S}\right) \tag{168}$$

Hence, the total volume of gas adsorbed at a given pressure is the sum of all these volumes:

$$V = \frac{V_m}{S} \sum_{i=0}^{\infty} i s_i = V_m \frac{\sum_{i=0}^{\infty} i s_i}{\sum_{i=0}^{\infty} s_i} \tag{169}$$

To explicitly obtain the amount of gas adsorbed as a function of pressure, we have to express  $S_i$  in terms of the gas pressure. To proceed with this, we need to make a further assumption beside the assumptions made so far about the ideality of layers (so that Langmuir kinetics could be applied). One of the assumptions is that the heat of adsorption of the second and subsequent layers is the same and equal to the heat of liquefaction,  $E_L$ .

$$E_2 = E_3 = \dots = E_i = \dots = E_L \quad (170)$$

The other assumption is that the ratio of the rate constants of the second and higher layers is equal to each other, that is:

$$\frac{b_2}{a_2} = \frac{b_3}{a_3} = \dots = \frac{b_i}{a_i} = g \quad (171)$$

where the ratio  $g$  is assumed constant. This ratio is related to the vapor pressure of the adsorbate. With these two additional assumptions, one can solve the surface coverage that contains one layer of molecule ( $s_1$ ) in terms of  $s_0$  and pressure as follows:

$$s_1 = \frac{a_1}{b_1} = P s_0 \exp(\varepsilon_1) \quad (172)$$

where  $\varepsilon_1$  is the reduced energy of adsorption of the first layer, defined as

$$\varepsilon_1 = \frac{E_1}{R_g T} \quad (173)$$

Similarly the surface coverage of the section containing  $i$  layers of molecules is:

$$s_i = \frac{a_i}{b_i} s_0 g \cdot \exp(\varepsilon_1 - \varepsilon_2) \left[ \left( \frac{P}{g} \right) \exp \varepsilon_L \right]^i \quad (174)$$

for  $i = 2, 3, \dots$ , where  $E_L$  is the reduced heat of liquefaction

$$\varepsilon_L = \frac{E_L}{R_g T} \quad (173)$$

Substituting these surface coverage into the total amount of gas adsorbed (eq. 169), we obtain:

$$\frac{V}{V_m} = \frac{C s_0 \sum_{i=0}^{\infty} i \cdot x^i}{s_0 (1 + C \sum_{i=1}^{\infty} x^i)} \quad (174)$$

where the parameter  $C$  and the variable  $x$  are defined as follows:

$$y = \frac{a_1}{b_1} P \exp \varepsilon_1 \quad (175)$$

$$x = \frac{P}{g} \exp \varepsilon_L \quad (176)$$

$$C = \frac{y}{x} = \frac{a_1 g}{b_1} e^{(\varepsilon_1 - \varepsilon_L)} \quad (177)$$



By using the following formulas (Abramowitz and Stegun, 1962)

$$\sum_{i=1}^{\infty} x^i = \frac{x}{1-x}; \sum_{i=1}^{\infty} ix^i = \frac{x}{(1-x)^2} \tag{178}$$

eq. (174) can be simplified to yield the following form written in terms of C and x:

$$\frac{V}{V_m} = \frac{Cx}{(1-x)(1-x+Cx)} \tag{179}$$

Eq. (179) can only be used if we can relate x in terms of pressure and other known quantities. This is done as follows. Since this model allows for infinite layers on top of a flat surface, the amount adsorbed must be infinity when the gas phase pressure is equal to the vapor pressure, that is  $P = P_0$  occurs when  $x = 1$ ; thus the variable x is the ratio of the pressure to the vapor pressure at the adsorption temperature:

$$x = \frac{P}{P_0} \tag{180}$$

With this definition, eq. (179) will become what is now known as the famous BET equation containing two fitting parameters, C and  $V_m$ :

$$\frac{V}{V_m} = \frac{CP}{(P_0 - P)(1 + (C - 1)(P / P_0))} \tag{181}$$

Fig. 17 shows plots of the BET equation (181) versus the reduced pressure with C being the varying parameter. The larger is the value of C, the sooner will the multilayer form and the convexity of the isotherm increases toward the low pressure range.

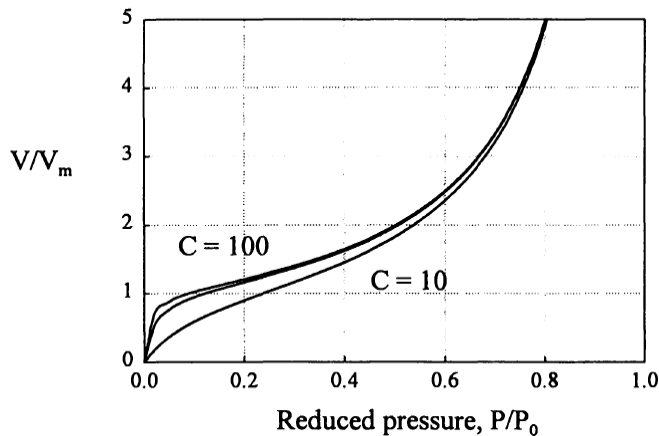


Fig. 17. Plots of the BET equation versus the reduced pressure (C = 10,50, 100)

Equating eqs.(180) and (176), we obtain the following relationship between the vapor pressure, the constant g and the heat of liquefaction:

$$P_0 = g \cdot \exp\left(-\frac{E_L}{R_s T}\right) \quad (182)$$

Within a narrow range of temperature, the vapor pressure follows the Clausius-Clapeyron equation, that is

$$P_0 = \alpha \cdot \exp\left(-\frac{E_L}{R_s T}\right) \quad (183)$$

Comparing this equation with eq.(182), we see that the parameter  $g$  is simply the pre-exponential factor in the Clausius-Clapeyron vapor pressure equation. It is reminded that the parameter  $g$  is the ratio of the rate constant for desorption to that for adsorption of the second and subsequent layers, suggesting that these layers condense and evaporate similar to the bulk liquid phase. The pre-exponential factor of the constant  $C$  (eq.177)

$$\frac{a_i g}{b_1} = \frac{a_i b_j}{b_1 a_j}; \text{ for } j > 1 \quad (184)$$

can be either greater or smaller than unity (Brunauer et al., 1967), and it is often assumed as unity without any theoretical justification. In setting this factor to be unity, we have assumed that the ratio of the rate constants for adsorption to desorption of the first layer is the same as that for the subsequent layers at infinite temperature. Also by assuming this factor to be unity, we can calculate the interaction energy between the first layer and the solid from the knowledge of  $C$  (obtained by fitting of the isotherm equation 3.3-18 with experimental data) The interaction energy between solid and adsorbate molecule in the first layer is always greater than the heat of adsorption; thus the constant  $C$  is a large number (usually greater than 100).

## 7. BDDT (Brunauer, Deming, Denting, Teller) classification

The theory of BET was developed to describe the multilayer adsorption. Adsorption in real solids has given rise to isotherms exhibiting many different shapes. However, five isotherm shapes were identified (Brunauer et al., 1940) and are shown in Fig.19. The following five systems typify the five classes of isotherm.

**Type 1:** Adsorption of oxygen on charcoal at -183 °C

**Type 2:** Adsorption of nitrogen on iron catalysts at -195°C (many solids fall into this type).

**Type 3:** Adsorption of bromine on silica gel at 79°C, water on glass

**Type 4:** Adsorption of benzene on ferric oxide gel at 50°C

**Type 5:** Adsorption of water on charcoal at 100°C

Type I isotherm is the Langmuir isotherm type (monolayer coverage), typical of adsorption in microporous solids, such as adsorption of oxygen in charcoal. Type II typifies the BET adsorption mechanism. Type III is the type typical of water adsorption on charcoal where the adsorption is not favorable at low pressure because of the nonpolar (hydrophobic) nature of the charcoal surface. At sufficiently high pressures, the adsorption is due to the capillary condensation in mesopores. Type IV and type V are the same as types II and III with the exception that they have finite limit as  $P \rightarrow P_0$  due to the finite pore volume of porous solids.

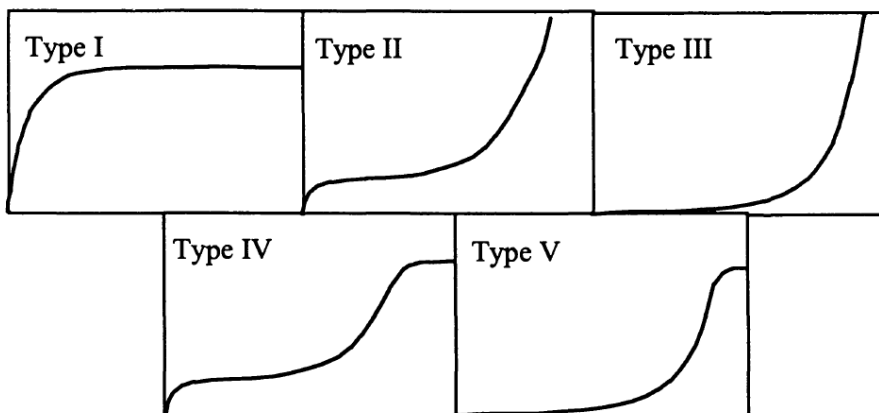


Fig. 19. BDDT classification of five isotherm shapes

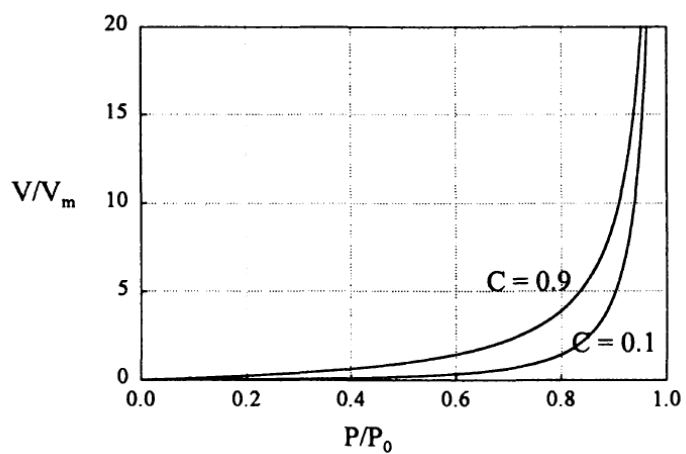


Fig. 20. Plots of the BET equation when  $C < 1$

The BET equation developed originally by Brunauer et al. (1938) is able to describe type I to type III. The type III isotherm can be produced from the BET equation when the forces between adsorbate and adsorbent are smaller than that between adsorbate molecules in the liquid state (i.e.  $E_s < E_L$ ). Fig. 20 shows such plots for the cases of  $C = 0.1$  and  $0.9$  to illustrate type III isotherm.

The BET equation does not cover the last two types (IV and V) because one of the assumptions of the BET theory is the allowance for infinite layers of molecules to build up on top of the surface. To consider the last two types, we have to limit the number of layers which can be formed above a solid surface. (Foo K.Y., Hameed B.H., 2009), (Moradi O. , et al. 2003). (Hirschfelder, and et al. 1954).

## 8. Conclusion

In following chapter thermodynamics of interface is frequently applied to derive relations between macroscopic parameters. Nevertheless, this chapter is included as a reminder. It presents a consist summary of thermodynamics principles that are relevant to interfaces in view of the topics discussed such as thermodynamics for open and close systems, Equilibrium between phases, Physical description of a real liquid interface, Surface free energy and surface tension of liquids, Surface equation of state, Relation of van der Waals constants with molecular pair potentials and etc in forthcoming and special attention is paid to heterogeneous systems that contain phase boundaries.

## 9. References

- Adamson, A.W. and Gast, A.P. (1997) *Physical Chemistry of Surfaces* (6th edn). Wiley, New York, USA.
- Abdullah M.A., Chiang L., Nadeem M., Comparative evaluation of adsorption kinetics and isotherms of a natural product removal by Amberlite polymeric adsorbents, *Chem. Eng. J.* 146 (3) (2009) 370–376.
- Ahmaruzzaman M. d., Adsorption of phenolic compounds on low-cost adsorbents: a review, *Adv. Colloid Interface Sci.* 143 (1–2) (2008) 48–67.
- Adam, N.K. (1968) *The Physics and Chemistry of Surfaces*. Dover, New York.
- Atkins, P.W. (1998) *Physical Chemistry* (6th edn). Oxford University Press, Oxford.
- Aveyard, R. and Haydon, D.A. (1973) *An Introduction to the Principles of Surface Chemistry*. Cambridge University Press, Cambridge.
- Dabrowski A., Adsorption—from theory to practice, *Adv. Colloid Interface Sci.* 93 (2001) 135–224.
- Dubinin M. M., Radushkevich L.V., The equation of the characteristic curve of the activated charcoal, *Proc. Acad. Sci. USSR Phys. Chem. Sect.* 55 (1947) 331–337.
- Erbil, H.Y. (1997) Interfacial Interactions of Liquids. In Birdi, K.S. (ed.). *Handbook of Surface and Colloid Chemistry*. CRC Press, Boca Raton.
- Foo K.Y., Hameed B.H., Recent developments in the preparation and regeneration of activated carbons by microwaves, *Adv. Colloid Interface Sci.* 149 (2009) 19–27.

- Foo K.Y., Hameed B.H., A short review of activated carbon assisted electrosorption process: An overview, current stage and future prospects, *J. Hazard. Mater.* 171 (2009) 54–60.
- Hirschfelder, J.O., Curtiss, C.F. and Bird, R.B. (1954) *Molecular Theory of Gases and Liquids*. Wiley, New York.
- Israelachvili, J. (1991) *Intermolecular & Surface Forces* (2nd edn). Academic Press, London.
- Levine, I.N. (1990) *Physical Chemistry* (3rd edn). McGraw-Hill, New York.
- Lyklema, J. (2005) *Fundamentals of interface and colloid science*, Elsevier Ltd.
- Lyklema, L. (1991) *Fundamentals of Interface and Colloid Science* (vols. I and II). Academic Press, London.
- Keller J.U., (2005) *Gas adsorption equilibria, Experimental Methods and Adsorptive Isotherms*, Springer Science, USA.
- Koopal L.K., Van Riemsdijk W. H., Wit J.C.M., Benedetti M.F., Analytical isotherm equation for multicomponent adsorption to heterogeneous surfaces, *J. Colloid Interface Sci.* 166 (1994) 51–60.
- Miladinovic N., Weatherley L.R., Intensification of ammonia removal in a combined ion-exchange and nitrification column, *Chem. Eng. J.* 135 (2008) 15–24.
- Moradi, O, Modarress, H.; Norouzi, M.; Effect of Lysozyme Concentration, pH and Ionic Strength and Its Adsorption on HEMA and AA Contact Lenses, "Iranian Polymer Journal", Vol.12, No.6, 477-484, 2003.
- Moradi, O, Modarress, H.; Norouzi, M.; Experimental Study of Albumin and Lysozyme Adsorption onto Acrylic Acid (AA) and 2-hydroxyethyl methacrylate (HEMA) surfaces, *Journal of colloid and interface science*, Vol. 261, No. 1, 16-19, 2004.
- Moradi, O. Heavy metal [Cr(VI), Cd(II) and Pb(II)] ions removal by modified jute: Characterization and Modeling, *J. Theoretical and Physical Chemistry*, Vol. 4, No. 3, 163-168, 2007.
- Moradi O., Aghaie M., Zare K., Monajjemi M., Aghaie, H., The Study of Adsorption Characteristics Cu<sup>2+</sup> and Pb<sup>2+</sup> Ions onto PHEMA and P(MMA-HEMA) Surfaces from Aqueous Single Solution, *Journal of Hazardous Materials*, 673-679, 170, 2009.
- Moradi O., Aghaie M., Zare K., Monajjemi M., Aghaie, H., The Studies of Equilibrium and Thermodynamic Adsorption of Pb(II), Cd(II) and Cu(II) Ions From Aqueous Solution onto SWCNTs and SWCNT -COOH Surfaces, Fullerenes, Nanotubes and Carbon Nanostructures, 18, 285-302, 2010.
- Mohan D., Pittman Jr C.U., Activated carbons and low cost adsorbents for remediation of tri- and hexavalent chromium from water, *J. Hazard. Mater.* 137 (2006) 762–811.
- Murrell, J.N. and Jenkins, A.D. (1994) *Properties of Liquids and Solutions* (2nd edn). Wiley.
- Ng J.C.Y., Cheung W.H., McKay G., Equilibrium studies of the sorption of Cu(II) ions onto chitosan, *J. Colloid Interface Sci.* 255 (2002) 64–74.
- Norde, W., (2003) *colloids and interfaces in life science*, Marcel Dekker, Inc., New York, USA.
- Pitzer, K.S. and Brewer L. (1961) *Thermodynamics* (2nd edn). McGraw-Hill, New York.
- Prausnitz, J.M., Lichtenthaler, R.N. and Azevedo E.G. (1999) *Molecular Thermodynamics of Fluid-Phase Equilibria* (3rd edn). Prentice Hall, Englewood Cliffs.
- Tempkin M. I., Pyzhev V., Kinetics of ammonia synthesis on promoted iron catalyst, *Acta Phys. Chim. USSR* 12 (1940) 327–356.

Scatchard, G. (1976) *Equilibrium in Solutions & Surface and Colloid Chemistry*. Harvard University Press, Cambridge.

Zeldowitsch J., Adsorption site energy distribution, *Acta Phys. Chim. URSS* 1 (1934) 961-973.

# Exergy, the Potential Work

Mofid Gorji-Bandpy  
*Noshirovani University of Technology*  
*Iran*

## 1. Introduction

The exergy method is an alternative, relatively new technique based on the concept of exergy, loosely defined as a universal measure of the work potential or quality of different forms of energy in relation to a given environment. An exergy balance applied to a process or a whole plant tells us how much of the usable work potential, or exergy, supplied as the input to the system under consideration has been consumed (irretrievably lost) by the process. The loss of exergy, or irreversibility, provides a generally applicable quantitative measure of process inefficiency. Analyzing a multi component plant indicates the total plant irreversibility distribution among the plant components, pinpointing those contributing most to overall plant inefficiency (Gorji-Bandpy&Ebrahimian, 2007; Gorji-Bandpy et al., 2011)

Unlike the traditional criteria of performance, the concept of irreversibility is firmly based on the two main laws of thermodynamics. The exergy balance for a control region, from which the irreversibility rate of a steady flow process can be calculated, can be derived by combining the steady flow energy equation (First Law) with the expression for the entropy production rate (Second Law).

Exergy analysis of the systems, which analyses the processes and functioning of systems, is based on the second law of thermodynamics. In this analysis, the efficiency of the second law which states the exact functionality of a system and depicts the irreversible factors which result in exergy loss and efficiency decrease, is mentioned. Therefore, solutions to reduce exergy loss will be identified for optimization of engineering installations (Ebadi&Gorji-Bandpy, 2005). Considering exergy as the amount of useful work which is brought about, as the system and the environment reach a balance due to irreversible process, we can say that the exergy efficiency is a criterion for the assessment of the systems. Because of the irreversibility of the heating processes, the resulting work is usually less than the maximum amount and by analyzing the work losses of the system, system problems are consequently defined. Grossman diagrams, in which any single flow is defined by its own exergy, are used to determine the flow exergy in the system (Bejan, 1988). The other famous flow exergy diagrams have been published by Keenan (1932), Reistad (1972) and Thirumaleshwar (1979). The famous diagrams of air exergy were published by Moran (1982) and Brodianskii (1973). Brodianskii (1973), Kotas (1995) and Szargut et al. (1988) have used the exergy method for thermal, chemical and metallurgical analysis of plants. Analysis of the technical chains of processes and the life-cycle of a product were respectively done by Szargut et al. (1988) and Comelissen and Hirs (1999). The thermoeconomy field, or in other words, interference of economical affairs in analyzing exergy, has been studied by Bejan (1982).

In this paper, the cycle of a power plant and its details, with two kind fuels, natural gas and diesel, have been analysed at its maximum load and the two factors, losses and exergy efficiency which are the basic factors of systems under study have been analysed.

## 2. Methodology

When a system is thermodynamically studied, based on the first principle of thermodynamics, the amount of energy is constant during the transfer or exchange and also, based on the second principle of thermodynamics, the degree of energy is reduced and the potential for producing work is lessened. But none of the mentioned principles are able to determine the exact magnitude of work potential reduction, or in other words, to analyse the energy quality. For an open system which deals with some heat resources, the first and second principles are written as follows (Bejan, 1988):

$$\frac{dE}{dt} = \sum_{i=0}^n \dot{Q}_i - \dot{W} + \sum_{in} \dot{m}h^0 - \sum_{out} \dot{m}h^0 \quad (1)$$

$$\dot{S}_{gen} = \frac{dS}{dt} - \sum_{i=0}^n \frac{\dot{Q}_i}{T_i} - \sum_{in} \dot{m}s + \sum_{out} \dot{m}s \geq 0 \quad (2)$$

In the above equations, enthalpy,  $h^0$ , is  $h + (V^2 / 2) + gz$ ,  $T_0$  is the surrounding temperature,  $E$ , internal energy,  $S$ , entropy, and  $\dot{W}$  and  $\dot{Q}$  are the rates of work and heat transfer.

For increasing the work transfer rate ( $\dot{W}$ ), consider the possibility of changes in design of system. Assumed that all the other interactions that are specified around the system ( $\dot{Q}_1, \dot{Q}_2, \dots, \dot{Q}_n$ , inflows and outflows of enthalpy and entropy) are fixed by design and only  $\dot{Q}_0$  floats in order to balance the changes in  $\dot{W}$ . If we eliminate  $\dot{Q}_0$  from equations (1) and (2), we will have (Bejan, 1988):

$$\dot{W} = -\frac{d}{dt}(E - T_0S) + \sum_{i=0}^n \left(1 - \frac{T_0}{T_i}\right) \dot{Q}_i + \sum_{in} \dot{m}(h^0 - T_0s) - \sum_{out} \dot{m}(h^0 - T_0s) - T_0\dot{S}_{gen} \quad (3)$$

When the process is reversible ( $\dot{S}_{gen} = 0$ ), the rate of work transfer will be maximum and therefore we will have:

$$\dot{W} = \dot{W}_{rev} - T_0\dot{S}_{gen} \quad (4)$$

Combination of the two principles results in the conclusion that whenever a system functions irreversibly, the work will be eliminated at a rate relative to the one of the entropy. The eliminated work caused by thermodynamic irreversibility,  $(\dot{W}_{rev} - \dot{W})$  is called "the exergy lost". The ratio of the exergy lost to the entropy production, or the ratio of their rates results in the principle of lost work:

$$\dot{W}_{lost} = T_0\dot{S}_{gen} \quad (5)$$

Since exergy is the useful work which derived from a material or energy flow, the exergy of work transfer,  $\dot{E}_w$ , would be given as (Bejan, 1988):



$$\begin{aligned} \dot{E}_w = \dot{W} - P_0 \frac{dV}{dt} = -\frac{d}{dt}(E + P_0V - T_0S) + \sum_{i=1}^n \left(1 - \frac{T_0}{T_i}\right) \dot{Q}_{in} \\ + \sum_{in} \dot{m}(h^0 - T_0s) - \sum_{out} \dot{m}(h^0 - T_0s) - T_0 \dot{S}_{gen} \end{aligned} \quad (6)$$

In most of the systems with incoming and outgoing flows which are considered of great importance, there is no atmospheric work,  $(P_0(dV / dt))$  and  $\dot{W}$  is equal to  $\dot{E}_w$  (Bejan, 1988):

$$\begin{aligned} (\dot{E}_w)_{rev} = \dot{W}_{rev} - P_0 \frac{dV}{dt} = -\frac{d}{dt}(E + P_0V - T_0S) + \sum_{i=1}^n \left(1 - \frac{T_0}{T_i}\right) \dot{Q}_{in} \\ + \sum_{in} \dot{m}(h^0 - T_0s) - \sum_{out} \dot{m}(h^0 - T_0s) \end{aligned} \quad (7)$$

The exergy lost, which was previously defined as the difference between the maximum rate of work transfer and rate of the real work transfer, can also be mentioned in another way, namely, the difference between the corresponding parameters and the available work (Figure 1):

$$\dot{W}_{lost} = (\dot{E}_w)_{rev} - \dot{E}_w = (\dot{E}_w)_{lost} \quad (8)$$

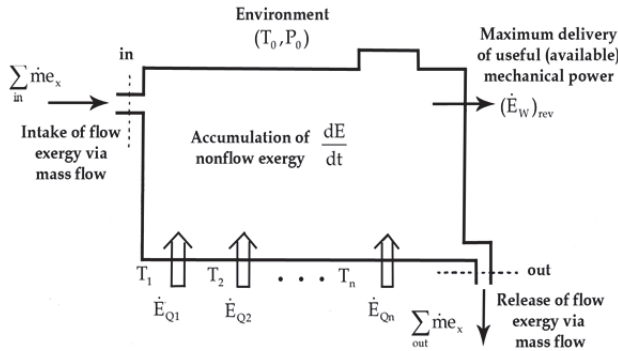


Fig. 1. Exergy transfer via heat transfer

In equation (6), the exergy transfer caused by heat transfer or simply speaking, the heat transfer exergy will be:

$$\dot{E}_Q = \dot{Q} \left(1 - \frac{T_0}{T}\right) \quad (9)$$

Using equation (1), the flow availability will be introduces as:

$$b = h^0 - T_0s \quad (10)$$

In installation analysis which functions uniformly, the properties do not changes with time and the stagnation exergy term will be zero, in equation (6):

$$\dot{E}_w = \sum_{i=0}^n (\dot{E}_Q)_i + \sum_{in} \dot{m}b - \sum_{out} \dot{m}b - T_0 \dot{S}_{gen} \quad (11)$$

The flow exergy of any fluid is defined as:

$$e_x = b - b_0 = h^0 - h_0^0 - T_0(s - s_0) \quad (12)$$

Substituting this definition into equation (11), we will have:

$$\dot{E}_w = \sum_{i=0}^n (\dot{E}_Q)_i + \sum_{in} \dot{m}e_x - \sum_{out} \dot{m}e_x - T_0 \dot{S}_{gen} \quad (13)$$

The flow exergy ( $e_x$ ) is the difference between the availabilities of a flow (b), in a specific condition and in the restricted dead state (in balance with the environment). Equation (13) is used to balance the exergy of uniform flow systems. The mechanisms which lead to the production of entropy and the elimination of exergy are listed as follows:

- heat transfer caused by limited temperature difference (Bejan, 1988):

$$\dot{S}_{gen} = \frac{\dot{Q}}{T_H T_L} (T_H - T_L) \quad (14)$$

- frictional flow (Reistad, 1972):

$$\dot{S}_{gen} = \dot{m} \int_{out}^{in} \left( \frac{v}{T} \right)_{h=const.} dp \quad (15)$$

- combining (Stepanov, 1955):

$$\frac{1}{\dot{m}_3} \dot{S}_{gen} \cong x \left[ \frac{1}{T} (h_3 - h_1) - \frac{v}{T} (P_3 - P_1) \right] + (1-x) \left[ \frac{1}{T} (h_3 - h_2) - \frac{v}{T} (P_3 - P_2) \right] \quad (16)$$

The efficiency of the second law that determines used exergy is divided into two groups: Elements efficiency (Pump and Turbine) and Cycle efficiency (Thermal efficiency and coefficient of performance). The definition of the efficiency of the second law is (Wark, 1955):

$$\eta_{II} = \frac{\text{availability of useful outgoing}}{\text{incoming availability}} \quad (17)$$

The definition of the efficiency of the second law is more practical for the uniform flow systems, and is determined as follows (Bejan, 1988):

$$\eta_{II} = \frac{\text{outgoing exergy rate}}{\text{incoming exergy rate}} \quad (18)$$

The second law emphasizes the fact that two features of the same concept of energy may have completely different exergies. Therefore, any feature of energy is defined by taking into account its own exergy. The efficiency of the second law will be used in calculating the reduction of ability in performing a certain amount of work.

### 3. Case study

In order to analyse the above theories, the consequences have been analysed on the Shahid Rajaii power plant in Qazvin of Iran. This power plant has an installed capacity of 1000 MW electrical energy, which consists of four 250 MW steam-cycle units (Rankin cycle with reheating and recycling) and has been working since 1994. The major fuel for the plant is the natural gas and is augmented with diesel fuel.

The Shahid Rajaii power plant consists of three turbines: high pressure, medium pressure, and low pressure. The 11-stage high pressure turbine has Curtis stage. The number of the stages in the medium pressure turbine is 11 reactionary stages and in the lower pressure, 2×5 reactionary stages. All of the turbines have a common shaft with a speed of 3000 rpm. The boiler is a natural circulation type in which there is a drum with no top. Other properties of the boiler are that the super-heater is three-staged and that the reheater and the economizer are both two-staged. Figure 2 illustrates the plant diagram overlooking the boiler furnace, cooling towers, attachment (circulation and discharge pumps, blowers, etc), turbine glands condenser and regulator lands, expansion valves and governors and feed water tanks.

In Table 1, properties of water and vapour in the main parts of the cycle have been shown. Maximum losses of cycle water in this plant are 5 kg/hr, which is negligible due to the minute amount. In analyzing the cycle and drawing diagrams, it is assumed that the temperature is 30°C, pressure is 90kPa and relative humidity is 30% as environmental conditions. Other assumptions are:

- kinetic and potential energies are neglected because they are not so important
- all elements of the cycle are considered to be adiabatic
- in this part, the combustion process of the boiler has been ignored.

With Figure 2, the conversion equation and energy balance of boiler will be written as (Jordan, 1997):

$$\dot{m}_1 = \dot{m}_2 \text{ and } \dot{m}_3 = \dot{m}_4 \quad (19)$$

Description	$P(kg/cm^2 \text{ abs})$	$T(^{\circ}C)$
Feed water incoming to boiler	150	247-202
Vapour incoming to HP turbine	140	838
Vapour incoming to reheater	17-40	358-287
Vapour incoming to IP turbine	15.2-37.3	538
Vapour incoming to LP turbine	8-3.5	320
Vapour incoming to condenser	0.241-0.960	64-45
Water outgoing from condenser pump	16-7	63-44

Table 1. Properties of water and vapour in cycle

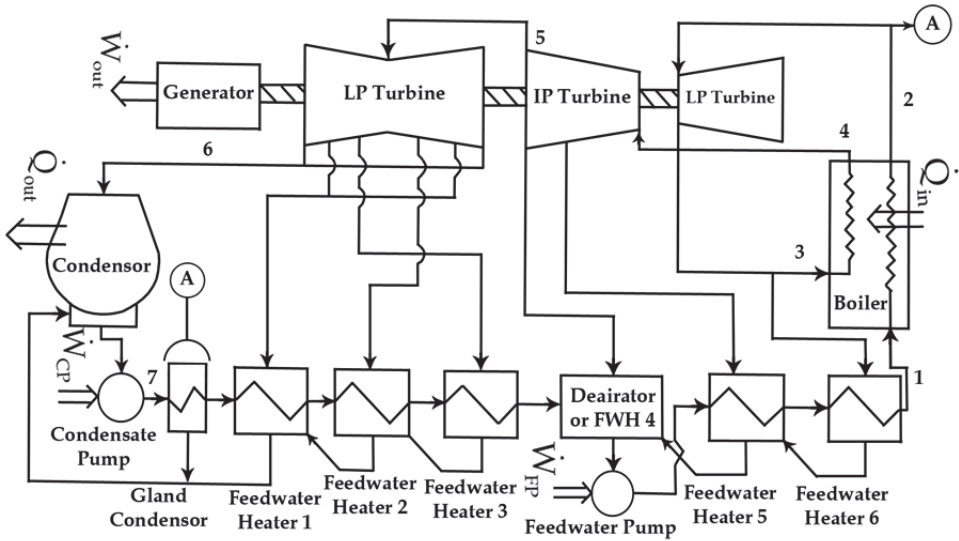


Fig. 2. The diagram of flow cycle of the plant

$$\dot{Q}_{in} = \dot{m}_1(h_2 - h_1) + \dot{m}_4(h_4 - h_3) \tag{20}$$

Energy efficiency is written as (Jordan, 1997):

$$\eta_I = \frac{\dot{W}_{net}}{\dot{Q}_{in}} \tag{21}$$

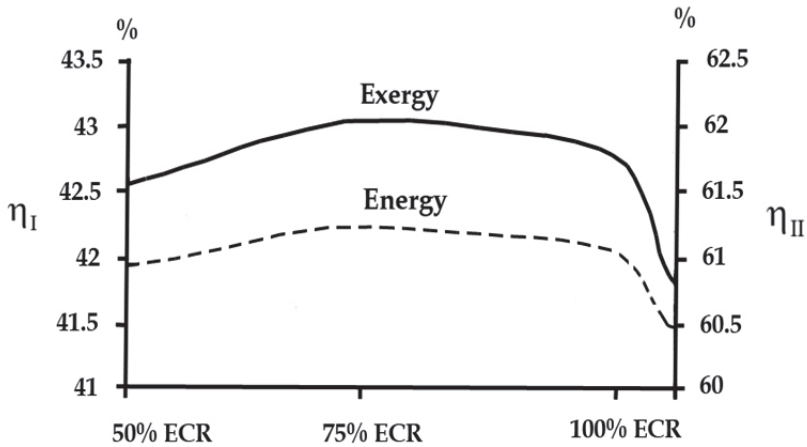


Fig. 3. The diagram of Qazvin power plant efficiency under different loads (natural gas)

If we use the lost work law for the closed cycle:

$$\dot{W}_{net} = \dot{E}_{Qin} - \dot{E}_w \quad (22)$$

$$(\dot{E}_w)_{lost} = \dot{E}_{Qin} - \dot{W}_{net} = T_0 \dot{S}_{gen} \quad (23)$$

Exergy efficiency of plant is:

$$\eta_{II} = \frac{\dot{E}_w}{\dot{E}_{Qin}} = \frac{\dot{W}_{net}}{\dot{E}_{Qin}} \quad (24)$$

Exerting energy and exergy balance equations for the plant cycle, and calculating the energy and efficiencies, Figure 3 consequently results.

As can be seen in Figure 3, under the maximum load, the exergy efficiency is 60.78% and the energy efficiency is 41.38%, relative to different minimum loads. Therefore, boiler analysis is done at maximum load.

### 3.1 Analysing the different elements of the cycle

Using energy balance, which is the basis of exergy balance, and implementing equations (9) and (13) and assuming the warm source temperature to be 950 K, the results of exergy lost and efficiency of all components of the plant cycle are shown in Table 5.

### 3.2 Energy and exergy efficiencies of the plant

In part one, the power plant efficiency has been calculated, overlooking the boiler combustion process and losses under different loads and Figure 3 was therefore mapped out. In order to more accurately calculate the efficiencies, it is necessary to consider the combustion process. The energy efficiency of the plan is the amount of produced net work divided by the fuel energy. In Table 2, percentage of mass for both Natural gas and Diesel fuel in this study has been shown.

Element	Natural gas	Diesel fuel
C	75.624	85
H	23.225	12
O	0	0.4
N	0.206	0.2
S	0	2.4
Ash	0	0
Moisure	0	0
Co <sub>2</sub>	0.945	0

Table 2. Percentage of mass for both natural gas and diesel fuel

$$\eta_I = \frac{\dot{W}_{net}}{\dot{m}_f \times LHV} \quad (25)$$

where LHV is the fuel low heating value. The exergy efficiency of the plant is the amount of outgoing exergy (produced net work) divided by the fuel exergy.

$$\eta_{II} = \frac{\dot{E}_w}{\dot{m}_f e_{ch.f}} \quad (26)$$

Where,  $e_{ch.f}$  is the specific chemical exergy of the fuel.

The natural gas and diesel fuel consumption are respectively 50010 kg/hr and 59130 kg/hr under the maximum load. The low heating values of the natural gas and diesel fuel are 41597 kJ/kg and 48588 kJ/kg. Assuming the natural gas as a perfect gas and using the tables of standard chemical exergy, the chemical exergy of the natural gas is calculated as (Szargut et al., 1988):

$$e_{ch.NG}^0 = \sum Y_i e_{ch,i}^0 = 50403 \frac{kJ}{kg} \quad (27)$$

And implementing the Szargut method, the chemical exergy of diesel fuel is calculated to be 45540 kJ/kg (Szargut et al., 1988):

$$e_{ch.oil} = \left( LHV + \dot{m}_{m,f} h_{fg} \right) \times \left[ 1.0401 + 0.1728 \frac{H}{C} + 0.0432 \frac{O}{C} + 0.2196 \frac{S}{C} \left( 1 - 2.0628 \frac{H}{C} \right) \right] \quad (28)$$

Where LHV is the fuel's low heating value,  $h_{fg}$  and  $m_{m,f}$  are the vaporization temperature of the hot water and the mass of the moisture content and S/C, H/C and O/C are the mass ratio of sulphur, hydrogen and oxygen, to carbon, respectively. The electric power needed for the attachments of the boiler such as fans and pumps is 3.83 MW or 4.28 MW for natural gas and diesel fuels. Feed water pumps and the condenser and other helping elements of the plant also respectively use 9.926 MW and 70.06 MW of the electrical energy. So the produced net work will be:

$$\dot{W}_{net,NG} = 263.53 - 6.926 - 3.83 = 252.774 \text{ MW} \quad (29a)$$

$$\dot{W}_{net,oil} = 261.95 - 7.06 - 4.28 = 250.61 \text{ MW} \quad (29b)$$

The heating and exergy efficiencies of the plant using the two fuels will be:

$$\eta_{I,NG} = \frac{252.774}{(50010 / 3600) \times 11605 (4.1868)} = 37.45\% \quad (30a)$$

$$\eta_{I,oil} = \frac{250.610}{(59130 / 3600) \times 41597} = 36.68\% \quad (30b)$$

$$\eta_{II,NG} = \frac{252.774}{(50010 / 3600) \times 50403} = 36.10\% \quad (31a)$$

$$\eta_{II,oil} = \frac{250.610}{(59130 / 3600) \times 45540} = 33.50\% \quad (31b)$$

As is obviously seen, the heating efficiency of the power plant changes from 36.68% to 37.45% and the exergy efficiency from 33.5% to 36.1%, when natural gas is replaced by diesel fuel. Therefore the exergy efficiency change is greater than that of energy efficiency.

### 3.3 Analysing the boiler

The boiler of this plant is designed based on the natural circulation, and high pressure cold water flow furnace and the water pipes have been appointed vertically. The design pressure of the boiler is 172 kg/cm<sup>2</sup>, the design pressure of the reheating system is 46 kg/cm<sup>2</sup> and the capacity is 840 ton/hrs. Two centrifugal fans (forced draught fan (FDF)) provide the needed air for the combustion.

The boiler is modeled for thermodynamic analysis. The air and gas fans, discharge pumps, and generally, the utilities which are work consuming are not considered in the model. Their effect is the total work which enters the control volume ( $\dot{W}_{m,B}$ ). Also the heaters and the de-super-heaters within the control volume have been ignored. The heat losses to the surrounding environment are introduced as  $\dot{Q}_{out,B}$ . The energy and exergy balance of the boiler referring to the equations (1) and (13) are written as:

$$\dot{m}_f h_f + \dot{W}_{in,B} + \dot{m}_a h_a - \dot{m}_g h_g + \sum_{in} \dot{m}_w h_w - \sum_{out} \dot{m}_w h_w - \dot{Q}_{out,B} = 0 \quad (32)$$

$$\dot{m}_f e_{ch,f} + \dot{W}_{in,B} + \dot{m}_a e_{x,g} - \dot{m}_g e_{x,g} + \sum_{in} \dot{m}_w e_{x,w} - \sum_{out} \dot{m}_w e_{x,w} - T_0 \dot{S}_{gen,B} = 0 \quad (33)$$

Indices a, g and respectively used for air, gas (combustion products) and water (vapour).

The energy and exergy efficiencies of the boiler are defined as:

In Table 3, thermodynamic properties of water and vapour for both Natural gas and Diesel fuel have been summarized.

fuel	property	Feed water	Super heated vapour	Hot reheated vapour	Cold reheated vapour
Natural Gas	$T(^{\circ}C)$	247.7	538	358.1	538
	$P(kPa)$	15640	13730	3820	3660
	$h(kj / kg)$	1075.6	3430.2	3116.2	3534.9
	$e_x(kj / kg)$	241.9	1439.3	1109.8	1345.9
	$\dot{m}(kg / h)$	840000	840000	751210	751210
Diesel Fuel	$T(^{\circ}C)$	247.4	538	357.1	538
	$P(kPa)$	15640	13730	3800	3640
	$h(kj / kg)$	1074.7	3430.2	3115.5	3535.0
	$e_x(kj / kg)$	241.8	1439.0	1109.1	1346.0
	$\dot{m}(kg / h)$	840000	840000	747010	747010

Table 3. Thermodynamic properties of incoming water and outgoing vapour for the boiler at maximum load

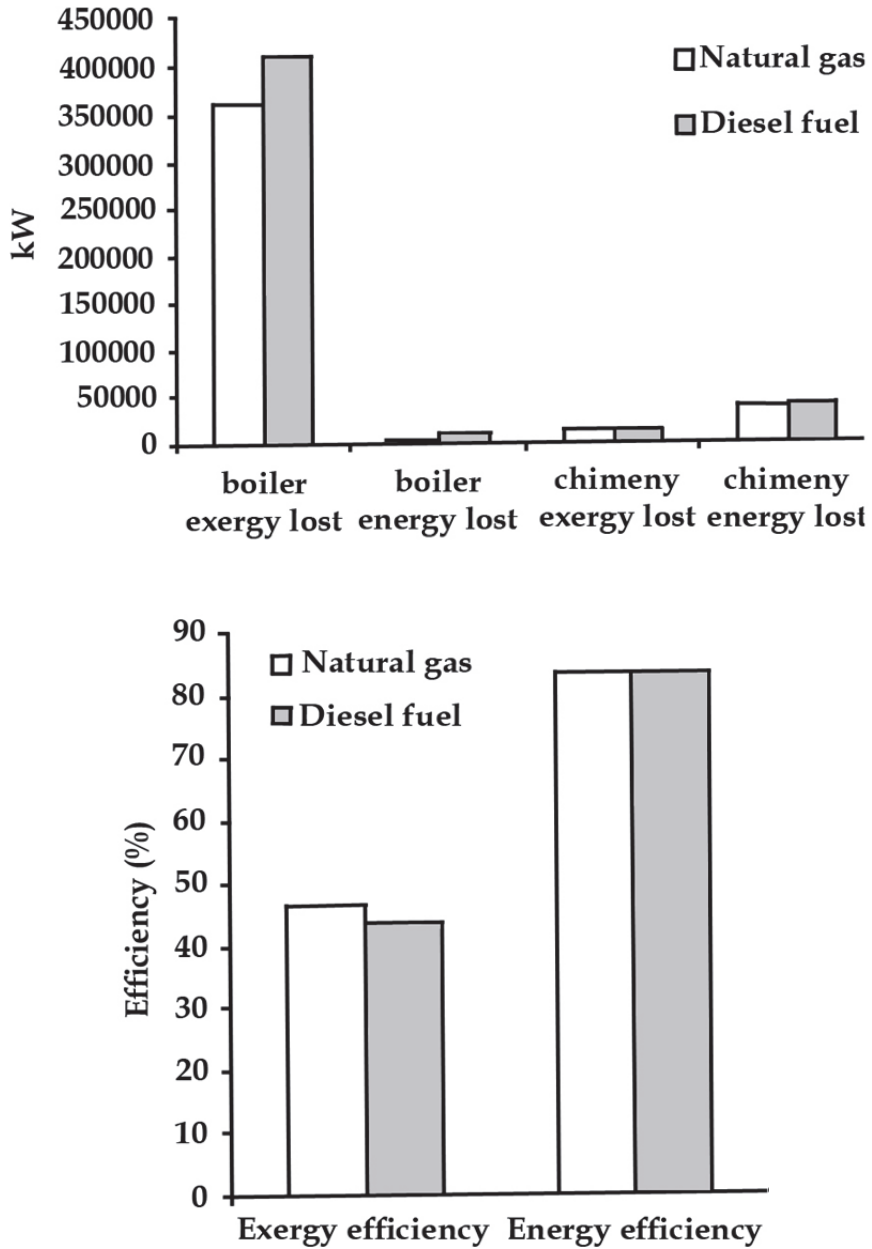


Fig. 4. The diagram of boiler exergy losses and efficiency



$$\eta_{I,B} = \frac{\sum_{out} \dot{m}_w h_w - \sum_{in} \dot{m}_w h_w}{\dot{m}_f LHV + \dot{m}_{dry\ air} h_a + \dot{w}_{in,B}} \quad (34)$$

$$\eta_{II,B} = \frac{\sum_{out} \dot{m}_w e_{x,w} - \sum_{in} \dot{m}_w e_{x,w}}{\dot{m}_f e_{ch,f} + \dot{w}_{in,B} + \dot{m}_{dry\ air} e_{x,a}} \quad (35)$$

In Table 4, the calculations of the enthalpy and exergy of the dry combustion gases, vapour and wet combustion gases have been summarized. Combustion gases do not have CO. Since the chemical exergy of CO is high, the exergy of the combustion products is negligible.

Fuel	Natural gas ( $M_{dry\ gas} = 29.77$ )		Diesel fuel ( $M_{dry\ gas} = 30.36$ )	
	Enthalpy	Exergy	Enthalpy	Exergy
Dry combustion gases	384.8	10.15	409.0	17.42
Vapour	740.4	314.30	801.6	417.19
Wet combustion gases	474.3	48.15	472.3	50.34

Table 4. Thermodynamic properties of dry and wet combustion gases and vapour outgoing from pre-heater

The result of calculating the last four equations is briefly shown in Figure 4.

### 3.4 The boiler processes

Two important processes happen in the boiler: Combustion and heat transfer. Therefore the internal exergy losses of the boiler ( $T_0 \dot{S}_{gen,B}$ ) are the losses of both exergy and heat transfer. Of course, there is a small exergy losses caused by friction which is calculated in the exergy caused by heat transfer. In the boiler, the exergy losses caused by friction have two different reasons; one is the pressure losses of the combusting gases (at most 4.4 kPa) which is to be neglected and the other is the pressure losses of the actuating fluid. These kinds of exergy losses, as was previously mentioned in the first part, are negligible compared to the losses due to heat transfer. Thus, there will be no specific analysis of friction; these two kinds of exergy losses together called the exergy losses due to heat transfer.

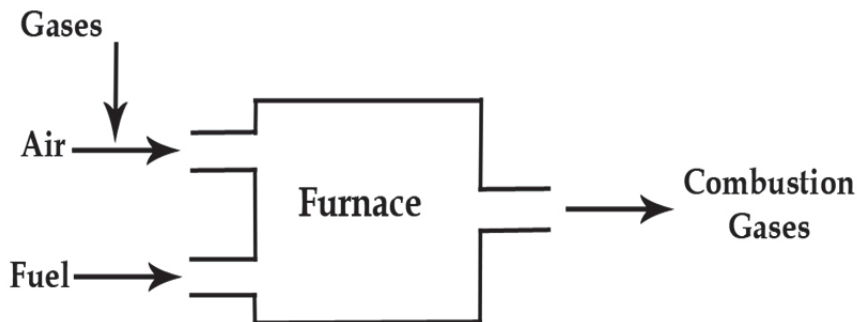


Fig. 5. The model of the boiler furnace

### 3.4.1 Combustion

To study the furnace, we assume that the combustion is in an isolated control volume. According to Figure 5, using the energy balance, we will get the enthalpy of the combustion gases:

$$\dot{Q}_{out,comb} = \dot{m}_f h_f + \dot{m}_a h_a + \dot{m}_r h_r - \dot{m}_p h_{p,ad} = 0 \quad (36)$$

By determining the temperature of the combustion products, using the iteration method, we can write the exergy balance equation of the furnace in order to determine the chemical exergy losses of the furnace.

$$T_0 \dot{S}_{gen,comb} = \dot{m}_f e_{ch,f} + \dot{m}_a e_{x,a} + \dot{m}_r e_{x,r} - \dot{m}_p e_{x,p,ad} \quad (37)$$

Therefore the combustion exergy efficiency will be:

$$\eta_{II,comb} = \frac{\dot{m}_p e_{x,p,ad}}{\dot{m}_f e_{ch,f} + \dot{m}_r e_{x,r} + \dot{m}_a e_{x,a}} \quad (38)$$

### 3.4.2 Heat transfer

In the boilers, heat transfer to the actuating fluid is classified into four categories:

- heat transfer in the pipes of the first and secondary economizers
- heat transfer in the pipes of the furnace walls
- heat transfer in the first, secondary and final super-heaters
- heat transfer in the first and secondary reheaters

Exergy losses caused by heat transfer occur in five main parts of the boiler; the evaporator, the economizer, the super-heater, the reheater and the air preheater. Exergy losses of the main five elements together with the furnace losses are the total losses of the boiler which were mentioned in Section 3.3.

$$T_0 \dot{S}_{gen,Q} = T_0 \dot{S}_{gen,B} - T_0 \dot{S}_{gen,comb} \quad (39)$$

### 3.5 The boiler elements

Each element of the boiler is, in fact, a heat exchanger. Therefore, when there is more one inlet or outlet to the heat exchanger, exergy balance is written as follows:

$$T_0 \dot{S}_{gen,element} = \sum_{in} \dot{m}_g e_{x,g} - \sum_{out} \dot{m}_g e_{x,g} + \dot{m}_w (e_{x,w,in} - e_{x,w,out}) \quad (40)$$

Therefore the heat transfer exergy efficiency will be:

$$\eta_{II,element} = \frac{\dot{m}_w (e_{x,w,out} - e_{x,w,in})}{\sum_{in} \dot{m}_g e_{x,g} - \sum_{out} \dot{m}_g e_{x,g}} \quad (41)$$

The results of exergy losses and efficiency of each element of the boiler are shown in Figure 6.

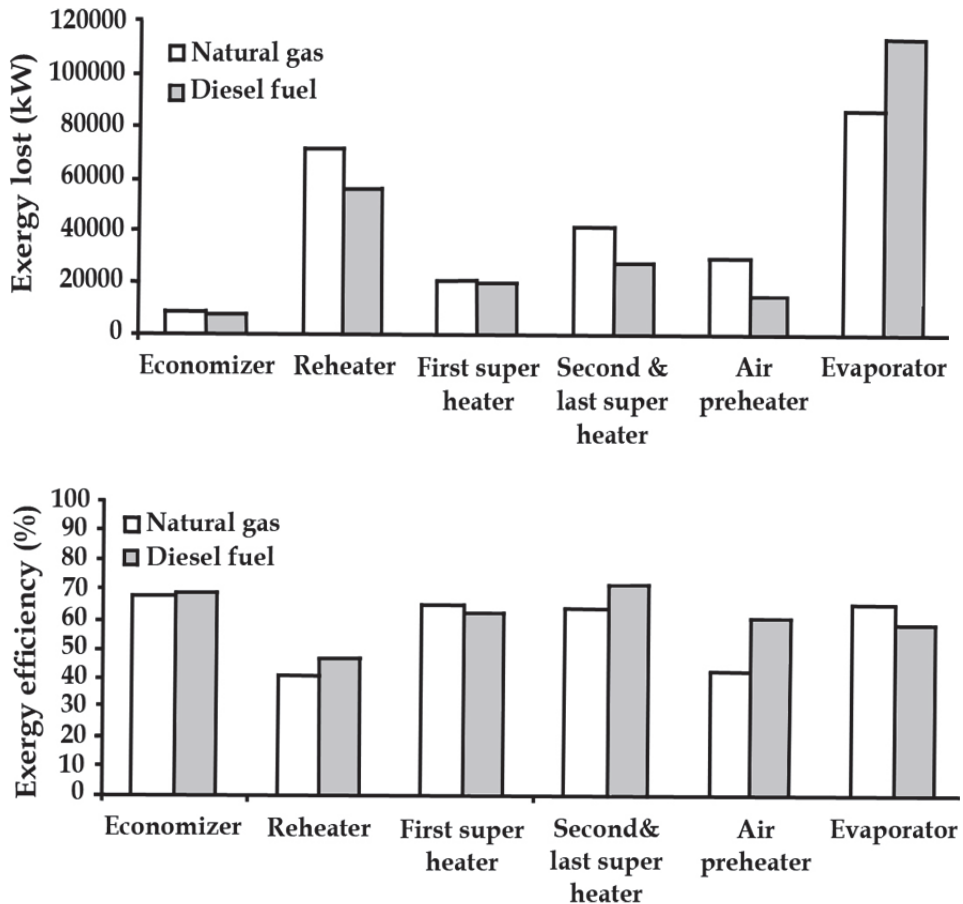


Fig. 6. The diagram of the exergy loss and efficiency

### 3.6 The correction factor of the boiler efficiency

All the measured data such as temperature, pressure, flow, etc and also the calculated magnitude of the exergy efficiency and losses were assumed under some specific environmental conditions. But the plant is not always under these specific conditions, and in fact there are some conditions under which the efficiency changes. These are divided into two parts: internal conditions such as excess air and moisture content and the external (environmental) conditions such as temperature and humidity.

An increase in the amount of air in the combustion process can easily decrease the adiabatic temperature of the flame so that the adiabatic exergy of the products is reduced and the exergy losses of the combustion increase. On the other hand, the combustion products have a lower temperature and greater mass flow as they flow inside the boiler, which leads to lower exergy losses in the heat transfer process.

The correction factor for the boiler exergy efficiency caused by the internal and external conditions is done using the equation (35). For instance, this factor due to the excess air is:

$$\Delta\eta_{II,B} = \frac{\sum_{out} \dot{m}_w e_{x,w} - \sum_{in} \dot{m}_w e_{x,w}}{\dot{m}_f e_{x,f} + \dot{m}_a e_{x,a} + \dot{W}_{in,B}} \Bigg|_{EAR}^x \quad (42)$$

Where EAR (Excess Air Ratio) is the ratio of the excess air to the time of the boiler test.

#### 4. Results and discussions

As is shown in Table 5, the lowest efficiency belongs to the gland condenser which is 28.9%. Its exergy loss at 46 kW is among the least. The highest exergy efficiency belongs to the high pressure and medium pressure turbines, and is 94.95%. In these turbines, 8617 kW of the 616505 kW of incoming exergy disappears and the rest is changed to work. The low pressure turbine has more exergy losses at 15303 kW and its efficiency is 87.47%, which is less than the high pressure turbine. The plant boiler, which is one of the most important elements of the cycle, has one of the lowest exergy efficiencies, 46.94%, which should be optimized.

Cycle elements	$(\dot{E}_w)_{lost}$ (kW)	$\eta_{II}$ (%)
Boiler	371522	46.94
Low Pressure Turbine	15303	87.47
Condenser	12867	60.05
High Pressure Turbine	8617	94.95
Generator	5222	98.06
Feed water pump	2147	60.00
Heater 5	2026	86.04
Heater 6	1563	91.42
Heater 4	1555	85.38
Heater 3	456	88.71
Heater 2	730	81.72
Heater 1	449	81.29
Feed water motor pump	313	94.50
Main drain	120	31.94
Condenser pump	62	64.57
Gland condenser	46	28.90
Condenser motor pump	15	92.00
Total	423013	-

Table 5. Exergy losses and exergy efficiency of the cycle elements of the plant

In Figure 4, exergy analysis shows that the major part of the exergy losses are because of the internal irreversibility of the boiler, while the chimney losses are less than one thirtieth of the boiler losses. Conversely, energy analysis shows that chimney losses are four times greater than boiler losses (heat transfer of the boiler to the surrounding area) and this is the major cause of efficiency reduction. Therefore, in order to reduce the energy losses, chimney losses are to be decreased, whereas the reduction of the internal exergy losses is more effective in the increase of exergy efficiency.

Figure 4 proves that energy efficiency is not dependent on the fuel type. The difference between the efficiencies of the two different fuels is 0.22% which is very minute. But the exergy efficiencies with two fuels are 3.02% different, based on Figure 4.

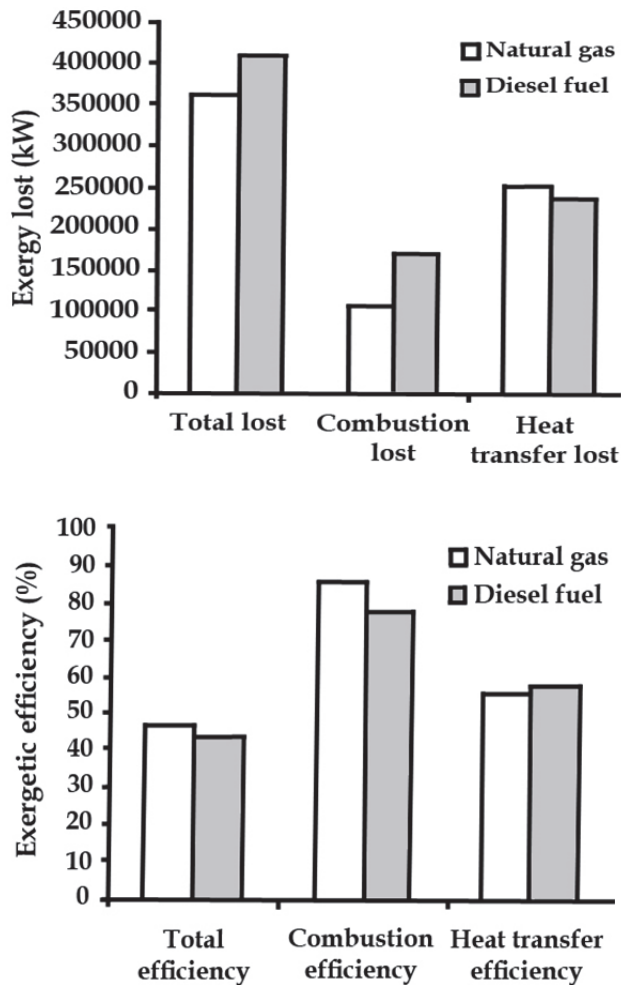


Fig. 7. The diagram of the exergy losses and exergetic efficiency of the different processes

Figure 7 demonstrates the comparison of the exergy losses and efficiencies of the different processes of the boiler.

The important result is that the exergy losses in the heat transfer process are greater than that in the combustion process. In other words, the heat transfer process is more irreversible. Exergy efficiency, as an evaluating standard for exergy losses, shows that the heat transfer process is more inconvenient than the combustion process. This result is generally true for all the plant boilers. In order to optimize the boilers, engineers should focus on heat transfer processes, optimization of the heat exchangers and increasing their exergy efficiencies. One of the methods to achieve this aim is Pinch which helps us in the exergy analysis and in determining the arrangement of the pipes in order to reduce the irreversibility of the heat transfer. When using diesel fuel, exergy losses in the combustion process are greater than

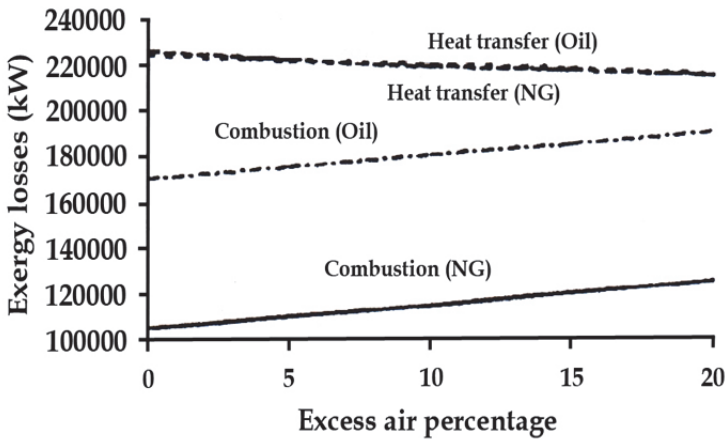


Fig. 8. The diagram of the internal exergy loss relative to the excess air percentage

In Figure 9, as the excess air percentage increases, the exergy efficiency of the boiler is reduced.

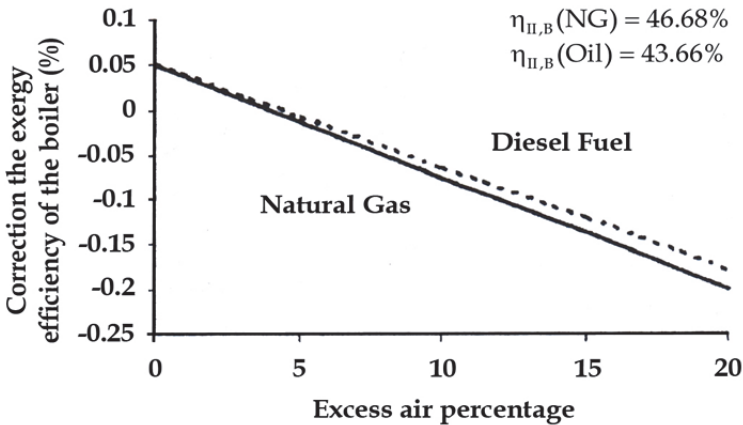


Fig. 9. The diagram of boiler efficiency based on the excess air percentage

when we use natural gas. But his is not the case for heat transfer. It is because the flow of combustion products is more when we use diesel fuel than when we use natural gas. On the other hand, the increase in the flow of combustion products is because of the great amount of incoming air to the furnace which leads to reduced exergy efficiency. In Figure 6, when we use natural gas, the exergy losses are greater in all elements than when we use diesel fuel, except the evaporator. The exergy efficiency, using diesel fuel, is less than when we use natural gas. The main reason behind the difference between the evaporator function and the other elements is that the combustion and heat transfer processes happen simultaneously in

the evaporator. The secondary and the final super-heaters have the highest exergy efficiency (71.67%) when using diesel fuel, with the exergy losses of 27657 kW. The second one is the evaporator using the natural gas, with exergy efficiency of 67.1% and exergy losses of 58800 kW. The reheater has the lowest efficiency using the natural gas (40.99%), with exergy losses of 70933 kW. The evaporator using diesel fuel has the highest exergy loss of 114071 kW and the economizer using diesel fuel has the lowest exergy losses of 7486 kW.

In Figure 8, as the excess air percentage becomes higher, the exergy losses of the natural gas increase at a lower rate. This reduction in the exergy losses of heat transfer based on the excess air percentage has a higher rate with natural gas than with diesel fuel. Generally, the gradients of exergy losses in the diagrams on combustion are more than in heat transfer.

In Figure 10, as the mass percentage of fuel humidity increases, the exergy efficiency of the boiler is reduced.

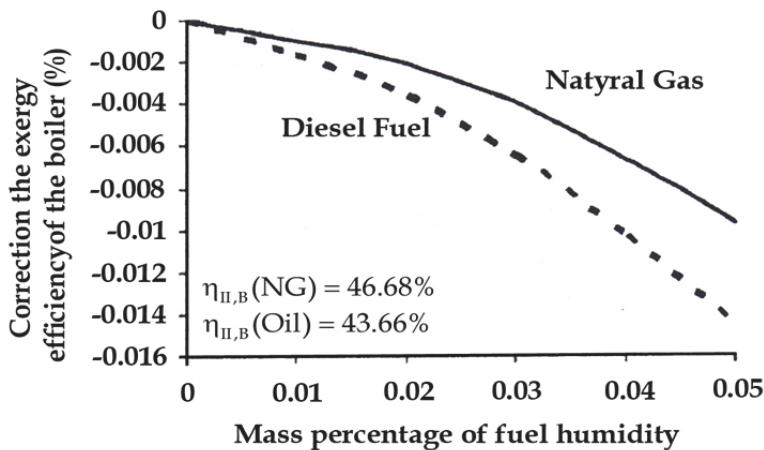


Fig. 10. The diagram of boiler efficiency changes based on the moisture content percentage

In Figure 10, humidity in diesel fuel causes more reduction in the boiler efficiency than in the natural gas. According to this Figure, it can be said that moisture content does not have a considerable effect upon the efficiency.

According to the Figure 11, we can say that as the relative air humidity percentage becomes higher, the boiler efficiency is reduced and the exergy efficiency reduction rate is more with diesel fuel than the natural gas.

Figure 12 shows that an increase in the air temperature from the assumed temperature of the environment (30°C) decreases the exergy efficiency, and under the other conditions, decrease or increase of the exergy efficiency is greater with the natural gas.

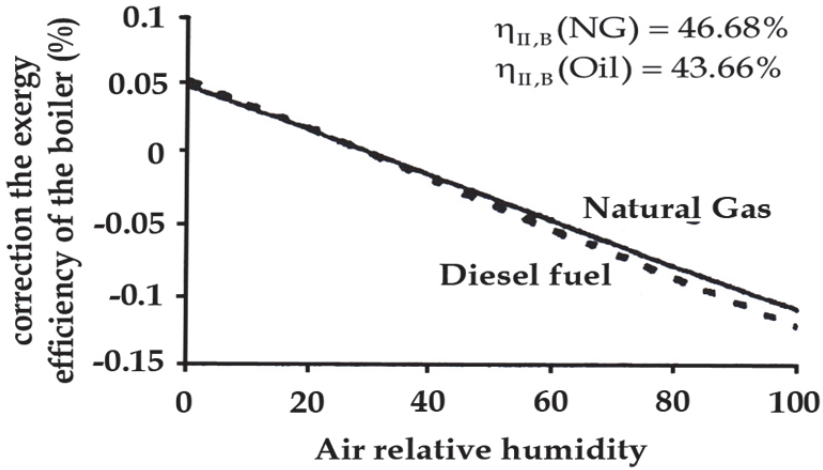


Fig. 11. The diagram of boiler efficiency changes based on the air relative humidity percentage

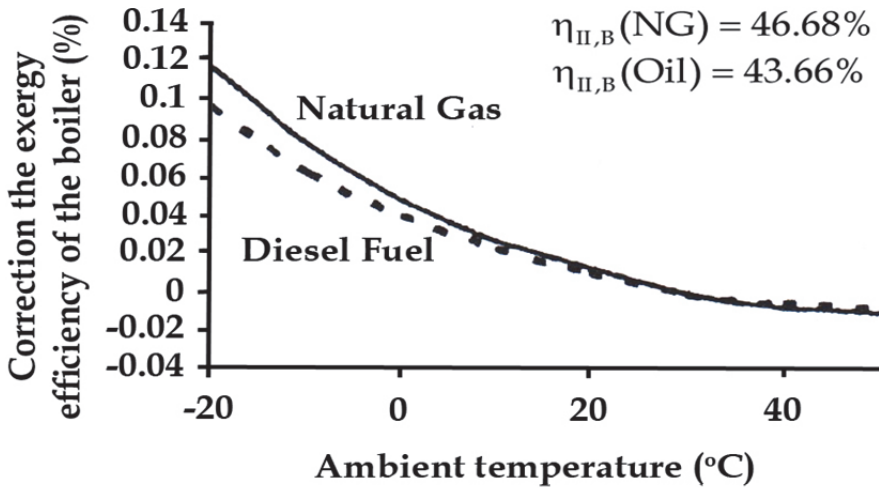


Fig. 12. The diagram of boiler efficiency changes based on the air temperature



## 5. Conclusions

The exergy analysis of the power plant cycle has shown that

- The total exergy efficiency of the plant is 36.1% for natural gas and 35.5% for diesel fuel.
- Among the main elements of the plant cycle, the greatest irreversibility (exergy losses or the least exergy efficiency) belongs to the boiler.
- The internal losses of the boiler which includes the heat transfer losses, the combustion losses and the friction losses, are 362899 kW for natural gas and 411127 MW for diesel fuel, and the exergy losses of the chimney, which are caused because of the combustion hot gases exiting it, are 12453 kW for the natural gas and 12668 kW for diesel fuel. The natural gas and diesel fuel, respectively, have chemical exergies of 700182 kW and 747995 kW.
- The comparison of the internal losses and the chimney losses shows that the outgoing exergy from the chimney is not considerable. Also, the exergy losses of the boiler are 46.68% for the natural gas and 43.66% for diesel fuel, under the assumed conditions ( $T_0 = 30^\circ\text{C}$  and  $P_0 = 90\text{ kPa}$ ).
- Analysing the boiler processes shows that the exergy losses caused by heat transfer (255999 kW for the natural gas and 239302 kW for diesel fuel) are greater than the exergy losses of the combustion (106900 kW for the natural gas and 171835 kW for diesel fuel) and those against the heat transfer process in the combustion process of the natural gas are less irreversible than in the diesel fuel combustion.
- After proving that the main cause of the exergy losses are the heat transfer has the boiler, the exergy analysis of the boiler elements shows that the reheater has the lowest exergy efficiency (40.99% for the natural gas and 46.6% for diesel fuel) and that the evaporator has the highest exergy losses (85800 kW for the natural gas and 114071 kW for diesel fuel). All these reports show that natural gas is better than diesel fuel in producing super heater vapour in the boiler.

## 6. References

- Bejan, A. (1982). Entropy Generation through Heat and Fluid Flow. Wiley, New York.
- Bejan, A. (1988). Advanced Engineering Thermodynamics. Wiley, New York.
- Brodianskii, V.M. (1973). Eksergeticheskii Method Termodinamicheskogo Analiza. *Energia*, Moscow. pp. 32.
- Comelissen, R.L. and Hirs, G.G. (1999). Thermodynamic Optimization of a Heat Exchanger. *International Journal of Heat and Mass Transfer*, Vol. 42, pp. (951-959).
- Ebadi, M.J. and Gorji-Bandpy, M. (2005). Exergetic analysis of gas turbine plants. *Int. J. Exergy*, Vol. 2, No. 1, pp. (31-39).
- Gorji-Bandpy, M. and Ebrahimian, V. (2007). Exergy analysis of a steam power plant: a case study in Iran. *Int. J. Exergy*, Vol. 4, No. 1, pp. (54-71).
- Gorji-Bandpy, M.; Goodarzian, H. and Biglari M. (2010). The Cost-effective Analysis of a Gas Power Plant. *Taylor&Francis Group, LLC*, Vol. 5, No. 4, pp. (348-358).
- Gorji-Bandpy, M.; Yahyazadeh-Jelodar, H. and Khalili, M.T. (2011). Optimization of Heat Exchanger Network. Vol. 31, Issue 5, pp. (770-784).

- Jordan, D.P. (1997). Macroscopic Thermodynamics. Part 9: Methodology, *Pergamon*.
- Keenan, J.H. (1932). A Steam Chart for Second Law Analysis. *Mech. Eng.*, Vol. 54, pp. (195-204)
- Kotas, T.J. (1995). The exergy Method of Thermal Plant Analysis. *Krieger Publishing Co.*, Malabar, Florida.
- Moran, M.J. (1982). Availability Analysis: A Guide to Efficient Energy Use. *Prentice-Hall*, Englewood Cliffs, NJ, p. (115).
- Reistad, G.M. (1972). A property diagram to illustrate irreversibility. *ASHRAE Trans.*, Vol. 78, Part II, pp. (97-101).
- Stepanov, V.S. (1995). Chemical energies and exergies of fuels. *Energy*, Vol. 20, No. 3, pp. (235-242).
- Szargut, J.; Morris, D.R. and Steward, F.R. (1988). Exergy Analysis of Thermal, Chemical and Metallurgical Process. *Hemisphere Publishing Corp.*, New York.
- Thirumaleshwar, M. (1979). Exergy method of analysis and its application. *Cryogenics*, Vol. 19, pp. (355-361).
- Wark Jr. (1995). Advanced Thermodynamics for Engineering. *McGraw-Hill*, New York.

# Dimensionless Parametric Analysis of Spark Ignited Free-Piston Linear Alternator

Jinlong Mao, Zhengxing Zuo and Huihua Feng  
*School of Mechanical Engineering,  
Beijing Institute of Technology, Beijing,  
China*

## 1. Introduction

The decreasing crude oil sources, the increasing concern about greenhouse gas pollution, and the strict emission standards force researchers into never-ending effort to design less polluting and more fuel efficient vehicles. Free-piston linear alternator (FPLA) which is a free-piston engine coupled to a linear alternator, converting piston's kinetic energy into electricity directly, has attracted considerable research interests recently by a number of research groups worldwide due to its potential advantages listed below:

1. High efficiency. 1) Direct conversion of piston motion into electric energy with shorter energy transferring path. 2) Reduction of friction between piston ring and cylinder wall due to the elimination of the dynamic piston side forces.
2. High power capacity. With a free-piston engine, heavy mechanical components as crankshaft, flywheel and camshaft are eliminated and the generator is integrated into the engine directly. Therefore, the power-train will have fewer components, lower weight and also occupies less space.
3. Controllable compression ratio. The stroke and thus the compression ratio can be dynamically controlled through the electrical load.
4. Multi-fuel feasibility. The variable compression ratio feature provides the opportunity for the engine to be operated with wide range of fuels [1].
5. Few moving parts. There is only one major moving part: the piston, connecting rod and translator assembly. As a consequence, the engine requires less lubrication and has less friction wear.
6. Good transient response. The free-piston engine reaches its operating point almost immediately since it does not have any energy accumulators, such as flywheel [2].

In the recent two decades, some literatures have already been published about the design and numerical simulation of FPLA. Martin Goertz and Lixin Peng evaluated several feasible hybrid power-train concepts and the results indicated that free-piston linear generator was one of the most promising candidates of the future power-train configuration [3]. West Virginia University had already demonstrated stable operation of a spark ignited FPLA prototype with bore of 36.5mm, maximum possible stroke of 50mm and 316W output power was produced at 79V working at full load [4]. Ehab Shoukry did some research about parametric study of a two-stroke direct injection linear engine using zero dimensional single zone models [5]. The European Union has been doing a subject of

Free-piston Energy Converter (FPEC) since 2002 aimed to develop an efficient new technology suitable for vehicle propulsion, auxiliary power unit and distributed power generation. The design and optimization of linear alternator and crankless engine, control strategy, numerical simulation and power management have been carried out systemically [6]. Dr. Peter Van Blarigan at Sandia National Laboratory presented the design of a dual piston free-piston engine generator with 40kW electric power output and the engine employed homogeneous charge compression ignition (HCCI) operating on a variety of hydrogen-containing fuels [7, 8]. Jakob Fredriksson and Ingemar Denbratt at Chalmers University of Technology investigated a two-stroke free-piston engine with different fuels using BOOST and SENKIN [9]. Czech Technical University had successfully developed a direct injection FPLA prototype recently, and steady operation had been realized based on precise motion control. When the prototype was running with frequency of 27Hz and compression of 9, the average power output was approximately 350W, but the efficiency had not been reported [10]. Mikalsen and Roskilly proposed a design of a single cylinder free-piston engine generator with gas-filled bounce chamber, simulated its working process and discussed the effects of parameters in a wide operating range to the engine's performance [11]. A novel approach of modeling the free-piston engines through the introduction of solution-dependent mesh motion in an engine CFD toolkit OpenFOAM was also presented [12].

Dimensionless analysis using Buckingham's  $\pi$  theory was performed base on the parameters of a spark ignited FPLA prototype (as can be seen in Fig.1) in the following contents. The objectives of this work are:

- To decrease the number of variables and still use the equations involved in the modeling of FPLA.
- To generalize the study of FPLA to a wide range of geometrical designs in order to find out the best geometric dimensions with most favorable performance.
- To establish a scale for the similitude concept so the experiments done on the existing prototype could be used to predict performances for FPLA with various sizes and dimensions.

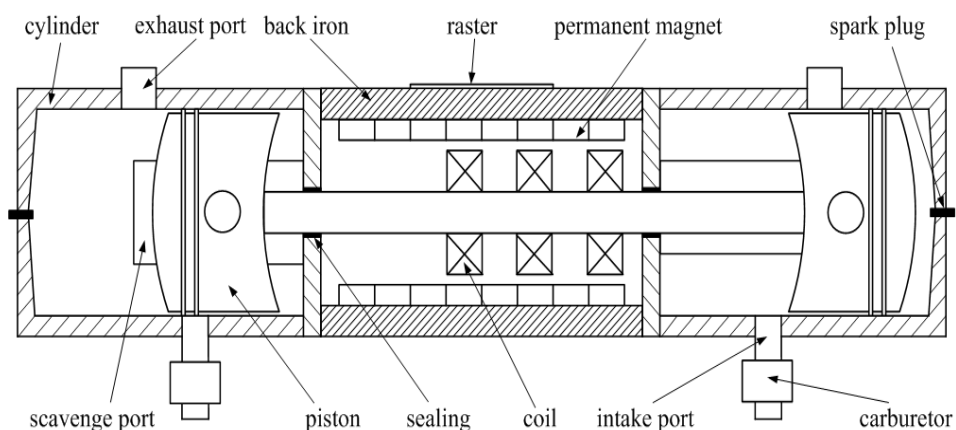


Fig. 1. FPLA configuration

## 2. System modeling

### 2.1 Dynamic modeling

The Modeling of FPLA and numerical simulation of its performance have already been done by a lot of researchers. Usually the model consists of a dynamic calculation of the piston motion and a thermodynamic calculation of the engine's thermodynamic events. The piston motion is governed by the interaction of forces that act on the piston simultaneously, which are in-cylinder pressure force from each cylinder, electromagnetic force, friction force and inertia force, as is shown in Fig.2.

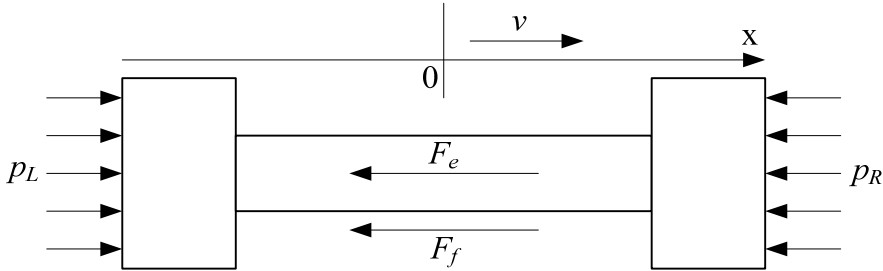


Fig. 2. Free body diagram for FPLA

The Newton's second law is applied:

$$m \frac{d^2x}{dt^2} = (p_L - p_R)A - F_f - F_e \quad (1)$$

Since the FPLA doesn't have a crankshaft, it is free of the dynamic friction caused by the crankshaft mechanism and the friction force is caused mainly by the interactions between cylinder and piston rings, cylinder and piston skirt, which is small compared with the load of the linear alternator. Therefore, the friction force is considered to be constant during the whole cycle in the calculation [13].

### 2.2 Modeling of the linear alternator

The linear alternator consists of two main components, a stator and a translator. The permanent magnets are mounted on the stator and the translator is the moving portion of the machine which is made up of coils. A schematic of a three-phase, "U" shaped linear alternator with permanent magnet (PM) excitation is shown in Fig.3.

The FPLA operates on the same basic physical principles as conventional rotary alternators. The principle that governs the voltage generating operation of the alternator is Faraday's law expressed as [14]:

$$\varepsilon_{ind} = -\frac{d\lambda}{dt} = -N_{coil} \frac{d\phi}{dt} \quad (2)$$

The permanent magnets create a magneto motive force (MMF) in the air gap between the stator and the winding coils as shown in Fig.4, and it can be described by the following mathematic equation:



Where,

$$a_0 = \frac{1}{\tau} \int_0^{2\tau} M_F(x) dx = 0$$

$$a_1 = \frac{1}{\tau} \int_0^{2\tau} M_F(x) \cos\left(\frac{\pi x}{\tau}\right) dx = 0$$

$$b_1 = \frac{1}{\tau} \int_0^{2\tau} M_F(x) \sin\left(\frac{\pi x}{\tau}\right) dx = \frac{4}{\pi} M_p \sin\left(\frac{\pi \tau_p}{2\tau}\right)$$

Then

$$M_F(x) = \frac{4}{\pi} M_p \sin\left(\frac{\pi \tau_p}{2\tau}\right) \sin\left(\frac{\pi x}{\tau}\right).$$

So the flux density in the air gap due to PM is:

$$B(x) = \frac{\mu_0}{g} M_F(x) = \frac{\mu_0}{g} \frac{4}{\pi} M_p \sin\left(\frac{\pi \tau_p}{2\tau}\right) \sin\left(\frac{\pi x}{\tau}\right) = B_m \sin\left(\frac{\pi x}{\tau}\right) \quad (5)$$

Where

$$B_m = \frac{\mu_0}{g} \frac{4}{\pi} M_p \sin\left(\frac{\pi \tau_p}{2\tau}\right).$$

Both experimental measurements and numerical calculation by finite element method showed that the flux in the air gap of the PM excited linear alternator in Fig.4 could be assumed to be sinusoidal supporting the above result [16].

Therefore, the flux contained in the differential element  $dx$  is:

$$d\phi = B(x) dA = B(x) H dx \quad (6)$$

Then the total flux contained in the coil of one phase at random position  $x$  is described by the following equation:

$$\begin{aligned} \lambda(x) &= \int_{x-\tau}^x N_{coil} H B(x) dx = \int_{x-\tau}^x N_{coil} \frac{\mu_0}{g} \frac{4}{\pi} M_p \sin\left(\frac{\pi \tau_p}{2\tau}\right) \sin\left(\frac{\pi x}{\tau}\right) dx \\ &= -\tau H N_{coil} M_p \frac{\mu_0}{g} \frac{8}{\pi^2} \sin\left(\frac{\pi \tau_p}{2\tau}\right) \cos\left(\frac{\pi x}{\tau}\right) \end{aligned} \quad (7)$$

Thus, the induced electromotive force produced in the coil of one phase is:

$$\varepsilon = -\frac{d\lambda}{dt} = H N_{coil} M_p \frac{\mu_0}{g} \frac{8}{\pi} \sin\left(\frac{\pi \tau_p}{2\tau}\right) \sin\left(\frac{\pi x}{\tau}\right) \frac{dx}{dt} \quad (8)$$

The induced current from the load circuit can be derived in the following equations:

$$\varepsilon(t) = (R_s + R_L)i_L(t) + L \frac{di_L(t)}{dt} \quad (9)$$

$$i_L(t) = \frac{\varepsilon(t)}{R_s + R_L} \left( 1 - e^{-\frac{R_s + R_L}{L}t} \right) \quad (10)$$

The magnetic force has the opposite direction to the direction of the translator's movement. According to Ampere's law, it is described in the following equation:

$$F_e = 2N_{coil}B(x)i_LH = 4H^2N_{coil}^2B_m^2 \frac{\left( 1 - e^{-\frac{R_s + R_L}{L}t} \right)}{R_s + R_L} \sin^2\left(\frac{\pi x}{\tau}\right) \frac{dx}{dt} \quad (11)$$

When it comes to three-phase linear alternator, the third phase is derived from another two according to the following equation [17]:

$$\sin \varphi = -\sin\left(\varphi + \frac{2}{3}\pi\right) - \sin\left(\varphi - \frac{2}{3}\pi\right) \quad (12)$$

So the total electromagnetic force produced by a three-phase linear alternator is:

$$\begin{aligned} F_e &= 4H^2N_{coil}^2B_m^2 \frac{\left( 1 - e^{-\frac{R_s + R_L}{L}t} \right)}{R_s + R_L} \frac{dx}{dt} \left( \sin^2\left(\frac{\pi x}{\tau} - \frac{2}{3}\pi\right) + \sin^2\left(\frac{\pi x}{\tau}\right) + \sin^2\left(\frac{\pi x}{\tau} + \frac{2}{3}\pi\right) \right) \\ &= 6H^2N_{coil}^2B_m^2 \left( 1 - e^{-\frac{R_s + R_L}{L}t} \right) \frac{1}{R_s + R_L} \frac{dx}{dt} \\ &= M \left( 1 - e^{-\frac{R_s + R_L}{L}t} \right) \frac{dx}{dt} \end{aligned} \quad (13)$$

Where

$$M = 6H^2N_{coil}^2B_m^2 \frac{1}{R_s + R_L}$$

In order to develop a first approach analysis, it's assumed that the alternator circuit works in resonance condition, which means that the current and voltage have the same phase, so that it is possible to consider the circuit as only a resistance [18]. Thus, the electromagnetic force is proportional to the speed of the translator:

$$F_e = M \frac{dx}{dt} \quad (14)$$



### 2.3 Thermodynamic modeling

The thermodynamic model is derived based on the first law of thermodynamics and ideal gas law. It consists of the calculation of the process of scavenging, compression, combustion, expansion and exhaust. The zero-dimensional, single zone model is used to describe the thermodynamic process.

Applying the first law of thermodynamics and ideal gas law on the cylinder as an open thermodynamic system, shown in Fig.5, and assuming that the specific heat  $c_v$  and the gas constant  $R$  are constant:

$$\frac{dU}{dt} = -p \frac{dV}{dt} + \frac{dQ}{dt} + \dot{H}_i - \dot{H}_e \quad (15)$$

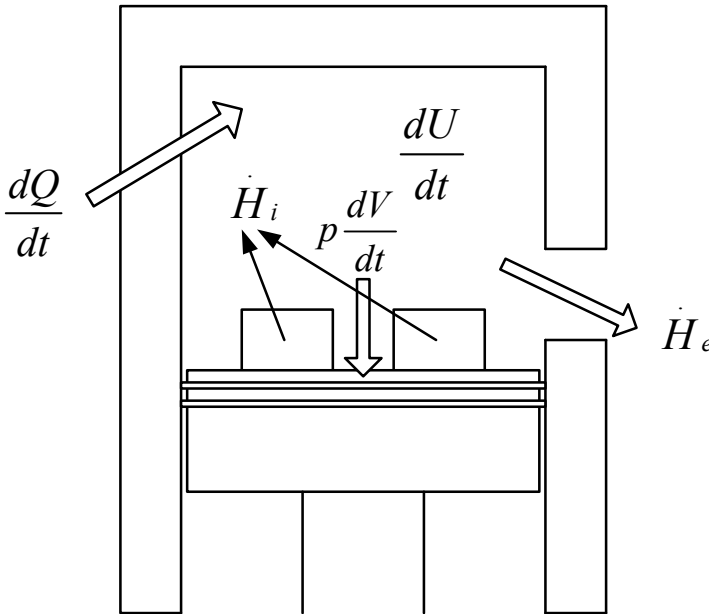


Fig. 5. Thermodynamic system of FPLA

For the case of compression and expansion process neglecting the crevice flow and the leakage, the first law of thermodynamics applied to the cylinder content becomes:

$$m_{in} \frac{d(c_v T)}{dt} = -p \frac{dV}{dt} + \frac{dQ}{dt} \quad (16)$$

Considering the cylinder content is ideal gas, and then at every instant the ideal gas law is satisfied:

$$pV = m_{in} RT \quad (17)$$

Substitution and mathematical manipulation yield the following equation which is used to calculate the in-cylinder pressure at each time step.

$$\frac{dp}{dt} = \frac{\gamma - 1}{V} \frac{dQ}{dt} - \gamma \frac{p}{V} \frac{dV}{dt} \quad (18)$$

In the combustion model, since the engine is crankless, a time-based Wiebe functions (as opposed to a conventional crank angle-based approach) is used to express the mass fraction burned in combustion process [4]:

$$\chi(t) = 1 - \exp\left(-a\left(\frac{t-t_0}{t_c}\right)^{1+b}\right) \quad (19)$$

and the heat rate released during combustion is:

$$\frac{\partial Q_c}{\partial t} = Q_{in} \frac{\partial \chi(t)}{\partial t} = Q_{in} \frac{1}{t_c} a(1+b) \left(\frac{t-t_0}{t_c}\right)^b \exp\left[-a\left(\frac{t-t_0}{t_c}\right)^{1+b}\right] \quad (20)$$

The in-cylinder heat transfer effect is modeled according to Hohenberg [19]:

$$\frac{\partial Q_{ht}}{\partial t} = hA_{cyl}(T - T_w) \quad (21)$$

$$h = 130V^{-0.06} \left(\frac{p}{10^5}\right)^{0.8} T^{-0.4} \left(\bar{U} + 1.4\right)^{0.8} \quad (22)$$

Since most heat transfer models, like the ones proposed by Woschni and Hohenberg, are made for Diesel engines. This means that they take radiative heat transfer effect into account, which is hardly present in premixed combustion. Hence, in the numerical simulations a factor of 0.5 is introduced to reduce the heat transfer coefficient [9].

So the total energy that is used to increase the in-cylinder pressure in equation (18) can be expressed in the following equation:

$$\frac{dQ}{dt} = \frac{\partial Q_c}{\partial t} - \frac{\partial Q_{ht}}{\partial t} \quad (23)$$

Exhaust blown down is modeled to be a polytropic expansion process while the exhaust port is opening and the scavenging ports are still covered by the piston [11].

$$\frac{dp}{dt} = \frac{n-1}{V} \frac{dQ}{dt} - n \frac{p}{V} \frac{dV}{dt} \quad (24)$$

For two-stroke spark ignition engines with under piston or crankcase scavenging, the scavenging efficiency is about 0.7~0.9 [20], a scavenging efficiency of 0.8 is introduced to evaluate the effects of incomplete scavenging effect. The moment the scavenging ports are open, the pressure and temperature are assumed to be the same with the scavenging conditions and the incoming gases mix entirely with the burned gases.

### 3. Dimensionless analysis

#### 3.1 Dimensionless modeling of FPLA

The dimension of a physical quantity is associated with combinations of mass, length, time, temperature and heat quantity, represented by symbols [M], [L], [T], [θ] and [H] respectively, each raised to rational powers. Since heat quantity is also a kind of energy, its dimension is replaced with [ML<sup>2</sup>T<sup>-2</sup>].

The basic variables that involved were defined based on the dynamic and thermodynamic modeling of FPLA. The variables and their dimensions are listed in Tab.1.

Variables	Symbol	Dimensions
Bore	$D$	[L]
Piston area	$A$	[L <sup>2</sup> ]
Volume of the cylinder	$V$	[L <sup>3</sup> ]
Effective stroke length	$L_{eff}$	[L]
Total stroke length	$L_{tot}$	[L]
Translator ignition position	$x_{ign}$	[L]
Mass of the translator	$m$	[M]
Load coefficient	$M$	[MT <sup>-1</sup> ]
Friction force	$F_f$	[MLT <sup>-2</sup> ]
Energy	$Q$	[ML <sup>2</sup> T <sup>-2</sup> ]
Scavenge pressure	$p_0$	[ML <sup>-1</sup> T <sup>-2</sup> ]
Scavenge temperature	$T_0$	[θ]
Pressure	$p$	[ML <sup>-1</sup> T <sup>-2</sup> ]
Gas constant	$R$	[θL <sup>-1</sup> T <sup>-2</sup> ]
Temperature	$T$	[θ]
Wall temperature	$T_w$	[θ]
Combustion duration	$t_c$	[T]
Compression ratio	$\varepsilon$	[1]
indicated efficiency	$\eta_i$	[1]
Effective efficiency	$\eta_e$	[1]
Time	$t$	[T]
Frequency	$f$	[T <sup>-1</sup> ]

Table 1. Basic variables and their dimensions

The effective stroke length is the distance between the upper edge of the exhaust port and the cylinder and the total stroke length is distance that the translator can travel from cylinder head to cylinder head [21].

Since the FPLA is controlled based on the displacement feedback of displacement sensor, the ignition timing is defined by the position of the translator where the controller gives out the ignition signals.

According to Buckingham's  $\pi$  theory [22], four fundamental reference variables which are independent from each other were chosen, and the other variables are described with these reference variables in their index form and every redefined variable is dimensionless. The four reference variables chosen were bore, mass of the translator, scavenging pressure and scavenging temperature.

Dividing the basic variables by the reference variables results in the dimensionless variables:

Dimensionless bore	$D^*=D/D=1$
Dimensionless piston area	$A^*=A/D^2$
Dimensionless cylinder volume	$V^*=V/D^3$
Dimensionless total stroke length	$L_{tot}^*=L_{tot}/D$
Dimensionless mass of the translator	$m^*=m/m=1$
Dimensionless scavenge air pressure	$p_0^*=p_0/p_0=1$
Dimensionless scavenging temperature	$T_0^*=T_0/T_0=1$
Dimensionless temperature	$T^*=T/T_0$
Dimensionless wall temperature	$T_w^*=T_w/T_0$
Dimensionless load coefficient	$M^*=M/(p_0^{0.5}D^{0.5}m^{0.5})$

Since the dimensionless translator ignition position is changing with different stroke length and fixed translator ignition position, in the dimensionless calculation, its actual dimensionless value taken is defined by the compression ratio the engine has already achieved when the spark plug ignites, as is described in Fig.6.

$$\epsilon_{ign} = \frac{V_{eff}}{V_{ign}} = \frac{D^2 L_{eff}}{D^2 (L_{tot} / 2 - x_{ign})} = \frac{L_{eff} / D}{(L_{tot} / 2 - x_{ign}) / D} = \frac{L_{eff}^*}{L_{tot}^* / 2 - x_{ign}^*} \quad (25)$$

Then the dimensionless translator ignition position can be deduced:

$$x_{ign}^* = \frac{L_{tot}^*}{2} - \frac{L_{eff}^*}{\epsilon_{ign}} \quad (26)$$

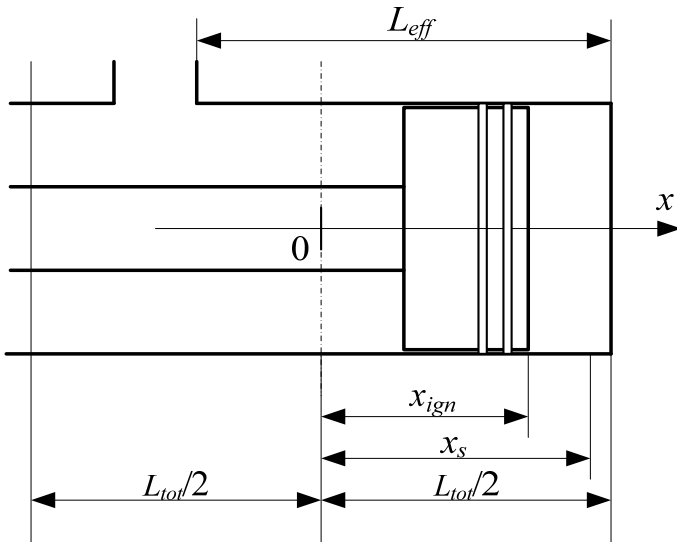


Fig. 6. Diagram of translator ignition position

Substituting all the dimensionless variables into equations (1) and (18), the following equations can be derived:

$$\frac{m}{m} \frac{d^2 \left( \frac{x}{D} \right)}{d \left( \frac{t}{t_r} \right)^2} = \frac{1}{p_0} \frac{A}{D^2} (p_L - p_R) - \frac{1}{p_0 D^2} F_f - \frac{M}{p_0^2 D^2 m^2} \frac{d \left( \frac{x}{D} \right)}{d \left( \frac{t}{t_r} \right)} \quad (27)$$

$$m^* \frac{d^2 x^*}{dt^{*2}} = (p_L^* - p_R^*) A^* - F_f^* - M^* \frac{dx^*}{dt^*} \quad (28)$$

$$\frac{dp^*}{dt^*} = \frac{d \left( \frac{p}{p_0} \right)}{d \left( \frac{t}{t_r} \right)} = -\gamma \frac{p}{p_0} \frac{d \left( \frac{V}{D^3} \right)}{d \left( \frac{t}{t_r} \right)} + \frac{\gamma - 1}{V} \frac{d \left( \frac{Q}{p_0 D^3} \right)}{d \left( \frac{t}{t_r} \right)} \quad (29)$$

$$\frac{dp^*}{dt^*} = -\gamma \frac{p^*}{V^*} \frac{dV^*}{dt^*} + \frac{\gamma - 1}{V^*} \frac{dQ^*}{dt^*} \quad (30)$$

Where  $t_r = m^2 p_0^{-2} D^{-2}$  and equations (28) (30) are the dimensionless form of the translator's dynamic and thermodynamic equations.

The dimensionless gas constant was deduced based on the ideal gas law:

$$R^* = \frac{p^* V^*}{m_{in}^* T^*} = \frac{p_0 D^3}{m_{in} T} = \frac{m T_0}{p_0 D^3} \frac{p V}{m_{in} T} = \frac{R}{p_0 D^3 / m T_0} \quad (31)$$

Since the in-cylinder temperature is strongly transient, the dimensionless in-cylinder temperature is acquired using the dimensionless ideal gas law:

$$T^* = \frac{p^* V^*}{m_{in}^* R^*} \quad (32)$$

So the dimensionless variables defined in equations (28) (30) and (31) are:

Dimensionless time	$t^* = t / t_r$
Dimensionless combustion duration	$t_c^* = t_c / t_r$
Dimensionless frequency	$f^* = f \cdot t_r$
Dimensionless friction force	$F_f^* = F_f / (p_0 \cdot D^2)$
Dimensionless pressure	$p^* = p / p_0$
Dimensionless energy	$Q^* = Q / (p_0 \cdot D^3)$
Dimensionless gas constant	$R^* = R / (p_0 \cdot D^3 / m / T_0)$

The combustion model and heat transfer equations can also be transferred to their dimensionless form using the dimensionless parameters we have already obtained:

$$\frac{\partial Q_c^*}{\partial t^*} = Q_{in}^* \frac{a(b+1)}{t_c^*} \left( \frac{t^* - t_0^*}{t_c^*} \right)^b \exp \left[ -a \left( \frac{t^* - t_0^*}{t_c^*} \right)^{b+1} \right] \quad (33)$$

$$h^* = 130V^{*-0.06} \left( \frac{P^*}{10^5} \right)^{0.8} T^{*-0.4} \left( \bar{U}^* + 1.4 \right)^{0.8} \quad (34)$$

$$\frac{\partial Q_{ht}^*}{\partial t^*} = h^* A_{cyl}^* (T^* - T_w^*) \quad (35)$$

All the equations were solved with a numerical simulating program in Matlab. The program starts calculation using a group of mathematic equations describing the dimensionless dynamic and dimensionless thermodynamic processes of FPLA, including the heat transfer rate, the in-cylinder pressure, the in-cylinder temperature, the fraction of fuel that burnt, the load and the work done, etc. At each time step, the program calls the dynamic subroutine and updates the displacement, the velocity, the averaged velocity and the acceleration. After the engine stabilizes, the program calculates the engine frequency, the compression ratio, the indicated power, the frictional power, the effective power output and the effective efficiency.

The dimensionless form of the output parameters are:

$$\text{Dimensionless compression ratio} \quad \varepsilon^* = \frac{L_{eff}^*}{L_{tot}^* / 2 - x_s^*} = \frac{L_{eff}}{L_{tot} / 2 - x_s} = \varepsilon$$

$$\text{Dimensionless frictional work} \quad W_f^* = 4F_f^* x_s^*$$

$$\text{Dimensionless averaged speed} \quad \bar{U}^* = 4x_s^* f^*$$

$$\text{Dimensionless indicated work} \quad W_i^* = \oint_{V^*} p^* dV^*$$

$$\text{Dimensionless effective work} \quad W_e^* = W_i^* - W_f^*$$

$$\text{Dimensionless indicated efficiency} \quad \eta_i^* = \frac{W_i^*}{Q_{in}^*} = \frac{W_i}{Q_{in}} = \eta_i$$

$$\text{Dimensionless effective efficiency} \quad \eta_e^* = \frac{W_i^* - W_f^*}{Q_{in}^*} = \frac{W_i - W_f}{Q_{in}} = \eta_e$$

$$\text{Dimensionless effective power output} \quad P_e^* = W_e^* f^*$$

$$\text{Dimensionless frictional power} \quad P_f^* = W_f^* f^*$$

The dimensionless compression ratio and efficiency have just the same value with their dimensional forms, which are used to validate the correctness of the dimensionless process.

### 3.2 Operating ranges

The variables of the dimensionless analysis were chosen to cover the normal operating ranges of two-stroke free-piston engine. The base case represented the parameters of the experimental spark ignited, two-stroke FPLA prototype that was built in Beijing Institute of Technology. The basic parameters are listed in Tab.2.

The effective stroke length to bore ratio, dimensionless load coefficient, dimensionless translator ignition position, dimensionless combustion duration and dimensionless input

energy (opening proportion of the throttle) were the variable input factors of the parametric study. The dimensionless operating matrixes of each parameter are listed in Tab.3.

Basic parameters	Value	Dimensionless value
$D/\text{mm}$	34	1
$L_{\text{eff}}/\text{mm}$	20	0.5882
$L_{\text{tot}}/\text{mm}$	36	1.0588
$R_l/\Omega$	2.5	×
$R_s/\Omega$	2	×
$M/N \cdot (\text{m s}^{-1})^{-1}$	55.3	0.7746
$m/\text{kg}$	1.74	1
$p_0/\text{bar}$	1	1
$T_0/\text{K}$	293	1
$T_w/\text{K}$	453	1.5461
$t_c/\text{ms}$	5	0.2210
$\varepsilon_{\text{ign}}$	4	×
$a$	30%	×

Table 2. Parameters of the FPLA prototype

Parameters	Value						
	case1	case2	case3	case4	case5	case6	case7
$L_{\text{eff}}^*$	0.5	0.6	0.7	0.8	0.9	1.0	1.1
$R_l/\Omega$	2	2.5	3	4	×	×	×
$M/N \cdot (\text{m s}^{-1})^{-1}$	62.2320	55.3174	49.7856	41.4880	×	×	×
$M^*$	0.8714	0.7746	0.6971	0.5809	×	×	×
$\varepsilon_{\text{ign}}$	3	4	5	6	×	×	×
$t_c/\text{ms}$	3	4	5	6	×	×	×
$t_c^*$	0.1326	0.1768	0.2210	0.2652	×	×	×
$a$	25%	30%	35%	×	×	×	×

Table 3. Dimensionless operating matrix of FPLA

The dimensionless effective stroke length was defined as the first independent variable while the dimensionless load coefficient, dimensionless translator ignition position, dimensionless combustion duration and dimensionless input energy were taken as second independent variables. When analyzing the effects of each second variable, the other second variables would be set to equal to the base case which was decided by the FPLA prototype.

## 4. CFD calculation of combustion process for validation

### 4.1 Combustion modeling of FPLA

Since the combustion model used in the numerical program is zero dimensional Wiebe function and some parameters like combustion duration are of great uncertainty, the accuracy of the numerical calculated results is suspectable. Nowadays, multi-dimensional CFD computational tools have become an integral part of the engine design process due mainly to advances in computing capabilities and improvements in the modeling techniques utilized. In this study, in order to validate the results of dimensionless analysis, a multi-

dimensional commercial CFD software AVL\_Fire was used to evaluate the effects of translator ignition position with different effective stroke length to bore ratio.

As the FPLA doesn't have a crankshaft mechanism, the dynamics of the piston is totally different from conventional engine. The dynamics were defined based on the results of a zero dimensional FPLA modeling mentioned in the former paragraphs. The dynamics and thermodynamics equations of FPLA in section 2 were solved using a numerical simulating program in Matlab, and some of the parameters were defined according to the experimental data measured. Then the dynamics of the FPLA were incorporated into AVL\_Fire to define the movement of the piston. The piston motion profile was described with two arrays of numbers, one of which represented the ECA (Here ECA is equivalent crank angle which is used to note the port timings. However, it is only a time notation since the free-piston engine does not have a crankshaft to define the piston's motion and  $ECA = t f 360$  [23, 24, 25]) and the other represented the displacement of the piston, and then the file was imported into the CFD code directly. Since there was no coupling between CFD code and free piston's motion, the dynamics was adjusted depending on the desired operating frequency and the stroke of the free-piston engine in the zero-dimensional FPLA simulating program. The dynamic mesh tool *Fame Engine* in AVL\_Fire was used to create the moving mesh according to the numerical simulated free-piston motion profile. The update of the volume was handled automatically at each time step based on the new positions of the piston.

Only compression, combustion and expansion processes of the free-piston engine were calculated in order to minimize the number of computational cells (intake port, scavenging ports, exhaust port and scavenging case were not included in the combustion process). The computational model of the cylinder is shown in Fig.7, and the basic geometry is defined based on the FPLA prototype. Due to the symmetry of the cylinder ports layout, it is only necessary to model half of the geometry in order to minimize the computational cost.

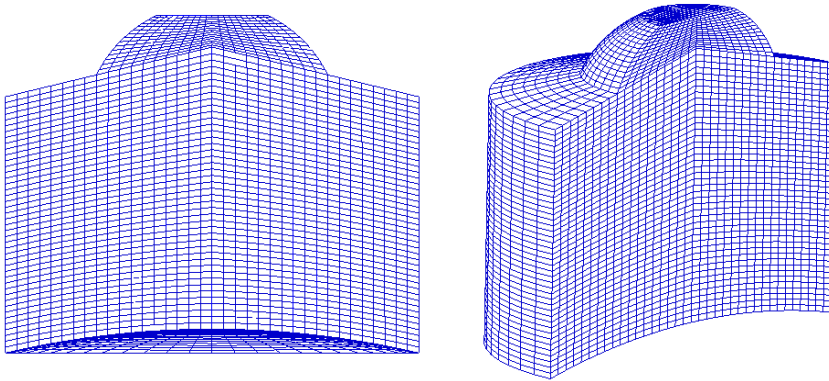


Fig. 7. Computational mesh of cylinder

#### 4.2 Boundary conditions and combustion model

The boundary conditions were chosen to reflect the physical conditions which exist in the validation model and the prototype engine. Constant wall temperatures were also used. The standard  $k-\varepsilon$  model was employed to capture turbulence. As the engine operates on a two



stroke cycle, the simulations run from the point of exhaust port closing to the point of exhaust port opening.

The initial conditions in the cylinder, such as pressure, temperature, EGR ratio and kinetic energy were defined based on the results of multi-dimensional scavenging calculation using CFD tools. In order to investigate the effects of ignition timings, the initial conditions were the same with base case. The input energy in the cylinder had just the same value as it is in the numerical simulation program.

Probability Density Function (PDF) model which takes into account the simultaneous effects of both finite rate chemistry and turbulence was chosen to describe the combustion process. The benefits of the PDF approach lie in the fact that it provides a complete statistical description of the scalar quantities under consideration. Thus, it allows first (mean values), second (variance), and even higher (skewness) order moments to be easily extracted, and that the term expressing the rate of chemical reaction appears in closed form.

### 4.3 CFD calculation cases

Based on the basic geometry of the FPLA, two kinds of effective stroke length to bore ratio and four ignition compression ratios were chosen in the CFD calculation. The other parameters are the same with the base case mentioned before.

As the piston dynamics is changing with different operating conditions, the piston motion profiles have to be defined first in the numerical simulation program, and then the required data in the CFD calculation can be derived, which are listed in Tab.4. The other parameters are based on the FPLA prototype. The revolution of the engine doesn't has real meanings as the free piston engine does not have a crankshaft, and it is bring forward just to complete the combustion process required by the CFD software.

Practically, the maximum compression ratio is confined by the geometry of the chamber since the roof of cylinder and piston is not flat, as is shown in Fig.7. Thus, we have to make sure that the compression ratios of the typical effective stroke length chosen do not exceed their maximum value.

$L_{\text{eff}}/\text{mm}$	$D/\text{mm}$	$L_{\text{eff}}/D$	$\varepsilon_{\text{ign}}$	$t_{\text{ign}}/\text{ECA}$	$\varepsilon$	$f/\text{Hz}$	EGR	$Q_{\text{in}}/\text{J}$	Revolution/rpm
23	34	0.6765	3	319.5	21.0	46.7	20%	17.84	2803.7
23	34	0.6765	4	328.5	22.2	45.0	20%	17.84	2697.8
23	34	0.6765	5	333.6	21.6	43.4	20%	17.84	2606.4
23	34	0.6765	6	337.4	20.7	42.2	20%	17.84	2531.6
35	34	1.0294	3	320.7	11.4	34.1	20%	27.15	2046.4
35	34	1.0294	4	328.6	13.8	33.7	20%	27.15	2025.0
35	34	1.0294	5	333.7	14.6	33.0	20%	27.15	1980.2
35	34	1.0294	6	335.4	14.5	32.1	20%	27.15	1928.0

Table 4. CFD calculation cases

The piston motion profiles with different operating conditions listed in Tab.4 are shown in Fig.8.

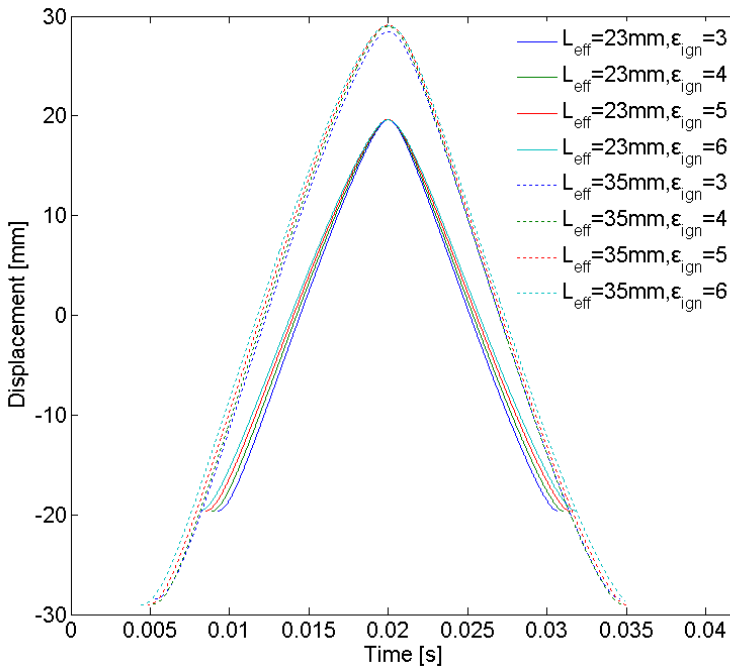


Fig. 8. Piston dynamics with different operating conditions

## 5. Results and discussion

### 5.1 Zero dimensional models validation

In order to make sure most part of the numerical simulation program is correct, the numerical simulated in-cylinder pressure is compared with experimental derived pressure in Fig.9. The thermodynamic model used in the numerical analysis is a single zone model. In single zone models the cylinder mixture composition, pressure and temperature are considered to be uniform and the energy release rate is modeled using experiential model. Matching the experimentally and numerically derived pressure profiles was a complicated task due to the simplification and unknown variables, such as the actual heat addition, the combustion duration and the actual load when the engine was running. Seen in Fig.9, the numerical simulation model used to simulate the operation of FPLA proved to be in agreement with the experimental data with a certain combination of the variables.

### 5.2 Effects of dimensionless parameters

#### 5.2.1 Effects of dimensionless effective stroke length

The stroke to bore ratio is one of the core design variables in internal combustion engines, relating combustion chamber surface area to its volume and piston area to stroke length [11]. In order to predict the performance of FPLA with various sizes and dimensions, seven cases of dimensionless effective stroke length were chosen as the basic variable to investigate the performance of the FPLA in wide operating ranges.

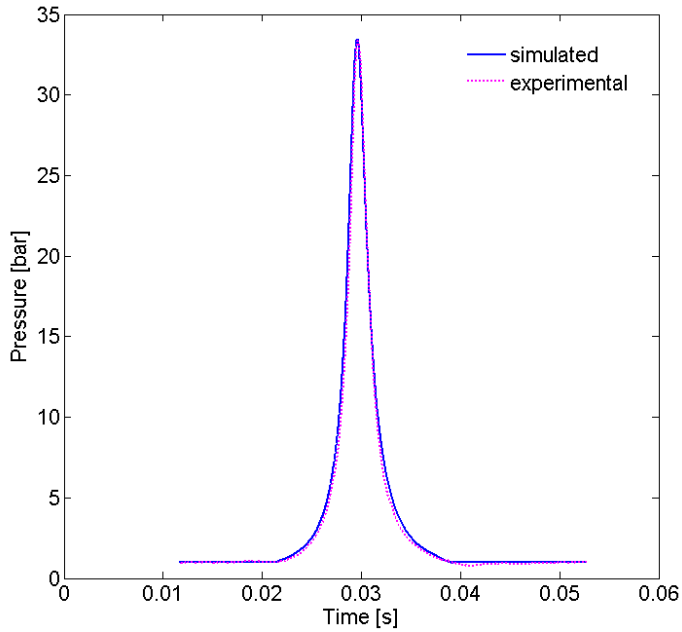


Fig. 9. Comparison of experimental and numerical simulated pressure data

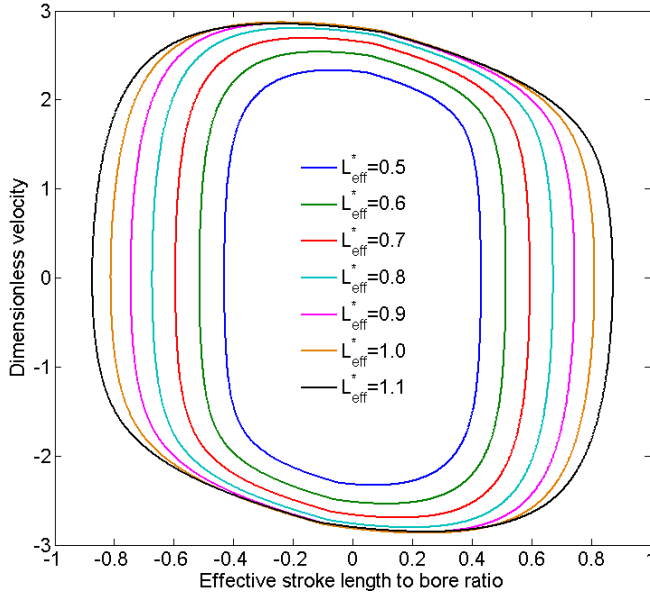


Fig. 10. Dimensionless velocity vs. dimensionless displacement with different effective stroke length to bore ratio

Fig.10 illustrates the change of dimensionless velocity versus dimensionless displacement for different effective stroke length to bore ratios while the other parameters remain the same with the base case. It can be seen that as the dimensionless effective stroke length increases, the dimensionless velocity increases. However, the profile of all the curves is similar to each other, which means this kind of free-piston engine has its own specific characteristics.

As is shown in Figs.11~13, the dimensionless compression ratio, dimensionless frequency and dimensionless in-cylinder peak pressure keep decreasing as the effective stroke length to bore ratio increases. These are because as the dimensionless effective stroke length increases, the translator has to travel longer strokes and more energy is wasted overcoming friction. And as the effective stroke length to bore ratio increases, the cylinder contains more charge and the charge would contain more energy during compression stroke; therefore the dimensionless compression ratio would decrease.

The highest dimensionless effective efficiency is achieved while  $L_{eff}^*$  is 0.8 under the basic working conditions, and the peak point is mainly affected by the other four dimensionless parameters, which will be discussed in the following sections.

As the effective stroke length to bore ratio increases, the dimensionless energy input to the engine every cycle increases as a result of increasing the volume of the cylinder. Since the dimensionless effective power output is also strongly determined by the dimensionless frequency and dimensionless effective efficiency, the highest effective power output is achieved while  $L_{eff}^*$  is 0.9 under the basic working conditions.

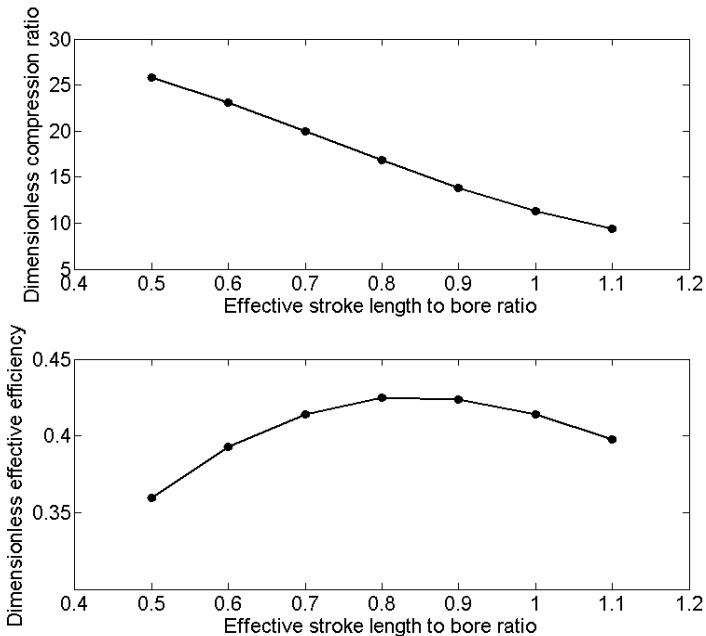


Fig. 11. Effects of effective stroke length to bore ratio to dimensionless compression ratio and dimensionless effective efficiency

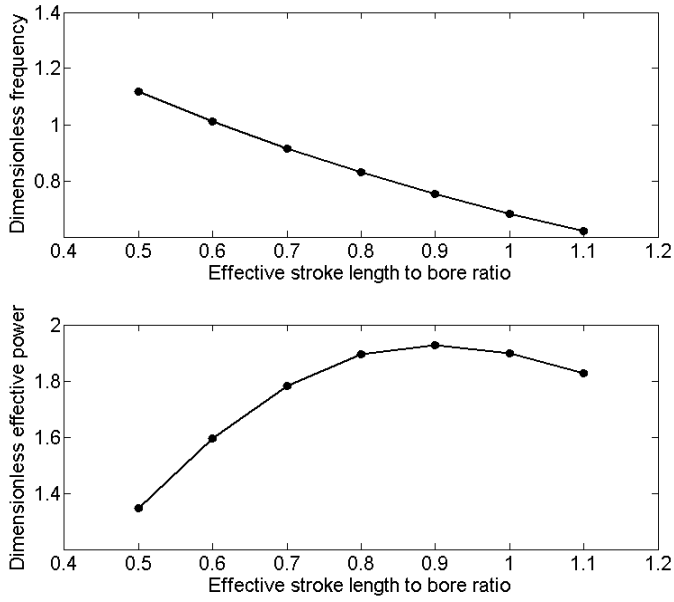


Fig. 12. Effects of effective stroke length to bore ratio to dimensionless frequency and dimensionless effective power output

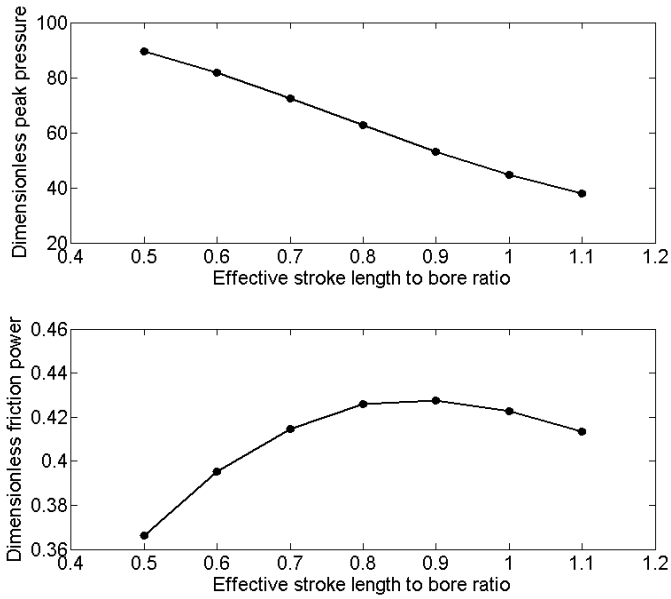


Fig. 13. Effects of effective stroke length to bore ratio to dimensionless peak pressure and dimensionless frictional power

### 5.2.2 Effects of dimensionless load coefficient

Increasing the dimensionless load coefficient means the load demand of the linear alternator is increasing and the electromagnetic force produced by the linear alternator is increasing. Four different dimensionless load coefficients ( $M^*1 > M^*2 > M^*3 > M^*4$ ) were chosen to investigate the effects of changing the load of the linear alternator. The load coefficient was varied by changing the value of the load resistance. According to the results calculated, the dimensionless load coefficient has large impact on different parameters studied and can affect the operating condition of FPLA.

According to Figs.14~15, as the dimensionless load coefficient increases, the dimensionless compression ratio and dimensionless frequency decrease since bigger electromagnetic force is acting on the translator. The highest dimensionless effective efficiency is changing with different dimensionless load coefficient and effective stroke length to bore ratio. As is shown in Fig.14, when the effective stroke length to bore ratio is less than 0.67, smaller dimensionless load coefficient would lead to a higher dimensionless effective efficiency and when the effective stroke length to bore ratio is more than 1.0, the larger the load coefficient the higher the dimensionless effective efficiency. The reason behind these is believed to be caused by the percentage of heat released before top dead center (TDC), which is strongly determined by the frequency of the translator.

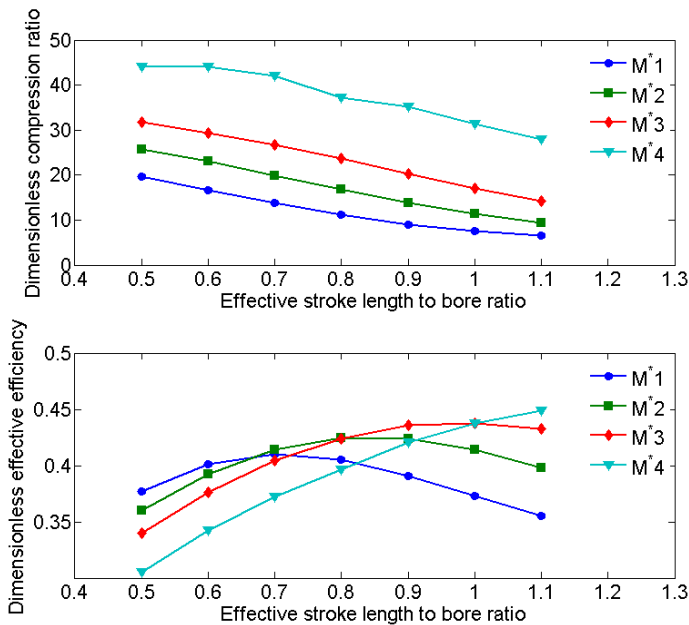


Fig. 14. Effects of dimensionless load coefficient to dimensionless compression ratio and dimensionless effective efficiency

As is shown in Fig.15, smallest dimensionless load coefficient lead to the highest dimensionless power output although the dimensionless effective efficiency is the lowest since the dimensionless frequency with smaller load coefficient is higher. It is more obvious when the effective stroke length to bore ratio is more than 1.0 since smaller load coefficient

lead to higher dimensionless effective efficiency and higher dimensionless frequency. Therefore, we can conclude that the main factor that controls the power output of FPLA is its frequency.

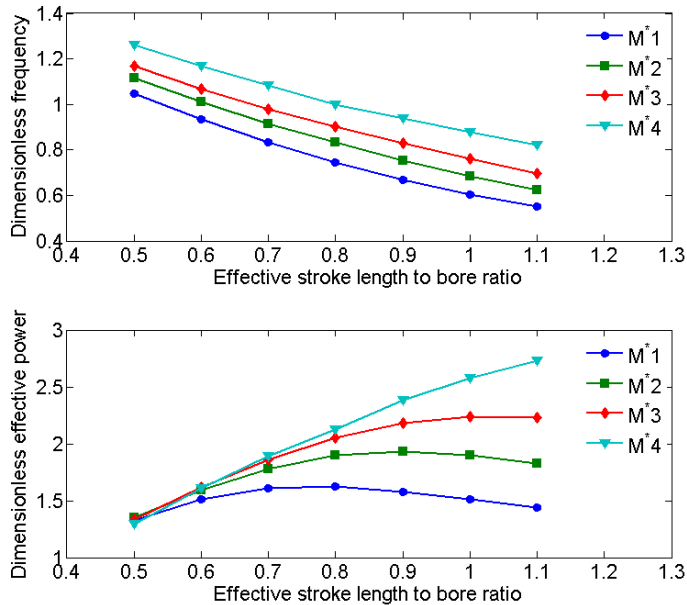


Fig. 15. Effects of dimensionless load coefficient to dimensionless frequency and dimensionless effective power output

### 5.2.3 Effects of dimensionless translator ignition position

Ignition timing is one of the major parameters that control the engine's operating conditions, such as frequency and compression ratio. Since the dimensionless ignition timing is changing with different dimensionless stroke length, the ignition timing is defined by the compression ratio the engine has already achieved when the spark plug ignites in the calculation, and it means that the lower the ignition compression ratio is the bigger the ignition advance is.

According to some literatures [3][5], it's held that an earlier combustion in diesel free-piston engines would lead to more waste of energy to reverse the translator, thus the efficiency and frequency would drop. However, according to the results of spark ignited FPLA obtained in this paper, with different effective stroke length to bore ratio the best ignition advance differs with each other, since an early ignition is associated with negative work in the compression stroke and a late ignition is associated with low peak in-cylinder pressure, as is shown in Fig.16.

As is described in Figs.17~18, with smaller effective stroke length to bore ratio (closer to 0.5), a bigger ignition advance would lead to higher dimensionless compression ratio, higher dimensionless effective efficiency, higher dimensionless frequency and higher dimensionless effective power output. The reason is that with small dimensionless effective

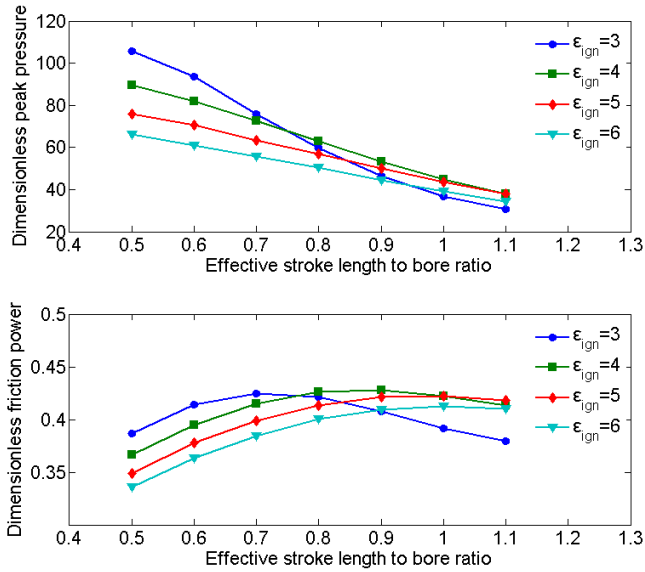


Fig. 16. Effects of dimensionless translator ignition position to dimensionless peak pressure and dimensionless frictional power

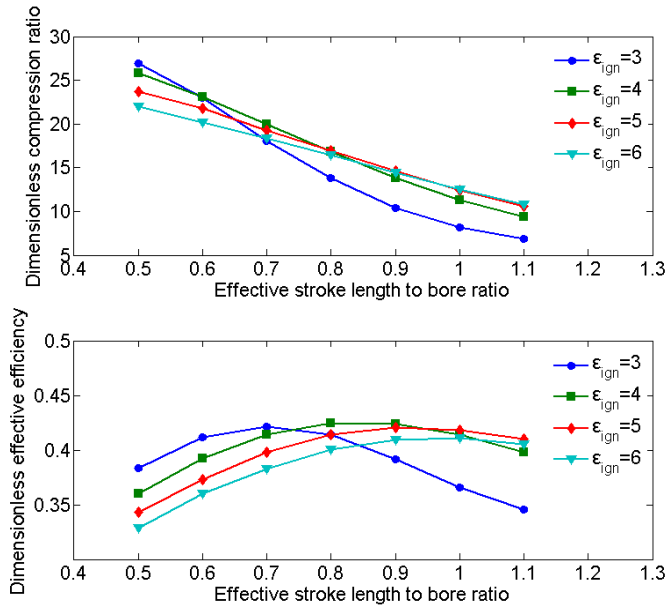


Fig. 17. Effects of dimensionless translator ignition position to dimensionless compression ratio and dimensionless effective efficiency



stroke length, the dimensionless frequency of FPLA is high and most of the energy is released after TDC. Thus, the in-cylinder peak pressure is higher with a bigger ignition advance, which will help improve the performance of the engine. With a high effective stroke length to bore ratio (closer to 1.1), the frequency of the engine decreases a lot since the translator has to travel a longer stroke and a bigger proportion of energy will be released before TDC, which is associated with negative work in the compression stroke. According to the results derived, when the dimensionless effective stroke length is longer than 1.0, the ignition compression ratio of 5 would leads to the best engine performance.

The dimensionless effective power output is determined by dimensionless effective efficiency and dimensionless frequency, as has been discussed before. As is shown in Fig.18, the biggest dimensionless power output is achieved when effective stroke length to bore ratio is 0.9 and ignition compression ratio is 4. Since the dimensionless frequency has little deviation with different ignition compression ratios, the dimensionless effective power output has similar trends with the dimensionless effective efficiency.

In order to analysis the effects of different ignition timings, the combustion duration was assumed to be invariant. However, the combustion duration is strongly depend on the working conditions of the engine, thus CFD tools were taken to analysis the effects of different ignition timings to verify the dimensionless results later.

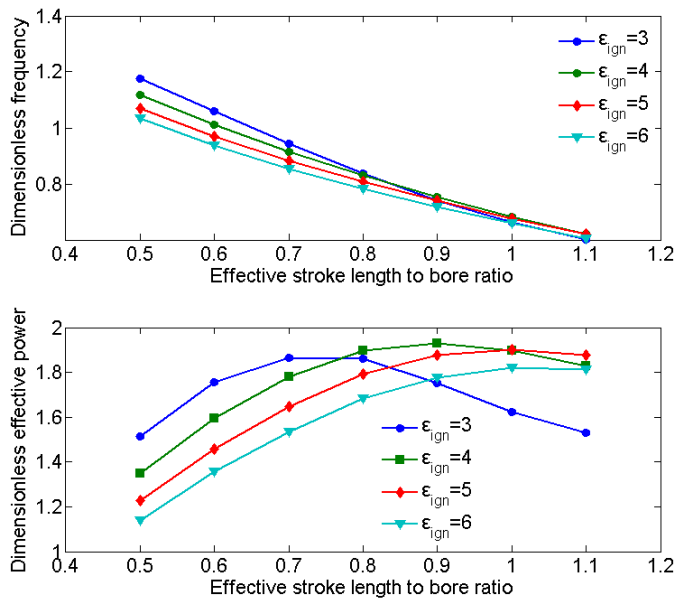


Fig. 18. Effects of dimensionless translator ignition position to dimensionless frequency and dimensionless effective power output

#### 5.2.4 Effect of dimensionless combustion duration

The modeling of the heat release in free-piston engine is one of the factors with the highest degree of uncertainty in the simulation model [11]. The piston motion of free-piston engines

differs significantly from that of conventional engines and very little research exists on how this influences the combustion process. In the dimensionless calculation, the heat release rate is defined by the combustion duration and shorter combustion duration will lead to a faster heat release rate. Based on the base case, four cases of combustion duration were chosen to instigate its effects to the engine's performances.

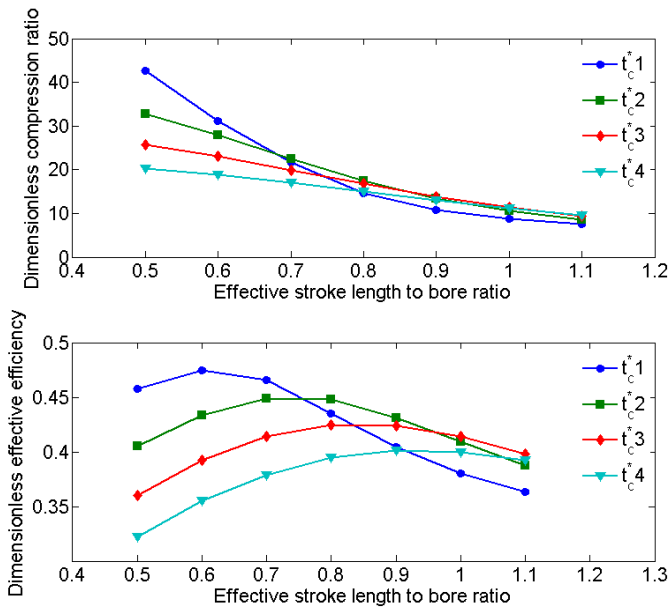


Fig. 19. Effects of dimensionless combustion duration to dimensionless compression ratio and dimensionless effective efficiency

Seen in Fig.19, a shorter combustion duration which means a faster heat release rate would lead to a higher compression ratio and higher effective efficiency when the dimensionless effective stroke length is less than 0.68 and 0.75. However, as the dimensionless effective stroke length increases, the dimensionless frequency will decrease and more energy will be released before TDC. For shorter combustion duration a lot more percentage of energy is released before TDC, which is associated with more negative work in the compression stroke. Thus, shorter combustion duration would lead to a lower dimensionless compression ratio and lower dimensionless effective efficiency with a longer dimensionless effective stroke length and fixed ignition compression ratio.

As is shown in Fig.20, shorter combustion duration leads to a higher frequency with smaller dimensionless effective stroke length and as dimensionless effective stroke length grows, shorter combustion duration leads to faster decreasing of dimensionless frequency as more energy is released before TDC to stop the translator. The dimensionless effective power output is determined by the dimensionless frequency and dimensionless effective efficiency and it has a similar trend with dimensionless efficiency.

Therefore, with a longer effective stroke length to bore ratio it is recommended to postpone the ignition timing to achieve a good performance of the free-piston engine.

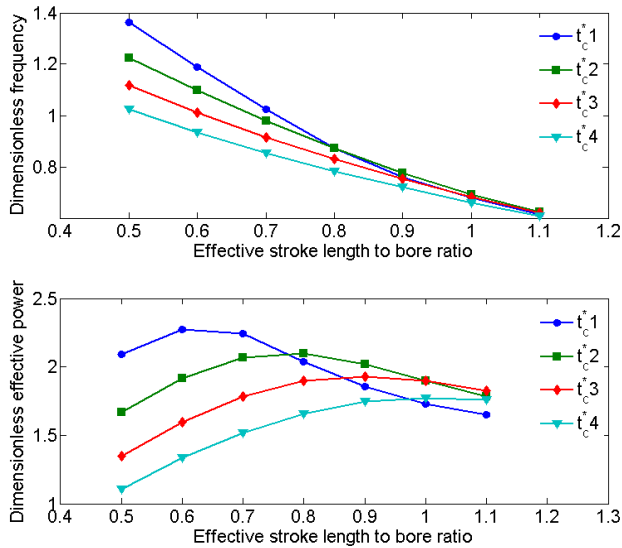


Fig. 20. Effects of dimensionless combustion duration to dimensionless frequency and dimensionless effective power output

### 5.2.5 Effects of dimensionless input energy

The free-piston engine investigated in this paper is a spark-ignited engine and the input energy is varied by changing the opening proportion of the throttle. For FPLA, a much narrow range of operating speeds is expected to be utilized, which is due to the electrical generating scheme employed by the device [23]. Therefore, the opening proportion of the throttle is confined to low speed range. According to the load of FPLA, efficient generation will be achieved by operation at a fixed oscillating rate.

The effects of different dimensionless input energy while other parameters remain the same with the base case are shown in Figs.21~22. As expected, with more input energy, the dimensionless frequency, dimensionless compression ratio and dimensionless effective power output of the engine are increasing since more energy is released in the combustion process.

The amount of energy input to the engine is strictly determined by the load of FPLA. If we keep increasing the amount of input energy, the current load coefficient is not suitable for the current load coefficient and the speed of the translator will keep increasing since extra energy cannot be extracted, and at last the piston will crush with the cylinder head, which is strictly forbidden. However, if we decrease the amount of input energy, the translator will stop since the amount of energy is not enough to sustain the stable operation of the engine. Therefore, the operation range of the engine is confined by the load of the linear alternator, and the amount of the input energy has to be adjusted with the load coefficient to obtain a higher efficiency or higher power output.

### 5.3 CFD calculated results

In order to verify the results of dimensionless translator ignition position of spark ignited free-piston engines, multi-dimensional CFD tools were used to calculate the combustion

process of the FPLA with four different ignition timings and two kinds of effective stroke length to bore ratio.

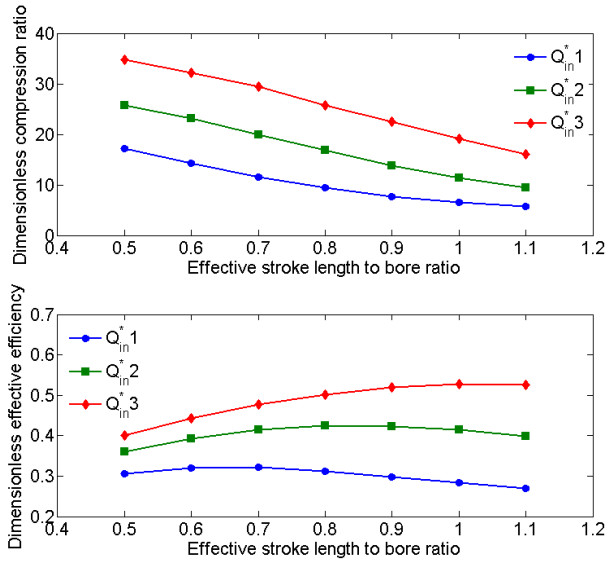


Fig. 21. Effects of dimensionless input energy to dimensionless compression ratio and dimensionless effective efficiency

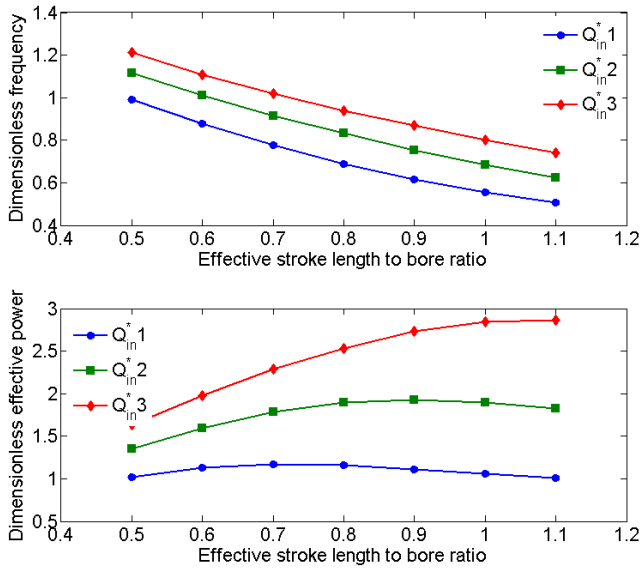


Fig. 22. Effects of dimensionless input energy to dimensionless frequency and dimensionless effective power output

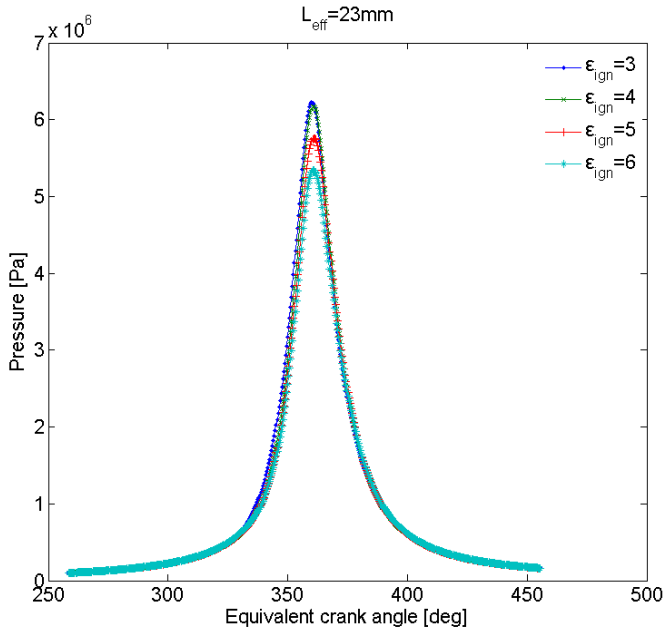


Fig. 23. In-cylinder pressure with different translator ignition position while  $L_{eff}^*=0.6765$

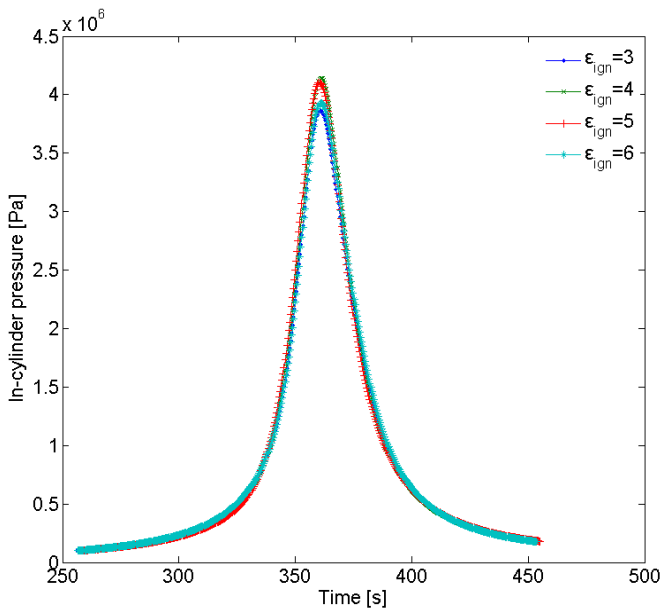


Fig. 24. In-cylinder pressure with different translator ignition position while  $L_{eff}^*=1.0294$

## Nomenclature

$a$	combustion constant	$R$	air gas constant
$A$	top area of the piston	$R_L$	load resistance
$A_{cyl}$	heat transfer area	$R_s$	internal resistance of coils
$b$	combustion form factor	$t$	time
$B$	magnetic induction intensity	$t_0$	time combustion begins
$c_V$	constant volume specific heat	$t_c$	combustion duration
$D$	cylinder diameter	$t_{ign}$	ignition timing
$f$	frequency	$T$	temperature
$F_e$	electromagnetic force	$T_0$	scavenge temperature
$F_f$	friction force	$T_w$	wall temperature
$g$	air gap length	$U$	internal energy
$h$	heat transfer coefficient	$\bar{U}$	mean piston speed
$h_m$	thickness of the permanent magnet	$V$	displaced volume of the cylinder
$H$	length of the coils cutting magnetic lines	$V_{eff}$	effectively compressed volume of the cylinder
$H_c$	magnetic field strength	$V_{ign}$	volume of the cylinder when ignite
$H_e$	enthalpy output	$W$	work done
$H_i$	enthalpy input	$W_e$	effective work
$i_L$	current in the load circuit	$W_f$	frictional work
$L$	induction	$W_i$	indicated work
$L_{tot}$	total stroke length	$x$	displacement of the translator
$L_{eff}$	effective stroke length	$x_{ign}$	translator ignition position
$m$	translator mass	$x_s$	half of maximum stroke length
$m_{in}$	mass of the charge	$a$	opening proportion of throttle
$M$	load coefficient	$\gamma$	specific heat ratio
$M_F$	mean magneto motive force	$\varepsilon$	compression ratio
$n$	polytrophic exponent	$\varepsilon_{ign}$	ignition compression ratio
$N_{coil}$	number of turns in the coil	$\varepsilon_{ind}$	induced voltage
$p$	in-cylinder absolute pressure	$\Phi$	flux passing through the coil
$p_0$	scavenge pressure	$\lambda$	total flux pass through the coil
$p_L$	pressures in the left cylinder	$\mu_0$	vacuum permeability
$p_R$	pressures in the right cylinder	$\tau$	pole pitch
$P_e$	effective power output	$\tau_p$	width of PM
$P_f$	frictional power	$\eta_e$	effective efficiency
$Q$	energy	$\eta_i$	indicated efficiency
$Q_c$	heat released in combustion	$\frac{dx}{dt}$	velocity
$Q_{ht}$	heat transfer	$\frac{d^2x}{dt^2}$	acceleration
$Q_{in}$	total input energy		

(The variable with superscript “\*” is its dimensionless form.)

The in-cylinder pressure curves with different ignition compression ratio while  $L_{eff}^* = 0.6765$  are shown in Fig.23. It is clear that smaller ignition compression ratio or bigger ignition advance leads to higher peak pressure which is in agreement with the dimensionless results.

The in-cylinder pressure curves with different ignition compression ratio while  $L_{eff}^*=1.0294$  are shown in Fig.24. The sequence of the peak pressure achieved with different ignition compression ratio is  $p_{\epsilon_{ign}=4} > p_{\epsilon_{ign}=5} > p_{\epsilon_{ign}=6} > p_{\epsilon_{ign}=3}$ , which supports the dimensionless results. The combustion duration calculated via CFD is about 4.4~5.6ms with different ignition timings and effective stroke length, which has some deviation with the value in numerical simulating program which is defined based on the heat release rate of FPLA prototype. The deviations can be eliminated by using an iterative procedure between the numerical simulating program and CFD calculation when calculating a specific free-piston engine.

## 6. Conclusion

A detailed dimensionless modeling and dimensionless parametric study of spark ignited FPLA was presented to build up a guideline for the design of FPLA prototype with desired operating performances. The parameters of the numerical simulation program were amended by comparing the simulated in-cylinder pressure with experimentally derived data. At last CFD calculation of the combustion process was carried out to verify the effects of translator ignition position with two kinds of typical effective stroke length to bore ratios. According to the dimensionless results, it can be concluded that:

1. For FPLA, a much narrow range of low operating speeds is expected to be utilized, which is due to the electrical generating scheme employed by the device. Therefore, a bigger stroke to bore ratio is favorable to decrease the to and fro frequency of the translator.
2. According to the load of FPLA, efficient power generation will be achieved by operating at a fixed oscillating rate. With smaller effective stroke length to bore ratio, bigger load coefficient is advantageous to achieve a higher effective efficiency while smaller load coefficient would lead to higher effective efficiency with bigger effective stroke length to bore ratio. Smaller load coefficient would lead to higher effective power output.
3. It has been found that an optimum ignition advance is available for the free-piston engine to achieve its best performance since earlier ignition is associated with more negative work in the compression stroke and a later ignition is associated with low peak in-cylinder pressures.
4. The efficiency of the engine is mainly associated with the proportion of the energy released before TDC which is associated with negative work to stop the translator. With a longer effective stroke length to bore ratio it is recommended to postpone the ignition timing to achieve a good performance of the engine.
5. According to the CFD calculated results with typical effective stroke length to bore ratio and ignition timings, the dimensionless results were reasonable.

## 7. References

- [1] Hannson J, Leksell M, Carlsson F. Minimizing power pulsation in a free piston energy converter. Proceedings of the 11<sup>th</sup> European Conference on Power Electronics and Applications (EPE05), Dresden, Germany, 2005
- [2] Mikalsen R, Roskilly AP. The control of a free-piston engine generator. Part 2: Engine dynamics and piston motion control. Appl Energy (2009), doi: 10.1016/j.apenergy.2009.06.035
- [3] Goertz M, Peng LX. Free piston engine its application and optimization. SAE paper 2000-01-0996, 2000

- [4] Atkinson C, Petreanu S, Clark NN, Atkinson RJ etc. Numerical simulation of a two-stroke engine-alternator combination. SAE Technical Paper 1999-01-0921, 1999
- [5] Shoukry E, Taylor S, Clark N. Numerical simulation for parametric study of a two-stroke direct injection linear engine. SAE paper 2002-01-1739, 2002
- [6] Max E. FPEC, Free piston energy converter. In Proceedings of the 21<sup>st</sup> Electric Vehicle Symposium & Exhibition, EVS21, Monaco, 2005
- [7] Blarigan PV, Paradiso N, Goldsborough SS. Homogeneous charge compression ignition with a free piston: A new approach to ideal Otto cycle performance. SAE paper 982484, 1998
- [8] Blarigan PV. Advanced internal combustion electrical generator. Proceedings of the 2002 U.S. hydrogen program review, NREL/CP-610-32405, 2002
- [9] Fredriksson J, Denbratt I. Simulation of a two-stroke free piston engine. SAE paper 2004-01-1871, 2004
- [10] Nemecek P, Vysoky O. Control of two-stroke free-piston generator. Proceeding of the 6<sup>th</sup> Asian control conference, 2006
- [11] Mikalsen R, Roskilly AP. The design and simulation of a two-stroke free piston engine for electric power generation. Appl. Therm. Eng. (2007), doi: 10.1061/j.applthermaleng.2007.04.009
- [12] Mikalsen R, Roskilly AP. A computational study of free-piston diesel engine combustion, Appl Energy (2008), doi: 10.1016/j.apenergy.2008.08.004
- [13] Xiao J et al. Motion characteristic of a free piston linear engine. Appl Energy (2009), doi:10.1016/j.apenergy.2009.07.005
- [14] Cawthorne WR, Famouri P, Chen JD. Development of a linear alternator-engine for hybrid electric vehicle application. IEEE transactions on vehicular technology, vol.48, NO.6, 1999
- [15] Wang JB, Howe H. A linear permanent magnet generator for a free-piston energy converter. 2005 IEEE International Conference on Electric Machines and Drives, p1521-1528, 2005
- [16] Deng Z, Bold I, Nasar SA. Fields in permanent magnet linear synchronous machines. IEEE Transactions on magnets. Vol. MAG-22, NO.2, 1986
- [17] Němeček P, Vysoký O. Modeling and control of free-piston generator. IFAC Mechatronic systems, Sydney, Australia, 2004
- [18] Caresana F, Comodi G, Pelagalli L. Design approach for a two-stroke free piston engine for electric power generation. Society of Automotive Engineers of Japan 2004-32-0037, 2004
- [19] Hohenberg GF. Advanced approaches for heat transfer calculations. SAE Special Publications. SP-449, pp. 61-79, 1979
- [20] Stone R. Introduction to internal combustion engine. ISBN 0-7680-0495-0, Society of Automotive Engineers, Inc. Warrendale, Pa, 1999
- [21] Nagy CT. Linear engine development for series hybrid electrical vehicles. Dissertation, West Virginia: West Virginia University, 2004
- [22] Buckingham, Edgar (1914). On Physically Similar Systems: Illustrations of the Use of Dimensional Analysis. Phys. Rev. 4: 345. doi:10.1103/PhysRev.4.345
- [23] Goldsborough SS, Blarigan PV. A numerical study of a free piston IC engine operating on homogeneous charge compression ignition combustion. SAE paper 990619, 1999
- [24] Goldsborough SS, Blarigan PV. Optimizing the scavenging system for a two-stroke cycle, free piston engine for high efficiency and low emissions: A computational approach. International Multidimensional Engine Modeling User's Group Meeting at the SAE Congress 2003, 2003
- [25] Bergman M, Fredriksson J, Golovitchev VI. CFD-Base Optimization of a Diesel-fueled Free Piston Engine Prototype for Conventional and HCCI Combustion. SAE 2008-01-2423, 2008



# Time Resolved Thermodynamics Associated with Diatomic Ligand Dissociation from Globins

Jaroslava Miksovska and Luisana Astudillo

*Department of Chemistry and Biochemistry, Florida International University Miami FL  
USA*

## 1. Introduction

Ligand-induced conformational transitions play an eminent role in the biological activity of proteins including recognition, signal transduction, and membrane trafficking. Conformational transitions occur over a broad time range starting from picosecond transitions that reflect reorientation of amino acid side chains to slower dynamics on the millisecond time-scale that correspond to larger domain reorganization (Henzler-Wildman et al., 2007). Direct characterization of the dynamics and energetics associated with conformational changes over such a broad time range remains challenging due to limitations in experimental protocols and often due to the absence of a suitable molecular probe through which to detect structural reorganization. Photothermal methods such as photoacoustic calorimetry (PAC) and photothermal beam deflection provide a unique approach to characterize conformational transitions in terms of time resolved volume and enthalpy changes (Gensch&Viappiani, 2003; Miksovska&Larsen, 2003). Unlike traditional spectroscopic techniques that are sensitive to structural changes confined to the vicinity of a chromophore, photothermal methods monitor overall changes in volume and enthalpy allowing for the detection of structural transitions that are spectroscopically silent (i.e. do not lead to optical perturbations of either intrinsic or extrinsic chromophores).

Myoglobin (Mb) and hemoglobin (Hb) play a crucial role in the storage and transport of oxygen molecules in vertebrates and have served as model systems for understanding the mechanism through which protein dynamics regulate ligand access to the active site, ligand affinity and specificity, and, in the case of hemoglobin, oxygen binding cooperativity. Mb and individual  $\alpha$ - and  $\beta$ - subunits of Hb exhibit significant structural similarities, i.e. the presence of a five coordinate heme iron with a His residue coordinated to the central iron (proximal ligand) and a characteristic "3-on-3" globin fold (Fig. 1)(Park et al., 2006; Yang&Phillips Jr, 1996). Both proteins reversibly bind small gaseous ligands such as O<sub>2</sub>, CO, and NO. The photo-cleavable Fe-ligand bond allows for the monitoring of transient deoxy intermediates using time-resolved absorption spectroscopy (Carver et al., 1990; Esquerra et al., 2010; Gibson et al., 1986) and time resolved X-ray crystallography (Milani et al., 2008; Šrajcar et al., 2001). Based on spectroscopic data and molecular dynamics approaches (Bossa et al., 2004; Mouawad et al., 2005), a comprehensive molecular mechanism for ligand migration in Mb was proposed including an initial diffusion of the photo-dissociated CO molecule into the internal network of hydrophobic cavities, followed by a return

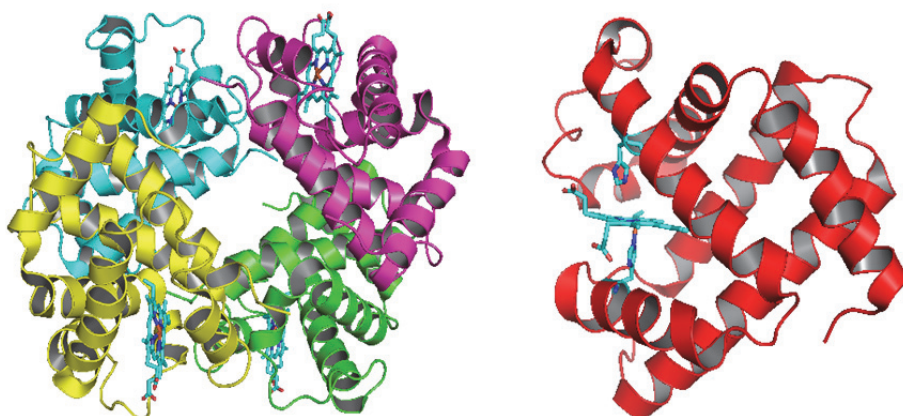
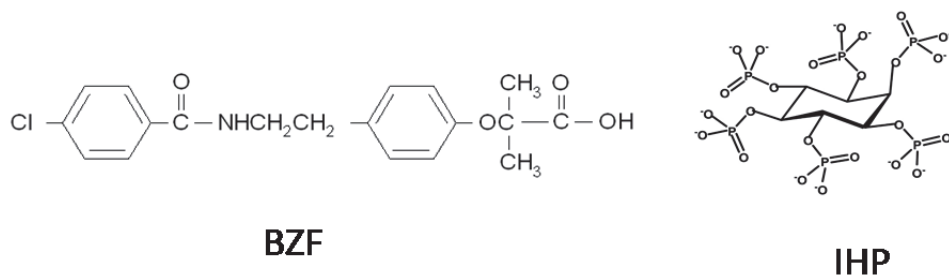


Fig. 1. Left: Ribbon representation of the tetrameric human Hb structure (PDB entry 1FDH). Right: horse heart Mb structure (PDB entry 1WLA). The heme prosthetic groups are shown as sticks. In the case of Mb, the distal and proximal histidine are visualized.

into the distal pocket and subsequent rebinding to heme iron or escape from the protein through a distal histidine gate. The ligand migration into internal cavities induces a structural deformation, which promotes a transient opening of a gate in the CO migration channel. Such transitional reorganization of the internal cavities is ultimately associated with a change in volume and/or enthalpy and thus can be probed using photothermal techniques. Indeed, CO photo-dissociation from Mb has been intensively investigated using PAC by our group and others (Belogortseva et al., 2007; Peters et al., 1992; Vetromile, et al., 2011; Westrick&Peters, 1990; Westrick et al., 1990) and these results lead to a thermodynamic description of the transient “deoxy intermediate” that is populated upon CO photo-dissociation.

The mechanism of ligand migration in Hb is more complex, since it is determined by the tertiary structure of individual subunits as well as by the tetramer quaternary structure. Crystallographic data have shown that the structure of the fully unliganded tense (T) state of Hb and the fully ligated relaxed (R) states differ at both the tertiary and quaternary level (Park et al., 2006). Crystallographic and NMR studies suggest that the fully ligated relaxed state corresponds to the ensemble of conformations with distinct structures (Mueser et al., 2000; Silva et al., 1992). Moreover, Hb interactions with diatomic ligands is modulated by physiological effectors such as protons, chloride, and phosphate ions, and non-physiological ligands including inositol hexakisphosphate (IHP) and bezafibrate (BZF) (Yonetani et al., 2002). Despite a structural homology between Hb and Mb, the network of internal hydrophobic cavities identified in Mb is not conserved in Hb suggesting distinct ligand migration pathways in this protein (Mouawad et al., 2005; Savino et al., 2009). Here we present thermodynamic profiles of CO photo-dissociation from human Hb in the presence of heterotropic allosteric effectors IHP and BZF. In addition, we include an acoustic study of oxygen photo-dissociation from Mb that has not been investigated previously using photothermal methods, despite the fact that oxygen is the physiological ligand for Mb.



Scheme 1.

## 2. Material and methods

Mb, Hb, inositol hexakisphosphate (IHP), and bezafibrate (BZF) were purchased from Sigma-Aldrich and used as received. Fe(III) tetrakis(4-sulfonatophenyl)porphine (Fe(III)4SP) was obtained from Frontier-Scientific Inc. Oxy-myoglobin (O<sub>2</sub>-Mb) samples were prepared by dissolving the protein in 50 mM HEPES buffer pH 7.0. The sample was then purged with Ar for 10 min and reduced by addition of a freshly prepared solution of sodium dithionite. The quality of the deoxy-myoglobin (deoxyMb) was verified by UV-visible spectroscopy. (O<sub>2</sub>-Mb) was obtained by bubbling air through deoxyMb sample. The CO bound hemoglobin sample was prepared by desolving Hb in 100 mM HEPES buffer pH 7.0 in a 0.5 x 1cm quartz cuvette. The concentration of allosteric effectors was 5 mM for BZF and 1 mM for IHP. The sample was then sealed with a septum cap and purged with Ar for 10 min, reduced with a small amount of sodium dithionite to prepare deoxyhemoglobin (deoxyHb), and subsequently bubbled with CO for approximately 1 min. Preparation of O<sub>2</sub>-Mb and CO-Hb adducts was checked by UV-vis spectroscopy (Cary50, Varian).

### 2.1 Quantum yield determination

The quantum yield ( $\Phi$ ) was determined as described previously (Belogortseva et al., 2007). All transient absorption measurements were carried out on 50  $\mu$ M samples in 50 or 100 mM HEPES buffer, pH 7.0, placed in a 2 mm path quartz cell. The cell was placed into a temperature controlled holder (Quantum Northwest) and the ligand photo-dissociation was triggered using a 532 nm output from a Nd:YAG laser (Minilite II, Continuum). The probe beam, an output from the Xe arc lamp (200 W, Newport) was propagated through the center of the cell and then focused on the input of a monochromator (Yvon-Jovin). The intensity of the probe beam was detected using an amplified photodiode (PDA 10A, Thornlabs) and subsequently digitized (Wave Surfer 42Xs, 400 MHz). The power of the pump beam was kept below 50  $\mu$ J to match the laser power used in photoacoustic measurements. The quantum yield was determined by comparing the change in the sample absorbance at 440 nm with that of the reference, CO bound myoglobin of known quantum yield ( $\Phi_{\text{ref}} = 0.96$  (Henry et al., 1983)) according to Eq 1:

$$\Phi = \frac{\Delta A_{\text{sam}} \Delta \epsilon_{\text{ref}} \Phi_{\text{ref}}}{\Delta A_{\text{ref}} \Delta \epsilon_{\text{sam}}} \quad (1)$$

where  $\Delta A_{\text{sam}}$  and  $\Delta A_{\text{ref}}$  are the absorbance change of the sample and reference at 440 nm, respectively, and  $\Delta \epsilon_{\text{sam}}$  and  $\Delta \epsilon_{\text{ref}}$  are the change in the extinction coefficient between the CO bound and reduced form of the sample and the reference, respectively.

## 2.2 Photoacoustic calorimetry

The photo-acoustic set-up used in our lab was described previously (Miksovská&Larsen, 2003). Briefly, the sample in a quartz cell was placed in a temperature controlled holder (Quantum Northwest). The 532 nm output from a Nd:YAG laser (7 ns pulse width, < 50  $\mu$ J power) was shaped using a narrow slit (100  $\mu$ m) and focused on the center of the quartz cell. An acoustic detector (1 MHz, RV103, Panametric) was coupled to the side of a quartz cell using a thin layer of honey and the detector output was amplified using an ultrasonic preamplifier (Panametrics 5662). The signal was then stored in a digitizer (Wave Surfer 42Xs, 400 MHz). The data were analyzed using Sound Analysis software (Quantum Northwest).

## 2.3 Data analysis

The excitation of the photocleavable iron-ligand bond in heme proteins generates at least two processes that contribute to the photoacoustic wave: the volume change due to the heat released during the reaction (Q), and the volume change ( $\Delta V'$ ) due to the photo-triggered structural changes (including bond cleavage / formation, electrostriction, solvation, etc.). The amplitude of the sample acoustic wave ( $A_{sam}$ ) can be expressed as:

$$A_{sam} = KE_a(Q \frac{\beta}{\rho C_p} + \Delta V') \quad (2)$$

where K is the instrument response constant,  $E_a$  is number of Einsteins absorbed,  $\beta$  is the expansion coefficient,  $\rho$  is the density, and  $C_p$  is the heat capacity. For water, the ( $\beta/C_p\rho$ ) term strongly varies with temperature mainly due to the temperature dependence of the  $\beta$  term. To evaluate the instrument response constant, the photo-acoustic traces are measured for a reference compound under experimental conditions (laser power, temperature, etc.) identical to those for the sample measurements. We have used Fe(III)4SP as a reference since it is non-fluorescent and photo-chemically stable. The amplitude of the reference acoustic trace can be described as:

$$A_{ref} = KE_a E_{hv} \frac{\beta}{\rho C_p} \quad (3)$$

where  $E_{hv}$  is the energy of a photon at 532 nm ( $E_{hv} = 53.7$  kcal mol<sup>-1</sup>). The amount of heat deposited to the solvent and the non-thermal volume changes can then be determined by measuring the acoustic wave for the sample and the reference for several temperatures and plotting the ratio of the sample and reference acoustic wave ( $\phi$ ) as a function of ( $C_p\rho/\beta$ ) according to Eq. 4:

$$\frac{A_{sam}}{A_{ref}} = \phi = Q + \Delta V' \frac{\rho C_p}{\beta} \quad (4)$$

For a multi-step process that exhibits volume and enthalpy changes on the time-scale between  $\sim 20$  ns to 5  $\mu$ s, the thermodynamic parameters for each individual step and corresponding lifetimes can be determined due to the sensitivity of the acoustic detector to the temporal profile of the pressure change. The time dependent sample acoustic signal  $E(t)_{obs}$  can be expressed as a convolution of the time dependent function describing the volume change  $H(t)$  with the instrument response  $T(t)$  function (the reference acoustic wave):

$$H(t) = \phi_1 e^{-\frac{t}{\tau_1}} + \frac{\phi_2 k_2}{(k_2 - k_1)} e^{-\frac{t}{\tau_1}} - e^{-\frac{t}{\tau_2}} \quad (5)$$

$$E(t)_{obs} = H(t) \otimes T(t) \quad (6)$$

where  $\phi_1$  and  $\phi_2$  correspond to the  $\left(\frac{A_{sam}}{A_{ref}}\right)$  term in Eq. 4 and the  $\tau_1$  and  $\tau_2$  are the lifetime for the first and subsequent step of the reaction, respectively. To retrieve thermodynamic and kinetic parameters, the reference trace is convoluted with the  $H(t)$  function using estimated parameters ( $\phi_i$  and  $\tau_i$ ) and the calculated  $E(t)_{calc}$  is compared with the  $E(t)_{obs}$ . The  $\phi_i$  and  $\tau_i$  values are varied until a satisfactory fit is obtained in terms of  $\chi^2$  and autocorrelation function. In practice, the lifetime for the prompt process is fixed to 1 ns, whereas other parameters are allowed to be varied.

For processes that occur with a quantum yield that is temperature dependent in the temperature range used in PAC measurements, the thermodynamic parameters for the fast phase ( $\tau < 20$  ns) are determined by plotting  $[E_{hv}(1-\Phi)]/\Phi$  versus  $(C_p \rho / \beta)$  according to Eq. 7 and the volume and enthalpy changes for the subsequent steps are obtained by plotting  $(\Phi E_{hv} / \Phi)$  versus  $(C_p \rho / \beta)$  according to Eq. 8 (Peters et al., 1992).

$$\frac{E_{hv}(1-\Phi)}{\Phi} = -\Delta H + \Delta V \frac{C_p \rho}{\beta} \quad (7)$$

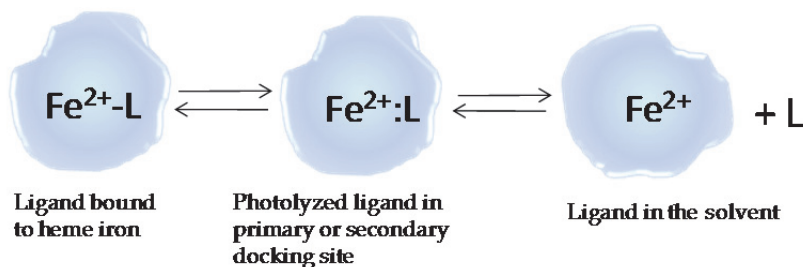
$$\frac{\Phi E_{hv}}{\Phi} = -\Delta H + \Delta V \left(\frac{C_p \rho}{\beta}\right) \quad (8)$$

where  $\Delta H$  and  $\Delta V$  correspond to the reaction enthalpy and volume change, respectively.

### 3. Results

Ligand migration in heme proteins is often described using the sequential three-state model (Henry et al., 1983) shown in Scheme 2. Upon cleavage of the coordination bond between the ligand and heme iron, the ligand is temporarily trapped within the protein matrix and then it either directly rebinds back to the heme iron in the so called "geminate rebinding" or diffuses from the protein matrix into the surrounding solvent. The subsequent bimolecular ligand binding to heme iron occurs on significantly longer time scales, hundreds of microseconds to milliseconds. The quantum yield for the geminate rebinding and for bimolecular association is strongly dependent on the character of the ligand and the protein. For example, CO rebinds to Mb predominantly through a bimolecular reaction with quantum yield close to unity ( $\Phi_{bim} = 0.96$ ) (Henry et al., 1983), whereas the quantum yield for bimolecular O<sub>2</sub> rebinding to heme proteins is significantly lower (Carver et al., 1990; Walda et al., 1994), and NO rebinds predominantly through geminate rebinding (Ye et al., 2002). To determine the thermodynamic parameters from acoustic data, the quantum yields for CO and O<sub>2</sub> bimolecular rebinding to Hb and Mb, respectively, have to be known. The quantum yield for O<sub>2</sub> binding to Mb was measured in the temperature range from 5 °C to 35 °C (Fig. 2) and the values show a weak temperature dependence with the quantum yield decreasing with increasing temperature. At 20 °C the quantum yield is  $0.09 \pm 0.01$  that is within the range of values reported previously ( $\Phi = 0.057$  (Walda et al., 1994) and  $\Phi = 0.12$  (Carver et al., 1990)). We have also measured the quantum yield for CO bimolecular rebinding to Hb, and to Hb in the presence of effector molecules (Fig. 2). The quantum yield increases linearly with temperature and at 20 °C, CO binds to Hb with quantum yield of

0.68 and in the presence of IHP and BZF 0.62 and 0.46, respectively. A similar quantum yield for CO bimolecular rebinding to Hb was reported previously by Unno et al. ( $\Phi_{\text{bim}}=0.7$  at 20 °C) (Unno et al., 1990) and by Saffran and Gibson ( $\Phi=0.7$  for CO binding to Hb and  $\Phi = 0.73$  for CO association to Hb in the presence of IHP at 40 °C) (Saffran&Gibson, 1977).



Scheme 2.

The photo-acoustic traces for  $\text{O}_2$  dissociation from Fe(II)Mb at pH 7.0 are shown in Fig. 3. At low temperatures (6 °C to 15 °C), the sample photoacoustic traces show a phase shift with respect to the reference trace indicating the presence of thermodynamic process(es) that occurs between 50 ns and  $\sim 5 \mu\text{s}$ . The sample traces were deconvoluted as described in the Materials and Methods section and the  $\phi_i$  values were plotted as a function of the temperature dependent factor ( $C_{\text{pp}}/\beta$ ) (Fig. 4). The extrapolated volume and enthalpy changes are listed in Table 1. The photo-cleavage of the Fe- $\text{O}_2$  bond is associated with a fast structural relaxation ( $\tau < 20$  ns) forming a transient “deoxy-Mb intermediate”. This initial transition is endothermic ( $\Delta H = 21 \pm 9 \text{ kcal mol}^{-1}$ ) and leads to a small volume contraction of  $-3.0 \pm 0.5 \text{ mL mol}^{-1}$ . This initial relaxation is followed by  $\sim 250$  ns kinetics that exhibit a volume increase of  $5.5 \pm 0.4 \text{ mL mol}^{-1}$  and a very small enthalpy change of  $-8.9 \pm 8.0 \text{ kcal mol}^{-1}$ . We associate the initial process with the photo-cleavage of the Fe- $\text{O}_2$  bond. A similar volume decrease of approximately  $-3 \text{ mL mol}^{-1}$  has been observed previously for the photo-dissociation of Fe-CO bond in Mb (Westrick&Peters, 1990; Westrick et al., 1990). The observed volume contraction reflects a fast relaxation of the heme binding pocket including: i) cleavage of the hydrogen bond between the distal histidine and oxygen molecule (Phillips&Schoenborn, 1981) ii) reorientation of distal residues within the heme binding pocket (Olson et al., 2007), and iii) fast migration of the photo-released ligand into the primary docking site and then into the internal cavities (Xe4 or Xe1) (Hummer et al., 2004). Also, the positive enthalpy change is consistent with the photo-cleavage of Fe- $\text{O}_2$  bond.

The subsequent 250 ns kinetics may reflect either the nanosecond geminate rebinding of the  $\text{O}_2$  molecule or the ligand diffusion from the protein matrix into the surrounding solvent. The kinetics for the geminate  $\text{O}_2$  rebinding were studied on femtosecond timescale by Petrich et al. (Petrich et al., 1988), and on picosecond and nanosecond timescales (Carver et al., 1990; Miller et al., 1996). These studies identified two distinct sub-states of the “deoxyMb” intermediate: a “barrier-less” and a “photolyzable” sub-state. In the “barrier-less” sub-state, oxygen rebinds to heme iron on sub-picosecond timescale whereas the oxygen association to the “photolyzable” substate occurs on nanosecond and microsecond timescales. Carver et al. (Carver et al., 1990) have reported the time constant for  $\text{O}_2$  nanosecond geminate rebinding to be  $52 \pm 14$  ns at room temperature. This kinetic step has a lifetime that is comparable to the time resolution of our PAC instrument ( $\tau \sim 50$  ns) and therefore it was not resolved in this study. The 250 ns step thus corresponds to the  $\text{O}_2$  escape

from the transient “deoxy-Mb” intermediate into the surrounding solvent and is approximately 3 times faster than the rate of the CO escape (Westrick et al., 1990), which suggests that O<sub>2</sub> diffuses from the protein matrix through a transient channel with a lower activation barrier than CO. This result is consistent with the transient absorption studies that estimated the rate for O<sub>2</sub> release to be approximately two times faster than that for CO (Carver et al., 1990). Interestingly, a similar time-constant of 200 ns to 300 ns was determined for CO escape from Mb at pH 3.5 (Angeloni&Feis, 2003). At acidic pH Mb adopts an open conformation with His 64 displaced toward the solvent giving a direct access to the distal cavity. These data suggest that the reorientation of His 64 may not be a rate limiting step for the O<sub>2</sub> escape.

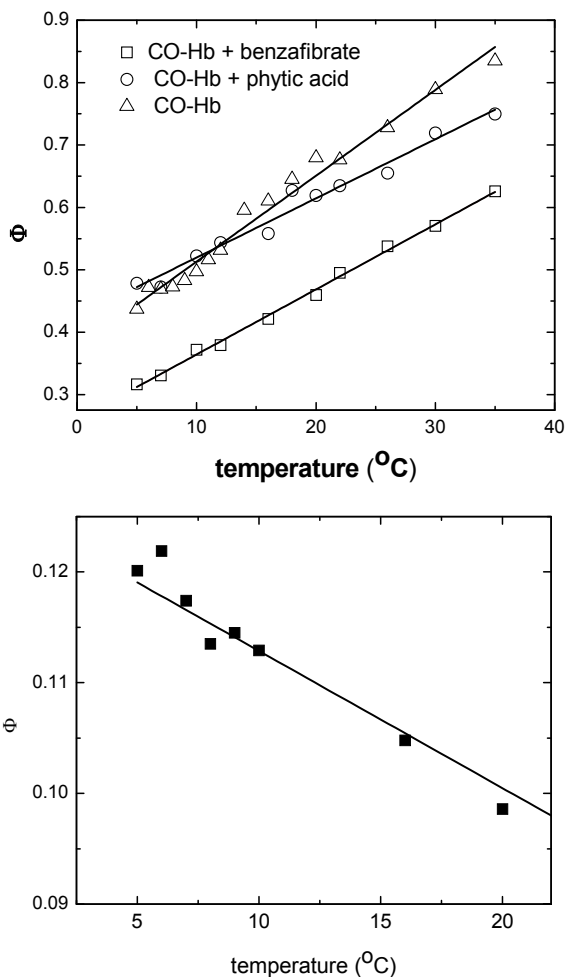


Fig. 2. Quantum yield for bimolecular photo-dissociation of O<sub>2</sub> from the O<sub>2</sub>-Mb complex (bottom) and CO from the CO-Hb complex (top) as a function of temperature. The error of quantum yield is  $\pm 0.05$ . The solid line demonstrates the trend.

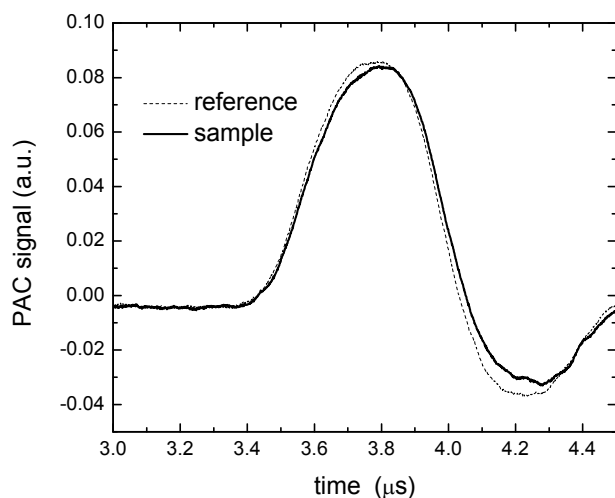


Fig. 3. PAC traces for  $O_2$  photo-dissociation from  $O_2$ -Mb complex at 6 °C. Conditions: 40  $\mu$ M Mb dissolved in 50 mM Hepes buffer pH 7.0. The absorbance of the reference compound, Fe(III)4SP, at excitation wavelength of 532 nm was identical as that of  $O_2$ -Mb.

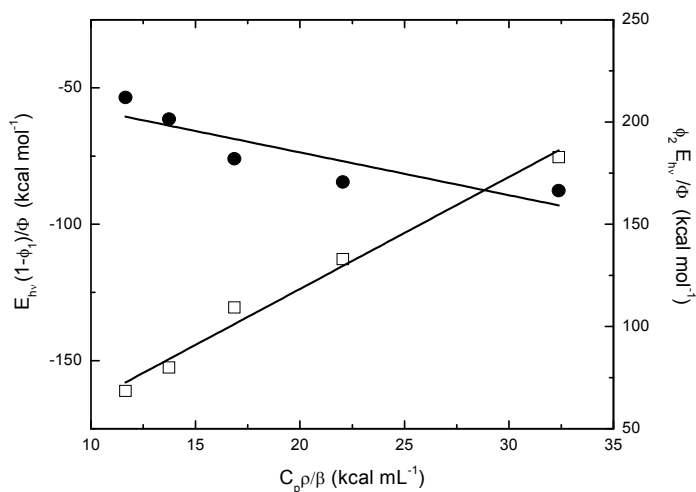


Fig. 4. Plot of the ratio of the acoustic amplitude for the photo-dissociation of the  $O_2$ -Mb complex and the reference compound as a function of  $(C_{p\rho}/\beta)$  term.  $\phi_1$  values that correspond to the prompt phase are shown as solid circles and the  $\phi_2$  values corresponding to the slow phase are shown as open squares. The data were fit with a linear curve and the corresponding volume and enthalpy changes were determined using Eq. 6 and Eq. 7.

The reaction volume change observed for the slow phase includes several factors: i) volume change due to the  $O_2$  escape into the surrounding solvent, ii) volume change associated with the heme hydration in deoxyMb and iii) volume change due to the structural changes. The



reaction volume can be expressed as the difference between the partial molar volume of products and reactants according to:  $\Delta V_{\text{slow}} = V^{\circ}_{\text{O}_2} + V^{\circ}_{\text{deoxyMb}} - V^{\circ}_{\text{O}_2\text{-Mb}} - V^{\circ}_{\text{H}_2\text{O}}$ , where  $V^{\circ}_{\text{O}_2}$  is the partial molar volume of oxygen,  $V^{\circ}_{\text{H}_2\text{O}}$  is the partial molar volume of water,  $V^{\circ}_{\text{deoxyMb}}$  is the partial molar volume of transient "deoxyMb" intermediate and  $V^{\circ}_{\text{O}_2\text{-Mb}}$  is the partial molar volume of oxy-Mb. Using  $V^{\circ}_{\text{O}_2} = 28 \text{ mL mol}^{-1}$  (Projahn et al., 1990) and  $V^{\circ}_{\text{H}_2\text{O}} = 15 \text{ mL mol}^{-1}$  (the partial molar volume of water scaled to the occupancy of water molecule hydrogen bound to distal histidine) (Belogortseva et al., 2007), we estimate that the  $\text{O}_2$  release from Mb results in a structural volume change ( $V^{\circ}_{\text{deoxyMb}} - V^{\circ}_{\text{O}_2\text{-Mb}}$ ) of  $-7.5 \text{ mL mol}^{-1}$ . This value is very similar to that reported previously for CO escape from Mb ( $\Delta V_{\text{structural}} = V^{\circ}_{\text{deoxyMb}} - V^{\circ}_{\text{CO-Mb}} = -6 \text{ mL mol}^{-1}$ ) (Vetromile, et al., 2011) demonstrating that the overall structural changes accompanying the ligand bound to ligand free transition in Mb are very similar for both ligands. This is in agreement with the close resemblance of the X-ray structure of both the CO-bound and  $\text{O}_2$ -bound Mb (Yang&Phillips Jr, 1996). The small enthalpy change measured for the 250 ns relaxation ( $\Delta H = -8.9 \pm 8.0 \text{ kcal mol}^{-1}$ ) includes the enthalpy change for  $\text{O}_2$  solvation ( $\Delta H_{\text{solv}} = -2.9 \text{ kcal mol}^{-1}$  (Mills et al., 1979)) and the enthalpy change associated with  $\text{H}_2\text{O}$  binding to the heme binding pocket ( $\Delta H_{\text{solv}} = -7 \text{ kcal mol}^{-1}$  (Vetromile, et al., 2011) indicating that the structural relaxation coupled to the ligand escape from the protein is entropy driven.

The overall enthalpy change for  $\text{O}_2$  dissociation from Mb was determined to be  $11.6 \pm 8.5 \text{ kcal mol}^{-1}$  and this value is in agreement with the value of  $10 \text{ kcal mol}^{-1}$  reported previously (Projahn et al., 1990). The overall reaction volume change determined here ( $\Delta V_{\text{overall}} = +2.5 \text{ mL mol}^{-1}$ ) is somewhat larger than the reaction volume change determined from the measurement of the equilibrium constant as a function of pressure ( $\Delta V = -2.9 \text{ mL mol}^{-1}$ ) (Hasinoff, 1974) and significantly smaller than the reaction volume change determined as a difference between the activation volume for oxygen binding and dissociation from Mb that was reported to be  $18 \text{ mL mol}^{-1}$  (Projahn et al., 1990). Unlike photoacoustic studies that allow for reaction volume determination at ambient pressure, the high pressure measurements of equilibrium constant and/or rate constants (to determine activation volumes) may cause a pressure induced protein denaturation and/or structural changes, which may influence the magnitude of reaction volume changes in high pressure studies.

	$\Delta V \text{ (mL mol}^{-1}\text{)}$	$\Delta H \text{ (kcal mol}^{-1}\text{)}$
Fast phase	$-3.0 \pm 0.5$	$20.5 \pm 8.5$
Slow phase	$5.5 \pm 0.4$	$-8.9 \pm 8.0$

Table 1. Volume and enthalpy changes associated with  $\text{O}_2$  dissociation from Mb in the temperature range 6 -  $10^\circ\text{C}$ .

We have also probed the thermodynamic parameters associated with the CO photo-dissociation from Hb and the impact of the binding of BZF and IHP on the thermodynamics associated with the ligand migration between the heme binding pocket and surrounding solvent. The photo-acoustic traces for CO photo-dissociation from Hb are shown in Fig. 5. The sample and the reference acoustic wave overlay in phase indicating that the observed thermodynamic processes take place within 50 ns upon photo-dissociation, which is consistent with the fast CO diffusion from the heme matrix into the surrounding solvent.

The fast ligand escape from the heme binding pocket was observed in the presence of effectors (data not shown). Previous transient absorption studies showed that the CO photo-release from the fully ligated R-state Hb is followed by three relaxations with lifetimes of 50 ns, 1  $\mu$ s, and 20  $\mu$ s that were assigned to the unimolecular geminate rebinding, the tertiary structural relaxation, and the R $\rightarrow$ T quaternary change, respectively (Goldbeck et al., 1996). The geminate rebinding occurs too fast to be resolved by our PAC detector, whereas the 20  $\mu$ s R $\rightarrow$ T transition, which strongly depends on the extent of heme ligation, is too slow to be resolved in PAC measurements. The 1  $\mu$ s relaxation is within the time-window accessible by our detection system, however we were unable to resolve this step. Since this relaxation was observed as a small perturbation of the deoxy-Soret band (Goldbeck et al., 1996), it may reflect the structural relaxation localized within the vicinity of the heme binding pocket, which does not lead to measurable volume and enthalpy changes.

The volume and enthalpy changes associated with the diffusion of the photo-dissociated ligand to the surrounding solvent can be determined from the plot of the ratio of the amplitude of the acoustic trace for CO photo-dissociation from Hb and the reference as a function of temperature according to Eq. 7 (Fig. 6). The extrapolated thermodynamic values

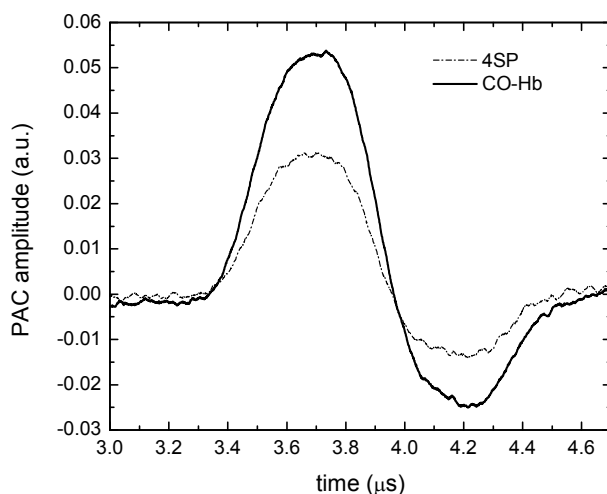


Fig. 5. PAC traces for the CO photo-dissociation from the CO-Hb complex and the reference compound Fe(III)4SP. Conditions: 40  $\mu$ M Hb in 100 mM HEPES buffer pH 7.0 and 20  $^{\circ}$ C. The absorbance of the reference compound matched the absorbance of the sample at 532 nm.

are shown in Table 2. The CO photo-release from Hb is associated with a positive volume change of  $21.5 \pm 0.9$  mL mol $^{-1}$  and enthalpy change of  $19.4 \pm 1.2$  kcal mol $^{-1}$ . These results are in agreement with the thermodynamic parameters reported previously by Peters et al:  $\Delta V = 23.4 \pm 0.5$  mL mol $^{-1}$  and  $\Delta H = 18.0 \pm 2.9$  kcal mol $^{-1}$  (Peters et al., 1992). Since the laser power used in this study was kept below 50  $\mu$ J, the low level of photo-dissociation was achieved that corresponds to 1 CO molecule per hemoglobin photo-released. Thus the observed thermodynamic parameters reflect the transition between fully ligated (CO) $_4$ Hb

and triple ligated  $(\text{CO})_3\text{Hb}$ . Consequently, the observed reaction enthalpy corresponds to the enthalpy change due to the cleavage of the Fe-CO bond ( $\Delta H_{\text{Fe-CO}}=17.5 \text{ kcal mol}^{-1}$  (Leung et al., 1987; Miksovská et al., 2005)), the enthalpy change due to the solvation of a CO molecule ( $\Delta H_{\text{solv}} = 2.6 \text{ kcal mol}^{-1}$  (Leung et al., 1987)), the enthalpy change of structural relaxation associated with the ligand release from the protein matrix, and enthalpy of the distal pocket hydration. The occupancy of water molecules in the distal pocket of deoxyHb was determined to be significantly lower than that in Mb ( $\sim 0.64$  for the Hb  $\alpha$ - chain and  $\sim 0.33$  for the Hb  $\beta$ -chain (Esquerra et al., 2010)). Using an average occupancy of 0.48, we estimate that the distal pocket hydration contributes to the overall enthalpy change by  $\sim -3 \text{ kcal mol}^{-1}$  (Vetromile, et al., 2011). Therefore, the structural relaxation coupled to the CO dissociation and diffusion into the surrounding solvent is accompanied by a small enthalpy change of  $2 \text{ kcal mol}^{-1}$ .

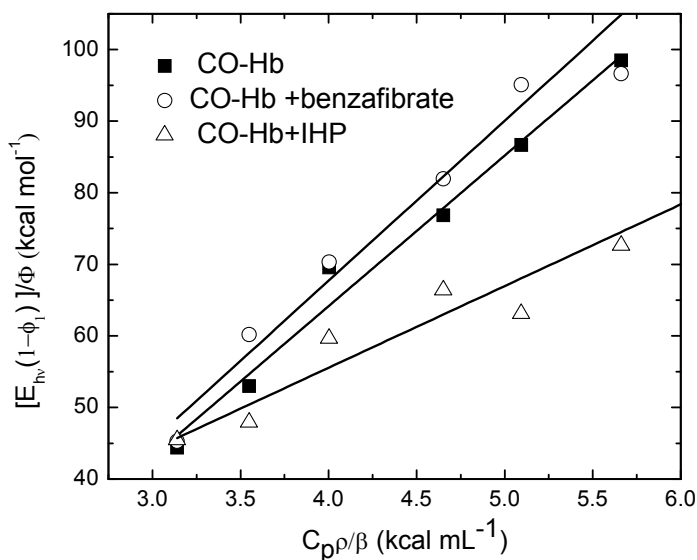


Fig. 6. The plot of the ratio of the acoustic amplitude for the CO photo-dissociation from the CO-Hb complex and the reference compound as a function of the temperature dependent factor ( $C_p\rho/\beta$ ) term. The reaction volume and enthalpy changes were extrapolated according to Eq. 5

Analogous to  $\text{O}_2$  photo-release from Mb, the observed reaction volume change for CO photorelease from Hb,  $\Delta V=21.5 \text{ mL mol}^{-1}$ , can be expressed as:  $\Delta V = V^\circ_{\text{CO}} + V^\circ_{(\text{CO})_3\text{Hb}} - V^\circ_{(\text{CO})_4\text{Hb}} - V^\circ_{\text{H}_2\text{O}}$ , where  $V^\circ_{\text{CO}}$  is the partial molar volume of CO and  $V^\circ_{(\text{CO})_3\text{Hb}}$  and  $V^\circ_{(\text{CO})_4\text{Hb}}$  are the partial molar volume of  $(\text{CO})_3\text{Hb}$  and  $(\text{CO})_4\text{Hb}$ , respectively. Using  $V^\circ_{\text{CO}} = 35 \text{ mL mol}^{-1}$  (Projahn et al., 1990) and  $V^\circ_{\text{H}_2\text{O}} = 9 \text{ mL mol}^{-1}$  (partial molar volume of water scaled by the average occupancy of the Hb chain), we estimate that upon release of one CO molecule per Hb, the protein undergoes a small contraction of  $-7 \text{ mL mol}^{-1}$ . The small volume change observed here is consistent with the minor structural changes due to deligation of Hb in the R-state as observed in the X-ray structure that are predominantly

localized in the the  $\alpha$ -chain and include reposition of the F-helix and shift of the EF and CD corner (Wilson et al., 1996).

	$\Delta H_{\text{prompt}}$ (kcal mol <sup>-1</sup> )	$\Delta V_{\text{prompt}}$ (mL mol <sup>-1</sup> )
CO-Hb	19.4 ± 1.2	21.5 ± 0.9
CO-Hb + BZF	21.7 ± 7.9	22.3 ± 1.7
CO-Hb + IHP	-9.9 ± 6.1	11.4 ± 1.3

Table 2. Volume and enthalpy changes associated with CO photo-dissociation from Hb.

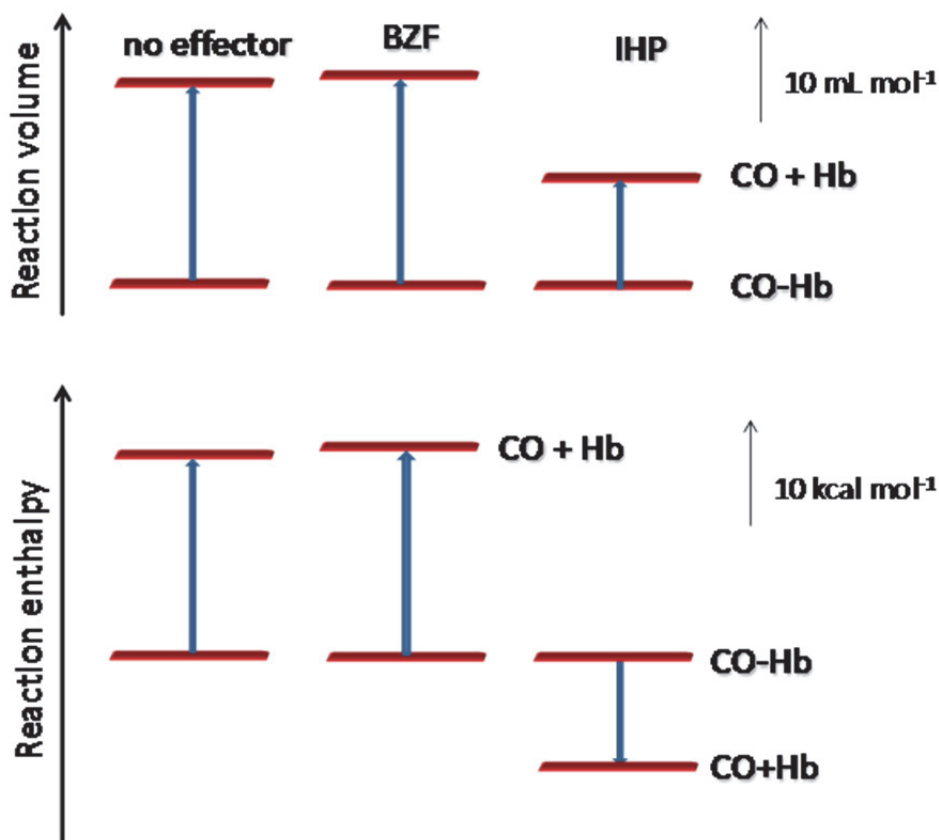


Fig. 7. The thermodynamic profile for CO photo-dissociation from Hb in the absence of effector and in the presence of BZF and IHP.

We have also determined volume and enthalpy changes associated with the CO photo-dissociation from Hb in the presence of heterogenous effectors BZT and IHP (Fig. 6) and the thermodynamic profiles for CO photo-dissociation from CO-Hb complex in the presence and absence of effectors are presented in Fig.7 . Both effectors bind to Hb in the T-state and R-state and modulate the interaction of Hb with diatomic ligands (Coletta et al., 1999b; Marden et al., 1990). For example, the binding of BZF or IHP to CO-Hb complex decreases the CO association rate approximately four or eight times, respectively (Marden et al., 1990), and decreases the affinity of R state deoxy-Hb for oxygen (Tsuneshige et al., 2002). Coletta et al (Coletta et al., 1999a) have reported that simultaneous binding of IHP and BZF effectors to Hb at ambient pressure leads to the Hb intermediate with tertiary T-like structure in the quaternary R- conformation. Recently, using NMR spectroscopy Song et al. have shown that binding of IHP to the fully ligated Hb increase the conformational fluctuation of the R-state in both the  $\alpha$ - and  $\beta$ -chain (Song et al., 2008).

The photoacoustic data presented here show that BZF binding to CO-Hb complex does not impact the reaction volume and enthalpy changes associated with CO photo-release. The crystal structure of horse CO-Hb in complex with BZF indicates that the structural changes due to BZF association to fully ligated Hb are localized in the  $\alpha$ -subunits (Shibayama et al., 2002). BZF binds to the surface of each  $\alpha$ -chain E-helix and decreases the distance between the heme iron and distal His and its binding site is surrounded by hydrophobic residues such as Ala 65, Leu 68, Leu 80 and Leu 83 (Shibayama et al., 2002). Such minor structural changes caused by BZF association are unlikely to alter the overall structural volume and enthalpy changes associated with the CO photo-release. However, due to the lower solubility of BZF, the effector concentration used in this study was 5 mM that results in a Hb fractional saturation of 0.25 (using  $K_D$  of 15 mM (Ascenzi et al., 1993)). Such lower fractional saturation may prevent detection of BZF induced changes in Hb conformational dynamics.

On the other hand, the binding of IHP has a significant impact on the observed volume and enthalpy changes (Table 2). The reaction volume decreases by 10 mL mol<sup>-1</sup> and the enthalpy change is more exothermic by nearly 30 kcal mol<sup>-1</sup> compared to the thermodynamic parameters determined in the absence of effector molecules. Such negative reaction volume and exothermic enthalpy change indicates that electrostriction of solvent molecules caused by reorganization of salt bridges or redistribution of charges on protein surface contributes to the overall reaction volume and enthalpy change associated with the CO photo-release. Indeed, IHP interacts with charged residues along the Hb central cavity. At the Hb T-state, the IHP binding site is located at the interface of the  $\beta$ -chains involving Val 1, His2, Lys 82 and His 141 from each chain (Riccio et al., 2001); whereas at the R-state Hb, the IHP molecule interacts with the charged residues Lys 99 and Arg 141 from each  $\alpha$ -chain (Laberge et al., 2005). In the absence of the X-ray structure of IHP bound fully ligated and partially photolyzed CO-Hb, it is difficult to point out the factors that contribute to the observed volume and enthalpy changes on the molecular level. Arg 141 forms a salt bridge with Asp 126 in the T-state deoxy Hb that is absent in the fully ligated R- state (Park et al., 2006). We speculate that the transition between the fully ligated (CO)<sub>4</sub>Hb and partially ligated (CO)<sub>3</sub>Hb may be associated with a repositioning of the Arg 141 side chain leading to a partial exposure of either the IHP molecule and/or the Arg 141 side chain to the surrounding solvent molecules. Also, the

ligand photo-release may be associated with the repositioning of the IHP molecule within the Hb central cavity. Based on a molecular dynamics simulation of IHP binding sites in south polar skua deoxyHb, an IHP migration pathway connecting the binding site at the interface between the  $\alpha$ -chains and the second binding site located between the  $\beta$ -chains was proposed suggesting that IHP interactions with Hb are dynamic and involve numerous positively charged residues situated along the central cavity (Riccio et al., 2001). Therefore, CO photo-release may trigger relocation of IHP within the central cavity resulting in larger exposure of IHP phosphate groups and/or charged amino acid residues and concomitant electrostriction of solvent molecules.

#### 4. Conclusion

The photoacoustic data for the ligand photo-dissociation from Mb shows that the structural volume changes associated with the O<sub>2</sub> diffusion from the Mb active site are similar to those determined previously for CO in agreement with the crystallographic data. On the other hand, the time constant for O<sub>2</sub> escape from the distal pocket to the surrounding solvent is two to three time faster than that for CO suggesting a distinct migration pathway for diatomic ligands in Mb. Our PAC study also indicates that IHP binding to Hb-CO complex alters the volume and enthalpy changes associated with the CO photo-dissociation from the heme iron indicating that the transition between the fully ligated (CO)<sub>4</sub>Hb and partially ligated (CO)<sub>3</sub>Hb complex is associated with the reorientation of IHP molecule within the central cavity and/ or charged amino acid residues interacting with IHP.

#### 5. Acknowledgement

This work was supported by J. & E. Biomedical Research Program (Florida Department of Health) and National Science Foundation (MCB 1021831).

#### 6. References

- Angeloni, L.&Feis, A. (2003). Protein relaxation in the photodissociation of myoglobin-CO complexes. *Photochem. Photobiol. Sci.*, 2, 7, pp. 730-740, 1474-905X
- Ascenzi, P., Bertollini, A., Santucci, R., Amiconi, G., Coletta, M., Desideri, A., Giardina, B., Polizio, F.&Scatena, R. (1993). Cooperative effect of inositol hexakisphosphate, bezafibrate, and clofibrate acid on the spectroscopic properties of the nitric oxide derivative of ferrous human hemoglobin. *J. Inorg. Biochem.*, 50, 4, pp. 263-272, 0162-0134
- Belogortseva, N., Rubio, M., Terrell, W.&Miksovska, J. (2007). The contribution of heme propionate groups to the conformational dynamics associated with CO photodissociation from horse heart myoglobin. *J. Inorg. Biochem.*, 101, 7, pp. 977-986, 0162-0134
- Bossa, C., Anselmi, M., Roccatano, D., Amadei, A., Vallone, B., Brunori, M.&Di Nola, A. (2004). Extended Molecular Dynamics Simulation of the Carbon Monoxide Migration in Sperm Whale Myoglobin. *Biophys. J.*, 86, 6, pp. 3855-3862, 0006-3495

- Carver, T. E., Rohlf, R. J., Olson, J. S., Gibson, Q. H., Blackmore, R. S., Springer, B. A. & Sligar, S. G. (1990). Analysis of the kinetic barriers for ligand binding to sperm whale myoglobin using site-directed mutagenesis and laser photolysis techniques. *J. Biol. Chem.*, 265, 32, pp. 20007-20020,
- Coletta, M., Angeletti, M., Ascenzi, P., Bertollini, A., Della Longa, S., De Sanctis, G., Priori, A. M., Santucci, R. & Amiconi, G. (1999a). Coupling of the Oxygen-linked Interaction Energy for Inositol Hexakisphosphate and Bezafibrate Binding to Human HbA0. *J. Biol. Chem.*, 274, 11, pp. 6865-6874,
- Coletta, M., Angeletti, M., Ascone, I., Boumis, G., Castellano, A. C., Dell Ariccia, M., Della Longa, S., De Sanctis, G., Priori, A. M., Santucci, R., Feis, A. & Amiconi, G. (1999b). Heterotropic Effectors Exert More Significant Strain on Monoligated than on Unligated Hemoglobin. *Biophys. J.*, 76, 3, pp. 1532-1536, 0006-3495
- Esquerra, R. M., Lopez-Pena, I., Tipgunlakant, P., Birukou, I., Nguyen, R. L., Soman, J., Olson, J. S., Kliger, D. S. & Goldbeck, R. A. (2010). Kinetic spectroscopy of heme hydration and ligand binding in myoglobin and isolated hemoglobin chains: an optical window into heme pocket water dynamics. *Phys. Chem. Chem. Phys.*, 12, 35, pp. 10270-10278, 1463-9076
- Gensch, T. & Viappiani, C. (2003). Time-resolved photothermal methods: accessing time-resolved thermodynamics of photoinduced processes in chemistry and biology. *Photoch. Photobiol. Sci.*, 2, 7, pp. 699-721, 1474-905X
- Gibson, Q. H., Olson, J. S., McKinnie, R. E. & Rohlf, R. J. (1986). A kinetic description of ligand binding to sperm whale myoglobin. *J. Biol. Chem.*, 261, 22, pp. 10228-10239,
- Goldbeck, R. A., Paquette, S. J., Björling, S. C. & Kliger, D. S. (1996). Allosteric Intermediates in Hemoglobin. 2. Kinetic Modeling of HbCO Photolysis. *Biochemistry*, 35, 26, pp. 8628-8639, 0006-2960
- Hasinoff, B. B. (1974). Kinetic activation volumes of the binding of oxygen and carbon monoxide to hemoglobin and myoglobin studied on a high-pressure laser flash photolysis apparatus. *Biochemistry*, 13, 15, pp. 3111-3117, 0006-2960
- Henry, E. R., Sommer, J. H., Hofrichter, J., Eaton, W. A. & Gellert, M. (1983). Geminate recombination of carbon monoxide to myoglobin. *J. Mol. Biol.*, 166, 3, pp. 443-451, 0022-2836
- Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M. & Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450, 7171, pp. 913-916, 1476-4687
- Hummer, G., Schotte, F. & Anfinrud, P. A. (2004). Unveiling functional protein motions with picosecond x-ray crystallography and molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.*, 101, 43, pp. 15330-15334,
- Laberge, M., Kövesi, I., Yonetani, T. & Fidy, J. (2005). R-state hemoglobin bound to heterotropic effectors: models of the DPG, IHP and RSR13 binding sites. *FEBS Lett.*, 579, 3, pp. 627-632, 0014-5793
- Leung, W. P., Cho, K. C., Chau, S. K. & Choy, C. L. (1987). Measurement of the protein-ligand bond energy of carboxymyoglobin by pulsed photoacoustic calorimetry. *Chem. Phys. Lett.*, 141, 3, pp. 220-224, 0009-2614

- Marden, M. C., Bohn, B., Kister, J.&Poyart, C. (1990). Effectors of hemoglobin. Separation of allosteric and affinity factors. *Biophys. J.*, 57, 3, pp. 397-403, 0006-3495
- Miksovská, J.&Larsen, R. W. (2003). Structure-function relationships in metalloproteins. *Methods Enzymol*, 360, pp. 302-329, 0076-6879
- Miksovská, J., Norstrom, J.&Larsen, R. W. (2005). Thermodynamic profiles for CO photodissociation from heme model compounds: effect of proximal ligands. *Inorg Chem*, 44, 4, pp. 1006-1014, 0020-1669
- Milani, M., Nardini, M., Pesce, A., Mastrangelo, E.&Bolognesi, M. (2008). Hemoprotein time-resolved X-ray crystallography. *IUBMB Life*, 60, 3, pp. 154-158, 1521-6551
- Miller, L. M., Patel, M.&Chance, M. R. (1996). Identification of Conformational Substates in Oxymyoglobin through the pH-Dependence of the Low-Temperature Photoproduct Yield. *J. Am. Chem. Soc.*, 118, 19, pp. 4511-4517, 0002-7863
- Mills, F. C., Ackers, G. K., Gaud, H. T.&Gill, S. J. (1979). Thermodynamic studies on ligand binding and subunit association of human hemoglobins. Enthalpies of binding O<sub>2</sub> and CO to subunit chains of hemoglobin A. *J. Biol. Chem.*, 254, 8, pp. 2875-2880,
- Mouawad, L., Maréchal, J.-D.&Perahia, D. (2005). Internal cavities and ligand passageways in human hemoglobin characterized by molecular dynamics simulations. *Biochim. Biophys. Acta*, 1724, 3, pp. 385-393, 0304-4165
- Mueser, T. C., Rogers, P. H.&Arnone, A. (2000). Interface sliding as illustrated by the multiple quaternary structures of liganded hemoglobin. *Biochemistry*, 39, 50, pp. 15353-15364, 0006-2960
- Olson, J. S., Soman, J.&Phillips, G. N. (2007). Ligand pathways in myoglobin: A review of trp cavity mutations. *IUBMB Life*, 59, 8-9, pp. 552-562, 1521-6551
- Park, S.-Y., Yokoyama, T., Shibayama, N., Shiro, Y.&Tame, J. R. H. (2006). 1.25 Å Resolution Crystal Structures of Human Haemoglobin in the Oxy, Deoxy and Carbonmonoxy Forms. *J. Mol. Biol.*, 360, 3, pp. 690-701, 0022-2836
- Peters, K. S., Watson, T.&Logan, T. (1992). Photoacoustic calorimetry study of human carboxyhemoglobin. *J. Am. Chem. Soc.*, 114, 11, pp. 4276-4278, 0002-7863
- Petrich, J. W., Poyart, C.&Martin, J. L. (1988). Photophysics and reactivity of heme proteins: a femtosecond absorption study of hemoglobin, myoglobin, and protoheme. *Biochemistry*, 27, 11, pp. 4049-4060, 0006-2960
- Phillips, S. E. V.&Schoenborn, B. P. (1981). Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin. *Nature*, 292, 5818, pp. 81-82,
- Projahn, H. D., Dreher, C.&Van Eldik, R. (1990). Effect of pressure on the formation and deoxygenation kinetics of oxymyoglobin. Mechanistic information from a volume profile analysis. *J. Am. Chem. Soc.*, 112, 1, pp. 17-22, 0002-7863
- Riccio, A., Tamburrini, M., Giardina, B.&di Prisco, G. (2001). Molecular Dynamics Analysis of a Second Phosphate Site in the Hemoglobins of the Seabird, South Polar Skua. Is There a Site-Site Migratory Mechanism along the Central Cavity? *Biophys. J.*, 81, 4, pp. 1938-1946, 0006-3495



- Saffran, W. A. & Gibson, Q. H. (1977). Photodissociation of ligands from heme and heme proteins. Effect of temperature and organic phosphate. *J. Biol. Chem.*, 252, 22, pp. 7955-7958,
- Savino, C., Miele, A. E., Draghi, F., Johnson, K. A., Sciara, G., Brunori, M. & Vallone, B. (2009). Pattern of cavities in globins: The case of human hemoglobin. *Biopolymers*, 91, 12, pp. 1097-1107, 1097-0282
- Shibayama, N., Miura, S., Tame, J. R. H., Yonetani, T. & Park, S.-Y. (2002). Crystal Structure of Horse Carbonmonoxyhemoglobin-Bezafibrate Complex at 1.55-Å Resolution. *J. Biol. Chem.*, 277, 41, pp. 38791-38796,
- Silva, M. M., Rogers, P. H. & Arnone, A. (1992). A third quaternary structure of human hemoglobin A at 1.7-Å resolution. *J. Biol. Chem.*, 267, 24, pp. 17248-17256,
- Song, X.-j., Simplaceanu, V., Ho, N. T. & Ho, C. (2008). Effector-Induced Structural Fluctuation Regulates the Ligand Affinity of an Allosteric Protein: Binding of Inositol Hexaphosphate Has Distinct Dynamic Consequences for the T and R States of Hemoglobin. *Biochemistry*, 47, 17, pp. 4907-4915, 0006-2960
- Šrajcar, V., Ren, Z., Teng, T.-Y., Schmidt, M., Ursby, T., Bourgeois, D., Pradervand, C., Schildkamp, W., Wulff, M. & Moffat, K. (2001). Protein Conformational Relaxation and Ligand Migration in Myoglobin: A Nanosecond to Millisecond Molecular Movie from Time-Resolved Laue X-ray Diffraction†. *Biochemistry*, 40, 46, pp. 13802-13815, 0006-2960
- Tsuneshige, A., Park, S. & Yonetani, T. (2002). Heterotropic effectors control the hemoglobin function by interacting with its T and R states--a new view on the principle of allostery. *Biophys. Chem.*, 98, 1-2, pp. 49-63, 0301-4622
- Unno, M., Ishimori, K. & Morishima, I. (1990). High-pressure laser photolysis study of hemoproteins. Effects of pressure on carbon monoxide binding dynamics for R- and T-state hemoglobins. *Biochemistry*, 29, 44, pp. 10199-10205, 0006-2960
- Vetromile, C. M., Miksovská, J. & Larsen, R. W. (2011). Time resolved thermodynamics associated with ligand photorelease in heme peroxidases and globins: Open access channels versus gated ligand release. *Biochim. Biophys. Acta*, pp. 1570-9639
- Walda, K. N., Liu, X. Y., Sharma, V. S. & Magde, D. (1994). Geminate recombination of diatomic ligands CO, O<sub>2</sub>, NO with myoglobin *Biochemistry*, 33, pp. 2198-2209,
- Westrick, J. A. & Peters, K. S. (1990). A photoacoustic calorimetric study of horse myoglobin. *Biophys. Chem.*, 37, 1-3, pp. 73-79, 0301-4622
- Westrick, J. A., Peters, K. S., Ropp, J. D. & Sligar, S. G. (1990). Role of the arginine-45 salt bridge in ligand dissociation from sperm whale carboxymyoglobin as probed by photoacoustic calorimetry. *Biochemistry*, 29, 28, pp. 6741-6746, 0006-2960
- Wilson, J., Phillips, K. & Luisi, B. (1996). The Crystal Structure of Horse Deoxyhaemoglobin Trapped in the High-affinity (R) State. *J. Mol. Biol.*, 264, 4, pp. 743-756, 0022-2836
- Yang, F. & Phillips Jr, G. N. (1996). Crystal Structures of CO-, Deoxy- and Met-myoglobins at Various pH Values. *J. Mol. Biol.*, 256, 4, pp. 762-774, 0022-2836
- Ye, X., Demidov, A. & Champion, P. M. (2002). Measurements of the Photodissociation Quantum Yields of MbNO and MbO<sub>2</sub> and the Vibrational Relaxation of the Six-Coordinate Heme Species. *J. Am. Chem. Soc.*, 124, 20, pp. 5914-5924, 0002-7863

Yonetani, T., Park, S. I., Tsuneshige, A., Imai, K. & Kanaori, K. (2002). Global allostery model of hemoglobin. Modulation of O<sub>2</sub> affinity, cooperativity, and Bohr effect by heterotropic allosteric effectors. *J. Biol. Chem.*, 277, 37, pp. 34508-34520, 0021-9258

# Some Applications of Thermodynamics for Ecological Systems

Eugene A. Silow<sup>1</sup>, Andrey V. Mokry<sup>1</sup> and Sven E. Jørgensen<sup>2</sup>

<sup>1</sup>*Institute of Biology, UNESCO Chair of Water Resources, Irkutsk State University,*

<sup>2</sup>*Chair of Environmental Chemistry, Copenhagen University,*

<sup>1</sup>*Russia*

<sup>2</sup>*Denmark*

## 1. Introduction

Now ecologists feel necessary to construct the theoretical building of system ecology, to break strong reductionistic tradition of ecology and to include the use of thermodynamics in a new holistic approach to study ecosystems, their structure, functioning and natural history. We tried to present here the current state of thermodynamic view on ecosystems.

The first law of thermodynamics proclaims constancy of the total energy of isolated system for all changes, taking place in this system: energy cannot be created or destroyed. According to the second law of thermodynamics in isolated system entropy is always increasing or remaining constant. All processes in the Universe are oriented to the equilibrium state. Nevertheless, biological systems, and, consequently, ecological systems create order from disorder, they create and support chemical and physical non-equilibrium state – the basis they live on.

In this chapter the general overview of ecosystem as thermodynamic system is given and the concept of Eco-Exergy is introduced. The use of this concept in ecology is demonstrated to be very fruitful. To make it easy for other researchers to use the Eco-Exergy the procedure of exergy evaluation for ecosystems is followed with special attention to dimensions used. The main applications of exergy in modern ecology are reviewed with special focus on practical use of Eco-Exergy, exergy index, and structural exergy for real ecosystem assessment and estimation of their health and disturbance.

Another application of irreversible thermodynamics (Prigogine's inventions) is discussed. The theory of hypercycles, developed for cycles of autocatalytic reactions and widely accepted in biochemistry and molecular biology can also be applied for ecological systems. The model of conjugated hypercycles, applied to ecological systems explains many aspects of their non-linear dynamics and can be used for analysis of oscillating processes in ecological systems.

## 2. Ecosystem as thermodynamic system

Ecosystem is an open system. It supports structure and functioning due to external energy input. Usually ecosystem consume solar energy in the form of relatively short-wave radiation (visible light), though we know some ecosystems (e.g., at great depth in ocean)

which use chemical energy. Nevertheless general rule is the reception of solar energy by green plants via photosynthesis. They assimilate approximately 0,01 - 3 % of energy of falling radiation and, using this energy, create organic matter (*primary production*) from inorganic compounds (water, carbon dioxide, nitrates, phosphates and a lot of minor substances). The by-product of photosynthesis is oxygen. Organisms, creating primary production are called *producers*. Energy, stored in organic matter is used by producers themselves and is dissipated during the processes of plants respiration, growth and reproduction in the form of heat. The remaining energy, accumulated in plant biomass is used by animals to support their structure and functioning. These processes are balanced at global level, as well as in healthy mature ecosystems. The rate of total organic matter production is called *gross primary production*. The difference between *gross primary production* and the rate of decomposition of this substance by plants themselves is called *net primary production*. Organisms, consuming plants (*consumers*) can utilize not more than 10 % of net primary production consumed, the rest being dissipated in the form of heat. Predators (*secondary consumers*) can use not more than 10 % of primary consumers production. The dead bodies of plants and animals, organic wastes, produced by the last etc. are decomposed and reduced to primary inorganic compounds, available for the new cycle of production/destruction by *decomposers* or *reducers* (bacteria and other microorganisms). The rate of total increase of ecosystem biomass (yield) is known as *productivity*. In healthy, mature, balanced ecosystem it is equal to zero (or relation of production to respiration, P/R, is equal to one). So, ecosystem consumes high quality energy of solar radiation, uses part of it to support itself and dissipates the rest in the form of heat, increasing the total entropy of whole system Sun - ecosystem - environment. We have seen the ecosystem is functioning according to both the first and the second laws of thermodynamics.

### 2.1 Ten basic laws of ecology

Previously (Jørgensen & Fath, 2004) eight basic laws of ecology were proposed, but now it was decided to split one of them into three laws (Jørgensen, 2006-2007, 2011). These ten laws or Ecological Laws (EL) are listed below.

1. Mass and energy conservations are valid for ecosystems.
2. All ecosystem processes are irreversible and are accompanied by entropy production and exergy destruction.
3. All ecosystems are open systems embedded in an environment from which they receive energy – matter input and discharge energy – matter output.
4. Ecosystems have many levels of organization and operate hierarchically.
5. The components in an ecosystem form a complex interactive, self-organizing ecological network.
6. The carbon based life on Earth, has a characteristic basic biochemistry which all organisms share.
7. Thermodynamically, carbon-based life has a viability domain determined between about 250-350 K.
8. After the initial capture of energy across a boundary, ecosystem growth and development is possible by 1) an increase of the physical structure (biomass), 2) an increase of the network (more cycling) or 3) an increase of information embodied in the system.
9. Biological processes use captured energy (input) to move further from thermodynamic equilibrium and maintain a state of low-entropy and high exergy relative to its

surrounding and to thermodynamic equilibrium (**The First Ecological Law of Thermodynamics**).

10. If the ecosystem is offered more pathways or combinations of pathways to move away from thermodynamic equilibrium, then the combinations of pathways that move the system most away from thermodynamic equilibrium (=yield the highest Eco-Exergy of the ecosystem) will win (**The Second Ecological Law of Thermodynamics**).

### 3. Eco-exergy

The exergy of a system is a measure of its deviation from thermodynamic equilibrium with the environment, and represents the maximum capacity of energy to perform useful work as the system proceeds to equilibrium, with irreversibility increasing its entropy at the expense of exergy (Ludovisi, 2009). Taken by itself, the total exergy of an ecosystem is a measure of the change in entropy content from the equilibrium and the actual state (Svirezhev, 2000).

We may distinguish between technological exergy and Eco-Exergy: technological exergy uses the environment as reference state and is useful to find the first class energy (work) that a power plant can produce, Eco-Exergy uses as reference state the same ecosystem with the same temperature and pressure but at thermodynamic - chemical equilibrium (Fig. 1). Below we use the terms exergy and Eco-Exergy as synonymic.

The development and maintenance of the far-from-equilibrium condition of ecosystems is due to the steady storage of free energy into complex organic structures, biosynthesized from simple inorganic compounds. Accordingly, the total exergy of an ecosystem actually reflects the accumulation of biomass into the system, irrespective of the distribution of biogenic matter among ecosystem components. Exergy is a measure of the free energy of a system with contributions from all components including the energy of organisms. The measure for exergy in ecology also includes a factor to weigh the "complexity" of the ecological species. Moving from macroscopic to microscopic information storage, the exergetic contribution due to information grows and becomes even three orders of magnitude higher than the physical one in the more complex living systems. The capacity of packaging information at the molecular level (DNA) that differs from one organism to another can be taken into account using Eco-Exergy function.

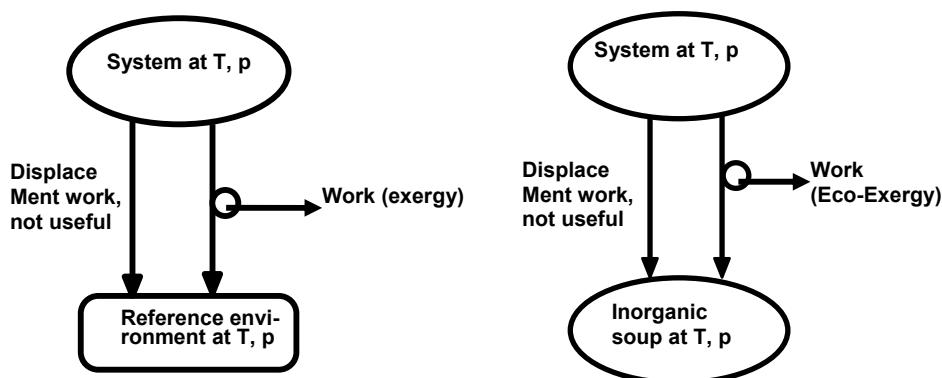


Fig. 1. Exergy is calculated for the system relatively to reference environment, Eco-Exergy relatively to the same system at the same temperature and pressure, but as inorganic solution without life and even organic molecules.

Here, we accept the following definition of exergy (according to Jørgensen, 1992; Svirezhev, 2000; Jørgensen & Svirezhev, 2004): *Exergy is the distance between the present state of the system and the state of it in thermodynamic equilibrium with the environment, measured in the units of energy.* It demonstrates the amount of work performed to create a given system from its primary components (in the case of ecological systems – from primary chemical compounds). *Exergy related to the total biomass (structural, specific or normalized exergy) measures the possibility of the ecosystem to accept and utilize external fluxes of energy.* It reflects the degree of ecosystem development or complexity, and has advantages in comparison with the total exergy such as independence from the total biomass of the ecosystem and possibility to serve as an indicator, demonstrating the level of evolutionary development of organisms in the ecosystem.

### 3.1 Eco-exergy calculation

According to **Prigogine's theorem**, an entropy production in every linear system without external influences is decreasing until it reaches minimum at steady state of dynamic equilibrium. Every living system is thermodynamic open system (EL 3) continuously converting potential chemical energy of organic matter into useful energy of creative processes (EL 9) and, in the end of ends, dispose to environment in the form of heat (EL 2). As a result of it, there is no thermodynamic equilibrium in living system. At temperatures, normal for life (see EL 7), living structures are labile and are destructed constantly. To compensate this destruction the permanent internal work in the form of synthesis is fulfilled. These working synthetic processes are processes producing negative entropy (negentropy), they create order with the use of chemical energy of low-entropy energy-rich compounds (consumers and reducers) or low-entropy energy-rich solar radiation (producers). Termination of these processes causes the loss of order, death. Dead body is in thermodynamic equilibrium with maximal entropy.

Exergy is the useful part of energy involved in some process, i.e. the maximal work fulfilled by the system during transit to thermodynamic equilibrium with environment state. This equilibrium means all components to be: 1) inorganic, 2) oxidized to maximum degree, 3) distributed homogenously (there is no gradients in the system). So, if we shall transfer the system into thermodynamic equilibrium with its environment, temperature and pressure will be equal for system and environment, so the only component exergy consists of will be chemical energy. Differences in temperature and pressure between system and environment are small, so we can ignore them in our calculations. The maximal input to exergetic constituent of ecological system will be done due to chemical energy, stored in organic matter and biological structures (Jørgensen et al., 2000; Jørgensen, Fath, 2011; Jørgensen, 2011). Taking into account this concept Eco-Exergy index can be calculated basing on chemical energy:  $\sum_i(\mu_c - \mu_{c,o})N_i$ , where  $i$  - the number of exergy containing components  $c$ ;  $\mu_c$  - chemical potential of component  $c$ ;  $\mu_{c,o}$  - chemical potential of component  $c$  in inorganic state. Eco-Exergy index for the system is calculated by reference of this system to the same system in the form of inorganic soup (i.e. - without life, structure, information, organic matter). The equation for exergy calculation was proposed by S.E. Jørgensen (Mejer & Jørgensen, 1977):

$$Ex = R \cdot T \cdot \sum_{i=0}^N [c_i \cdot \ln(c_i / c_{i,eq}) - (c_i - c_{i,eq})], [J] \quad (1)$$

where  $Ex$  - exergy, J;  $R$  - gas constant, J Mol<sup>-1</sup> K<sup>-1</sup>;  $T$  - temperature, K;  $c_i$  - concentration of component  $i$ , Mol;  $c_{i,eq}$  - concentration of the same component in the state of thermodynamic equilibrium with environment. Mol;  $N$  - number of components. The problem is to find value of  $c_{i,eq}$ . From one hand, exergy of compounds can be calculated on the basis of elementary composition. The disadvantage of this approach consists in that, firstly, input of inorganic components into total exergy of ecosystem is too low, secondly, higher and lower organisms have approximately the same stoichiometry (EL 6), so their exergy will be equal, that is in contradiction with information constituent of exergy. From another hand, the value  $c_{i,eq}$  can be assessed from the probability  $P_{i,eq}$  of discovering of component  $i$  in thermodynamic equilibrium state:

$$P_{i,eq} = \frac{c_{i,eq}}{\sum_{i=0}^N c_{i,eq}} \quad [1] \quad (2)$$

If we could find this probability, we can find the ratio of  $c_{i,eq}$  to current concentration.. As inorganic component  $c_0$  prevails in thermodynamic equilibrium state, we can rewrite (2) as:

$$P_{i,eq} \approx c_{i,eq} / c_{0,eq} \quad [1] \quad (3)$$

or

$$c_{i,eq} \approx P_{i,eq} \cdot c_{0,eq} \quad [\text{Mol}] \quad (4)$$

Chemical potential of dead organic matter ( $i=1$ ) can be found from classic thermodynamics as:

$$\mu_1 = \mu_{1,eq} + R \cdot T \cdot \ln(c_1 / c_{1,eq}), \quad [\text{J Mol}^{-1}] \quad (5)$$

where  $\mu$  - chemical potential. The difference  $\mu_i - \mu_{i,eq}$  is known for detritus. From (5) we see:

$$c_{1,eq} = c_1 \cdot \exp[-(\mu_1 - \mu_{1,eq}) / (R \cdot T)] . \quad [\text{Mol}] \quad (6)$$

From (3) at  $i=1$ :

$$P_{1,eq} \approx [c_1 / c_{0,eq}] \cdot \exp[-(\mu_1 - \mu_{1,eq}) / (R \cdot T)] . \quad [1] \quad (7)$$

For biological components ( $i=2, 3, 4, \dots, N$ ) probability  $P_{i,eq}$  is composed from probability of detritus production  $P_{1,eq}$  and probability  $P_{i,a}$  of genetic information collection to determine protein structure. Supposing this events independent:

$$P_{i,eq} = P_{1,eq} \cdot P_{i,a} \quad (i \geq 2). \quad [1] \quad (8)$$

Equation (1) we can rewrite taking into account (4) as:

$$Ex \approx R \cdot T \cdot \sum_{i=0}^N [c_i \cdot \ln(c_i / (P_{i,eq} \cdot c_{0,eq})) - (c_i - P_{i,eq} \cdot c_{0,eq})] . \quad [\text{J}] \quad (9)$$

Then:

$$Ex \approx RT \sum_{i=0}^N [c_i (\ln(1/P_{i,eq}) - \ln(c_{0,eq}/c_i)) - (c_i - P_{i,eq}c_{0,eq})] \quad \text{[J]} \quad (10)$$

From (3), as  $c_i \gg c_{i,eq}$ , then  $1/P_{i,eq} \approx c_{0,eq}/c_{i,eq} \gg c_{0,eq}/c_i$ . Consequently

$$\left| \ln(1/P_{i,eq}) \right| \gg \left| \ln(c_{0,eq}/c_i) \right|, \quad \text{[1]} \quad (11)$$

after that, we can ignore the second logarithm in the sum:

$$Ex \approx R \cdot T \cdot \sum_{i=0}^N [c_i \cdot (\ln(1/P_{i,eq}) - 1) + P_{i,eq} \cdot c_{0,eq}]. \quad \text{[J]} \quad (12)$$

Also  $1 \ll \ln(1/P_{i,eq})$  and  $P_{i,eq} \cdot c_{0,eq} \approx 0$ , then:

$$Ex \approx -R \cdot T \cdot \sum_{i=1}^N c_i \cdot \ln P_{i,eq}. \quad \text{[J]} \quad (13)$$

where summation starts from 1, as  $P_{0,eq} \approx 1$ . Taking to (13)  $P_{i,eq}$  from (8) and  $P_{1,eq}$  from (7), we have the following expression to calculate exergy:

$$\frac{Ex}{R \cdot T} \approx - \sum_{i=1}^N c_i \ln \left( \frac{c_1}{c_{0,eq}} \right) + \left( \frac{\mu_1 - \mu_{1,eq}}{R \cdot T} \right) \sum_{i=1}^N c_i - \sum_{i=2}^N (c_i \ln(P_{i,a})) \quad \text{[Mol]} \quad (14)$$

In (14) the first sum is insufficient in comparison with the rest two, so:

$$\frac{Ex}{R \cdot T} \approx \left( \frac{\mu_1 - \mu_{1,eq}}{R \cdot T} \right) \cdot \sum_{i=1}^N c_i - \sum_{i=2}^N (c_i \cdot \ln(P_{i,a})). \quad \text{[Mol]} \quad (15)$$

Exergy in (15) sufficiently depends on organism complexity, as it is connected with information stored in genetic code. This equation can be used to calculate exergy of ecosystem components. If we take detritus ( $i=1$ ), we know that free energy released from it equals approximately  $18,7 \text{ kJ} \cdot \text{g}^{-1}$ . Taking  $T=300 \text{ K}$ ,  $R=8,31 \text{ J} \cdot \text{Mol}^{-1} \cdot \text{K}^{-1}$ , and average Mol mass of detritus about  $105 \text{ g} \cdot \text{Mol}^{-1}$ , we obtain the following for detritus exergy in  $\text{m}^3$  of water:

$$Ex_1 = 18,7 \cdot c_1 \quad \text{[kJ m}^{-3}\text{]}, \quad \text{or} \quad \frac{Ex_1}{R \cdot T} = 7,34 \cdot 10^5 \cdot c_1 \quad \text{[g m}^{-3}\text{]} \quad (16)$$

So, we can rewrite (15) as

$$\frac{Ex}{R \cdot T} \approx 7,34 \cdot 10^5 \cdot c_i - \sum_{i=2}^N (c_i \cdot \ln(P_{i,a})). \quad \text{[g m}^{-3}\text{]} \quad (17)$$

Now we are to find  $P_{i,a}$  – the probability of creation of specific genetic information, characteristic for the organism given. Originally (Jørgensen, 1992, 2002; Jørgensen et al.,



2000)  $P_{i,a}$  was determined basing on number of informative (structural) genes (each gene codes the sequence of 700 aminoacids in average) for various taxonomic groups:

$$P_{i,a} = 20^{-700g} \quad (i \geq 2), \quad [1] \quad (18)$$

where  $g$  - число генов. Then exergy of typical single cell alga (850 genes approximately) can be calculated as:

$$\frac{Ex_{alga}}{R \cdot T} \approx 7,34 \cdot 10^5 c_i - c_i \ln 20^{-700 \times 850} \approx 25,2 \cdot 10^5 c_i \quad [\mathbf{g} \cdot \mathbf{m}^{-3}] \quad (19)$$

If we relate values of different components of ecosystem exergy to one of detritus ( $7,34 \cdot 10^5$ ), we shall get relative recalculation coefficient  $\beta_i$ . Corresponding coefficients were calculated for many systematic groups and published (Jørgensen, 1992; Bendoricchio & Jørgensen, 1997; Jørgensen et al., 2000). These coefficients reflect relative complexity of organisms (simpler organisms have lower  $\beta$  values). Later, with the use of new genetic data and some indirect methods of  $\beta$  values assessment, ratio of non-informative genes to total genes number and others, new list of  $\beta$  was composed and published (Jørgensen et al., 2005; Jørgensen, 2007). New  $\beta$  values are added every year (table 1).

Therefore, total exergy of ecosystem, based on chemical energy of organic matter (biomass) and information, stored in living organisms (recalculating coefficient  $\beta$ ), can be calculated as:

$$Ex / RT = \sum_{i=1}^N c_i \cdot \beta_i \quad [\mathbf{g} \text{ detritus equivalent } \mathbf{m}^{-3}] \quad (20)$$

This exergy now is often called Eco-Exergy (sometimes - exergy index) to distinguish it from physical or technological exergy (Marques et al, 2003; Jørgensen, 2006, 2007).

Another indicator of ecosystem state, based on Eco-Exergy, was proposed – structural or specific exergy (structural or specific Eco-Exergy). Structural exergy ( $Ex_{str}$ ) is the exergy related to total biomass (Silow, 1998, 1999, 2006; Xu et al., 1999, 2001, 2004, 2005; Marques et al., 2003; Jørgensen, 2006a). Unlike total exergy it does not depends on biomass and it reflects the ability of ecosystem to accept and utilize the flow of energy from external sources, serving simultaneously as indicator of ecosystem development, its complexity and level of evolutionaty development of biological species composed in it.

$$Ex_{str} = \left( \sum_{i=1}^N c_i \cdot \beta_i \right) \cdot \left( \sum_{i=1}^N c_i \right)^{-1}. \quad [1] \quad (21)$$

We can measure the following aspects of an ecosystem state with the Eco-Exergy: 1) the distance from thermodynamic equilibrium, i.e. general measure of total complexity of ecosystem; 2) structure (biomass and network size) and functions (available information) of ecosystem; 3) ability of ecosystem to survive (expressed via biomass and information of system).

Structural exergy reflects: 1) efficiency of energy use by organisms; 2) relative information content of ecosystem and, 3) consequently, the ability of ecosystem to regulate interactions between organisms or groups of organisms.

Group	Exergy conversion factor, $\beta$	Group	Exergy conversion factor, $\beta$
Minimal cell	5,8	Brachiopoda	109
Bacteria	8,5–12	Seedless vascular plants	158
Archaea	13,8	Rotifera	163
Yeasts	18	Insecta	167–446
Alga	15–298	Chironomida	300
Cyanobacteria	15	Moss	174
Dynophyta	18	Crustaceans	230–300
Green microalgae	20	Cladocera	232
Diatoms	66	Copepoda	240
Macrophyta (alga)	67–298	Amphipoda	290
Rhodophyta	92	Mollusca	297–450
Protozoa	31–97	Bivalves	297
Amoeba	38	Gastropoda	312–450
Gastrotricha	97	Gymnosperm	314
Fungi	61	Macrophytes (Phanerogam)	356–520
Nemertina	76	Flowering plants	393–543
Worms	91–133	Fish	499–800
Cnidaria	91	Amphibia	688
Plathelminthes	120	Reptilia	833
Oligochaeta	130	Aves	980
Nematoda	133	Mammalia	2127
Sponges	98	Homo sapiens	2173

Table 1. Exergy/Biomass Conversion factors for different groups of organisms, after Silow & Mokry, 2010

### 3.2 Eco-exergy and structural exergy applications in ecology and environmental science

#### 3.2.1 Eco-exergy in theoretical ecology and in aquatic ecology

We have seen above exergy approach was demonstrated to be very fruitful during the analysis of the application of thermodynamic principles and laws to the main fundamental concepts of ecology at the end of the XX century. The analysis of three thermodynamic laws expressions in ecological rules together with exergy analysis led to formulation of the 10 Ecological Laws, in particular the Fourth (Ecological) Law of Thermodynamics, EL9 (Patten et al., 1997; Jørgensen et al., 1999; Straškraba et al., 1999; Jørgensen, 2006b).

Non-equilibrium thermodynamics models based on the concept of exergy provided a common basis for representing many aspects of ecosystem development and response to environmental impacts as a single measure (Pykh et al., 2000). The use of exergy made possible the investigation of the flows of an ecosystem in terms of exergy and to arrange the system as a hierarchically ordered sequence of systems, thermodynamically embedded in each other (Nielsen, 2000). Experiments with mathematical models supported the

hypothesis that an ecosystem can coordinate the most complex behaviour in the case of high level of exergy of the systems at the edge of oscillation before entering into the chaotic situation (Mandal et al., 2007). The thermodynamic notion of exergy was shown to give better insight both to the patterns of nonlinear ecosystem behaviour and to comparison of the patterns in ecological modelling (Svirezhev & Steinbom, 2001).

### 3.2.2 Eco-exergy and aquatic ecology

There are very few researches devoted to analysis of plankton communities with the aid of exergy. The implications of body sizes of phytoplankton and zooplankton for total system dynamics by optimizing exergy as a goal function for system performance indicator with mathematical models have been analyzed (Ray et al., 2001). A structurally dynamic model based on phosphorus nutrient limitation has been developed for Lake Mogan located nearby Ankara, Turkey. Exergy was applied as a goal function to consider the dynamic adaptation and the seasonality of plankton species (e.g., size shifts) (Zhang et al., 2003a, b).

The ecosystem of the North Sea integrity was approved to be reflected in exergy capture, storage capacity, cycling, matter losses, and heterogeneity (the diatom/non-diatom ratio of planktonic algae was used) with ecosystem model. Its feasibility was assessed as an ecosystem model of the North Sea, for the Elbe plume, after prior satisfactory calibration. The modeling effort suggested that drastic nutrient load reduction from the Elbe alone would have a limited effect on the larger German Bight: even a 60% reduction scenario would only lead to moderate changes in all five indicators (Windhorst et al., 2005).

More representative and multiple are applications of exergy to benthos communities. Exergy was used in optimization models of phytobenthos (Nielsen, 1997). Exergy concept allowed the finding of the best adapted water plants species in a given environmental condition and to explain in a satisfactory way the observed distributions of them in the Lagoon of Venice, Italy (Coffaro et al., 1997).

Exergy storage was estimated for benthic communities of sandy and muddy bottoms of the North Adriatic Sea subjected to experimental disturbance, induced by means of a controlled trawl fishing haul. The results showed a decrease of local exergy content in the disturbed area, with the minimum, both in sandy and muddy bottom, one month after the experimental disturbance. The exergy of the benthic community increased to the reference level, i.e., the surrounding control area, in accordance with the proposed hypothesis on the dynamics of exergy storage during a systems' development (Libralato et al., 2006).

The changes of exergy and specific exergy were studied with data of benthic macrofauna in the Mondego estuary (Western Portugal). Estimates for the exergy indices provided useful indications for the evaluation of environmental impact due to the eutrophication process (Fonseka et al., 2002).

Export of exergy was estimated for benthic communities on the South-Western Atlantic Coast of France. This export was mainly composed of the migration of grazing fish during the warm season, and of cultivated bivalves during the cold season (Leguerrier et al., 2007).

In the following study a self-organizing map for patterning exergy of benthic macroinvertebrate communities of 650 sampling sites in the Netherlands, including 855 species was implemented. Using these datasets, authors have calculated exergy of five trophic functional groups for each sampling site on the basis of the biomass data. Exergy of

different trophic groups responded differently to different water types reflecting characteristics of target ecosystems (Park et al., 2006).

Eco-Exergy and Specific Eco-Exergy were used to characterize the state of the community during the recovery process after damage to the benthic communities caused by ecological engineering Yangtze River, China (Zhang et al., 2009). Changes of the macro-benthic community structure (Venice lagoon, Italy) over almost 70 years were pictured, showing a sharp decrease in its diversity and system efficiency, estimated with the use of exergy (Pranovi et al., 2008).

### **3.2.3 Eco-exergy as an ecosystem health indicator**

The idea to use exergy as an indicator of ecosystem health was proposed by S.E. Jørgensen (1992, 1999, 2002, 2006a,b), but the first applications of exergy as an ecosystem health indicator were fulfilled with mathematical models. Mathematical modelling with the use of exergy was shown to be applicable to explain ecosystem reactions (Jørgensen & Padišak, 1996), and to facilitate the estimation of parameters of models. The first pioneer papers describing the application of exergy indicators for natural aquatic ecosystems were published in 1997 (Xu et al., 1997; Marques et al., 1997). In 1998 the first application of exergy analysis to the results of field and laboratory experiments with real aquatic ecosystem was published (Silow, 1998). This work was continued by few more publications (Xu et al., 1999a, 1999b; Silow & Oh, 2004). The possibility to use such parameters as structural exergy and exergy for estimation of ecosystem state and its changes under various external influences was demonstrated for real natural and experimental ecosystems. These parameters were shown to reflect the state of ecosystem and they can indicate the degree of ecosystem adaptation, decreasing when important for ecosystem functioning components were depressed or eliminated. S.E. Jørgensen (2006a) proposed to use Eco-Exergy, specific Eco-Exergy and ecological buffer capacities as Ecological indicators for ecosystem development and health assessment.

Exergy is now often used for eutrophication assessment (Xu et al., 1999, 2001, 2011a; Fonseca et al., 2002; Marques et al., 2003; Ye et al., 2007), for ecological engineering purposes (Nunneri et al., 2008), for ecosystem health assessment (Vassallo et al., 2006; Libralato et al., 2006; Xu et al., 2011b). Exergy and specific exergy indices as Ecological indicators of the trophic state of lake ecosystems were tested on a set of lakes (Ludovisi & Poletti, 2003). The ecosystem maturity was estimated for Lake Trasimeno (Ludovisi et al., 2005).

### **3.3 Case study; eco-exergy analysis of lake Baikal state**

The first works were fulfilled with the use of mathematic models. Different sensitivity of under-ice and open water plankton communities to contaminant additions was demonstrated. This can be related to different structural exergy content in plankton community. Exergy buffer capacity seems to be a more realistic measure for pliability of ecosystem reaction to external factors than biomass buffer capacity (Silow, 1998, 1999). In field researches the structural exergy of benthic communities at control (pristine) site, and in the region of Baikalsk Pulp and Paper Combine wastewaters discharge region at the same depths and kind of sediments was shown to differ strongly (structural exergy in polluted area was much lower than in pristine one), while biomass and total exergy behaved in not such an expressive way (Silow & Oh, 2004; Silow, 2006). The next step was the analysis of

exergy and structural exergy of plankton community response to different chemical stressors analyzed in mesocosms experiments. Results obtained with mesocosms and microcosms demonstrate structural exergy decrease in experiments proportionally to a value of the added toxicant concentration, while other parameters (biomasses of components, total biomass of community, total exergy) fluctuated (Silow & Oh, 2004; Silow, 2006). Here we present the results of exergy calculations for natural plankton community of the lake Baikal.

Yearly average values of structural exergy during 1951–1999 fluctuated around their long-term average within the limits “long-term average  $\pm$  mean square deviation” ( $154,9 \pm 26,0$ ) without any trends. More interesting is the picture for total eco-exergy for the same period.

It demonstrates well expressed linear trend of increase with  $r^2 = 0.31$  (Fig. 2).

We have also analysed the long-term dynamics of exergetic parameters for four limnological seasons at Baikal: inverted stratification (limnological Winter, under-ice season, February – April), spring overturn (limnological Spring, ice melting, May – June), direct stratification (limnological Summer, July – October), fall overturn (limnological Autumn, November – January)<sup>1</sup>. Analysis of eco-exergy and structural exergy behaviour during different seasons cleared that the positive trend of eco-exergy is observed during limnological Summer (Fig. 3, 4).

Dynamics of pelagic plankton biomass in Baikal for 1951–1999 is given in Fig. 2. There is neither expressed directional change of total biomass, nor changes of biomasses of different components (only slight positive trend of zooplankton biomass). Long-term oscillations of individual components are easily observed. Taking into account all discussed above and remembering the relative constancy of the total biomass of pelagic plankton, we can try to explain the tendency of its exergy to increase according to three listed above strategies (EL8 – increase of biomass, increase of network, increase of information). According to the first strategy it is the primary production increase, based on the mass development of small sized alga in summer period. Actually it is observed in the lake (Izmesyeva & Silow, 2010). According to the second strategy it might be some recently observed structural changes in the plankton community (Hampton et al., 2008; Moore et al., 2009; Silow, 2010), and the third strategy is realized through the growth of share of larger zooplankton, like Cladocerans (Pislegina & Silow, 2010). Total biomass of plankton community and its individual components remains constant, while the total exergy of the community tends to increase. This increase can be explained with the principles of S.E. Jørgensen (section 2.1) – the principles of exergy maximization (EL9 and EL10) via the growth of solar exergy consuming capacity, sophistication of ecosystem networking and increase of ecosystem information storage (EL8).

The calculated values of structural exergy for different seasons in the lake Baikal plankton for the second half of XX century, on the basis of long-term monitoring data, fluctuate within their natural limits (long-term average  $\pm$  mean square deviation) and do not demonstrate any positive or negative trends (Fig. 2, Fig. 4). It points to the lack of expressed unfavourable changes in the lake Baikal pelagic.

---

<sup>1</sup> Lake Baikal is dimictic lake, characterized by two periods of stratification – inverted, when upper layer (0–50 m) of water has the temperature 0–1 °C, layer 50–250 m – 1–4 °C, direct, when temperature of upper layer decreases from 12 °C at surface to 5–6 °C at 50 m, layer 50–250 m – 4–5 °C, and two overturns with temperature at 0–250 m is about 4 °C. Below 250 m temperature is constant about 3,3 °C.

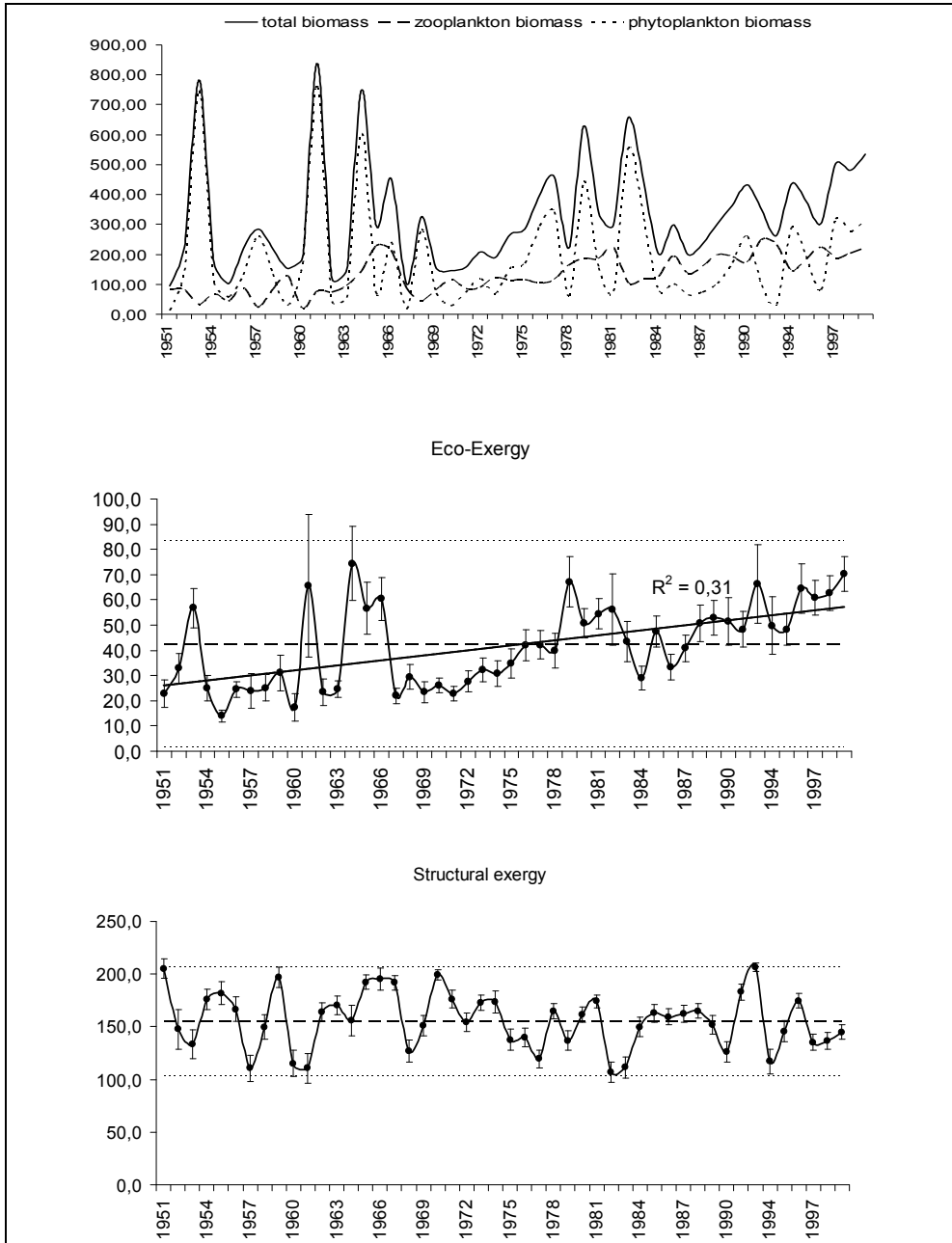


Fig. 2. Long-term dynamics of year-average biomasses of components (mg m<sup>-3</sup>), exergy (g detritus eq m<sup>-3</sup>) and structural exergy of lake Baikal plankton. Dotted line - long-term average.

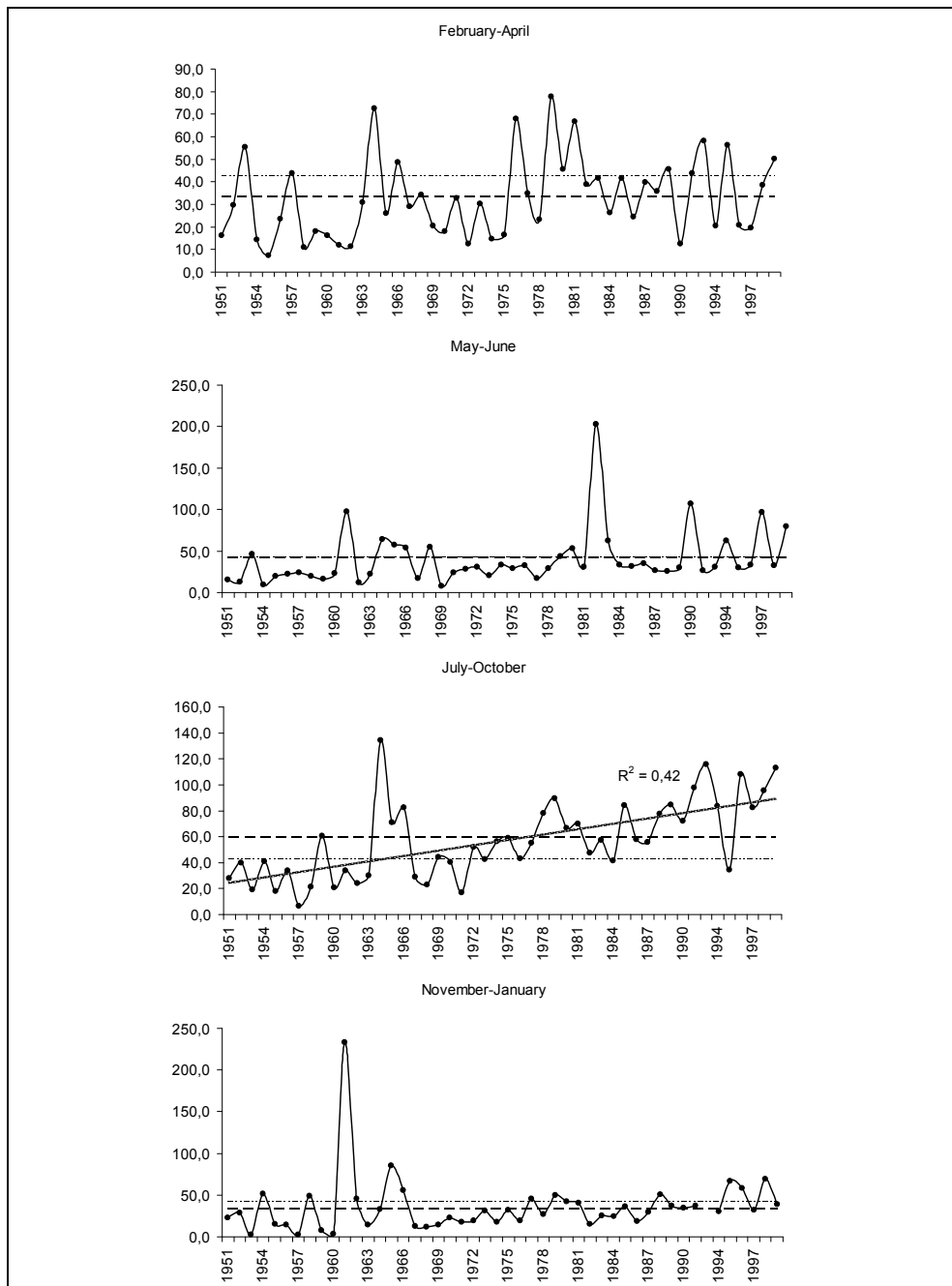


Fig. 3. Long-term dynamics of total exergy (g detritus eq m<sup>-3</sup>) of lake Baikal plankton for different seasons.

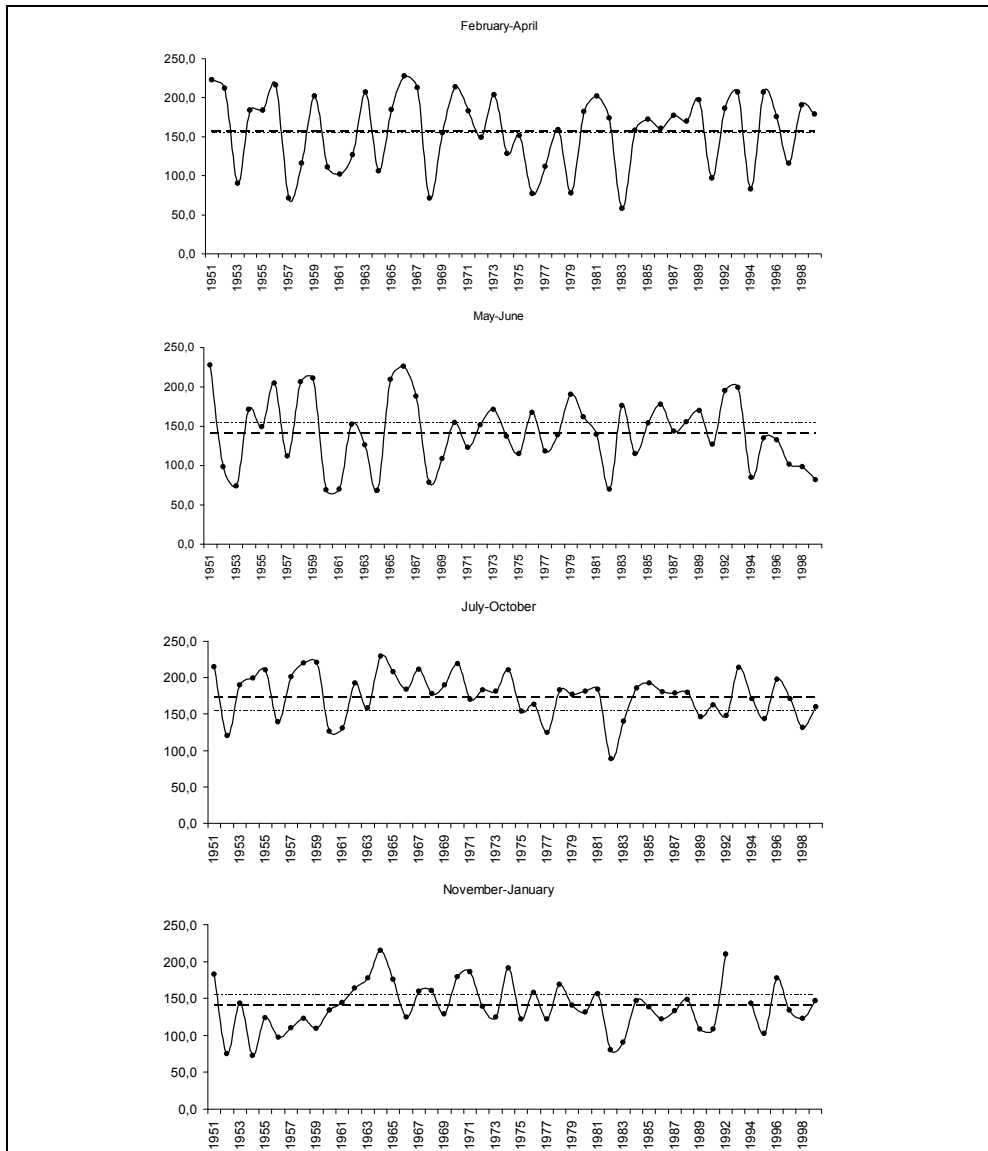


Fig. 4. Long-term dynamics of structural exergy of lake Baikal plankton for different seasons.

As we know from above (sections 3.1, 3.2.4) structural exergy reflects ecosystem health and ability of it to withstand to external influences, e.g. human impact. It is seen from Fig. 4 average long-term under-ice structural exergy (157,2) is practically equal to long-term year-average (154,9), summer structural exergy (175,9) is sufficiently higher than year-average ,



while overturn seasons exergy is lower - 142,5 and 139,4 for spring and fall overturns. It means summer community is much more resistant to external disturbances than under-ice one, and during overturns Baikal is especially sensitive to pollution and other kinds of human impact. It is in good concordance with previously obtained results of our experiments with mathematical models (Silow et al., 1995, 2001; Silow, 1999) and mesocosms (Silow et al., 1991; Silow & Oh, 2004).

## 4. Hypercycles

One of the characteristic features of ecosystems behaviour are cyclic changes of component biomasses and numbers. They are observed both in natural objects and in artificial ecosystems. Application of Lotka - Volterra and energy flow models to systems containing more than two components demonstrates reducing of oscillation amplitude and stabilization of components parameters at fixed levels (Limburg, 1985; Ruan, 2001; Mougi Nishimura, 2007). Some researchers make their systems to oscillate via chaotic (Stone & He, 2007) or stochastic (Abta et al., 2008) behaviour of its components.

Another fact hardly simulated with the use of existing approaches is "paradox of plankton" - coexisting of two or more species in the same ecological niche (Hutchinson, 1961). To explain it researchers are forced to find any, though very small differences in ecological characteristics of these species, such as optimal temperatures or oxygen contents, growth rates, nutrient thresholds for growth, mortality (Ebenhöh, 1988), time and duration of mass development (Nikolaev, 1986), albeit the last can be not the cause but effect of coexistence.

### 4.1 Description of models

Nicolis & Prigozhin (1977) basing on the following suggestions:



where  $A$  - food,  $X$  - component biomass,  $N$  - general organic matter content in closed system,  $k, d$  - growth and death rate coefficients, obtained

$$\frac{dX}{dt} = kX(N - X) - dX . \quad (23)$$

This equation is identical to those used for description of autocatalytic processes, where component  $X$  serves as catalyst for self-creation from substance  $A$ . Such autocatalytic and self-reproduction units can be regarded in cycles called hypercycles (Eigen & Schuster, 1979).

Ecosystem also can be described as hypercycle, where every next trophic level obtains material from previous level to reproduce itself as autocatalyst. Example of such structure is given in Fig. 5. Phytoplankton obtains nutrients from bacteria to create organic matter with the use of external energy (solar irradiation). Zooplankton feeding on phytoplankton obtains organic matter for growth. Bacteriae in turn get food supply from zooplankton corpses and faeces. Of course, this scheme is idealized as bacteriae obtain food from

phytoplankton extracellular products and dead phytoplankton cells, zooplankton can consume not only phytoplankton but also bacteriae etc. Nevertheless this scheme represents the most important ways of energy and matter transfer in closed ecosystem including producers, consumers and reducers. Dynamics of components is determined by following system of equations

$$\begin{aligned} dx_1 / dt &= f(x_1, \mu_1, \phi(x_3)) - g(x_1, \varphi(x_2)), \\ dx_2 / dt &= f(x_2, \mu_2, \phi(x_1)) - m(x_2), \\ dx_3 / dt &= f(x_3, \mu_3, \phi(x_2)) - m(x_3), \end{aligned} \quad (24)$$

where  $x_i$  - biomasses,  $f$  - growth functions,  $m$  - death functions,  $g$  - grazing function,  $\phi$  - effectiveness of energy and matter conversion (for phytoplankton - relation between bacteriae concentration and nutrients availability) between components,  $\varphi$  - effectiveness of grazing,  $\mu$  - maximum growth rate. Indices  $i$  mean: 1 - phytoplankton, 2 - zooplankton, 3 - bacteriae. Function parameters were calculated at the ecosystem stability condition  $dx_i/dt=0$ . Also we have investigated a system including two species of phytoplankton ( $x_{11}$ ,  $x_{12}$ ) competing for nutrient supply and two species of zooplankton ( $x_{21}$ ,  $x_{22}$ ), and bacteriae (Fig. 6). Starting biomasses for these newly introduced species at stability state were  $x_{12}=0,33 x_{11}$ ,  $x_{22}=0,1 x_{21}$ . System was described by the following equations:

$$\begin{aligned} dx_{11} / dt &= f(x_{11}, \mu_{11}, \phi(x_3), \xi(x_{11}, x_{12})) - g(x_{11}, \varphi(x_{21})), \\ dx_{12} / dt &= f(x_{12}, \mu_{12}, \phi(x_3), \xi(x_{12}, x_{11})) - g(x_{12}, \varphi(x_{22})), \\ dx_{21} / dt &= f(x_{21}, \mu_{21}, \phi(x_{11})) - m(x_{21}), \\ dx_{22} / dt &= f(x_{22}, \mu_{22}, \phi(x_{12})) - m(x_{22}), \\ dx_3 / dt &= f(x_3, \mu_3, \phi(x_{21}, x_{22})) - m(x_3), \end{aligned} \quad (25)$$

where  $\xi$  - competition for nutrients function.

#### 4.2 Behaviour of models

Dynamics of model (24) after external influence shows its returning to the stability point. Such behaviour is characteristic for stable non-linear systems (Gnauck, 1979).

There are no oscillations in this system, it always returns to stable state after initial biomasses changes. To make the system oscillate it is necessary to imitate input of nutrients or toxicants into it (Fig. 7). We can remind similar situation obtained by group of R. Pal et al. (2009). Their phytoplankton - zooplankton - nutrients model (with much more sophisticated mathematics, than ours) demonstrated oscillations under toxification. In other works oscillations of ecosystem components were caused by externally driven forces (oscillating environment) (Eladydi & Sacker, 2006; Koszalka et al., 2007). Hypercycles with not more than three components are shown to remain stable with equilibrium concentrations of components regardless with initial concentrations (Köppers, 1985).

Model (25) demonstrates oscillation behaviour around stability point but never reaches it (Fig. 8). It is in good accordance, e.g. with the biomass fluctuations and species population

fluctuations, caused by increase of community size and complexity (Fowler, 2009). It may be connected with the competition for resources (in our case - for nutrients, released by bacteria), as in many works similar results were obtained. For example, in system of two plant populations, competing for one nutrient (Damgaard, 2004), two predators, competing for one prey (Saleem et al., 2003; Yu et al., 2011), three microbial populations, competing for resources (Li, 2001).

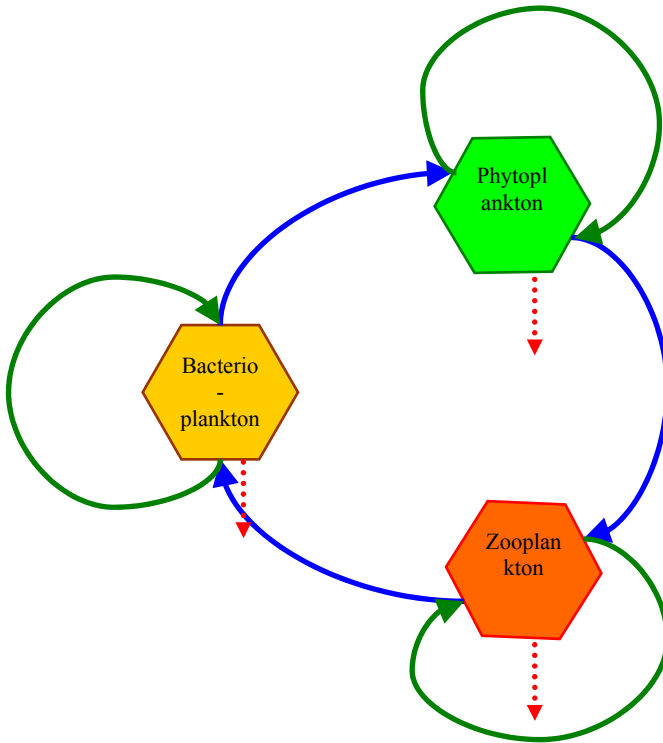


Fig. 5. Idealized scheme of ecosystem as hypercycle. Blue arrows - flows of energy and matter between ecosystem components, green - self-reproduction of components, yellow arrow - income of low entropy (high exergy) energy from external source (Sun), dotted arrows - dissipation of energy in the form of high entropy (low exergy) heat energy.

## 5. Conclusion

It is now becoming clear that the movement away from thermodynamic equilibrium, and the subsequent increase in organization during ecosystem growth and development, is a result of system components and configurations that maximize the flux of useful energy and the amount of stored exergy. Both empirical data, as well as theoretical models, support these conclusions. Exergy is widely used in ecology to analyze theoretical problems and to solve applied tasks. The most perspective use of exergy parameters in recent ecology is the use of them as ecosystem health indicators.

Exergy, and, especially structural exergy, is shown to be a good health indicator for ecosystems in many model, experimental, field and complex case studies.

The application of exergy calculations to long-term dynamics of the lake Baikal plankton demonstrates the steady state of plankton community structural exergy and well observed increase of its total exergy.

Presented model of two competing hypercycles shows: 1) simultaneous coexistence of two ecologically equivalent phytoplankton species obtaining nutrients from the one source; 2) auto-oscillations of all the components included in model in constant environmental conditions, similar to those observed in real ecosystems.

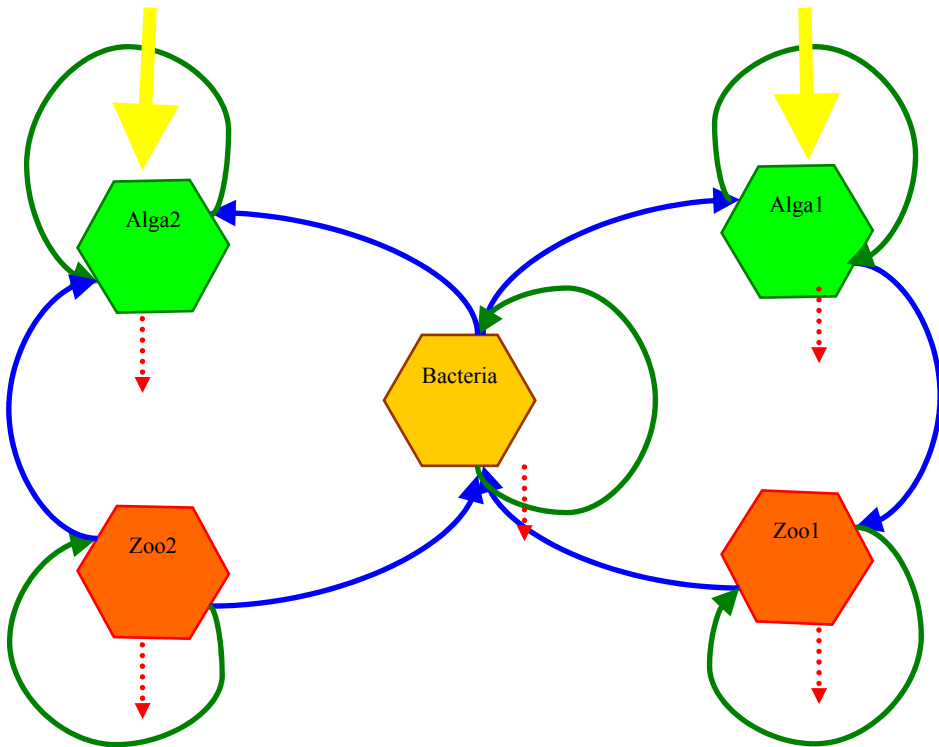


Fig. 6. Idealized scheme of an ecosystem as system of coupled hypercycles.

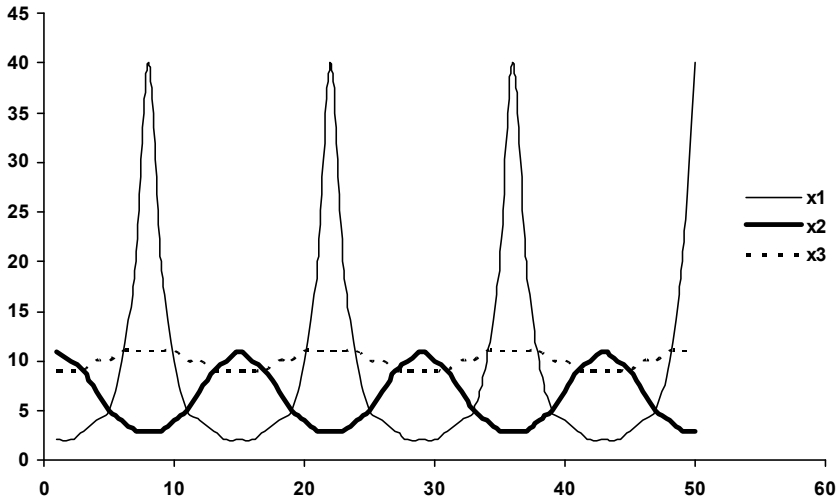


Fig. 7. Dynamics of system (25) at income of pesticide, causing death of 10 % of phytoplankton, together with nutrients (10% of ecosystem storage). Y-axis - biomass ( $\text{kJ m}^{-2}$ ), X-axis - years.

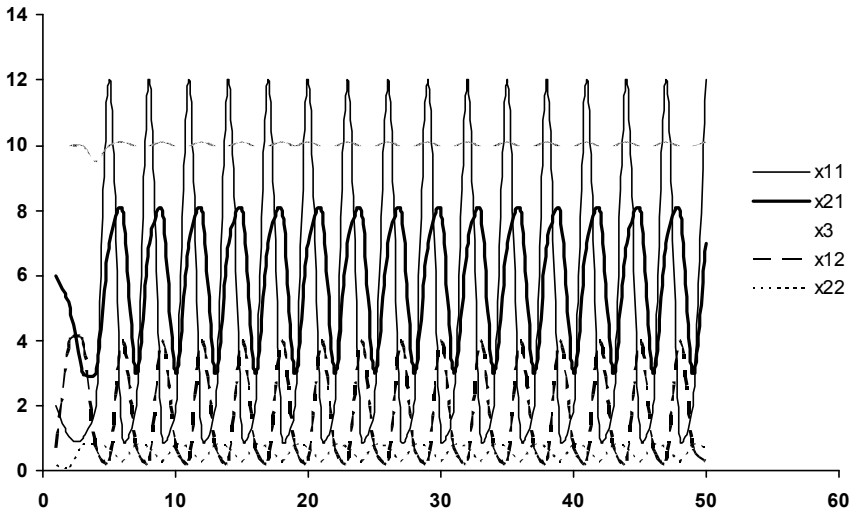


Fig. 8. Dynamics of system (26) without external perturbations.

## 6. Acknowledgments

Authors are greatly thankful to Professor Viktorija Zgela for her patience and permanent help we received from her during the preparation of manuscript. Authors are pleased to acknowledge the Analytical Institutional Program “The Development of the Research Potential of Higher School (2009–2011)”, supported this research with contracts № 2.1.1/1359 and № 2.2.1.1/5901, the Federal Targeted Program “Scientific and Pedagogical Staff for Innovative Russia” for 2009 – 2013, supported this research with contract № 02.740.11.0018.

## 7. References

- Abta, R., Schiffer, M. & Ben-Ishay, A. & Shnerb, N.M. (2008) Stabilization of metapopulation cycles: Towards a classification scheme. *Theoretical Population Biology*, Vol. 74, pp. 273 – 283, ISSN 0040-5809.
- Bendoricchio, G. & Jørgensen, S.E. (1997) Exergy as goal function of ecosystem dynamic. *Ecological Modelling*, Vol. 102, pp. 5–15, ISSN 0304-3800.
- Coffaro, G., Bocci, M. & Bendoricchio, G. (1997) Application of structural dynamic approach to estimate space variability of primary producers in shallow marine water. *Ecological Modelling*, Vol. 102, pp. 97–114, ISSN 0304-3800.
- Damgaard, C. (2004) Dynamics in a discrete two-species competition model: coexistence and over-compensation. *Journal of Theoretical Biology*, Vol. 27, pp. 197–203, ISSN: 0022-5193.
- Ebenhöh, W. (1988) Coexistence of and unlimited number of algal species in a model system. *Theoretical population biology*, Vol. 34, pp. 130 – 144, ISSN 0040-5809.
- Eigen, M. & Schuster, P. (1979) *The Hypercycle. A Principle of Natural Self-Organization*. Springer Verlag, ISBN 0-387-09293-5, Berlin-Heidelberg-New York, Germany-USA.
- Eladydi, S. & Sacker, R.J. (2006) Periodic difference equations, population biology and the Cushing-Henson conjectures. *Mathematical Biosciences*. Vol. 201, pp. 195 – 207, ISSN 0025-5564.
- Fonseca, J.C., Pardal, M.A., Azeiteiro, U.M. & Marques, J.C. (2002) Estimation of ecological exergy using weighing parameters determined from DNA contents of organisms – A case study. *Hydrobiologia*, Vol. 475/476, pp. 79–90, ISSN 0018-8158.
- Fowler, M.S. (2009) Increasing community size and connectance can increase stability in competitive communities. *Journal of Theoretical Biology*, Vol. 258, pp. 179–188, ISSN: 0022-5193.
- Gnauck, A. H. (1979) Grundlagen der Stabilitätsanalyse limnischer ökosysteme. *Acta hydrochimica et hydrobiologica*, Vol. 7, N 1, pp. 5 – 25, ISSN 0323-4320.
- Hampton, S.E., Izmet'eva, L.R., Moore, M.V., Katz, S.L., Dennis, B. & Silow, E.A. (2008) Sixty years of environmental change in the in the world's largest freshwater lake – Lake Baikal, Siberia. *Global Change Biology*, Vol. 14, pp. 1947–1958, ISSN 1365-2486.
- Hutchinson, G. (1961) The paradox of the plankton. *American Naturalist*, Vol. 95, pp. 137 – 146, ISSN 0003-0147.
- Izmestyeva, L. & Silow, E. (2010) Long-term dynamics of summer community of Baikal phytoplankton and climate change. In: *13th World Lake Papers*. ILEC, Shiga, 2010, Available from: [http://wldb.ilec.or.jp/data/ilec/WLC13\\_Papers/others/9.pdf](http://wldb.ilec.or.jp/data/ilec/WLC13_Papers/others/9.pdf)
- Jørgensen, S.E. (1992) Parameters, ecological constraints and exergy. *Ecological Modelling*, Vol. 62, pp. 163–170, ISSN 0304-3800.

- Jørgensen, S.E. (1999) State-of-the-art of Ecological Modelling with emphasis on development of structural dynamic models. *Ecological Modelling*, Vol. 120, pp. 75–96, ISSN 0304-3800.
- Jørgensen, S.E. (2002) *Integration of Ecosystem Theories: A Pattern (Ecology and Environment)*. 3d edition. Springer, ISBN 978-0-7923-4523-7, New York, USA.
- Jørgensen, S.E. (2006a) Application of holistic thermodynamic indicators. *Ecological Indicators*, 2006, Vol. 6, pp. 24–29, ISSN: 1470-160X.
- Jørgensen, S.E. (2006b) Description of aquatic ecosystem's development by Eco-Exergy and exergy destruction. *Ecological Modelling*, Vol. 204, pp. 22–28, ISSN 0304-3800.
- Jørgensen, S.E. (2006-2007) *Integrated Ecosystem Theory*, *Ann. – Eur. Acad. Sci.*, EAS Publishing House, Liege, ISBN 978-0-7923-4523-7.
- Jørgensen, S.E. (2011) *Introduction to Systems Ecology*. Taylor & Francis Group/CRC Press, Boca Raton, USA. (In press)
- Jørgensen, S.E. & Fath, B. (2004) Application of thermodynamic principles in ecology. *Ecological Complexity*, Vol. 1, pp. 267-280, ISSN 1476-945X.
- Jørgensen, S.E. & Fath, B.D. (2011) *Fundamentals of Ecological Modelling: Applications in Environmental Management and Research*, Elsevier, 4<sup>th</sup> ed., ISBN 978-0-444-53567-2, Amsterdam, The Netherlands.
- Jørgensen, S.E. & Padišak, J. (1996) Does the intermediate disturbance hypothesis comply with thermodynamics? *Ecological Modelling*, Vol. 323, pp. 9–21, ISSN 0304-3800.
- Jørgensen, S.E. & Svirezhev, Y.M. (2004) *Towards a Thermodynamic Theory for Ecological Systems*, Elsevier, ISBN 978-0-08-044166-5, Amsterdam, The Netherlands.
- Jørgensen, S.E., Patten, B.C. & Straškraba, M. (1999) Ecosystem emerging: 3. Openness. *Ecological Modelling*, Vol. 117, pp. 41–64, ISSN 0304-3800.
- Jørgensen, S.E., Patten, B.C. & Straškraba, M. (2000) Ecosystems emerging: 4. Growth. *Ecological Modelling*, Vol. 126, pp. 249-284, ISSN 0304-3800.
- Jørgensen, S.E., Ladegaard, N., Debeljak, M. & Marques, J.C. (2005) Calculations of exergy for organisms. *Ecological Modelling*, Vol. 185, pp. 165–175, ISSN 0304-3800.
- Köppers, B.O. (1985) *Molecular theory of evolution. Outline of a physico-chemical theory of the origin of life*. Springer-Verlag, ISBN 0387155287, Berlin, Germany.
- Koszalka, I., Bracco, A., Pasquero, C. & Provenzale, A. (2007) Plankton cycles disguised by turbulent advection. *Theoretical Population Biology*, Vol. 72, pp. 1 – 6, ISSN 0040-5809.
- Leguerrier, D., Degre, D. & Niquil, N. (2007) Network analysis and inter-ecosystem comparison of two intertidal mudflat food webs (Brouage Mudflat and Aiguillon Cove, SW France). *Estuarine Coastal Shelf Science*, Vol. 74, pp. 403–418, ISSN: 0272-7714.
- Li, B. (2001) Periodic coexistence in the chemostat with three species competing for three essential resources. *Mathematical Biosciences*, Vol. 174, pp. 27-40, ISSN 0025-5564.
- Libralato, S., Torricelli, P. & Pranovi, F. (2006) Exergy as ecosystem indicator: An application to the recovery process of marine benthic communities. *Ecological Modelling*, Vol. 192, pp. 571–585, ISSN 0304-3800.
- Limburg K. E.: (1985) Increasing complexity and energy flow in models of food webs. *Ecological Modelling*, Vol. 29, pp. 5 - 25, ISSN 0304-3800.
- Ludovisi, A. (2009) Exergy vs information in ecological successions: Interpreting community changes by a classical thermodynamic approach. *Ecological Modelling*, Vol. 220, pp. 1566–1577, ISSN 0304-3800.

- Ludovisi, A. & Poletti, A. (2003) Use of thermodynamic indices as Ecological indicators of the development state of lake ecosystems. 2. Exergy and specific exergy indices. *Ecological Modelling*, Vol. 159, pp. 223-238, ISSN 0304-3800.
- Ludovisi, A., Pandolfi, P. & Taticchi, M.I. (2005) The strategy of ecosystem development: Specific dissipation as an indicator of ecosystem maturity. *Journal of Theoretical Biology*, Vol. 235, pp. 33-43, ISSN 0022-5193.
- Mandal, S., Ray, S., Ray, S. & Jørgensen, S.E. (2007) Investigation of thermodynamic properties in an ecological model developing from ordered to chaotic states. *Ecological Modelling*, Vol. 204, pp. 40-46, ISSN 0304-3800.
- Marques, J.C., Pardal, M.A., Nielsen, S.N. & Jørgensen, S.E. (1997) Analysis of the properties of exergy and biodiversity along an estuarine gradient of eutrophication. *Ecological Modelling*, Vol. 102, pp. 155-167, ISSN 0304-3800.
- Marques, J.C., Nielsen, S.N., Pardal, M.A. & Jørgensen, S.E. (2003) Impact of eutrophication and river management within a framework of ecosystem theories. *Ecological Modelling*, Vol. 166, pp. 147-168, ISSN 0304-3800.
- Mejer, H.F. & Jørgensen, S.E. (1977) Energy and ecological buffer capacity. In *State of the Art of Ecological Modelling Environmental Sciences and Applications*, Jørgensen, S.E., Ed., pp. 829-846, International Society for Ecological Modelling, Copenhagen, Denmark.
- Moore, M.V., Hampton, S.E., Izmet'eva, L.R., Silow, E.A., Peshkova, E.V. & Pavlov, B.K. (2009) Climate Change and the World's "Sacred Sea" – Lake Baikal, Siberia. *BioScience*, Vol. 59, pp. 405-417, ISSN 0006-3568.
- Mougi, A. A & Nishimura, K. (2007) A resolution of the paradox of enrichment. *Journal of Theoretical Biology*, Vol. 248, pp. 194-201, ISSN 0022-5193.
- Nicolis, G. & Prigogzhine, I. (1977) *Self-organization in nonequilibrium systems. From dissipative structures to order through fluctuations*. J. Wiley & Sons, ISBN 0471024015, New York, USA.
- Nielsen, S.N. (1997) Examination and optimization of different exergy forms in macrophyte societies. *Ecological Modelling*, Vol. 102, pp. 115-127, ISSN 0304-3800.
- Nielsen, S.N. (2000) Thermodynamics of an ecosystem interpreted as a hierarchy of embedded systems. *Ecological Modelling*. Vol. 135, pp. 279-289, ISSN 0304-3800.
- Nikolaev, I.I. (1986) Some mechanisms of taxonomically related species coexistence in planktonic communities. *Hydrobiological Journal*, Vol. 22, N 4, pp. 9 - 16, 0375-8990, (in Russian).
- Nunneri, C., Lenhart, H.J., Burkhard, B. & Windhorst, W. (2008) Ecological risk as a tool for evaluating the effects of offshore wind farm construction in the North Sea. *Regional Environmental Change*, Vol. 8, pp. 31-43, ISSN 1436-3798.
- Pal, R., Basu, D. & Banerjee, M. (2009) Modelling of phytoplankton allelopathy with Monod-Haldane-type functional response - A mathematical study. *BioSystems*, Vol. 95, pp. 243-253, ISSN 0303-2647.
- Park, Y.S., Lek, S., Scardi, M., Verdonshot, P.F.M. & Jørgensen, S.E. (2006) Patterning exergy of benthic macroinvertebrate communities using self-organizing maps. *Ecological Modelling*, Vol. 195, pp. 105-113, ISSN 0304-3800.
- Patten, B.C., Straškraba, M. & Jørgensen, S.E. (1997) Ecosystems emerging: 1. Conservation. *Ecological Modelling*, Vol. 96, pp. 221-284, ISSN 0304-3800.
- Pislegina, E. & Silow, E. (2010) Long-term dynamics of Baikal zooplankton and climate change. *13th World Lake Conference Papers*, ILEC, Shiga, 2010, Available from: [http://wldb.ilec.or.jp/data/ilec/WLC13\\_Papers/others/10.pdf](http://wldb.ilec.or.jp/data/ilec/WLC13_Papers/others/10.pdf)



- Pranovi, F., da Ponte, F. & Torricelli, P. (2008) Historical changes in the structure and functioning of the benthic community in the lagoon of Venice. *Estuarine Coastal Shelf Science*, Vol. 76, pp. 753–764, ISSN 0272-7714.
- Pykh, Y.A., Kennedy, E.T. & Grant, W.E. (2000) An overview of system analysis methods in delineating environmental quality indices. *Ecological Modelling*, Vol. 130, pp. 25–38, ISSN 0304-3800.
- Ray, S., Berec, L., Straškraba, M. & Jørgensen, S.E. (2001) Optimization of exergy and implications of body sizes of phytoplankton and zooplankton in an aquatic ecosystem model. *Ecological Modelling*, Vol. 140, pp. 219–234, ISSN 0304-3800.
- Ruan, S. (2001) Oscillations in plankton models with nutrient recycling. *Journal of Theoretical Biology*, Vol. 208, pp. 15–26, ISSN 0022-5193.
- Saleem, M., Tripathi, A.K. & Sadiyal, A.H. (2003) Coexistence of species in a defensive switching model. *Mathematical Biosciences*, Vol. 181, pp. 145 – 164, ISSN 0025-5564.
- Silow, E.A. (1998) The changes of ecosystem goal functions in stressed aquatic communities. *Journal of Lake Science*, Vol. 10, 421–435, ISSN 1003-5427.
- Silow, E.A. (1999) The use of two lumped models for the analysis of consequences of external influences on the lake Baikal ecosystem. *Ecological Modelling*, 1999, Vol. 121, pp. 103–113, ISSN 0304-3800.
- Silow, E.A. (2006) Possibility of using goal functions for estimation of “health” of ecological aquatic systems: exergy. *Siberian Ecological Journal*, Vol. 3, pp. 269–284, ISSN 0869-8619.
- Silow, E. (2010) Lake Baikal as possible sentinel of the Climate Change. *13th World Lake Conference Papers*, ILEC, Shiga, 2010, Available from:  
[http://wldb.ilec.or.jp/data/ilec/WLC13\\_Papers/S2/s2-6.pdf](http://wldb.ilec.or.jp/data/ilec/WLC13_Papers/S2/s2-6.pdf)
- Silow, E.A. & Mokry, A.V. (2010) Exergy as a Tool for Ecosystem Health Assessment. *Entropy*, Vol. 12, pp. 902-925, EISSN 1099-4300.
- Silow, E.A., Oh, I.H. (2004) Aquatic ecosystem assessment using exergy. *Ecological Indicators*, Vol. 4, pp. 189–198, ISSN 1470-160X.
- Silow E. A., Baturin, V. A. & Stom, D. J. (2001) Prediction of Lake Baikal ecosystem behaviour using an ecosystem disturbance model. *Lakes & Reservoirs: Research and Management*, Vol. 6, N 1, pp. 33–36, ISSN 1440-1770.
- Silow, E. A., Stom, D. J. & Basharova, N. I. *et al.* (1991) Influence of biogenous elements on the lake Baikal plankton community. *Acta hydrochimica et hydrobiologica*, Vol. 19, N 6, pp. 629 – 634, ISSN 0323-4320.
- Silow, E. A., Gurman, V. J., Stom, D. J., Rosenraukh, D. M. & Baturin, V. I. (1995) Mathematical models of lake Baikal ecosystem. *Ecological Modelling*, Vol. 82, pp. 27 – 39, ISSN 0304-3800.
- Stone, L. & He, D. (2007) Chaotic oscillations and cycles in multi-trophic ecological systems. *Theoretical Population Biology*, Vol. 248, pp. 382 – 390, ISSN 0040-5809.
- Straškraba, M., Jørgensen, S.E. & Patten, B.C. (1999) Ecosystem emerging: 2. Dissipation. *Ecological Modelling*, Vol. 117, pp. 3–39, ISSN 0304-3800.
- Svirezhev, Y.M. (2000) Thermodynamics and ecology, *Ecological Modelling*, Vol. 132, pp. 11–22, ISSN 0304-3800.
- Svirezhev, Y.M. & Steinborn, W.H. (2001) Exergy of solar radiation: information approach. *Ecological Modelling*, Vol. 145, pp. 101–110, ISSN 0304-3800.
- Vassallo, P., Fabiano, M., Vezzulli, L., Sandulli, R., Marques, J.C. & Jørgensen, S.E. (2006) Assessing the health of coastal marine ecosystems: Aholistic approach based on

- sediment micro and meio-benthic measures. *Ecological Indicators*, Vol. 6, pp. 525–542, ISSN: 1470-160X.
- Windhorst, W., Colijn, F., Kabuta, S., Laane, R.P.W.M. & Lenhart, H.-J. (2005) Defining a good ecological status of coastal waters—A case study for the Elbe plume. In *Managing European Coasts: Past, Present, and Future*, Vermaat, J.E., Salomons, W., Bouwer L. Eds., Springer-Verlag, pp. 59–74, ISBN 978-3-540-23454-8, Berlin, Germany.
- Xu, F.L. (1997) Exergy and structural exergy as Ecol. Indic. for the development state of the Lake Chaohu ecosystem. *Ecological Modelling*, Vol. 99, pp. 41–49, ISSN 0304-3800.
- Xu, F.L., Jørgensen, S.E. & Tao, S. (1999a) Ecological indicators for assessing freshwater ecosystem health. *Ecological Modelling*, Vol. 116, pp. 77–106.
- Xu, F.L., Jørgensen, S.E., Tao, S. & Li, B.G. (1999b) Modelling the effects of ecological engineering on ecosystem health of a shallow eutrophic Chinese Lake (Lake Chao). *Ecological Modelling*, Vol. 117, pp. 239–260, ISSN 0304-3800.
- Xu, F.L., Dawson, R.W., Tao, S., Cao, J. & Li, B.G. (2001) A method for ecosystem health assessment: An Ecological Modeling Method (EMM) and its application. *Hydrobiologia*, Vol. 443, pp. 159–175, ISSN 0018-8158.
- Xu, F.L., Zhao, Z.Y., Zhan, W., Zhao, S.S., Dawson, R.W. & Tao, S. (2005) An ecosystem health index methodology (EHIM) for lake ecosystem health assessment. *Ecological Modelling*, Vol. 188, pp. 327–339, ISSN 0304-3800.
- Xu, F., Yang, Z.F. & Zhao, Y.W. (2011b) Ecosystem health assessment of the plant-dominated Baiyangdian Lake based on eco-exergy. *Ecological Modelling*, Vol. 222, p. 201–209, ISSN 0304-3800.
- Xu, F.L., Lam, K.C., Zhao, Z.Y., Zhan, W., Chen, Y.D. & Tao, S. (2004) Marine coastal ecosystem health assessment: A case study of the Tolo Harbour, Hong Kong, China *Ecological Modelling*, Vol. 173, pp. 355–370, ISSN 0304-3800.
- Xu, F.L., Wang, J.J., Chen, B., Qin, N., Wu, W.J., He, W., He, Q.S. & Wang, Y. (2011a) The variations of exergies and structural exergies along eutrophication gradients in Chinese and Italian lakes. *Ecological Modelling*, Vol. 222, pp. 337–350, ISSN 0304-3800.
- Ye, C., Xu, Q., Kong, H., Shen, Z. & Yan, C. (2007) Eutrophication conditions and ecological status in typical bays of Lake Taihu in China. *Environmental Monitoring and Assessment*, Vol. 135, pp. 217–225, ISSN 0964-4016.
- Yu, H., Zhong, S. & Agarwal, R.P. (2011) Mathematical analysis and chaos in an ecological model with an impulsive control strategy. *Communications in Nonlinear Science and Numerical Simulation*, Vol. 16, pp. 776–786, ISSN 1007-5704.
- Zhang, F.J., Tong, C.F., Xie Z.F. & Lu, J.J. (2009) Different Ecological indicators used in a recolonization field study of benthic communities and its compliance with ecosystem theories. *Estuarine Coastal Shelf Science*, Vol. 81, pp. 463–469, ISSN 0272-7714.
- Zhang, J., Jørgensen, S.E., Tan, C.O. & Beklioglu, M. (2003a) A structurally dynamic modelling—Lake Mogan, Turkey as a case study. *Ecological Modelling*, Vol. 164, pp. 103–120, ISSN 0304-3800.
- Zhang, J., Jørgensen, S.E., Beklioglu, M. & Ince, O. (2003b) Hysteresis in vegetation shift—Lake Mogan prognoses. *Ecological Modelling*, Vol. 164, pp. 227–238, ISSN 0304-3800.

# Statistical Thermodynamics of Material Transport in Non-Isothermal Mixtures

Semen Semenov<sup>1</sup> and Martin Schimpf<sup>2</sup>

<sup>1</sup>*Institute of Biochemical Physics RAS,*

<sup>2</sup>*Boise State University, Boise*

<sup>1</sup>*Russia*

<sup>2</sup>*USA*

## 1. Introduction

This chapter outlines a theoretical framework for the microscopic approach to material transport in liquid mixtures, and applies that framework to binary one-phase systems. The material transport in this approach includes no hydrodynamic processes related to the macroscopic transfer of momenta. In analyzing the current state of thermodynamic theory, we indicate critically important refinements necessary to use non-equilibrium thermodynamics and statistical mechanics in the application to material transport in non-isothermal mixtures.

## 2. Thermodynamic theory of material transport in liquid mixtures: Role of the Gibbs-Duhem equation

The aim of this section is to outline the thermodynamic approach to material transport in mixtures of different components. The approach is based on the principle of local equilibrium, which assumes that thermodynamic principles hold in a small volume within a non-equilibrium system. Consequently, a small volume containing a macroscopic number of particles within a non-equilibrium system can be treated as an equilibrium system. A detailed discussion on this topic and references to earlier work are given by Gyarmati (1970). The conditions required for the validity of such a system are that both the temperature and molecular velocity of the particles change little over the scale of molecular length or mean free path (the latter change being small relative to the speed of sound). For a gas, these conditions are met with a temperature gradient below  $10^4 \text{ K cm}^{-1}$ ; for a liquid, where the heat conductivity is greater, the speed of sound higher and the mean free path is small, this condition for local equilibrium is more than fulfilled, provided the experimental temperature gradient is below  $10^4 \text{ K cm}^{-1}$ .

Thermodynamic expressions for material transport in liquids have been established based on equilibrium thermodynamics (Gibbs and Gibbs-Duhem equations), as well as on the principles of non-equilibrium thermodynamics (thermodynamic forces and fluxes). For a review of these models, see (De Groot, 1952; De Groot, Mazur, 1962; Kondepudi, Prigogine, 1999; Haase, 1969).

Non-equilibrium thermodynamics is based on the entropy production expression

$$\Sigma = \bar{J}_e \cdot \nabla \left( \frac{1}{T} \right) - \sum_{i=1}^N \bar{J}_i \cdot \nabla \left( \frac{\mu_i}{T} \right) \quad (1)$$

where  $\bar{J}_e$  is the energy flux,  $\bar{J}_i$  are the component material fluxes,  $N$  is the number of the components,  $\mu_i$  are the chemical potentials of components, and  $T$  is the temperature. The energy flux and the temperature distribution in the liquid are assumed to be known, whereas the material concentrations are determined by the continuity equations

$$\frac{\partial n_i}{\partial t} = -\nabla \bar{J}_i \quad (2)$$

Here  $n_i$  is the numerical volume concentration of component  $i$  and  $t$  is time. Non-equilibrium thermodynamics defines the material flux as

$$\bar{J}_i = -n_i L_i \nabla \left( \frac{\mu_i}{T} \right) - n_i L_{iQ} \nabla \left( \frac{1}{T} \right) \quad (3)$$

where  $L_i$  and  $L_{iQ}$  are individual molecular kinetic coefficients. The second term on the right-hand side of Eq. (3) represents the cross effect between material flux and heat flux.

The chemical potentials are expressed through component concentrations and other physical parameters (De Groot, 1952; De Groot, Mazur, 1962; Kondepudi, Prigogine, 1999):

$$\nabla \mu_k = \sum_{l=1}^2 \frac{\partial \mu_k}{\partial n_l} \nabla n_l - \bar{v}_k \nabla P + \frac{\partial \mu_k}{\partial T} \nabla T \quad (4)$$

Here  $P$  is the internal macroscopic pressure of the system and  $\bar{v}_k = -\partial \mu_k / \partial P$  is the partial molecular volume, which is nearly equivalent to the specific molecular volume  $v_k$ . Substituting Eq. (4) into Eq. (3), and using parameter  $q_i = L_{iQ} / L_i$ , termed the molecular heat of transport, we obtain the equation for component material flux:

$$\bar{J}_i = -\frac{n_i L_i}{T} \left[ \sum_{k=1}^N \frac{\partial \mu_i}{\partial n_k} \nabla n_k - \bar{v}_i \nabla P + \left( \frac{\partial \mu_i}{\partial T} - \frac{\mu_i + q_i}{T} \right) \nabla T \right] \quad (5)$$

Defining the relation between the heat of transport and thermodynamic parameters is a key problem because the Soret coefficient, which is the parameter that characterizes the distribution of components concentrations in a temperature gradient, is expressed through the heat of transport (De Groot, 1952; De Groot, Mazur, 1962). A number of studies that offer approaches to calculating the heat of transport are cited in (Pan S et al., 2007).

Eq. (5) must be augmented by an equation for the macroscopic pressure gradient in the system. The simplest possible approach is to consider the pressure to be constant (De Groot, 1952; De Groot, Mazur, 1962; Kondepudi, Prigogine, 1999; Haase, 1969; Landau, Lifshitz, 1959), but pressure cannot be constant in a system with a non-uniform temperature and concentration. This issue is addressed with a well-known expression referred to as the

Gibbs-Duhem equation (De Groot, 1952; De Groot, Mazur, 1962; Kondepudi, Prigogine, 1999; Haase, 1969; Landau, Lifshitz, 1959; Ghorayeb, Firoozabadi, 2000; Pan S et al., 2007):

$$\nabla P = \sum_{i=1}^N n_i \left( \sum_{k=1}^N \frac{\partial \mu_i}{\partial n_k} \nabla n_k + \frac{\partial \mu_i}{\partial T} \nabla T \right) \quad (6)$$

The Gibbs-Duhem equation defines the macroscopic pressure gradient in a thermodynamic system. In equilibrium thermodynamics the equation defines the potentiality of the thermodynamic functions (Kondepudi, Prigogine, 1999). In equilibrium thermodynamics the change in the thermodynamic function is determined only by the initial and final states of the systems, without consideration of the transition process itself. In non-equilibrium thermodynamics, Eq. (5) plays the role of expressing mechanical equilibrium in the system. According to the Prigogine theorem (De Groot, 1952; De Groot, Mazur, 1962; Kondepudi, Prigogine, 1999; Haase, 1969), pressure gradient cancels the volume forces expressed as the gradients of the chemical potentials and provides mechanical equilibrium in a thermodynamically stable system. However, in a non-isothermal system, the same authors considered a constant pressure and the left- and right-hand side of Eq. (6) were assumed to be zero simultaneously, which is both physically and mathematically invalid.

Substituting Eq. (6) into Eq. (5) we obtain the following equation for material flux:

$$\bar{J}_i = -\frac{\phi_i L_i}{v_i T} \left[ (1 - \phi_i) \left( \sum_{k=1}^N \frac{\partial \mu_i}{\partial \phi_k} \nabla \phi_k + \frac{\partial \mu_i}{\partial T} \nabla T \right) - \sum_{k \neq i} \frac{v_i \phi_k}{v_k} \sum_{l=1}^N \frac{\partial \mu_k}{\partial \phi_l} \nabla \phi_l + \frac{\partial \mu_k}{\partial T} \nabla T - (\mu_i + q_i) \frac{\nabla T}{T} \right] \quad (7)$$

In Eq. (7), the numeric volume concentrations of the components are replaced by their volume fractions  $\phi_i = n_i v_i$ , which obey the equation

$$\sum_{i=1}^N \phi_i = 1 \quad (8)$$

Using Eq. (8) and the standard rule of differentiation of a composite function

$$\frac{\partial \mu_k [\phi_1, \phi_1 (\dots, \phi_l, \dots)]}{\partial \phi_1} \nabla \phi_1 = \frac{\partial \mu_k}{\partial \phi_1} \nabla \phi_1 + \frac{\partial \mu_k}{\partial \phi_1} \frac{\partial \phi_1}{\partial \phi_1} \nabla \phi_1 = 2 \frac{\partial \mu_k}{\partial \phi_1} \nabla \phi_1 \quad (9)$$

we can eliminate  $\phi_1$  and obtain Eq. (7) in a more compact form:

$$\bar{J}_i = -\frac{L_i}{T} \sum_k \frac{\phi_i}{v_i} \left[ \phi_k \left( 2 \sum_{l>1}^N \frac{\partial \mu_{ik}^*}{\partial \phi_l} \nabla \phi_l + \frac{\partial \mu_{ik}^*}{\partial T} \nabla T \right) - (\mu_i + q_i) \frac{\nabla T}{T} \right] \quad (10)$$

Here  $\phi_1$  is expressed through the other volume fractions using Eq. (8), and the following combined chemical potential is introduced:

$$\mu_{ik}^* = \mu_i - \frac{v_i}{v_k} \mu_k \quad (11)$$

We note that the volume fraction selected for elimination is arbitrary (any other volume fraction can be eliminated in the same manner), and that in subsequent mathematical

expressions, we express the volume fraction of the first component through that of the others using Eq. (8).

Equations for the material fluxes are usually augmented by the following equation, which relates the material fluxes of components (De Groot, 1952; De Groot, Mazur, 1962; Kondepudi, Prigogine, 1999; Haase, 1969; Ghorayeb, Firoozabadi, 2000; Pan S et al., 2007):

$$\sum_{i=1}^N v_i \bar{J}_i = 0 \quad (12)$$

Eq. (12) expresses the conservation mass in the considered system and the absence of any hydrodynamic mass transfer. Also, Eq. (12) is used to eliminate one of the components from the series of component fluxes expressed by Eq. (10). That material flux that is replaced in this way is arbitrary, and the resulting concentration distribution will depend on which flux is selected. The result is not significant in a dilute system, but in non-dilute systems this practice renders an ambiguous description of the material transport processes.

In addition to being mathematically inconsistent with Eq. (12) because there are  $N+1$  equations [i.e.,  $N$  Eq. (10) plus Eq. (12)] for  $N-1$  independent component concentrations, Eq. (10) predicts a drift in a pure liquid subjected to a temperature gradient. Thus, at  $\phi_i = 1$  Eq. (10) predicts

$$\bar{J}_i = -\frac{L_i}{T} \frac{(\mu_i + q_i)}{v_i} \frac{\nabla T}{T} \quad (13)$$

This result contradicts the basic principle of local equilibrium, and the notion of thermodiffusion as an effect that takes place in mixtures only. Moreover, Eq. (13) indicates that the achievement of a stationary state in a closed system is impossible, since material transport will occur even in a pure liquid.

The contradiction that a system cannot reach a stationary state, as expressed in Eq. (13), can be eliminated if we assume

$$q_i = -\mu_i \quad (14)$$

With such an assumption Eq. (10) can be cast in the following form:

$$\bar{J}_i = -\frac{L_i}{T} \sum_k \frac{\phi_i \phi_k}{v_i} \left( 2 \sum_{l>1}^N \frac{\partial \mu_{ik}^*}{\partial \phi_l} \nabla \phi_l + \frac{\partial \mu_{ik}^*}{\partial T} \nabla T \right) \quad (15)$$

Because the kinetic coefficients are usually calculated independently from thermodynamics, the material fluxes expressed by Eq. (15) cannot satisfy Eq. (12) for the general case. But in a closed and stationary system, where  $\bar{J}_i = 0$ , Eqs. (12) and (15) become consistent. In this case, any component flux can be expressed by Eq. (15) through summation of the other equations. The condition of mechanical equilibrium for an isothermal homogeneous system, as well as the use of Eqs. (1) – (6) for non-isothermal systems, are closely related to the principle of local equilibrium (De Groot, 1952; De Groot, Mazur, 1962; Kondepudi, Prigogine, 1999; Haase, 1969). As argued in (Duhr, Braun, 2006; Weinert, Braun, 2008), thermodiffusion violates local equilibrium because the change in free energy across a particle is typically comparable to the thermal energy of the particle. However, their calculations predict that

even for large (micron size) particles, the energy difference is no more than a few percent of  $kT$ . But the local equilibrium is determined by processes at molecular level, as will be discussed below, and this argumentation cannot be accepted.

### 3. Dynamic pressure gradient in open and non-stationary systems: Thermodynamic equations of material transport with the Soret coefficient as a thermodynamic parameter

Expressing the heats of transport by Eq. (14), we derived a set of consistent equations for material transport in a stationary closed system. However, expression for the heat of transport itself cannot yield consistent equations for material transport in a non-stationary and open system.

In an open system, the flux of a component may be nonzero because of transport across the system boundaries. Also, in a closed system that is non-stationary, the component material

fluxes  $\bar{J}_i$  can be nonzero even though the total material flux in the system,  $\bar{J} = \sum_{i=1}^N v_i \bar{J}_i$ , is

zero. In both these cases, the Gibbs-Duhem equation can no longer be used to determine the pressure in the system, and an alternate approach is necessary.

In previous works (Schimpf, Semenov, 2004; Semenov, Schimpf, 2005), we combined hydrodynamic calculations of the kinetic coefficients with the Fokker-Planck equations to obtain material transfer equations that contain dynamic parameters such as the cross-diffusion and thermal diffusion coefficients. In that approach, the macroscopic gradient of pressure in a binary system was calculated from equations of continuity of the same type as expressed by Eqs. (2) and (8). This same approach may be used for solving the material transport equations obtained by non-equilibrium thermodynamics.

In this approach, the continuity equations [Eq. (2)] are first expressed in the form

$$\frac{\partial \phi_i}{\partial t} = \nabla \frac{\phi_i L_i}{T} \left( 2 \sum_{k>1}^N \frac{\partial \mu_i}{\partial \phi_k} \nabla \phi_k - \bar{v}_i \nabla P + \frac{\partial \mu_i}{\partial T} \nabla T \right) \quad (16)$$

Summing Eq. (16) for each component and utilizing Eq. (8) we obtain the following equation for the dynamic pressure gradient in an open non-stationary system:

$$\nabla P = \left[ \bar{J} T + \sum_{i=1}^N \phi_i L_i \left( 2 \sum_{k>1}^N \frac{\partial \mu_i}{\partial \phi_k} \nabla \phi_k + \frac{\partial \mu_i}{\partial T} \nabla T \right) \right] \left/ \sum_{i=1}^N \phi_i L_i \bar{v}_i \right. \quad (17)$$

Substituting Eq. (17) into Eq. (16) we obtain the material transport equations:

$$\frac{\partial \phi_i}{\partial t} = \nabla \frac{\phi_i L_i}{T} \left\{ \left[ \bar{J} T + \sum_{j=1}^N \phi_j v_j L_j \left( 2 \sum_{k>1}^N \frac{\partial \mu_{ij}^*}{\partial \phi_k} \nabla \phi_k + \frac{\partial \mu_{ij}^*}{\partial T} \nabla T \right) \right] \left/ \sum_{k=1}^N \phi_k v_k L_k \right. \right\} \quad (18)$$

Comparing Eq. (18) with Eq. (15) for a stationary mixture shows that former contains an

additional drift term  $\frac{-\phi_i v_i L_i \bar{J}}{\sum_{k=1}^N \phi_k v_k L_k}$  proportional to the total material flux through the open

system. The term  $\frac{\bar{J}T}{\sum_{k=1}^N \phi_k v_k L_k}$  in Eq. (17) describes the contribution of that drift to the pressure

gradient. This additional component of the total material flux is attributed to barodiffusion, which is driven by the dynamic pressure gradient defined by Eq. (17). This dynamic pressure gradient is associated with viscous dissipation in the system. Parameter  $\bar{J}$  is independent of position in the system but is determined by material transfer across the system boundaries, which may vary over time.

If the system is open but stationary, molecules entering it through one of its boundary surfaces can leave it through another, thus creating a molecular drift that is independent of the existence of a temperature or pressure gradient. This drift is determined by conditions at the boundaries and is independent of any force applied to the system. For example, the system may have a component source at one boundary and a sink of the same component at opposite boundary. As molecules of a given species move between the two boundaries, they experience viscous friction, which creates a dynamic pressure gradient that induces barodiffusion in all molecular species. The pressure gradient that is induced by viscous friction in such a system is not considered in the Gibbs-Duhem equation.

Equations (6), (7), and (15) describe a system in hydrostatic equilibrium, without viscous friction caused by material flux due to material exchange through the system boundaries. Unlike the Gibbs-Duhem equation, Eq. (17) accounts for viscous friction forces and the resulting dynamic pressure gradient. For a closed stationary system, in which  $\bar{J} = 0$  and  $\frac{\partial \phi}{\partial t} = 0$ , Eq. (18) is transformed into

$$\sum_{k>1}^N \left( 2 \sum_{j=1}^N \frac{\partial \mu_{ij}^*}{\partial \phi_k} \right) \nabla \phi_k + \left( \sum_{j=1}^N \frac{\partial \mu_{ij}^*}{\partial T} \right) \nabla T = 0 \quad (19)$$

There are thermal diffusion experiments in which the system experiences periodic temperature changes. An example is the method used described by (Wiegand, Kohler, 2002), where thermodiffusion in liquids is observed within a dynamic temperature grating produced using a pulsed infrared laser. Because this technique involves changing the wall temperature, which changes the equilibrium adsorption constant, material fluxes vary with time, creating a periodicity in the inflow and outflow of material. A preliminary analysis shows that material fluxes to and from the walls have relaxation times on the order of a few microseconds until equilibrium is attained, and that such non-stationary material fluxes can be observed using dynamic temperature gratings.

The Soret coefficient is a common parameter used to characterize material transport in temperature gradients. For binary systems, Eq. (19) can be used to define the Soret coefficient as

$$S_T = - \frac{\frac{\partial \mu_{21P}^*}{\partial T}}{2\phi_2(1-\phi_2) \frac{\partial \mu_{21P}^*}{\partial \phi_2}} \quad (20)$$



where subscript  $P$  is used to indicate that the derivatives are taken at constant pressure, as is the case in Eqs. (4) and (6). We can solve Eqs. (19) to express the “partial” Soret coefficient  $S_T^k$  for the  $k$ 'th component through a factor of proportionality between  $\nabla\phi_k$  and  $\nabla T$ .

#### 4. Statistical mechanics of material transport: Chemical potentials at constant volume and pressure and the Laplace component of pressure in a molecular force field

The chemical potential at constant volume can be calculated using a modification of an expression derived in (Kirkwood, Boggs, 1942; Fisher, 1964):

$$\mu_{iV} = \mu_{0i} + \int_0^1 d\lambda \sum_{j=1}^N \frac{\phi_j}{v_j} \int_{V_{out}^i}^{\infty} g_{ij}(\vec{r}, \lambda) \Phi_{ij}(r) dv \quad (21)$$

Here

$$\mu_{0i} = -\frac{3}{2}kT \ln \frac{2\pi m_i kT}{h^2} + kT \ln \frac{\phi_i}{v_i} - kT \ln Z_{rot}^i - kT \ln Z_{vib}^i \quad (22)$$

is the chemical potential of an ideal gas of the respective non-interacting molecules (related to their kinetic energy),  $h$  is Planck's constant,  $m_i$  is the mass of the molecule,  $Z_{rot}^i$  and  $Z_{vib}^i$  are its rotational and vibrational statistical sums, respectively, and  $V_{out}^i$  is the volume external to a molecule of the  $i$ 'th component. The molecular vibrations make no significant contribution to the thermodynamic parameters except in special situations, which will not be discussed here. The rotational statistical sum for polyatomic molecules is written as (Landau, Lifshitz, 1980)

$$Z_{rot} = \frac{\sqrt{\pi}}{\gamma h^3} \sqrt{(8\pi^2 kT)^3 I_1 I_2 I_3} \quad (23)$$

where  $\gamma$  is the symmetry value, which is the number of possible rotations about the symmetry axes carrying the molecule into itself. For  $H_2O$  and  $C_2H_5OH$ ,  $\gamma = 2$ ; for  $NH_3$ ,  $\gamma = 3$ ; for  $CH_4$  and  $C_6H_6$ ,  $\gamma = 12$ .  $I_1, I_2$ , and  $I_3$  are the principal values of the tensor of the moment of inertia.

In Eq. (21), parameter  $\lambda$  describes the gradual “switching on” of the intermolecular interaction. A detailed description of this representation can be found in (Kirkwood, Boggs, 1942; Fisher, 1964). Parameter  $r$  is the distance between the molecule of the surrounding liquid and the center of the considered molecule;  $g_{ij}(\vec{r}, \lambda)$  is the pair correlative function, which expresses the probability of finding a molecule of the surrounding liquid at  $\vec{r}$  ( $r = |\vec{r}|$ ) if the considered molecule is placed at  $r = 0$ ; and  $\Phi_{ij}$  is the molecular interaction potential, known as the London potential (Ross, Morrison, 1988):

$$\Phi_{ij} = -\varepsilon_{ij} \left( \frac{\sigma_{ij}}{r} \right)^6 \quad (24)$$

Here  $\varepsilon_{ij}$  is the energy of interaction and  $\sigma_{ij}$  is the minimal molecular approach distance. In the integration over  $V_{out}^i$ , the lower limit is  $r = \sigma_{ij}$ . There is no satisfactory simple method for calculating the pair correlation function in liquids, although it should approach unity at infinity. We will approximate it as

$$g_{ij}(r, \lambda) = 1 \quad (25)$$

With this approximation we assume that the local distribution of solvent molecules is not disturbed by the particle under consideration. The approximation is used widely in the theory of liquids and its effectiveness has been shown. For example, in (Bringuier, Bourdon, 2003, 2007), it was used in a kinetic approach to define the thermodiffusion of colloidal particles. In (Schimpf, Semenov, 2004; Semenov, Schimpf, 2000, 2005) the approximation was used in a hydrodynamic theory to define thermodiffusion in polymer solutions. The approximation of constant local density is also used in the theory of regular solutions (Kirkwood, 1939). With this approximation we obtain

$$\mu_{iV} = \mu_{0i} + \sum_{j=1}^N \frac{\phi_j}{v_j} \int_{V_{out}^i}^{\infty} \Phi_{ij}(r) dv \quad (26)$$

The terms under the summation sign are a simple modification of the expression obtained in (Bringuier, Bourdon, 2003, 2007).

In our calculations, we will use the fact that there is certain symmetry between the chemical potentials contained in Eq. (11). The term  $\frac{v_i}{v_k} \mu_k$  can be written as  $N_{ik} \mu_k$ , where  $N_{ik} = \frac{v_i}{v_k}$  is the number of the molecules of the  $k$ 'th component that can be placed within the volume  $v_i$  but are displaced by a molecule of  $i$ 'th component. Using the known result that free energy is the sum of the chemical potentials we can say that  $N_{ik} \mu_k$  is the free energy or chemical potential of a virtual molecular particle consisting of molecules of the  $k$ 'th component displaced by a molecule of the  $i$ 'th component. For this reason we can extend the results obtained in the calculations of molecular chemical potential  $\mu_{iV}$  of the second component to calculations of parameter  $N_{ik} \mu_{kV}$  by a simple change in the respective designations  $i \rightarrow k$ . Regarding the concentration of these virtual particles, there are at least two approaches allowed:

- we can assume that the volume fraction of the virtual particles is equal to the volume fraction of the real particles that displace molecules of the  $k$ 'th component, i.e., their numeric concentration is  $\frac{\phi_i}{v_i}$ . This approach means that only the actually displaced molecules are taken into account, and that they are each distinguishable from molecules of the  $k$ 'th component in the surrounding liquid.
- we can take into account the indistinguishability of the virtual particles. In this approach any group of the  $N_{ik}$  molecules of the  $k$ 'th component can be considered as a virtual particle. In this case, the numeric volume concentration of these virtual molecules is  $\frac{\phi_k}{v_i}$ .

We have chosen to use the more general assumption b).

Using Eqs. (21) and (22), along with the definition of a virtual particle outlined above, we can define the combined chemical potential at constant volume  $\mu_{ikV}^*$  as

$$\mu_{ikV}^* = +kT \left( \frac{3}{2} \ln \frac{m_i}{m_{N_{ik}}} + \ln \frac{\phi_i}{\phi_k} + \ln \frac{Z_{rot}^i}{Z_{rot}^{N_{ik}}} \right) + \sum_{j=1}^N \frac{\phi_j}{v_j} \int_{V_{out}^i}^{\infty} \Phi_{ij}(r) dv - \sum_{j=1}^N \frac{\phi_j}{v_j} \int_{V_{out}^i}^{\infty} \Phi_{kj}^{N_{ik}}(r) dv \quad (27)$$

where  $m_{N_{ik}} = m_k N_{ik}$  and  $Z_{rot}^{N_{ik}}$  are the mass and the rotational statistical sum of the virtual particle, respectively. In Eq. (27), the total interaction potential  $N_{ik} \Phi_{kj}$  of the molecules included in the virtual particle is written as  $\Phi_j^{N_{ik}}$ . We will use the approximation

$$\Phi_j^{N_{ik}} = N_{ik} \Phi_{kj} = -\varepsilon_{kj} \left( \frac{\sigma_{ij}}{r} \right)^6 \quad (28)$$

This approximation corresponds to the virtual particle having the size of a molecule of the  $i$ 'th component and the energetic parameter of the  $k$ 'th component.

In further development of the microscopic calculations it is important that the chemical potential be defined at constant pressure. Chemical potentials at constant pressure are related to those at constant volume  $\mu_{iV}$  by the expression

$$\nabla \mu_{ip} = \nabla \mu_{iV} + \int_{V_{out}^i} \nabla \Pi_i dv \quad (29)$$

Here  $\Pi_i$  is the local pressure distribution around the molecule. Eq. (29) expresses the relation between the forces acting on a molecular particle at constant versus changing local pressure. This equation is a simple generalization of a known equation (Haase, 1969) in which the pressure gradient is assumed to be constant along a length about the particle size.

Next we calculate the local pressure distribution  $\Pi_i$ , which is widely used in hydrodynamic models of kinetic effects in liquids (Ruckenstein, 1981; Anderson, 1989; Schimpf, Semenov, 2004; Semenov, Schimpf, 2000, 2005). The local pressure distribution is usually obtained from the condition of the local mechanical equilibrium in the liquid around  $i$ 'th molecular

particle, a condition that is written as  $\nabla \left[ \Pi_i + \sum_{j=1}^N \frac{\phi_j}{v_j} \Phi_{ij}(r) \right] = 0$ . In (Semenov, Schimpf, 2009;

Semenov, 2010) the local pressure distribution is used in a thermodynamic approach, where it is obtained by formulating the condition for establishing local equilibrium in a thin layer of thickness  $l$  and area  $S$  when the layer shifts from position  $r$  to position  $r+dr$ . In this case, local equilibrium expresses the local conservation of specific free energy

$$F_i(r) = \Pi_i(r) + \sum_{j=1}^N \frac{\phi_j}{v_j} \Phi_{ij}(r) \quad \text{in such a shift when the isothermal system is placed in a force}$$

field of the  $i$ 'th molecule.

In the layer forming a closed surface, the change in the free energy is written as:

$$dF_i(r) = \nabla \left[ \sum_{j=1}^N \frac{\phi_j}{v_j} \Phi_{ij}(r) + \nabla \Pi_i \right] l S dr + \sum_{j=1}^N \left[ \frac{\phi_j}{v_j} \Phi_{ij}(r) \right] dS = 0 \quad (30)$$

where we consider changes in free energy due to both a change in the parameters of the layer volume ( $dV = Sdr$ ) and a change  $dS$  in the area of the closed layer. For a spherical layer, the changes in volume and surface area are related as  $dV = 2rdS$ , and we obtain the following modified equation of equilibrium for a closed spherical surface:

$$\nabla \left[ \sum_{j=1}^N \frac{\phi_j}{v_j} \Phi_{ij}(r) + \Pi_i \right] + 2 \sum_{j=1}^N \frac{\phi_j}{v_j} \frac{\Phi_{ij}(r)}{r} \bar{r}_0 = 0 \quad (31)$$

where  $\bar{r}_0$  is the unit radial vector. The pressure gradient related to the change in surface area has the same nature as the Laplace pressure gradient discussed in (Landau, Lifshitz, 1980). Solving Eq. (31), we obtain

$$\Pi_i = - \sum_{j=1}^N \frac{\phi_j}{v_j} \left[ \Phi_{ij}(r) + \int_{\infty}^r \frac{2\Phi_{ij}(r')}{r'} dr' \right] \quad (32)$$

Substituting the pressure gradient from Eq. (32) into Eq. (29), and using Eqs. (24), (27), and (28), we obtain a general expression for the gradient in chemical potential at constant pressure in a non-isothermal and non-homogeneous system. We will not write the general expression here, rather we will derive the expression for binary systems.

## 5. The Soret coefficient in diluted binary molecular mixtures: The kinetic term in thermodiffusion is related to the difference in the mass and symmetry of molecules

In this section we present the results obtained in (Semenov, 2010, Semenov, Schimpf, 2011a). In diluted systems, the concentration dependence of the chemical potentials for the solute and solvent is well-known [e.g., see (Landau, Lifshitz, 1980)]:  $\mu_2(\phi) = kT \ln \phi$ , and  $\mu_1$  is practically independent of solute concentration  $\phi = \phi_2$ . Thus, Eq. (20) for the Soret coefficient takes the form:

$$S_T = - \frac{\partial \mu_p^*}{2kT} \quad (33)$$

where  $\mu_p^*$  is  $\mu_{21p}^*$ .

The equation for combined chemical potential at constant volume [Eq. (28)] using assumption b) in Section 3 takes the form

$$\mu_V^* = -kT \left( \frac{3}{2} \ln \frac{m_2}{m_{N_1}} - \ln \frac{\phi}{1-\phi} + \ln \frac{Z_{rot}^1}{Z_{rot}^{N_1}} \right) + 4\pi \int_R^{\infty} \frac{\Phi_{21}(r) - \Phi_{11}^{N_1}(r)}{v_1} r^2 dr \quad (34)$$

where  $N_1 = N_{21}$  is the number of solvent molecules displaced by molecule of the solute,  $\Phi_{11}^{N_1}$  is the potential of interaction between the virtual particle and a molecule of the solvent. The relation  $\phi_1 = 1 - \phi$  is also used in deriving Eq. (34). Because  $\ln[\phi/(1-\phi)] \rightarrow -\infty$  at  $\phi \rightarrow 0$ , we expect the use of assumption a) in Section 3 for the concentration of virtual particles will yield a reasonable physical result.

In a dilute binary mixture, the equation for local pressure [Eq. (32)] takes the form

$$\Pi_i = -\sum_{j=1}^N \frac{\Phi_{ij}(r)}{v_1} + \int_{\infty}^r \frac{2\Phi_{i1}(r')}{v_1 r'} dr' \quad (35)$$

where index  $i$  is related to the virtual particle or solute.

Using Eqs. (29), (34), we obtain the following expression for the temperature-induced gradient of the combined chemical potential of the diluted molecular mixture:

$$\nabla \mu_p^* = -k\nabla T \left[ \frac{3}{2} \ln \frac{m_2}{m_{N_1}} + \ln \frac{Z_{rot}^1}{Z_{rot}^{N_1}} \right] + \int_{v_{out}} \frac{\alpha_1 dv}{v_1} \nabla_g T \int_{\infty}^r \frac{\Phi_{21}(r') - \Phi_{11}^{N_1}(r')}{r'} dr' \quad (36)$$

Here  $\alpha_1$  is the thermal expansion coefficient for the solvent and  $\nabla_g T$  is the tangential component of the bulk temperature gradient. After substituting the expressions for the interaction potentials defined by Eqs. (23), (24), and (28) into Eq. (36), we obtain the following expression for the Soret coefficient in the diluted binary system:

$$S_T = \frac{1}{2T} \left[ \frac{3}{2} \ln \left( \frac{m_2}{m_{N_1}} \right) + \ln \left( \frac{\gamma_{N_1} \sqrt{(I_1 I_2 I_3)_2}}{\gamma_2 \sqrt{(I_1 I_2 I_3)_{N_1}}} \right) \right] + \frac{\pi^2 \alpha_1 \sigma_{12}^3 \varepsilon_{12}}{18 v_1 k T} \left( \frac{\varepsilon_{11}}{\varepsilon_{12}} - 1 \right) \quad (37)$$

In Eq. (37), the subscripts 2 and  $N_1$  are used again to denote the real and virtual particle, respectively.

The Soret coefficient expressed by Eq. (37) contains two main terms. The first term corresponds to the temperature derivative of the part of the chemical potential related to the solute kinetic energy. In turn, this kinetic term contains the contributions related to the translational and rotational movements of the solute in the solvent. The second term is related to the potential interaction of solute with solvent molecules. This potential term has the same structure as those obtained by the hydrodynamic approach in (Schimpf, Semenov, 2004; Semenov, Schimpf, 2005).

According to Eq. (37), both positive (from hot to cold wall) and negative (from cold to hot wall) thermodiffusion is possible. The molecules with larger mass ( $m_2 > m_{N_1}$ ) and with a stronger interactions between solvent molecules ( $\varepsilon_{11} > \varepsilon_{12}$ ) should demonstrate positive thermodiffusion. Thus, dilute aqueous solutions are expected to demonstrate positive thermophoresis. In (Ning, Wiegand, 2006), dilute aqueous solutions of acetone and dimethyl sulfoxide were shown to undergo positive thermophoresis. In that paper, a very high value of the Hildebrand parameter is given as an indication of the strong intermolecular interaction for water. More specifically, the value of the Hildebrand parameter exceeds by two-fold the respective parameters for other components.

Since the kinetic term in the Soret coefficient contains solute and solvent symmetry numbers, Eq. (37) predicts thermodiffusion in mixtures where the components are distinct only in symmetry, while being identical in respect to all other parameters. In (Wittko, Köhler, 2005) it was shown that the Soret coefficient in the binary mixtures containing the isotopically substituted cyclohexane can be in general approximated as the linear function

$$S_T = S_{iT} + a_m \Delta M + b_i \Delta I \quad (38)$$

where  $S_{iT}$  is the contribution of the intermolecular interactions,  $a_m$  and  $b_i$  are coefficients, while  $\Delta M$  and  $\Delta I$  are differences in the mass and moment of inertia, respectively, for the molecules constituting the binary mixture. According to Eq. (37), the coefficients are defined by

$$a_m = \frac{3}{4Tm_{N_1}} \quad (39)$$

$$b_i = \frac{(\gamma_{N_1})^2}{4T(\gamma_2)^2 \sqrt{(I_1 I_2 I_3)_{N_1}}} \quad (40)$$

In (Wittko, Köhler, 2005) the first coefficient was empirically determined for cyclohexane isomers to be  $a_m = 0.99 \cdot 10^{-3} K^{-1}$  at room temperature ( $T=300$  K), while Eq. (39) yields  $a_m = 0.03 \cdot 10^{-3} K^{-1}$  ( $M_1 = 84$ ). There are several possible reasons for this discrepancy. The first term on the right side of Eq. (38) is not the only term with a mass dependence, as the second term also depends on mass. The empirical parameter  $a_m$  also has an implicit dependence on mass that is not in the theoretical expression given by Eq. (39). The mass dependence of the second term in Eq. (37) will be much stronger when a change in mass occurs at the periphery of the molecule.

A sharp change in molecular symmetry upon isotopic substitution could also lead to a discrepancy between theory and experiment. Cyclohexane studied in (Wittko, Köhler, 2005) has high symmetry, as it can be carried into itself by six rotations about the axis perpendicular to the plane of the carbon ring and by two rotations around the axes placed in the plane of the ring and perpendicular to each other. Thus, cyclohexane has  $\gamma_{N_1} = 24$ . The partial isotopic substitution breaks this symmetry. We can start from the assumption that for the substituted molecules,  $\gamma_2 = 1$ . When the molecular geometry is not changed in the substitution and only the momentum of inertia related to the axis perpendicular to the ring plane is changed, the relative change in parameter  $b_i$  can be written as

$$\frac{(\gamma_{N_1})^2 \sqrt{(I_1 I_2 I_3)_2} - (\gamma_2)^2 (I_1 I_2 I_3)_1}{4T(\gamma_2)^2 \sqrt{(I_1 I_2 I_3)_{N_1}}} = \frac{(\gamma_{N_1})^2 (m_2 - m_{N_1})}{4T(\gamma_2)^2 m_{N_1}} + \frac{(\gamma_{N_1})^2 - (\gamma_2)^2}{4T(\gamma_2)^2} \quad (41)$$

Eq. (41) yields

$$a_m = \frac{1}{4Tm_{N_1}} \left[ 3 + \left( \frac{\gamma_{N_1}}{\gamma_2} \right)^2 \right] \quad (42)$$

Using the above parameters and Eq. (42), we obtain  $a_m \approx 5.7 \cdot 10^{-3} K^{-1}$ , which is still about six-times greater than the empirical value from (Wittko, Köhler, 2005). The remaining discrepancy could be due to our overestimation of the degree of symmetry violation upon isotopic substitution. The true value of this parameter can be obtained with  $\gamma_2 \approx 2-3$ . One should understand that the value of parameter  $\gamma_2$  is to some extent conditional because the isotopic substitutions occur at random positions. Thus, it may be more relevant to use Eq. (42) to evaluate the characteristic degree of symmetry from an experimental measurement of  $a_m$  rather than trying to use theoretical values to predict thermodiffusion.

## 6. The Soret coefficient in diluted colloidal suspensions: Size dependence of the Soret coefficient and the applicability of thermodynamics

While thermodynamic approaches yield simple and clear expressions for the Soret coefficient, such approaches are the subject of rigorous debate. The thermodynamic or “energetic” approach has been criticized in the literature. Parola and Piazza (2004) note that the Soret coefficient obtained by thermodynamics should be proportional to a linear combination of the surface area and the volume of the particle, since it contains the parameter  $\mu_{ik}^*$  given by Eq. (11). They argue that empirical evidence indicates the Soret coefficient is directly proportional to particle size for colloidal particles [see numerous references in (Parola, Piazza, 2004)], and is practically independent of particle size for molecular species. By contrast, Duhr and Braun (2006) show the proportionality between the Soret coefficient and particle surface area, and use thermodynamics to explain their empirical data. Dhont et al (2007) also reports a Soret coefficient proportional to the square of the particle radius, as calculated by a quasi-thermodynamic method.

Let us consider the situation in which a thermodynamic calculation for a large particle as said contradicts the empirical data. For large particles, the total interaction potential is assumed to be the sum of the individual potentials for the atoms or molecules which are contained in the particle

$$\Phi_{i1}^*(r) = \int_{V_{in}^i} \frac{dV_{in}}{v_i} \Phi_{i1}(|\vec{r}_i - \vec{r}|) \quad (43)$$

Here  $V_{in}^i$  is the internal volume of the real or virtual particle and  $\Phi_{i1}(|\vec{r}_i - \vec{r}|)$  is the respective intermolecular potential given by Eq. (24) or (28) for the interaction between a molecule of a liquid placed at  $\vec{r}$  ( $r = |\vec{r}|$ ) and an internal molecule or atom placed at  $\vec{r}_i$ . Such potentials are referred to as Hamaker potential, and are used in studies of interactions between colloidal particles (Hunter, 1992; Ross, Morrison, 1988). In this and the following sections,  $v_i$  is the specific molecular volume of the atom or molecule in a real or virtual particle, respectively.

For a colloidal particle with radius  $R \gg \sigma_{ij}$ , the temperature distribution at the particle surface can be used instead of the bulk temperature gradient (Giddings et al, 1995), and the curvature of the particle surface can be ignored in calculating the respective integrals. This corresponds to the assumption that  $r' \approx R$  and  $dv \approx 4\pi R^2 dr$  in Eq. (36). To calculate the Hamaker potential, the expression calculated in (Ross, Morrison, 1988), which is based on the London potential, can be used:

$$\Phi_{i1}^*(y) = -\frac{\varepsilon_{i1}}{6} \frac{\sigma_{21}^3}{v_2} \left( \frac{1}{y} + \frac{1}{2+y} + \ln \frac{y}{2+y} \right) \quad (44)$$

Here  $y = \frac{x}{\sigma_{21}}$ , and  $x$  is the distance from the particle surface to the closest solvent molecule surface. Using Eqs. (36) and (44) we can obtain the following expression for the Soret coefficient of a colloidal particle:

$$S_T = \frac{\pi^2 \alpha_1 R \sigma_{21}^2 \varepsilon_{21}}{2(n+2)v_2 kT} \frac{\sigma_{21}^3}{v_1} \left( \frac{\varepsilon_{11}}{\varepsilon_{21}} - 1 \right) \quad (45)$$

Here  $n$  is ratio of particle to solvent thermal conductivity. The Soret coefficient for the colloidal particle is proportional to  $\frac{R\sigma_{21}^5}{v_1 v_2}$ . In practice, this means that  $S_T$  is proportional to

$\frac{R}{\sigma_{21}}$  since the ratio  $\frac{\sigma_{21}^6}{v_1 v_2}$  is practically independent of molecular size. This proportionality is consistent with hydrodynamic theory [e.g., see (Anderson, 1989)], as well as with empirical data. The present theory explains also why the contribution of the kinetic term and the isotope effect has been observed only in molecular systems. In colloidal systems the potential related to intermolecular interactions is the prevailing factor due to the large value of  $\frac{R\sigma_{21}^2}{v_1}$ . Thus, the colloidal Soret coefficient is  $\frac{R}{\sigma_{21}}$  times larger than its molecular counterpart. This result is also consistent with numerous experimental data and with hydrodynamic theory.

## 7. The Soret coefficient in diluted suspensions of charged particles: Contribution of electrostatic and non-electrostatic interactions to thermodiffusion

In this section we present the results obtained in (Semenov, Schimpf, 2011b). The colloidal particles discussed in the previous section are usually stabilized in suspensions by electrostatic interactions. Salt added to the suspension becomes dissociated into ions of opposite electric charge. These ions are adsorbed onto the particle surface and lead to the establishment of an electrostatic charge, giving the particle an electric potential. A diffuse layer of charge is established around the particle, in which counter-ions are accumulated. This diffuse layer is the electric double layer. The electric double layer, where an additional pressure is present, can contribute to thermodiffusion. It was shown in experiments that particle thermodiffusion is enhanced several times by the addition of salt [see citations in (Dhont, 2004)].

For a system of charged colloidal particles and molecular ions, the thermodynamic equations should be modified to include the respective electrostatic parameters. The basic thermodynamic equations, Eqs. (4) and (6), can be written as

$$\nabla \mu_i = \sum_{k=1}^N \frac{\partial \mu_i}{\partial n_k} \nabla n_k - \bar{v}_i \nabla P + \frac{\partial \mu_i}{\partial T} \nabla T + e_i \bar{E} \quad (46)$$



$$\nabla P = \sum_{i=1}^N n_i \left( \sum_{k=1}^N \frac{\partial \mu_i}{\partial n_k} \nabla n_k + \frac{\partial \mu_i}{\partial T} \nabla T + e_i \bar{E} \right) \quad (47)$$

where  $e_i = -\frac{\partial \mu_i}{\partial \Phi}$  is the electric charge of the respective ion,  $\Phi$  is the macroscopic electrical potential, and  $\bar{E} = -\nabla \Phi$  is the electric field strength. Substituting Eq. (47) into Eq. (46) we obtain the following material transport equations for a closed and stationary system:

$$\bar{J}_i = 0 = -\frac{L_i}{T} \sum_k^N \frac{\phi_i \phi_k}{v_i} \left( \sum_{l=1}^N \frac{\partial \mu_{ik}^*}{\partial \phi_l} \nabla \phi_l + \frac{\partial \mu_{ik}^*}{\partial T} \nabla T - \frac{\partial \mu_{ik}^*}{\partial \Phi} \bar{E} \right) \quad (48)$$

where

$$-\frac{\partial \mu_{ik}^*}{\partial \Phi} = e_i - N_{ik} e_k \quad (49)$$

We will consider a quaternary diluted system that contains a background neutral solvent with concentration  $\phi_1$ , an electrolyte salt dissociated into ions with concentrations  $\phi_{\pm} = n_{\pm} v_{\pm}$ , and charged particles with concentration  $\phi_2$  that is so small that it makes no contribution to the physicochemical parameters of the system. In other words, we consider the thermophoresis of an isolated charged colloidal particle stabilized by an ionic surfactant. With a symmetric electrolyte, the ion concentrations are equal to maintain electric neutrality

$$v_- \phi_+ = v_+ \phi_- \quad (50)$$

In this case we can introduce the volume concentration of salt as  $\phi_s = \phi_+ \left( 1 + \frac{v_-}{v_+} \right) = \phi_- \left( 1 + \frac{v_+}{v_-} \right)$  and formulate an approximate relationship in place of the exact form expressed by Eq. (8):

$$\phi_s + \phi_1 = 1 \quad (51)$$

Here the volume contribution of charged particles is ignored since their concentration is very low, i.e.  $\phi_2 \ll \phi_s \ll \phi_1$ . Due to electric neutrality, the ion concentrations will be equal at any salt concentration and temperature, that is, the chemical potentials of the ions should be equal:  $\mu_+ = \mu_-$  (Landau, Lifshitz, 1980).

Using Eqs. (48) - (51) we obtain equations for the material fluxes, which are set to zero:

$$\bar{J}_2 = 0 = -\frac{\phi_2 L_2}{v_2 T} \left[ \frac{\partial \mu_{21}^*}{\partial \phi_2} \nabla \phi_2 + 3 \frac{\partial \mu_{21}^*}{\partial \phi_s} \nabla \phi_s + \frac{\partial \mu_{21}^*}{\partial T} \nabla T + e_2 \bar{E} \right] \quad (52)$$

$$\bar{J}_- = 0 = -\frac{\phi_- L_-}{v_- T} \left( 3 \frac{\partial \mu_{-1}^*}{\partial \phi_s} \nabla \phi_s + \frac{\partial \mu_{-1}^*}{\partial T} \nabla T - e \bar{E} \right) \quad (53)$$

$$\bar{J}_+ = 0 = -\frac{\phi_+ L_+}{v_+ T} \left( 3 \frac{\partial \mu_{+1}^*}{\partial \phi_s} \nabla \phi_s + \frac{\partial \mu_{+1}^*}{\partial T} \nabla T + e \bar{E} \right) \quad (54)$$

where  $e_+ = -e_- = e$  (symmetric electrolyte). We will not write the equation for the flux of background solvent  $\bar{J}_1$  because it yields no new information in comparison with Eqs. (52) - 54), as shown above. Solving Eqs. (52) - (54), we obtain

$$\nabla \phi_s = -\nabla T \frac{\partial(\mu_{+1}^* + \mu_{-1}^*)}{\partial T} \Big/ 3 \frac{\partial(\mu_{+1}^* + \mu_{-1}^*)}{\partial \phi_s} \quad (55)$$

$$2e\bar{E} = 3 \frac{\partial(\mu_{-1}^* - \mu_{+1}^*)}{\partial \phi_s} \nabla \phi_s + \frac{\partial(\mu_{-1}^* - \mu_{+1}^*)}{\partial T} \nabla T \quad (56)$$

Eq. (55) allows us to numerically evaluate the concentration gradient as

$$\nabla \phi_s \approx \phi_s S_T^s \nabla T \quad (57)$$

where  $S_T^s \approx 10^{-3}$  is the characteristic Soret coefficient for the salts. Salt concentrations are typically around  $10^{-2}$ - $10^{-1}$  mol/L, that is  $\phi_s \approx 10^{-4}$  or lower. A typical maximum temperature gradient is  $\nabla T \approx 10^4$  K/cm. These values substituted into Eq. (57) yield  $\nabla \phi_s \approx 10^{-4} - 10^{-3}$  cm<sup>-1</sup>. The same evaluation applied to parameters in Eq. (56) shows that the first term on the right side of this equation is negligible, and the equation for thermoelectric power can be written as

$$\bar{E} \approx \frac{\partial(\mu_{-1}^* - \mu_{+1}^*)}{\partial T} \frac{\nabla T}{2e} = \frac{v_+ - v_-}{2ev_1} \frac{\partial \mu_1}{\partial T} \nabla T \quad (58)$$

For a non-electrolyte background solvent, parameter  $\partial \mu_1 / \partial T$  can be evaluated as  $\partial \mu_1 / \partial T \approx \alpha_1 kT$ , where  $\alpha_1$  is the thermal expansion coefficient of the solvent (Semenov, Schimpf, 2009; Semenov, 2010). Usually, in liquids the thermal expansion coefficient is low enough ( $\alpha_1 \approx 10^{-3}$  K<sup>-1</sup>) that the thermoelectric field strength does not exceed 1 V/cm. This electric field strength corresponds to the maximum temperature gradient discussed above. The electrophoretic velocity in such a field will be about  $10^{-5}$ - $10^{-4}$  cm/s. The thermophoretic velocities in such temperature gradients are usually at least one or two orders of magnitude higher.

These evaluations show that temperature-induced diffusion and electrophoresis of charged colloidal particle in a temperature gradient can be ignored, so that the expression for the Soret coefficient of a diluted suspension of such particles can be written as

$$S_{2T} = -\frac{\nabla \phi_2}{\phi_2 \nabla T} = -\frac{\frac{\partial \mu_{21P}^*}{\partial T}}{\phi_2 \frac{\partial \mu_{21P}^*}{\partial \phi_2}} = -\frac{1}{kT} \frac{\partial \mu_{21P}^*}{\partial T} \quad (59)$$

Eq. (59) can also be used for microscopic calculations.

For an isolated particle placed in a liquid, the chemical potential at constant volume can be calculated using a modified procedure mentioned in the preceding section. In these calculations, we use both the Hamaker potential and the electrostatic potential of the electric double layer to account for the two types of the interactions in these systems. The chemical potential of the non-interacting molecules plays no role for colloid particles, as was shown above.

In a salt solution, the suspended particle interacts with both solvent molecules and dissolved ions. The two interactions can be described separately, as the salt concentration is usually very low and does not significantly change the solvent density. The first type of interaction uses Eqs. (25) and the Hamaker potential [Eq. (44)].

For the electrostatic interactions, the properties of diluted systems may be used, in which the pair correlative function has a Boltzmann form (Fisher, 1964; Hunter, 1992). Since there are two kinds of ions, Eq. (21) for the "electrostatic" part of the chemical potential at constant volume can be written as

$$\mu_2^e = -4\pi n_s \int_0^\infty d\lambda \int_R^\infty \left( e^{\frac{\lambda\Phi_e}{kT}} - e^{-\frac{\lambda\Phi_e}{kT}} \right) \Phi_e(r) r^2 dr = -4\pi n_s kT \int_R^\infty \left( e^{\frac{\Phi_e}{kT}} + e^{-\frac{\Phi_e}{kT}} - 2 \right) r^2 dr \quad (60)$$

where  $n_s = \frac{\phi_s}{v_+ + v_-}$  is the numeric volume concentration of salt, and  $\Phi_e = e\Phi$  is the electrostatic interaction energy.

Eq. (32) expressing the equilibrium condition for electrostatic interactions is written as

$$\nabla \left[ (n_+ - n_-) \Phi_e(r) + \Pi \right] + 2(n_+ - n_-) \Phi_e(r) \frac{\vec{r}_0}{R} = 0 \quad (61)$$

where  $\vec{r}_0$  is the unit radial vector. In Eq. (61) it is assumed that the particle radius is much larger than the characteristic thickness of the electric double layer. Solving Eq. (62) assuming a Boltzmann distribution for the ion concentration, as in (Ruckenstein, 1981; Anderson, 1989), we obtain

$$\Pi_e = n_s kT \left( e^{\frac{\Phi_e}{kT}} + e^{-\frac{\Phi_e}{kT}} - 2 \right) - \frac{2n_s}{R} \int_\infty^r \left( e^{\frac{\Phi_e}{kT}} - e^{-\frac{\Phi_e}{kT}} \right) \int_\infty^{r'} \Phi_e(r'') dr'' dr' \quad (62)$$

Substituting the pressure gradient calculated from Eq. (62) into Eq. (29), utilizing Eq. (60), and considering the temperature-induced gradients related to the temperature dependence of the Boltzmann exponents, we obtain the temperature derivative in the gradient of the chemical potential for a charged colloidal particle, which is related to the electrostatic interactions in its electric double layer:

$$\frac{\partial \mu_{2P}^e}{\partial T} = \frac{4\pi n_s kR}{(n+2)} \int_R^\infty dr \int_\infty^r \left( e^{\frac{\Phi_e}{kT}} + e^{-\frac{\Phi_e}{kT}} \right) \frac{\Phi_e^2(r')}{(kT)^2} dr' \quad (63)$$

Here  $n$  is again the ratio of particle to solvent thermal conductivity. For low potentials ( $\Phi_e < kT$ ), where the Debye-Hueckel theory should work, Eq. (63) takes the form

$$\frac{\partial \mu_{2P}^e}{\partial T} = \frac{8\pi n_s kR}{(n+2)} \int_R^\infty dr \int_\infty^r \frac{\Phi_e^2(r')}{(kT)^2} dr' \quad (64)$$

Using an exponential distribution for the electric double layer potential, which is characteristic for low electrokinetic potentials  $\zeta$ , we obtain from Eq. (64)

$$\frac{\partial \mu_{2P}^e}{\partial T} = \frac{8\pi n_s kR \lambda_D^2}{(n+2)} \left( \frac{e\zeta}{kT} \right)^2 \quad (65)$$

where  $\lambda_D$  is the Debye length [for a definition of Debye length, see (Landau, Lifshitz, 1980; Hunter, 1992)].

Calculation of the non-electrostatic (Hamaker) term in the thermodynamic expression for the Soret coefficient is carried out in the preceding section [Eq. (45)]. Combining this expression with Eq. (65), we obtain the Soret coefficient of an isolated charged colloidal particle in an electrolyte solution:

$$S_T = \frac{8\pi n_s R \lambda_D^2}{T(n+2)} \left( \frac{e\zeta}{kT} \right)^2 + \frac{\pi^2 \alpha_1 R \sigma_{21}^2 \varepsilon_{21} \sigma_{21}^3}{2(n+2)v_2 kT v_1} \left( \frac{\varepsilon_{11}}{\varepsilon_{21}} - 1 \right) \quad (66)$$

This thermodynamic expression for the Soret coefficient contains terms related to the electrostatic and Hamaker interactions of the suspended colloidal particle. The electrostatic term has the same structure as the respective expressions for the Soret coefficient obtained by other methods (Ruckenstein, 1981; Anderson, 1989; Parola, Piazza, 2004; Dhont, 2004). In the Hamaker term, the last term in the brackets reflects the effects related to displacing the solvent by particle. It is this effect that can cause a change in the direction of thermophoresis when the solvent is changed. However, such a reverse in the direction of thermophoresis can only occur when the electrostatic interactions are relatively weak. When electrostatic interactions prevail, only positive thermophoresis can be observed, as the displaced solvent molecules are not charged, therefore, the respective electrostatic term is zero. The numerous theoretical results on electrostatic contributions leading to a change in the direction of thermophoresis are wrong due to an incorrect use of the principle of local equilibrium in the hydrodynamic approach [see discussion in (Semenov, Schimpf, 2005)].

The relative role of the electrostatic mechanism can be evaluated by the following ratio:

$$\frac{8 n_s v_2 \lambda_D^2 v_1}{\pi \alpha_1 T \sigma_{21}^2 \sigma_{21}^3} \frac{(e\zeta)^2}{(\varepsilon_{11} - \varepsilon_{21}) kT} \quad (67)$$

The physicochemical parameters contained in Eq. (67) are separated into several groups and are collected in the respective coefficients. Coefficient  $\frac{n_s v_2}{\alpha_1 T}$  contains the parameters related to concentration and its change with temperature,  $\frac{\lambda_D^2}{\sigma_{21}^2}$  is the coefficient reflecting the respective lengths of the interaction,  $\frac{v_1}{\sigma_{21}^3}$  reflects the geometry of the solvent molecules, and

$\frac{(e\zeta)^2}{(\varepsilon_{11} - \varepsilon_{21})kT}$  is the ratio of energetic parameters for the respective interactions. Only the first two of these four terms are always significantly distinct from unity. The characteristic length of the interaction is much higher for electrostatic interactions. Also, the characteristic density of ions or molecules in a liquid, which are involved in their electrostatic interaction with the colloidal particle, is much lower than the density of the solvent molecules. The values of these respective coefficients are  $\frac{\lambda_D^2}{\sigma_{21}^2} \geq 10^3$  and  $\frac{n_s v_2}{\alpha_1 T} \approx 10^{-3}$  for typical ion concentrations in water at room temperature. The energetic parameter may be small, ( $\sim 0.1$ ) when the colloidal particles are compatible with the solvent. Characteristic values of the energetic coefficient range from  $0.1-10$ . Combining these numeric values, one can see that the ratio given by Eq. (67) lies in a range of  $0.1-10$  and is governed primarily by the value of the electrokinetic potential  $\zeta$  and the difference in the energetic parameters of the Hamaker interaction  $\varepsilon_{11} - \varepsilon_{21}$ . Thus, calculation of the ratio given by Eq. (67) shows that either the electrostatic or the Hamaker contribution to particle thermophoresis may prevail, depending on the value of the particle's energetic parameters. In the region of high Soret coefficients, particle thermophoresis is determined by electrostatic interactions and is positive. In the region of low Soret coefficients, thermophoresis is related to Hamaker interactions and can have different directions in different solvents.

## 8. Material transport equation in binary molecular mixtures: Concentration dependence of the Soret coefficient

In this section we present the results obtained in (Semenov, 2011). In a binary system in which the component concentrations are comparable, the material transport equations defined by Eq. (18) have the form

$$\frac{\partial \phi}{\partial t} = \nabla \left[ L_2 \phi (1 - \phi) \left( 2 \frac{\partial \mu^*}{\partial \phi} \nabla \phi + \frac{\partial \mu^*}{\partial T} \nabla T \right) \right] / T \left( 1 - \phi + \frac{L_2 v_2}{L_1 v_1} \phi \right) \quad (68)$$

Eq. (68) can be used in the thermodynamical definition of the Soret coefficient [Eq. (59)]. The mass and thermodiffusion coefficients can be calculated in the same way as the Soret coefficient. The microscopic models used to calculate the Soret Coefficient in (Ghorayeb, Firoozabadi, 2000; Pan S et al., 2007) ignore the requirement expressed by Eq. (10) and cannot yield a description of thermodiffusion that is unambiguous. Although the material transport equations based on non-equilibrium thermodynamics were used, the fact that the chemical potential at constant pressure must be used was not taken into account. In these articles there is also the problem that in the transition to a dilute system the entropy of mixing does not become zero, yielding unacceptably large Soret coefficients even for pure components. An expression for the Soret coefficient was obtained in (Dhont et al, 2007; Dhont, 2004) by a quasi-thermodynamic method. However, the expressions for the thermodiffusion coefficient in those works become zero at high dilution, where the standard expression for osmotic pressure is used. These results contradict empirical observation.

Using Eq. (27) with the notion of a virtual particle outlined above, and substituting the expression for interaction potential [Eqs. (24, 28)], we can write the combined chemical potential at constant volume  $\mu_v^*$  as

$$\begin{aligned} \mu_v^* = & -kT \left( \frac{3}{2} \ln \frac{m_2}{m_{N_1}} - \ln \frac{\phi}{1-\phi} + \ln \frac{Z_{rot}^2}{Z_{rot}^{N_1}} \right) + \\ & + \frac{\phi}{v_2} \left[ \int_{V_{out}^2}^{\infty} \Phi_{22}(r) dv - \int_{V_{out}^1}^{\infty} \Phi_{12}^{N_1}(r) dv \right] + \frac{1-\phi}{v_1} \left[ \int_{V_{out}^2}^{\infty} \Phi_{21}(r) dv - \int_{V_{out}^1}^{\infty} \Phi_{11}^{N_1}(r) dv \right] \end{aligned} \quad (69)$$

In order to proceed to the calculation of chemical potentials at constant pressure using Eq. (29), we must calculate the local pressure distribution  $\Pi_i$  using Eq. (32). We can subsequently use Eqs. (29) and (33) to obtain an expression for the gradient of the combined chemical potential at constant pressure in a non-isothermal and non-homogeneous system:

$$\begin{aligned} \nabla \mu_p^* = & \left[ \frac{kT}{\phi(1-\phi)} - a \left( \frac{\varepsilon_{11} + \beta \varepsilon_{22}}{\varepsilon_{12}} - 1 - \beta \right) \right] \nabla \phi + \\ & + a \left[ \alpha_2 \beta \phi \left( 1 - \frac{\varepsilon_{22}}{\varepsilon_{12}} \right) - \alpha_1 (1-\phi) \left( 1 - \frac{\varepsilon_{11}}{\varepsilon_{12}} \right) \right] \nabla T - \\ & - k \left( \frac{3}{2} \ln \frac{m_2}{m_{N_1}} - \ln \frac{\phi}{1-\phi} + \ln \frac{Z_{rot}^2}{Z_{rot}^{N_1}} \right) \nabla T \end{aligned} \quad (70)$$

Here  $\alpha_i$  is the thermal expansion coefficient for the respective component,  $\beta = \frac{v_1 \sigma_{22}^3}{v_2 \sigma_{12}^3}$  is the parameter characterizing the geometrical relationship between the different component molecules, and  $a = \frac{\pi^2 \sigma_{12}^3 \varepsilon_{12}}{9v_1}$  is the energetic parameter similar to the respective parameter in the van der Waals equation (Landau, Lifshitz, 1980) but characterizing the interaction between the different kinds of molecules. Then, using Eqs. (20), (70), we can write:

$$S_T = \tau \frac{(1-\phi)S_{1T} - \phi S_{2T} + S_T^{kin}}{4(\phi - 1/2)^2 + \tau - 1} \quad (71)$$

where  $\tau = T/T_c$  is the ratio of the temperature at the point of measurement to the critical

temperature  $T_c = \frac{a}{k} \left( \frac{\varepsilon_{11} + \beta \varepsilon_{22}}{\varepsilon_{12}} - 1 - \beta \right)$ , where phase layering in the system begins.

Assuming that  $\beta \approx 1$ , the condition for parameter  $T_c$  to be positive is as  $\varepsilon_{11} + \varepsilon_{22} > 2\varepsilon_{12}$ . This means that phase layering is possible when interactions between the identical molecules are stronger than those between different molecules. When  $\varepsilon_{11} + \varepsilon_{22} < 2\varepsilon_{12}$ , the present theory predicts absolute miscibility in the system.

At temperatures lower than some positive  $T_c$ , when  $\tau < 1$  only solutions in a limited concentration range can exist. In this temperature range, only mixtures with  $\phi \leq \phi_1^*$ ,  $\phi \geq \phi_2^*$  can exist, where  $\phi_{1,2}^* = (1 \pm \sqrt{1-\tau})/2$ , which is equivalent to the equation that defines the boundary for phase layering in phase diagrams for regular solutions, as discussed in

(Kondepudi, Prigogine, 1999).  $S_{iT} = \alpha_i a [(\varepsilon_{ii}/\varepsilon_{12}) - 1] / 2kT$  is the “potential” Soret coefficient related to intermolecular interactions in dilute systems. These parameters can be both positive and negative depending on the relationship between parameters  $\varepsilon_{ii}$  and  $\varepsilon_{12}$ . When the intermolecular interaction is stronger between identical solutes, thermodiffusion is positive, and vice versa. This corresponds to the experimental data of Ning and Wiegand (2006).

When simplifications are taken into account, the equations expressed by the non-equilibrium thermodynamic approach are equivalent to expressions obtained in our hydrodynamic approach (Schimpf, Semenov, 2004; Semenov, Schimpf, 2005). Parameter  $S_T^{kin}$  in Eq. (71) is the kinetic contribution to the Soret coefficient, and has the same form as the term in square brackets in Eq. (37). In deriving this “kinetic” Soret coefficient, we have made different assumptions regarding the properties and concentration of the virtual particles for different terms in Eq. (70).

In deriving the temperature derivative of the combined chemical potential at constant pressure in Eq. (70) we used assumption a) in Section 4, which corresponds to zero entropy of mixing. Without such an assumption a pure liquid would be predicted to drift when subjected to a temperature gradient. Furthermore, the term that corresponds to the entropy of mixing  $-k \ln[\phi/(1-\phi)]$  will approach infinity at low volume fractions, yielding unacceptably high negative values of the Soret coefficient. However, in deriving the concentration derivative we must accept assumption b) because without this assumption the term related to entropy of mixing in Eq. (70) is lost. Consequently, the concentration derivative becomes zero in dilute mixtures and the Soret coefficient approaches infinity.

Thus, we are required to use different assumptions regarding the properties of the virtual particles in the two expressions for diffusion and thermodiffusion flux. This situation reflects a general problem with statistical mechanics, which does not allow for the entropy of mixing for approaching the proper limit of zero at infinite dilution or as the difference in particle properties approaches zero. This situation is known as the Gibbs paradox.

In a diluted system, at  $\phi \ll 1$ , Eq. (71) is transformed into Eq. (37) at any temperature, provided  $\phi \ll \phi_1^*$ . At  $|\tau| \gg 1$ , when the system is miscible at all concentrations,  $S_T$  is a linear function of the concentration

$$S_T = (1 - \phi)S_{T1} - \phi S_{T2} + S_T^{kin} \quad (72)$$

Eq. (72) yields the main features for thermodiffusion of molecules in a one-phase system. It describes the situation where the Soret coefficient changes its sign at some volume fraction. Thus a change in sign with concentration is possible when the interaction between molecules of one component is strong enough, the interaction between molecules of the second component is weak, and the interaction between the different components has an intermediate value. Ignoring again the kinetic contribution, the condition for changing the sign change can be written as  $(\varepsilon_{22} + \varepsilon_{11})/2 > \varepsilon_{12} > \varepsilon_{11}$  or  $(\varepsilon_{22} + \varepsilon_{11})/2 < \varepsilon_{12} < \varepsilon_{11}$ . A good example of such a system is the binary mixture of water with certain alcohols, where a change of sign was observed (Ning, Wiegand, 2006).

## 9. Conclusion

Upon refinement, a model for thermodiffusion in liquids based on non-equilibrium thermodynamics yields a system of consistent equations for providing an unambiguous

description of material transport in closed stationary systems. The macroscopic pressure gradient in such systems is determined by the Gibbs-Duhem equation. The only assumption used is that the heat of transport is equivalent to the negative of the chemical potential. In open and non-stationary systems, the macroscopic pressure gradient is calculated using modified material transport equations obtained by non-equilibrium thermodynamics, where the macroscopic pressure gradient is the unknown parameter. In that case, the Soret coefficient is expressed through combined chemical potentials at constant pressure. The resulting thermodynamic expressions allow for the use of statistical mechanics to relate the gradient in chemical potential to macroscopic parameters of the system.

This refined thermodynamic theory can be supplemented by microscopic calculations to explain the characteristic features of thermodiffusion in binary molecular solutions and suspensions. The approach yields the correct size dependence in the Soret coefficient and the correct relationship between the roles of electrostatic and Hamaker interactions in the thermodiffusion of colloidal particles. The theory illuminates the role of translational and rotational kinetic energy and the consequent dependence of thermodiffusion on molecular symmetry, as well as the isotopic effect. For non-dilute molecular mixtures, the refined thermodynamic theory explains the change in the direction of thermophoresis with concentration in certain mixtures, and the possibility of phase layering in the system. The concept of a Laplace-like pressure established in the force field of the particle under consideration plays an important role in microscopic calculations. Finally, the refinements make the thermodynamic theory consistent with hydrodynamic theories and with empirical data.

## 10. List of symbols

$a$	Energetic parameter characterizing the interaction between the different kinds of molecules
$a_m$	Empiric coefficient in Eq. (38)
$b_i$	Empiric coefficient in Eq. (38)
$\bar{E}$	Electric field strength
$e_i$	Electric charge of the respective ion
$g_{ij}$	Pair correlation function for respective components
$h$	Planck constant
$I_1, I_2, I_3$	and Principal values of the tensor of the moment of inertia
$\bar{J}$	Total material flux in the system
$\bar{J}_e$	Energy flux
$\bar{J}_i$	Component material fluxes
$k$	Boltzmann constant
$L_i$ and $L_{iQ}$	Individual molecular kinetic coefficients
$l$	Thickness of a spherical layer around the particle
$m_i$	Molecular mass of the respective component
$m_{N_1}$	Mass of the virtual particle
$N$	Number of components in the mixture



$N_{ik}$	Number of the molecules of the $k$ 'th component that are displaced by a molecule of $i$ 'th component
$N_1 = N_{21}$	Number of solvent molecules displaced by the solute in binary systems
$n$	Ratio of particle to solvent thermal conductivity
$n_s$	Numeric volume concentration of salt
$n_i$	Numeric volume concentration of the respective component
$P$	Internal macroscopic pressure of the system
$q_i$	Molecular heat of transport
$\vec{r}$	Coordinate of the correlated molecule when the considered particle is placed at $r = 0$
$\vec{r}_0$	Unit radial vector
$\vec{r}_i$	Coordinate of internal molecule or atom in the particle
$R$	Radius of a colloidal particle
$S$	Surface area of a spherical layer around the particle
$S_T$	Soret coefficient in binary systems
$S_{iT}$	Contribution of the intermolecular interactions in Eq. (38) and in the Soret coefficient for diluted systems.
$S_T^s \approx 10^{-3}$	Characteristic Soret coefficient for the salts
$S_T^{kin}$	Contribution of kinetic energy to the Soret coefficient
$T$	Temperature
$T_c$	Critical temperature, where phase layering in binary systems begins
$t$	Time
$V_{out}^i$	Volume external to a molecule of the $i$ 'th component
$V_{in}^i$	Internal volume of a molecule or atom of the $i$ 'th component
$\bar{v}_k$	Partial molecular volume of respective component
$v_k$	Its specific molecular volume
$x$	Distance from the colloid particle surface to the closest solvent molecule surface
$y$	Dimensionless distance from the colloid particle surface to the closest solvent molecule surface
$Z_{rot}$	Rotational statistical sum for polyatomic molecules
$Z_{rot}^i$	Rotational statistical sum for the respective component
$Z_{vib}^i$	Vibrational statistical sum for the respective component
$Z_{rot}^{N_{ik}}$	Rotational statistical sum for the virtual particle of the molecules $k$ 'th component displaced by the molecule of $i$ 'th component
$\alpha_i$	Thermal expansion coefficient for the respective component
$\beta$	Parameter characterizing the geometrical relationship between the different component molecules

$\Delta I$	Difference in the moment of inertia for the molecules constituting the binary mixture
$\Delta M$	Difference in the mass for the molecules constituting the binary mixture
$\varepsilon_{ij}$	Energy of interaction between the molecules of the respective components
$\Phi_{ij}(r)$	Interaction potential for the respective molecules
$\Phi_j^{N_{ik}}$	Total interaction potential of the atoms or molecules included in the respective virtual particle
$\Phi_{i1}^*(r)$	Hamaker potential of isolated colloid particle
$\Phi$	Macroscopic electrical potential
$\Phi_e = e\Phi$	Electrostatic interaction energy
$\phi = \phi_2$	Volume fraction of the second component in binary mixtures
$\phi_i$	Volume fraction of the respective component
$\phi_{1,2}^*$	Boundary values of stable volume fractions in binary systems below the critical temperature
$\gamma_i$	Molecular symmetry number for the respective component
$\gamma_{N_1}$	Molecular symmetry number for the virtual particle in binary mixture
$\lambda$	Parameter which describes the gradual “switching on” of the intermolecular interaction
$\lambda_D$	Debye length
$\mu_i$	Chemical potential of the respective component
$\mu_{0i}$	Chemical potential of the ideal gas of the molecules or atoms of the respective component
$\mu_{ik}^* = \mu_i - \frac{v_i}{v_k} \mu_k$	Combined chemical potential for the respective components
$\mu_P^* = \mu^* = \mu_{21P}^*$	Combined chemical potential at the constant pressure for the binary systems
$\mu_{iP}, \nabla \mu_{iV}$	Chemical potentials of the respective component at the constant pressure and volume, respectively
$\mu_2^e$	Electrostatic contribution to the chemical potential at the constant volume for the charged colloid particle
$\mu_{2P}^e$	Electrostatic contribution to the chemical potential at the constant pressure for the charged colloid particle
$\Pi_i$	Local pressure distribution around the respective molecule or particle
$\Pi_e$	Electrostatic contribution to the local pressure distribution around the charged colloid particle
$\sigma_{ij}$	Minimal molecular approach distance
$\zeta$	Electrokinetic potential
$\tau = T/T_c$	Ratio of the temperature at the point of measurement to the critical temperature

## 11. References

- Anderson, J.L. (1989) Colloid Transport by Interfacial Forces. *Annual Review of Fluid Mechanics*. Vol. 21, 61-99.
- Bringuier, E., Bourdon, A. (2003). Colloid transport in nonuniform temperature. *Physical Review E*, Vol. 67, No. 1 (January 2003), 011404 (6 pages).
- Bringuier, E., Bourdon, A. (2007). Colloid Thermophoresis as a Non-Proportional Response. *The Journal of Non-equilibrium Thermodynamics*. Vol. 32, No. 3 (July 2007), 221-229, ISSN 0340-0204
- De Groot, S. R. (1952). *Thermodynamics of Irreversible Processes*. North-Holland, Amsterdam, The Netherlands.
- De Groot, S. R., Mazur, P. (1962). *Non-Equilibrium Thermodynamics*. North-Holland, Amsterdam, The Netherlands.
- Dhont, J. K. G. (2004). Thermodiffusion of interacting colloids. *The Journal of Chemical Physics*. Vol.120, No. 3 (February 2004) 1632-1641.
- Dhont, J. K. G. et al, (2007). Thermodiffusion of Charged Colloids: Single-Particle Diffusion. *Langmuir*, Vol. 23, No. 4 (November 2007), 1674-1683.
- Duhr, S., Braun, D. (2006). Thermophoretic Depletion Follows Boltzmann Distribution. *Physical Review Letters*, Vol. 96, No. 16 (April 2006) 168301 (4 pages)
- Duhr, S., Braun, D. (2006). Why molecules move along a temperature gradient. *Proceedings of National Academy of USA*, Vol. 103, 19678-19682.
- Fisher, I. Z. (1964). *Statistical Theory of Liquids*. Chicago University Press, Chicago, USA.
- Ghorayeb, K., Firoozabadi, A. (2000). Molecular, pressure, and thermal diffusion in nonideal multicomponent mixtures. *AIChE Journal*, Vol. 46, No. 5 ( May 2000), 883-891.
- Giddings, J. C. et al. (1995). Thermophoresis of Metal Particles in a Liquid. *The Journal of Colloid and Interface Science*. Vol. 176, No. 454-458.
- Gyarmati, I. (1970). *Non-Equilibrium Thermodynamics*. Springer Verlag, Berlin, Germany.
- Haase, R. (1969). *Thermodynamics of Irreversible Processes*, Addison-Wesley: Reading, Massachusetts, USA.
- Hunter, R. J. (1992). *Foundations of Colloid Science*. Vol. 2, Clarendon Press, London, Great Britain.
- Kirkwood, I. , Boggs, E. (1942). The radial distribution function in liquids. *The Journal of Chemical Physics*., Vol. 10, n.d., 394-402.
- Kirkwood, J. G. (1939). Order and Disorder in Liquid Solutions. *The Journal of Physical Chemistry*, Vol. 43, n.d., 97-107
- Kondepudi, D, Prigogine, I. (1999). *Modern Thermodynamics: From Heat Engines to Dissipative Structures*, ISBN 0471973947, John Wiley and Sons, New York, USA.
- Landau, L. D., Lifshitz, E. M. (1954). *Mekhanika Sploshnykh Sred* (Fluid Mechanics) (GITTL, Moscow, USSR) [Translated into English (1959, Pergamon Press, Oxford, Great Britain)].
- Landau, L. D., Lifshitz, E. M. (1980). *Statistical Physics*, Part 1, English translation, Third Edition, Lifshitz, E. M. and Pitaevskii, L. P., Pergamon Press, Oxford, Great Britain.
- Ning, H., Wiegand, S. (2006). Experimental investigation of the Soret effect in acetone/water and dimethylsulfoxide/water mixtures. *The Journal of Chemical Physics*. Vol. 125, No. 22 (December 2006), 221102 (4 pages).

- Pan S et al. (2007). Theoretical approach to evaluate thermodiffusion in aqueous alkanol solutions. *The Journal of Chemical Physics*, Vol. 126, No. 1 (January 2007), 014502 (12 pages).
- Parola, A., Piazza, R. (2004). Particle thermophoresis in liquids. *The European Physical Journal*, Vol.15, No. 11(November2004), 255-263.
- Ross, S. and Morrison, I. D. (1988) *Colloidal Systems and Interfaces*, John Wiley and Sons, New York, USA.
- Ruckenstein, E. (1981). Can phoretic motion be treated as interfacial tension gradient driven phenomena? *The Journal of Colloid and Interface Science*, Vol. 83No. 1 (September 1981), 77-82.
- Schimpf, M. E., Semenov, S. N. (2004). Thermophoresis of Dissolved Molecules and Polymers: Consideration of the Temperature-Induced Macroscopic Pressure Gradient. *Physical Review E*, Vol. 69, No. 1 (October 2004), 011201 (8 pages).
- Semenov, S. N., Schimpf, M. E. (2005). Molecular thermodiffusion (thermophoresis) in liquid mixtures. *Physical Review E*, Vol. 72, No. 4 (October 2005) 041202 (9 pages).
- Semenov, S. N., Schimpf, M. E. (2000) . Mechanism of Polymer Thermophoresis in Nonaqueous Solvents. *The Journal of Physical Chemistry B*, Vol. 104, No. 42 (July 2000),9935-9942.
- Semenov, S. N., Schimpf, M. E. (2009). Mass Transport Thermodynamics in Nonisothermal Molecular Liquid Mixtures. *Physics – Uspekhi*, Vol. 52, No. 11 (November 2009), 1045-1054.
- Semenov, S. N. (2010). Statistical thermodynamic expression for the Soret coefficient *Europhysics Letters*, Vol. 90, No. 5 (June 2010), 56002 (6 pages).
- Weinert, F. M., Braun, D. (2008). Observation of Slip Flow in Thermophoresis. *Physical Review Letters*. Vol. 101 (October 2008), 168301(4 pages).
- Semenov, S. N., Schimpf, M. E. (2011). Internal degrees of freedom, molecular symmetry and thermodiffusion. *Comptes Rendus Mecanique*, doi:10.1016/j.crme.2011.03.011.
- Semenov, S. N., Schimpf, M. E. (2011).Thermodynamics of mass transport in diluted suspensions of charged particles in non-isothermal liquid electrolytes. *Comptes Rendus Mecanique*, doi:10.1016/j.crme.2011.03.002.
- Semenov, S. N. (2011). Statistical thermodynamics of material transport in non-isothermal binary molecular systems. Submitted to *Europhysics Letters*.
- Wiegand, S., Kohler, W. (2002). Measurements of transport coefficients by an optical grating technique. In: *Thermal Nonequilibrium Phenomena in Fluid Mixtures* (Lecture Notes in Physics, Vol. 584, W. Kohler, S. Wiegand (Eds.), 189-210, ISBN 3-540-43231-0, Springer, Berlin, Germany).
- Wittko, G., Köhler, W. (2005). Universal isotope effect in thermal diffusion of mixtures containing cyclohexane and cyclohexane-d12. *The Journal of Chemical Physics*, Vol. 123, No. 6 (June 2005), 014506 (6 pages).

# Thermodynamics of Surface Growth with Application to Bone Remodeling

Jean-François Ganghoffer

LEMMA – ENSEM, 2, Avenue de la Forêt de Haye,  
France

## 1. Introduction

In physics, surface growth classically refers to processes where material reorganize on the substrate onto which it is deposited (like epitaxial growth), but principally to phenomena associated to phase transition, whereby the evolution of the interface separating the phases produces a crystal (Kessler, 1990; Langer, 1980). From a biological perspective, *surface growth* refers to mechanisms tied to accretion and deposition occurring mostly in hard tissues, and is active in the formation of teeth, seashells, horns, nails, or bones (Thompson, 1992). A landmark in this field is Skalak (Skalak et al., 1982, 1997) who describe the growth or atrophy of part of a biological body by the accretion or resorption of biological tissue lying on the surface of the body. Surface growth of biological tissues is a widespread situation, with may be classified as either fixed growth surface (e.g. nails and horns) or moving growing surface (e.g. seashells, antlers). Models for the kinematics of surface growth have been developed in (Skalak et al., 1997), with a clear distinction between cases of fixed and moving growth surfaces, see (Ganghoffer et al., 2010a,b; Garikipati, 2009) for a recent exhaustive literature review.

Following the pioneering mechanical treatments of elastic material surfaces and surface tension by (Gurtin and Murdoch, 1975; Mindlin, 1965), and considering that the boundary of a continuum displays a specific behavior (distinct from the bulk behavior), subsequent contributions in this direction have been developed in the literature (Gurtin and Struthers, 1990; Gurtin, 1995, Leo and Sekerka, 1989) for a thermodynamical approach of the surface stresses in crystals; configurational forces acting on interfaces have been considered e.g. in (Maugin, 1993; Maugin and Trimarco, 1995) – however not considering surface stress -, and (Gurtin, 1995; 2000) considering specific balance laws of configurational forces localized at interfaces.

Biological evolution has entered into the realm of continuum mechanics in the 1990's, with attempts to incorporate into a continuum description time-dependent phenomena, basically consisting of a variation of material properties, mass and shape of the solid body. One outstanding problem in developmental biology is indeed the understanding of the factors that may promote the generation of biological form, involving the processes of growth (change of mass), remodeling (change of properties), and morphogenesis (shape changes), a classification suggested by Taber (Taber, 1995).

The main focus in this chapter is the setting up of a modeling platform relying on the thermodynamics of surfaces (Linford, 1973) and configurational mechanics (Maugin, 1993)

for the treatment of surface growth phenomena in a biomechanical context. A typical situation is the external remodeling in long bones, which is induced by genetic and epigenetic factors, such as mechanical and chemical stimulations. The content of the chapter is the following: the thermodynamics of coupled irreversible phenomena is briefly reviewed, and balance laws accounting for the mass flux and the mass source associated to growth are expressed (section 2). Evolution laws for a growth tensor (the kinematic multiplicative decomposition of the transformation gradient into a growth tensor and an accommodation tensor is adopted) in the context of volumetric growth are formulated, considering the interactions between the transport of nutrients and the mechanical forces responsible for growth. As growth deals with a modification of the internal structure of the body in a changing referential configuration, the language and technique of Eshelbian mechanics (Eshelby, 1951) are adopted and the driving forces for growth are identified in terms of suitable Eshelby stresses (Ganghoffer and Haussy, 2005; Ganghoffer, 2010a). Considering next surface growth, the thermodynamics of surfaces is first exposed as a basis for a consistent treatment of phenomena occurring at a growing surface (section 3), corresponding to the set of generating cells in a physiological context. Material forces for surface growth are identified (section 4), in relation to a surface Eshelby stress and to the curvature of the growing surface. Considering with special emphasis bone remodeling (Cowin, 2001), a system of coupled field equations is written for the superficial density of minerals, their concentration and the surface velocity, which is expressed versus a surface material driving force in the referential configuration. The model is able to describe both bone growth and resorption, according to the respective magnitude of the chemical and mechanical contributions to the surface driving force for growth (Ganghoffer, 2010a). Simulations show the shape evolution of the diaphysis of the human femur. Finally, some perspectives in the field of growth of biological tissues are mentioned.

As to notations, vectors and tensors are denoted by boldface symbols. The inner product of two second order tensors is denoted  $(\mathbf{A}\cdot\mathbf{B})_{ij} = A_{ik}B_{kj}$ . The material derivative of any function is denoted by a superposed dot.

## 2. Thermodynamics of irreversible coupled phenomena: a survey

We consider multicomponent systems, mutually interacting by chemical reactions. Two alternative viewpoints shall be considered: in the first viewpoint, the system is closed, which in consideration of growth phenomena means that the nutrients are included into the overall system. The second point of view is based on the analysis of a solid body as an open system exchanging nutrients with its surrounding; hence growth shall be accounted for by additional source terms and convective fluxes.

### 2.1 Multiconstituents irreversible thermodynamics

We adopt the thermodynamic framework of open systems irreversible thermodynamics, which shall first be exposed in a general setting, and particularized thereafter for growing continuum solid bodies. Recall first that any extensive quantity  $A$  with volumetric density  $a = a(\mathbf{x}, t)$  satisfies a prototype balance law of the form

$$\frac{\partial a(\mathbf{x}, t)}{\partial t} = -\nabla \cdot \mathbf{J}_a(\mathbf{x}, t) + \sigma_a(\mathbf{x}, t) \quad (1)$$

with  $\mathbf{J}_a(\mathbf{x}, t)$  the flux density of  $a(\mathbf{x}, t)$  and  $\sigma_a(\mathbf{x}, t)$  the local production (or destruction) of  $a(\mathbf{x}, t)$ . The particular form of the flux and source depend on the nature of the considered extensive quantity, as shall appear in the forthcoming balance laws. We consider a system including  $n$  constituents undergoing  $r$  chemical reactions; the local variations of the partial density of a given constituent  $k$ , quantity  $\rho_k$ , obey the local balance law (Vidal et al., 1994)

$$\frac{\partial \rho_k}{\partial t} = -\nabla \cdot (\rho_k \mathbf{u} + \mathbf{J}_k) + M_k \sum_{\alpha=1..r} \nu_{\alpha k} J_\alpha \quad (2)$$

with  $\mathbf{u} := \frac{1}{\rho} \sum_{k=1}^n \rho_k \mathbf{u}_k$  the local barycentric velocity,  $M_k$  the molar mass, and  $\nu_{\alpha k}$  the stoichiometric coefficients in the reaction  $\alpha$ , such that the variation of the mass  $dm_k$  of the species  $k$  due to chemical reactions expresses as

$$dm_k = M_k \sum_{\alpha=1..r} \nu_{\alpha k} \xi_\alpha, \quad k=1..n \quad (3)$$

wherein  $\xi_\alpha$  denotes the degree of advancement of reaction  $\alpha$ . The molar masses  $M_k$  satisfy the global conservation law (due to Lavoisier)

$$\sum_{k=1}^n \nu_{\alpha k} M_k = 0, \quad \alpha = 1..r \quad (4)$$

Observe that the total flux of mass is the sum of a convective flux  $\rho_k \mathbf{u}$  and a diffusive flux  $\mathbf{J}_k$ ; the mass production is identified as the contribution  $M_k \sum_{\alpha=1..r} \nu_{\alpha k} J_\alpha$ . In this viewpoint,

the system is in fact closed, since the balance law satisfied by the global density  $\rho = \sum_{k=1}^n \rho_k$

writes (Vidal et al., 1994) accounting for the relation  $\sum_{j=1}^n \mathbf{J}_j = \sum_{j=1}^n \rho_j \mathbf{u}_j = \mathbf{0}$ , as

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{u}) + \sum_{k=1}^n \sum_{\alpha=1..r} M_k \nu_{\alpha k} J_\alpha \equiv -\nabla \cdot (\rho \mathbf{u}) \quad (5)$$

This balance law does not involve any source term for the total density. Instead of using the partial densities of the system constituents, one can write balance equations for the number of moles of constituent  $k$ ,  $n_k = m_k / M_k$ , with  $m_k$  the mass of the same constituent. The molar concentration is defined as  $c_k = n_k / V$ , its inverse being called the partial molar volume. The partial mole number  $n_k$  satisfies the balance equation

$$\frac{\partial n_k}{\partial t} = -\text{div} \mathbf{J}_k + \frac{\partial_i n_k}{\partial t} \quad (6)$$

with  $\mathbf{J}_k$  the flux of species  $k$  and  $\frac{\partial_i n_k}{\partial t}$  its production term, given by De Donder definition of the rate of progress of the  $j^{\text{th}}$  chemical reaction

$$\frac{\partial_i n_k}{\partial t} = \sum_{j=1}^r v_{kj} \dot{\xi}_j \quad (7)$$

The two previous equalities enter into Gibbs relation as

$$\rho \dot{u} = \theta \rho \dot{s}_e + \theta \rho \dot{s}_i + \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} - \sum_k \mu_k \frac{\rho}{M} \operatorname{div} \mathbf{J}_k + \sum_k \mu_k \frac{\rho}{M} \sum_j v_{kj} \dot{\xi}_j \quad (8)$$

with  $\theta$  the temperature and  $\mu_k$  the chemical potential of constituent  $k$ . The chemical affinity in the sense of De Donder is defined as the force conjugated to the rate  $\dot{\xi}_j$

$$A_j = - \sum_k \left( \mu_k \frac{\rho}{M} \right) v_{kj} = - \sum_k (\mu_k / V) v_{kj} \quad (9)$$

Hence, Gibbs relation can be rewritten in order to highlight the variation of entropy

$$\rho \dot{s} = \frac{1}{\theta} \rho \dot{u} - \frac{1}{\theta} \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} + \sum_k \frac{1}{\theta} \left( \mu_k \frac{1}{V} \right) \operatorname{div} \mathbf{J}_k + \sum_j \frac{1}{\theta} A_j \dot{\xi}_j \quad (10)$$

The local balance of internal energy traduces the first principle of thermodynamics as

$$\rho \dot{u} = -\nabla \cdot \mathbf{J}_q + \rho \dot{w}$$

with  $\mathbf{J}_q$  the heat flux, and the term  $\rho \dot{w}$  is relative to all forms of work. One shall isolate the flux-like contributions in the entropy variation, which after a few transformations writes

$$\rho \dot{s} = \rho \dot{s}_e + \rho \dot{s}_i = \frac{1}{\theta} \nabla \cdot \mathbf{J}_q + \sum_k \nabla \cdot \left( \mathbf{J}_k \frac{\mu_k}{\theta} \frac{1}{V} \right) - \sum_k \mathbf{J}_k \cdot \nabla \left( \frac{\mu_k}{\theta} \frac{1}{V} \right) + \frac{1}{\theta} (\rho \dot{w} - \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}) + \sum_j \frac{1}{\theta} A_j \dot{\xi}_j$$

The contribution  $-\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} / \theta$  (involving the virtual power of internal forces) is further decomposed into

$$\frac{1}{\theta} \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} = \frac{1}{\theta} \boldsymbol{\sigma} : \nabla \mathbf{u} = \nabla \cdot \left( \frac{1}{\theta} \boldsymbol{\sigma} : \mathbf{u} \right) - \mathbf{u} \cdot \nabla \cdot \left( \frac{1}{\theta} \boldsymbol{\sigma} \right)$$

Hence, the rate of the entropy density decomposes into

$$\begin{aligned} \rho \dot{s} &= \rho \dot{s}_e + \rho \dot{s}_i = -\nabla \cdot \left( \mathbf{J}_q + \sum_k \mathbf{J}_k \frac{\mu_k}{\theta} \frac{1}{V} \right) + \mathbf{J}_q \cdot \nabla \left( \frac{1}{\theta} \right) - \\ &\sum_k \mathbf{J}_k \cdot \nabla \left( \frac{\mu_k}{\theta} \frac{1}{V} \right) + \frac{1}{\theta} (\rho \dot{w} - \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}) + \sum_j \frac{1}{\theta} A_j \dot{\xi}_j \end{aligned} \quad (11)$$

This writing allows the identification of the divergential contribution to the exchange entropy, hence to the entropy flux

$$\mathbf{J}_s = \mathbf{J}_q + \sum_k \mathbf{J}_k \frac{\mu_k}{\theta} \frac{1}{V} \quad (12)$$



and of the internal entropy production

$$\rho \dot{s}_i = \mathbf{J}_q \cdot \nabla \left( \frac{1}{\theta} \right) - \sum_k \mathbf{J}_k \cdot \nabla \left( \frac{\mu_k}{\theta} \right) - \frac{1}{\theta} (\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} - \rho \dot{w}) + \sum_j \frac{1}{\theta} A_j \dot{\xi}_j \quad (13)$$

which is due to the gradient of intensive variables (temperature, chemical potential), to the irreversible mechanical power spent and to chemical reactions.

An alternative to the previous writing of the internal entropy production bearing the name of Clausius-Duhem inequality is frequently used; as a starting point, the first principle is written as

$$\rho \dot{u} = -\nabla \cdot \mathbf{J}_q + \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} + \sum_k (\mu_k / V) \dot{n}_k \quad (14)$$

One has assumed in this alternative that the mechanical power  $\rho \dot{w} = \boldsymbol{\sigma}^{eq} : \nabla \mathbf{u}$  does not include a flux contribution, hence only the heat diffusion contributes to the flux of internal energy. The contribution  $\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} + \sum_k (\mu_k / V) \dot{n}_k$  is identified to the term  $\rho \dot{w}$ . Previous

equality combined with the second principle, equality  $\rho \dot{s} = -\nabla \cdot \left( \frac{\mathbf{J}_q}{\theta} \right) + \rho \dot{s}_i$  (the entropy flux resumes to the sole heat flux), delivers after a few manipulations the variation of the internal energy as

$$\rho \dot{u} = \left( \theta \rho \dot{s} - \mathbf{J}_q \cdot \frac{\nabla \theta}{\theta} - T \rho \dot{s}_i \right) + \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} + \sum_k (\mu_k / V) \dot{n}_k \quad (15)$$

Hence, the internal entropy production is identified as

$$\theta \rho \dot{s}_i = -\rho (\dot{u} - T \dot{s}) - \mathbf{J}_q \cdot \frac{\nabla \theta}{\theta} + \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} + \sum_k (\mu_k / V) \dot{n}_k \quad (16)$$

which is conveniently rewritten in terms of Helmholtz free energy density  $\psi := u - Ts$  as

$$\theta \rho \dot{s}_i = -\rho (\dot{\psi} + s \dot{\theta}) - \mathbf{J}_q \cdot \frac{\nabla \theta}{\theta} + \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} + \sum_k (\mu_k / V) \dot{n}_k \quad (17)$$

This is at variant with the point of view adopted next, which consists in insulating a growing solid body from the external nutrients, identified as one the chemical species, but accounted for in a global manner as a source term.

## 2.2 General balance laws accounting for mass production due to growth

In the case of mass being created / resorbed within a solid body considered as an open system from a general thermodynamic point of view, one has to account for a source term  $\pi$  being produced (by a set of generating cells) at each point within the time varying volume  $\Omega_t$ ; a convective term is also added, corresponding to the transport of nutrients by the velocity field of the underlying continuum. For any quantity  $a$ , the convective flux is locally defined in terms of its surface density as  $\mathbf{F}(a) = a \mathbf{v}$ ; the overall convective flux of  $a$  across the closed surface  $\partial \Omega_t$  expresses then as

$$\phi_{conv}(a) = - \int_{\partial\Omega_t} a(\mathbf{v} - \mathbf{w}) \cdot \mathbf{n} dA \quad (18)$$

(the minus accounts for the unit exterior normal  $\mathbf{n}$ ). This diffusive flux corresponds to a macroscopic flux

$$\phi_{diff}(a) = - \int_{\partial\Omega_t} \mathbf{J}(a) \cdot \mathbf{n} dA \quad (19)$$

The density of microscopic flux  $\mathbf{J}(a)$  is associated to an invisible motion of molecules within a continuum description, hence must be described by a specific constitutive law. It does not depend on the velocity of the points of  $\partial\Omega_t$ .

The convective derivative along the vector field  $\mathbf{w}$  of the field  $a = a(\mathbf{x}, t)$  writes

$$\frac{\delta_w a}{\delta t} = \left( \frac{\partial a}{\partial t} \right)_x + \nabla a \cdot \mathbf{w} \quad (20)$$

In the case  $\mathbf{w}$  coincides with the velocity of the material particles, previous relation delivers the definition of the material (or particular) derivative

$$\frac{da}{dt} \equiv \frac{\delta_v a}{\delta t} = \frac{\delta_w a}{\delta t} + \nabla a \cdot (\mathbf{v} - \mathbf{w}) \quad (21)$$

The derivative of the volume integral  $A := \int_{\Omega_t} a dx$  is next calculated, according to Leibniz rule:

$$\frac{D}{Dt} \int_{\Omega_t} a dx = \int_{\Omega_t} \frac{\partial a}{\partial t} dx + \int_{\partial\Omega_t} a(\mathbf{w} \cdot \mathbf{n}) dA \quad (22)$$

with  $\mathbf{w}$  the velocity field of the points on  $\partial\Omega_t$ , which is associated to a variation of the domain occupied by the material points of the growing solid body (Figure 1).

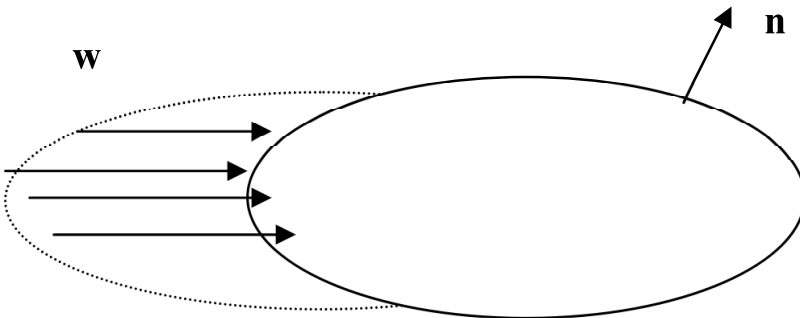


Fig. 1. Domain variation due to the virtual velocity field  $\mathbf{w}$

A global balance equation can next be written, according to the natural physical rule: the balance of any quantity is the sum of the production / destruction term and of the flux; this yields

$$\frac{D}{Dt} \int_{\Omega_t} a dx = \int_{\Omega_t} \Pi dx - \int_{\partial\Omega_t} \{a(\mathbf{v} - \mathbf{w}) + \mathbf{J}(a)\} \cdot \mathbf{n} dA \tag{23}$$

The first term on the r.h.s. corresponds to mass production, the second contribution to convection of the produced mass through the boundary  $\partial\Omega(t)$ , and the third contribution to diffusion through the boundary of the moving volume  $\partial\Omega(t)$ . One can see that only the relative velocity of particles w.r. to the surface velocity matters. Combining this identity with (22) gives

$$\int_{\Omega_t} \frac{\partial a}{\partial t} dx + \int_{\partial\Omega_t} a(\mathbf{w} \cdot \mathbf{n}) dA = \int_{\Omega_t} \Pi dx - \int_{\partial\Omega_t} \{a(\mathbf{v} - \mathbf{w}) + \mathbf{J}(a)\} \cdot \mathbf{n} dA \tag{24}$$

The corresponding local balance law is obtained after elimination of the velocity  $\mathbf{w}$ , hence

$$\frac{\delta_v a}{\delta t} + a \operatorname{div} \mathbf{v} = \Pi - \operatorname{div} \mathbf{J}(a) \Leftrightarrow \frac{\partial a}{\partial t} + \operatorname{div}(a\mathbf{v}) = \Pi - \operatorname{div} \mathbf{J}(a) \tag{25}$$

**Mass balance:** the mass balance equation is deduced from the identification  $a = \rho$ , the actual density. Hence, (23) gives

$$\frac{D}{Dt} \int_{\Omega_t} \rho dx = \int_{\Omega_t} \pi dx - \int_{\partial\Omega_t} \mathbf{J}(\rho) \cdot \mathbf{n} dA \tag{26}$$

The strong form of the balance law of mass writes finally

$$\frac{\delta_v \rho}{\delta t} = \Pi - \operatorname{div} \mathbf{J}(\rho) - \rho \nabla \cdot \mathbf{v} \tag{27}$$

The **mass balance in Eulerian format** is given in terms of the actual density  $\rho$  by the following reasoning: we first write the general form of the balance of mass in physical space as

$$\frac{D}{Dt} \int_{\Omega_t} \rho dx = \int_{\Omega_t} \left( \frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} \right) dx = \int_{\Omega_t} \pi dx + \int_{\partial\Omega_t} m ds \equiv \int_{\Omega_t} \Gamma \rho dx \tag{28}$$

with  $\rho(\mathbf{x}, t)$  the actual density,  $\pi$  the physical source of mass, and  $m := \mathbf{m} \cdot \mathbf{n}$  the scalar physical mass flux across the boundary, projection of the flux (vector)  $\mathbf{m}$ . The previous balance law is quite general, as we account for both the variation of the integration volume through the term  $\rho \nabla \cdot \mathbf{v}$ , and for the source and flux of mass reflected by the right hand side of (28). Localization of previous integral equation gives

$$\frac{D\rho}{Dt} = \pi + \nabla \cdot \mathbf{m} - \rho \nabla \cdot \mathbf{v} \tag{29}$$

with  $\mathbf{v}(\mathbf{x}, t) := \left( \frac{\partial \mathbf{x}}{\partial t} \right)_x$  the Eulerian velocity, which proves identical to (27); the same balance law has been obtained in (Epstein and Maugin, 2000) starting from its Lagrangian counterpart.

In the sequel, we shall extensively use the following expression of the material derivative of integrals of specific quantities (defined per unit mass)  $a = a(\mathbf{x}, t)$ , obtained using the mass balance (28)

$$\frac{D}{Dt} \int_{\Omega_t} \rho a dx = \int_{\Omega_t} \left\{ \rho \frac{Da}{Dt} + a(\pi + \nabla \cdot \mathbf{m}) \right\} dx \quad (30)$$

The comparison of (27) with (29) gives the identification of fluxes  $\mathbf{J}(\rho) \equiv -\mathbf{m}$ ; the balance law is further consistent with (and equivalent to) the writing (Ganghoffer and Haussy, 2005)

$$\dot{\rho} + \rho \operatorname{div}(\mathbf{v}) = \Phi_\rho + \sigma_\rho$$

with  $\Phi_\rho \equiv \nabla \cdot \mathbf{m}$  the total flux of conduction and  $\sigma_\rho \equiv \pi$  the volumetric source of mass. Observe the difference with the treatment of section 1 considering overall a closed system with no internal sources, reflected by equation (1.5): this first point of view considers the nutrients responsible for growth as part of the system, whereas they appear as external sources in the second viewpoint.

Expressing the total mass of the domain  $\Omega_t$  as  $m(\Omega_t) = \int_{\Omega_t} \rho(x) dx$ , the mass variation due to

the transport phenomena is written as the following integral accounting for source terms, allowing the identification of the production term

$$\left( \frac{dm}{dt} \right)_{\text{source}} := \int_{\Omega_t} \pi dx = \int_{\Omega_t} dx \Rightarrow \pi = \Gamma \rho$$

The time variation of chemical concentration of nutrients is due to exchange through the boundary accounted for by a flux  $\int_{\partial\Omega_t} \mathbf{j}_k \cdot \mathbf{n} ds$ , and to a source term due to growth

$$\int_{\Omega_t} \Gamma \rho dx = \int_{\Omega_t} \operatorname{Tr}(\dot{\mathbf{F}}_g \cdot \mathbf{F}_g^{-1}) \rho dx, \text{ hence (see (28))}$$

$$\frac{D}{Dt} \int_{\Omega_t} \rho dx = \int_{\Omega_t} \operatorname{Tr}(\dot{\mathbf{F}}_g \cdot \mathbf{F}_g^{-1}) \rho dx = \int_{\Omega_t} \rho \dot{n}_k dx + \int_{\partial\Omega_t} \mathbf{j}_k \cdot \mathbf{N} ds \rightarrow \rho \dot{n}_k + \operatorname{div} \mathbf{j}_k = \rho (\dot{\mathbf{F}}_g \cdot \mathbf{F}_g^{-1} : \mathbf{I}) = \rho \Gamma \quad (31)$$

The last equality is nothing else than  $\Gamma = (\pi + \nabla \cdot \mathbf{m}) / \rho$  - a consequence of (28) - expressed in material format, with the identifications  $\Gamma := \operatorname{Tr}(\mathbf{D}_g)$ ,  $\mathbf{m}_i = \mathbf{j}_i$ ;  $\pi = \rho \dot{n}_k$ . The global form

$$\frac{D}{Dt} \int_{\Omega_R} \rho dx = \frac{D}{Dt} \int_{\Omega_R} \rho J dX \text{ also fits within the general balance law for an open system, relation}$$

(30) with  $a \equiv 1$ .

**Balance of momentum:** the Eulerian version of the balance of momentum writes (Epstein and Maugin, 2000)

$$\frac{D}{Dt} \int_{\Omega_t} \rho \mathbf{v} dx = \int_{\Omega_t} \mathbf{f} dx + \int_{\partial\Omega_t} \mathbf{n} \cdot \boldsymbol{\sigma} d\sigma_t + \int_{\Omega_t} \pi \mathbf{v} dx + \int_{\partial\Omega_t} \mathbf{n} \cdot (\mathbf{m} \otimes \mathbf{v}) d\sigma_t \quad (32)$$

with  $\boldsymbol{\sigma}$  Cauchy stress and  $\mathbf{f}$  body forces per unit physical volume. Localizing (32) gives using the mass balance (29)

$$\rho \frac{D\mathbf{v}}{Dt} = \mathbf{f} + \text{div} \boldsymbol{\sigma} + (\mathbf{m} \cdot \nabla) \mathbf{v} \quad (33)$$

**Balance of kinetic and internal energy:** the first law of thermodynamics for an open system has to account for the contributions to kinetic and internal energies due to the incoming material. Denoting  $u$  the specific internal energy density, one may write the energy balance in the actual configuration as

$$\begin{aligned} \frac{D}{Dt} \int_{\Omega_t} \rho \left( u + \frac{1}{2} \mathbf{v}^2 \right) dx &= \int_{\Omega_t} \left\{ (\mathbf{f} \cdot \mathbf{v} + r) + \pi \left( u + \frac{1}{2} \mathbf{v}^2 \right) \right\} dx \\ &+ \int_{\partial\Omega_t} \mathbf{n} \cdot \left\{ \boldsymbol{\sigma} \cdot \mathbf{v} + \mathbf{m} \left( u + \frac{1}{2} \mathbf{v}^2 \right) - \mathbf{q} \right\} d\sigma(\mathbf{x}) \end{aligned} \quad (34)$$

with  $r$  the volumetric heat supply (generated by growth), and  $\mathbf{q}$  the heat flux across  $\partial\Omega_t$ . This writing of the energy balance can be simplified using the balance of kinetic energy with volumetric density  $k$ , obtained by multiplying (33) by the velocity and integrating over  $\Omega_t$ , hence

$$\begin{aligned} k &:= \frac{1}{2} \rho \frac{D\mathbf{v}^2}{Dt} = \mathbf{f} \cdot \mathbf{v} + \mathbf{v} \cdot \text{div} \boldsymbol{\sigma} + \mathbf{v} \cdot (\mathbf{m} \cdot \nabla) \mathbf{v} \equiv \mathbf{f} \cdot \mathbf{v} + \mathbf{v} \cdot \text{div} \boldsymbol{\sigma} + \mathbf{m} \cdot \nabla \frac{\mathbf{v}^2}{2} \Rightarrow \\ &\int_{\Omega_t} \frac{1}{2} \rho \frac{D\mathbf{v}^2}{Dt} dx = \int_{\Omega_t} \left( \mathbf{f} \cdot \mathbf{v} + \mathbf{v} \cdot \text{div} \boldsymbol{\sigma} + \mathbf{m} \cdot \nabla \left( \frac{\mathbf{v}^2}{2} \right) \right) dx \end{aligned}$$

The left hand side of previous equality can be expressed versus the material derivative of the total kinetic energy of the growing body, using the general equality (30) with  $a = \frac{1}{2} \mathbf{v}^2$ , hence (35)

$$\begin{aligned} \frac{DK}{Dt} &= \int_{\Omega_t} \left\{ \rho \frac{D\left(\frac{\mathbf{v}^2}{2}\right)}{Dt} + \frac{\mathbf{v}^2}{2} (\pi + \nabla \cdot \mathbf{m}) \right\} dx = \int_{\Omega_t} \left( \mathbf{f} \cdot \mathbf{v} + \mathbf{v} \cdot \text{div} \boldsymbol{\sigma} + \nabla \cdot \left( \mathbf{m} \frac{\mathbf{v}^2}{2} \right) + \pi \frac{\mathbf{v}^2}{2} \right) dx = \\ &\int_{\Omega_t} \left( \mathbf{f} \cdot \mathbf{v} - \boldsymbol{\sigma} : \nabla \mathbf{v} + \pi \frac{\mathbf{v}^2}{2} \right) dx + \int_{\partial\Omega_t} \mathbf{n} \cdot \left\{ \boldsymbol{\sigma} \cdot \mathbf{v} + \mathbf{m} \frac{\mathbf{v}^2}{2} \right\} d\sigma \end{aligned} \quad (35)$$

Using again (30) delivers similarly the material derivative of the total energy (left-hand side in (34)) as (the total internal energy is denoted  $U$ )

$$\begin{aligned} \frac{D}{Dt}(U+K) &= \int_{\Omega_t} \rho \frac{D}{Dt} \left( u + \frac{1}{2} \mathbf{v}^2 \right) dx + \int_{\Omega_t} \left( u + \frac{1}{2} \mathbf{v}^2 \right) (\pi + \nabla \cdot \mathbf{m}) dx = \int_{\Omega_t} \left\{ (\mathbf{f} \cdot \mathbf{v} + r) + \pi \left( u + \frac{1}{2} \mathbf{v}^2 \right) \right\} dx \\ &+ \int_{\partial \Omega_t} \mathbf{n} \cdot \left\{ \boldsymbol{\sigma} \cdot \mathbf{v} + \mathbf{m} \left( u + \frac{1}{2} \mathbf{v}^2 \right) - \mathbf{q} \right\} d\sigma(x) \end{aligned}$$

Using the balance of kinetic energy (35) allows isolating the material derivative of the internal energy

$$\frac{DU}{Dt} = \int_{\Omega_t} (-\boldsymbol{\sigma} : \nabla \mathbf{v} + r + \pi u) dx + \int_{\partial \Omega_t} \mathbf{n} \cdot (\mathbf{m}u - \mathbf{q}) d\sigma(x)$$

Its strong form is given by localization using the general equality (30) with the identification  $a \equiv u$

$$\rho \frac{Du}{Dt} = -\boldsymbol{\sigma} : \nabla \mathbf{v} + r + \mathbf{m} \nabla \cdot u - \nabla \cdot \mathbf{q} \quad (36)$$

The Lagrangian counterpart of previous balance laws has been expressed in (Epstein and Maugin, 2000).

**Dissipation and second principle:** the dissipation inequality writes in global form as

$$\frac{D}{Dt} \int_{\Omega_t} \rho s dx \geq \int_{\Omega_t} (\pi s + \theta^{-1} r) dx - \int_{\partial \Omega_t} \mathbf{n} \cdot \frac{\mathbf{q}}{\theta} d\sigma(\mathbf{x}) \Rightarrow \int_{\Omega_t} \left( \rho \frac{Ds}{Dt} + s \nabla \cdot \mathbf{m} \right) dx \geq \int_{\Omega_t} (\theta^{-1} r) dx - \int_{\partial \Omega_t} \mathbf{n} \cdot \frac{\mathbf{q}}{\theta} d\sigma(\mathbf{x})$$

Hence, the local dissipation inequality localizes as Clausius-Duhem inequality

$$\rho \frac{Ds}{Dt} \geq \theta^{-1} r - \operatorname{div} \left( \frac{\mathbf{q}}{\theta} \right) - s \nabla \cdot \mathbf{m} \quad (37)$$

The previous balance laws are general balance laws in the framework of open systems irreversible thermodynamics; we shall in the next section make the fluxes and source terms involved in those balance laws more specific, in order to identify an evolution law for the volumetric growth of solid bodies.

### 3. Volumetric growth

The kinematics of growth is elaborated from the classical multiplicative decomposition (Rodriguez et al., 1994) of the transformation gradient

$$\mathbf{F} = \nabla_{\mathbf{X}} \mathbf{x}(\mathbf{X}, t) \rightarrow J := \det(\mathbf{F}) \quad (3.1)$$

with  $\mathbf{X}, \mathbf{x}$  the Lagrangian end Eulerian positions in the referential and actual configurations denoted  $\Omega_R, \Omega_t$  respectively, as the product of the growth deformation gradient  $\mathbf{F}_g$  and the growth accommodation mapping  $\mathbf{F}_a$

$$\mathbf{F} = \mathbf{F}_a \cdot \mathbf{F}_g \quad (3.2)$$

The transformation gradients  $\mathbf{F}_a, \mathbf{F}_g, \mathbf{F}$  define the mappings of the tangent spaces to the various configurations. The Jacobean of the growth mapping informs about the nature of growth:

$$J_g := \det(\mathbf{F}_g) \tag{3.3}$$

Hence  $J_g < 1$  describes growth, whereas  $J_g > 1$  represents resorption. Growth essentially occurs between the referential and the actual configurations.

Adopting the framework of hyperelasticity, the first Piola-Kirchhoff stress  $\mathbf{P}$  expresses from the strain energy density per unit volume in the reference configuration  $W(\mathbf{F}_a; \mathbf{X})$  with argument the reversible part of the transformation gradient (a possible explicit dependence upon the Lagrangian variable is included for heterogeneous media) as

$$\mathbf{P} := \partial_{\mathbf{F}} W(\mathbf{F}_a; \mathbf{X}) \tag{3.4}$$

A more explicit (compared to (38)) expression of the dissipation accounting for heat and matter exchanges is obtained by considering the general form of the balance of energy and entropy: let denote  $u$  and  $s$  the density of internal energy and entropy per unit mass respectively; the first and second principles of thermodynamics write (Munster, 1970)

$$\rho \dot{u} = -\nabla \cdot \mathbf{J}_q - p_i + \mathbf{J}_k \cdot \mathbf{F}_k \ ; \ \rho \dot{s} = -\nabla \cdot \mathbf{J}_s + \sigma_s \tag{3.5}$$

with  $\mathbf{J}_q$  the heat diffusion flux,  $\mathbf{J}_s := \frac{1}{\theta}(\mathbf{J}_q - \mu_i \mathbf{J}_i)$  the total entropy flux,  $\mathbf{J}_k$  the diffusion flux of the  $k$ -specie,  $\mathbf{F}_k(\mathbf{x}, t)$  an external force acting on the  $k$ -specie, and  $\sigma_s$  the entropy production, always positive (it is dissipated). Introducing the free energy density per unit mass  $\psi := u - \theta s$ , with  $s$  the entropy density, we then immediately obtain the rate of variation of the free energy density

$$\rho \dot{\psi} = \rho s \dot{\theta} - \nabla \cdot \mathbf{J}_q + \theta \nabla \cdot \mathbf{J}_s + \mathbf{J}_k \cdot \mathbf{F}_k - p_i - \sigma_s \tag{3.6}$$

The positivity of the entropy production  $\sigma_s$  in previous inequality then expresses as

$$\rho \dot{\psi} \leq -p_i + \mathbf{J}_k \cdot \mathbf{F}_k + \nabla \cdot (\mathbf{J}_q - \mu_i \mathbf{J}_i) \tag{3.7}$$

The principle of virtual power  $\frac{dK}{dt} = P_e + P_i$  ( $K$  is the kinetic energy,  $P_e, P_i$  being the virtual power of external and internal forces respectively), leads to the global form of previous inequality in Eulerian format:

$$\frac{dK}{dt} + \int_{\Omega} \rho \dot{\psi} dx \leq P_e + \int_{\Gamma_k} \underline{E}_k + \Phi_m + \Phi_q \tag{3.8}$$

with  $\Phi_q := \int_{\partial\Omega} \mathbf{J}_q \cdot \mathbf{n} d\sigma$  and  $\Phi_m := - \int_{\partial\Omega} \mu_i \mathbf{J}_i \cdot \mathbf{n} d\sigma$  respectively the flux of heat and mass through the boundary of  $\Omega$ . Previous inequality traduces the fact that the flux of mechanical work

and mass increases the kinetic and internal free energy of the system, the difference being dissipated.

The second principle may be rewritten after a few manipulations in terms of a dynamical Eshelby stress accounting for all sources of energies (mechanical, chemical, thermal): the free energy density is taken to depend on the elastic part of the transformation gradient  $\mathbf{F}_a$ , the concentration of chemical specie  $n_k$  and the temperature  $\theta$ , so that Clausius-Duhem inequality (3.7) becomes in material format:

$$\begin{aligned} \frac{D}{Dt}(\rho J \psi) &\leq \mathbf{T} : \dot{\mathbf{F}} + \nabla \cdot (\mathbf{J}_q - \mu_i \mathbf{J}_i) \rightarrow \\ \rho J \psi \mathbf{I} : \dot{\mathbf{F}}_g \cdot \mathbf{F}_g^{-1} &+ \rho J \frac{\partial \psi}{\partial \mathbf{F}_a} \dot{\mathbf{F}}_a + \rho J \frac{\partial \psi}{\partial n_k} \dot{n}_k \\ + \rho J \frac{\partial \psi}{\partial \theta} \dot{\theta} &\leq \mathbf{T} : (\dot{\mathbf{F}}_a \cdot \mathbf{F}_g + \mathbf{F}_a \cdot \dot{\mathbf{F}}_g) + \text{Div} \mathbf{J}_q - \mu_i \text{Div} \mathbf{J}_i - \mathbf{J}_i \text{Grad} \mu_i \end{aligned} \quad (3.9)$$

The balance of biochemical energy expresses that the time variation of chemical concentration of nutrients is due to exchange through the boundary accounted for by the term  $\int_{\partial \Omega_R} \mathbf{J}_\mu \cdot \mathbf{N} dA$  and to a source term due to growth  $\int_{\Omega_R} \Gamma \rho J dX = \int_{\Omega_R} \text{Tr}(\dot{\mathbf{F}}_g \cdot \mathbf{F}_g^{-1}) \rho J dX$ , hence

$$\begin{aligned} \frac{D}{Dt} \int_{\Omega_i} \rho dx &= \int_{\Omega_R} \text{Tr}(\dot{\mathbf{F}}_g \cdot \mathbf{F}_g^{-1}) \rho J dX = \int_{\Omega_R} \rho J \dot{n}_k dX \\ + \int_{\partial \Omega_R} \mathbf{J}_k \cdot \mathbf{N} dA &\rightarrow \rho J \dot{n}_k + \text{Div} \mathbf{J}_k = \rho J (\dot{\mathbf{F}}_g \cdot \mathbf{F}_g^{-1} : \mathbf{I}) \rho J \end{aligned} \quad (3.10)$$

The last equality is nothing else than  $\Gamma = (\pi + \nabla \cdot \mathbf{m}) / \rho$  expressed in material format, identifying  $\Gamma := \text{Tr}(\mathbf{D}_g)$ ,  $\mathbf{M}_i = \mathbf{J}_i$ . Accordingly, (3.9) becomes

$$\begin{aligned} 0 &\leq \left( \mathbf{T} \cdot \mathbf{F}_g^t - \rho J \frac{\partial \psi}{\partial \mathbf{F}_a} \right) \dot{\mathbf{F}}_a - \left( \frac{\partial \psi}{\partial n_k} - \mu_k \right) \rho J \dot{n}_k - \mathbf{J}_i \text{Grad} \mu_i \\ + \left( \mathbf{F}_a^t \cdot \mathbf{T} \cdot \mathbf{F}_g^t - \rho J \psi \mathbf{I} - \rho J \frac{\partial \psi}{\partial n_k} \right) &: \dot{\mathbf{F}}_g \cdot \mathbf{F}_g^{-1} + \rho J s \dot{\theta} \end{aligned} \quad (3.11)$$

Since previous equality must hold true for arbitrary variations  $\dot{\mathbf{F}}_a, \dot{n}_k$ , the following constitutive equations for the first Piola-Kirchhoff stress and the chemical potential are obtained

$$\mathbf{T} = \rho J \frac{\partial \psi}{\partial \mathbf{F}_a} \cdot \mathbf{F}_g^{-t}; \quad \mu_k = \frac{\partial \psi}{\partial n_k} \quad (3.12)$$

Especially, (3.12)<sub>1</sub> is an alternative to (3.4) using the specific free energy instead of a strain energy potential; observe that  $\psi$  is expressed per unit mass, in contrast to  $W(\mathbf{F}_a; \mathbf{X})$  in (3.4), expressed per unit referential volume.

The residual dissipation then writes from (3.11)



$$0 \leq \left\{ \rho J s \dot{\theta} - \mathbf{J}_i \text{Grad} \mu_i \right\} + \rho J \left( \mathbf{F}_a^t \cdot \frac{\partial \psi}{\partial \mathbf{F}_a} - (\psi + \mu_k) \mathbf{I} \right) : \mathbf{L}_g \quad (3.13)$$

The dissipation splits into the sum of the thermal and chemical dissipation and the intrinsic (mechanical) dissipation

$$\rho J s \dot{\theta} - \mathbf{J}_i \text{Grad} \mu_i \geq 0; \left( \mathbf{F}_a^t \cdot \frac{\partial \psi}{\partial \mathbf{F}_a} - (\psi + \mu_k) \mathbf{I} \right) : \mathbf{L}_g \geq 0 \quad (3.14)$$

From (3.14), and as a generalization of the growth models initially written in a purely mechanical context, relations (3.3) and (3.4), one is entitled to write a general growth model according to

$$\mathbf{L}_g = f(\tilde{\Sigma}_a) \quad (3.15)_1$$

with the Eshelby stress accounting for both mechanical and chemical energy contributions

$$\tilde{\Sigma}_a := \rho \mathbf{F}_a^t \cdot \frac{\partial \psi}{\partial \mathbf{F}_a} - \rho (\psi + \mu_k) \mathbf{I} \quad (3.15)_2$$

Thereby, the Eshelby stress accounts for the change of domain induced by growth; this is further reflected in the material driving force for growth, including the (material) divergence of Eshelby stress (Ganghoffer, 2010a, b). The exchange of matter is accounted for by the number of moles (with corresponding driving forces the chemical potentials), which may obey specific kinetic equations, of evolution diffusion type in a general setting (Ganghoffer, 2010a, b).

Simulations of volumetric growth based on this formalism have been done for academic situations in (Ganghoffer, 2010b). The objective of the present contribution is rather to unify volumetric and surface growth under a common umbrella, basing on the framework of Eshelby stress and material forces.

#### 4. Surface growth: A review of the thermodynamics

Surface thermodynamics is clearly a pluridisciplinary topic, which has its origins in the study of liquids, and touches various disciplines, such as metallurgy (grain boundary energy), fracture mechanics (fracture energy, mechanics (surface stress), physics of fluids (surface tension) and of solids (surface stress). Surface thermodynamic data are important parameters for specialists in each of those fields, with however a different acceptance of the term.

The thermodynamics of surfaces has a long history, tracing back to Gibbs; an interface exists when a thin inhomogeneous element of material forms a transition zone separating two phases of different materials (denoted  $\alpha, \beta$  in the sequel), as pictured in figure 2. The transition zone between the bulk phases will be denoted by the Greek letter  $\sigma$  in the sequel. The aim of this section is not to give a detailed account in each of those fields, but rather to provide the reader with a broad overview of the basic surface thermodynamics and to review the major underlying parameters and their possible source of variation.

Different viewpoints have been considered in the literature as to the geometry of the surface (this coinage used in Linford refers to the surface, as opposed to the bulk phases): the

surface phase is considered as two-dimensional by Gibbs, and coined the mathematical dividing surface, as the neat separation between fluid and solid phases. Gibbs viewpoint may be called the *surface excess approach (at fixed volume)*, in which the composite system (bulk phases and the interface) is the sum of the reference system without the interface and a correction; the difference of any quantity between the actual and the reference system leads to an *interfacial excess quantity*.

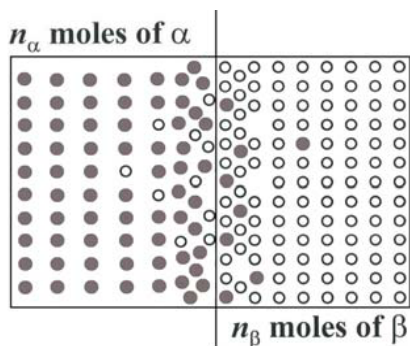


Fig. 2. Formation of an interface from a fixed number of moles of  $\alpha$  and  $\beta$ .

Important to this viewpoint is the fact that the reference and actual systems have the same volume.

Guggenheim considered the surface phase as a three dimensional body of finite small thickness, and is commonly coined the *surface phase approach*. A third approach has been introduced by Goodrich, relying on Guggenheim vision, but with the interfaces between the surface phase and the two bulk phases identified to the walls of a confining vessel. A last vision at variant with Gibbs treatment advocates that both the actual and reference systems have the same mass, but possibly different volumes: it bears the name *Surface excess approach (at fixed mass)*, and was hardly considered in the literature, although rapidly mentioned by Gibbs in 1878. One drawback of the Guggenheim model is that the volume of the interfacial region  $V^\sigma$  is arbitrary, and has nothing to do with the volume change that occurs during the formation of the interface; this difficulty is not apparent in Gibbs approach, for which the excess volume  $V^\sigma$  is always zero.

For liquids, the situation is simple, as a single scalar parameter, the surface tension, is sufficient. Three parameters are required to characterize the thermodynamics of surfaces: the reversible work to produce unit area of new surface, sometimes called the *specific surface work* (the counterpart of the surface tension in liquids), the specific surface Helmholtz energy, as the change of energy of the surface region (as opposed to the bulk phases), and the surface stress tensor, defined as the reversible work required to produced a unit area of new surface by deformation. In order to avoid some existing confusion in the early literature (this is due to the oversimplified situation that prevails for liquids), those three parameters are next introduced in a distinct manner.

The thermodynamics of surfaces is based on the setting up of *excess quantities*. The reader is referred to (Linford, 1973) and (Couchman and Linford, 1980) for more details on the topic. Hence, the excess (Helmholtz) free energy is defined through its differential

$$dF^\sigma = -S^\sigma dT + g dA + \mu_k dN_k^\sigma$$

The quantity  $g$  accounts for both the creation of new surface (with a fixed number of atoms) and the elastic deformation of the surface (also with a fixed number of atoms). The addition of atoms (particles) on the surface is accounted for by the last term. Considering two phases  $\alpha, \beta$  with a separating interface  $\sigma$  in-between, one can write the differential of the total number of particles as

$$dN = dN^\alpha + dN^\beta + dN^\sigma$$

The superficial excess or molar superficial concentrations are then defined as  $n^\sigma := N^\sigma / A$ , with  $A$  the area of the interface. Any extensive quantity  $Z$  can be decomposed as

$$Z = Z^\alpha + Z^\beta + Z^\sigma = z^\alpha V^\alpha + z^\beta V^\beta + z^\sigma A$$

with  $z^\sigma := Z^\sigma / A$  the *superficial excess* quantity. Regarding surface quantities, one makes a distinction between:

- The superficial energy  $\gamma$  ( $J / m^2$ ) - a scalar - accounting for the creation of a new surface (irreversible phenomena), with a constant number of particles.
- The purely elastic variation of the surface area, expressed by a superficial stress  $\tilde{\sigma}$ , dual to an elastic surface strain  $\tilde{\epsilon}$ .

The excess total internal energy writes

$$dU^\sigma = \theta dS^\sigma + g dA + \mu_k dN_k^\sigma$$

For the whole system, using the previous decomposition

$$Z = Z^\alpha + Z^\beta + Z^\sigma$$

one has

$$dU = \theta dS - p dV + g dA + \mu_k dN_k$$

The variation of the free energy is

$$dF = -S d\theta - p dV + g dA + \mu_k dN_k$$

Hence,  $g$  is defined as the partial derivative  $g = \left( \frac{\partial F}{\partial A} \right)_{T, V, N_i}$ . Combining both relations

$$dF^\sigma = -S^\sigma d\theta + g dA + \mu_k dN_k^\sigma ; F^\sigma = \gamma A + \mu_k N_k^\sigma$$

gives

$$S^\sigma dT + N_k^\sigma d\mu_k - (g - \gamma) dA + A d\gamma = 0$$

This leads to the differential

$$d\gamma = -s^\sigma d\theta - n_k^\sigma d\mu_k + (g - \gamma)dA / A$$

expressing the variation of the superficial energy. In the case of an isothermal surface stretch with a constant chemical potential, one gets the Couchmann-Everett formula

$$g = \gamma + A \left( \frac{\partial \gamma}{\partial A} \right)_{T, \mu_i}$$

In the case of a purely elastic stretch, previous formula specializes to the relation

$$g = \gamma + A \left( \frac{\partial \gamma}{\partial A} \right)_{T, \mu_i}^{el}.$$

The reversible work needed to form a unit area of new surface is defined at constant temperature and pressure as the partial derivative of the Gibbs free energy of the entire system (bulk phases and surface), quantity  $G(P, T, n_i, A)$ , with respect to the formed area  $A$ , at constant temperature  $T$ , pressure  $P$ , and number of moles of each component  $n_i$ , viz

$$dG = -VdP - SdT + \mu_i dN_i + \gamma dA$$

whereby the multiplicative factors of the differential elements on the right-hand side of  $dG$  are the partial derivatives

$$V = -\frac{\partial G}{\partial P}; S = -\frac{\partial G}{\partial T}; \mu_i = \frac{\partial G}{\partial N_i}; \gamma = \frac{\partial G}{\partial A}$$

The partial derivatives are evaluated with all three other variables being held fixed. The specific surface work  $\gamma$  includes two contributions, the change of Gibbs free energy per unit area for the surface region, denoted  $g^\sigma$ , and the change per unit area of surface created from the surrounding bulk phases, evaluated as the sum  $\mu_i n_i^\sigma$  over all components,

$$\gamma = g^\sigma - \mu_i n_i^\sigma$$

with  $n_i^\sigma := N_i / A$ , the surface excess of the  $i^{\text{th}}$  species. In terms of the Helmholtz energy of the whole system  $F$ , one has the similar relation involving the Helmholtz free energy per unit area  $f^\sigma$ , viz

$$\gamma = f^\sigma + Pe - \mu_i n_i^\sigma$$

with  $e$  the thickness of the surface; in most cases, the parameter  $e$  is small, and one may neglect the contribution  $Pe$ , hence one has the identification  $f^\sigma = g^\sigma$ . The last two formulas are expressions of Gibb's adsorption equation, with the derivation due to Mullins, which is next reproduced. We consider a system with  $n$  components consisting of a solid phase  $\sigma$  in contact with a fluid phase and a solid phase acting as a thermal bath at temperature  $T$  and as a chemical reservoir for each component; accordingly, the components concentrations can

be adjusted to maintain the chemical potentials at fixed specified values  $\mu_i$ . Imagine a modification of the temperature by  $dT$ , and of the  $i^{\text{th}}$  chemical potential by  $d\mu_i$ , at fixed surface area;  $dn_i$  particles from the bulk will enter the solid phase  $\sigma$  from the bulk, and the change of Helmholtz free energy  $F^\sigma$  will be

$$dF^\sigma = -S^\sigma dT + (\mu_i + d\mu_i)dN_{i,\sigma} \approx -S^\sigma dT + \mu_i dN_{i,\sigma}$$

with  $S^\sigma$  the entropy of the phase  $\sigma$ . Consider next a new system for which  $T$  and  $\mu_i$  are returned to their initial values, but with the surface area increased by  $dA^\sigma$ , and modify thereafter the temperature by  $dT$ , and the  $i^{\text{th}}$  chemical potential by  $d\mu_i$ ; the variation of free energy of this system of larger area is

$$dF'^\sigma \approx -S'^\sigma dT + \mu_i dN'_{i,\sigma}$$

Subtracting both variations of Helmholtz free energy by unit surface gives

$$\frac{dF'^\sigma - dF^\sigma}{dA^\sigma} \approx -\frac{(S'^\sigma - S^\sigma)}{dA^\sigma} dT + \mu_i \frac{d(N'_{i,\sigma} - N_{i,\sigma})}{dA^\sigma}$$

Introducing therein the definitions of the specific surface Helmholtz energy  $f^\sigma := (dF'^\sigma - dF^\sigma) / dA^\sigma$ , the specific surface entropy  $s^\sigma := (S'^\sigma - S^\sigma) / dA^\sigma$ , and the surface excess  $n_i^\sigma := d(N'_{i,\sigma} - N_{i,\sigma}) / dA^\sigma$  leads to

$$df^\sigma = -s^\sigma dT + \mu_i d\Gamma_i$$

But one can also express the specific surface Helmholtz energy as  $f^\sigma = \gamma + \mu_i \Gamma_i$ , hence

$$df^\sigma = d\gamma + \mu_i d\Gamma_i + \Gamma_i d\mu_i = -s^\sigma dT + \mu_i d\Gamma_i$$

and thus finally

$$d\gamma = -s^\sigma dT - \Gamma_i d\mu_i$$

The same identity was derived by (Goodrich, 1969) for a one-component system using the method of Lagrange multipliers. The reversible work needed to generate a unit area of new surface by stretching at constant pressure and temperature represents the surface stress tensor, denoted  $\tilde{\sigma}_{ij}$ . It is related to  $\gamma$  by

$$\tilde{\sigma}_{ij} = \gamma \delta_{ij} + \frac{\partial \gamma}{\partial \tilde{\varepsilon}_{ij}}$$

The second order tensor  $\tilde{\varepsilon}_{ij}$  is the strain (a small perturbation scheme is presently adopted) induced by the component  $\tilde{\sigma}_{ij}$  acting in the  $j^{\text{th}}$  direction per unit length of the edge normal

to the  $i^{\text{th}}$  direction, with both indices  $i, j$  lying in the plane of the surface. Previous equation is valid for an anisotropic solid, and reduces in the case of an isotropic surface to the previously written Couchmann-Everett formula, with  $g$  half the trace of the surface stress tensor. The proof of previous formula follows Mullins derivation: let imagine a unit cube with edges parallel to the axes  $x, y, z$ , and perform two distinct operations on it:

- i. Stretch the cube reversibly along the  $x$  axis by an amount  $dx$ , with the  $y$  edge fixed, but allowing the edge  $z$  to vary its length. The surface in the  $xz$  plane may then change by an inflow (or outflow) of material from the bulk, increasing (or decreasing), the cube height; denote  $W_0$  the work expanded in this transformation. Let next separate the stretched cube along the  $xy$  plane, requiring the work  $W_2 = 2(\gamma + d\gamma)(1 + dx)$ , with  $d\gamma$  the variation of the specific surface work  $\gamma$  due to the stretch  $dx$  (factor 2 arises since two surfaces are created, and the factor  $(1 + dx)$  since the specific surface work applies per unit surface area).
- ii. Separate the original unit cube into two parts along an  $xy$  plane, requiring the work  $W_3 = 2\gamma$ , and stretch each half by  $dx$  in the  $x$  direction, at fixed  $y$  edge, but varying  $z$  edge. Let  $W_1$  be the work expanded in this operation. The final configuration is the same as that obtained in the first process, hence the same total work has been expanded, hence  $W_0 + W_2 = W_1 + W_3$ , viz

$$W_0 + 2(\gamma + d\gamma)(1 + dx) = 2\gamma + W_1$$

The difference  $W_1 - W_0$  is the work due to the stretching operations of both processes, and can be equalized to the  $x$ -component  $\tilde{\sigma}_{xx}$  of a force in the newly formed surface times the distance  $2dx$  through which this force acts, hence

$$2\tilde{\sigma}_{xx}dx = W_1 - W_0$$

The strain  $\Delta\varepsilon_{xx} = dx$  (since the other side has unit length), hence

$$2\gamma + 2\gamma\Delta\varepsilon_{xx} + 2\Delta\gamma + 2\Delta\gamma\Delta\varepsilon_{xx} = 2\gamma + 2\tilde{\sigma}_{xx}\Delta\varepsilon_{xx}$$

Due to the equalities

$$\Delta\gamma\Delta\varepsilon_{xx} \approx 0; \Delta\gamma / \Delta\varepsilon_{xx} \approx d\gamma / d\varepsilon_{xx}$$

it finally results

$$\tilde{\sigma}_{xx} = \gamma + \frac{d\gamma}{d\varepsilon_{xx}}$$

Similar analogous processes with the stretching replaced by shear lead to the relation (Linford, 1973)

$$\tilde{\sigma}_{xy} = \frac{d\gamma}{d\varepsilon_{xy}}$$

Combining the stretch and shear processes then lead to the expression of the surface stress tensor

$$\tilde{\sigma}_{ij} = \gamma \delta_{ij} + \frac{d\gamma}{d\varepsilon_{ij}}$$

represented by a 2 by 2 symmetrical matrix (3 independent components). For an isotropic material or a crystal with a threefold (or greater axis of symmetry), it follows as shown by Shuttleworth (using the principle of virtual work) the isotropic surface stress

$$\tilde{\sigma}_{ij} = g \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Lastly, consider a square section in the xy plane of side  $(A^\sigma)^{1/2}$  and imagine an extension of the x edge by  $\varepsilon_{xx}(A^\sigma)^{1/2}$ ; the required work is  $W_1 = \tilde{\sigma}_{xx}(A^\sigma)^{1/2} \varepsilon_{xx}(A^\sigma)^{1/2}$ . Extend next the y edge by  $\varepsilon_{yy}(A^\sigma)^{1/2}$ , with an expanded work given by

$$W_2 = \tilde{\sigma}_{yy}(A^\sigma)^{1/2} (1 + \varepsilon_{xx}) \varepsilon_{yy}(A^\sigma)^{1/2}$$

Assuming the deformation is reversible and isothermal, the total work spent is the variation of surface energy, which expresses for a high symmetry isotropic crystal as

$$d(A^\sigma \gamma) = W_1 + W_2 = \tilde{\sigma}_{xx} \varepsilon_{xx} A^\sigma + \tilde{\sigma}_{yy} (1 + \varepsilon_{xx}) \varepsilon_{yy} A^\sigma \approx g dA^\sigma$$

due to the equality  $A^\sigma (\varepsilon_{xx} + \varepsilon_{yy}) = dA^\sigma$ . Therefore, one has

$$g dA^\sigma = \gamma dA^\sigma + A^\sigma d\gamma \Rightarrow g = \gamma + \frac{d\gamma}{dA^\sigma}$$

Note that the last term vanishes for liquids; as a corollary, liquid films can easily be stretched since atoms can move from the bulk to the surface without additional energy costs. The opposite situation prevails for solids, as they shear and their structure changes with an overall additional energy contribution.

The Gibbs approach towards interfacial excess quantities is as previously mentioned valid only at fixed volume; a parallel approach that is valid at fixed mass instead has been developed in (Muller and Kern, 2001), which is next exposed. The bulk phases  $\alpha, \beta$  are initially separated and interface-free, and are in a thought experiment imagined to be joined along a plane to generate the  $\alpha / \beta$  interface. Since mass is conserved, any change in the thermodynamic quantities of the whole system are due to the new  $\alpha / \beta$  interface, coined *excess values* of the corresponding quantities, denoted with a subscript  $\gamma$  to distinguish them from Gibbs approach at fixed volume. The differential of the Gibbs energy of the system before and after formation of the interface successively writes (for a constant number of molecules)

$$dG_1 = (V_\alpha + V_\beta)dP - (S_\alpha + S_\beta)dT ; dG_2 = (V_\alpha + V_\beta + V_\gamma)dP - (S_\alpha + S_\beta + S_\gamma)dT + \gamma^* dA$$

with  $\gamma^*$  the *reorganization surface energy*, although commonly referred to as the interfacial tension in the literature; it is a mechanical positive quantity, that may depend upon interface curvature. Note that the number of atoms is the same in the reference and final states, in contrast with Gibbs approach. Hence, the variation of the excess Gibbs free energy between states 1 and 2 for the fixed masses  $m_\alpha, m_\beta$  is

$$dG_\gamma = dG_2 - dG_1 = V_\gamma dP - S_\gamma dT + \gamma^* dA$$

which may be interpreted from an energetic point of view as follows: the term  $V_\gamma dP$  is the mechanical work done against the external force field, the contribution  $S_\gamma dT$  represents the heat of formation of the interface, and  $\gamma^* dA$  is the mechanical work done against the internal force field of both phases  $\alpha, \beta$  by motion of the molecules from the bulk to generate a new interface. The *excess free energy* of formation of the interface, potential  $G_\gamma$ , is the additional free energy required to form the interface from fixed masses of the pre-existing bulk phases  $\alpha, \beta$ . The above equations implicitly use the conservation of mass, equation

$$n_{total} = n_\alpha + n_\beta$$

and the definition of the excess interfacial volume  $V_\gamma$  from the contributions to the total volume after interface formation (balance law for the volume)

$$V_{total} = V_\alpha + V_\beta + V_\gamma$$

In contrast to this treatment, Gibbs assumes a conservation of the total volume as  $V_{total} = V_\alpha + V_\beta$ , but with addition of the new mass  $n^\sigma$  such that

$$n_{total} = n_\alpha + n_\beta + n^\sigma$$

As a compensation for the volume change accompanying the formation of the interface; hence,  $n^\sigma$  is a supply of material from outside the system, with the sense that the Gibbs volume is an open thermodynamic volume.

Due to its status as a state function, the previous differential OF  $G_\gamma$  allows writing relations between partial derivatives as the analogues of the bulk phase Maxwell relations

$$\left(\frac{\partial \gamma^*}{\partial T}\right)_{P, A, n_{\alpha, \beta}} = -\left(\frac{\partial S}{\partial A}\right)_{P, T, n_{\alpha, \beta}} = -S_\gamma^* ; \left(\frac{\partial \gamma^*}{\partial P}\right)_{T, A, n_{\alpha, \beta}} = \left(\frac{\partial V^*}{\partial A}\right)_{P, T, n_{\alpha, \beta}} = V_\gamma \cdot$$

$$\left(\frac{\partial V_\gamma}{\partial T}\right)_{P, A, n_{\alpha, \beta}} = -\left(\frac{\partial S}{\partial P}\right)_{A, T, n_{\alpha, \beta}}$$

The introduced quantities  $S_\gamma^*, V_\gamma$  are respectively the *interfacial excess entropy* and the *specific interfacial excess volume*; the compact notation  $n_{\alpha, \beta}$  stands for the two quantities  $\{n_\alpha, n_\beta\}$ .



The specific interfacial excess energy is obtained by simply integrating the differential  $dG_\gamma = V_\gamma dP - S_\gamma dT + \gamma^* dA \Rightarrow (G_\gamma^*)_{n_{\alpha,\beta}} = \gamma^*$  at constant pressure and temperature, introducing the specific interfacial excess energy  $G_\gamma^* := G_\gamma / A$ . Last relation implies that the temperature and pressure dependence of  $\gamma^*$  can be determined from those of  $G_\gamma^*$ . The specific interfacial excess energy is obtained from a Legendre transform to  $dG_\gamma = V_\gamma dP - S_\gamma dT + \gamma^* dA$  and substitution of the previous interfacial Maxwell relations, thus

$$(U_\gamma^*)_{n_{\alpha,\beta}} = \gamma^* + TS_\gamma^* - PV_\gamma^*$$

It immediately results the specific interfacial excess enthalpy

$$(H_\gamma^*)_{n_{\alpha,\beta}} = \gamma^* + TS_\gamma^* \equiv \gamma^* - T \left( \frac{\partial \gamma^*}{\partial T} \right)_{P, A, n_{\alpha,\beta}}$$

with  $H_\gamma^*$  identified as the surface energy, which is the sum of the interfacial tension and the heat supplied by the surrounding for an isothermal creation of new interface. The advantages of this last approach in comparison to Gibbs treatment is that it leads to non-nil interfacial volumes, analogues of the Maxwell relations for bulk phases can be derived, and the temperature and pressure dependence of the interfacial tension can be accessed from a comparison between simple formulae and experiments.

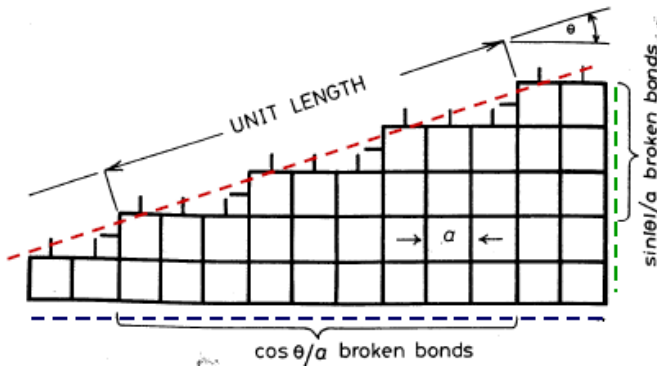


Fig. 3. The broken bond model for surface energy

The reversible work to form new surface area, parameter  $\gamma$ , is for a solid generally orientation dependent, although not for a liquid. This surface energy parameter has been up to now considered under the thermodynamic continuum viewpoint; we next examine two other viewpoints, the atomistic approach and Wulff plot. The atomistic approach considers the interaction between atoms to calculate the surface energy; arrangement of atoms in crystals are such that one can order atoms according to the energy required to remove atoms from the bulk: first nearest neighbours requiring more energy compared to second and third nearest neighbours. For a crystal lattice presenting dislocations, the number of broken bonds



with anisotropic surface tension  $\gamma(\mathbf{n})$

### 5. Model of surface growth with application to bone remodeling

The present model aims at describing radial bone remodeling, accounting for chemical and mechanical influences from the surrounding. Our approach of bone growth typically follows the streamlines of continuum mechanical models of bone adaptation, including the time-dependent description of the external geometry of cortical bone surfaces in the spirit of free boundary value problems – a process sometimes called net ‘surface remodeling’ - and of the bone material properties, sometimes coined net ‘internal remodeling’ (Cowin, 2001).

#### 5.1 Material driving forces for surface growth

In the sequel, the framework for surface growth elaborated in (Ganghoffer, 2010) will be applied to describe bone modeling and remodeling. As a prerequisite, we recall the identification of the driving forces for surface growth. We consider a tissue element under grow submitted to a surface force field  $\mathbf{f}_S$  (surface density) and to line densities  $p_\tau, p_\nu$  defined as the projections onto the unit vectors  $\boldsymbol{\tau}_g, \mathbf{v}_g$  resp. along the contour of the open growing surface  $S_g$  (Figure 5); hence, those line densities are respectively tangential and normal to the surface  $S_g$  (forces acting in the tangent plane).

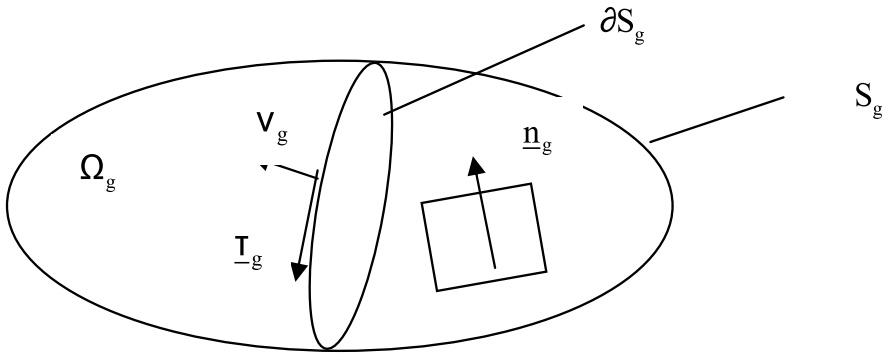


Fig. 5. Tissue element under growth: elements of differential geometry.

Focusing on the surface behavior, the potential energy of the growing tissue element is set as the expression

$$\begin{aligned}
 V = & \int_{\Omega_g} W_0(\mathbf{F}) dx_g + \int_{S_g} \psi^S(\tilde{\mathbf{F}}, \mathbf{N}; \mathbf{X}_S) d\sigma_g + \int_{S_g} \mu_k n_k^\sigma d\sigma_g \\
 & - \int_{S_g} \mathbf{f}_S \cdot \tilde{\mathbf{x}} d\sigma_g - \int_{\partial S_g} p_\tau \tilde{\mathbf{x}} \cdot \boldsymbol{\tau}_g dl_g - \int_{\partial S_g} p_\nu \tilde{\mathbf{x}} \cdot \mathbf{v}_g dl_g
 \end{aligned}
 \tag{5.1}$$

Thereby, the growing solid surface is supposed to be endowed with a volumetric density  $W_0(\mathbf{F})$  depending upon the transformation gradient  $\mathbf{F} := \nabla_{\mathbf{X}} \mathbf{x}$ , a surface energy with density  $\psi^S(\tilde{\mathbf{F}}, \mathbf{N}; \mathbf{X}_S)$  per unit reference surface, depending upon the surface gradient  $\tilde{\mathbf{F}}$ , the unit normal vector  $\mathbf{N}$  to  $S_g$ , and possibly explicitly upon the surface position vector

$\mathbf{X}_S$  on  $S_g$  (no tilde notation is adopted here since the support of  $\mathbf{X}_S$  is strictly restricted to the surface  $S_g$ ), and chemical energy  $\mu_k n_k^\sigma$ , with  $\mu_k$  the chemical potential of the surface concentration of species  $n_k^\sigma$ . The surface gradient  $\tilde{\mathbf{F}}$  maps material lengths (or material tangent vectors) onto the deformed surface; it is elaborated as the surface projection of  $\mathbf{F}$  (onto the tangent plane to  $\Omega_a$ ), viz

$$\tilde{\mathbf{F}} := \mathbf{F} \cdot \mathbf{P}$$

The tissue element under grow is submitted to a surface force field  $\mathbf{f}_S$  (surface density) and to line densities  $p_\tau, p_\nu$  defined as the projections onto the unit vectors  $\boldsymbol{\tau}_g, \boldsymbol{\nu}_g$  resp. along the contour of the open growing surface  $S_g$  (Figure 5); Hence, those line densities are respectively tangential and normal to the surface  $S_g$  (forces acting in the tangent plane).

The variation of the previously built potential energy of the growing tissue element  $V$  is next evaluated, assuming applied forces act as dead loads, using the fact that the variation is performed over a changing domain (Petryk and Mroz, 1986), here the growing surface  $S_g$ . We refer to the recent work in (Ganghoffer, 2010a) giving the detailed calculation of the material forces for surface growth, very similar to present developments.

The variation of the volumetric term (first term on the right hand side of  $\delta V$ ) can be developed from the equalities (A2.1) through (A2.3) given in (Ganghoffer, 2010a, Appendix 2):

$$\delta \left( \int_{\Omega_g} W_0(\mathbf{F}, \mathbf{X}_g) dx_g \right) = \int_{\partial\Omega_g} (\boldsymbol{\Sigma} \cdot \delta \mathbf{X}_g + \mathbf{p} \cdot \delta \mathbf{x}) \cdot \mathbf{N} d(\partial\Omega_g) + v.t. \quad (5.2)$$

with volumetric terms denoted as ‘v.t.’ that will not be expressed here, as we are mostly interested in surface growth. The r.h.s. in previous identity is a pure surface contribution involving the volumetric Eshelby stress built from the volumetric strain energy density and the so-called canonical momentum

$$\boldsymbol{\Sigma} := W_0 \mathbf{I} - \mathbf{F}^t \cdot \mathbf{p} \quad \mathbf{p} := \frac{\partial W_0}{\partial \nabla x} \quad (5.3)$$

As we perform material variations over an assumed fixed actual configuration, the contribution of the canonical momentum vanishes ( $\delta \mathbf{x} = \mathbf{0}$ ). Observe that the volumetric Eshelby stress  $\boldsymbol{\Sigma}$  triggers surface growth in the sense of the boundary values taken by the normal Eshelby-like traction  $\boldsymbol{\Sigma} \cdot \mathbf{N}$ . The variation of the surface energy contribution  $\psi^S$  can be expanded using the surface divergence theorem (equality (3.15) in Ganghoffer, 2010a) as

$$\delta \left( \int_{S_g} \psi^S(\tilde{\mathbf{F}}, \mathbf{N}; \mathbf{X}_S) d\sigma_g \right) = \int_{S_g} \left[ \nabla_S \cdot \tilde{\boldsymbol{\Sigma}} - \boldsymbol{\Pi} \mathbf{K}^t \cdot \partial_N \psi^S + (\partial_{X_S} \psi^S)_{\text{expl}} + \tilde{\mathbf{F}}^T \cdot \mathbf{f}_S \right] \cdot \delta \mathbf{X}_S d\sigma_g \quad (5.4)$$

The surface energy momentum tensor (of Eshelby type) is then defined as the second order tensor

$$\tilde{\mathbf{T}} := \partial_{\tilde{\mathbf{F}}} \psi^S \rightarrow \tilde{\boldsymbol{\Sigma}} := \tilde{\mathbf{F}}^T \cdot \tilde{\mathbf{T}} - \psi^S \mathbf{I}_S \quad (5.5)$$

basing on the *surface stress*  $\tilde{\mathbf{T}}$ . The Lagrangian curvature tensor is defined as  $\mathbf{K} := -\nabla_R \mathbf{N}$ . The chemical potential as the partial derivative of the surface energy density with respect to the superficial concentration

$$\mu_k := \frac{\partial \psi^S}{\partial n_k^\sigma} \Big|_{X,F,N} \equiv \mu_k(n_k^\sigma) \tag{5.6}$$

The contributions arising from the domain variation due to surface growth are considered as irreversible.

The material surface driving force (for surface growth) triggers the motion of the surface of the growing solid; it is identified from the material variation of  $V$  as the vector acting on the variation of the surface position

$$\tilde{\mathbf{Y}}_g := \boldsymbol{\Sigma} \cdot \mathbf{N} + \nabla_S \cdot \tilde{\boldsymbol{\Sigma}} - \mathbf{P} \cdot \mathbf{K}^t \cdot \partial_N \psi^S + \mu_k \nabla_S n_k^\sigma - \mathbf{f}_S \tag{5.7}$$

itself built from the *surface stress*  $\tilde{\mathbf{T}} := \partial_{\tilde{\mathbf{F}}} \psi^S$ , and on the curvature tensor  $\mathbf{K} := -\nabla_R \mathbf{N}$  in the referential configuration.

**5.2 Bone remodeling**

Bone is considered as a homogeneous single phase continuum material; from a microstructural viewpoint, bone consists mainly of hydroxyapatite, a type-I collagen, providing the structural rigidity. The collageneous fraction will be discarded, as the mineral carries most of the strain energy (Silva and Ulm, 2002). The ultrastructure may be considered as a continuum, subjected to a portion of its boundary to the chemical activity generated by osteoclasts, generating an overall change of mass of the solid (the mineral fraction) given by

$$\frac{d}{dt} \int_{\Omega_g} \rho_g dx_g = \int_{S_g} \rho_g \mathbf{V}_S \cdot \mathbf{N} d\sigma_g$$

The quantity  $\rho_g \mathbf{V}_S \cdot \mathbf{N} d\sigma_g$  therein represents the molar flux of bone material being dissolved, hence

$$\rho_g V_N d\sigma_g = M J d\sigma_g \tag{5.8}$$

with  $V_N$  the normal surface velocity,  $M$  the bone mineral molar mass, and  $J \equiv \rho_g V_N / M$  the molar influx of minerals (positive in case of bone apposition, and negative when resorption occurs). Clearly, the previous expression shows that the knowledge of the normal surface growth velocity determines the molar influx of minerals. Estimates of the order of magnitude of the dissolution rate given in (Christoffersen at al., 1997), for a pH of 7.2 (although much higher compared to the pH for which bone resorption takes place) and at a temperature of 310K, are indicative of values of the molar influx in the interval  $J \in [10^{-9}, 1.8 \cdot 10^{-8}] \text{ mol} \cdot \text{s}^{-1} \cdot \text{m}^{-2}$ . The osteoclasts responsible for bone resorption attach to the bone surface, remove the collageneous fraction of the material by transport phenomena, and diffuse within the material. This osteoclasts activity occurs at a typical scale of about  $50 \mu\text{m}$ ,

which is much larger compared to the characteristic size of the ultrastructure; the resorption phase takes typically 21 days (the complete remodeling cycle lasts 3 months). The osteoclasts, generate an acid environment causing simultaneously the dissolution of the mineral - hydroxyapatite, a strong basic mineral  $[Ca_3(PO_4)_2]_3 Ca(OH)_2$ , abbreviated HA in the sequel - and the degradation of the collagenous fraction of the material. The metabolic processes behind bone remodeling are very complicated, with kinetics of various chemical substances, see (Petryl and Danesova, 1999).

The pure chemical driving force represents the difference of the chemical potential externally supplied  $\mu_e$  (biochemical activity generated by the osteoclasts) with the chemical potential of the mineral of the solid phase, denoted  $\mu_{\min}$ ; it can be estimated from the change of activity of the  $H^+$  cation (Silva and Ulm, 2002):

$$\Delta\mu := \mu_e - \mu_{\min} = R\theta \ln \frac{[H^+]_{eq}^2}{[H^+]_{ex}^2} \quad (5.9)$$

This chemical driving force is the affinity conjugated to the superficial concentration of minerals, denoted  $n^\sigma(t)$  in the sequel. The conversion to mechanical units  $\Delta\mu$  is done, considering a density of HA  $\rho = 3000 \text{ kg} / \text{m}^3$  (5.1), hence  $(\rho / M)\Delta\mu = -20 \text{ MPa}$ , according to (Silva and Ulm, 2002); the negative value means that the dissolution of HA is chemically more favorable (bone resorption occurs).

Relying on the biochemical description given thereabove, bone remodeling is considered as a pure surface growth process. In order to analyze the influence of mechanical stress on bone remodeling, a simple geometrical model of a long bone as a hollow homogeneous cylinder is introduced, endowed with a linear elastic isotropic behavior (the interstitial fluid phase in the bone is presently neglected). This situation is representative of the diaphysal region of long bones (Cowin and Firozbakhsh, 1981), such as the human femur (figure 6).

According to experiments performed by (Currey, 1988), the elastic modulus is assumed to scale uniformly versus the bone density according to

$$E = E_{\max} \rho_S(t)^p \quad (5.10)$$

with  $\rho_S(t)$  the surface density of mineral,  $E_{\max} = 15 \text{ GPa}$  (Reilly and Burstein, 1975) the maximum value of the tensile modulus, and  $p$  a constant exponent, here taken equal to 3 (Currey, 1988; Ruinerman et al., 2005).

Following the representation theorems for isotropic scalar valued functions of tensorial arguments, the surface strain energy density  $\psi_{mech}^S(\tilde{\mathbf{F}}, \mathbf{N}; \mathbf{X}_S)$  of mechanical origin is selected as a function of the curvature tensor invariants, viz the mean and Gaussian curvatures, the invariants of the surface Cauchy-Green tensor  $\tilde{\mathbf{C}} := \tilde{\mathbf{F}}^t \cdot \tilde{\mathbf{F}}$  and of its square. The following simple form depending on the second invariant of the linearized part of  $\tilde{\mathbf{C}} \cong \mathbf{I} + 2\tilde{\boldsymbol{\varepsilon}}$  is selected, adopting the small strain framework, viz, hence

$$\psi_{mech}^S(\tilde{\boldsymbol{\varepsilon}}) = \frac{A}{2} \text{Tr}(\tilde{\boldsymbol{\varepsilon}})^2 + B(\tilde{\boldsymbol{\varepsilon}} : \tilde{\boldsymbol{\varepsilon}}) \quad (5.11)$$

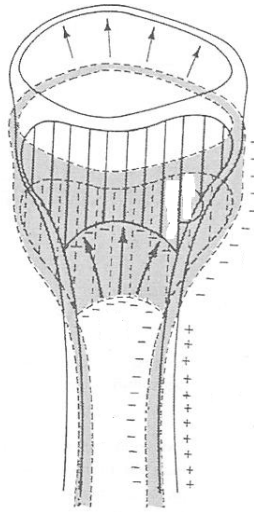


Fig. 6. Modeling occurring during growth of the proximal end of the femur. Frontal section of the original proxima tibia is indicated as the stippled area. The situation after a growth of 21 days is superimposed. Bone formation (+) and bone resorption zones indicated [Weiss, 1988].

with  $\tilde{\epsilon} = \mathbf{P}\cdot\boldsymbol{\epsilon} = \mathbf{I}_S\cdot\boldsymbol{\epsilon} - \boldsymbol{\epsilon}\cdot(\mathbf{e}_r \otimes \mathbf{e}_r)$  the surface strain (induced by the existing volumetric strain), and  $A, B$  mechanical properties of the surface, expressing versus the surface density of minerals and the maximum value of the traction modulus as (the Poisson ratio is selected as  $\nu = 0.3$ )

$$A = \frac{E_{\max}\rho_S(t)^3 \nu}{(1 - 2\nu)(1 + \nu)}; B = \frac{E_{\max}\rho_S(t)^3}{2(1 + \nu)} \tag{5.12}$$

As the surface of bone undergoes resorption, its mechanical properties are continuously changing from the bulk behavior, due to the decrease of mineral density as reflected in (5.10). The surface stress results from (5.11), (5.12) as

$$\tilde{\mathbf{T}} \equiv \tilde{\boldsymbol{\sigma}} := \frac{\partial \psi_{\text{mech}}^S}{\partial \tilde{\boldsymbol{\epsilon}}} = A\tilde{\boldsymbol{\epsilon}} + 2B\text{tr}(\tilde{\boldsymbol{\epsilon}})\mathbf{I}_S \tag{5.13}$$

The unknowns of the remodeling problem are the normal velocity of the bone surface  $V_N(t)$ , the surface density of minerals  $\rho_S(t)$  and its superficial concentration. We shall herewith simulate the resorption of a hollow bone submitted to a composite applied stress, consisting of the superposition of an axial and a radial component, as

$$\boldsymbol{\sigma} = \sigma_{rr}\mathbf{e}_r \otimes \mathbf{e}_r + \sigma_{zz}\mathbf{e}_z \otimes \mathbf{e}_z \tag{5.14}$$

in the cylindrical basis  $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_z)$ ; this applied stress generates a preexisting homogeneous stress state within the bulk material, inducing a surface stress given by

$$\tilde{\mathbf{o}} = \mathbf{P} \cdot \boldsymbol{\sigma} = \sigma_{zz} \mathbf{e}_z \otimes \mathbf{e}_z$$

The radial component of Eshelby stress  $\Sigma_{rr}$  is then easily evaluated from the preexisting homogeneous stress state. Straightforward calculations deliver then the driving force for surface remodeling, as the sum of a chemical and a mechanical contribution due to the applied axial stress:

$$\tilde{Y}_{gN} = \frac{1}{r_i(t)} \left\{ \frac{1}{8} \Delta \mu n^\sigma(t) + \frac{A + 2B}{8(A+B)B} \sigma_{zz}^2 \right\} \quad (5.15)$$

with the material coefficients  $A, B$  given in (5.12), and the axial stress  $\sigma_{zz}$  possibly function of time. A simple linear relation of the velocity of the growing surface to the driving force is selected, viz

$$V_N(t) = C \tilde{Y}_{gN}(t) \quad (5.16)$$

with  $C$  a positive parameter; the positive sign is due to the velocity direction being opposite to the outer normal (the inner radius is increasing). The chemical contribution leads by itself to resorption, hence the normal velocity has to be negative; the mechanical contribution in (5.15) brings a positive contribution to the driving force for bone growth, corresponding to apposition of new bone when the neat balance of energy is favorable to bone growth. An estimate of the amplitude of the normal velocity is given from the expression of the rate of dissolution of HA in (5.8) as

$$J = \rho_g V_N / M = 10^{-8} \text{ mol} \cdot \text{s}^{-1} \cdot \text{m}^{-2} \Rightarrow V_N = JM / \rho_g \approx 3.3 \cdot 10^{-12} \text{ m} / \text{s} \approx 0.286 \mu\text{m} / \text{day}$$

selecting a molar mass  $M \approx 1.004 \text{ kg} / \text{mol}$ , following (Silva and Ulm, 2002). This value is an initial condition for the radius evolution (its rate is prescribed), leading to  $C = 3.5 \cdot 10^{-23} \text{ m}^2 \cdot \text{kg}^{-1} \cdot \text{s}$ ; it is however much lower compared to typical values of the bulk growth velocity, about  $10 \mu\text{m} / \text{day}$ .

The mass balance equation for the surface density of minerals  $\rho_s$  writes

$$\dot{\rho}_s + \rho_s \nabla_s \cdot \tilde{\mathbf{V}} = \Gamma^S \rho_s \quad (5.17)$$

expressing as the following conservation law

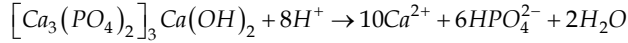
$$\frac{\dot{\rho}_s}{\rho_s} - \frac{V_N}{r_i(t)} = \Gamma_0^S \Leftrightarrow \rho_s(t) = \rho_s^0 \frac{r_0}{r_i(t)} \exp(\Gamma_0^S t) \quad (5.18)$$

The initial surface density of minerals  $\rho_s^0 = \rho_s(t_0)$ , is evaluated from the bulk density of HA, viz  $3000 \text{ kg} / \text{m}^3$ , and the estimated thickness of the attachment region of osteoclasts, about  $7 \mu\text{m}$  (Blair, 1998), hence  $\rho_s^0 \approx 2.1 \cdot 10^{-2} \text{ kg} / \text{m}^2$ .

The surface growth rate of mass  $\Gamma_0^S$  is here assumed to be constant (it represents a datum) and can be identified to the rate of dissolution of HA, adopting the chemical reaction model of (Blair, 1998):  $\Gamma_0^S$  is estimated by considering that 80% of the superficial minerals have been dissolved in a 2 months period, hence  $\Gamma_0^S \approx -2.2 \cdot 10^{-7} \text{ s}^{-1}$ . The dissolution of HA is in



reality a rather complex chemical reaction (Blair, 1998) that is here simply modeled as a single first order kinetic reaction



The kinetic equation is chosen as:

$$\frac{\partial n^\sigma(t)}{\partial t} = -\tilde{\gamma}\rho_s(t)n^\sigma(t) = -\tilde{\gamma}r_i(t)\frac{\rho_s^0}{r_0}\exp(\Gamma_0^S t)n^\sigma(t) \quad (5.19)$$

incorporating the density of minerals. The rate coefficient of dissolution of HA, namely the parameter  $\tilde{\gamma}$ , is taken at room temperature from literature values available for CHA (carbonated HA, similar to bone), viz  $\tilde{\gamma} \approx 2.2 \cdot 10^{-4} s^{-1}$  (Hankermeyer et al., 2002).

### 5.3 Simulation results

The present model involves a dependency of the triplet of variables  $\{r_i(t), \rho_s(t), n^\sigma(t)\}$  solution of the set of equations (5.15) through (5.19) on a set of parameters, arising from initial conditions satisfied by those variables:

- The initial concentration of minerals  $n_0^\sigma$  is taken as unity, viz  $n_0^\sigma = 1 \text{ mol.m}^{-3}$ .
- The initial radius  $r_0 := r_i(0)$  is estimated as  $r_0 = 1.6 \text{ cm}$  for the diaphysis of the human femur (Huiskes and Sloof, 1981). The evolution versus time of the internal radius obtained by time integration of the normal velocity expressed in (5.16).

The evolution versus time of some variables of interest is next shown, considering a time scale conveniently expressed in days. Numerical simulations of bone resorption are to be performed for three stress levels in the normal physiological range,  $\sigma \in \{1MPa, 2MPa, 5MPa\}$ . The surface velocity (Figure 7) shows an acceleration of the resorption process with time, which is enhanced by the stress level, as expected from the higher magnitude of the driving force.

The density and concentration vanish over long durations, meaning that the bone has been completely dissolved (Figure 8).

An order of magnitude of the simulated radial surface velocity is about  $10 \mu\text{m} / \text{day}$  for a stress level of 1MPa (Cowin, 2001). The superficial density of minerals and its concentration are both weakly dependent upon stress; the density of minerals decreases by a factor two (for low stresses; the resorption is enhanced by the applied stress) over a period of one month resorption period.

Considering an imposed stress function of time, the surface driving force is seen to vanish for a critical stress  $\sigma_{zz}^{crit}(t)$ , depending upon the density and concentration, given from (5.18), (5.19) as

$$\sigma_{zz}^{crit}(t) \approx 9.4 \cdot 10^{10} \rho_s(t)^{3/2} n^\sigma(t)^{1/2} \quad (5.16)$$

This expression gives an order of magnitude of the stress level above which bone apposition (growth) shall take place; when the critical stress is reached, the chemical and mechanical driving forces do balance, and the bone microstructure is stable.

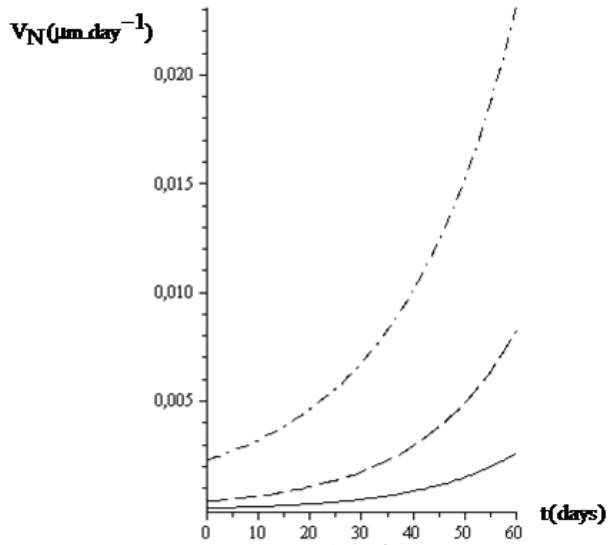


Fig. 7. Evolution vs. time of the surface growth velocity for three stress levels:  $\sigma_{zz} = 1\text{ MPa}$  (thick line),  $\sigma_{zz} = 2\text{ MPa}$  (dashed line),  $\sigma_{zz} = 5\text{ MPa}$  (dash-dotted line).

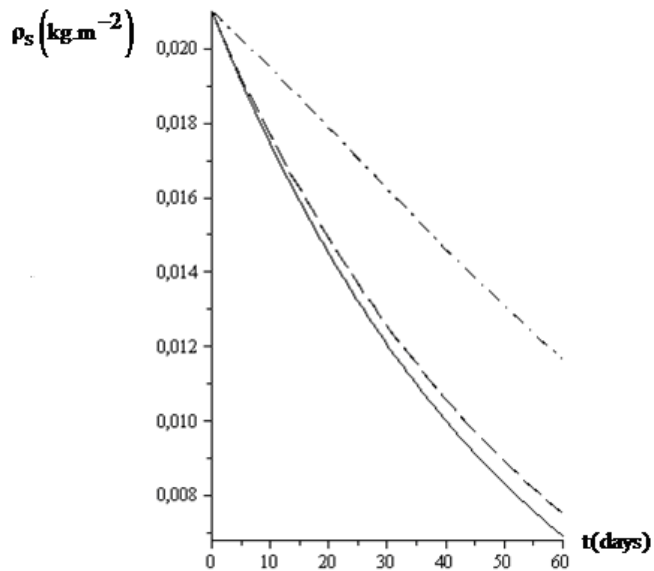


Fig. 8. Evolution of the superficial density of HA versus time for three stress levels.  $\sigma_{zz} = 1\text{ MPa}$  (thick line),  $\sigma_{zz} = 2\text{ MPa}$  (dashed line),  $\sigma_{zz} = 5\text{ MPa}$  (dash-dotted line).

For an applied stress  $\sigma_{zz} = 0.2MPa$  lying slightly above the critical stress expressed in (5.16), growth will occur due to mineralization (the chemical driving force in (5.9) favors apposition of new bone on the surface), as reflected by the simulated decrease of the internal radius over the first week (Figure 9).

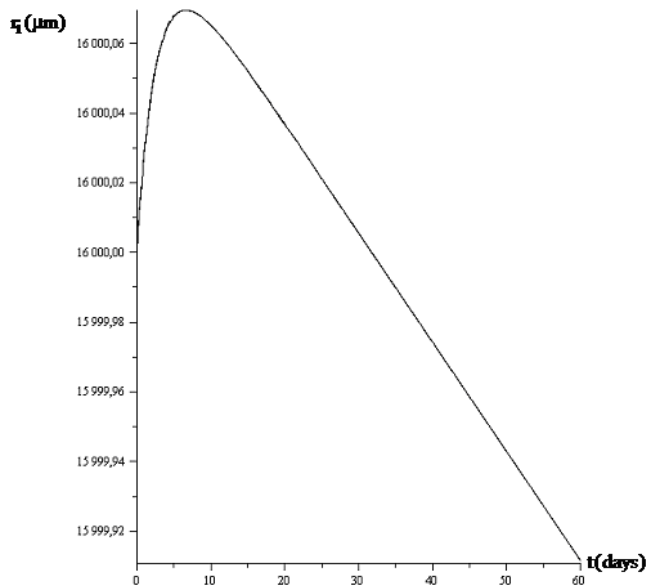


Fig. 9. Evolution of the internal radius of the diaphysis of the human femur (in microns) versus time. Applied stress above the critical stress level:  $\sigma_{zz} = 0.02MPa$ .

Apposition of new bone would occur in the absence of mechanical stimulus, under the influence of a pure chemical driving force; in that case, the internal radius will decrease very rapidly (Figure 9) and tends to an asymptotic value (for long times) after about two weeks growth. For a non vanishing axial stress above the critical stress in (5.16), the driving force is negative in the first growth period, and becomes thereafter positive due to the decrease of the surface density of minerals, indicating that growth takes over from bone resorption. Hence, the developed model is able to encompass both situations of growth and resorption, according to the level of applied stress (the nature of the stress, compressive or under traction, does not play a role according to (5.15)), which determine the mechanical contribution of the overall driving force for growth.

## 6. Concluding remarks

Surface growth is by essence a pluridisciplinary field, involving interactions between the physics and mechanics of surfaces and transport phenomena. The literature survey shows different strategies for treating superficial interactions, hence recognizing that no unitary viewpoint yet exists. The present contribution aims at providing a pluridisciplinary approach of surface growth focusing on

A macroscopic model of bone external remodeling has been developed, basing on the thermodynamics of surfaces and with the identified configurational driving forces promoting surface evolution. The interactions between the surface diffusion of minerals and the mechanical driving factors have been quantified, resulting in a relatively rich model in terms of physical and mechanical parameters. Applications of the developed formalism to real geometries

Works accounting for the multiscale aspect of bone remodeling have emerged in the literature since the late nineteen's considering cell-scale (a few microns) up to bone-scale (a few centimeters) remodeling, showing adaptation of the 3D trabeculae architecture in response to mechanical stimulation, see the recent contributions (Tsubota et al., 2009; Coelho et al., 2009) and the references therein. It is likely that one has in the future to combine models at both micro and macro scales in a hierarchical approach to get deeper insight into the mechanisms of Wolff's law.

The present modeling framework shall serve as a convenient platform for the simulation of bone remodeling with the consideration of real geometries extracted from CT scans. The predictive aspect of those simulations is interesting in a medical context, since it will help doctors in adapting the medical treatment according to short and long term predictions of the simulations.

## 7. References

- Blair, H.C. (1998). How the osteoclast degrades bone. *Bioessays*, Vol.20, pp.837-46
- Carter, D.R. ; Orr, T.E. & Fyrhie, D.P. (1989). Relationships between loading history and femoral cancellous bone architecture. *J. Biomech.*, Vol.22, No.3, pp.231-244
- Couchman, P.R. & Linford, R.G. (1980). Aspects of solid surface thermodynamics: relationships of the Shuttleworth type. *Jnal of Electroanalytical Chemistry*. Vol.115, pp.143-148
- Cowin, S.C. (2001). *Bone Mechanics Handbook*. CRC Press
- Eshelby, J. D. (1951). The force on an elastic singularity. *Phil. Trans. R. Soc.*, Vol.A244, pp.87-112
- Frost, H.M. (1964). Dynamics of bone remodeling. *Bone Biodynam.*, Vol.2, pp.315-333
- Ganghoffer, J.F. & Haussy, B. (2005). Mechanical modeling of growth considering domain variation. Part I: constitutive framework. *Int. J. Solids Struct.*, Vol. 42, No.15, pp.4311-4337
- Ganghoffer, J.F. (2010a). Mechanical modeling of growth considering domain variation— Part II: Volumetric and surface growth involving Eshelby tensors. *J. Mech. Phys. Solids*, Vol.58, No.9, pp.1434-1459
- Ganghoffer, J.F. (2010b). On Eshelby tensors in the context of the thermodynamics of open systems: Application to volumetric growth. *Int. J. Engng Sci.*, doi:10.1016/j.ijengsci.2010.04.003
- Garikipati., K. (2009). The kinematics of biological growth. *Appl. Mech. Rev.* 62 (3), 030801, doi:10.1115/1.3090829
- Goodrich, F.C. (1969). In *Surface and Colloid Science*, Vol. 1, Ed. By E. Matijevic, p. 1
- Gurtin, M.E. & Murdoch, A.I. (1975). A continuum theory of elastic material Surfaces. *Archive for Rational Mechanics and Analysis*, Vol.57, No.4, pp.291-323
- Gurtin, M.E. & Struthers, A. (1990). Multiphase thermomechanics with interfacial structure. Part 3. *Arch. Rat. Mech. Anal.*, Vol.112, pp.97-160

- Gurtin, M.E. (1995). On the nature of configurational forces. *Arch. Rat. Mech. Anal.*, Vol. 131, pp.67-100
- Gurtin, M.E. (2000). In: *Configurational forces as Basic Concepts of Continuum Physics*. Springer, New York
- Huiskes, R. & Sloff, T.J. (1981). Geometrical and mechanical properties of the human femur. *Biomechanics VII-A: 7th international congress on biomechanics*, Ed. A. Morecki, Vol.3A, pp.57-64
- Huiskes, R.; Weinans, H., Grootenboer, H. J., Dalstra, M., Fudala, B. & Slooff, T.J. (1987). *J. Biomech.*, Vol.20, pp.1135-50
- Kessler, D.A. (1990). Coupled-map lattice model for crystal growth. *Phys. Rev.*, Vol.A 42, pp.6125-6128
- Kondepudi, D. & Prigogine, I. (1998). *Modern Thermodynamics: From Heat Engines to Dissipative Structures*. Wiley
- Langer, J.S. (1980). Instabilities and pattern formation in crystal growth. *Rev. Mod. Phys.*, Vol.52, pp.1-28
- Leo, P.H. & Sekerka, R.F. (1989). The effect of surface stress on crystal-melt and crystal-crystal equilibrium. *Acta Metall.*, Vol.37, No.12, pp.3119-3138
- Linford, R.G. (1973). *Surface thermodynamics of solids. Solid State. Surface Science*, Vol.2, pp.1-152
- Lu, T.W.; O'Connor, J. J., Taylor, S. J. G. & Walker, P. S. (1997). Influence of muscle activity on the forces in the femur: an *in vivo* study. *J. Biomech.* Vol.30, No.11, pp.1101-1106
- Maugin, G.A. & Trimarco, C. (1995). The dynamics of configurational forces at phase-transition fronts. *Meccanica* Vol.30, pp.605-619
- Maugin, G.A. (1993). *Material Inhomogeneities in elasticity*. Chapman et al., London
- Mindlin, R.D. (1965). Second gradient of strain and surface-tension in linear elasticity. *Int. J. Solids Struct.*, Vol.1, pp.417-438
- Muller, P. & Kern, R. (2001). In *Stress and Strain in epitaxy: theoretical concept measurements*. Ed. M. Handbucken, J.P. Deville, Elsevier
- Mullins, W.W. (1963). in *Metals surfaces, structures, energetics, and kinetics. Amer. Soc. Metals. Metals Park, Ohio*, p. 7
- Munster, A. (1970). *Classical Thermodynamics*, John Wiley and Sons
- Reilly, D.T. & Burstein, A.H., 1975. The elastic and ultimate properties of compact bone tissue, *J. Biomech.*, Vol.8, pp.393-405
- Rice, J.C.; Cowin, S.C. & Bowman, J.A. (1988). On the dependence of the elasticity and strength of cancellous bone on apparent density. *J. Biomech.*, Vol.21, pp.155-168
- Ruimerman, R.; Hilbers, P., Van Rietbergen, B. & Huiskes, R. (2005). A theoretical framework for strain-related trabecular bone maintenance and adaptation. *J. Biomech.*, Vol.38, pp.931-941
- Silva, E.C.C.M. & Ulm, F.J. (2002). A bio-chemo-mechanics approach to bone resorption and fracture. Proc. 15<sup>th</sup> ASCE Engineering Mechanics Conference. 02-05/06/2002, Columbia University, New York, USA
- Skalak, R., Dasgupta, G. & Moss, M. (1982). Analytical description of growth. *J. Theor. Biol.*, Vol.94, pp.555-577
- Skalak, R., Farrow, D.A. & Hoger, A. (1997). Kinematics of surface growth. *J. Math. Biol.*, Vol.35, pp.869-907

- Taber, L. (1995). Biomechanics of growth, remodeling and morphogenesis. *Appl. Mech. Rev.*, Vol.48, pp.487-545
- Thompson, D.W. (1992). *On Growth and Form*. Dover reprint of 1942. 2nd edition
- Vidal, C.; Dewel, P. & Borckmans, P. (1994). *Au-delà de l'équilibre*. Hermann
- Weinans, H.; Huskes, R. & Grootenboer, H.J. (1992). The behavior of adaptive bone remodeling simulation models. *J. Biomech.*, Vol.25, pp.1425-1441
- Yang, G.; Kaber, J., Van Rietbergen, B., Odgaard, A., Huiskes, R. & Cowin, S.C. (1999). The anisotropic Hooke's law for cancellous bone and wood. *J. Elast.*, Vol.53, pp.125-146

# Thermodynamic Aspects of CVD Crystallization of Refractory Metals and Their Alloys

Yu. V. Lakhotkin  
*Frumkin Institute of Physical Chemistry and Electrochemistry,  
Russian Academy of Sciences, Moscow,  
Russia*

## 1. Introduction

The low-temperature chemical vapor deposition (CVD) of refractory metals by the hydrogen reduction of their fluorides is known as one of the perspective technique for the production of high quality metallic coatings [1]. The CVD of tungsten has been more extensively studied due to unique combination of its features such as low deposition temperature (750-900 K), high growth rate (up to 5 mm/h), a good purity and high density of tungsten deposit [2, 3]. Up to now there is a great interest to CVD tungsten alloys due to their physical-mechanical properties [4, 5].

The thermodynamic analysis of the CVD processes is useful to define the optimal deposition conditions. The understanding of the gas phase phenomena controlling the metals and alloys deposition requires the knowledge of the gaseous mixture composition and surface reaction kinetics which lead to the deposit growth. This chapter contains the calculated and known thermochemical parameters of V, Nb, Ta, Mo, W, Re fluorides, the compositions of gas and solid phases as result of the equilibrium of the hydrogen and fluorides for the metals VB group (V, Nb, Ta), VIB group (Mo, W), VII group (Re). A particular attention is paid to the theoretical aspects of tungsten alloys crystallization.

## 2. Estimation of thermochemical constants

The accuracy of thermodynamic analysis depends on the completeness and reliability of thermochemical data. Unfortunately, a limited number of the transition metal fluorides have been characterized thermochemically or have been studied by a spectroscopic technique. The experimental data were completed with the evaluated thermochemical constants for fluorides in different valent and structural states. The calculated data were obtained by the interpolation procedure based on the periodic law. The interpolation was performed on properties of a number of the compounds that represent the electron-nuclei analogies [6]. The unknown enthalpy of the fluorides formation was calculating via energy of halids atomization as following:

$$\Omega (MX_n) = \Delta_f H (M_{at}) + n \Delta_f H (X_{at}) - \Delta_f H (MX_n), \quad (1)$$

The atomization energies of isovalent fluorides, chlorides and oxides of 4, 5, 6 period metals were discussed in [7]. It can be emphasised that the chlorides and oxides are studied well by experimental way. These curves are calling as “two-hilled” curves. Quantum-mechanical interpretation of these dependences can be found in [8, 9].

$$\Omega (MX_n) = \varphi (Z_m), \quad (2)$$

$$\Delta Z_m \Omega (MF_n) / \Delta Z_m \Omega (MF_s) = \varphi (Z_m, n), \quad (3)$$

$$\Delta n_m \Omega (MF_n) / \Delta Z_m \Omega (MCl_n) = \varphi (Z_m, n), \quad (4)$$

$$\Omega (MF_n) / \Omega (MCl_n) = \varphi (Z_m, n), \quad (5)$$

$$\Omega (MF_n) / \Omega (MO_{n/2}) = \varphi (Z_m, n), \quad (6)$$

$$\Omega [M (Z_m) F_n] / \Omega [M (Z_m + 32) F_n] = \varphi (Z_m, n), \text{ where } Z_m = 39-48 \quad (7)$$

$$\Omega (MX_n) = \varphi (Z_x) = A_n \psi (Z_x) + B_n, \quad X = F, Cl, Br, I, \quad (8)$$

$$\Omega (MF_n) = \varphi (n) = \psi [\Omega (MCl_n)] : \Omega (MF_n) = C \Omega (MCl_n) + D \quad (9)$$

$$E (MF_n) = \varphi (n) = \psi [D (MF_n)] : E (MF_n) = L \Omega (MF_n) + N, \quad (10)$$

where  $A_n, B_n, C, D, L, N$  -const.

These sequences are the dependencies of energies of halids atomization (2, 8-10), one of ratio of loss of energies of fluoride and chloride atomization (3, 4) from atom number of metal  $Z_m$  (2-7), from halid  $Z_x$  (8) and from valent state  $n$  (3-7, 9, 10).

All sequences were analyzed in order to determine the probable regions for interpolation by linear function. For example, the estimation of unknown atomization energies can be performed by the use of the sequence (2) within following region:

$\Omega (MF)$  where  $Z_m$  corresponds to (III-IV-V) and (VI-VII-VIII-I) groups;

$\Omega (MF_2)$  where  $Z_m$  corresponds to (V-VI) and (VI-VII-VIII) groups;

$\Omega (MF_n, n \geq 3)$  where  $Z_m$  corresponds to (V-VI-VII) groups.

The sublimation heat  $\Delta_s H (MX_n)$  and entrophy  $S (MX_n)$  were analyzed:

$$\Delta_s H (MX_n) = \varphi (Z_m, Z_x, n), \quad (11)$$

$$S (MX_n) = \varphi (Z_m, Z_x, n). \quad (12)$$

All calculated thermochemical constants together with most reliable literature data are collected in tables 1, 2. The accuracy of the estimation data is  $\pm 30$  kJ/mol for atomization energy and  $\pm 4$  J/mol K for atomization entrophy. The accurate thermochemical data of W-F-H components are collected in the table 3, due to their importance for this analysis.

The literature review shows that the formation enthalpy is determined for several fluorides enough reliable which are taken as milestone points. Among them are  $AlF_3, UF_4, UF_5, ScF_3, CrF_2, MnF_2, TiF_4, FeF_3$  and other [28, 29]. Table 1 contains also the thermochemical constants for polymer fluorides. Most reliable thermochemical data among the fluoride associations were obtained for  $Al_2F_6, Fe_2F_8, Cr_2F_4$ . The thermochemical data for tungsten fluorides are collected in table 2 because of the special importance for this investigation. Of course these data will be more full and reliable in the progress of fluoride chemistry.



№	Substance	$\Delta_f H^\circ_{298}$ (g)	$\Delta_f H^\circ$ (s)	$\Delta_s H^\circ_{298}$
1	V	514,1±4,2 [10]	0	514,1±4,2
2	VF	2,5±63 [11]	-	-
3	VF <sub>2</sub>	≤-514±28 [7]	≤-899±28 [7]	385±28
4	VF <sub>3</sub>	-878,0±48,1 [12]	-1263,1±48,1 [7]	385,1
5	VF <sub>4</sub>	≤-1241±8 [7]	-1412,1 [10]	169,1±8,0
6	VF <sub>5</sub>	≤-1429,7±5,0 [11]	-	-
7	V <sub>2</sub> F <sub>6</sub>	-1963,4±48,0 [13]	-	-
8	V <sub>2</sub> F <sub>8</sub>	-2746,7±20,9 [14]	-	-
1	Nb	721,9±4,2 [10]	-	721,9±4,2
2	NbF	228±25 [7]	-	-
3	NbF <sub>2</sub>	≤-226±21 [7]	-	-
4	NbF <sub>3</sub>	≤-754±16 [7]	-	-
5	NbF <sub>4</sub>	-1257±22 [7]	-1506±21 [7]	249±22
6	NbF <sub>5</sub>	-1711,7±6,3 [12]	-1813,8±0,6 [10]	102,1±6,9
7	Nb <sub>3</sub> F <sub>15</sub>	-5342,0±4,6 [15]	-	-
1	Ta	785,4±4,2 [10]	0	785,4±4,2
2	TaF	289,3±12,5 [16]	-	
3	TaF <sub>2</sub>	-287,2±12,5 [16]	-	
4	TaF <sub>3</sub>	-810,9±12,5 [16]	-	
5	TaF <sub>4</sub>	-1275,7±12,5 [16]	-	
6	TaF <sub>5</sub>	-1774,8±12,5 [16]	-1901,8±0,8 [10]	127±13,3
7	Ta <sub>3</sub> F <sub>15</sub>	-5611,2±5,4 [15]		
1	Mo	655,8±3,4 [7]	0	655,8±3,4
2	MoF	271,7±9,2 [17]	-	-
3	MoF <sub>2</sub>	-168,0±12,1 [17]	-	-
4	MoF <sub>3</sub>	-591,5±14,6 [17]	-909,6±19,7 [18]	318,1±34,3
5	MoF <sub>4</sub>	-953,0±16,3 [17]	-1149,0±14,6 [18]	196,0±30,9
6	MoF <sub>5</sub>	-1240,2±35,9 [17]	-1394,4±4,6 [19]	154,2±40,5
7	MoF <sub>6</sub>	-1556,2±0,8 [10]	-	-
8	Mo <sub>3</sub> F <sub>15</sub>	-4091,0±9,6 [20]	-	-
1	W	856,1±4,2 [10]	0	856,1±4,2
2	WF	≤385 [21]	-	
3	WF <sub>2</sub>	-86,2±13,4 [21]	-	
4	WF <sub>3</sub>	-507,1±11,7 [21]	-	
5	WF <sub>4</sub>	-928,8±10,5 [21]	-1206,2±7,5 [18]	277,4±13
6	WF <sub>5</sub>	-1293,3±8,4 [21]	-1446,8±8,4 [10]	153,5±16,8
7	WF <sub>6</sub>	-1721,5±0,7 [10]	-	

№	Substance	$\Delta_f H^{\circ}_{298}$ (g)	$\Delta_f H^{\circ}$ (s)	$\Delta_s H^{\circ}_{298}$
8	W <sub>3</sub> F <sub>15</sub>	-4244,0±8,4 [15]	-	
1	Re	775,0±6,3 [10]	0	775,0±6,3
2	ReF	343±60 [7]	-	
3	ReF <sub>2</sub>	-116±46 [7]	-	
4	ReF <sub>3</sub>	-354±36 [7]	-	
5	ReF <sub>4</sub>	-733±33 [7]	-995±33 [7]	263±26
6	ReF <sub>5</sub>	-962±29 [7]	-1142±18 [7]	180±29
7	ReF <sub>6</sub>	-1353,5±12,6 [10]	-	
8	ReF <sub>7</sub>	-1410±11 [22]	-1450,5±10,9 [22]	40,5±21,9
9	Re <sub>2</sub> F <sub>8</sub>	-1854,2±33,4 [23]		
10	Re <sub>3</sub> F <sub>15</sub>	-3337,7±17,6 [24]		
1	F	79,43±1,05 [10]	-	-
2	F <sub>2</sub>	0	-	-
1	H	217,77±0,02 [10]	-	-
2	H <sub>2</sub>	0	-	-
1	HF	-270,4±1,2 [10]	-	-

Table 1. Enthalpy of forming  $\Delta_f H$  (kJ/mol) and sublimation  $\Delta_s H$  (kJ/mol) of system M-F-H components in gas (g) and solid (s) states.

№	Substance	$S^{\circ}_{298}$ (g)	$S^{\circ}_{298}$ (s)
1	V	182,010±0,033 [10]	28,88±0,33 [10]
2	VF	230±4 [10]	-
3	VF <sub>2</sub>	254,4 [12]	76,220 [25]
4	VF <sub>3</sub>	283,05 [12]	96,99 [10]
5	VF <sub>4</sub>	305±4 [7]	126,13 [10]
6	VF <sub>5</sub>	331,0±2,9 [10]	-
7	V <sub>2</sub> F <sub>6</sub>	397,0±17 [13]	-
8	V <sub>2</sub> F <sub>8</sub>	456±17 [14]	-
1	Nb	186,000±0,033 [10]	36,53±0,21 [10]
2	NbF	241,4 [12]	-
3	NbF <sub>2</sub>	281,6 [12]	-
4	NbF <sub>3</sub>	296,2 [12]	-
5	NbF <sub>4</sub>	325,5 [12]	100±4 [10]
6	NbF <sub>5</sub>	323,8 [12]	157,3±2,1 [10]
7	Nb <sub>3</sub> F <sub>15</sub>	683,0±16,7 [15]	-
1	Ta	184,927±0,033 [10]	41,47±0,17 [10]
2	TaF	240,8 [12]	-

N <sub>o</sub>	Substance	S <sup>o</sup> <sub>298</sub> (g)		S <sup>o</sup> <sub>298</sub> (s)
3	TaF <sub>2</sub>	290,9	[12]	-
4	TaF <sub>3</sub>	308,1	[12]	-
5	TaF <sub>4</sub>	336,1	[12]	-
6	TaF <sub>5</sub>	332,7	[12]	169,7±16,7 [10]
7	Ta <sub>3</sub> F <sub>15</sub>	720,6±14,6	[15]	-
1	Mo	181,663±0,029	[10]	28,59±0,21 [10]
2	MoF	243,53	[12]	-
3	MoF <sub>2</sub>	275,9	[12]	-
4	MoF <sub>3</sub>	301,3	[12]	93±12 [7]
5	MoF <sub>4</sub>	319,3	[12]	100±12 [7]
6	MoF <sub>5</sub>	327,7±1,7	[10]	125±12 [10]
7	MoF <sub>6</sub>	350,3±1,2	[10]	-
8	Mo <sub>3</sub> F <sub>15</sub>	580,6±16,7	[20]	-
1	W	173,675±0,029	[10]	32,65±0,33 [10]
2	WF	250,6±4,2	[10]	-
3	WF <sub>2</sub>	285,8	[12]	-
4	WF <sub>3</sub>	314,2	[12]	-
5	WF <sub>4</sub>	330,1	[12]	103,3±8,4 [10]
6	WF <sub>5</sub>	343,1	[12]	146±13 [10]
7	WF <sub>6</sub>	353,5±1,3	[10]	-
8	W <sub>3</sub> F <sub>15</sub>	631±12	[15]	-
1	Re	188,643±0,029	[10]	36,49±0,33 [10]
2	ReF	251±4	[7]	-
3	ReF <sub>2</sub>	285±4	[7]	-
4	ReF <sub>3</sub>	308,8	[12]	-
5	ReF <sub>4</sub>	333,9±6,3	[10]	146,4±8,4 [10]
6	ReF <sub>5</sub>	337,6±6,3	[10]	175,7±8,4 [10]
7	ReF <sub>6</sub>	363,6±2,1	[10]	-
8	ReF <sub>7</sub>	360	[22]	-
9	Re <sub>2</sub> F <sub>8</sub>	497±17	[23]	
10	Re <sub>3</sub> F <sub>15</sub>	736±17	[24]	
1	F	158,489±0,021	[10]	-
2	F <sub>2</sub>	202,52±0,25	[10]	-
1	H	114,494±0,021	[10]	-
2	H <sub>2</sub>	130,395±0,021	[10]	-
1	HF	173,512±0,033	[10]	-

Table 2. Entropy data S<sup>o</sup><sub>298</sub> (J/K mol) of system M-F-H components in gas (g) and solid (s) states.

Komponents	$\Delta H_f^{o,298}$ , kJ/mol	$S_{o,298}$ , J/mol·K	$C_p = \alpha + \beta T + \gamma T^2 + \delta T^{-2}$ , J/mol·K				References
			$\alpha$	$10^3 \beta$	$10^5 \gamma$	$10^{-5} \delta$	
H <sub>2</sub>	0,0	130,4	32,02	-7,36	0,58	1,34	[10, 26]
H	217,8	114,5	20,77	0,0	0,0	0,0	[10, 26]
F <sub>2</sub>	0,0	202,5	26,42	22,36	-1,25	-0,63	[10, 26]
F	79,4	158,5	25,08	-7,86	0,42	-0,33	[10, 26]
HF	-270,4	173,5	30,01	-3,47	3,47	-0,25	[10, 26]
WF <sub>6</sub>	-1719,9	357,2	117,46	83,60	-5,02	-16,93	[10, 21, 26]
WF <sub>5</sub>	-1292,0	342,8	114,95	54,76	-3,26	-14,13	[12, 21]
WF <sub>4</sub>	-928,0	329,8	80,67	56,85	-3,43	-7,98	[12, 21]
WF <sub>3</sub>	-506,6	313,9	65,63	36,28	-2,21	-7,57	[12, 21]
WF <sub>2</sub>	-86,1	285,5	53,92	26,41	-1,59	-3,85	[12, 21]
WF	384,6	250,4	30,72	13,92	-0,79	-1,55	[12, 14]
W <sub>2</sub> F <sub>8</sub>	-2042,4	414,5	166,36	112,86	-6,81	-22,40	[7, 27]
W <sub>2</sub> F <sub>10</sub>	-2829,4	497,0	202,31	142,12	-8,61	-28,34	[7, 27]
W <sub>3</sub> F <sub>15</sub>	-4244,0	631,2	295,00	247,94	-15,29	-42,39	[7, 15]
HWF <sub>5</sub>	-1383,9	352,2	98,65	103,25	-5,60	-15,01	[27]

Table 3. Standart thermochemical constants of W-F-H components.

### 3. Equilibrium states in M-F systems

Temperature dependencies of equilibrium compositions in the M-F systems (M = V, Nb, Ta, Mo, W, Re) are presented at the Fig.1. The data represent the thermodynamic stability of the refractory metal fluorides with different valencies both monomer and polymer states depending on the place of the metal in the Periodic table. The gas phase composition depends on both the heat of the fluoride formation and the vaporation heat of the fluorides. The thermodynamic analysis of M-F systems shows that the highest fluorides of the metals are stable at temperatures up to 2000 K. The exceptions contain the fluorides VF<sub>5</sub>, MoF<sub>6</sub>, ReF<sub>6</sub> that decompose slightly at the high temperature range and their thermal stability increase according to the following order: VF<sub>5</sub> > MoF<sub>6</sub> > ReF<sub>6</sub>.

The gas low-valent fluoride concentrations, which depend upon the metal place in the periodic system, rise with the increase of atomic number within each group and decrease with the increase of atomic number within each period. Thus tantalum fluorides are most strongly bonded halids and vanadium fluorides are most unstable among considered fluorides. It is necessary to note that partial pressures of low valent fluorides in Re-F system are close to each other but low valent fluorides in Ta-F system have very different concentrations.

Nevertheless the vaporation temperature of fluorides varies depending upon the metal place in the periodic system in opposite direction than the gas low-valent fluorides concentration. The most refractory fluorides are VF<sub>2</sub> and VF<sub>3</sub> (above 1500 K), the low-valent fluorides of Nb and Mo possess the mean vaporation temperature (900-1100 K). Th low-

valent fluorides of tantalum, tungsten, rhenium have the lowest vaporation temperature (500-550 K).

The peculiarity of the fluorides is the possibility of their polymerization. It is known that dimers or threemers are observed in gas state but tetramer clusters of Nb, Ta, Mo, W fluorides and chains of V, Re fluoride polymers are forming in solid state [30]. For example, fluorides  $W_2F_8$ ,  $W_2F_{10}$  and  $Mo_2F_6$ ,  $Mo_2F_8$ ,  $Mo_2F_{10}$  exist in W-F and Mo-F system, correspondingly. The main structural state of Nb, Ta, Mo, W, Re fluoride polymers are threemers but vanadium pentafluoride does form polymer state.  $M_3F_{15}$  polymers are forming by the single M-F-M bonds but the fluoride dimers have double fluorine bridge bonds. The exception are dimer molecules  $V_2F_6$ ,  $V_2F_8$ ,  $Re_2F_8$  with the M-M bonds. All polymer states are presented in tables 1-3.

#### 4. Equilibrium states in M-F-H systems

The equilibrium analysis of the metal-fluorine-hydrogen (M-F-H) systems for the temperature range 400-2000 K, total pressure of  $1.3 \times 10^5$  Pa and 2 kPa and for fluoride to hydrogen ratio from 1:3 to 1:100 have been calculated using a special procedure based on the search of entropy extremum for the polycomponent mixture [7, 31]. All experimental and calculated thermochemical constants of the fluorides and the characteristics of the fluoride phase transitions were involved into the data set. The equilibrium compositions of M-F-H systems (M=V, Nb, Ta, Mo, W, Re) for the optimal total pressure and the optimal reagent ratio are presented at the Fig.2.

The comparison of the results presented at the Fig. 1 and Fig.2 shows that the addition of hydrogen to VB metal pentafluorides decrease concentrations of the highest fluorides in monomer and polymer states (except of  $V_2F_6$ ) and rise the concentration of lower-valent fluorides. The large difference is observed for V-F-H system and small difference - for Ta-F-H system.

The hydrogen addition to tungsten, molibdenium and rhenium hexafluorides leads to the decrease of  $MF_x$  concentration,  $7 \leq x \leq 3$ , and to a small increase of di- and monofluorides concentration.

The source of VB group metals formed from M-F-H systems are highest fluorides and polymers. The VI group metals are the product of hexa-, penta- and tetrafluoride decomposition, but all known rhenium fluorides produce the metallic deposit. The variation of the external conditions (total pressure and fluoride to hydrogen ratio) influence on the gas phase composition according to the law of mass action and Le Chatelier principle.

Fig. 3 presents the equilibrium yield of solid metallic deposit from the mixtures of their fluorides with hydrogen as a function of the temperature. It is shown that metallic Re, Mo, W may be deposited from M-F-H system at temperatures above 300 K. Yields of Nb and Ta were varied in the temperature range from 800 K to 1300 K. Metallic V may be not deposited from M-F-H system until 1700 K due to the high sublimation temperature of  $VF_2$  and  $VF_3$ . It was established that the moving force (supersaturation) of the metal crystallization in M-F-H system increase in the order for following metals: Re, Mo, W, Nb, Ta, V. These thermodynamic results are in agreement with experimental data reviewed in [7, 32, 33].

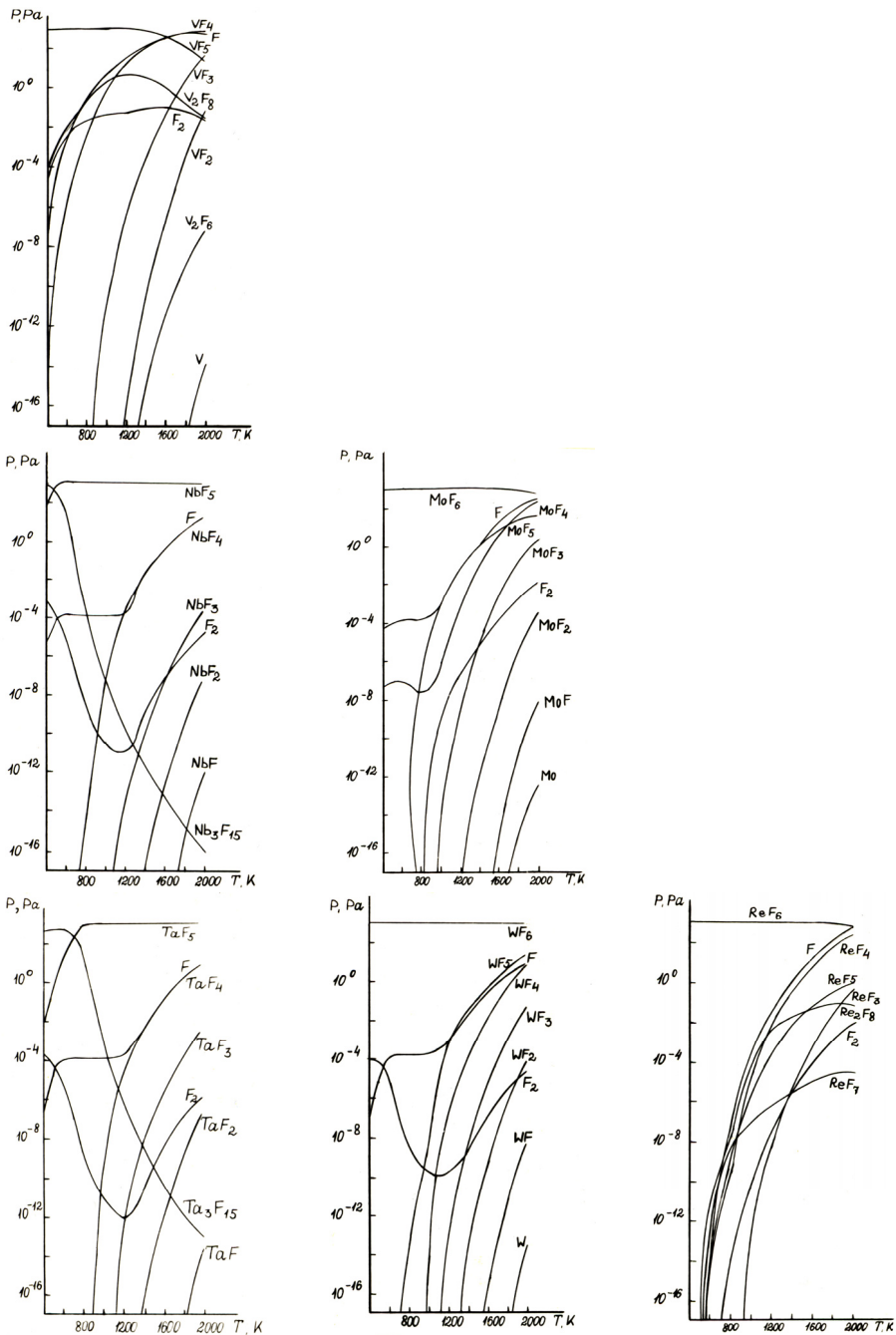


Fig. 1. Equilibrium gaseous composition in M-F systems at total pressure of 2 kPa [7].

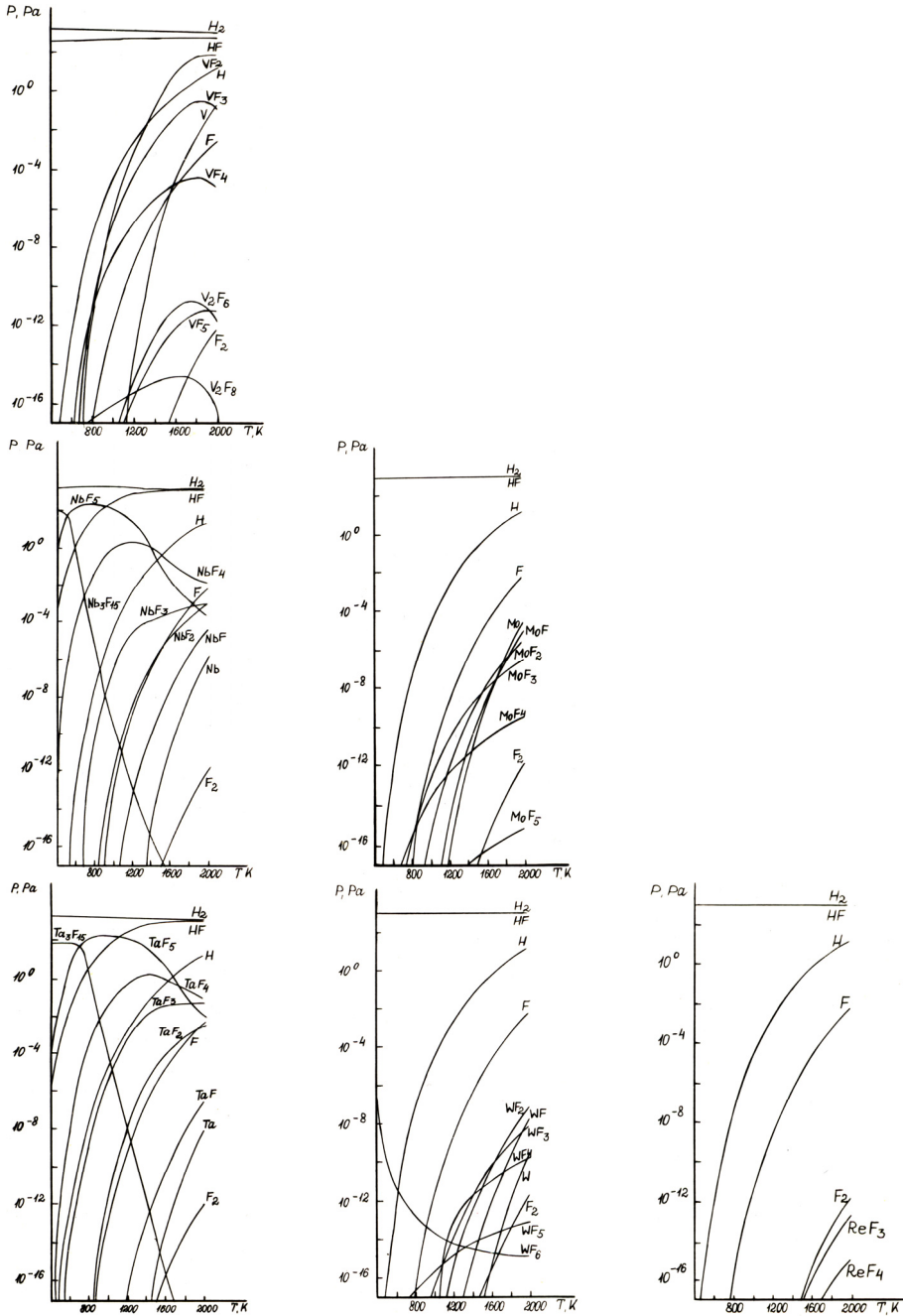


Fig. 2. Equilibrium gaseous composition in M-F-H systems at total pressure of 2 kPa and hydrogen to highest fluoride initial ratio of 10 [31].

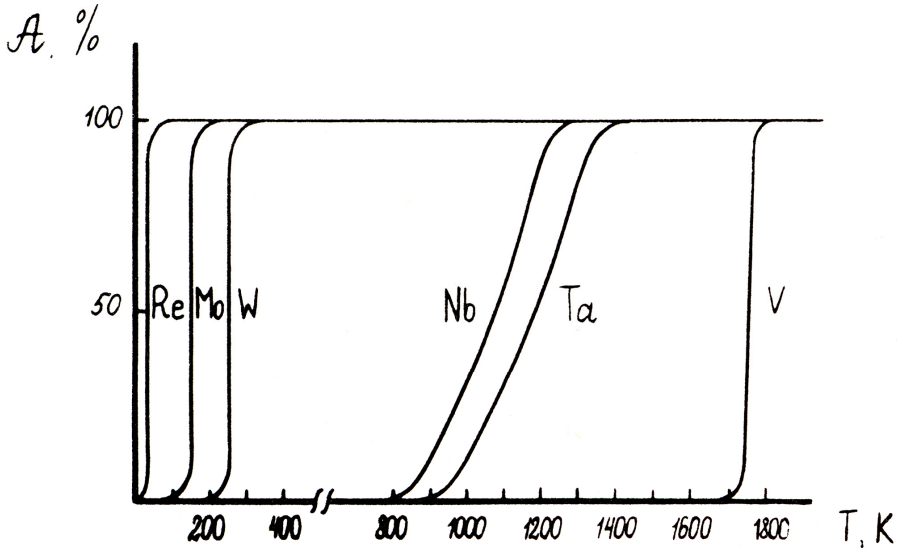


Fig. 3. Yield of metals (V, Nb, Ta, Mo, W, Re) from the equilibrium mixtures of their fluorides with hydrogen (1:10) as a function of the temperature [31].

### 5. Equilibrium composition of solid deposit in W-M-F-H systems

A thermodynamics of alloy co-deposition is often considered as a heterogeneous equilibrium of gas and solid phases, in which solid components are not bonded chemically or form the solid solution. The calculation of the solid solution composition requires the knowledge of the entropy and enthalpy of the components mixing. The entropy of mixing is easily calculated but the enthalpy of mixing is usually determined by the experimental procedure. For tungsten alloys, these parameters are estimated only theoretically [34]. A partial enthalpy of mixing can be approximated as the following:

$$\Delta H_m = (h_{1,i} + h_{2,i} T + h_{3,i} x_i) \times (1 - x_i)^2,$$

where  $h_{1,i}$ ,  $h_{2,i}$ ,  $h_{3,i}$  - polynomial's coefficients,  $T$  - temperature,  $x_i$  - mole fraction of solution component.

The surface properties of tungsten are sharply different from the bulk properties due to strongest chemical interatomic bonds. Therefore, there is an expedience to include the crystallization stage in the thermodynamic consideration, because the crystallization stage controls the tungsten growth in a large interval of deposition conditions. To determine the enthalpy of mixing of surface atoms we use the results of the desorption of transition metals on (100) tungsten plane presented at the Fig. 4. [35]. The crystallization energy can be determined as the difference between the molar enthalpy of the transition metal sublimation



from (100) tungsten surface and sublimation energy of pure metal. These values are presented in the table 4 in terms of polynomial's coefficients, which were estimated in the case of the infinite dilute solution. The peculiarity of the detail calculation of polynomial's coefficients is discussed in [7]. The data predict that the co-crystallization of tungsten with Nb, V, Mo, Re will be performed more easily than the crystallization of pure tungsten. The crystallization of W-Ta alloys has the reverse tendency. Certainly the synergetic effects will influence on the composition of gas and solid phases.

No	M	$\Delta H_m^0 \ominus 298 \text{ K}$ $x_i = 0$	$h_{1, i}$ kJ/mol	$h_{2, i}$ kJ/mol	$h_{3, i}$ kJ/mol	$x_i$
1	W	0	0	0	0	1,0000-0,9375
	Ta	36,4±10,9	36,4	-0,00042	72,7	0,0000-0,0625
2	W	0	0	0	0	1,0000-0,9375
	Nb	-225,7±50,2	-225,7	-0,00025	-451,4	0,0000-0,0625
3	W	0	0	0	0	1,0000-0,9375
	V	-434,7±50,2	-434,7	-0,00017	-1304,2	0,0000-0,0625
4	W	0	0	0	0	1,0000-0,9375
	Mo	-467,7±10,9	-467,7	-0,00117	-935,5	0,0000-0,0625
5	W	0	0	0	0	1,0000-0,9375
	Re	-220,3±10,9	-220,3	-0,00058	-440,5	0,0000-0,0625

Table 4. Excess partial "enthalpy of mixing" atoms for crystallization of W-M binary solid solution and  $h_i$  polynomial's coefficients for  $x_i = 0 - 0.0625$  and  $T = 298 - 2500 \text{ K}$  [7, 31].

Therefore the thermodynamic calculation for gas and solid composition of W-M-F-H systems were carried out for following cases:

1. without the mutual interaction of solid components;
2. for the formation of ideal solid solution
3. for the interaction of binary solution components on the surface.

The temperature influence on the conversion of VB group metal fluorides and their addition to the tungsten hexafluoride – hydrogen mixture is presented at the Fig.5 a,b,c. If the metal interaction in the solid phase is not taken into account, the vanadium pentafluoride is reduced by hydrogen only to lower-valent fluorides. It should be noted that metallic vanadium can be deposited at temperatures above 1700 K. Equilibrium fraction of  $\text{NbF}_5$  conversion achieves 50% at 1400 K, and of  $\text{TaF}_5$  – at 1600 K (Fig. 5 a,b,c, curves 1).

The thermodynamic consideration of ideal solid solution shows that tungsten-vanadium alloys may deposit at the high temperature range ( $T \geq 1400 \text{ K}$ ) and metallic vanadium is deposited in mixture with lower-valent fluorides of vanadium (Fig. 5 a, curves 2). The beginnings of formation of W-Nb and W-Ta ideal solid solutions are shifted to lower temperature by about 100 K (Fig. 5 b,c, curves 2) in comparison with the case (1).

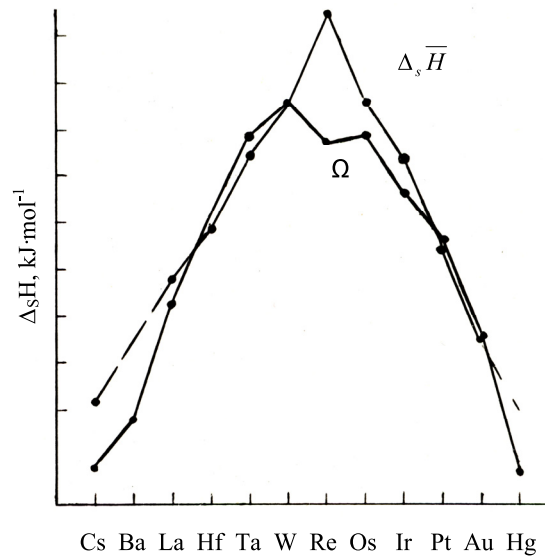
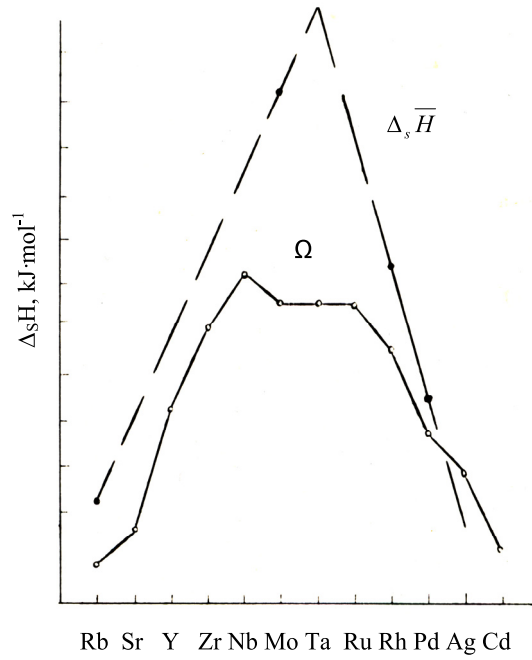


Fig. 4. Partial molar enthalpy of 4d и 5d atoms sublimation ( $\Delta_s \bar{H}$ ) from tungsten plane (100) and atomization energy ( $\Omega$ ) of transition metals in dependence on their place in periodic table [35]

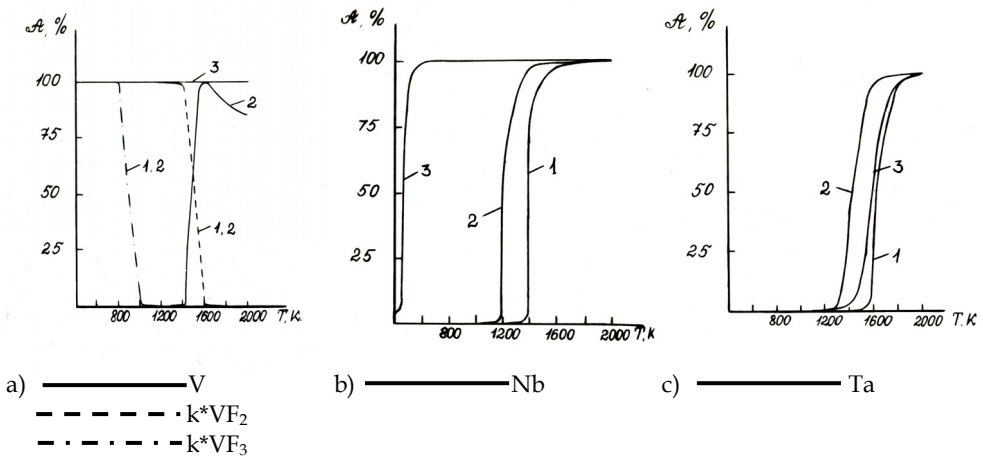


Fig. 5. Equilibrium yield of VB metals during crystallization with tungsten at initial ratio

WF6:MF5:H2=10, total pressure of 2 kPa calculated for following cases:

1. without the mutual interaction of solid components;
2. for the formation of ideal solid solution
3. for the interaction of binary solution components on the surface.

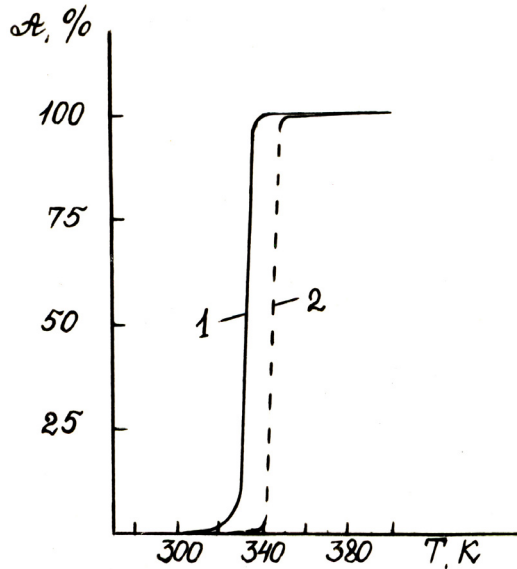


Fig. 6. Temperature influence on equilibrium yield of tungsten in W-Re-F-H (1) and W-F-H (2) systems at total pressure of 2 kPa and gaseous composition of  $(WF_6 + 6\% ReF_6) : H_2 = 10$

Taking into account the interaction of component of alloys during crystallization, the formation of W-V and W-Nb alloys possibly takes place at the temperatures above 300 K

(Fig. 5 a,b, curves 3). Temperature boundary shown at the Fig. 5 is shifted in reverse direction for the W-Ta system (Fig. 5 c, curves 3). It should be noted, that the calculation results performed for cases (2) and (3) (for ideal and nonideal solid solution) for the W-Ta system are almost identical due to the small enthalpy of mixing [35].

The influence of rhenium and molybdenum on the equilibrium yield of tungsten in the M-W-F-H systems is observed for W-Re and W-Mo alloys deposition. The  $\text{ReF}_6$  addition to the gas mixture with  $\text{WF}_6$  increase insignificantly the yield of tungsten in spite of strong atom interaction during the crystallization according to thermodynamic calculations (Fig. 6). This effect is still smaller for the case of W-Mo co-deposition. However equilibrium yield of metals for their co-deposition with tungsten and the energy of the interaction of metallic components during the crystallization have the common tendency. The knowledge of refined data of process energies will allow us to obtain a more realistic situation.

## 6. The application fields of the coatings

The thermodynamic background presented above is very useful for production of the coatings based on tungsten, tungsten alloys with Re, Mo, Nb, Ta, V and tungsten compounds (for example tungsten carbides). The tungsten coatings have found wide application in thin-film integral circuits when preparing the Ohmic contacts in the production of the silicon-, germanium-, and gallium-arsenide-based Schottky-barrier diodes. The tungsten selective deposition technology is perspective in the production of conducting elements at dielectric substrates [36]. Tungsten films are used for covering hot cathodes, improving their emission characteristics, and as protective coatings for anodes in extra-high-power microwave devices. The CVD-tungsten coatings are used as independent elements in electronics.

The X-ray bremsstrahlung in modern clinical tomographs and other X-ray units is obtained by using tungsten or W-Re coatings at rotating anodes made of molybdenum or carbon-carbon composite materials. In the nuclear power engineering, tungsten was shown to be a good material for enveloping nuclear fuel particles because of low diffusion permeability of the envelope for the fuel. The tungsten- and W-Re alloy-coatings [2, 3, 5] are extremely stable in molten salts and metals used as coolants in high-temperature and nuclear machinery, e.g., in heat pipes with lithium coolant and in thermonuclear facilities. Tungsten emitters with high emission uniformity, elevated high-temperature grain orientation and microstructure stability are of interest for their use in thermionic energy converters.

High-temperature technical equipment cannot go without tungsten crucibles, capillaries, and other works that can be easily prepared by the CVD techniques. Tungsten is used as a coating for components of jet engines, fuel cell electrodes, filters and porous components of ion engines, etc. [2] The CVD-alloying of tungsten coatings with rhenium allows to improve significantly their operating ability, especially under the temperature or load cycling. Tungsten compounds have a wide field of application. The tungsten-carbide composites deposited by using the fluoride technology occupy a niche among coatings with a thickness of 10 to 100 mkm; they are unique in respect of strengthening practically any material, starting with carbon, tool, and stainless steels, titanium alloys, and finishing with hard alloys. CVD method permits to coat complicated shape components (which cannot be coated using PVD-method or plasma sputtering of carbide powders with binder). Below we list the most promising fields of applications [37].

In the first place we can mention the strengthening of the oil and gas and drilling equipment (pumps, friction and erosion assemblies). The problems of hydrogen- sulfide corrosion,

wear of movable units, and erosion of immobile parts of drilling bits operating underground take special significance because their replacement is very expensive. The carbide coatings can be deposited inside cylinders and on the outer surfaces of components of rotary or piston oil pumps. Numerous units in the oil and gas equipment, for example, block bearings, solution-supplying channels in drilling bits, backings directing the sludge flow, etc. require the strengthening of their working surfaces.

Another application in this field is the coating of metal-metal gaskets in the high- and ultrahigh-pressure stop and control valves. In addition to intense corrosion, abrasion and erosion wear, the working surfaces of ball cocks and dampers are subject of seizing under high pressure; W-C-coatings prevent the seizure. An important advantage of the carbide coatings is their accessibility for the quality of surface polishing, due to the initial smooth morphology. The examples mentioned above relate not only to oil and gas but also to chemical industry. The W-C-coatings are promising for working in contact with hydrogen-sulfide-rich oil, acids, molten metals, as well as chemically aggressive gases. Due to their high wear and corrosion resistance, these coatings can be use instead of hard chromium.

The abrasion mass extrusion and the metal shape draft require expensive extrusion tools; the product price depends on the working surface quality and life time. The extrusion tools must often have sophisticated shape inappropriate for coating with PVD or PACVD methods. Therefore, W-C-coating prepared by a thermal CVD-method is promising in strengthening these tools. Strengthening of spinneret for drawing wires or complicated section of steel, copper, matrices for aluminum extrusion, ceramic honeycomb structures for the porous substrate of catalytic carriers may give the same effect. Also, very perspective is the deposition of strengthening coatings onto components of equipment for the pressing of powdered abrasion materials. One may also mention the strengthening of knife blade used for cutting paper, cardboard, leather, polyethylene, wood, etc [38].

In addition to the surface strengthening, the W-C-coatings can function as high-temperature glue for mounting diamond particles in a matrix when preparing diamond tools or diamond cakes (conglomerates) in drilling bits [39]. The above-given examples demonstrate the variety of applications for tungsten, its alloys and carbides in mechanical engineering, chemical, gas and oil industry, metallurgy, and microelectronics.

## 7. Conclusion

1. A number of unknown thermochemical constants of refractory metal fluorides were calculated and collected in this chapter.
2. The systematic investigation of equilibrium states in the M-F, M-F-H (M = V, Nb, Ta, Mo, W, Re) systems was carried out. It was demonstrated that the equilibrium concentrations of highest fluorides in the M-F systems are determined by the place of metal in the periodic table. They rise with the increase of atomic number within each group and decrease with the increase of atomic number within each period. The low valent fluoride concentrations have the opposite tendency. It was shown that the equilibrium yield of Re, Mo, W deposition from the M-F-H systems achieve 100% at room temperature, equilibrium yield of Nb, Ta and V deposition - at temperatures above 1300 K, 1600 K and 1700 K, respectively.
3. The solid compositions of the W-M-F-H systems were calculated by taking into account the formation of ideal, nonideal solid solution, the mechanical mixture of solid

components and the atom interaction on the growing surface during the crystallization. It was established that only an introduction in the thermodynamic calculation of atom interaction on the growing surface, which increase in the following sequence: Ta, W, Re, Nb, V, Mo, results in a rise of yield of VB group metals under their co-deposition with tungsten, excepting W-Ta system. This may explain the experimentally observed tungsten yield rise under its alloying with rhenium and molibdenium.

4. The thermodynamic analysis, performed by taking into account the formation of solid lower-valent fluorides and excess enthalpy of atom interaction during crystallization, showed that the moving force of CVD of the alloys from the W-M-F-H systems (the supersaturation in these systems) increase in order: Ta, Nb, V, Mo, W, Re.
5. A lot of applications of tungsten coatings, deposited from tungsten hexafluoride and hydrogen mixture at low temperature, as well as tungsten alloys and carbides are reviewed in this chapter.

## 8. Acknowledgments

This work was supported by the Russian Foundation for Basic Research, project No. 09-08-182.

## 9. Appendix 1

Description of symbols used in the text

Symbol	Description
$\Omega$	Atomization energy
M	Metal
X	Halid
n	Valency of metal
$\Delta_f H$	Formation enthalpy
at	Atom
$\varphi$	Function
$Z_m$	Atomic number of metal
$Z_x$	Atomic number of halid
$\psi$	Functional
$\Delta_s H$	Sublimation enthalpy
S	Entropy
$\Delta_f H^{\circ}_{298} (g)$	Standart formation enthalpy at 298 K at gaseous state
$\Delta_f H^{\circ} (s)$	Standart formation enthalpy at 298 K at solid state
$\Delta_s H^{\circ}_{298}$	Standart sublimation enthalpy at 298 K
$S^{\circ}_{298} (g)$	Standart entropy at 298 K at gaseous state
$S^{\circ}_{298} (s)$	Standart entropy at 298 K at solid state
$C_p$	Specific heat at constant stress
$\Delta H_m$	Partial enthalpy of mixing
$\Delta_s \bar{H}$	Partial molar enthalpy
$\Delta H^{\circ}_m$	Standart mixing enthalpy

## 10. References

- [1] Korolev Yu. M., Stolyarov V. I., Vosstanovlenie fluoridov tugoplavkikh metallov vodorodom (Metallurgii, Moskva, 1981) 184 p. (in Russian).
- [2] Krasovskii A.I., Chuzhko R.K., Tregulov V.R., Balakhovskii O.A., Fluoridnii process polucheniya volframa (Nauka, Moskva, 1981) 260 p. (in Russian).
- [3] Pons M., Benezech A., Huguet P., et al., J. Phys. France, Vol. 5, N. 8 (1995) pp. 1145-1160.
- [4] Lakhotkin Yu.V., Krasovskii A.I., Volfram-renieviae pokritiya (Nauka, Moskva, 1989) 158 p. (in Russian).
- [5] Lakhotkin Yu.V., Protection of Metals, Vol. 44, N. 4 (2008) pp. 319-332.
- [6] Blokhinzev D.I. Osnovi kvantovoi mekhaniki (Nauka, Moskva, 1983) 664 p. (in Russian).
- [7] Malandin M.B., Lakhotkin Yu.V., Kuzmin V.P., Problemi fizicheskogo metallovedeniya (MIFI, Moskva, 1991) pp. 35-47. (in Russian).
- [8] Bersuker I.B., Elektronnoe stroenie i svoystva koordinatsionnykh soedinenii. Vvedenie v teoriyu (Khimiya, Leningrad, 1976) 312 p. (in Russian).
- [9] Charkin O.P., Stabilnost i struktura gasoobrasnykh neorganicheskikh molekul, radikalov i ionov (Nauka, Moskva, 1980) 280 p. (in Russian).
- [10] Termicheskie konstanty veshchestv. Spravochnik (Izd. AN SSSR, Moskva, 1962-1981) 10 t. (in Russian).
- [11] Gurvich L.V., Veiz I.V., Medvedev V.A. et al., Termodinamicheskie svoystva individualnykh veshchestv (Nauka, Moskva, 1978-1982) 4 t. (in Russian).
- [12] Molekulyarnye postoyannye neorganicheskikh soedinenii. Spravochnik (Khimiya, Leningrad, 1979) 448 p. (in Russian).
- [13] Sidopov L.N., Sholz V.B., J. Fiz. Khimii, T. 45, N. 2 (1971) pp. 275-280 (in Russian).
- [14] Sidopov L.N., Denisov M.Ya., Akishin P.A. et al., J. Fiz. Khimii, T. 40, N. 5 (1966) pp. 1151-1154 (in Russian).
- [15] Gusarov A.V., Pervov V.S., Gotkis I.S. et al., DAN SSSR, T. 216, N. 6 (1974) pp. 1296-1299 (in Russian).
- [16] Lau K.H., Hildenbrand D.L. J. Chem. Phys., Vol. 71, N. 4 (1979) pp. 1572-1577.
- [17] Hildenbrand D.L. J. Chem. Phys., Vol. 65, N. 2 (1976) pp. 614-618.
- [18] Alikhanyan A.S., Pervov V.S., Malkerova N.P. et al., J. Neorganicheskoi khimii, T. 23, N. 6 (1978) pp. 1483-1485 (in Russian).
- [19] Nuttal R.L., Kilday M.Y., Churney K.L., Natt. Bur. Stand. Rep. 73-281, (1973).
- [20] Gotkis I.S., Gusarov A.V., Pervov V.S., et al., Koordinatsionnaya khimiya T. 4, Vip. 5 (1978) pp. 720-724 (in Russian).
- [21] Hildenbrand D.L. J. Chem. Phys., Vol. 62, N. 8 (1975) pp. 3074-3079.
- [22] Burgess J., Fawcett J., Peacock R.D. et al., J. Chem. Soc., Dalton Trans., N. 14, (1976) pp. 1363-1364.
- [23] Politov Yu.A., Alikhanyan A.S., Butzki V.D., et al., DAN SSSR, T. 309, N. 4 (1989) pp. 897-899 (in Russian).
- [24] Politov Yu.A., Alikhanyan A.S., Butzki V.D., et al., J. Neorganicheskoi khimii, T. 32, N. 2 (1987) pp. 520-523 (in Russian).
- [25] Stout J.W., Boo W.O.J. J. Chem. Phys., Vol. 71, N. 1 (1979) pp. 1-8.
- [26] Stull D.R., Prophet H. JANAF Thermochemical Tables. NSRDS-NBS 37 US, (NBS, Washington, DC, 1971).
- [27] Arara R., Pollard R. J. Electrochem. Soc., Vol. 138, N. 5 (1991) pp. 1523-1537.

- [28] Boltalina O.V., Borzsevskii A.Ya., Sidorov L.N. J. Fiz. Chimii, T. 66, Vip. 9 (1992) pp. 2289-2309 (in Russian).
- [29] Amatucci G.G., Pereira N., Journal of Fluorine Chemistry, Vol.128 Iss. 4 (2007) pp. 243-262.
- [30] Peacock R.D. Adv. In Fluorine Chem., N. 7 (1973) pp. 113-145 .
- [31] Lakhotkin Yu.V. Journal de Physique. Colloque C5, supplement au Journal de Physique II 5 (1995) pp. 199-204.
- [32] Lakhotkin Yu.V., Goncharov V.L., Protection of Metals, Vol. 44, N. 7 (2008) pp. 637-643.
- [33] Lakhotkin Yu.V., Kuzmin V.P., Goncharov V.L., Protection of Metals and Physical Chemistry of Surfaces, Vol. 45, N. 7 (2009) pp. 833-837.
- [34] Kaufman L., Bernstein Kh., Raschet diagramm sostojnij s pomoschyu EVM (Mir, Moskva, 1972) 328 p. (in Russian).
- [35] Plummer E.W., Rhodin T.N., J. Chem. Phys., Vol. 49, N. 8, (1968) pp. 3479-3496.
- [36] Bell D.A., Falconer J.L., J. Electrochem. Soc., Vol. 142, Iss. 7 (1995) pp. 2401-2404.
- [37] Lakhotkin Yu. V., Kuzmin V.P. Tungsten carbide coatings and method for production the same. Patent EP 1 158 070 A1.28.11. Bulletin 2001/48. 28.11.2001.
- [38] Lakhotkin Yu. V., Aleksandrov S.A., Zhuk Yu. N. Self-sharpening cutting tool with hard coating. Patent US. 20050158589 A1. July 21. 2005.
- [39] Lakhotkin Yu. V., Kuzmin V.P. Adhesive composite coating for diamond and diamond-containing materials and method for production said coating. Patent EP 1 300 380 A1. Bulletin 2003/15. 09.04.2003.



# Effect of Stagnation Temperature on Supersonic Flow Parameters with Application for Air in Nozzles

Toufik Zebbiche  
University SAAD Dahleb of Blida,  
Algeria

## 1. Introduction

The obtained results of a supersonic perfect gas flow presented in (Anderson, 1982, 1988 & Ryhming, 1984), are valid under some assumptions. One of the assumptions is that the gas is regarded as a calorically perfect, i. e., the specific heats  $C_p$  is constant and does not depend on the temperature, which is not valid in the real case when the temperature increases (Zebbiche & Youbi, 2005b, 2006, Zebbiche, 2010a, 2010b). The aim of this research is to develop a mathematical model of the gas flow by adding the variation effect of  $C_p$  and  $\gamma$  with the temperature. In this case, the gas is named by calorically imperfect gas or *gas at high temperature*. There are tables for air (Peterson & Hill, 1965) for example that contain the values of  $C_p$  and  $\gamma$  versus the temperature in interval 55 K to 3550 K. We carried out a polynomial interpolation of these values in order to find an analytical form for the function  $C_p(T)$ .

The presented mathematical relations are valid in the general case independently of the interpolation form and the substance, but the results are illustrated by a polynomial interpolation of the 9<sup>th</sup> degree. The obtained mathematical relations are in the form of nonlinear algebraic equations, and so analytical integration was impossible. Thus, our interest is directed towards to the determination of numerical solutions. The dichotomy method for the solution of the nonlinear algebraic equations is used; the Simpson's algorithm (Démidovitch & Maron, 1987 & Zebbiche & Youbi, 2006, Zebbiche, 2010a, 2010b) for numerical integration of the found functions is applied. The integrated functions have high gradients of the interval extremity, where the Simpson's algorithm requires a very high discretization to have a suitable precision. The solution of this problem is made by introduction of a condensation procedure in order to refine the points at the place where there is high gradient. The Robert's condensation formula presented in (Fletcher, 1988) was chosen. The application for the air in the supersonic field is limited by the threshold of the molecules dissociation. The comparison is made with the calorically perfect gas model.

The problem encounters in the aeronautical experiments where the use of the nozzle designed on the basis of the perfect gas assumption, degrades the performances. If during the experiment measurements are carried out it will be found that measured parameters are differed from the calculated, especially for the high stagnation temperature. Several reasons

are responsible for this deviation. Our flow is regarded as perfect, permanent and non-rotational. The gas is regarded as calorically imperfect and thermally perfect. The theory of perfect gas does not take account of this temperature.

To determine the application limits of the perfect gas model, the error given by this model is compared with our results.

## 2. Mathematical formulation

The development is based on the use of the conservation equations in differential form. We assume that the state equation of perfect gas ( $P=\rho RT$ ) remains valid, with  $R=287.102$  J/(kg K). For the adiabatic flow, the temperature and the density of a perfect gas are related by the following differential equation (Moran, 2007 & Oosthuisen & Carscallen, 1997 & Zuker & Bilbarz, 2002, Zebbiche, 2010a, 2010b).

$$\frac{C_p}{\gamma} dT - \frac{RT}{\rho} d\rho = 0 \quad (1)$$

Using relationship between  $C_p$  and  $\gamma$  [ $C_p=\gamma R/(\gamma-1)$ ], the equation (1) can be written at the following form:

$$\frac{d\rho}{\rho} = \frac{dT}{T [\gamma(T)-1]} \quad (2)$$

The integration of the relation (2) gives the adiabatic equation of a perfect gas at high temperature.

The sound velocity is (Ryhming, 1984),

$$a^2 = \left( \frac{dP}{d\rho} \right)_{\text{entropy}=\text{constant}} \quad (3)$$

The differentiation of the state equation of a perfect gas gives:

$$\frac{dP}{d\rho} = \rho R \frac{dT}{d\rho} + R T \quad (4)$$

Substituting the relationship (2) in the equation (4), we obtain after transformation:

$$a^2(T) = \gamma(T) R T \quad (5)$$

Equation (5) proves that the relation of speed of sound of perfect gas remains always valid for the model at high temperature, but it is necessary to take into account the variation of the ratio  $\gamma(T)$ .

The equation of the energy conservation in differential form (Anderson, 1988 & Moran, 2007) is written as:

$$C_p dT + V dV = 0 \quad (6)$$

The integration between the stagnation state ( $V_0 \approx 0, T_0$ ) and supersonic state ( $V, T$ ) gives:

$$V^2 = 2 H(T) \tag{7}$$

Where

$$H(T) = \int_T^{T_0} C_p(T) dT \tag{8}$$

Dividing the equation (6) by  $V^2$  and substituting the relation (7) in the obtained result, we obtain:

$$\frac{dV}{V} = - \frac{C_p(T)}{2 H(T)} dT \tag{9}$$

Dividing the relation (7) by the sound velocity, we obtain an expression connecting the Mach number with the enthalpy and the temperature:

$$M(T) = \frac{\sqrt{2 H(T)}}{a(T)} \tag{10}$$

The relation (10) shows the variation of the Mach number with the temperature for calorically imperfect gas.

The momentum equation in differential form can be written as (Moran, 2007, Peterson & Hill1, 1965, & Oosthuisen & Carscallen, 1997):

$$V dV + \frac{dP}{\rho} = 0 \tag{11}$$

Using the expression (3), the relationship (10), can be written as:

$$\frac{d\rho}{\rho} = F_\rho(T) dT \tag{12}$$

Where

$$F_\rho(T) = \frac{C_p(T)}{a^2(T)} \tag{13}$$

The density ratio relative to the temperature  $T_0$  can be obtained by integration of the function (13) between the stagnation state  $(\rho_0, T_0)$  and the concerned supersonic state  $(\rho, T)$ :

$$\frac{\rho}{\rho_0} = \text{Exp} \left( - \int_T^{T_0} F_\rho(T) dT \right) \tag{14}$$

The pressure ratio is obtained by using the relation of the perfect gas state:

$$\frac{P}{P_0} = \left( \frac{\rho}{\rho_0} \right) \left( \frac{T}{T_0} \right) \quad (15)$$

The mass conservation equation is written as (Anderson, 1988 & Moran, 2007)

$$\rho V A = \text{constant} \quad (16)$$

The taking logarithm and then differentiating of relation (16), and also using of the relations (9) and (12), one can receive the following equation:

$$\frac{dA}{A} = F_A(T) dT \quad (17)$$

Where

$$F_A(T) = C_p(T) \left[ \frac{1}{a^2(T)} - \frac{1}{2H(T)} \right] \quad (18)$$

The integration of equation (17) between the critical state ( $A^*$ ,  $T^*$ ) and the supersonic state ( $A$ ,  $T$ ) gives the cross-section areas ratio:

$$\frac{A}{A^*} = \text{Exp} \left( \int_T^{T^*} F_A(T) dT \right) \quad (19)$$

To find parameters  $\rho$  and  $A$ , the integrals of functions  $F_\rho(T)$  and  $F_A(T)$  should be found. As the analytical procedure is impossible, our interest is directed towards the numerical calculation. *All parameters  $M$ ,  $\rho$  and  $A$  depend on the temperature.*

The critical mass flow rate (Moran, 2007, Zebbiche & Youbi, 2005a, 2005b) can be written in non-dimensional form:

$$\frac{\dot{m}}{A^* \rho_0 a_0} = \int_{A^*}^A \left( \frac{\rho}{\rho_0} \right) \left( \frac{a}{a_0} \right) M \cos(\theta) \frac{dA}{A^*} \quad (20)$$

As the mass flow rate through the throat is constant, we can calculate it at the throat. In this section, we have  $\rho = \rho^*$ ,  $a = a^*$ ,  $M = 1$ ,  $\theta = 0$  and  $A = A^*$ . Therefore, the relation (20) is reduced to:

$$\frac{\dot{m}}{A^* \rho_0 a_0} = \left( \frac{\rho^*}{\rho_0} \right) \left( \frac{a^*}{a_0} \right) \quad (21)$$

The determination of the velocity sound ratio is done by the relation (5). Thus,

$$\frac{a}{a_0} = \left[ \frac{\gamma(T)}{\gamma(T_0)} \right]^{1/2} \left[ \frac{T}{T_0} \right]^{1/2} \quad (22)$$

The parameters  $T$ ,  $P$ ,  $\rho$  and  $A$  for the perfect gas are connected explicitly with the Mach number, which is the basic variable for that model. For our model, the basic variable is the temperature because of the implicit equation (10) connecting  $M$  and  $T$ , where the reverse analytical expression does not exist.

### 3. Calculation procedure

In the first case, one presents the table of variation of  $C_p$  and  $\gamma$  versus the temperature for air (Peterson & Hill, 1965, Zebbiche 2010a, 2010b). The values are presented in the table 1.

T (K)	$C_p$ (J/(KgK))	$\gamma(T)$	T (K)	$C_p$ (J/(Kg K))	$\gamma(T)$	T (K)	$C_p$ J/(Kg K)	$\gamma(T)$
55.538	1001.104	1.402	833.316	1107.192	1.350	2111.094	1256.813	1.296
.	.	.	888.872	1119.078	1.345	2222.205	1263.410	1.294
222.205	1001.101	1.402	944.427	1131.314	1.340	2333.316	1270.097	1.292
277.761	1002.885	1.401	999.983	1141.365	1.336	2444.427	1273.476	1.291
305.538	1004.675	1.400	1055.538	1151.658	1.332	2555.538	1276.877	1.290
333.316	1006.473	1.399	1111.094	1162.202	1.328	2666.650	1283.751	1.288
361.094	1008.281	1.398	1166.650	1170.280	1.325	2777.761	1287.224	1.287
388.872	1011.923	1.396	1222.205	1178.509	1.322	2888.872	1290.721	1.286
416.650	1015.603	1.394	1277.761	1186.893	1.319	2999.983	1294.242	1.285
444.427	1019.320	1.392	1333.316	1192.570	1.317	3111.094	1297.789	1.284
499.983	1028.781	1.387	1444.427	1204.142	1.313	3222.205	1301.360	1.283
555.538	1054.563	1.374	1555.538	1216.014	1.309	3333.316	1304.957	1.282
611.094	1054.563	1.370	1666.650	1225.121	1.306	3444.427	1304.957	1.282
666.650	1067.077	1.368	1777.761	1234.409	1.303	3555.538	1308.580	1.281
722.205	1080.005	1.362	1888.872	1243.883	1.300			
777.761	1093.370	1.356	1999.983	1250.305	1.298			

Table 1. Variation of  $C_p(T)$  and  $\gamma(T)$  versus the temperature for air.

For a perfect gas, the  $\gamma$  and  $C_p$  values are equal to  $\gamma=1.402$  and  $C_p=1001.28932$  J/(kgK) (Oosthuisen & Carscallen, 1997, Moran, 2007 & Zuker & Bilbarz, 2002).. The interpolation of the  $C_p$  values according to the temperature is presented by relation (23) in the form of Horner scheme to minimize the mathematical operations number (Zebbiche, 2010a, 2010b):

$$C_p(T) = a_1 + T(a_2 + T(a_3 + T(a_4 + T(a_5 + T(a_6 + T(a_7 + T(a_8 + T(a_9 + T(a_{10})))))))))) \quad (23)$$

The interpolation ( $a_i$   $i=1, 2, \dots, 10$ ) of constants are illustrated in table 2.

$I$	$a_i$	$I$	$a_i$
1	1001.1058	6	3.069773 10 <sup>-12</sup>
2	0.04066128	7	-1.350935 10 <sup>-15</sup>
3	-0.000633769	8	3.472262 10 <sup>-19</sup>
4	2.747475 10 <sup>-6</sup>	9	-4.846753 10 <sup>-23</sup>
5	-4.033845 10 <sup>-9</sup>	10	2.841187 10 <sup>-27</sup>

Table 2. Coefficients of the polynomial  $C_p(T)$ .

A relationship (23) gives undulated dependence for temperature approximately low than  $\bar{T} = 240$  K . So for this field, the table value (Peterson & Hill, 1965), was taken

$$\bar{C}_p = C_p(\bar{T}) = 1001.15868 \text{ J / (kg K)}$$

Thus:

for  $T \leq \bar{T}$ , we have  $C_p(T) = \bar{C}_p$

for  $T > \bar{T}$ , relation (23) is used.

The selected interpolation gives an error less than  $\varepsilon=10^{-3}$  between the table and interpolated values.

Once the interpolation is made, we determine the function  $H(T)$  of the relation (8), by integrating the function  $C_p(T)$  in the interval  $[T, T_0]$ . Then,  $H(T)$  is a function with a parameter  $T_0$  and it is defined when  $T \leq T_0$ .

Substituting the relation (23) in (8) and writing the integration results in the form of Horner scheme, the following expression for enthalpy is obtained

$$H(T) = H_0 - [c_1 + T(c_2 + T(c_3 + T(c_4 + T(c_5 + Tc_6 + T(c_7 + T(c_8 + T(c_9 + T(c_{10})))))))))] \tag{24}$$

Where

$$H_0 = T_0(c_1 + T_0(c_2 + T_0(c_3 + T_0(c_4 + T_0(c_5 + T_0(c_6 + T_0(c_7 + T_0(c_8 + T_0(c_9 + T_0(c_{10})))))))))) \tag{25}$$

and

$$c_i = \frac{a_i}{i} \quad (i = 1, 2, 3, \dots, 10)$$

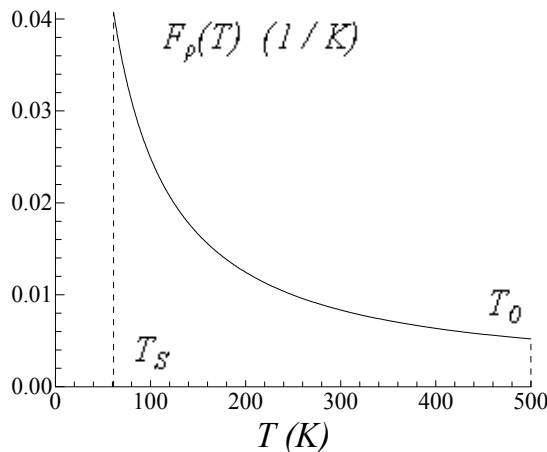


Fig. 1. Variation of function  $F_p(T)$  in the interval  $[T_S, T_0]$  versus  $T_0$ .

Taking into account the correction made to the function  $C_p(T)$ , the function  $H(T)$  has the following form:

For  $T_0 < \bar{T}$ ,  $H(T) = \bar{C}_p (T_0 - T)$

For  $T_0 > \bar{T}$ , we have two cases:

if  $T > \bar{T}$  :  $H(T) =$  relation (24)

if  $T \leq \bar{T}$  :  $H(T) = \bar{C}_p(\bar{T} - T) + H(\bar{T})$

The determination of the ratios (14) and (19) require the numerical integration of  $F_p(T)$  and  $F_A(T)$  in the intervals  $[T, T_0]$  and  $[T, T^*]$  respectively. We carried out preliminary calculation of these functions (Figs. 1, 2) to see their variations and to choice the integration method.

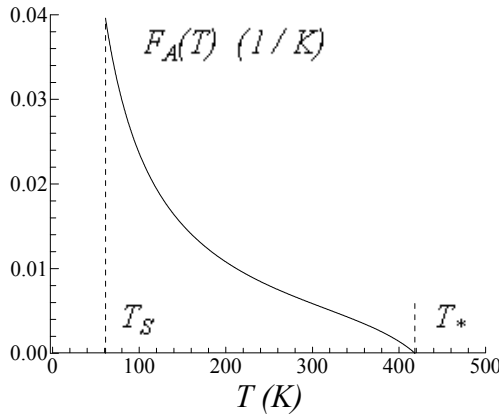


Fig. 2. Variation of the function  $F_A(T)$  in the interval  $[T_S, T^*]$  versus  $T_0$

Due to high gradient at the left extremity of the interval, the integration with a constant step requires a very small step. The tracing of the functions is selected for  $T_0=500$  K (low temperature) and  $M_S=6.00$  (extreme supersonic) for a good representation in these ends. In this case, we obtain  $T^*=418.34$  K and  $T_S=61.07$  K. the two functions presents a very large derivative at temperature  $T_S$ .

A Condensation of nodes is then necessary in the vicinity of  $T_S$  for the two functions. The goal of this condensation is to calculate the value of integral with a high precision in a reduced time by minimizing the nodes number. The Simpson’s integration method (Démidovitch & Maron, 1987 & Zebbiche & Youbi, 2006) was chosen. The chosen condensation function has the following form (Zebbiche & Youbi, 2005a):

$$s_i = b_1 z_i + (1 - b_1) \left[ 1 - \frac{\tanh[b_2 \cdot (1 - z_i)]}{\tanh(b_2)} \right] \tag{26}$$

Where

$$z_i = \frac{i - 1}{N - 1} \quad 1 \leq i \leq N \tag{27}$$

Obtained  $s_i$  values, enable to find the value of  $T_i$  in nodes  $i$ :

$$T_i = s_i (T_D - T_G) + T_G \quad (28)$$

The temperature  $T_D$  is equal to  $T_0$  for  $F_\rho(T)$ , and equal to  $T^*$  for  $F_A(T)$ . The temperature  $T_G$  is equal to  $T^*$  for the critical parameter, and equal to  $T_S$  for the supersonic parameter. Taking a value  $b_1$  near zero ( $b_1=0.1$ , for example) and  $b_2=2.0$ , it can condense the nodes towards left edge  $T_S$  of the interval, see figure 3.

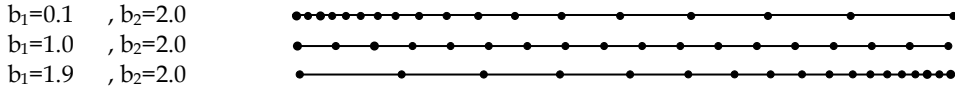


Fig. 3. Presentation of the condensation of nodes

### 3.1 Critical parameters

The stagnation state is given by  $M=0$ . Then, the critical parameters correspond to  $M=1.00$ , for example at the throat of a supersonic nozzle, summarize by:

When  $M=1.00$  we have  $T=T^*$ . These conditions in the relation (10), we obtain:

$$2 H(T^*) - a^2(T^*) = 0 \quad (29)$$

The resolution of equation (29) is made by the use of the dichotomy algorithm (Démidovitch & Maron, 1987 & Zebbiche & Youbi, 2006), with  $T^* < T_0$ . It can choose the interval  $[T_1, T_2]$  containing  $T^*$  by  $T_1=0$  K and  $T_2=T_0$ . The value  $T^*$  can be given with a precision  $\varepsilon$  if the interval of subdivision number  $K$  is satisfied by the following condition:

$$K = 1.4426 \text{Log} \left( \frac{T_0}{\varepsilon} \right) + 1 \quad (30)$$

If  $\varepsilon=10^{-8}$  is taken, the number  $K$  cannot exceed 39. Consequently, the temperature ratio  $T^*/T_0$  can be calculated.

Taking  $T=T^*$  and  $\rho=\rho^*$  in the relation (14) and integrating the function  $F_\rho(T)$  by using the Simpson's formula with condensation of nodes towards the left end, the critical density ratio is obtained.

The critical ratios of the pressures and the sound velocity can be calculated by using the relations (15) and (22) respectively, by replacing  $T=T^*$ ,  $\rho=\rho^*$ ,  $P=P^*$  and  $a=a^*$ ,

### 3.2 Parameters for a supersonic Mach number

For a given supersonic cross-section, the parameters  $\rho=\rho_S$ ,  $P=P_S$ ,  $A=A_S$ , and  $T=T_S$  can be determined according to the Mach number  $M=M_S$ . Replacing  $T=T_S$  and  $M=M_S$  in relation (10) gives

$$2 H(T_S) - M_S^2 a^2(T_S) = 0 \quad (31)$$

The determination of  $T_S$  of equation (31) is done always by the dichotomy algorithm, excepting  $T_S < T^*$ . We can take the interval  $[T_1, T_2]$  containing  $T_S$ , by ( $T_1=0$  K, and  $T_2=T^*$ ).

Replacing  $T=T_S$  and  $\rho=\rho_S$  in relation (14) and integrating the function  $F_\rho(T)$  by using the Simpson's method with condensation of nodes towards the left end, the density ratio can be obtained.



The ratios of pressures, speed of sound and the sections corresponding to  $M=M_S$  can be calculated respectively by using the relations (15), (22) and (19) by replacing  $T=T_S$ ,  $\rho=\rho_S$ ,  $P=P_S$ ,  $a=a_S$  and  $A=A_S$ .

The integration results of the ratios  $\rho^*/\rho_0$ ,  $\rho_S/\rho_0$  and  $A_S/A^*$  primarily depend on the values of  $N$ ,  $b_1$  and  $b_2$ .

### 3.3 Supersonic nozzle conception

For supersonic nozzle application, it is necessary to determine the thrust coefficient. For nozzles giving a uniform and parallel flow at the exit section, the thrust coefficient is (Peterson & Hill, 1965 & Zebbiche, Youbi, 2005b)

$$C_F = \frac{F}{P_0 A^*} \quad (32)$$

Where

$$F = m V_E = m M_E a_E \quad (33)$$

The introduction of relations (21), (22) into (32) gives as the following relation:

$$C_F = \gamma(T_0) M_E \left( \frac{a_E}{a_0} \right) \left( \frac{\rho^*}{\rho^*} \right) \left( \frac{a^*}{a_0} \right) \quad (34)$$

The design of the nozzle is made on the basis of its application. For rockets and missiles applications, the design is made to obtain nozzles having largest possible exit Mach number, which gives largest thrust coefficient, and smallest possible length, which give smallest possible mass of structure.

For the application of blowers, we make the design on the basis to obtain the smallest possible temperature at the exit section, to not to destroy the measuring instruments, and to save the ambient conditions. Another condition requested is to have possible largest ray of the exit section for the site of instruments. Between the two possibilities of construction, we prefer the first one.

### 3.4 Error of perfect gas model

The mathematical perfect gas model is developed on the basis to regarding the specific heat  $C_p$  and ratio  $\gamma$  as constants, which gives acceptable results for low temperature. According to this study, we can notice a difference on the given results between the perfect gas model and developed here model. The error given by the *PG* model compared to our *HT* model can be calculated for each parameter. Then, for each value  $(T_0, M)$ , the  $\varepsilon$  error can be evaluated by the following relationship:

$$\varepsilon_y(T_0, M) = \left| 1 - \frac{y_{PG}(T_0, M)}{y_{HT}(T_0, M)} \right| \times 100 \quad (35)$$

The letter  $y$  in the expression (35) can represent all above-mentioned parameters. As a rule for the aerodynamic applications, the error should be lower than 5%.

#### 4. Application

The design of a supersonic propulsion nozzle can be considered as example. The use of the obtained dimensioned nozzle shape based on the application of the *PG* model given a supersonic uniform Mach number  $M_S$  at the exit section of rockets, degrades the desired performances (exit Mach number, pressure force), especially if the temperature  $T_0$  of the combustion chamber is higher. We recall here that the form of the nozzle structure does not change, except the thermodynamic behaviour of the air which changes with  $T_0$ . Two situations can be presented.

The first situation presented is that, if we wants to preserve the same variation of the Mach number throughout the nozzle, and consequently, the same exit Mach number  $M_E$ , is necessary to determine by the application of our model, the ray of each section and in particular the ray of the exit section, which will give the same variation of the Mach number, and consequently another shape of the nozzle will be obtained.

$$M_S (HT) = M_S (PG) \quad (36)$$

$$M_S (PG) = \frac{\sqrt{2H[T_{S(HT)}]}}{a[T_{S(HT)}]} \quad (37)$$

$$\frac{A_S (HT)}{A_*} = e^{\int_{T_S (HT)}^{T_*} F_A(T) dT} > \frac{A_S (PG)}{A_*} \quad (38)$$

The relation (36) indicates that the Mach number of the *PG* model is preserved for each section in our calculation. Initially, we determine the temperature at each section; witch presents the solution of equation (37). To determine the ratio of the sections, we use the relation (38). The ratio of the section obtained by our model will be superior that that determined by the *PG* model as present equation (38). Then the shape of the nozzle obtained by *PG* model is included in the nozzle obtained by our model. The temperature  $T_0$  presented in equation (38) is that correspond to the temperature  $T_0$  for our model.

The second situation consists to preserving the shape of the nozzle dimensioned on the basis of *PG* model for the aeronautical applications considered the *HT* model.

$$\frac{A_S (HT)}{A_*} = \frac{A_S (PG)}{A_*} \quad (39)$$

$$M_S (HT) < M_S (PG) \quad (40)$$

The relation (39) presents this situation. In this case, the nozzle will deliver a Mach number lower than desired, as shows the relation (40). The correction of the Mach number for *HT* model is initially made by the determination of the temperature  $T_S$  as solution of equation (38), then determine the exit Mach number as solution of relation (37). The resolution of equation (38) is done by combining the dichotomy method with Simpson's algorithm.

### 5. Results and comments

Figures 4 and 5 respectively represent the variation of specific heat  $C_p(T)$  and the ratio  $\gamma(T)$  of the air versus the temperature up to 3550 K for HT and PG models. The graphs at high temperature are presented by using the polynomial interpolation (23). We can say that at low temperature until approximately 240 K, the gas can be regarded as calorically perfect, because of the invariance of specific heat  $C_p(T)$  and the ratio  $\gamma(T)$ . But if  $T_0$  increases, we can see the difference between these values and it influences on the thermodynamic parameters of the flow.

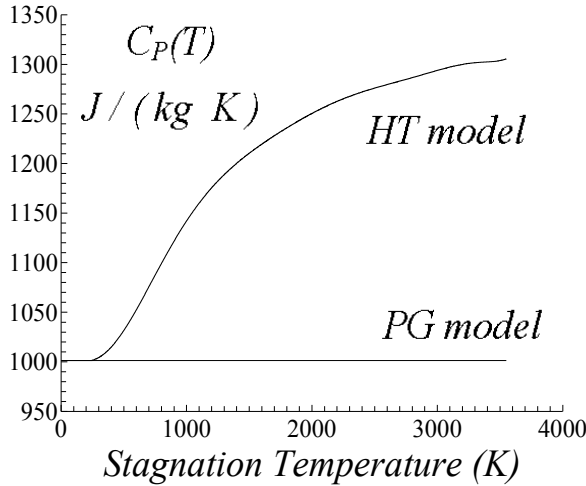


Fig. 4. Variation of the specific heat for constant pressure versus stagnation temperature  $T_0$ .

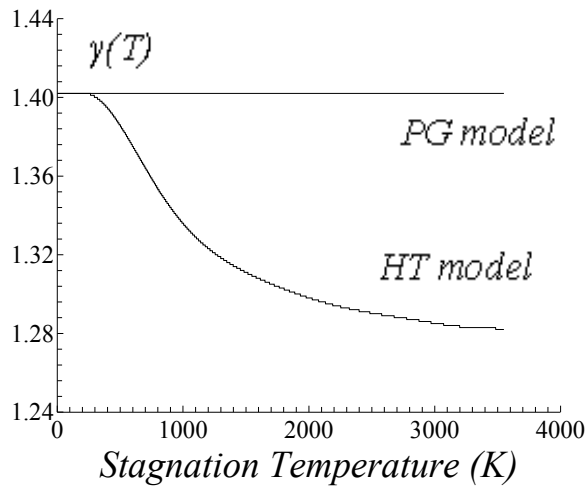


Fig. 5. Variation of the specific heats ratio versus  $T_0$ .

### 5.1 Results for the critical parameters

Figures 6, 7 and 8 represent the variation of the critical thermodynamic ratios versus  $T_0$ . It can be seen that with enhancement  $T_0$ , the critical parameters vary, and this variation becomes considerable for high values of  $T_0$  unlike to the *PG* model, where they do not depend on  $T_0$ . For example, the value of the temperature ratio given by the *HT* model is always higher than the value given by the *PG* model. The ratios are determined by the choice of  $N=300000$ ,  $b_1=0.1$  and  $b_2=2.0$  to have a precision better than  $\varepsilon=10^{-5}$ . The obtained numerical values of the critical parameters are presented in the table 3.

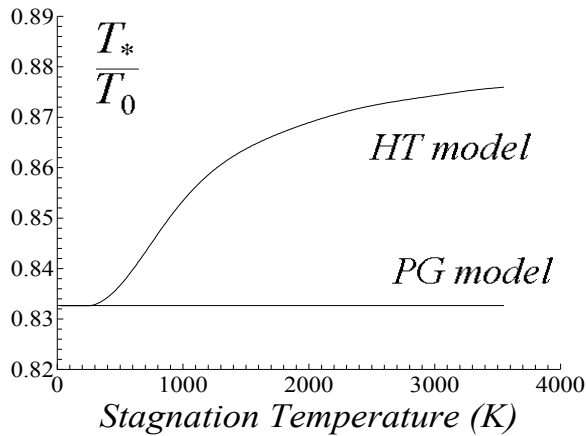


Fig. 6. Variation of  $T^*/T_0$  versus  $T_0$ .

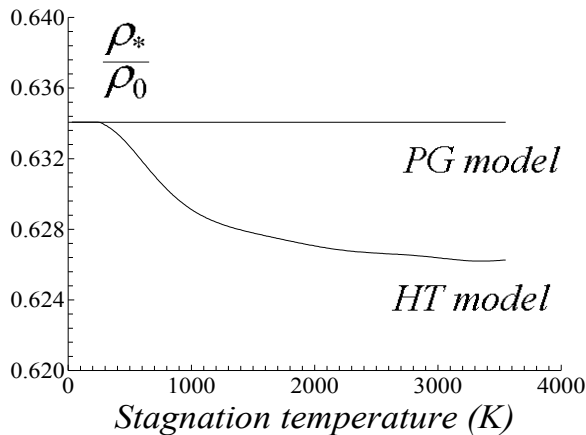


Fig. 7. Variation of  $\rho^*/\rho_0$  versus  $T_0$ .

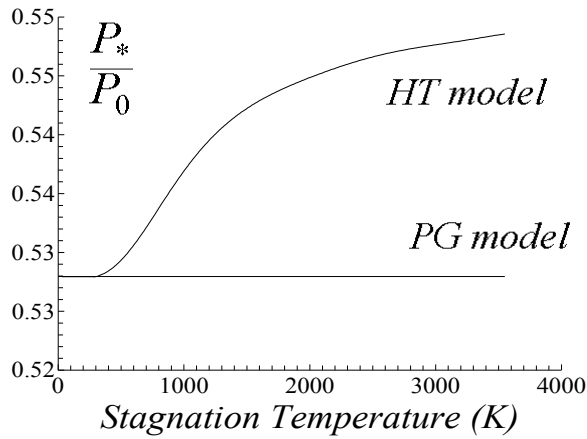


Fig. 8. Variation of  $P^*/P_0$  versus  $T_0$ .

Figure 9 shows that mass flow rate through the critical cross section given by the perfect gas theory is lower than it is at the HT model, especially for values of  $T_0$ .

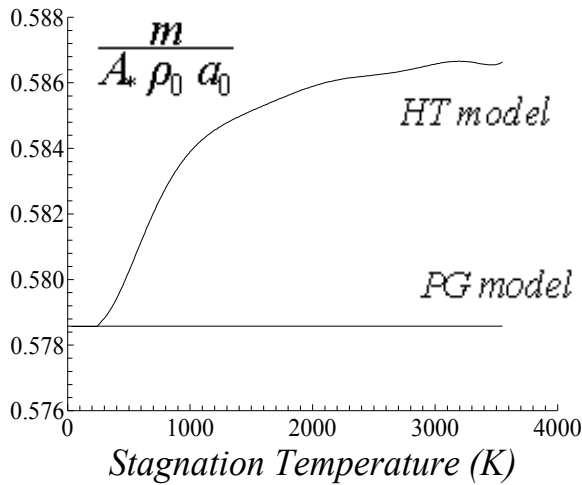


Fig. 9. Variation of the non-dimensional critical mass flow rate with  $T_0$ .

Figure 10 presents the variation of the critical sound velocity ratio versus  $T_0$ . The influence of the  $T_0$  on this parameter can be found.

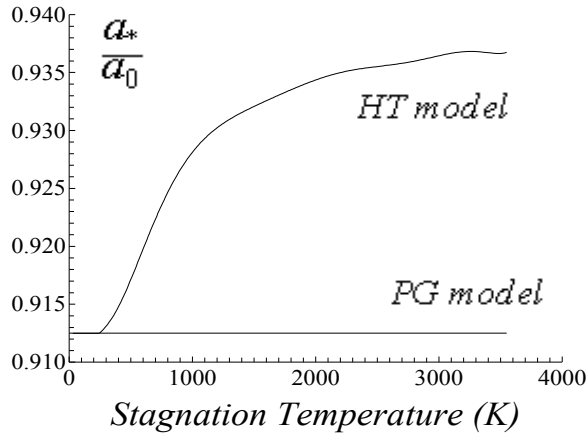


Fig. 10. Effect of  $T_0$  on the velocity sound ratio.

	$\frac{T_*}{T_0}$	$\frac{P_*}{P_0}$	$\frac{\rho_*}{\rho_0}$	$\frac{a_*}{a_0}$	$\frac{m}{A_* \rho_0 a_0}$
PG ( $\gamma=1.402$ )	0.8326	0.5279	0.6340	0.9124	0.5785
$T_0=298.15$ K	0.8328	0.5279	0.6339	0.9131	0.5788
$T_0=500$ K	0.8366	0.5293	0.6326	0.9171	0.5802
$T_0=1000$ K	0.8535	0.5369	0.6291	0.9280	0.5838
$T_0=2000$ K	0.8689	0.5448	0.6270	0.9343	0.5858
$T_0=2500$ K	0.8722	0.5466	0.6266	0.9355	0.5862
$T_0=3000$ K	0.8743	0.5475	0.6263	0.9365	0.5865
$T_0=3500$ K	0.8758	0.5484	0.6262	0.9366	0.5865

Table 3. Numerical values of the critical parameters at high temperature.

## 5.2 Results for the supersonic parameters

Figures 11, 12 and 13 presents the variation of the supersonic flow parameters in a cross-section versus Mach number for  $T_0=1000$  K, 2000 K and 3000 K, including the case of perfect gas for  $\gamma=1.402$ . When  $M=1$ , we can obtain the values of the critical ratios. If we take into account the variation of  $C_p(T)$ , the temperature  $T_0$  influences on the value of the thermodynamic and geometrical parameters of flow unlike the *PG* model.

The curve 4 of figure 11 is under the curves of the *HT* model, which indicates that the perfect gas model cool the flow compared to the real thermodynamic behaviour of the gas, and consequently, it influences on the dimensionless parameters of a nozzle. At low temperature and Mach number, the theory of perfect gas gives acceptable results. The obtained numerical values of the supersonic flow parameters, the cross section area ratio and sound velocity ratio are presented respectively if the tables 4, 5, 6, 7 and 8.

$T/T_0$	M=2.00	M=3.00	M=4.00	M=5.00	M=6.00
PG ( $\gamma=1.402$ )	0.5543	0.3560	0.2371	0.1659	0.1214
$T_0=298.15$ K	0.5544	0.3560	0.2372	0.1659	0.1214
$T_0=500$ K	0.5577	0.3581	0.2386	0.1669	0.1221
$T_0=1000$ K	0.5810	0.3731	0.2481	0.1736	0.1269
$T_0=1500$ K	0.6031	0.3911	0.2594	0.1810	0.1323
$T_0=2000$ K	0.6163	0.4058	0.2694	0.1873	0.1366
$T_0=2500$ K	0.6245	0.4162	0.2778	0.1928	0.1403
$T_0=3000$ K	0.6301	0.4233	0.2848	0.1977	0.1473
$T_0=3500$ K	0.6340	0.4285	0.2901	0.2018	0.1462

Table 4. Numerical values of the temperature ratio at high temperature

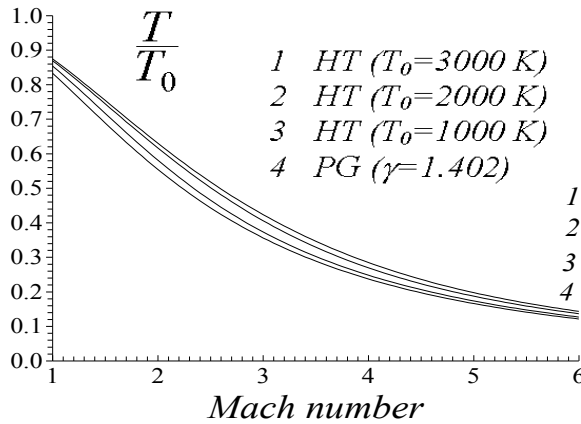


Fig. 11. Variation of  $T/T_0$  versus Mach number.

$\rho/\rho_0$	M=2.00	M=3.00	M=4.00	M=5.00	M=6.00
PG ( $\gamma=1.402$ )	0.2304	0.0765	0.0278	0.0114	0.0052
$T_0=298.15$ K	0.2304	0.0765	0.0278	0.0114	0.0052
$T_0=500$ K	0.2283	0.0758	0.0276	0.0113	0.0052
$T_0=1000$ K	0.2181	0.0696	0.0250	0.0103	0.0047
$T_0=1500$ K	0.2116	0.0636	0.0220	0.0089	0.0041
$T_0=2000$ K	0.2087	0.0601	0.0197	0.0077	0.0035
$T_0=2500$ K	0.2069	0.0581	0.0182	0.0069	0.0030
$T_0=3000$ K	0.2057	0.0569	0.0173	0.0063	0.0027
$T_0=3500$ K	0.2049	0.0560	0.0166	0.0058	0.0024

Table 5. Numerical values of the density ratio at high temperature

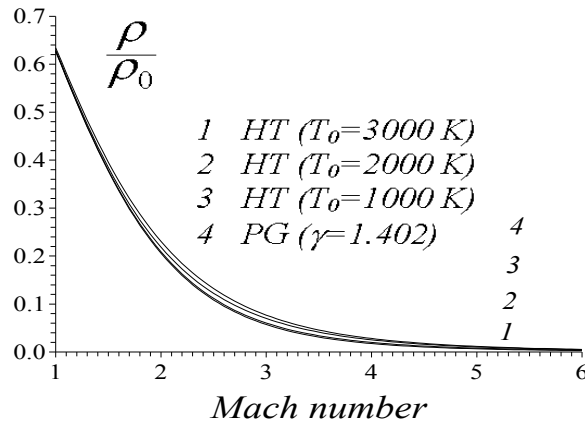


Fig. 12. Variation of  $\rho/\rho_0$  versus Mach number.

$P/P_0$	$M=2.00$	$M=3.00$	$M=4.00$	$M=5.00$	$M=6.00$
$PG (\gamma=1.402)$	0.1277	0.0272	0.0066	0.0019	0.0006
$T_0=298.15 \text{ K}$	0.1277	0.0272	0.0066	0.0019	0.0006
$T_0=500 \text{ K}$	0.1273	0.0271	0.0065	0.0018	0.0006
$T_0=1000 \text{ K}$	0.1267	0.0259	0.0062	0.0017	0.0006
$T_0=1500 \text{ K}$	0.1276	0.0248	0.0057	0.0016	0.0005
$T_0=2000 \text{ K}$	0.1286	0.0244	0.0053	0.0014	0.0004
$T_0=2500 \text{ K}$	0.1292	0.0242	0.0050	0.0013	0.0004
$T_0=3000 \text{ K}$	0.1296	0.0240	0.0049	0.0004	0.0003
$T_0=3500 \text{ K}$	0.1299	0.0240	0.0048	0.0011	0.0003

Table 6. Numerical values of the Pressure ratio at high temperature.

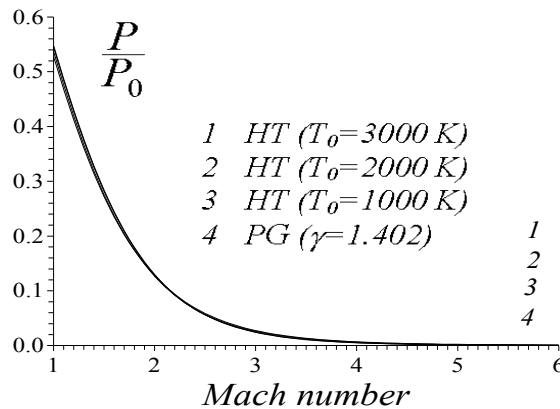


Fig. 13. Variation of  $P/P_0$  versus Mach number.



$A/A^*$	$M=2.00$	$M=3.00$	$M=4.00$	$M=5.00$	$M=6.00$
$PG (\gamma=1.402)$	1.6859	4.2200	10.6470	24.7491	52.4769
$T_0=298.15 \text{ K}$	1.6859	4.2195	10.6444	24.7401	52.4516
$T_0=500 \text{ K}$	1.6916	4.2373	10.6895	24.8447	52.6735
$T_0=1000 \text{ K}$	1.7295	4.4739	11.3996	26.5019	56.1887
$T_0=1500 \text{ K}$	1.7582	4.7822	12.6397	29.7769	63.2133
$T_0=2000 \text{ K}$	1.7711	4.9930	13.8617	33.5860	72.0795
$T_0=2500 \text{ K}$	1.7795	5.1217	14.8227	37.2104	81.2941
$T_0=3000 \text{ K}$	1.7851	5.2091	15.5040	40.3844	90.4168
$T_0=3500 \text{ K}$	1.7889	5.2727	16.0098	43.0001	98.7953

Table 7. Numerical Values of the cross section area ratio at high temperature.

Figure 14 represent the variation of the critical cross-section area section ratio versus Mach number at high temperature. For low values of Mach number and  $T_0$ , the four curves fuses and start to be differs when  $M>2.00$ . We can see that the curves 3 and 4 are almost superposed for any value of  $T_0$ . This result shows that the  $PG$  model can be used for  $T_0<1000 \text{ K}$ .

Figure 15 presents the variation of the sound velocity ratio versus Mach number at high temperature.  $T_0$  value influences on this parameter.

Figure 16 shows the variation of the thrust coefficient versus exit Mach number for various values of  $T_0$ . It can be seen the effect of  $T_0$  on this parameter. We can found that all the four curves are almost confounded when  $M_E<2.00$  approximately. After this value, the curves begin to separates progressively. The numerical values of the thrust coefficient are presented in the table 9.

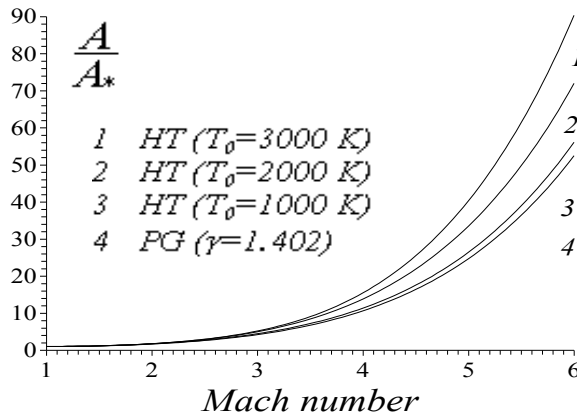


Fig. 14. Variation of the critical cross-section area ratio versus Mach number.

$a/a_0$	$M=2.00$	$M=3.00$	$M=4.00$	$M=5.00$	$M=6.00$
$PG (\gamma=1.402)$	0.7445	0.5966	0.4870	0.4074	0.3484
$T_0=298.15 \text{ K}$	0.7450	0.5970	0.4873	0.4076	0.3486
$T_0=500 \text{ K}$	0.7510	0.6019	0.4913	0.4110	0.3515
$T_0=1000 \text{ K}$	0.7739	0.6245	0.5103	0.4268	0.3651
$T_0=1500 \text{ K}$	0.7862	0.6408	0.5254	0.4398	0.3762
$T_0=2000 \text{ K}$	0.7923	0.6501	0.5354	0.4489	0.3841
$T_0=2500 \text{ K}$	0.7959	0.6556	0.5420	0.4553	0.3898
$T_0=3000 \text{ K}$	0.7985	0.6595	0.5465	0.4600	0.3942
$T_0=3500 \text{ K}$	0.7998	0.6618	0.5495	0.4632	0.3973

Table 8. Numerical values of the sound velocity ratio at high temperature.

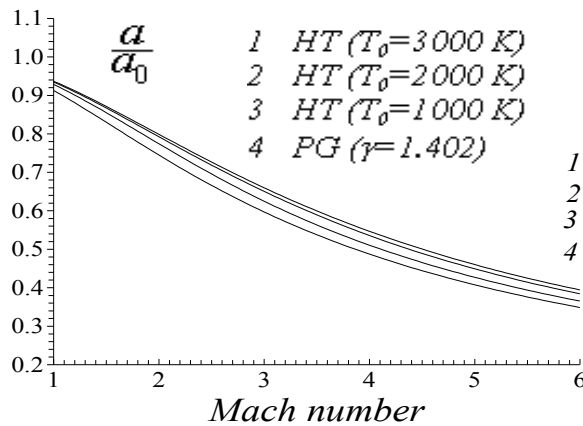


Fig. 15. Variation of the ratio of the velocity sound versus Mach number.

$C_F$	$M=2.00$	$M=3.00$	$M=4.00$	$M=5.00$	$M=6.00$
$PG (\gamma=1.402)$	1.2078	1.4519	1.5802	1.6523	1.6959
$T_0=298.15 \text{ K}$	1.2078	1.4518	1.5800	1.6521	1.6957
$T_0=500 \text{ K}$	1.2076	1.4519	1.5802	1.6523	1.6958
$T_0=1000 \text{ K}$	1.2072	1.4613	1.5919	1.6646	1.7085
$T_0=1500 \text{ K}$	1.2062	1.4748	1.6123	1.6871	1.7317
$T_0=2000 \text{ K}$	1.2048	1.4832	1.6288	1.7069	1.7527
$T_0=2500 \text{ K}$	1.2042	1.4879	1.6401	1.7221	1.7694
$T_0=3000 \text{ K}$	1.2038	1.4912	1.6479	1.7337	1.7828
$T_0=3500 \text{ K}$	1.2033	1.4936	1.6533	1.7422	1.7932

Table 9. Numerical values of the thrust coefficient at high temperature

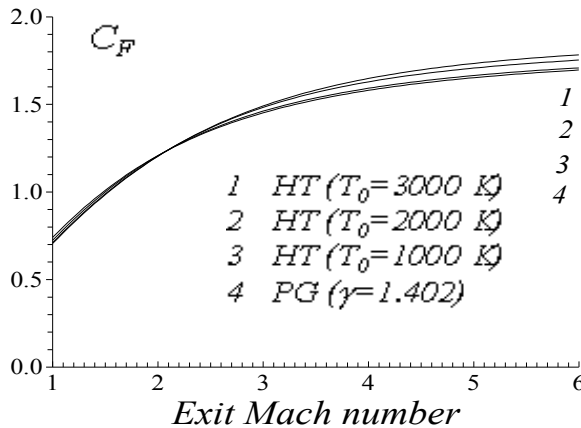


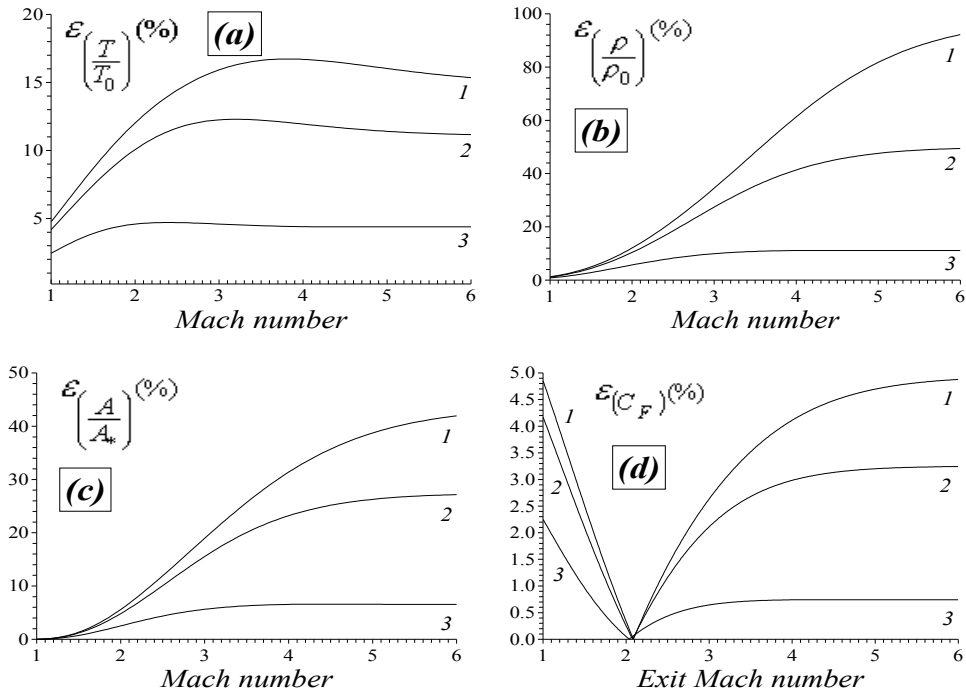
Fig. 16. Variation of  $C_F$  versus exit Mach number.

### 5.3 Results for the error given by the perfect gas model

Figure 17 presents the relative error of the thermodynamic and geometrical parameters between the  $PG$  and the  $HT$  models for several  $T_0$  values.

It can be seen that the error depends on the values of  $T_0$  and  $M$ . For example, if  $T_0=2000 \text{ K}$  and  $M=3.00$ , the use of the  $PG$  model will give a relative error equal to  $\epsilon=14.27 \%$  for the temperatures ratio,  $\epsilon=27.30 \%$  for the density ratio, error  $\epsilon=15.48 \%$  for the critical sections ratio and  $\epsilon=2.11 \%$  for the thrust coefficient. For lower values of  $M$  and  $T_0$ , the error  $\epsilon$  is weak. The curve 3 in the figure 17 is under the error 5% independently of the Mach number, which is interpreted by the use potential of the  $PG$  model when  $T_0<1000 \text{ K}$ .

We can deduce for the error given by the thrust coefficient that it is equal to  $\epsilon=0.0 \%$ , if  $M_E=2.00$  approximately independently of  $T_0$ . There is no intersection of the three curves in the same time. When  $M_E=2.00$ .



Curve 1      Error compared to HT model for ( $T_0=3000$  K)  
 Curve 2      Error compared to HT model for ( $T_0=2000$  K)  
 Curve 3      Error compared to HT model for ( $T_0=1000$  K)

(a): Temperature ratio. (b): Density ratio. (c): Critical sections ratio. (d): Thrust coefficient.

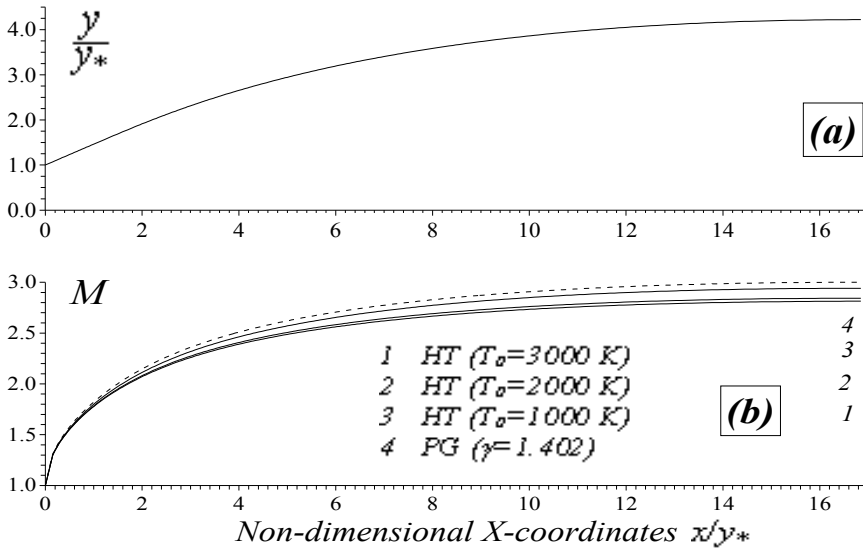
Fig. 17. Variation of the relative error given by supersonic parameters of *PG* versus Mach number.

#### 5.4 Results for the supersonic nozzle application

Figure 18 presents the variation of the Mach number through the nozzle for  $T_0=1000$  K, 2000 K and 3000 K, including the case of perfect gas presented by curve 4. The example is selected for  $M_5=3.00$  for the *PG* model. If  $T_0$  is taken into account, we will see a fall in Mach number of the dimensioned nozzle in comparison with the *PG* model. The more is the temperature  $T_0$ , the more it is this fall. Consequently, the thermodynamics parameters force to design the nozzle with different dimensions than it is predicted by use the *PG* model. It should be noticed that the difference becomes considerable if the value  $T_0$  exceeds 1000 K.

Figure 19 present the correction of the Mach number of nozzle giving exit Mach number  $M_5$ , dimensioned on the basis of the *PG* model for various values of  $T_0$ .

One can see that the curves confound until Mach number  $M_5=2.0$  for the whole range of  $T_0$ . From this value, the difference between the three curves 1, 2 and 3, start to increase. The curves 3 and 4 are almost confounded whatever the Mach number if the value of  $T_0$  is lower than 1000 K. For example, if the nozzle delivers a Mach number  $M_5=3.00$  at the exit section, on the assumption of the *PG* model, the *HT* model gives Mach number equal to  $M_5=2.93$ , 2.84 and 2.81 for  $T_0=1000$  K, 2000 K and 3000 K respectively. The numerical values of the correction of the exit Mach number of the nozzle are presented in the table 10.



(a): Shape of nozzle, dimensioned on the consideration of the PG model for  $M_s=3.00$ .

(b): Variation of the Mach number at high temperature through the nozzle.

Fig. 18. Effect of stagnation temperature on the variation of the Mach number through the nozzle.

$M_s$ (PG $\gamma=1.402$ )	1.5000	2.0000	3.0000	4.0000	5.0000	6.0000
$M_s$ ( $T_0=298.15$ K)	1.4995	1.9995	2.9995	3.9993	4.9989	5.9985
$M_s$ ( $T_0=500$ K)	1.4977	1.9959	2.9956	3.9955	4.9951	5.9947
$M_s$ ( $T_0=1000$ K)	1.4879	1.9705	2.9398	3.9237	4.9145	5.9040
$M_s$ ( $T_0=1500$ K)	1.4830	1.9534	2.8777	3.8147	4.7727	5.7411
$M_s$ ( $T_0=2000$ K)	1.4807	1.9463	2.8432	3.7293	4.6372	5.5675
$M_s$ ( $T_0=2500$ K)	1.4792	1.9417	2.8245	3.6765	4.5360	5.4209
$M_s$ ( $T_0=3000$ K)	1.4785	1.9388	2.8121	3.6454	4.4676	5.3066
$M_s$ ( $T_0=3500$ K)	1.4778	1.9368	2.8035	3.6241	4.4216	5.2237

Table 10. Correction of the exit Mach number of the nozzle.

Figure 20 presents the supersonic nozzles shapes delivering a same variation of the Mach number throughout the nozzle and consequently given the same exit Mach number  $M_s=3.00$ . The variation of the Mach number through these 4 nozzles is illustrated on curve 4 of figure 18. The three other curves 1, 2, and 3 of figure 15 are obtained with the HT model use for  $T_0=3000$  K, 2000 K and 1000 K respectively. The curve 4 of figure 20 is the same as it is in the figure 13a, and it is calculated with the PG model use. The nozzle that is calculated according to the PG model provides less cross-section area in comparison with the HT model.

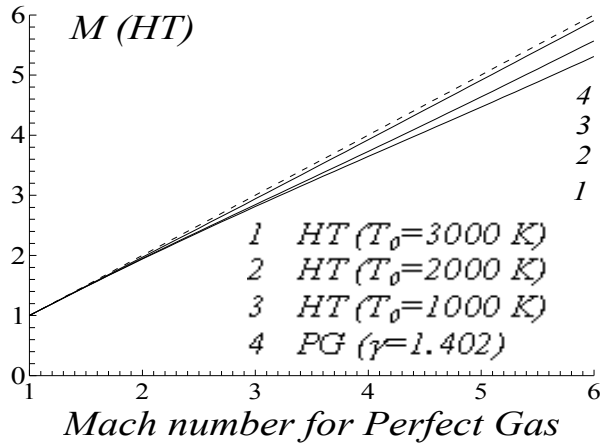


Fig. 19. Correction of the Mach number at High Temperature of a nozzle dimensioned on the perfect gas model.

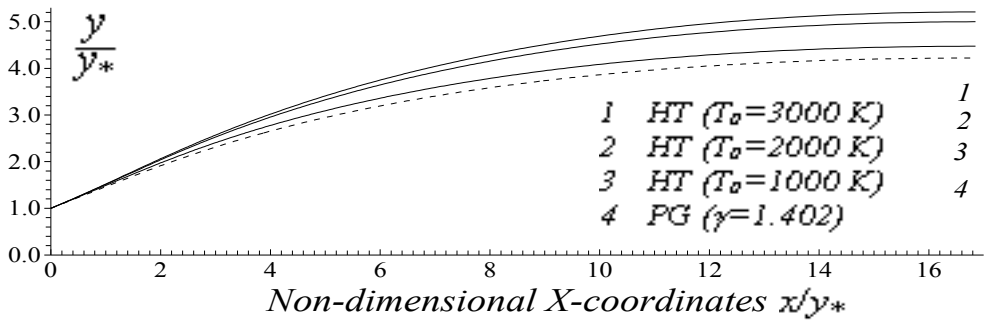


Fig. 20. Shapes of nozzles at high temperature corresponding to same Mach number variation throughout the nozzle and given  $M_S=3.00$  at the exit.

## 6. Conclusion

From this study, we can quote the following points:

If we accept an error lower than 5%, we can study a supersonic flow using a perfect gas relations, if the stagnation temperature  $T_0$  is lower than 1000 K for any value of Mach number, or when the Mach number is lower than 2.0 for any value of  $T_0$  up to approximately 3000 K.

The *PG* model is represented by an explicit and simple relations, and do not request a high time to make calculation, unlike the proposed model, which requires the resolution of a nonlinear algebraic equations, and integration of two complex analytical functions. It takes more time for calculation and for data processing.

The basic variable for our model is the temperature and for the *PG* model is the Mach number because of a nonlinear implicit equation connecting the parameters  $T$  and  $M$ .

The relations presented in this study are valid for any interpolation chosen for the function  $C_p(T)$ . The essential one is that the selected interpolation gives small error.

We can choose another substance instead of the air. The relations remain valid, except that it is necessary to have the table of variation of  $C_p$  and  $\gamma$  according to the temperature and to make a suitable interpolation.

The cross section area ratio presented by the relation (19) can be used as *a source of comparison for verification of the dimensions calculation of various supersonic nozzles*. It provides a uniform and parallel flow at the exit section by the method of characteristics and the Prandtl Meyer function (Zebbiche & Youbi, 2005a, 2005b, Zebbiche, 2007, Zebbiche, 2010a & Zebbiche, 2010b). The thermodynamic ratios can be used to determine the design parameters of the various shapes of nozzles under the basis of the *HT* model.

We can obtain the relations of a perfect gas starting from the relations of our model by annulling all constants of interpolation except the first. In this case, the *PG* model becomes a particular case of our model.

## 7. Acknowledgment

The author acknowledges Djamel, Khaoula, Abdelghani Amine, Ritadj Zebbiche and Fettoum Mebrek for granting time to prepare this manuscript.

## 8. References

- Anderson J. D. Jr. (1982), *Modern Compressible Flow. With Historical Perspective*, (2<sup>nd</sup> edition), Mc Graw-Hill Book Company, ISBN 0-07-001673-9. New York, USA.
- Anderson J. D. Jr. (1988), *Fundamentals of Aerodynamics*, (2<sup>nd</sup> edition), Mc Graw-Hill Book Company, ISBN 0-07-001656-9, New York, USA.
- Démidovitch B. et Maron I. (1987), *Éléments de calcul numérique*, Editions MIR, ISBN 978-2-7298-9461-0, Moscou, USSR.
- Fletcher C. A. J. (1988), *Computational Techniques for Fluid Dynamics: Specific Techniques for Different Flow Categories*, Vol. II, Springer Verlag, ISBN 0-387-18759-6, Berlin, Heidelberg.
- Moran M. J., (2007). *Fundamentals of Engineering Thermodynamics*, John Wiley & Sons Inc., 6<sup>th</sup> Edition, ISBN 978-8-0471787358, USA
- Oosthuisen P. H. & Carscallen W. E., (1997), *Compressible Fluid Flow*. Mc Grw-Hill, ISBN 0-07-0158752-9, New York, USA.
- Peterson C.R. & Hill P. G. (1965), *Mechanics and Thermodynamics of Propulsion*, Addition-Wesley Publishing Company Inc., ISBN 0-201-02838-7, New York, USA.
- Ralston A. & Rabinowitz P. A. (1985). *A First Course in Numerical Analysis*. (2<sup>nd</sup> Edition), McGraw-Hill Book Company, ISBN 0-07-051158-6, New York, USA.
- Ryhming I. L. (1984), *Dynamique des fluides*, Presses Polytechniques Romandes, Lausanne, ISBN 2-88074-224-2, Suisse.
- Zebbiche T. (2007). Stagnation Temperature Effect on the Prandtl Meyer Function. *AIAA Journal*, Vol. 45 N° 04, PP. 952-954, April 2007, ISSN 0001-1452, USA
- Zebbiche T. & Youbi Z. (2005a). Supersonic Flow Parameters at High Temperature. Application for Air in nozzles. *German Aerospace Congress 2005, DGLR-2005-0256, 26-29 Sep. 2005, ISBN 978-3-8322-7492-4, Friedrichshafen, Germany.*

- Zebbiche T. & Youbi Z., (2005b). Supersonic Two-Dimensional Minimum Length Nozzle Conception. Application for Air. *German Aerospace Congress 2005, DGLR-2005-0257, 26-29 Sep. 2005, ISBN 978-3-8322-7492-4, Friedrichshafen, Germany.*
- Zebbiche T. & Youbi Z. (2006), Supersonic Plug Nozzle Design at High Temperature. Application for Air, *AIAA Paper 2006-0592, 44<sup>th</sup> AIAA Aerospace Sciences Meeting and Exhibit, 9-12 Jan. 2006, ISBN 978-1-56347-893-2, Reno Nevada, Hilton, USA.*
- Zebbiche T., (2010a). Supersonic Axisymmetric Minimum Length Conception at High Temperature with Application for Air. *Journal of British Interplanetary Society (JBIS), Vol. 63, N° 04-05, PP. 171-192, May-June 2010, ISBN 0007-084X, 2010.*
- Zebbiche T., (2010b). *Tuyères Supersoniques à Haute Température.* Editions Universitaires Européennes. ISBN 978-613-1-50997-1, Dudweiler Landstrabe, Sarrebruck, Germany.
- Zucker R. D. & Bilbarz O. (2002). *Fundamentals of Gas Dynamics*, John Wiley & Sons. ISBN 0-471-05967-6, New York, USA



# Statistical Mechanics That Takes into Account Angular Momentum Conservation Law - Theory and Application

Illia Dubrovskyi  
*Institute for Metal Physics National Academy of Science  
 Ukraine*

## 1. Introduction

The fundamental problem of statistical mechanics is obtaining an ensemble average of physical quantities that are described by phase functions (classical physics) or operators (quantum physics). In classical statistical mechanics the ensemble density of distribution is defined in the phase space of the system. In quantum statistical mechanics the space of functions that describe microscopic states of the system play a role similar to the classical phase space. The probability density of the system detection in the phase space must be normalized. It depends on external parameters that determine the macroscopic state of the system.

An in-depth study of the statistical mechanics foundations was presented in the works of A.Y. Khinchin (Khinchin, 1949, 1960). For classical statistical mechanics an invariant set was introduced. It would be mapped into itself by transforming with the Hamilton equations. The phase point of the isolated system remains during the process of the motion at the invariant set at all times. If the system is in the stationary equilibrium state, this invariant set has a finite measure. The Ergodic hypothesis asserts that in this case the probability  $dP(\mathbf{R})$  to detect this system at any point  $\mathbf{R}$  of the phase space is:

$$dP(\mathbf{R}) = \frac{\varphi_{\Sigma}(\mathbf{R})d\Gamma}{\Omega_{\Sigma} \left[ (2\pi\hbar)^{3N} N! \right]} \quad (1)$$

where  $\Omega_{\Sigma}$  - the measure (phase volume) of the invariant set  $\Sigma$ ;  $\varphi_{\Sigma}(\mathbf{R})$  - the characteristic function of the invariant set, which is equal to one if the point  $\mathbf{R}$  belongs to this set, and is equal to zero in all other points of the phase space;  $d\Gamma = \prod_{i=1}^N d\mathbf{p}_i d\mathbf{r}_i$  - the phase space volume element. The number of distinguishable states in a phase space volume element  $d\Gamma$  is  $\left[ (2\pi\hbar)^{3N} N! \right]^{-1}$ . The system that will be under consideration is a collection of  $N$  structureless particles. The averaged value of a phase function  $F(\mathbf{R})$  is  $\bar{F} = \int_{\Gamma} F(\mathbf{R})dP(\mathbf{R})$ . Here the integral goes over all phase space  $\Gamma$ . This is microcanonical distribution. A characteristic function often would be presented as  $\delta[f(\mathbf{R}) - z]$ , where  $f(\mathbf{R})$  is a phase function and  $z$  is it's fixed value.

A hypersurface in a hyperspace is a set with zero measure. Therefore the invariant set is determined as a thin layer that nearly envelops the hypersurface in the phase space. The

determining equations of this hypersurface are the equalities that fix the values of controllable motion integrals. A controllable motion integral is a phase function, the value of which does not vary with the motion of the system and can be measured. An isolated system universally has the Hamiltonian that does not depend on the time explicitly, and is the controllable motion integral. A fixed value of the Hamiltonian is the energy of the system. The kinetic energy of majority of systems is a positive definite quadric form of all momenta. It determines a closed hypersurface in the subspace of momenta of the phase space. If motions of all particles are finite, the hypersurface of the fixed energy is closed and the layer that envelops it has the finite measure. Then this hypersurface can determine the invariant set of the system. A finiteness of motions of particles as a rule is provided by enclosing the system in an envelope that reflects particles without changing their energy, if the system is considered as isolated. It is common in statistical mechanics to consider the layer enveloping the energy hypersurface as the invariant set. But A.Y. Khinchin (Khinchin, 1949) shows that other controllable integrals of the system, if they exist, must be taken into account. In the general case an isolated system can have another two vector controllable integrals. That is the total momentum of the system, and the total angular momentum relative to the system's mass centre. The total momentum is a sum of all momenta of particles. If the volume of the system is bounded by an external field or an envelope, the total momentum does not conserve. In the absence of external fields the total momentum conservation cannot make particle motions finite. Therefore the total momentum cannot be a controllable motion integral that determines the invariant set.

The angular momentum is another case. A vector of angular momentum relative to the mass center always is conserved in an isolated system. If this vector is nonzero, a condition should exist that provides a limitation of a gas expansion area. For example, nebulas do not collapse because they rotate, and do not scatter because of the gravitation. In the system of charged particles in a uniform magnetic field the conservation of the angular momentum provides a limitation of a gas expansion area (confinement of plasma). If a gas system is enclosed into envelope, and total system has nonzero angular momentum, the vector of the angular momentum should be conserved. However an envelope can have the non-ideal form and surface. That is the cause of the failure to consider the angular momentum of the gas as a controllable motion integral (Fowler, & Guggenheim, 1939). But if the cylindrical envelope rotates and the gas rotates with the same angular velocity deviations of the angular momentum of the gas from the fixed value as the result of reflections of particles from the envelope should be small and symmetric with respect to a sign. These fluctuations are akin to energy fluctuations for a system that is in equilibrium with a thermostat. Therefore the angular momentum conservation in specific cases can determine the invariant set and the thermodynamical natures of the system together with the energy conservation. Taking into account all controllable motion integrals is the necessary condition of the validity of the Ergodic hypothesis (Khinchin, 1949).

There is a contradiction in physics at the present time. Firstly, it has been proven that in the equilibrium state a system spin can exist only if the system is rigid and can rotate as a whole (Landau, & Lifshitz, E.M., 1980a). Therefore a gas, which supposed not be able to rotate as a whole, cannot have any angular momentum and spin. Based on this reasoning R.P. Feynman proves that an electron gas cannot have diamagnetism (the Bohr - van Leeuwen theorem) (Feynman, Leighton, & Sands, 1964). On the other hand, it is well known that density of a gas in a rotating centrifuge is non-uniform. This effect is used for the separation

of isotopes (Cohen, 1951). The experiment by R. Tolman, described in the book (Pohl, 1960), is a proof of the existence of the electron gas angular momentum. In this experiment a coil was rotated and then sharply stopped. An electrical potential was observed that generated a moment of force, which decreased to zero the angular momentum of electron gas.

The contradiction described above requires creation of statistical mechanics for non-rigid systems taking into account the nonzero angular momentum conservation. This statistical mechanics differs from common one in many respects. If the angular momentum relative to the axis that passes through the mass centre conserves, the system is spatially inhomogeneous. This means that passage to the thermodynamical limit makes no sense, a spatial part of the system is not a subsystem that similar to the total system, specific quantities such as densities or susceptibilities have no physical meaning.

The microcanonical distribution is seldom used directly when the computations and the justifications of thermodynamics are done. The more usable Gibbs distribution can be deduced from microcanonical one (Krutkov, 1933; Zubarev, 1974). The Gibbs assembly describes a system that is in equilibrium with environment. These systems do not have motion integrals because they are non-isolated. All elements of the Gibbs assembly must have equal values of parameters that are determined by the equilibrium conditions. In usual thermodynamics this parameters are the temperature and the chemical potential. The physical interpretation of these parameters is getting by statistical mechanics. A rotating system can be in equilibrium only with rotating environment. The equilibrium condition in this case is apparent. That is equality of the both angular velocities of the system and of the environment. The Gibbs assembly density of distribution and thermodynamical functions in the case of a rotating classical system will be obtained in the second section of this work. It was done (Landau, & Lifshitz E.M., 1980a) but an object, to which this distribution is applied, is incomprehensible, because an angular velocity of an equilibrium gas has not been determined.

In quantum statistical mechanics the invariant set is the linear manifold of the microscopic states of the system in which the commutative operators that correspond to the controllable motion integrals have fixed eigenvalues. The phase volume of system in this case is the dimension of the manifold, if this dimension is limited. It directly determines the number of distinguishable microstates of the system that are accessible and equiprobable. The role of the angular momentum conservation in quantum statistical mechanics is similar to one in classical statistical mechanics. The method of computing this phase volume will be also proposed in the second section of this work. The Gibbs assembly density of distribution and thermodynamical functions in the case of a rotating quantum system also will be obtained.

In the third section of this work statistical mechanics of an electron gas in a magnetic field is considered. This question was investigated by many during the last century. Many hundreds experimental and theoretical works were summarized in the treatises (Lifshits, I.M. et al., 1973; Shoenberg, 1984). However, together with successful theoretical explanations of many experimental effects some paradoxes and discrepancies with observed facts remain unaccounted.

“Finally, it is shown that the presence of free electrons, contrary to the generally adopted opinion, will not give rise to any magnetic properties of the metals”. This sentence ends a short report on the presentation “Electron Theory of Metals” by N. Bohr, given at the meeting of the Philosophical Society at Cambridge. It was well-known that a charged particle in a uniform magnetic field moves in a circular orbit with fixed centre in such a way

that the time average value of the magnetic moment, generated by this motion, is directed opposite to the magnetic field and equal to the derivative of the kinetic energy with respect to the magnetic field. N. Bohr computed the magnetic moment of an electron gas by statistical mechanics with the density of distribution that is determined only by a Hamiltonian. Zero result of this theory (Bohr – van Leeuwen theorem) is the first paradox. Many attempts of derivation and explanation of this were summarized in the treatise (van Vleck, 1965). The most widespread explanation was that the magnetization generated by the electrons moving far from the bound is cancelled by the near-boundary electrons that reflect from the bound. But this explanation is not correct because, when formulae are derived in statistical mechanics, any peculiarities of the near-boundary states shall not be taken into account. Another paradox of the common theory went unnoticed. It is well known that a uniform magnetic field restricts an expanse of a charged particles gas in the plane perpendicular to the field. But from common statistical mechanics it follows that the gas uniformly fills all of the bounded area. The diamagnetism of some metals also was left non-explained.

L.D. Landau (Landau, 1930) explained the diamagnetism of metals as a quantum effect. He solved the quantum problem of an electron in a uniform magnetic field. The cross-section of the envelope perpendicular to the magnetic field is a rectangle with the sides  $2L_x$  and  $2L_y$ . The solutions are determined by three motion integrals. The first is energy that takes the values  $\varepsilon_{np} = \hbar\omega_c(n+1/2) + p^2/2m$ , where  $\omega_c$  is the cyclotron frequency  $\omega_c = eH/m$ , ( $-e$ ) is the charge and  $m$  is the mass of an electron,  $H$  is the magnetic induction,  $n$  is a positive integer or zero. The second is the  $z$ -component of the momentum  $p$ . The third motion integral is the Cartesian coordinate of the centre of the classical orbit. It takes the values  $y_j = (\pi\hbar/eHL_x)j$ , where  $j=0, \pm 1, \pm 2, \dots \pm eHL_xL_y/\pi\hbar$ . The thermodynamical potential with this energy spectrum, when the spin degeneracy is taken into account, is:

$$\Omega = -k_B T \int dp \frac{L_z}{2\pi\hbar} \sum_{n=0}^{\infty} \frac{eHS}{\pi\hbar} \ln \left[ 1 + \exp\left(\frac{\mu - \varepsilon_{np}}{k_B T}\right) \right] \quad (2)$$

Here  $k_B$  is the Boltzmann constant,  $T$  is the temperature,  $S = 4L_xL_y$ ,  $\mu$  is the chemical potential. If in this formula the summation over  $n$  is changed to the integration, the result  $\Omega_0$  does not depend on  $H$ , and the magnetic moment  $\mathfrak{M} = -(\partial\Omega_0/\partial H) = 0$ . That agrees to the classical and paradoxical Bohr – van Leeuwen theorem. L.D. Landau uses the Euler – Maclaurin summation formula in the first order and obtains the amendment that depends on the magnetic field. In the limit  $T \rightarrow 0$  the thermodynamical potential has appearance (Abrikosov, 1972):

$$\Omega = \Omega_0 + V \frac{e^2 H^2 p_F}{24\pi^2 \hbar m}, \quad (3)$$

where  $p_F = \sqrt{2m\zeta} = (3\pi^2)^{1/3} (N/V)^{1/3} \hbar$ , and  $\zeta$  is the Fermi energy,  $V = 4L_xL_yL_z$  is the volume. This result cannot be correct because the magnetic moment does not depend on the Plank constant  $\hbar$  and thus it cannot be a quantum effect. This problem is simpler for a two-dimensional gas. In this case the common formula of the thermodynamical potential has the form:

$$\Omega_{2D} = -k_B T \sum_{n=0}^{\infty} \frac{eHS}{\pi \hbar} \ln \left[ 1 + \exp \left( \frac{\mu - \varepsilon_n}{k_B T} \right) \right]. \quad (4)$$

When  $T \rightarrow 0$ , this sum can be computed without to change the summation to the integration.

$$\Omega_{2D} = -\frac{\zeta^2 m S}{2\pi \hbar^2} + \frac{e^2 H^2 S}{8\pi m}. \quad (5)$$

It is suggested that the Fermi level is filled. In this case the magnetic moment does not depend not only on the Plank constant, but also on the number of the electrons. Therefore the fundamental formula for the thermodynamical potential is incorrect.

In the third section of this work the diamagnetism of an electron gas is investigated with taking into account the conservation of zero value of the total angular momentum in classical and quantum statistical mechanics. The paradoxes described above are eliminated; however many other theories should be reconsidered.

## 2. Statistical mechanics of rotating gas

For the computation of average values of macroscopic quantities it is necessary to derive a formula of the phase volume as a function of macroscopic parameters. This function is called "structural function" by Khinchin (Khinchin, 1949) and "number of accessible states (or complexions)" by Fowler (Fowler, & Guggenheim, 1939). It determines the normalizing factor in the probability density of the microcanonical distribution (1). In usual theory this function is essential to the derivation of formulae that connect statistical physics with thermodynamics.

The system that will be considered is a collection of  $N$  structureless particles. If forces of interaction between particles manifest themselves only at distances considerably smaller than the average distance between particles, the interaction energy of particles is essential only in a small fraction of the phase volume. Therefore, the interaction of particles can be neglected or be taken into account as a perturbation in calculating the phase volume and the average values (Uhlenbeck & Ford, 1963). Otherwise, if particles interact by a long-range force, this interaction needs to be considered using a mean field method. This is a model of an ideal gas under an external field. Meanwhile, this external field can be also a periodical crystal field. In the commonly considered cases the Hamiltonian and other phase functions of the system can be presented as the sum of identical terms, each of which depends on the coordinates and momenta of a single particle. Such phase function is said to be a summatory function.

For integration characteristic functions over a phase space the method by Krutkov will be used. The main idea of this method is to make the Laplace transformation of the  $\delta$ -function with respect the value of the summatory function. Then the product of the  $N$  like exponents from the terms of the summatory function would be integrated over variables of the phase space. The inverse transformation would be made by the saddle-point method with using the large parameter  $N$ .

Let us write several equalities with a characteristic function. If the system can be divided into two independent subsystems described by non-overlapping groups of phase variables, so that  $\Sigma = \Sigma_1 \otimes \Sigma_2$ ,  $\mathbf{R} = \mathbf{R}_1 + \mathbf{R}_2$ , and the determining functions possess the values independently, then

$$\Omega_{\Sigma} = \Omega_{\Sigma_1} \cdot \Omega_{\Sigma_2}, \quad \varphi_{\Sigma}(\mathbf{R}) = \varphi_{\Sigma_1}(\mathbf{R}_1) \cdot \varphi_{\Sigma_2}(\mathbf{R}_2),$$

$$dP(\mathbf{R}) = dP_1(\mathbf{R}_1) \cdot dP_2(\mathbf{R}_2) = \left[ \varphi_{\Sigma_1}(\mathbf{R}_1) / \Omega_1 (2\pi\hbar)^{3N_1} \right] \cdot \left[ \varphi_{\Sigma_2}(\mathbf{R}_2) / \Omega_2 (2\pi\hbar)^{3N_2} \right] d\Gamma_1 d\Gamma_2 \quad (6)$$

Here the multiplier  $(N!)^{-1}$  is not taken into account because it cannot be introduced logically in classical statistical mechanics. Considering the fact that a density of distribution for a system is equal to the product of densities of distribution for subsystems, a conclusion is drawn in the treatise (Landau, & Lifshitz, E.M., 1980a) that the logarithm of the density of distribution should be an additive motion integral and, hence, it should be a linear combination of the additive controllable motion integrals, such as the energy, the momentum and the angular momentum. However, as it follows from the formula (6) this is incorrect for the microcanonical distribution, since the logarithm of the characteristic function is meaningless. A system in a thermostat does not have any motion integrals. If the invariant set is determined by some conservation laws, its characteristic function is

$$\varphi_{\Sigma}(\mathbf{R}) = \prod_i \varphi_i(\mathbf{R}), \quad (7)$$

where  $\varphi_i(\mathbf{R})$  is the characteristic function that is determined by the conservation law number  $i$ . Let us denote a set, at which the phase function  $A(\mathbf{R})$  is equal to  $a$ , by  $\Sigma_A^a$ , its characteristic function by  $\varphi_A^a(\mathbf{R})$ , and its measure by  $\Omega_A^a$ . It is supposed that this measure is limited and is not equal to zero. Then

$$\int_{\Gamma} \varphi_A^a(\mathbf{R}) d\Gamma = \Omega_A^a, \quad \int_{\Gamma} \varphi_A^a(\mathbf{R}) f(\mathbf{R}, A(\mathbf{R})) d\Gamma = \int_{\Gamma} \varphi_A^a(\mathbf{R}) f(\mathbf{R}, a) d\Gamma, \quad (8)$$

$$\int_A \int_{\Gamma} \varphi_A^a(\mathbf{R}) f(a, \mathbf{R}) d\Gamma da = \int_A \int_{\Gamma} \varphi_A^a(\mathbf{R}) f(A(\mathbf{R}), \mathbf{R}) d\Gamma da = \int_{\Gamma} f(A(\mathbf{R}), \mathbf{R}) d\Gamma$$

Here  $A$  is the range of values of the function  $A(\mathbf{R})$ . The prevalent formula  $\varphi_A^a = \Delta a \delta(A - a)$  (Landau, & Lifshitz, E.M., 1980a; Uhlenbeck, & Ford, 1963) satisfies to the equalities (7) and (8) but does not satisfy to the equality (6). It is more frequently considered the separation on subsystems that conserves the total value of the function. If  $A(\mathbf{R}) = A_1(\mathbf{R}_1) + A_2(\mathbf{R}_2) = a$ , then  $\varphi_A^a = \int_A \varphi_{A_1}^{a-x} \varphi_{A_2}^x dx$ , and the prevalent formula is correct.

## 2.1 Classical statistical thermodynamics of rotating gas

The formula for average values, when the conservation of the angular momentum is taken into account, has the form:

$$\bar{F} = \left[ (2\pi\hbar)^{3N} N! \right]^{-1} \frac{\Delta E \Delta \bar{L}}{\Omega} \int_{\Gamma} F(\mathbf{R}) \delta(H(\mathbf{p}, \mathbf{r}) - E) \delta(L(\mathbf{p}, \mathbf{r}) - \bar{L}) d\Gamma \quad (9)$$

$$\Omega = \Delta E \Delta \bar{L} \int_{\Gamma} \delta(H(\mathbf{p}, \mathbf{r}) - E) \delta(L(\mathbf{p}, \mathbf{r}) - \bar{L}) d\Gamma$$

In these formulae the axis  $\mathbf{Z}$  is parallel to the angular momentum  $\bar{\mathbf{L}}$ . The angular momentum of the gas is  $L(\mathbf{p}, \mathbf{r}) = \sum_{i=1}^N l(\mathbf{p}_i, \mathbf{r}_i)$ . The effective Hamiltonian of the gas with the fixed angular momentum in the cylindrical coordinates can be obtained from the usual formula by substitution  $l_1 = \bar{L} - \sum_{i=2}^N l_i$  with reduction of the quadratic form to the standard appearance. It is:

$$\mathcal{H}(\mathbf{p}, \mathbf{r}) = \frac{1}{2m} \sum_{i=2}^N \left( p_{z,i}^2 + p_{r,i}^2 + \frac{l_i^2}{r_i^2} + U(\sqrt{r_i^2 + z_i^2}) \right) + \frac{\bar{L}^2}{2m \sum_{i=1}^N r_i^2}, \quad (10)$$

where  $U(\sqrt{r^2 + z^2})$  is potential energy that confines the particles in bounded volume, and the second term of the Hamiltonian is the potential energy of the centrifugal force that leads to a collapse of rotating nebula into disk. This Hamiltonian is not a summatory function. Therefore the Krutkov's method cannot be used for the subsequent computations.

Let us consider equilibrium of a gas with a rotating rigid body. The rigid body can be determined as the body in which the rotatory degree of freedom can not transfer energy and angular momentum to the internal degrees of freedom. This possibility arises when this body is a cylindrical rotating envelope with non-ideal surface filled by a gas. The state of the gas is characterized by two parameters: the temperature  $T$  and the angular velocity  $\omega$ . For introducing the statistical parameter that corresponds to the thermodynamical temperature it is necessary to deduce the canonical Gibbs distribution for a system that is in equilibrium with a thermostat. That can be done, for example, by a method developed by Krutkov (Krutkov, 1933; Zubarev, 1974). The conditions of the equilibrium between the rotating envelope and the gas are apparent. Those are the equalities of the temperature and the angular velocity. Let us determine the angular velocity of a gas. An angular velocity of a particle is a stochastic quantity with an average value  $\bar{\omega}$ . The sum  $\sum_{i=1}^N \omega_i$  is the Gaussian random variable with the average value  $N\bar{\omega} = \omega_g$ . That is the angular velocity of a gas. The conditions of the equilibrium between the rotating envelope and the gas are the equality of the temperatures and

$$\omega_g = \omega. \quad (11)$$

The total system can be considered as motionless if it will be described in the rotating reference frame, when the right part of the equality (11) is zero. The hollow cylinder is the envelope, the thermostat, and it keeps the gas spin. It should be named "termospinstat". The potential energy of the centrifugal force  $U_{cf} = (-m\omega^2 r^2/2)$  is added in the Hamiltonian of the gas particle in the rotating reference frame (Landau, & Lifshitz, E.M., 1980a). The average angular momentum of the gas  $m\omega \overline{\sum_{i=1}^N r_i^2} = \omega N m \bar{r}^2 = \bar{L}$  depends on the angular velocity nonlinearly because the gas moment of inertia  $\bar{I}(\omega) = N m \bar{r}^2$  is the function of the angular velocity. This function can be obtained from the Gibbs distribution for a gas in the system of reference that rotates with the angular velocity  $\omega$ :

$$\begin{aligned} dP_G(\mathbf{R}) &= \frac{d\Gamma}{(2\pi\hbar)^{3N} N! Z_N} \exp \left[ \frac{1}{k_B T} \left( -\frac{1}{2m} \sum_{i=1}^N \left( p_{z,i}^2 + p_{r,i}^2 + \frac{l_i^2}{r_i^2} \right) + \frac{\omega^2 m}{2} \sum_{i=1}^N r_i^2 - U_0 \right) \right] \\ Z_N &= \left\{ (2\pi\hbar)^{3N} (N!) \right\}^{-1} \int_{\Gamma} \exp \left[ \frac{1}{k_B T} \left( -\frac{1}{2m} \sum_{i=1}^N \left( p_{z,i}^2 + p_{r,i}^2 + \frac{l_i^2}{r_i^2} \right) + \frac{\omega^2 m}{2} \sum_{i=1}^N r_i^2 - U_0 \right) \right] d\Gamma = \quad (12) \\ \left\{ (2\pi\hbar)^{3N} (N!) \right\}^{-1} &\left\{ \frac{h_z}{\omega^2} \sqrt{(2\pi k_B T)^5 m} \left[ 1 - \exp \left( -\frac{\omega^2 m \mathcal{R}^2}{2k_B T} \right) \right] \right\}^N = \exp \left[ -\frac{1}{k_B T} F \left( T, \frac{\omega^2}{2} \right) \right], \end{aligned}$$

where  $h_z$  and  $\mathcal{R}$  are the dimensions of the envelope, and  $U_0 = Nm\mathcal{R}^2\omega^2/2$  is an appending constant that does the energy positive. Going to the thermodynamics (it rather can be entitled by “thermospindynamics”) it is naturally to consider  $\sigma = \omega^2/2$  as an external parameter and the moment of inertia  $I = m\sum_{i=1}^N r_i^2$  as a characteristic of the gas. Then

$$\bar{I} = Nm\bar{r}^2 = -\partial F/\partial\sigma. \quad (13)$$

The formulae of the isotopes separation (Cohen, 1951) can be obtained from the distribution (12). If  $m(\omega\mathcal{R})^2 \ll k_B T$  the formula (12) can be presented as:

$$\begin{aligned} F &\approx F_0 + \frac{1}{2}N\sigma m\mathcal{R}^2 - \frac{N}{24} \frac{\sigma^2 m^2 \mathcal{R}^4}{k_B T} = F_0 + \bar{I}_0\sigma - \frac{N}{24} \frac{\sigma^2 m^2 \mathcal{R}^4}{k_B T} \\ F_0 &= -Nk_B T \ln \left[ \frac{e\pi h_z \mathcal{R}^2}{N} \left( \frac{mk_B T}{2\pi\hbar^2} \right)^{3/2} \right], \quad \bar{I}_0 = \frac{Nm\mathcal{R}^2}{2}, \end{aligned} \quad (14)$$

where  $F_0$  - is the free energy of the ideal gas that does not rotate. Hence it follows that the parameters  $h_z$ ,  $\mathcal{R}^2$ , and correspondingly  $P_z$ ,  $P_s$  should be introduced instead of the volume  $V$  and the pressure  $P$ . Other thermodynamical equations are changed also. The parameter of expansion in the formula (14) can be of the order of unity when  $m = 10^{-25}$  kg,  $\mathcal{R} = 1$  m,  $T = 100$  K,  $\omega = 100$  s<sup>-1</sup>.

## 2.2 Quantum statistical thermodynamics of rotating gas

The characteristic function of the invariant set that takes into account conservation of the angular momentum in quantum statistical mechanics can be presented as a set of diagonal elements of the operator:

$$\hat{\varphi}(E, \bar{L}) = (2\pi)^{-2} \int_0^{2\pi} \exp\left[(i\tau + \vartheta)(\hat{\mathcal{H}} - E)\right] d\tau \int_0^{2\pi} \exp\left[(i\alpha + \eta)(\hat{L} - \bar{L})\right] d\alpha \quad (15)$$

in the space of microstates of the system. Here  $\hat{\mathcal{H}} = \sum_{i=1}^N \hat{h}_i$  is Hamiltonian of gas;  $\hat{h}\hat{L} = \hbar \sum_{i=1}^N \hat{l}_i$  is the operator of the total angular momentum of gas;  $E$ ,  $\hbar\bar{L}$  are values of these quantities for the considered macroscopic state;  $\vartheta$ ,  $\eta$  are real numbers which will be defined below. As usually, let us assume that energies of one-particle states, and, hence, both eigenvalues of the operator  $\hat{\mathcal{H}}$  and gas energy  $E$ , are expressed by the dimensionless positive integers. This formula would be generalized by the transition to representation of secondary quantization. In this representation function of one-particle states are eigenfunctions of the one-particle Hamiltonian  $\hat{h}_i$  and angular momentum. These functions are numbered using index  $\Psi$  which consists of a pair of quantum numbers  $(t, l)$ , where  $\varepsilon(t)$  is energy and  $\hbar l(\Psi)$  is the angular momentum at the state  $\Psi$ . Let us suppose that only two quantum numbers determine the state. Both the production and annihilation operators are determined by the kind of statistics. The operator (15) should be replenished



by multiplier that describes the conservation of the particle number. Then the measure of the invariant set for  $N$  -particle system is:

$$C(N, E, \bar{L}; \chi, \vartheta, \eta) = (2\pi)^{-3} \int_0^{2\pi} \int_0^{2\pi} \int_0^{2\pi} \exp[-N(i\phi + \chi) - E(i\tau + \vartheta) - \bar{L}(i\alpha + \eta)] \times \\ \times \text{Sp} \left\{ \exp \left[ \sum_{\Psi} \left[ ((i\phi + \chi) + (i\tau + \vartheta)) \varepsilon(\Psi) + (i\alpha + \eta) l_z(\Psi) \right] \begin{Bmatrix} a_{\Psi}^{\dagger} a_{\Psi} \\ b_{\Psi}^{\dagger} b_{\Psi} \end{Bmatrix} \right] \right\} d\phi d\tau d\alpha = \quad (16) \\ (2\pi)^{-3} \oint \oint \oint x^{-N-1} y^{-E-1} z^{-\bar{L}-1} \prod_{\Psi} (1 \pm xy^{\varepsilon(\Psi)} z^{l_z(\Psi)})^{\pm 1} dx dy dz .$$

Here the following variables are entered:

$$x = \exp(i\phi + \chi), \quad y = \exp(i\tau + \vartheta), \quad z = \exp(i\alpha + \eta). \quad (17)$$

Thus, integrals are rearranged into integrals along contours which enclose the origin of coordinates. If  $\bar{L} \neq 0$  then the axis  $\mathbf{Z}$  is parallel to the angular momentum  $\bar{\mathbf{L}}$  and  $\bar{L} > 0, \eta > 0$ . The lower operator or sign should be taken for the Bose statistics, and upper ones should be taken for the Fermi statistics. The expression of the measure of the invariant set in the case of quantum statistics is obtained from the apparent formula of the characteristic function (15). This expression is similar to the initial one in the Darwin - Fowler method (Fowler, & Guggenheim, 1939), where it was substantiated as a mathematical device. These contour integrals can be computed by the saddle-point method, when  $\bar{L} \neq 0$ . The saddle-points determine the values of  $\chi, \vartheta, \eta$ . It can be shown that taking into account the conservation of the angular momentum does not change statistic mechanics of this model of an ideal gas when  $\bar{L} = 0$ .

Let us describe a quantum gas in a termostpinstat with the temperature  $T$  and the angular velocity  $\omega$ . The potential of the centrifugal force would be introduced in the Hamiltonian only in classical statistical mechanics (Landau, & Lifshitz, E.M., 1980a). Other method should be used for quantum theory. Wave functions in a system of reference that rotates with the angular velocity  $\omega$  depend on the time, and should be determined from the Schrödinger equation:

$$i\hbar \frac{\partial \Psi}{\partial t} = \frac{1}{2m} \left( \hat{p}_z^2 + \hat{p}_r^2 + \frac{\hbar^2 \hat{l}^2}{r^2} \right) \Psi . \quad (18)$$

Dependence of the wave function on the time should be  $\psi = \exp\{i[z p_z / \hbar + l(\varphi - \omega t)]\}$ . Then  $\psi = \exp\{i[p_z / \hbar + l(\varphi - \omega t)]\} J_{|l|} \left( r \sqrt{2m(\varepsilon - \hbar\omega l) - p_z^2 / \hbar} \right)$  and  $J_{|l|}(x)$  is the Bessel function. The boundary condition  $\psi(\mathcal{R}) = 0$  determines the energy spectrum:

$$\varepsilon_{\nu} = \frac{p_z^2}{2m} + \frac{\hbar^2}{2m\mathcal{R}^2} j^2(|l|, \nu) - \hbar\omega l , \quad (19)$$

where  $j(|l|, \nu)$  is the null of the Bessel function  $J_{|l|}(x)$  with number in the order of increasing  $\nu$ . When  $\nu > 1$ ,  $j(|l|, \nu) \approx \pi(\nu + |l|/2)$  and  $j(|l|, 1) \approx |l|$ . This spectrum is quasicontinuous

because a distance between nearest-neighbor levels is proportional to  $\mathcal{R}^{-2}$ . The lowest level has value  $(-m\omega^2\mathcal{R}^2/2)$  when  $l = (m\mathcal{R}^3\omega/\hbar)$ . Then the reference point of energy should be altered by this value. It conforms to the appearance of the centrifugal force potential in the classical system. Energies of states with  $l > 0$  are lower than the ones of states with equal  $|l|$  and  $l < 0$ . Then in the gas the part of particles with  $l > 0$  should be more than half, and as result a circular current of the probability density should exist. This describes rotating of the system. With increasing argument modulus of extremes of the Bessel functions decrease. If the values of energy  $\varepsilon$  and positive angular momentum are fixed the value of the null number  $\nu$  in the rotating system should be lower than this value in the motionless system. Therefore, the gas density increases with distance from the axis in rotating system.

Let us compute the thermodynamical potential  $\Omega = \pm k_B T \sum_{\nu l} \ln \left\{ 1 \mp \exp \left[ (\mu - \varepsilon_{\nu l}) / k_B T \right] \right\}$  of ideal rotating gas when  $\exp(\mu/k_B T) \ll 1$ . Nulls of Bessel functions would be approximated by formula  $j(|l|, \nu) \approx \pi(\nu + |l|/2)$ , but when  $\nu = 1$  then  $j(|l|, 1) \approx |l|$ . The computation is performed by passing from summation over  $\nu$  and  $l$  to integration over  $j$  and  $l$ . This approximation for non-rotating gas leads to the result that differs from the common result by the multiplier  $(4/\pi)$ . The result of computation for the rotating gas is:

$$\Omega = -V \frac{4(k_B T)^{5/2} m^{3/2}}{\pi(2\pi)^{3/2} \hbar^3} \exp\left(\frac{\mu}{k_B T}\right) \left\{ 1 - \frac{\omega^2 \mathcal{R}^2}{3} \left(\frac{m}{2k_B T}\right) + \frac{\omega^4 \mathcal{R}^4}{10} \left(\frac{m}{2k_B T}\right)^2 \right\}. \quad (20)$$

If this result is compared with that of Eq. (14), it can be shown that amendments differ only by coefficients.

### 3. Statistical mechanics of electron gas in magnetic field

The review of the current status of this theory is in the paper (Vagner et al., 2006). There are some inaccuracies in this problem consideration besides disregard of the angular momentum conservation. To clarify the problem, in the first subsection we consider formulations of the one-particle problem in classical and quantum mechanics and its simplest application to the statistical mechanics. For simplicity, we will restrict ourselves to the case of a two-dimensional gas on a plane perpendicular to the uniform magnetic field  $\mathbf{H} = (0, 0, H)$ . As will be shown the magnetization of electron gas is nonuniform. We will suppose that the magnetization is small as against the uniform field, and will not regard effect of it. Then magnetic induction  $H$  is proportional to the external magnetic field strength by the coefficient  $\mu_0$ . Where it is needed, we imply the plane to be of finite "thickness"  $\delta z$ , and, for example, the equations of electrodynamics are written for three-dimensional space.

#### 3.1 Two-dimensional electron ideal gas in uniform magnetic field

This problem traditionally is considered in quasiclassical theory (Lifshitz, I.M. et al., 1973; Shoenberg, 1984). Some corrections will be inserted in this consideration in the section 3.1.1. In this section classical statistic mechanics of ideal gas will be discussed. In the next section the new correction will be obtained from the consistent quantum theory.

### 3.1.1 Classical statistical mechanics of ideal gas in magnetic field

The Hamiltonian of an electron in a magnetic field has the form:

$$\hbar = (1/2m)(\mathbf{p} + e\mathbf{A})^2 \quad (21)$$

where  $m$  and  $(-e)$  are the mass and the charge of an electron,  $\mathbf{p}$  is momentum, and  $\mathbf{A}$  is vector potential of the magnetic field:

$$\mathbf{A} = \frac{1}{2}[\mathbf{H}\mathbf{r}] = \frac{1}{2}(-yH, xH, 0). \quad (22)$$

This Hamiltonian does not have the translation symmetry. This symmetry, seemingly, should be, if the magnetic field is uniform at an unlimited plane. But a uniform magnetic field at unlimited plane is impossible because an electrical current that generates it according to Maxwell equation should envelope a part of this plane. It is asserted (Landau, & Lifshitz, E.M. 1980b; Vagner et al., 2006) that the Hamiltonian (21) with the vector potential (22) would be converted by gauge transformation  $\mathbf{A} \rightarrow \mathbf{A} + \nabla f(x, y, z)$ . If the function  $f = -eHxy/2$ , then Hamiltonian will be in the Landau form :

$$\hbar_L = (1/2m)[(p_x - eHy)^2 + p_y^2], \quad (23)$$

and will have the translation symmetry in the direction of the axis  $\mathbf{X}$  in return for axial symmetry. That is strange assertion because the symmetry is the physical property of the system rather than of a method of it description. In fact the transformation to the Hamiltonian (23) in classical mechanics is result of the canonical transformation of the Hamilton variables with the generating function  $\Phi = p'_x x + p'_y y - eHxy/2$ . Then  $p_x = p'_x - (1/2)eHy$ ,  $p_y = p'_y - (1/2)eHx$ ,  $x' = x$ ,  $y' = y$ . Therefore the  $p_x, p_y$  (in fact  $p'_x, p'_y$ ) in the Landau Hamiltonian (23) are not the momentum components in the Cartesian coordinates and the absence of the  $x$ -coordinate in this Hamiltonian does not lead to the momentum  $x$ -component conservation. In quantum mechanics the unitary transformation with operator  $\exp[-ieHxy/2\hbar]$  is equivalent to this canonical transformation. The boundary is created by the line of intersection of the plane with a solenoid that generates the magnetic field. Electrons, orbits of which transverse this boundary, will be extruded from the area, and the gas will evaporate. The more realistic problem is gas in the area with a reflecting boundary.

An isolated electron has three motion integrals. Those are the angular momentum relative to the centre of area and two coordinates of the centre electron orbit:

$$l_z = xp_y - yp_x, \quad X = -\frac{p_y}{eH} + \frac{x}{2}, \quad Y = \frac{p_x}{eH} + \frac{y}{2} \quad (24)$$

Two motion integrals that have the physical importance would be created from it: energy  $E$  and squared centre electron orbit distance from the centre of area  $R^2$ :

$$R^2(\mathbf{p}, \mathbf{r}) = X^2 + Y^2 = \frac{1}{e^2 H^2} \left[ \left( p_x + \frac{eHy}{2} \right)^2 + \left( p_y - \frac{eHx}{2} \right)^2 \right]; \quad (25)$$

$$\varepsilon(\mathbf{p}, \mathbf{r}) = \frac{e^2 H^2}{2m} \left( R^2 + \frac{2}{eH} l_z \right) = \frac{e^2 H^2}{2m} \rho^2 = \hbar(\mathbf{p}, \mathbf{r}).$$

Here  $\rho$  is the radius of the electron orbit. The motion integral  $R^2$  is proportional to the Hamiltonian with opposite direction of the magnetic field. The values of  $R^2$  and  $\varepsilon$  should be discrete by the rules of the quasiclassical quantization:

$$R_k^2 = 2\lambda^2 \left( k + \frac{1}{2} \right), \quad \varepsilon_n = \hbar\omega_c \left( n + \frac{1}{2} \right); \quad \lambda = \sqrt{\frac{\hbar}{eH}} = \sqrt{\frac{\hbar}{m\omega_c}}, \quad \omega_c = \frac{eH}{m}. \quad (26)$$

Here  $\lambda$  is magnetic length and  $\omega_c$  is cyclotron frequency. Every energy level  $E_n$  is degenerated because the integer  $k$  that determines the position of the orbit can take any values from zero to  $k_1$ , where  $k_1$  is determined by the condition that electron orbit does not have common points with the boundary. That condition would be expressed by formula when the boundary is a circle with radius  $\mathfrak{R}$ . Then  $R_{k_{\max}} + \rho_n < \mathfrak{R}$  or  $\lambda(\sqrt{2k_1+1} + \sqrt{2n+1}) < \mathfrak{R}$ . Those are the “ordinary” states. When  $k_1 < k < k_2$ , where  $\lambda(\sqrt{2k_2+1} - \sqrt{2n+1}) < \mathfrak{R}$ , the states are nominated “near-boundary state”. Their energies are not described by the formula (26). The instant magnetic moment is determined by the formula:

$$\mu_z = -\frac{e}{2}(xv_y - yv_x) = -\frac{\partial}{\partial H} \hbar(\mathbf{p}, \mathbf{r}) \quad (27)$$

The first determination is valid for any negative charged particle with and without an external magnetic field. The second equality is valid when the vector potential has the form (22), or when the equality  $v_i = \partial \hbar / \partial p_i'$  is transformed by canonical transformation. For the ordinary states the averaged over the orbit magnetic moment is  $\mu_z = -\varepsilon_n / H$ . The trajectories of the near-boundary states are composed from arcs and envelop the all area. Their magnetic moment is positive. If in the area exists a potential field  $U(r)$  the orbit centers of the ordinary states also move along equipotential lines and the energy values depend on  $k$ . The degeneration of the energy levels goes off. The angular momentum does not represent the electron motion, but rather the position of the electron orbit. From Eqs. (25-26) it follows that  $l_z = (eH/2)(\rho^2 - R^2) = \hbar(n - k)$ . Then  $l_z > 0$  if the orbit envelops the centre, and  $l_z < 0$  if the area centre locates out the orbit. If  $\rho_n \ll \mathfrak{R}$  a large share of the states with energy  $\varepsilon_n$  have angular momentum  $l_z < 0$ .

Going to consideration of the ideal gas with electron-electron collisions, let us suppose that the interaction does by a central force. Then total energy and angular momentum are conserved. It is generally believed that the area is filled by uniform and motionless positive charged background that neutralizes the electrostatic interaction. This assumption is inconsistently. If electron gas is in equilibrium with motionless background, its angular

momentum should be equal to zero. But then it should be nonuniform as is evident from the foregoing consideration. It should be regarded more comprehensively. Let us go to classical statistic mechanics for a gas of charged particles in magnetic field. The characteristic function of the total system (gas and background) is:

$$\varphi(\mathbf{R}_g + \mathbf{R}_b) = \delta(\mathfrak{H}_g(\mathbf{R}_g) + \mathfrak{H}_b(\mathbf{R}_b) - \varepsilon) \delta(L_g(\mathbf{R}_g)) \delta(L_b(\mathbf{R}_b)) \quad (28)$$

Here the indexes  $g$  and  $b$  denote the quantities those relating to the gas and to the background,  $\mathfrak{H}$  is a Hamiltonian,  $\varepsilon$  is the value of the total energy. To provide of Gibbs distribution for the gas it is needed to integrate the function  $\varphi(\mathbf{R}_g + \mathbf{R}_b)$  over the phase space of the background. Then  $\delta(L_g(\mathbf{R}_g))$  is factored out from integral. The function from  $\mathfrak{H}_g(\mathbf{R}_g)$  can be factored from integral by the method Krutkov (Krutkov, 1933; Zubarev, 1974) as  $\exp(-\mathfrak{H}_g(\mathbf{R}_g)/k_B T)$ . Then:

$$Z_N = \frac{1}{(2\pi\hbar)^{3N} N! \Gamma_g} \int \delta(L_g(\mathbf{R}_g)) \exp(-\mathfrak{H}_g(\mathbf{R}_g)/k_B T) d\Gamma_g. \quad (29)$$

The Hamiltonian  $\mathfrak{H}_g(\mathbf{R}_g)$  would be represented in the form:

$$\mathfrak{H}_g(\mathbf{R}_g) = \sum_{i=1}^N \mathfrak{h}_i = \sum_{i=1}^N \left( \frac{1}{2m} \left( p_{ri}^2 + \frac{l_{zi}^2}{r_i^2} \right) + \frac{e^2 H^2 r_i^2}{8m} \right) + \frac{eH}{2m} \sum_{i=1}^N l_{zi}. \quad (30)$$

Let us substitute Hamiltonian (30) to the formula (29) and take into account formula  $\sum_{i=1}^N l_{zg} = L_g$  and the Eqs. (8). Then it is obtained:

$$Z_N = \frac{1}{(2\pi\hbar)^{3N} N! \Gamma_g} \int \exp \left\{ -\frac{1}{k_B T} \sum_{i=1}^N \left( \frac{1}{2m} \left( p_{ri}^2 + \frac{l_{zi}^2}{r_i^2} \right) + \frac{e^2 H^2 r_i^2}{8m} \right) \right\} \delta(L_g) d\Gamma_g \quad (31)$$

The integration over  $l_{z1}$  leads to change  $l_{z1} \rightarrow -\sum_{i=2}^N l_{zi}$  and after reduction of the quadratic form to the standard appearance the Hamiltonian of ideal gas that is collection of  $N-1$  particles in harmonic potential field is obtained. Obviously, that taking into account conservation of the zero value of the angular momentum eliminates the paradoxes that were mentioned at the Introduction. The magnetic moment  $\mu_z = (-\partial/\partial H)(-k_B T \ln Z_N) \neq 0$ . The gas is confined by the magnetic field. But that confinement provides to the inconsistency of the model that neglects of the electrostatic interaction because the uniform background cannot neutralize it. The model that regards this interaction will be considered below.

### 3.1.2 Quantum problem of electron in magnetic field at bounded area

The quasiclassical description of an electron in a magnetic field would not give the correct picture of the probability density distribution and the current density. It would not also describe the alternation of the energy spectrum when a perturbation does the classical

motion nonperiodical. But that problem in quantum mechanics also is considered insufficiently. As suggested in the paper (Vagner et al., 2006) the density of the probability current of the wave function  $\psi(\mathbf{r}) = \exp[i\Phi(\mathbf{r})]|\psi|$  is

$$\mathbf{j} = \frac{\hbar}{m} |\psi|^2 \left( \nabla\Phi + \frac{e}{\hbar} \mathbf{A} \right). \quad (32)$$

The eigenfunctions  $\psi_n(\mathbf{r})$  cannot be chosen so that  $\mathbf{j}=0$  because any vector potential  $\mathbf{A}$  cannot be equal to a gradient of a continuous function. Then it is necessary for stationarity of the states that the current lines to be closed in the area under consideration. The boundary condition best suited to the research of this problem is null of the function on the circumference that bounds the area:

$$\psi(\mathcal{R}, \varphi) \equiv 0 \quad (33)$$

This condition retains the greatest possible symmetry. The current lines in this case are concentric circumferences. The density of the current can be zero only at separate circumferences. Therefore the magnetization is nonuniform. The magnitudes of the eigenfunctions should have the axial symmetry. The localization of the electron cannot coincide with any classical orbit because the uncertainties of values of the orbit centre coordinates (24) should satisfy to Heisenberg uncertainty relation.

The Hamiltonian is as follows:

$$\begin{aligned} \hat{h} &= \frac{1}{2m} \left[ \left( \hat{p}_x - \frac{eHy}{2} \right)^2 + \left( \hat{p}_y + \frac{eHx}{2} \right)^2 \right] + U(r) = \hat{h}_0 + \frac{\omega_c}{2} \hat{l}_z, \\ \hat{h}_0 &= \frac{1}{2m} \left( \hat{p}_r^2 + \frac{\hat{l}_z^2}{r^2} \right) + \frac{m\omega_c^2 r^2}{8} + U(r). \end{aligned} \quad (34)$$

The operators  $\hat{h}_0$  and  $\hat{l}_z$  commute with each other and each of them with the Hamiltonian  $\hat{h}$ . The potential energy  $U(r)$  is created by the interaction with other electrons and with a neutralizing background. The nonuniform magnetization  $\mathbf{M}(r)$  should be neglected. If the potential energy  $U(r)$  also will be neglected the eigenfunctions of the Hamiltonians  $\hat{h}$  and  $\hat{h}_0$  will have the form:

$$\begin{aligned} \Psi &= \exp(i\ell\varphi) \psi_{|\ell\alpha}(r); \quad \ell = 0, \pm 1, \pm 2, \dots \\ \psi_{|\ell\alpha}(r) &= \frac{A}{\lambda} \left( \frac{r}{\lambda\sqrt{2}} \right)^{|\ell|} \exp\left( -\frac{r^2}{4\lambda^2} \right) \Phi\left( \alpha, |\ell| + 1; \frac{r^2}{2\lambda^2} \right). \end{aligned} \quad (35)$$

Here  $\hbar l$  is eigenvalue of the operator of the angular momentum,  $\Phi(a, c; x)$  is the degenerate hypergeometric function,  $A$  is normalizing factor that depends on  $|\ell|$  and  $\alpha$ . The eigenvalues of energy are expressed by  $\alpha$  and  $l$ : for the Hamiltonian  $\hat{h}_0$  that is  $\varepsilon_0 = \hbar\omega_c [ -\alpha + (|\ell| + 1)/2 ]$ , and for  $\hat{h}$  that is  $\varepsilon = \hbar\omega_c [ -\alpha + (|\ell| + 1)/2 ] + \hbar\omega_c (l/2)$ . The permissible values of  $\alpha$  are determined by the boundary condition. In the common theory (Landau, & Lifshitz, E.M., 1980b; Vagner et al., 2006) that is normability of the eigenfunctions at an infinity plane. Then it

is necessary that  $\alpha = -n_r$ , where  $n_r$  is integer or zero.  $\Phi(-n_r, |l| + 1; x) = [n_r! |l|! / (|l| + n_r)!] L_{n_r}^{|l|}(x)$ , where  $L_{n_r}^{|l|}(x)$  is the Laguerre polynomial. The energy spectrum of the operator  $\hat{h}_0$  is  $\varepsilon_0(n_0) = (\hbar\omega_c/2)(2n_r + |l| + 1) = (\hbar\omega_c n_0/2)$ , where  $n_0$  is integer. The energy spectrum of the operator  $\hat{h}$  is  $\varepsilon(n) = (\hbar\omega_c/2)(2n_r + l + |l| + 1) = \hbar\omega_c(n + 1/2)$  where  $n$  is integer. There are two kinds of a degeneracy of energy levels of those Hamiltonians. The degeneracy of the first kind arise as result of the formulae:  $n_0 = 2n_r + |l| + 1$ , and  $n = n_r + l$  when  $l > 0$ . Every level is degenerated with multiplicity that equals to its number. The perturbation  $U(r)$  would be described as a power series without the linear term. Then the term that is proportional to  $r^2$  in the spectrum of the Hamiltonian  $\hat{h}_0$  should change the distance between levels without elimination of the degeneracy. In the spectrum of the Hamiltonian  $\hat{h}$  the level would be split. The other terms of potential field series eliminate the degeneracy in the first order of the perturbation theory. The second kind of the degeneracy is inherent only to the spectrum of the Hamiltonian  $\hat{h}$  in absence of any perturbation. Every level is degenerated with infinity multiplicity because when  $l < 0$  the energy value does not depend on  $l$ . At a bounded area the multiplicity would be limited but very large and would depend on the magnetic field. That will be shown below. This degeneracy eliminates by any potential field. The modulo  $l$  would have many integer value, and when  $\langle l|U|l \rangle \geq \hbar\omega_c$ , if that's possible at some  $l$ , the spectrum would be quasicontinuous. That is essential because the explanations of many experimental effects rely on the discreteness of the Landau spectrum. In so doing ones do not study the model stability relative to the electrostatic interaction that is unavoidable perturbation. The other inconsistency of the common model is the large-scale negative total angular momentum of the ground state when all levels have the identical and large-scale multiplicity of degeneracy. Let us study the degeneracy eliminating by the boundary condition (33) in the absence of other perturbation. The polynomial  $L_{n_r}^{|l|}(x)$  has  $n$  simple zeroes, which, if  $n_r \gg 1$ , are expressed by formula:

$$\zeta(n_r, |l|; i) \approx \frac{j^2(|l|, i)}{2|l| + 4n_r + 2}. \tag{36}$$

Here  $j(|l|, i)$  is a null of the Bessel function  $J_{|l|}(x)$  with number in the order of increasing  $i$ . Then the function  $\psi_{|l|(-n_r)}$  has  $(n_r + 1)$  extremes that are decrease modulo. The extreme number  $(n_r + 1)$  has the sign  $(-1)^{n_r}$ , and further the function tends to zero asymptotically. Obviously, this function cannot satisfy to the boundary condition (33). When  $\alpha = -n_r - \gamma$  the degenerate hypergeometric function has the form:

$$\begin{aligned} \Phi(-n_r - \gamma, |l| + 1; x) &= [(n_r)! |l|! / (|l| + n_r)!] L_{n_r}^{|l|}(\gamma; x) - \gamma A_{n_r}(\gamma, x) - (-1)^n \gamma(1 - \gamma) T_{n_r}(x), \\ A_{n_r}(\gamma, x) &= -(-x)^{n_r + 1} [\Gamma(n_r + 1 + \gamma) \Gamma(|l| + 1)] [\Gamma(|l| + n_r + 2) \Gamma(1 + \gamma) (n_r + 1)!]^{-1}, \end{aligned} \tag{37}$$

where  $L_{n_r}^{|l|}(\gamma; x)$  is the polynomial, the nulls of which are less than  $\zeta(n_r, |l|; i)$  by quantities that are proportional to  $\gamma$ , and  $T_{n_r}(x)$  is the infinite series that at  $x \gg \zeta(n_r, |l|; n_r)$  is proportional to  $x^{-(|l| + 1 + n_r)} \exp(x)$ . Then the function (35) will have one more null at the large-

scale value  $x$ . This null  $X$  tends to infinity when  $\gamma$  tends to zero, and it would be shown that

$$\gamma \approx \frac{X^{2n_r+|l|+1}}{(|l|+n_r)!(n_r)!} \exp(-X), \quad (38)$$

when  $\gamma \ll 1$ . When  $\gamma \rightarrow 1$ ,  $X \rightarrow \zeta(n_r+1, |l|; n_r+1)$ . Then the boundary condition (33) would be satisfied when  $(\mathcal{R}^2/2\lambda^2) = X$ . (In the mathematical handbook (Erdélyi, 1953) it is written that the function  $\Phi(\alpha, |l|+1; x)$  at  $\alpha = -n_r - \gamma$  has  $n_r + 1$  nulls, but all these nulls are determined by the formula that is like to the formula (36). Then the boundary condition would be satisfied at arbitrarily large values  $(\mathcal{R}^2/2\lambda^2)$  only when energy has also large-scale value). The value of the null  $\zeta(n_r+1, |l|; n_r+1)$  increases when the value of  $|l|$  increases. The maximal value  $|l_1|$ , at which the inequality  $(\mathcal{R}^2/2\lambda^2) \geq \zeta(n_r+1, |l_1|; n_r+1)$  is fulfilled, determines the number of the eigenstates of the Hamiltonian  $\hat{h}_0$  that have  $\gamma \ll 1$  (ordinary states). It can be shown that when  $|l_1| \gg n_r$ , then  $|l_1| \approx (\mathcal{R}^2/2\lambda^2) - (n_r+1)(\mathcal{R}/2\lambda)$ . The quantum number of the energy  $n$  is equal to  $n_r$ , because when  $l < 0$  the energy depends on  $l$  only by  $\gamma(n, l)$ . This number of the ordinary states is consistent with the estimate that was obtained in classic mechanics theory in section 3.1.1 and in the work (Landau, 1930). Those ordinary states are quasi degenerated. The multiplicity of this degeneration is proportional to the magnetic field. Then the main term in the thermodynamical potential, as it was computed by Landau (Landau, 1930), should not depend on the magnetic field, and the Bohr – van Leeuwen theorem is vindicated. When  $|l| > |l_1|$  and  $l < 0$  the other nulls that are described by the formula (36) would satisfy the boundary condition. Those are the near-boundary states, and in this case are not restrictions on the values of  $\gamma$  besides  $0 < \gamma < 1$ . There are  $n$  near-boundary states.

In the section 3.2 it will be shown that for statistical mechanics of the electron gas in the magnetic field the Hamiltonian  $\hat{h}_0$  has fundamental importance. Let us study the spectrum of this Hamiltonian with boundary condition (33). Then

$$\begin{aligned} \varepsilon_0(n_0, n_r) &= \frac{\hbar\omega_c}{2} [n_0 + 2\gamma(n_0, n_r)], \quad n_0 = 2n_r + |l| + 1, \\ \gamma(n_0, n_r) &= \frac{(\mathcal{R}^2/2\lambda^2)^{n_0}}{(n_0 - n_r - 1)!(n_r)!} \exp(-\mathcal{R}^2/2\lambda^2). \end{aligned} \quad (39)$$

The degenerate levels are transformed in zonule. It follows from formula (39) that the zonule upper edge is determined by minimum value of  $(n_0 - n_r - 1)!(n_r)!$  that is roughly  $[(n_0/2)!]^2$ . In the vicinity of  $n_r = n_0/2$  a shift of energy  $\gamma$  most slowly varies with  $n_r$ . It means that a density of states is the highest in the vicinity of the zonule upper edge. The zonule lower edge is shifted by  $\gamma_{\min} = \gamma_{\max} [(n_0/2)!]^2 / n_0! \ll \gamma_{\max}$ . The zonule width  $\gamma_{\max}(n_0)$  increases with  $n_0$  if  $(\mathcal{R}^2/2\lambda^2) > (n_0/2)$ . If  $\gamma_{\max}(n_0) < 1/2$  then number of states on each interval of energy values of width  $\hbar\omega_c/2$  is equal to  $n_0$ , (the spin will be taken into account subsequently). For  $\gamma_{\max}(n_0) > 1/2$  zonules overlap, gaps in the spectrum disappear, and the number of states in the interval becomes less than  $n_0$ , i.e., grows of the density of states is decelerated with energy increase. That is transitive area of energy. For higher



values of energy not the greatest but other nulls, which are described by formula (36), satisfy the boundary condition. With the approximate formula for the nulls of the Bessel function  $j(|l|, \nu) \approx \pi(\nu + |l|/2)$  the second in magnitude null of the function (35) with  $\nu = n_r$  is  $[\pi^2 n_0^2 / 8(n_0 + 2\gamma')] \approx (\pi^2 n_0 / 8)$ . When it will be so that  $\pi^2 n_0 / 8 \geq \mathcal{R}^2 / 2\lambda^2$  the boundary condition should be satisfied only not the greatest but other nulls of the degenerate hypergeometric function. Corresponding values of energy look like:

$$\varepsilon_0(n_0) = \frac{\hbar\omega_c}{2}(n_0 + 2\gamma') = \frac{\hbar\omega_c}{2} \left( \frac{\mathcal{R}^2}{\lambda^2} \right)^{-1} j^2(|l|, \nu) = \varepsilon_0(|l|, \nu) = \frac{\hbar^2}{2m\mathcal{R}^2} j^2(|l|, \nu) \quad (40)$$

where  $\nu = n_r \leq n/2$  and  $\gamma'$  is not described by the formula (39). They coincide with eigenvalues energy in a circular potential well with reflecting boundaries (see formula (19)). The spectrum becomes quasicontinuous as distances between the nearest levels are proportional to  $\mathcal{R}^2$ . The density of states does not depend on energy, as well as for two-dimensional gas of free particles. The function (35) would be expanded in series over the Bessel functions (Erdélyi, 1953) and the first term is proportional to  $J_{|l|}((r/\hbar)\sqrt{2\varepsilon m})$ . The parameter of this expansion is  $(n_0)^{-1/2}$ . Hence, the wave functions in this approximation also coincide with free electron wave functions. This form of the spectrum would be illustrated by the classical consideration. If the formula (34) would be considered as the classical Hamiltonian, the ultimate energy for which the classical accessible area is determined by the parabolic potential is equal to  $m\omega_c^2 \mathcal{R}^2 / 8$ . That within a factor  $\xi$  that is close to 1 coincides with the energy of the transitive area and with energy of the transition to quasicontinuous spectrum  $(2/\pi^2)m\omega_c^2 \mathcal{R}^2$ .

These results can be described as energy spectrum breakdown into two bands. A spectrum lower part is denoted as a magnetic band, and the upper one will be denoted as a conduction band. Bands are not separated by a gap or sharp boundary, but far from transitive area the density of states and wave functions differ substantially. Fine structure of the density of states in the lower part of magnetic band represents the narrow zonule separated by gaps. The total width of the allowed zonule and gap is equal to  $\hbar\omega_c/2$ . Number of states in an interval with number  $n_0$  is equal to  $n_0$ . In the transitive area gaps disappear and in the conduction band the spectrum is quasicontinuous. The maximum of magnitude of a wave function in the magnetic band is localized within a ring of width about  $\lambda\sqrt{2}$  and about  $R = \lambda\sqrt{2n_0}$  in radius. The density of states in the magnetic band, averaged over the interval  $\hbar\omega_c/2$ , is  $2n_0/\hbar\omega_c = 4\varepsilon_0/(\hbar\omega_c)^2$  and grows with energy. In the transitive area this grows is decelerated, and in the conduction band the density of states does not depend on energy and is equal to  $m\mathcal{R}^2/2\hbar^2$  (without spin consideration). It is believed that the magnetic band is finished when  $n_0 = \mathcal{R}^2/4\lambda^2 = n_b$  or  $\varepsilon_0 = m\omega_c^2 \mathcal{R}^2 / 8 = \varepsilon_b$ , as would be expected from the quasiclassical consideration. Then the density of states will be continuous.

### 3.2 Statistical mechanics of electron gas in uniform magnetic field with regard for electrostatic interaction

The quantum-mechanical average value of the magnetic moment in the ordinary eigenstate of the Hamiltonian  $\hat{h}$  in absence of the potential energy  $U(r)$  is:

$$\mu_z = -(\partial\varepsilon(n,l)/\partial H) = -\{(\hbar e/m)(n + \gamma) + (\hbar e H/m)(\partial\gamma/\partial H)\} \quad (41)$$

It is a negative quantity because the positive term that proportional to  $(\partial\gamma/\partial H)$  is small. In the paper (Landau, 1930) were taken into account only ordinary states. Then the quantum-statistical average value of the gas magnetic moment must be:

$$\mathfrak{M} = -\sum_{nl} \frac{\partial\varepsilon(n,l)}{\partial H} n(n,l) = -\sum_{nl} \frac{\partial\varepsilon(n,l)}{\partial H} \left\{ D(H) \left[ 1 + \exp\left(\frac{\varepsilon(n,l) - \mu}{k_B T}\right) \right]^{-1} \right\} < 0 \quad (42)$$

Here  $n(n,l)$  is the average occupation number of the state  $\Psi(n,l)$ ,  $D(H)$  is the multiplicity of degeneracy that in this case does not depend on the state energy and depend on the magnetic field. But in this work the magnetic moment of the gas was computed as

$$\begin{aligned} \mathfrak{M} &= -\frac{\partial\Omega}{\partial H} = k_B T \frac{\partial}{\partial H} \sum_{nl} \left\{ D(H) \ln \left[ 1 + \exp\left(\frac{\mu - \varepsilon(n,l)}{k_B T}\right) \right] \right\} = \\ &k_B T \sum_{nl} \left\{ \frac{\partial D(H)}{\partial H} \ln \left[ 1 + \exp\left(\frac{\mu - \varepsilon(n,l)}{k_B T}\right) \right] \right\} - \sum_{nl} \frac{\partial\varepsilon(n,l)}{\partial H} n(n,l). \end{aligned} \quad (43)$$

When this result is compared with the formula (42) it is apparent that the thermodynamical potential  $\Omega$  is determined incorrectly.

Let us obtain the density of distribution for an electron gas that is at equilibrium with thermostat, which is described by classical mechanics. The conservation of the zero value of the angular momentum also will be taken into account. The characteristic function of this system is:

$$\begin{aligned} \varphi &= (2\pi)^{-3} I_E \int_0^{2\pi} \int_0^{2\pi} \exp \left[ \left( N - \sum_{\Psi} a_{\Psi}^+ a_{\Psi} \right) (i\phi + \chi) + \left( \sum_{\Psi} l_z(\Psi) a_{\Psi}^+ a_{\Psi} \right) (i\alpha + \eta) \right] d\phi d\alpha \\ I_E &= 2\pi \int_0^{\infty} \varphi_1(\varepsilon) \varphi_2(E - \varepsilon) d\varepsilon = \\ &\int_0^{\infty} \delta[\mathfrak{H}_h - (E - \varepsilon)] \int_0^{2\pi} \exp \left[ \left( \varepsilon - \sum_{\Psi} [\varepsilon_0(\Psi) + (\hbar\omega_c/2)l_z(\Psi)] a_{\Psi}^+ a_{\Psi} \right) (i\tau + \vartheta) \right] d\tau d\varepsilon. \end{aligned} \quad (44)$$

This formula is obtained on a basis of the properties of characteristic functions that was described in the formulae (6 - 8), and the quantum characteristic function (see formula (15)). Here  $E$  is a total energy,  $\varepsilon$  is the energy of the electron gas,  $\mathfrak{H}_h$  is the Hamiltonian of the thermostat that is a summatory function of the classical variables. Values of energy are expressed by the dimensionless positive integers, and the angular momentum is measured in  $\hbar$ . The expression of the Hamiltonian  $\hat{h}$  as the sum  $\hat{h}_0 + (\omega_c/2)\hat{l}_z$  is taken into account. Then the second equality in (44) would be rewritten as:

$$I_E = \int_0^{\infty} \delta[\mathfrak{H}_h - (E - \varepsilon)] \int_0^{2\pi} \exp \left[ \left( \varepsilon - \sum_{\Psi} \varepsilon_0(\Psi) a_{\Psi}^+ a_{\Psi} \right) (i\tau + \vartheta) \right] d\tau d\varepsilon \quad (45)$$

by using the formula (8). Let us generalize the Krutkov method (Krutkov, 1933; Zubarev, 1974) for this case. We calculate the Laplace transformation with respect the total energy  $E$ . To do this let us multiply the function  $(2\pi)^{-1} I_E$  (formulae (44 - 45) by  $\exp(-\mathcal{G}E)$  and integrate over  $E$  between 0 and  $\infty$ . Then the integrations over  $\mathcal{E}$  and  $\tau$  can be performed, and the result is:

$$I_E = \int_0^\infty \delta[\mathcal{H}_{th} - (E - \mathcal{E})] \int_0^{2\pi} \exp\left[\left(\mathcal{E} - \sum_\Psi \varepsilon_0(\Psi) a_\Psi^+ a_\Psi\right)(i\tau + \mathcal{G})\right] d\tau d\mathcal{E} \quad (46)$$

where  $\mathcal{L}\{f(E), \mathcal{G}\} = \mathcal{L}_\mathcal{G}[f(E)]$  is Laplace transform of a function  $f(E)$ . The integration over the thermostat phase space leads to the formula:

$$(2\pi)^{-1} [t_{th}(\mathcal{G})]^{N_{th}} \exp\left(-\mathcal{G} \sum_\Psi \varepsilon_0(\Psi) a_\Psi^+ a_\Psi\right), \quad t_{th}(\mathcal{G}) = \int_{\Gamma_1} \exp(-\mathcal{G}h_{th}) d\Gamma_1 \quad (47)$$

where  $\Gamma_1$  is the phase space of a thermostat particle and  $h_{th}$  is its Hamiltonian and  $N_{th}$  is the number of thermostat particles. As result of inverse Laplace transformation we obtain:

$$\int_{\Gamma_{th} 0}^\infty \varphi_1(\mathcal{E}) \varphi_2(E - \mathcal{E}) d\mathcal{E} d\Gamma_{th} = (2\pi)^{-1} \int_{a-i\infty}^{a+i\infty} \exp\left[N_{th} \left(\mathcal{G} \frac{E}{N_{th}} + \ln t(\mathcal{G})\right)\right] \exp\left[-\mathcal{G} \sum_\Psi \varepsilon_0(\Psi) a_\Psi^+ a_\Psi\right] d\mathcal{G} \quad (48)$$

This integration can be performed by the saddle-point method because  $N_{th}$  is a large-scale number. The saddle-point  $\mathcal{G}_0 = (k_B T)^{-1}$  is determined by the first exponent, and the second exponent that depend on variables-operators of the quantum subsystem can be factored from integral at  $\mathcal{G} = (k_B T)^{-1}$  as well as in the classical case. The result should be substituted into (44). Then the non-normalized statistical operator is obtained:

$$\hat{\rho} = \exp\left[-(k_B T)^{-1} \sum_\Psi \varepsilon_0(\Psi) a_\Psi^+ a_\Psi\right] \int_0^{2\pi} \exp\left[\left(N - \sum_\Psi a_\Psi^+ a_\Psi\right)(i\phi + \chi)\right] d\phi \times \int_0^{2\pi} \exp\left[\left(\sum_\Psi L_z(\Psi) a_\Psi^+ a_\Psi\right)(i\alpha + \eta)\right] d\alpha. \quad (49)$$

Here the first multiplier is the common formula of the statistical operator for the quantum Gibbs distribute (Landau, & Lifshitz, E.M., 1980a; Zubarev, 1974). The second multiplier would be computed by the Darwin - Fowler method (Fowler, & Guggenheim, 1939) like as in the formula (16) and describes the conservation of the particle number. If the grand canonical ensemble is considered, then the statistical operator of particle number

$$\hat{\rho}_N = \exp\left[-\mu(k_B T)^{-1} \sum_\Psi a_\Psi^+ a_\Psi\right] \quad (50)$$

would be obtained from this multiplier by the Krutkov method. The last multiplier in formula (49) cannot be computed by those methods because it does not have any large-scale parameter. This multiplier imposes constraints on ensembles that the total angular momentum equal to zero. If Hamiltonian and an operator that should be averaged have the

commutative term that is proportional to the total angular momentum operator, this term should be eliminated when the averaging is performed. That is the reason for the change  $\varepsilon$  to  $\varepsilon_0$  in the Hamiltonian  $\sum_{\psi} \varepsilon_0 (\Psi) a_{\psi}^{\dagger} a_{\psi}$  in the formula (48).

The model that will be considered below is described by the Hamiltonian:

$$\hat{\mathcal{H}} = \sum_{i=1}^N \hat{k}_{0i}; \quad \hat{k}_0 = \frac{1}{2m} \left( \hat{p}_r^2 + \frac{\hat{l}_z^2}{r^2} \right) + \frac{m\omega_c^2 r^2}{8} + U(r) \quad (51)$$

Here  $U(r)$  is potential of the self-consistent electric field that describes interaction of an electron with other electrons and neutralizing background that have the form of circle with radius  $\mathcal{R}$ .

The electron density in the magnetic field should be distributed in such a way as to shield the external potential  $m\omega_c^2 r^2/8$ . This shielding cannot be perfect. The current that would be generated by a residual potential would be compensated by the diffusion current that generated by an inhomogeneity of the electron density. Because the kinetic energy in the Hamiltonian (51) has the standard form, electrons density distribution and self-consistent potential can be calculated by the density functional method, described in the work (March, 1983). In this approximation, energy of the ground state of electron gas is presented in the form of functional of the gas density, i.e., number of particles per unit of area  $\rho(r)$ :

$$E_0[\rho(\mathbf{r})] = \left( \pi \hbar^2 / 2m \right) \int_C \rho^2(\mathbf{r}) d\mathbf{r} + \left( e^2 / 8\pi \varepsilon_0 \right) \int_C \int_C \rho(\mathbf{r}) \rho(\mathbf{r}') |\mathbf{r} - \mathbf{r}'|^{-1} d\mathbf{r} d\mathbf{r}' - \left( e^2 / 4\pi \varepsilon_0 \right) \int_C \rho(\mathbf{r}) [u(\mathbf{r}) + u_0] d\mathbf{r} + \left( m\omega_c^2 / 8 \right) \int_C \rho(\mathbf{r}) r^2 d\mathbf{r}. \quad (52)$$

Here  $\pi \hbar^2 \rho^2 / 2m$  is density of kinetic energy of two-dimensional degenerated Fermi gas (each state is supposed to be twice degenerated), the second addend describes energy of electrons Coulomb interaction,  $\varepsilon_0$  is vacuum inductivity, the third addend is energy of electrostatic interaction with the neutralizing background, the last addend described gas potential energy in the effective harmonic potential depending on magnetic field. The neutralizing background has the charge density  $e\rho_0 Y(\mathcal{R} - r)$  and generate the potential  $eu(\mathbf{r})$ . This potential would be expressed by the elliptical integrals and is approximated by the parabola and a repulsing bound. The step function  $Y(x) = 1$  at  $x > 0$  and  $Y(x) = 0$  at  $x \leq 0$ . The constant  $eu_0$  would be assigned so that the potential energy in the effective Hamiltonian would be zero at  $r = 0$ . All integrations are made over the area  $C$ , which is a circle of radius  $\mathcal{R}$ . The functional  $E_0[\rho(\mathbf{r})]$  should be minimized under the following supplementary condition:

$$\int_C \rho(\mathbf{r}) d\mathbf{r} = \pi \mathcal{R}^2 \rho_0 \quad (53)$$

The quadratic term of the residual potential is of chief interest. Therefore the minimization would be performed in the quadratic approximation. The test function has the form:

$$\rho(r) = \rho_0 + \alpha \mathcal{R}^{-2} - \beta \mathcal{R}^{-4} r^2, \quad (54)$$

and as result we obtain:

$$\alpha = \frac{(m\omega_c \mathcal{R}^2)^2 a_0}{12\pi\hbar^2 \mathcal{R}} \left(1 - \frac{4a_0}{3\mathcal{R}}\right), \quad \beta = \frac{(m\omega_c \mathcal{R}^2)^2 a_0}{6\pi\hbar^2 \mathcal{R}} \left(1 - \frac{4a_0}{3\mathcal{R}}\right), \quad (55)$$

where  $a_0 = 4\pi\epsilon_0\hbar^2/me^2$  is Bohr radius. The Lagrange multiplier  $\zeta$ , which has meaning of the chemical potential or the Fermi energy is obtained:

$$\zeta = (\pi\hbar^2 \rho_0/m) + m\omega_c^2 \mathcal{R} a_0/12 = \zeta_0 + m\omega_c^2 \mathcal{R} a_0/12, \quad (56)$$

where  $\zeta_0$  is the Fermi energy of two-dimensional electron gas in absence of a magnetic field. The residual potential energy has the form  $(4a_0/3\mathcal{R})(m\omega_c^2 r^2/8)$ . It is expressed conveniently by new frequency  $\omega_r = \omega_c \sqrt{4a_0/3\mathcal{R}}$ . Then one-particle effective Hamiltonian will have the form that was study in the previous subsection:

$$\hat{h}_r = \frac{1}{2m} \left( \hat{p}_r^2 + \frac{\hat{l}_z^2}{r^2} \right) + \frac{m\omega_r^2 r^2}{8} \quad (57)$$

The Coulomb interaction and the electron density inhomogeneous that are neglected commonly decrease the frequency  $\omega_c \rightarrow \omega_r$  or another way the effective charge  $e \rightarrow e_r$  by factor  $\sqrt{4a_0/3\mathcal{R}} \approx 10^{-4}$ . This factor does not depend on the electron density if this density satisfies to the condition of the degeneration. The operator of the total magnetic moment  $\mathcal{O}\hat{l} = -(e_r/2m) \sum_i^N (\hat{l}_i + e_r H r_i^2/2)$ . When the quantum-statistical average is computed,  $\sum_i^N \hat{l}_i = 0$ , and  $\mathcal{O}\hat{l} = \langle \mathcal{O}\hat{l} \rangle = - \left\langle \sum_i^N \partial \hat{h}_r / \partial H \right\rangle$ . The energy magnetic splitting over spin does not alter. Hence it should be considered separately.

The electron gas always interacts with electromagnetic field that in the ordinary circumstance has zero temperature. This leads to the fact that the gas passes to the ground state by the spontaneous photon irradiation. The role of statistical mechanics is that it imposes a constraint on a value of the angular momentum in the ground state. The energy of the ground state with zero angular momentum would be computed by the spectrum of the one-particle states that was described in the previous subsection. The energy levels, the degeneracy multiplicity, and the boundary of the magnetic band are determined only the orbital motion. Then the energy of the magnetic band is:

$$E_{mb} = \hbar\omega_r \sum_{n=1}^{n_b} (n + 2\gamma_{n|n}) n \approx \frac{\hbar\omega_r n_b^3}{3} = \frac{e^4 a_0^2}{108\hbar^2 m} \mathcal{R}^4 H^4 \quad (58)$$

The number of states in the magnetic band equals to:

$$N_b = n_b (n_b + 1) \approx n_b^2 = (e^2 a_0 / 12\hbar^2) \mathcal{R}^3 H^2 \quad (59)$$

The quantity  $n_b$  is discrete; therefore the formulae (58 - 59) describe a smoothed function. The distance between the neighboring spikes in  $E_{mb}$  is  $\hbar\omega_r n_b^2$ . If  $N_b$  is less than the number of electrons  $N = \pi\mathcal{R}^2 \rho_0$  the energy of the ground state with a glance the formula (56) is:

$$E_0 = E_{mb} + \frac{m\mathcal{R}^2}{\hbar^2} \int_{\varepsilon_b}^{\zeta} \varepsilon d\varepsilon - \frac{m\mathcal{R}^2}{\hbar^2} \left( \frac{e\hbar}{2m} H \right)^2 = \quad (60)$$

$$\frac{m\mathcal{R}^2 \zeta_0^2}{2\hbar^2} + \frac{\pi\rho_0 a_0 e^2}{12m} \mathcal{R}^3 H^2 - \frac{a_0^2 e^4}{864\hbar^2 m} \mathcal{R}^4 H^4 - \frac{e^2}{4m} \mathcal{R}^2 H^2.$$

Here the first term is the energy of the electron gas in the absence of a magnetic field. The second term that makes the main addend in the magnetic moment is the product of the Fermi energy of the electron gas  $\zeta_0$  and the additional term in it, which depends on the magnetic field (formula (56)). The third term in the formula (60) is created by the addends  $E_{mb}$ , the squared additional term of the Fermi energy, and the negative addend in the result of integration  $-(m\mathcal{R}^2 \varepsilon_b^2 / 2\hbar^2)$ . The last term in this formula describes the energy lowering by the spin polarization. It always is smaller than the second term. The magnetic moment is:

$$\mathcal{M} = -\frac{\partial(E_0 - E_p)}{\partial H} - \frac{E_p}{H} = -\frac{\pi\rho_0 a_0 e^2}{6m} \mathcal{R}^3 H + \frac{a_0^2 e^4}{216\hbar^2 m} \mathcal{R}^4 H^3 + \frac{e^2}{4m} \mathcal{R}^2 H \quad (61)$$

where  $E_p = -(e\mathcal{R}H)^2 / 4m$ . The first term is the diamagnetic moment of the orbital motion in the states of the conduction band. It is proportional to the electron density and third power of the radius as distinct from the result of the Landau theory (see to formula (5)). This formula would be obtained by other way. Let us consider the density of electric current (formula (32)) in the ground state with zero angular momentum:

$$j_\varphi = -\frac{e_r \hbar}{2m} \left\langle \sum_{i=1}^N \left\{ \left[ \frac{\hat{l}_i}{r_i} + \frac{e_r H r_i}{2\hbar} \right] \delta(\mathbf{r}_i - \mathbf{r}) + \delta(\mathbf{r}_i - \mathbf{r}) \left[ \frac{\hat{l}_i}{r_i} + \frac{e_r H r_i}{2\hbar} \right] \right\} \right\rangle \quad (62)$$

The magnetic moment is:

$$\mathcal{M} = \pi \int_0^{\mathcal{R}} j_\varphi(r) r^2 dr = -\frac{e_r \hbar}{m} \sum_{nl} n(n,l) l \pi \int_0^{\mathcal{R}} |\psi_{nl}(r)|^2 r dr - \frac{\pi e_r^2 H}{2m} \int_0^{\mathcal{R}} \sum_{nl} |\psi_{nl}(r)|^2 n(n,l) r^3 dr \quad (63)$$

The first term equals to zero because  $2\pi \int_0^{\mathcal{R}} |\psi_{nl}(r)|^2 r dr \equiv 1$  and  $\sum_{nl} n(n,l) l = 0$ . Let us change in the second addend  $\sum_{nl} |\psi_{nl}(r)|^2 n(n,l) = \rho(r)$ . Then the integration gives both the first and the second terms of the formula (61). If the sum  $\sum_{nl} n(n,l) l$  will be computed in the common theory, the linear with respect to the magnetic field terms in the formula (63) will cancel.

The orbital diamagnetic susceptibility decreases with increasing of the magnetic field. This function also has a spikes that is caused by the addend  $\partial^2 E_{mb} / \partial H^2$ . When  $N_p$  will be equal to  $N = \pi \mathcal{R}^2 \rho_0$  the diamagnetic susceptibility will be zero. When the magnetic field is stronger, the Fermi energy remains on the level  $n_f = \sqrt{N}$  of the magnetic band and increases proportionally to magnetic field. The magnetic moment remains constant. The paramagnetic moment also does not increase because the density of states on the Fermi level decreases linearly. The value of the magnetic induction whereby the saturation of the magnetic moment will start is:

$$H_s = \frac{\hbar}{e} \left( \frac{12\pi\rho_0}{a_0\mathcal{R}} \right)^{1/2} = 5.15 \cdot 10^{-10} \left( \frac{\rho_0}{\mathcal{R}} \right)^{1/2} \text{ T} \quad (64)$$

If  $\rho_0 = 10^{19} \text{ m}^{-2}$ ,  $\mathcal{R} = 0.1 \text{ m}$ , then  $H_s = 5,15 \text{ T}$  and  $\mathcal{M}_s = -2.92 \cdot 10^{-2} \text{ J} \cdot \text{T}^{-1}$ .

#### 4. Conclusion

The fundamental theory of statistical mechanics requires taking into account the law of the angular momentum conservation. The fulfilment of this requirement does not introduce any essential alterations into statistical thermodynamics, when the angular momentum of the system equal zero and the system Hamiltonian is a positive definite quadric form of all momenta. An equilibrium isolated system would have nonzero angular momentum only if an attraction of particles can resist centrifugal forces, as it is in nebulas. A gas can be in equilibrium with a rotating envelope that is a termospinstat. The condition of this equilibrium is the equality of the average value of sum of particles angular velocities to the angular velocity of the envelope. The Gibbs density of distribution and the thermodynamical functions are generalized for this case. If a system has the angular momentum equal to zero, the conservation of this value is important only when the Hamiltonian or/and an averaged quantity depend on the angular momentum. The problem of an electron gas in a uniform magnetic field is considered with taking into account the conservation of the zero value of the angular momentum. This consideration eliminates the paradoxical statement of the conventional theory that diamagnetic moment of the gas equals zero in classical as well as quantum physics (the Bohr - van Leeuwen theorem). The new formulae for the magnetic moment of the electron gas are obtained. It also leads to the effect of confinement of two-dimensional gas of charged particles by magnetic field. This results in effect of a non-uniform density of a gas, which decreases with distance from a center according both to classical as well as quantum theory. Then the model of noninteracting charged particles does not have areas of application. Many theories should be reconsidered, if they are founded on this model and on the statistical mechanics which does not take into account the angular momentum conservation law.

#### 5. References

- Abrikosov, A.A. (1972). *Introduction to the Theory of Normal Metals*, Nauka, Moscow.
- Cohen, K. (1951). *The Theory of Isotope Separation as Applied to the Large Scale Production of U235*, McGraw - Hill, New York.
- Erdélyi, A. (1953). *Higher Transcendental Functions, based, in part, on notes left by Harry Bateman*, Vol. 1, Mc Graw-Hill book company, INC, New York Toronto London.
- Feynman, R.P., Leighton R.B., & Sands M. (1964). *The Feynman lectures on physics*, Vol. 2, Addison-Wesley Publishing Company, Inc. Reading, Massachusetts. Palo Alto. London
- Fowler, R.H., & Guggenheim, E.A. (1939). *Statistical Thermodynamics*, Cambridge University Press, London.
- Khinchin, A.Y. (1949). *Mathematical Foundation of Statistical Mechanics*, Ed. Dover, NewYork.
- Khinchin, A.Y. (1960). *Mathematical Foundation of Quantum Statistics*, Ed. Dover, NewYork.

- Krutkov, Y.A. (1933). *Zs. Phys.* v. 81, p. 377, & *Supplement by Editor*, In translation into Russian of the book by H. A. Lorentz, *Statistical Theory in Thermodynamics*, (1935), ONTI, Leningrad – Moscow.
- Landau, L.D. (1930). Diamagnetism of Metals, In: *Collected papers*, Vol. 1, p. 47, (1969), Nauka, Moscow.
- Landau, L.D., & Lifshitz, E.M. (1980a). *Statistical Physics* (3<sup>rd</sup> rev., part I), Pergamon Press, New York.
- Landau, L.D., & Lifshitz, E.M. (1980b). *Quantum mechanics. Nonrelativistic theory*, Pergamon Press, New York.
- Lifshits, I.M., Azbel, M.Y., & Kaganov, M.I. (1973). *Electron Theory of Metals*, Consultant Bureau, ISBN 030-610-8739, New York.
- March, N.H. (1983). Origins: theory by Thomas – Fermi, In: *Theory of the inhomogeneous electron gas*, Ch. 1, S. Lundqvist & N.H. March (Ed.), Plenum Press, New York & London.
- Pohl, R.W. (1960). *Elektrizitätslehre*, Springer – Verlag, Berlin – Göttingen – Heidelberg.
- Shoenberg, D. (1984). *Magnetic oscillations in metals*, Cambridge University Press, Cambridge.
- van Vleck J.H. (1965). *Theory of Electric and Magnetic Susceptibilities*, Oxford University Press, Oxford.
- Uhlenbeck, G., & Ford, G. (1963). *Lectures in Statistical Mechanics*, American Mathematical Society, Providence, Rhode Island.
- Vagner, I.D., Gvozdkov, V.M., & Wyder, P. (2006). Quantum mechanics of electrons in strong magnetic field. *HIT Journal of Science and Engineering*, Vol. 3, No.1, pp. 5-55, Holon Institute of Technology.
- Zubarev, D.N. (1974). *Nonequilibrium Statistical Thermodynamics*, Consultant Bureau, New York.



# The Role and the Status of Thermodynamics in Quantum Chemistry Calculations

Llored Jean-Pierre  
CREA/Ecole Polytechnique  
France

## 1. Introduction

This chapter aims at understanding what is at stake when thermodynamics is used in chemical quantum calculations. Quantum chemistry and thermodynamics seem to be two incommensurable scientific worlds the assumptions and the statements of which are thoroughly different. So my questions are: How was thermodynamics been integrated into the chemical quantum background at the very beginning of quantum chemistry? What are its role and status in current *Density Functional Theory* and others quantum methods used in chemistry?

I refer both to history and epistemology to grasp this entanglement of scientific approaches. First of all, I propose to analyze how thermodynamics became involved in chemistry. In this respect, I will point out how the concept of energy provides the old chemical affinity with a quantitative tool to understand chemical transformations. The birth of thermochemistry aroused opposition between two old rival conceptions of matter that framed the history of chemistry, that is to say the aggregate and the 'mixt' stances<sup>1</sup>.

I will then highlight that this duality of conceptions was still at stake when Mulliken and Pauling created two different quantum chemical approaches. In this context, thermodynamics was not just used as a mere tool to calibrate methods; it also guided the contrivance of new quantum concepts or parameters from the outset. Following this line of reasoning, I will query how the concept of 'state', be it electronic or thermodynamic, allows us to bridge thermodynamics to quantum chemistry in a different way. I will indicate why and how the second law of thermodynamics is reflexively of importance to understand molecular calculations and to better grasp the relation between a molecular "whole" and its respective parts.

These investigations are widened by a global overview of the ways thermodynamic parameters are currently involved in workaday quantum methods in order to describe molecular reactivity.

To conclude, the paper will query the status of thermodynamics in predictive quantum methods. I will insist on the status of the concept of energy and the heuristic power of the second law of thermodynamics on quantum grounds.

---

<sup>1</sup> A 'mixt' is a chemical combination composed of elements but not bearing the same properties as the constitutive elements. Conversely, an aggregate is a mere additive combinations of elements and their properties.

## 2. The integration of thermodynamics into chemical grounds: From a qualitative to a quantitative *affinity*

The new rules of the French Royal Academy of sciences (1699), Wilhelm Homberg's work on the interchangeability of 'average' -now called 'neutral'- salts, the mechanist philosophy influences at the end of the seventeenth century, and as well Paracelsus and the alchemists' traditions, paved the way for the empirical production of affinity tables during the 18th century. From Etienne-François Geoffroy (1718) to Bergman (1775), these tables were multiplied; some chemists, such as Guyton de Morveau (1773), developed the first experimental devices to quantify these affinities (Mi Gyung, 2003; Partington, 1962).

A shift of the explanatory function of the principles – Aristotelian, Paracelsian, or other, which previously accounted for qualities and chemical transmutations, towards the state of union between two chemical substances and the concept of process which implies union and disunion, gradually occurred (Bensaude-Vincent & Stengers, 1996). This major epistemological upheaval led to the attraction between chemical bodies being operationally redefined within the context of salts chemistry. The key question of the force or power which governed the chemical combinations remained rather unclear and mysterious according to Henri Sainte Claire Deville (Deville, 1864) until the chemists integrated knowledge of calorific and thermodynamics into their own practices.

Using a new calorimeter with mercury, J.T. Silbermann and P.A Favre showed for the first time in 1852 that a chemical decomposition could involve a release of heat. At the same time, Julius Thomsen published a paper entitled *Les bases d'un système thermochimique*<sup>2</sup> in the Annals of Poggendorf which upset the generally accepted ideas. The differentiation between combination and decomposition defended by Claude-Louis Berthollet could not be maintained anymore. A chemical act which produces heat was said to occur spontaneously. The concept of chemical reaction understood as an observable and measurable phenomenon was thus worked out by means of mathematical equations, and new experimental practices related to an innovative thermal instrumentation. Pierre Duhem reported a sentence of Thomsen according to whom: "When the chemical combination occurs, it releases a quantity of heat proportional to the affinity of the two chemical bodies"<sup>3</sup> (Duhem, 1893). Thomsen originally argued that the heat of a reaction was the true measure of affinity (Kragh, 1984). The chemical act became a work to refer to the physicists' vocabulary but a work reinterpreted from within the current framework of chemical knowledge and laboratory practices. In 1873, Marcellin Berthelot precisely applied the Principle of maximum work to a chemical reaction (Médoire & Tachoire, 1994). He stated that in the absence of external energy, every chemical change tends towards the production of the greatest quantity of heat (Nye, 1993).

As Thermochemistry began to develop, chemists paid attention to other facts which first appeared foreign from each other. In 1852, Edmond Fremy and Henri Becquerel showed that the production of ozone was an incomplete reaction, a conclusion that Berthelot and Pan de Saint Gilles also reached for the esterification reaction ten years later. The chemical reaction appeared limited and dependent on the time factor, Sainte Claire Deville and his collaborators widened and strengthened those findings thanks to many experiments

---

<sup>2</sup> 'The foundations of a thermodynamic system', my translation.

<sup>3</sup> The French original sentence is : '*Lorsque la combinaison se produit, il se dégage une quantité de chaleur proportionnelle à l'affinité des deux corps.*'

(Daumas, 1946). After many attempts, Maximilian Güldberg and Peter Waage asserted in 1861 that they were able to "find for each element and each chemical combination, numbers which express their relative affinity"<sup>4</sup> (Güldberg & Waage, 1867). Güldberg and Waage quickly connected the emerging concept of chemical equilibrium with the notion of affinity so as to designate the chemical force which was supposed to lead to the equilibrium. They then established the crucial chemical law of mass action while studying reaction rates, and the effects of time, temperature and mass factors.

The development of the energy approach in chemistry was the result of a fortuitous combination of independent works proposed by Wilhelm Hortsman in Germany, by Josiah Willard Gibbs in America and by Bakhuis Roozeboom and J.H. Van't Hoff in Holland. Hortsman integrated Rudolf Clausius' considerations on isolated systems into chemistry. In so doing, he rediscovered in 1873 the law of mass action by means of calculation without having any idea that it had already been found on other grounds. The same year, Gibbs, published a paper entitled '*On the equilibrium of heterogeneous substances*', within which he proposed a mathematical description of chemical equilibrium. This work remained mostly unknown by chemists because they didn't have the necessary basic mathematical knowledge to grasp it. In 1882, Hermann von Helmholtz rediscovered Gibbs' results -which he totally ignored- using the theory of heat published by J. Clark Maxwell in 1871. All these publications gave rise to new chemical concepts which dealt with energy changes in a chemical system submitted to the action of the various forces that led to an equilibrium. One must have distinguished, according to Helmholtz, between the part of energy which appeared only as heat and the part which could be freely converted into other kinds of work, i.e. the "free energy". Subsequently, the production of a decrease in free energy enabled chemists to explain chemical stability (Kondepudi & Prigogine, 1998). In 1884, Pierre Duhem introduced the notion of internal thermodynamic potential by analogy with classical mechanics (Duhem, 1902).

Applications to experimental chemistry by the Dutch school, for example, Roozeboom had to cope with difficulties in interpreting hydrobromic acid decomposition in the presence of water in the gas phase. His colleague physicist J.D. Van der Waals suggested to him to use Gibbs's work and helped him to put forward the so-called *phases rule*. Van't Hoff established the law of equilibrium variation depending on temperature and gave to the measure of the affinity as the expression of the maximum work that the system must be able to provide under defined conditions. According to Van't Hoff, affinity was the leading force which produced chemical transformation. The change of affinity sign accompanied the change in the direction of the reaction which occurred at the transition point (Kragh & Weininger, 1996). From that time onwards, researchers gradually moved their attention to other factors of equilibrium. In 1888, Henry Le Chatelier proposed a way to predict how a chemical equilibrium moved according to the variation of the factors on which it depended. Chemical affinity became therefore one of the many aspects of the chemical act allowing improved forecasts and performances.

At the beginning of the twentieth century, chemists attempted to know not loner why, but how matter is transformed. Chemical kinetics studied the process of transformation of matter. Swante Arrhenius introduced the concept of energy activation, researches gradually

---

<sup>4</sup> The French original sentence is : " (...) trouver pour chaque élément et pour chaque combinaison chimique, des nombres qui expriment leur affinité relative"

turned to focus on the question of the energy transfer and the direction of collisions between chemical bodies. Wilhelm Ostwald succeeded in describing chemical equilibrium without making any reference to atoms (Ostwald, 1919). Two antagonistic approaches of matter were at stake. Thermochemistry revolved around energy and denied any reality to atoms whereas chemical kinetics was based on the atomic assumption. Thomsen, for instance, used structural theory to assign heats of formation to specific bond types found in organic molecules. In this respect, he tried to reduce chemical properties to a mere juxtaposition of atomic properties. Others, like F.W. Clarke tried to connect the heat of formation with the one and only number of atomic linkages within the molecule. By doing so, he tried to connect valence with affinity (Weininger, 2001). All the attempts that tried to understand affinity thanks to additive and reductive descriptions failed.

To sum up this first part, I would like to emphasize that the integration of thermodynamics within the frameworks of chemistry was made possible because chemists were looking for a quantitative measure of affinity. The way thermodynamics became thermochemistry depended on the instrumentation and the practices that chemists contrived to tackle the challenge of affinity. As the philosopher Joseph Rouse points out: 'Practices are not just pattern of action, but the meaningful configurations of the world within which actions can take place intelligibly, and thus practices incorporate the objects that they are enacted with and on and the settings in which they are enacted'. (Rouse, 1996, p.135). Thermodynamics was thus integrated into chemical projects and then transformed by such integration because it made chemists goals achievable and intelligible within such new practical backgrounds.

I suggest we should take more distance and consider the whole history of chemistry to analyze the way this integration actually took place. Let us widen the circle to grasp what is at stake behind this integration and how the duel between different conceptions of matter will remain active at the very beginning of quantum chemistry. This study will enable us to understand the role of thermodynamics in the first chemical quantum calculations.

### **3. The integration of thermodynamics into first quantum methods: The reviving of the aggregate/'mixt' duel**

#### **3.1 Two conceptions of matter and the thermodynamics embodiment within chemical practices**

First and foremost, I would like to develop the opposition of conceptions of matter we previously stressed. Duhem's claim for an energy description of molecules that need not rely on any atomic assumption reminds us of other historical oppositions.

In the seventeenth century for instance, Nicolas Lemery in his famous *Cours de Chymie*, tried to account for chemical transformations by means of a multitude of corpuscles with different forms. Gabriel-François Venel argued that this reductive approach was unable to explain and predict chemical properties. Venel asserted that chemists studied 'mixt' whereas mere 'aggregates' came under mechanics. Venel used Georg Ernest Stahl's distinction between an aggregate which was defined as a mere sum of various substances that continued to exist in the whole compound, and a 'mixt' within which reactants disappeared to form an emergent new whole with specific properties. Two conceptions of matter were at odds in this context and became progressively more important within the debate. On the one hand, mechanics considered matter to be homogeneous, without qualities and necessarily informed by something from outside. This kind of matter representation solely

described by its form and motion could not account for the world of chemical activities and diversity according to Venel. On the other hand, most chemists considered matter to be heterogeneous and able to act and react (Bensaude-Vincent & Simon, 2008). More often than not, chemists pragmatically used one description or the other according to their laboratory goals. As Bensaude-Vincent and Simon write: 'We prefer to see this duel between the two approaches as a characteristic feature of the history of chemistry. Chemists have always been confronted with this interpretative dichotomy, and, depending on the period, they have opted for a version of atomism or an elementary approach, or else have tried to reconcile the two.' (Bensaude-Vincent & Simon, 2008, p. 128).

Not only did thermodynamics enable chemists to construe a quantitative version of affinity but it also fitted very well into the cultural background that had been framing chemists' activities for a long time. Thermodynamics embodiment within chemical practices was thus at least twofold; it provided chemists with quantitative tools for understanding chemical reaction while recasting old oppositions of matter representations. Along with this perspective, thermodynamics could easily be integrated into the usual chemical way of thinking about matter while reconfiguring it. As Rouse claims (1996, p.157): 'In order to understand how scientific knowledge is situated within practices, we need to take account of how practices are connected to one another, for knowledge will be established only through these interconnections. Scientific knowing is not located in some privileged type of practice, whether it be experimental manipulation, theoretical modeling, or reasoning from evidence, but in the ways these practices and others become intelligible together.'

Duhem focused his work on the dichotomy between the 'mixt' and the aggregate referring to Aristotle's philosophy (Needham, 1996). Like Sainte-Claire Deville and Berthelot, but not because of the same positivist reasons, he rejected atomism then deeply rooted in structural organic chemistry. According to the structural molecular paradigm, the physical arrangement of the constituent elements accounted for the properties of the whole compound. Since Lavoisier, chemists have been explaining the properties of compounds by reference to the nature, the proportion and, more recently, the bonds of its constitutive parts, be they atoms or elements: a logic that runs from simple to complex frameworks in post-Lavoisian chemistry (Bensaude-Vincent & Simon, 2008). Conversely, the holistic energy approach used compounds to explain the properties of the elements. In this respect, atomism had a weak explanatory power because it could not completely illuminate chemical processes. According to Duhem, chemical formula could make chemists believe that substances remained unchanged when they entered into combinations whereas they only existed potentially within them (Duhem, 1902). Joseph Earley has recently proposed an argument on the same lines. He uses the example of sea water in which salt and water cease to exist in their actual states—because for instance of solvation— but they can be reproduced by distillation (Earley, 2007). When the 'mixt' ceases to exist, it is made to reproduce its separate constituents as Venel might have asserted. In this respect, water and salt potentially exist in sea water but do not actually exist within it. Duhem then undertook to retranslate Aristotle's concept of power into that of the thermodynamic potential (Duhem, 1902). Measurable properties and mathematics allowed him to describe chemical reaction within the context of thermochemistry.

Duhem rejected both the idea of valence taken as an intrinsic atomic property and the concept of atomicity. According to him, the whole components could only give rise to valence information but not the contrary. The opposition between a holistic approach of

chemical bodies on the one hand and the aggregative atomic description on the other hand will appear of primary importance at the very beginning of quantum chemistry. I propose to study how Linus Pauling and Robert Sanderson Mulliken created the first chemical quantum approaches in the context described before and how they integrated thermodynamics and quantum mechanics into chemistry.

### 3.2 The ‘mixt’ and the aggregate: A framework for the embodiment of thermodynamics into quantum chemistry?

Both standardization and precision were required if thermodynamic bond measurements were to play a significant role in calibrating innovative methods and stabilizing new theories about affinity as well as about valence or the chemical bond (Servos, 1990). The Russian-Polish Wojciech Swietolawski played a leading role in this challenge (Médard & Tachoire, 1994). His work provided chemists with more accurate average bond energies that legitimized heat of reactions calculations. Weininger clearly shows how those thermodynamic data made researchers get to grips with valence within the atomist conception. He points out for instance how Morris Kharash used the Niels Bohr’s orbit model to propose a physical picture of thermodynamic quantities. This heuristic approach validated by Swientoslowski’s data enabled him to derive heats of combustion for hydrocarbons in quite good agreement with experiment (Weininger, 2001). But it was Linus Pauling who succeeded in bridging valence, atomic theory and thermochemistry.

Pauling’s work constitutively entangled thermodynamics with the Pauli Exclusion Principle, Heisenberg and Dirac’s approach of resonance, structural chemistry and Born’s probabilistic description (Pauling, 1928). We should bear in mind that he was first trained as a crystallographer to understand the way he shaped his experimental and theoretical crowded network that was the *Valence Bond Theory*. The use of both accurate thermodynamic and crystallographic data enabled Pauling to notice that the covalent radii sum of the bonded atoms approximated bond lengths very well. He then linked bond energies with experimental heats of formation of gaseous molecules (Pauling, 1932). The key step was to choose a set of molecules that could supply the data necessary for extracting those bond energies (Weininger, 2001). This approach allowed him to express the total energy of formation of the molecule as a mere sum of energy terms characteristic of the different bonds assuming that the molecule was obtained from separate atoms (Pauling, 1932). The referent molecules only had to have a single Lewis electronic structure (Pauling & Sherman, 1933a, 1933b). Atoms are the basic units of Pauling’s system, this atomic standpoint shaped the way he used thermodynamic data.

To understand Pauling’s molecular description, one needs: (1) to connect the molecular structure to its constitutive atoms; (2) to study how those atoms interact from within the molecule. This model retains the integrity of the atoms inside the molecule, a molecule is considered as an *aggregate* of atoms. Each atom has stable atomic orbits - 2s, 2p for instance - that will be used to form stable bonds inside a molecule or to induce ad hoc directed valence (Pauling, 1931; Slater, 1931). He stated that bonds resulted from the overlapping of two atomic eigenfunctions, the larger the overlap is, the stronger the bond gets.

The study of diatomic molecule enabled Pauling to propose the concept of ‘normal’ covalent bond and to express what he called the ‘normal’ covalent molecular wave function as a mere sum of covalent and ionic terms so as to provide his electronegativity concept with a quantum counterpart (Pauling, 1932). Thermochemistry was once again a touchstone for the

validity of this quantum mechanical treatment of chemical bonding; it was as not just a mere tool to calibrate methods. Empirical data really aroused Pauling's creativity and guided him to adapt his quantum work. By applying the rules for the electron-pair bond, Pauling removed the apparent incompatibility between chemistry and quantum theory (Gavroglu & Simões, 1994). Pauling answered more directly the concerns of the chemists by stressing the three-dimensional structure of molecules, the electrons being the bonding officers of the atoms. The valence bond approach which he developed with Slater was more quickly acknowledged by chemists because resonance corresponded to their usual representations and structural formula (Llored & Bitbol, 2010).

Mulliken proposed a very different quantum approach based on molecular spectroscopy. With regard to the concept of valence considered as an intrinsic property of the atom, Mulliken opposed the notion of 'energy state' deduced from molecular spectra on the basis of an *electronic configuration*, *i.e.*, of a distribution of the molecular electrons in different orbits. In this description, each orbit is delocalized over all the nuclei and can contribute, depending on each specific case, a stabilizing or destabilizing energy contribution to the total energy of the molecule (Llored, 2010). The sum of the energy contributions of each electron in its orbit determined whether the electronic configuration allowed for the existence of a stable molecule, *i.e.*, whether its energy was stabilizing overall. For Mulliken, the atom did not exist as a component in a molecule. His concept of molecular state suggested molecular variability of energy and geometry that could not even be considered within the approaches of Lewis and Irving Langmuir. Mulliken proved that the spectral states of the molecules could be obtained from that of their molecular ions by the mere addition of an electron without changing the quantum numbers and, thus, worked out his molecular *Aufbauprinzip* (Llored, 2010). This close connection between the quantum theory and the spectral studies gave birth to the correlation diagrams of 1932 (Mulliken, 1932b). Those diagrams made it possible to consider the degree of likeness between a molecule and its separated atoms or its united atom - a fictitious atom obtained by the coalescence of the two atoms such as helium He for two hydrogen H atoms - thanks, in particular, to empirical knowledge of the inter-nuclear distances, energy dissociation and of the charges of the nuclei. The molecule from then on was considered as a composite, *i.e.*, a new entity rather than a mere aggregate of individualized atoms. He wrote: 'In the 'molecular' point of view advanced here, the existence of the molecule as a distinct individual built up of nuclei and electrons is emphasized, whereas according to the usual atomic point of view the molecule is regarded as composed of atoms or of ions held together by valence bonds. From the molecular point of view, it is a matter of secondary importance to determine through what intermediate mechanism (union of atoms or ions) the finished molecule is most conveniently reached. It is really not necessary to think of valence bonds as existing in the molecule (Mulliken, 1931). Despite their irreducible differences, Duhem's thermodynamic potential echoed the electronic states developed by Mulliken insofar as both considered a molecule from an energy standpoint as a 'mixt' not as an 'aggregate'. The '*electronic state*', the '*binding capacity*', the '*promotion*' of an electron, '*the energy-bonding-power*', are among the many concepts Mulliken built to explain the capacity of the electrons to be linked to nuclei to form a molecule seen as a whole (Harré & Llored, 2011).

The semantic shift from the concept of molecular *orbit* to that of molecular *orbital* -MO- occurred in 1932. The concept of orbital took all its significance from Max Born's probabilistic interpretation that the square of a molecular orbital corresponded to the

probability density of finding this electron at a certain location in space. Mulliken wrote: 'By an atomic orbital is meant an orbital corresponding to the motion of an electron in the field of a single nucleus plus other electrons, while a molecular orbital corresponds to the motion of an electron in the field of two or more nuclei plus other electrons. Both atomic and molecular orbitals may be thought of as defined in accordance with the Hartree method of the self-consistent field, in order to allow so far as possible for the effects of other electrons than the one whose orbital is under consideration.' (Mulliken, 1932a).

At the very beginning of his investigations, Mulliken mainly used molecular spectroscopy data. He seldom referred to thermochemistry except for necessary calibration requirements. It is important to notice nevertheless that thermodynamics was influential when he envisaged the study of larger molecules by using group theory. I think it is important not only to check if his holistic molecular conception changed the way thermodynamics became involved in chemical quantum works; but also to compare it to Pauling's own use of thermal data.

Mulliken's studies of hyperconjugation are a relevant case study to grasp the role and the status of thermodynamics in such a chemical quantum background (Mulliken et al., 1941). Mulliken's calculations taken in connection with thermal and bond distance data indicated the conjugating power of chemical groups such as the landmark methyl group. With respect to strength and stability, he could then label the single or the multiple bonds of a conjugated system as acceptor and donor bonds, respectively. The thermal data allowed him to postulate that the hyperconjugation energy of saturated hydrocarbons was to a good approximation a function only of the numbers of different types of bonds. Using localized and non-localized molecular orbitals, he described the conjugation or resonance energy as the energy of delocalisation. In order to approximate quantitative calculations, he wrote the molecular orbital as a Linear Combination of Atomic Orbitals -LCAO- within the Hartree-Fock self-consistent field approach -labelled LCAO MO SCF-.

Unlike Pauling, he systematically used heats of combustion rather than bond energies referring to Karash and W.G. Brown's corrected tables mainly construed by using hydrogenation heats data. Mulliken and al. wrote: 'Our procedure for deriving conjugation energy from thermal data is similar to that of Pauling and Sherman who, assuming additivity of bond energies (with corrections for special groups), compute energies of formation and interpret deviations therefrom as resonance energies. However, we shall work with heats of combustion.' (Mulliken et al., 1941).

Heats of combustion enabled Mulliken to put forward formula to calculate conjugation energies from heats of combustion that fitted the available consistent data for gaseous saturated hydrocarbons - except methane - with considerable accuracy - mostly better than 1 kcal. The current practice of research then involved a rich set of corrections within which quantum formalism, approximations, chemical knowledge and thermochemistry were deeply intertwined in order to create a stabilized composite knowledge of conjugation energy for particular types of molecules. For instance, Mulliken tailored Lennard-Jones's curves to make them fit the empirical data, he then determined wave function coefficients by defining and substituting new parameters in the secular determinant, and finally extracted from the computed conjugation energies some energy quantities - the third-order conjugation energy - to make a direct comparison with observed conjugation energy. By trial and error, a host of other corrections and readjustments enabled him to determine the total conjugation energy and to compare it to thermodynamic outcomes. Mulliken and al.



wrote (p.56): 'Perhaps the most uncertain feature of our analysis is the derivation from thermal data. (...). Our empirical parameters, our bond order curve, and our numerical conclusions would then be so strongly altered, since they are decidedly sensitive to variations in the empirical conjugation energies to which they are fitted. Nevertheless, their self-consistency gives a distinct support to our numerical results, since we have found that such self-consistency is not easy to attain.' (Mulliken et al., 1941). The authors called for more accurate thermal and bond distances data, those researches got into an endless and open circle of refinements that linked calculations with empirical data. It is of importance to notice that this work led the authors to provide Hückel's resonance parameter ' $\beta$ ' with a new interpretation that allowed a more satisfactory understanding of energy interactions within unsaturated molecules. This theoretical accommodation was then confirmed by spectroscopic data. Thermodynamics not only took part in a motley complex of scientific practices that made it possible for a quantum chemist to calculate molecular properties and to predict chemical reactivity, but it also partly altered the meaning of the theoretical quantum background. I wish to emphasize that thermodynamics was not a mere tool for calibrating a semi-empirical method but a constitutive active part of a techno scientific network that Mulliken and others shaped to study a molecule understood as a 'mixt'.

In addition to this conclusion, there are other interesting facts we should take a look at. Mulliken and Parr studied the decrease in ' $\pi$ ' electron energy for the change from a Kekulé to a proper benzene structure by using a *completely theoretical method* (Mulliken & Parr, 1951). In order to make a comparison with the ordinary empirical resonance energy, they had to make several corrections that involved: (1) the 'compressive energy' needed to adjust the lengths of the single and double Kekulé's bonds to those of the proper benzene; (2) hyperconjugation and related effects. They discussed the corrections and estimated their magnitudes before concluding that a reliable value could only be obtained for the compression energy. Following this line of reasoning, they determined that the computer net resonance energy was 36.5 kcal. This outcome agreed, with the uncertainties due to the omitted correction terms, with the value 41.8 kcal of the empirical resonance energy ' $\Delta$ ' based on thermodynamic data. They then used ' $\Delta$ ' as the point of departure of the calculation of the actual heat of formation ' $\Delta H_f^\circ$ ' of benzene from the value given by a standard formula for nonresonating hydrocarbons. They proposed a new standard formula containing corrections for the mutual effects of neighboring carbon-carbon bonds while discussing its significance. This analysis allowed them to clarify what was meant by 'resonance energy' and to query the significance of 'nonresonating' structures and repulsion terms in their own theory. They always sought to identify the conditions that made it possible for a chemist to make a clean-cut comparison between theory and experiment. In quite that light, thermodynamic data guided the way they wrote equations relating theoretical energy quantities to a sum of empirically based terms. This work allowed them to define new useful concepts such as 'standard hydrocarbon' - held with  $\Delta = 0$  kcal - that fostered calculations and comparisons. To sum up, they continually queried their model and its meaning. Thermochemistry, quantum chemical methods, chemical practices and culture, computers, instruments were *constitutively intertwined*, and they were *interactively stabilized*. Modelling is an *open-ended process* that includes thermochemistry as a foundation to create a new quantum account of a molecular 'mixt'. As Andrew Pickering asserts: 'Existing culture constitutes the surface of emergence for the intentional structure of scientific practice, and such practices consists in the reciprocal tuning of human and material, tuning that can itself

reconfigure human intentions. The upshot is, on occasion, the reconfiguration and extension of scientific culture.’ (Pickering, 1995). The *dialectics of resistances and accommodations* between thermochemistry and the quantum chemical model made Mulliken continuously recast his approach so as to stabilize a great amount of tables and concepts about molecular properties. He produced a great number of tables throughout his academic life. From spectroscopic to conjugation energy tables as well as from correlation diagrams to Mulliken-Walsh ones, he knitted a network of data thanks to a constitutive interaction between theory and experiment.

I claim that this difference of practice from Pauling to Mulliken was in a way a consequence of the two conceptual schemes at stake. On the one hand, the aggregative Pauling’s approach focused on a reified chemical bond that resulted in valence electrons share. Pauling was indeed interested by the formation energy of a molecule from its parts. On the other hand, Mulliken used chemical reaction combustion data because he considered the way the ‘whole’ molecule reacted and released energy by thermal transfer in the presence of other chemical reactants and their surroundings. Pauling’s bottom-up analysis collapsed Mulliken’s holistic way of thinking. I think that my statement is to be qualified insofar as we should wonder if pragmatic reasons were also at stake concerning this choice of data. Heats of combustion corrected tables probably were more useful for Mulliken than others.

At that time, chemical affinity turned out to play no role in the integration of thermodynamics into quantum methods simply because researchers’ presumptions did not consider it as a challenge to face anymore. On the contrary, the duality of the two conceptions of matter were still at work and underpinned the way Mulliken and Pauling were using thermochemistry while doing quantum chemistry. So I emphasize that the way thermodynamics became involved in quantum chemistry partly depended on different human stories and skills -Pauling was first a chemist and crystallographer whereas Mulliken was trained as a chemist and a spectroscopist. Others were mathematicians, organic chemists, and so on. But it also depended on different representations of matter – the aggregate and the ‘mixt’. Practices of research, human skills and goals, human and non human agency, time, concepts and representations interactively took part in the integration of thermodynamics into the earlier quantum realm.

Before I move on to modern quantum chemistry, I would like to further examine the relation between earlier quantum methods and thermodynamics by querying the concept of ‘state’, be it electronic, quantum or thermodynamic.

### **3.3 The concept of ‘state’ and the relation between quantum chemical methods and thermodynamics**

Quantum chemistry is the result of a deep entanglement of scientific and human practices within which thermodynamics was an active generator of concepts and a tool for method calibration. If we want to query the role and status of thermodynamics in quantum chemistry, it is necessary to consider the practices of research from which they originate, *i.e.*, the techno-scientific closure which combines quantum mechanics, approximations, instrumental and algorithmic techniques, chemical know-how, and the use of Principles which do not belong to quantum theory such as the Pauli Principle. The predictive capacity of these chemical quantum approaches does not only rely on the molecular wave function but also on a host of approximations and compromises that make it possible for numerical properties and molecular landscapes to be calculated (Llored, 2010, 2012).

It is of interest to point out that quantum formalism gives rise to miscellaneous chemical quantum approaches depending on both chemical cultural resources and practical scientific backgrounds. It is astonishing however to notice that an atomic approach such as that of Pauling could have successfully developed on quantum grounds. The notion of atomic parts within a molecule is indeed deprived of meaning in quantum mechanics. The holistic approach of Mulliken seems much more understandable in a holistic, contextual and non-representationalist quantum theory. The final results reached by those methods are not pure quantum physics applications. This is a crucial point to bear in mind.

Let us deepen our study of Mulliken's molecular orbital framework to illuminate his fine-grained relation with thermodynamics. Mulliken first worked on the couplings between orbital kinetic moments and of spin suggested by Friedrich Hund. In 1927, Hund developed an approach radically different from the work developed by Walter Heitler and Fritz London and generalized the study of Oyvind Burrau to diatomic molecules. Rather than built a molecular wave function from those describing isolated atoms, he proposed to describe each electron in the total molecular electric field of the nuclei and other electrons. Hund focused on the evolution of electronic energy during the transfer of an orbit around the joined nuclei to an orbit around the separate atoms isolated from each other. On the basis of works developed by Erwin Schrödinger, Pascual Jordan and Max Born, Hund was able to describe the exact stationary states of the two subsystems knowing those of the system by using linear combination. He wrote: 'We investigate a system with one degree of freedom as an analogous for a molecule with several atoms, using quantum mechanics. Its potential energy has several minima. We can relate the stationary states of such a system to those of partial systems that result when the separation between the minima becomes infinite or when the potential energy separating them becomes infinite. In agreement with this (and in opposition to the classical theory) we obtain an adiabatic relation between the states of two separated atoms or ions, the states of a two-atomic molecule and the states of the atom that would result when the nuclei are united. This relation allows for a qualitatively valid term system of the molecule and for an explanation of the terms 'polar molecule' and 'ion lattice'.' (Hund, 1927). The new quantum theory thus allowed him to explain the adiabatic passage between two stationary states of the same system. Hund made this result suitable for the study of molecules and proposed an interpolation between the quantum states of the isolated atoms, the united atom and the molecule. Hund further added: 'The complete transition from the case of nuclei separated by a large distance to the case of a small separation cannot be done adiabatically in the classical model. If we start in the case of nuclei separated by a large distance with some given quantum numbers, then we first arrive at orbit type II, but for a certain internuclear distance this type is no longer possible. The classical motion becomes a limiting motion. The same occurs when we approach from the other side, with nuclei placed close together; for a certain distance between the nuclei, orbit type I becomes impossible and the motion becomes a limit. An adiabatic transition going over the limiting case is not possible because of the vanishing frequency.' (Hund, 1927).

Within the framework of thermodynamics, a system is involved in an adiabatic process if it does not exchange any thermal energy - any heat - with the outside. It can exchange only work. In mechanics, an adiabatic process is characterized by the fact that within infinitely slow changes of external parameters, the system evolves through successive states of equilibrium. In this kind of process, some quantities remain invariant, physicists call them

adiabatic invariants. The adiabatic hypothesis, which was originally developed by Paul Ehrenfest, considers that the quantum conditions must always be such that the adiabatic invariants of classical mechanics are equal to an integer multiple of the quantum of action. You can infer the values of the states of a system from quantum states of another system that can be reached by an adiabatic transformation. The difficulty related to the conservation of quantities when changing orbits, evoked by Hund, disappears when the problem is studied within the framework of quantum theory. We realize that beyond semantic diversity of words such as 'state' or 'adiabatic', what is at stake is the way quantum physics can encompass classical physics as a limited case in precise contexts. Researchers were inventing a new quantum chemical scheme, while using general scientific and linguistic devices to link it with different previous theories. The notion of 'state' related to that of the 'equilibrium state' involved in thermodynamics is not tantamount to that of a 'quantum state' that only provides scientists with the calculation of the probability of each set of 'observables' from within a precise experiment context (Bitbol, 1998). The quantum state is related to a predictive symbolism that enables scientists to study holistic systems constitutively entangled with apparatus, that is to say the study of which cannot be separated from the context of measurement. Thermodynamics and quantum chemistry are nevertheless holistic, the former is descriptive at a macroscopic level, the latter is predictive at a microscopic one. In this respect, it is not surprising that scientists tried and try to bridge those approaches in what we call different levels of our universe. What may the link between the two levels be? What are the necessary pre-conditions for tuning them? What may be the link between an energy quantum study of a molecule understood as a 'whole' at a microscopic level, and the energy of a set of molecules at a level described by thermodynamics?

Dealing with relations between a molecule and its parts, G.K. Vemulapalli noticed that: 'While properties of the whole are not the sums or products of the properties of parts, the states of the system can be obtained by adding the states of parts. Because properties in turn can be derived from the states, it appears that we have shown that properties of wholes are completely determined by parts. But there are two problems here. (1) It is true that the states of the system are composed of states of the parts, but there are also weighting factors in the composition. There are the constants  $\lambda$  in the linear combination. What factors determine these constants? (2) Just as in the molecular wave function, an atomic wave function may also be represented by a sum of an arbitrary set of functions. Thus one may claim that an atomic function is a linear combination of molecular functions or atomic states (parts) reduced to molecular states (wholes!).' (Vemulapalli, 2003). If we set apart that the notion of properties is open to criticism in quantum contexts and the linguistic traps related to it, the author's insight is relevant to query the interrelation between levels of description studied by quantum chemistry.

The arbitrary character of the relation between the whole and its parts is highlighted. It remains more than ever present in current semi-empirical or *ab initio* methods of molecular orbital calculation that depend on the choice of atomic or molecular orbital used. Mulliken developed the *fragment method* in 1933, two fragments could interact provided they had the same kind of symmetry and that the energy gap, measured by spectroscopy, was not too high. For the ethylene molecule ' $C_2H_4$ ', Mulliken considered two fragments ' $CH_2$ ' and determined a suitable molecular orbital by using the irreducible representations of ethylene. He could thus propose a representation of a molecular orbital of ethylene by increasing

order of energy as well as its correlation diagram thanks to those of the two fragments. In doing so, he included all the characteristics of the molecular orbital diagram of the ethylene molecule and checked it using molecular spectroscopy. Mulliken could just as easily have considered a fragment " $C_2$ " and another " $H_4$ " of adapted symmetries. The relation on whole " $C_2H_2$ " with its parts was of secondary interest. The fundamental choice relates to the nature and the extent of the basis sets of the calculation. Vemulapalli threw light on the role of the weighting coefficients appearing in front of the orbitals of the key basis sets. These coefficients determined by the Variational Principle are those which minimize the molecular potential energy. How is this minimum of energy justified? What can explain the use of the Variational Principle? A quantum principle?

Vemulapalli referred to the second law of thermodynamics to explain why the studied molecular system continuously eliminates its excess energy by interactions with its environment. An energy transformation into local entropy returns legitimates the use of the Variational Principle. Vemulapalli added: 'Thus we are led to conclude that it doesn't matter what the states of the parts are, but it does matter the surroundings soak up the excess energy of the molecule, increasing entropy, and make the molecule settle down into the lowest energy state. It is that part of the universe coupled to the system, and the varieties of interactions between the system (molecules) and the surroundings that determines the structure of the molecule. Holism thus appears as the root of the apparent reduction of properties of a molecule to its parts through coupling states. We are able to follow a reductionist program in calculating molecular properties, but what we are able to do is a gift of holism.' (Vemulapalli, 2003). A molecule is always in relation with its surroundings, it can at least emit a photon even in a strong vacuum. So the study at a molecular level requires a study of interactions at an upper level while microscopic descriptions require quantum predictions. Levels of description need one another, they are co-stabilized. The Variational Principle that underpins Mulliken's work at a molecular level can find a justification within the context of thermodynamics. It is an *a posteriori* analysis that allows us to widen our understanding of the possible links between thermodynamics and quantum chemistry from another point of view, that of inter-levels relations.

To sum up, we have focused our work on the way thermodynamics was used from within the earlier quantum chemical methods. We have shown that the opposition between the 'aggregate' and the 'mixt' was still at stake when explaining the integration of thermodynamics into quantum chemistry. Taking distance from linguistic traps concerning words such as 'state' or 'adiabatic', and by reflecting upon the relations between the levels of scientific description - a molecule to its alleged constitutive atoms or the macroscopic and microscopic scales -, we confirm that epistemology can provide us with another kind of understanding of the interrelations between thermodynamics and quantum chemistry. I would like to turn now to modern quantum methods and to examine how they involved thermodynamics. I choose to develop the example of the density functional theory - DFT - which has been widely used for twenty years in research laboratories.

#### 4. The role and status of thermodynamics in modern quantum chemistry

Kohn-Sham density functional theory has become one of the most popular tools in electronic-structure theory due to its excellent performance-cost ratio as compared with correlated wave function theory, WFT. Within this theory, the molecular space is divided

into grids of cubes; researchers define an electronic density for each point of this space. It is a holistic approach that enables quantum chemists to calculate molecular geometry or total energy exhaustively thanks to its electronic density – ' $\rho(\mathbf{r})$ ' -, provided that its Ground-State is not degenerate. The total energy is in consequence a *functional* of the electronic density that is to say a function the basic variable of which is the electronic density function (Kohn et al., 1996). Several authors have applied the Variational Principle to the total energy with the purpose of determining the exact electronic density that minimizes it. Approximations are required because the exact electronic density cannot be reached. The accuracy of a DFT calculation depends upon the quality of the exchange–correlation – XC - functional. This functional is used to account for the exchange-correlation energy term  $-E_{XC}$ . This energy contains not only the non-classical effects of self-interaction, exchange and correlation, which are contributions to the potential energy of the system, but also a portion belonging to the kinetic energy. The past two decades have seen remarkable progress in the development and validation of XC density functionals.

The first generation of functionals is called the local spin density approximation – LSDA -, in which density functionals depend only on local spin densities. Although LSDA gives accurate predictions for solid-state physics, it is not a useful model for chemistry due to its severe overbinding of chemical bonds and underestimation of barrier heights. The second generation of density functionals is called the generalized gradient approximation – GGA -, in which functionals depend both on the electronic density and its gradient. GGA functionals have been shown to give more accurate predictions for thermochemistry than LSDA ones, but they still underestimate barrier heights (Truhlar & Zhao, 2008a). In third-generation functionals, a Laplacian term density is added in the functional form; such functionals are called meta-GGAs. LSDAs, GGAs, and meta-GGAs are “local” functionals because the electronic energy density at a single spatial point depends only on the behavior of the electronic density and kinetic energy at and near that point. Local functionals can be mixed with nonlocal Hartree–Fock – HF - exchange as justified by the adiabatic connection theory (Becke, 1993). Functionals containing HF exchange are usually called hybrid functionals, and they are often more accurate than local functionals for main group thermochemistry (Truhlar & Zhao, 2008a, 2008b). This field of research aims at creating new density functionals with broader applicability to chemistry by including, for instance, non-covalent interactions. The crucial step is the calibration of new functionals against benchmark databases or best theoretical estimates (Goerigk & Grimme, 2010). Let us consider a case study developed by Truhlar and Zhao in order to understand the role and the status of thermochemistry in such a current context.

The most popular density functional, 'B3LYP', an hybrid GGA, has some serious shortcomings among which is its underestimation of barrier heights by an average of 4.4 kcal/mol for a database of 76 barrier heights. This underestimation is usually ascribed to the self-interaction error (unphysical interaction of an electron with itself) in local DFT (Truhlar & Zhao, 2008a). Truhlar and Zhao change parameters and include new ones while shaping a new mathematical functional form that takes physical phenomena into account. In so doing, they design a new functional by trial and error. They then use databases to appraise the reliability of a new functional within a defined purpose. Two databases gather all the thermodynamic quantities: (1) the data base 'TC177' is a composite database consisting of 177 data for main-group thermochemistry including atomization energies, ionization potentials, electron affinities, proton affinities of conjugated polyenes, and hydrocarbon

thermochemistry among others data; (2) 'DBH76' is database of 76 diverse barrier heights concerning for instance nucleophilic substitution and hydrogen transfer. Truhlar and Zhao then discuss the performance of new functionals for these databases, they conclude that functionals labeled 'MO6-2X' and 'MO5-2X' are the '*best performers*' for the main-group thermochemistry and barrier heights. They propose cases study to exemplify their statement. The isomerization energy of octane involves stereoelectronic effects; none of the previous functionals gives the right sign for the isomerization energy from 2,2,3,3-tetramethylbutane to *n*-octane. The functional 'B3LYP' gives an error of 10 kcal/mol while 'M05-2X' predicts the right sign because this later allows a better description of medium-range XC energies, which are manifested here as attractive components of the non-covalent interaction of geminal methyl and methylene groups (Truhlar & Zhao, 2008a). On the basis of 496 data in 32 databases, they recommend different '*best functionals*' designed to transition metal thermochemistry, main-group thermochemistry, kinetics, non-covalent interactions.

Choosing a functional of electron density depends upon: (1) the necessary accuracy; (2) the chemical system; (3) the time of calculation. It also requires choosing a set of functions called a basis to achieve calculations for each atom. The basis change according to the type of atoms and different effects such as diffusion, polarization, pseudo potentials for core electrons, and the size of functions -double, triple zeta-. The functional and its relative basis set define a level of calculation, the process of which requires choosing a computer program such as Gaussian type or Turbomole to be processed. If calculations are not convergent, researchers can change the functional, the size of the grids and convergence thresholds in order to optimize geometry or to calculate molecular energy. Each step reveals know-how, chemical culture and pragmatic compromises. Notwithstanding their basic differences, the ways thermochemistry is involved within molecular orbital approximation or DFT approach are quite similar. Modeling includes thermochemistry as a tool for calibration but also as a heuristic guide for theoretical parameters adjustments inside functionals or wavefunctions or for the design of new quantum methods (Grimme et al., 2007). The structure, within which calculations are made, is well framed by the Variational Principle. We thus realize that thermodynamic quantities partly shape current quantum practices of optimization of geometry and calibration. Calculations help researchers to find out the energy surface associated with a particular chemical reaction. The knowledge of the minimum points on an energy surface makes it possible for a chemist to interpret thermodynamic data. Besides, thermodynamics can retroactively justify minimization of energy as we have already explained. Thermodynamics and energy surface are thus interconnected to determine transition structure and reaction pathways. Modelling structural configurations is of importance in this context and the quantum calculations of entropy play a leading role in such descriptions and predictions.

Before I conclude, I would like to focus on a last case study to widen and deepen my enquiry. Let us consider how thermodynamic quantities are used to model solvation effects and to scrutinize a chemical reaction mechanism within the DFT calculation background. I will refer to a study about zinc-thiolate complexes reactivity depending on the zinc ligands (Picot et al., 2008). Some calculations are shaped by thermodynamic quantities especially designed for quantum context, that is to say that do not exist in classic thermodynamics. It is typically the case of the zero-point vibrational energy labeled 'ZPVE'. The molecular vibration energy is not equal to zero at absolute zero -0 K-, it is a quantum

mechanical effect which is a consequence of the Uncertainty Principle. Once a stationary point is localized, be it an energy minimum or a transition state, its energy turns out to be less important than the experimental energy of the molecule. For comparison with experimentally obtained thermochemical data, zero-point vibrational energy is required to convert total electronic energies obtained from *ab initio* quantum mechanical studies into 0 K enthalpies. The currently accepted practice is to employ self-consistent-field harmonic frequencies that have been scaled to reproduce experimentally observed fundamental frequencies (Grev et al., 1991). This procedure introduces systematic errors that result from a recognizable flaw in the method, namely that the correct ZPVE -G (0)- is not one half the sums of the fundamental vibrational frequencies. The use of scaling factors is therefore required (Grev et al., 1991); they depend upon the level of description and its computer data processing. It is then possible to calculate other thermodynamic quantities related to a chemical reaction such as the gas phase Gibbs's free energy from the equation:

$$\Delta G_{\text{gas}} = \Delta E_{\text{elec}} + \Delta \text{ZPVE} + \Delta E_{\text{T}} - T\Delta S$$

$\Delta E_{\text{elec}}$ ,  $\Delta \text{ZPVE}$ ,  $\Delta E_{\text{T}}$  and  $\Delta S$  are the differences of electronic energy, zero-point vibrational energy, thermal energy and entropy between the products and the reactants, respectively (Picot et al., 2008).

The solvation free energy of each compound is determined by calculations depending on a model. This quantity is always defined as the required amount of energy necessary to transfer a molecule of gaseous solute into the solvent. The crucial step is to appraise how the solvent gets involved in a chemical reaction. Its action can be direct if some molecules of solvent take part in the chemical process or indirect if the solvent -then labeled the 'bulk medium'- only modifies reactants reactivity compared with that of the same molecules in the gas phase. Whatever the context may be, the solvation free energy is calculated from the equation (Leach, 2001):

$$\Delta G_{\text{solv}} = \Delta G_{\text{elec}} + \Delta G_{\text{vdw}} + \Delta G_{\text{cav}}$$

$\Delta G_{\text{elec}}$  quantifies the interaction between the solvent and the solute, it is all the more important as the ionicity or polarity is great.  $\Delta G_{\text{vdw}}$  takes into account Van der Waals interactions between the two. To finish,  $\Delta G_{\text{cav}}$  quantifies the cavity occupied by the solute while counting solvent reorganization around the cavity and the necessary work to fight against solvent pressure when the cavity is created. It is possible to encompass the two last terms within the equation:

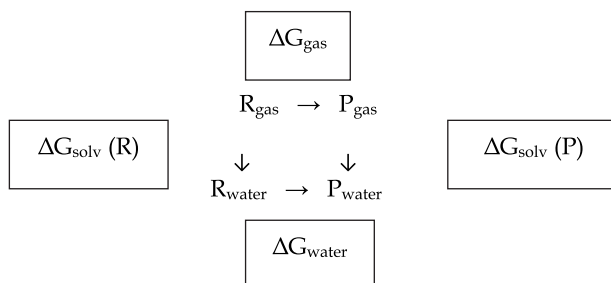
$$\Delta G_{\text{vdw}} + \Delta G_{\text{cav}} = a S + b$$

$a$  and  $b$  are constants, and  $S$  is the area of contact between the solute and the solvent. The different models that enable chemists to calculate  $\Delta G_{\text{solv}}$  mostly differs by the way they appraise  $\Delta G_{\text{elec}}$ . From earlier models developed by Born (1920) and Onsager (1936) to the PCM model -Polarisable Continuum Method-, the form of the cavity and the study of polarization between the solvent and the solute were continuously modified and improved (Barone et al., 2004; Cossi et al., 2002). The surface of the cavity was divided into fine-grained fragments labeled 'tesseræ', the wavefunction of solute is determined by Self-Consistent Field iteration. Two others models were performed, the COSMO theory -Conductor-Like Screening Model- and C-PCM approach -Conductor-Like PCM-. Modeling



the interactions between the solute and the solvent is a challenge for current quantum chemists. In this context, thermodynamic quantities are the heuristic framework that shapes quantum investigations for achieving better models. The calculation of such thermodynamic quantities stir up: (1) new polarization descriptions and understanding; (2) the creation of new algorithms and cavity topological models (Barone et al., 2004); (3) the continuous recasting of levels of description and software to optimize geometry or to calculate energy quantities (Takano & Houk, 2005); (4) the modelling of the electronic density of the solute especially outside the cavity.

It is then easy to express the free energy of chemical reaction in water using the following classic thermodynamic cycle (Picot et al., 2008):



This cycle in turn implied the following formula:

$$\Delta G_{\text{water}} = \Delta G_{\text{gas}} + \Delta G_{\text{solv}} (\text{P}) - \Delta G_{\text{solv}} (\text{R})$$

Let us analyze how those thermodynamic quantities guide Picot et al. during their investigation of zinc-thiolate complexes alkylation. This short study will allow us to grasp thermodynamics role and status in workaday chemical quantum practices of research.

They first need biomimetic models that are appropriate for both structural and mechanistic studies. Based on the experimental data, they search for a consistent series of zinc complexes in which the ligands, the electric charge, and the availability of hydrogen bonding to the atom of sulfur can be varied. They choose the Gaussian 03 software and a level of calculation for the geometry optimizations using basis especially designed for each atom or physical contraction, diffusion or polarization. For each possible mechanistic pathway -see figure 1 below-, they scrutinize each stationary point by using frequency analysis. Each transition state -labeled TS1-3 in the mechanisms presented below- was verified by stepping along the reaction coordinate and confirming that the transformation occurred.

They then calculate the gas phase Gibbs free energy, and use C-PCM model to calculate the solvation free energy within a precise set of levels of calculations. They can finally work out the react free energy in aqueous phase. They assess the adequacy of the chemical modeling and of the level of computation against observed databases of zinc complexes. They thus propose all the necessary thermodynamic quantities to analyze the chemical reaction - figure 2 below.

Those thermodynamic quantities guide the authors along their line of enquiry. They compared energy barriers required to reach transition states in order to elucidate all the influencing parameters such as the global charge of the complex, the hydrogen bond, the role of zinc ligands and that of the solvent. In doing so, they confirm that their

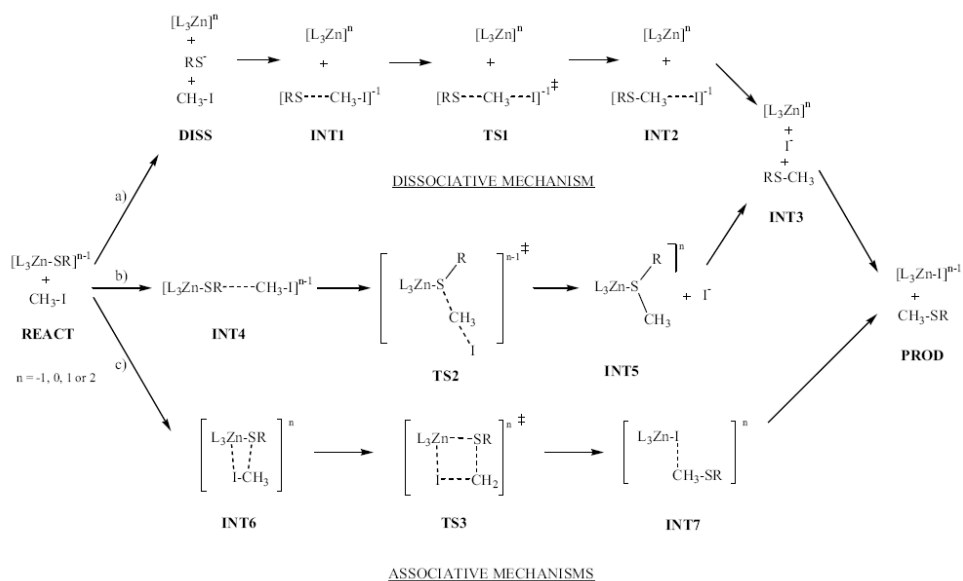


Fig. 1. Possible mechanistic pathways for the alkylation of a zinc-bound thiolate by methyl iodide. (Picot et al., 2008).

	1		2		3		4	
	$\Delta G_{\text{gas}}$	$\Delta G_{\text{water}}$	$\Delta G_{\text{gas}}$	$\Delta G_{\text{water}}$	$\Delta G_{\text{gas}}$	$\Delta G_{\text{water}}$	$\Delta G_{\text{gas}}$	$\Delta G_{\text{water}}$
<b>REACT</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Pathway a)								
<b>DISS</b>	128.9	23.0	116.5	19.1	-38.5	2.4	-40.0	-0.1
<b>INT1</b>	128.0	35.3	117.0	30.3	-39.5	14.6	-39.5	11.2
<b>TS1</b>	130.7	40.9	123.3	39.7	-36.7	20.3	-33.2	20.6
<b>INT2</b>	105.5	7.4	101.8	11.1	-61.9	-13.2	-54.7	-8.0
<b>INT3</b>	107.4	-7.5	108.9	-6.2	-60.0	-28.2	-47.6	-25.3
Pathway b)								
<b>INT4</b>	4.2	10.9	6.3	10.9				
<b>TS2</b>	32.3	26.5	34.8	26.9	4.8	23.2	8.6	25.0
<b>INT5</b>	87.5	-6.1	94.4	-2.4	-53.6	-14.1	-43.6	-11.0
Pathway c)								
<b>INT6</b>	6.9	12.4	6.6	14.0				
<b>TS3</b>	54.4	56.4	55.5	57.3				
<b>INT7</b>	-11.4	-8.4	-10.2	-7.5				
<b>PROD</b>	-17.7	-20.7	-16.3	-19.3	-17.4	-22.3	-5.0	-19.4

Fig. 2. Relative  $\Delta G_{\text{gas}}$  and  $\Delta G_{\text{water}}$  in kcal.mol<sup>-1</sup>. (Picot et al., 2008).

computational outcomes are in agreement with several experimental studies. They for instance show that the net electronic charge of the complex plays a significant role not only on its reactivity, but especially on the mechanism of thiolate alkylation. They finally discuss the nature of the pathways depending on all those energy considerations. Once again, geometry and molecular configurations of the transition state are modeled and assumed to make those predictions become achievable. The entropic contribution is thus of primary importance to query such chemical potential mechanisms.

Thermodynamics is thus a tool for calibrating levels of computation (Curtis et al., 1997; Trulhar & Zhao, 2008b), but it also shapes solvation modeling and the basic reasoning of mechanistic investigation (Takano & Houk, 2005). In a way, thermodynamics embeds a wide class of quantum activities of seeking and predicting. It provides quantum chemical methods with necessary conditions for reasoning and inventing new methods for calculations (Grimme et al., 2010).

## 5. Conclusion

The study of both earlier and recent quantum chemical methods highlights the way that thermodynamics is intertwined with quantum methods within a large network of scientific practices that includes computation, chemistry, spectroscopy, crystallography, physics, and so on. As Rouse claims concerning scientific practices (1996, p. 177): 'What results is not a systematic unification of the achievements of different scientific disciplines but a complex and partial overlap and interaction among the ways those disciplines develop over time.' Chemists connect ways of doing science and transform them within ongoing open-ended processes of research. As we have pointed out, thermodynamics was transmuted into thermochemistry through chemical practices, and conversely chemical instrumentation and ways of modeling were transformed by thermochemistry.

The role of thermodynamics is undoubtedly to validate models and methods while stirring up techno scientific creativity. The status of thermodynamics within quantum chemical methods is that of a reference framework that enables chemists to carry out their semi-empirical calculations or to create new *ab initio* predictions for thermodynamic data. This conclusion can be widened by considering other methods such as metadynamics, AIM - Atoms in Molecules - and so on.

This study also points out that alleged incommensurable scientific worlds such as thermodynamics and quantum mechanics, the assumptions, the formalisms and the natures - descriptive or predictive - of which are completely different, can constitutively interact to form the composite field of quantum chemistry. Epistemological queries thus arise concerning inter-levels description of what we call 'reality' and the way scientific fields and knowledge can be mutually stabilized. To this extent, this study also stresses the importance of an epistemology that focuses its attention on scientific practices while including historical insights.

It is interesting to notice that chemical affinities reappear in the latest quantum chemical background. Truhlar and Zhao, among others, refer to affinities -electron affinities, proton affinities of different molecules- in their benchmark databases. Thermodynamics was first introduced in chemistry, we have shown, because it provided chemists with a notion of quantitative affinity. This concept went astray in earlier chemical quantum works and then reappeared from within databases or concepts that help current quantum chemists to shape

their functionals according to thermochemistry and to investigate chemical reactivity. Further epistemological investigations are considered necessary to open up the reviving role of the concept of affinity to scrutiny in modern chemistry.

## 6. Acknowledgments

I would like to thank Rom Harré for his second reading of this paper, his advice and his generosity. I also would like to thank Miss Zgela, the Publishing Process Manager in charge for the book, for her help during the different steps of the publishing process.

## 7. References

- Barone, V., Improta, R. & Rega, N. (2004). Computation of protein pK's values by an integrated density functional theory/polarizable continuum model approach. *Theoretical Chemistry Accounts*, 111, pp. 237-245.
- Becke, A. D. (1993). Density-functional thermochemistry III. The role of exact exchange. *Journal of Chemical Physics*, 98, pp. 5648-5652.
- Bensaude-Vincent, B. & Stengers I. (1996). *History of chemistry*. Harvard University Press.
- Bensaude-Vincent, B. & Simon, J. (2008). *Chemistry, the impure science*. Imperial College Press.
- Bitbol, M. (1998). *L'aveuglante proximité du réel*. Flammarion. Paris.
- Cossi, M. et al. (2002). New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *Journal of Chemical Physics*, 117, pp. 43-54.
- Curtiss, L.A et al. (1997). Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *Journal of Chemical Physics*, 106.
- Daumas M. (1946). *L'Acte chimique, Essai sur l'histoire de la philosophie chimique*. Editions du Sablon, Bruxelles.
- Duhem P. (1902). *Le Mixte et la combinaison chimique*, réédité au Corpus des œuvres de philosophie en langue française. Fayard, Paris, 1985.
- Earley, J.E. (2005). Why there is no salt in the sea. *Foundations of Chemistry*, 7, pp. 85-102.
- Gavroglu, K. & Simões A. (1994). The Americans, the Germans, and the beginnings of quantum chemistry : The confluence of diverging traditions. *Historical Studies in the Physical Sciences*, 27, (1), pp. 47-110.
- Goerigk, L & Grimme, S. (2010). A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions - Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *Journal of Chemical Theory and Computation*, 6, pp. 107-126.
- Grev, R.S, Janssen, C.L & Schaefer, H.F. (1991). Concerning zero-point vibrational energy corrections to electronic energies. *Journal of Chemical Physics*, 95, pp. 5128-5132.
- Grimme, S. et al. (2010). A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics*, 132, 154104.
- Guldberg C.M & Waage P. (1867). *Etudes sur les Affinités Chimiques*, Christiana.
- Kragh, H. (1984). Julius Thomsen and Classical Thermochemistry. *British Journal for the History of Science*, 17, pp.255-272.

- Kragh, H., & Weineiger, S.J. (1996). Sooner Silence than Confusion: The Tortuous Entry of Entropy into Chemistry. *Historical Studies in the Physical and Biological Sciences*, 27, pp. 91-130.
- Kohn, W.; Becke, A. D.; Parr, R. G. (1996). Density functional theory of electronic structure. *Journal of Physical Chemistry*, 100, pp. 12974-12980.
- Kondepudi D. & Prigogine I. (1998). *Modern Thermodynamics: from Heat Engines to Dissipative Structures*, John Wiley & Sons, Chichester and New York.
- Harré, R. & Llored, J.P. (2011). Mereologies as the Grammars of Chemical Discourses. *Foundations of Chemistry*, 13, pp.63-76.
- Hund, F. (1927). On the Interpretation of Molecular Spectra I. *Zeitschrift für Physik*, 40.
- Hund, F. (1974). *The History of Quantum Theory*. Harrap London.
- Leach, A.R. (2001). *Molecular Modeling. Principles and Applications*. 2<sup>nd</sup> Edition. Pearson Education. Prentice-Hall.
- Llored, J.P. (2010). Mereology and quantum chemistry: the approximation of molecular orbital. *Foundations of Chemistry*, 12, pp.203-221.
- Llored, J.P. (2012, forthcoming). Towards a practical epistemology for chemistry, In: *Philosophy of Chemistry : practical roots, methods and concepts*, Llored, J.P. (ed), Cambridge Scholars Publishing, Cambridge.
- Llored, J.P. & Bitbol, M. (2010). Molecular orbital: Dispositions or Predictive Structure?, In: *Quantum biochemistry*, Matta, C.F. (ed), Wiley-VCH.
- Médoire, L.A & Tachoire H. (1994). *Histoire de la thermochimie. Prélude à la thermodynamique chimique*. Aix-en-Provence : Publications de l'Université de Provence.
- Mulliken, R.S. (1931). Bonding power of electrons and theory of valence. *Chemical Review*, 9, p 369.
- Mulliken, R.S. (1932a). Electronic structures of polyatomic molecules and valence II. General consideration. *Physical Review*, 41, p. 50.
- Mulliken, R.S. (1932b). Interpretation of Band Spectra, Part III. Electron Quantum Numbers and States of Molecules and their Atoms. *Review of Modern Physics*, 4, pp. 4-6.
- Mulliken, R.S. (1932c). Electronic structures of polyatomic molecules and valence III. Quantum theory of the double bond. *Physical Review*, 41, p. 754.
- Mulliken, R.S; Rieke, C.A & Brown, W.G. (1941). Hyperconjugation. *The Journal of the American Chemical Society*, 63, pp.41-56.
- Mulliken, R.S & Parr, R.G. (1951). LCAO Molecular Computation of Resonance Energies of Benzene and Butadiene, with General Analysis of Theoretical Versus Thermodynamical Resonance Energies. *Journal of Chemical Physics*, 19, n°10, pp.1271-78.
- Mulliken, R.S. (1967). Spectroscopy, molecular orbital and chemical bonding (Nobel lecture). *Science*, 157, pp. 13-24.
- Needham, P. (1996). Aristotelian Chemistry : A prelude to Duhemian Metaphysics. *Studies in the History and Philosophy of Science*, 27, pp.251-269.
- Neese, F., Schwabe, T & Grimme, S. (2007). Analytic derivatives for perturbatively corrected "double hybrid" density functionals: Theory, implementation, and applications. *The Journal of Chemical Physics*, 126, 124115.
- Nye, Mary J. (1993). *From Chemical Philosophy to Theoretical Chemistry: Dynamics of Matter and Dynamics of Disciplines: 1800-1950* (section: From Chemical Affinity to Chemical Thermodynamics, pp.116-20). University of California Press.

- Mi Gyung, K. (2003). *Affinity, that elusive dream*. The MIT Press. Cambridge Massachusetts, London.
- Ostwald, W. (1919). *L'évolution d'une science. La chimie*, traduction française, Flammarion, Paris, première édition en 1909.
- Partington, J.R. (1962). *A History of Chemistry*. vol III. Macmillan (ed), London.
- Pauling, L. (1928). The shared-electron chemical bond' *Proceedings of the National Academy of Sciences*, 14, pp.359-362.
- Pauling, L. (1931). The nature of the chemical bond. Application of results obtained from the quantum mechanics and from a theory of paramagnetic susceptibility to the structure of molecules. *Journal of American Chemical Society*, 53, pp.1367-1400.
- Pauling, L. (1932). The Nature of the Chemical Bond. IV. The Energy of Single Bonds and the relative Electronegativity of Atoms. *Journal of the American Chemical Society*, 54, pp. 3570-82.
- Pauling, L. & Sherman, J. (1933a). The Nature of the Chemical Bond. VI. Calculation from Thermodynamical Data of the Energy of Resonance of Molecules among several Electronic Structures. *Journal of Chemical Physics*, 1, pp.606-17.
- Pauling, L. & Sherman, J. (1933b). The Nature of the Chemical Bond. VII. The calculation of resonance energy in conjugated systems. *Journal of Chemical Physics*, 1, pp.679-86.
- Picot, D., Ohanessian, G. & Frison, G. (2008). The Alkylation Mechanism of Zinc-Bound Thiolates Depends upon the Zinc Ligands. *Inorganic Chemistry*, 47, pp. 8167-8178.
- Pickering, A. (1995). *The Mangle of Practice. Time, Agency and Science*. The University of Chicago Press. Chicago.p.21. ISBN 0-226-66802-9.
- Rouse, J. (1996). *Engaging Science; How to Understand Its Practices Philosophically*. Cornell University Press. Ithaca and London.
- Sainte Claire Deville H. (1914). *Leçons sur la Dissociation. Professées devant la Société Chimique de Paris le 18 mars et le 1<sup>er</sup> avril 1864*, Collection Les Classiques de la Science, Paris.
- Servos, J.W. (1990). *Physical Chemistry from Ostwald to Pauling : The Making of Science in America*. Princeton: Princeton University Press.
- Slater, J.C. (1931). Directed valence in polyatomic molecules. *Physical Review*, 37, pp. 481-489.
- Swietolawski, W. (1920). The Thermochemistry of Hydrocarbons according to P. W. Zubow's Data. *Journal of the American Chemical Society*, 42, pp.1312-21.
- Takano, Y. & Houk, K. J. (2005). *Chemical Theory Computations*, 1 (1), pp. 70-77. doi:10.1021/ct049977a.
- Vemulapalli, G.K. (2003). Property reduction in chemistry. Some lessons, In: *Chemical Explanation. Characteristics, Development, Autonomy*. Joseph E. Earley (Ed), Annals of the New York Academy of sciences, 988, Issue 1, p. 95.
- Weininger, S.J. (2001). Affinity, Additivity and the Reification of the Bond, In: *Tools and Modes of Representation in the Laboratory Sciences*, Ursula Klein ed., Boston Studies in the Philosophy of Science, Kluwer Academic Publishers.
- Zhao, Y. & Truhlar, D.G. (2008a). Density Functionals with Broad Applicability in Chemistry. *Accounts of Chemical Research*, 41, n°2, pp. 157-167.
- Zhao, Y. & Truhlar, D.G. (2008b). Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Journal of Chemical Theory and Computation*, 4, pp. 1849-1868.

# Thermodynamics of ABO<sub>3</sub>-Type Perovskite Surfaces

Eugene Heifets<sup>1</sup>, Eugene A. Kotomin<sup>1,2</sup>, Yuri A. Mastrikov<sup>2</sup>,  
Sergej Piskunov<sup>3</sup> and Joachim Maier<sup>1</sup>

<sup>1</sup>Max Planck Institute for Solid State Research, Stuttgart,

<sup>2</sup>Institute of Solid State Physics, University of Latvia, Riga,

<sup>3</sup>Department of Computer Science, University of Latvia, Riga,

<sup>1</sup>Germany

<sup>2,3</sup>Latvia

## 1. Introduction

The ABO<sub>3</sub>-type perovskite manganites, cobaltates, and ferrates (A= La, Sr, Ca; B=Mn, Co, Fe) are important functional materials which have numerous high-tech applications due to their outstanding magnetic and electrical properties, such as colossal magnetoresistance, half-metallic behavior, and composition-dependent metal-insulator transition (Coey et al., 1999; Haghiri-Gosnet & Renard, 2003). Owing to high electronic and ionic conductivities, these materials show also excellent electrochemical performance, thermal and chemical stability, as well as compatibility with widely used electrolyte based on yttrium-stabilized zirconia (YSZ). Therefore they are among the most promising materials as cathodes in solid oxide fuel Cells (SOFCs) (Fleig et al., 2003) and gas-permeation membranes (Zhou, 2009). Many of the above-mentioned applications require understanding and control of surface properties. An important example is LaMnO<sub>3</sub> (LMO). Pure LMO has a cubic structure above 750 K, whereas below this temperature the crystalline structure is orthorhombic, with four formula units in a primitive cell. Doping of LMO with Sr allows one to increase both the ionic and electronic conductivity as well as to stabilize the cubic structure down to room temperatures - necessary conditions for improving catalytic performance of LMO in electrochemical devices, e.g. cathodes for SOFCs. In optimal compositions of La<sub>1-x<sub>b</sub></sub>Sr<sub>x<sub>b</sub></sub>MnO<sub>3</sub> (LSM) solid solution the bulk concentration of Sr reaches x<sub>b</sub>≈0.2 .

Understanding of LMO and LSM basic properties (first of all, energetic stability and reactivity) for pure and adsorbate-covered surfaces is important for both the low-temperature applications (e.g., spintronics) and for high-temperature electrochemical processes where understanding the mechanism of oxygen reduction on the surfaces is a key issue in improving the performance of SOFC cathodes and gas-permeation membranes at relatively high (~800 C) temperatures. First of all, it is necessary to determine which LMO/LSM surfaces are the most stable under operational conditions and which terminations are the energetically preferential? For example, the results of our simulations described below show that the [001] surfaces are the most stable ones in the case of LMO (as

compared to [011] and others). However, the [001] surfaces have, in turn, two different terminations: LaO or MnO<sub>2</sub>. We will compare stabilities of these terminations under different environmental conditions (temperature and partial pressure of oxygen gas). Another important question to be addressed is, how Sr doping affects relative stabilities of the LMO surfaces? These issues directly influence the SOFC cathode performance. Answering these questions requires a thermodynamic analysis of surfaces under realistic SOFC operational conditions which is in the main focus of this Chapter. Such a thermodynamic analysis is becoming quite common in investigating structure and stability of various crystal surfaces (Examples of thermodynamic analyses of binary and ternary compounds can be found in Reuter & Scheffler, 2001a, 2001b; Bottin et al., 2003; Heifets et al., 2007a, 2007b, Johnson et al., 2004).

The thermodynamic analysis requires careful calculations of energies for two-dimensional slabs terminated by surfaces with various orientations and terminations. The required energies could be calculated using *ab initio* methods of the atomic and electronic structure based on density functional theory (DFT). In this Chapter, we present the results obtained using two complementary *ab initio* DFT approaches employing two different types of basis sets (BS) representing the electronic density distribution: plane waves (PW) and linear combination of atomic orbitals (LCAO). Both techniques were used to calculate the atomic and electronic structures of a pure LMO whereas investigation of the Sr influence on the stability of different (001) surfaces was performed within LCAO approach.

After studying the stabilities of various surfaces, the next step is investigating the relevant electrochemical processes on the most stable surfaces. For this purpose, we have to evaluate the adsorption energies for O<sub>2</sub> molecules, O atoms, the formation energies of O vacancies in the bulk and at the stable perovskite surfaces. These energies, together with calculated diffusion barriers of these species and reactions between them, allow us to determine the mechanism of incorporation of O atoms into the cathode materials. However, such mechanistic and kinetic analyses lie beyond the scope of this Chapter (for more details see e.g. Matrikov et al., 2010). Therefore, we limit ourselves here only to the thermodynamic characterization of the initial stages of the oxygen incorporation reaction, which include formation of stable adsorbed species (adsorbed O atoms, O<sub>2</sub> molecules) and formation of oxygen vacancies. The data for formation of both oxygen vacancies and adsorbed oxygen atoms and molecules have been collected using plane wave based DFT.

## 2. Computational details

The employed thermodynamic analysis relies on the energies obtained by DFT computations of the electronic structure of slabs terminated by given surfaces using the above-mentioned two types of basis sets. All calculations are performed with spin-polarized electronic densities, complete neglect of spin polarization results in considerable errors in material properties (Kotomin et al, 2008)).

The plane wave calculations were performed with VASP 4.6.19 code (Kresse & Hafner, 1993; Kresse & Furthmüller, 1996; Kresse et al., 2011), which implements projector augmented wave (PAW) technique (Bloechl, 1994; Kresse & Joubert, 1999), and generalized gradient approximation (GGA) exchange-correlation functional proposed by Perdew and Wang (PW91) (Perdew et al., 1992). Calculations were done with the cut-off energy of 400 eV. The core orbitals on all atoms were described by PAW pseudopotentials, while electronic



wavefunctions of valence electrons on O atoms and valence and core-valence electrons on metal atoms were explicitly evaluated in our calculations.

We found that seven- and eight-plane *slabs* infinite in two (x-y) directions are thick enough to show convergence of the main properties. The periodically repeated slabs were separated along the z-axis by a large vacuum gap of 15.8 Å. All atomic coordinates in slabs were allowed to relax. To avoid problems with a slab dipole moment and to ensure having identical surfaces on both sides of slabs, we employed the symmetrical seven-layer slab MnO<sub>2</sub>(LaO-MnO<sub>2</sub>)<sub>3</sub> in our plane-wave simulations, even though it has a Mn excess relative to La and a higher oxygen content. Such a choice of the slab structure however only slightly changes the calculated energies. For example, the energy for dissociative oxygen adsorption on the [001] MnO<sub>2</sub>-terminated surface



is -2.7 eV for eight-layers (LaO-MnO<sub>2</sub>)<sub>4</sub> slab and -2.2 eV for the symmetrical seven-layer MnO<sub>2</sub>-(LaO-MnO<sub>2</sub>)<sub>3</sub> slab. The use of symmetrical slabs also allows decoupling the effects of different surface terminations and saving computational time due to the possibility to exploit higher symmetry of the slabs. The simulations were done using an extended  $2\sqrt{2} \times 2\sqrt{2}$  surface unit cell and a  $2 \times 2$  Monkhorst-Pack k-point mesh in the Brillouin zone (Monkhorst & Pack, 1976). Such a unit cell corresponds to 12.5% concentration (coverage) of the surface defects in calculations of vacancies and adsorbed atoms and molecules.

The choice of the magnetic configuration only weakly affects the calculated surface relaxation and surface energies (Evarestov, et. al., 2005; Kotomin et al, 2008; Mastrikov et al., 2009). Relevant magnetic effects are sufficiently small ( $\approx 0.1$ eV) as do not affect noticeably relative stabilities of different surfaces; these values are much smaller than considered adsorption energies and vacancy formation energies. As for slabs the ferromagnetic (FM) configuration has the lowest energy, we performed all further plane-wave calculations with FM ordering of atomic spins.

The quality of plane-wave calculations can be illustrated by the results for the bulk properties (Evarestov, et. al., 2005; Mastrikov et al., 2009). In particular, for the low-temperature orthorhombic structure the A-type antiferromagnetic (A-AFM) configuration (in which spins point in the same direction within each [001] plane, but opposite in the neighbor planes) is the energetically most favorable one, in agreement with experiment. The lattice constant of both the cubic and orthorhombic phase exceeds the experimental value only by 0.5%. The calculated cohesive energy of 30.7 eV is also close to the experimental value (31 eV).

In our *ab initio* LCAO calculations we use DFT-HF (*i.e.*, density functional theory and Hartree-Fock) hybrid exchange-correlation functional which gave very good results for the electronic structure in our previous studies of both LMO and LSM (Evarestov et al., 2005; Piskunov et al., 2007). We employ here the hybrid B3LYP exchange-correlation functional (Becke, 1993). The simulations were carried out with the CRYSTAL06 computer code (Dovesi, et. al., 2007), employing BS of the atom-centered Gaussian-type functions. For Mn and O, all electrons are explicitly included into calculations. The inner core electrons of Sr and La are described by small-core Hay-Wadt effective pseudopotentials (Hay & Wadt, 1984) and by the nonrelativistic pseudopotential (Dolg et al., 1989), respectively. BSs for Sr and O in the form of 311d1G and 8-411d1G, respectively, were optimized by Piskunov et al., 2004. BS for Mn was taken from (Towler et al., 1994) in the form of 86-411d41G, BS for La is

provided in the CRYSTAL code's homepage (Dovesi, et. al., 2007) in form 311-31d3f1, to which we added an  $f$ -type polarization Gaussian function with exponent optimized in LMO ( $\alpha=0.475$ ). The reciprocal space integration was performed by sampling the Brillouin zone with the  $4 \times 4$  Monkhorst-Pack mesh (Monkhorst & Pack, 1976). In our LCAO calculations, nine-layer symmetrical slabs (terminated on both sides by either [001]  $\text{MnO}_2$  or  $\text{La}(\text{Sr})\text{O}$  surfaces) were used. The calculations were carried out for cubic phases and for A-AFM magnetic ordering of spins on Mn atoms. All atoms have been allowed to relax to the minimum of the total energy. This approach was initially tested on bulk properties as well, the experimentally measured atomic, electronic, and low-temperature magnetic structure of pure LMO and LSM ( $x_b=1/8$ ) were very well reproduced (Piskunov et al., 2007).

### 3. Thermodynamic analysis of surface stability

As was mentioned above, understanding of many surface related phenomena requires preliminary investigation of the relative stabilities of various crystalline surfaces. Usually (especially for high-temperature processes such as catalysis in electrochemical devices), determining the structure with the lowest internal energy is not sufficient. The internal energy characterizes only systems with a constant chemical composition, while atomic diffusion and atomic exchange between environment and surfaces occur at high temperatures. Thus, we have to take into account the exchange of atoms between the bulk crystal, its surface, and the gas phase, into our analysis of surface stability. Such processes are included into the Gibbs free energies at the thermodynamic level of description. Therefore, we have to calculate the surface Gibbs free energy (SGFE)  $\Omega^i$  for the LMO and LSM surfaces of various orientations and terminations. The SGFE is a measure of the excess energy of a semi-infinite crystal in contact with matter reservoirs with respect to the bulk crystal (Bottin et al., 2003; Heifets et al., 2007a, 2007b; Johnston et al., 2004; Mastrovich et al., 2009; Padilla & Vanderbilt, 1997, 1998; Piskunov et al., 2008; Pojani et al., 1999; Reuter & Scheffler, 2001b, 2004). The SGFEs are functions of chemical potentials of different atomic species. The most stable surface has a structure, orientation and composition with the lowest SGFE among all possible surfaces.

#### 3.1 Method of analysis for LMO surfaces

Introducing the chemical potentials  $\mu_{\text{La}}$ ,  $\mu_{\text{Mn}}$ , and  $\mu_{\text{O}}$  for the La, Mn, and O atomic species, respectively, the SGFE per unit cell area  $\Omega^i$  corresponding to the  $i$  termination is defined as

$$\Omega^i = \frac{1}{2} [G_{\text{slab}}^i - N_{\text{La}}^i \mu_{\text{La}} - N_{\text{Mn}}^i \mu_{\text{Mn}} - N_{\text{O}}^i \mu_{\text{O}}] \quad (2)$$

where  $G_{\text{slab}}^i$  is the Gibbs free energy for the slab terminated by surface  $i$ ,  $N_{\text{La}}^i$ ,  $N_{\text{Mn}}^i$ , and  $N_{\text{O}}^i$  denote numbers of La, Mn, and O atoms in the slab. Here we assume that the slab is symmetrical and has the same orientation, composition, and structure on both sides. The SGFE per unit area is represented by

$$\omega^i = \frac{\Omega^i}{A} \quad (3)$$

The thermodynamic part of the description below follows the well known chemical thermodynamics formalism developed originally by Gibbs in 1875 (see Gibbs, 1948) for

perfect bulk and surfaces and extended by Wagner & Schottky, 1930 (also Wagner, 1936) for point defects.

The chemical potential  $\mu_{LaMnO_3}$  of LMO (in the considered orthorhombic or cubic phase) is equal to the sum of the chemical potentials of each atomic component in the LMO crystal:

$$\mu_{LaMnO_3} = \mu_{La} + \mu_{Mn} + 3\mu_O \quad (4)$$

Owing to the requirement for the surface of each slab to be in equilibrium with the bulk LMO, the chemical potential is equal to the specific bulk crystal Gibbs free energy accordingly to

$$\mu_{LaMnO_3} = \delta_{LaMnO_3}^{bulk} \quad (5)$$

Eq. (4) imposes restrictions on  $\mu_{La}$ ,  $\mu_{Mn}$ , and  $\mu_O$ , leaving only two of them as independent variables. We use in following  $\mu_O$  as one of the independent variables because we consider oxygen exchange between the LaMnO<sub>3</sub> crystal and gas phase and have to account for strong dependence of this chemical potential on  $T$  and  $pO_2$ . As another independent variable, we use  $\mu_{Mn}$ . We will simplify the equation for the SGFE and eliminate the chemical potentials  $\mu_{La}$  and  $\mu_{LaMnO_3}$  by substituting this expression for the LMO bulk chemical potential:

$$\Omega^i = \frac{1}{2} \left[ G_{slab}^i - N_A^i \delta_{LaMnO_3}^{bulk} \right] - \Gamma_{A,Mn}^i \mu_{Mn} - \Gamma_{A,O}^i \mu_O, \quad (6)$$

where  $\Gamma_{A,a}^i$  are the Gibbs excesses in the  $i$ -terminated surface of components  $a$  with respect to the number of ions in A type sites (for ABO<sub>3</sub> perovskites) of the slabs (Gibbs,1948; Johnston et al., 2004) :

$$\Gamma_{A,a}^i = \frac{1}{2} \left( N_a^i - N_A^i \frac{N_a^{bulk}}{N_A^{bulk}} \right) \quad (7)$$

Here A type of sites are occupied solely by La atoms in LMO, so  $N_A = N_{La}$  for LMO. This will become somewhat more complicated in solid solutions such as LSM (see the next subsection).  $N_A^{bulk}$  is the number of A-sites in unit cell in the bulk.  $N_a^{bulk}$  is the number of  $a$  atoms in unit cell in the bulk.

The Gibbs free energies per unit cell for crystals and slabs are defined as

$$g_j = E_j + E_j^{vibr} - T s_j + p v_j \quad (8)$$

where  $E_j$  is the static component of the crystal energy,  $E_j^{vibr}$  is the vibrational contribution to the crystal energy,  $v_j$  volume, and  $s_j$  entropy. All these values are given per formula unit in  $j$ -type (=La,Mn, LMO...) crystals. We can reasonably assume that the applied pressure is not higher than ~100 atm. in practical cases. The volume per lattice molecule in LaMnO<sub>3</sub> is ~64 Å<sup>3</sup>. Then the largest  $p v_j$  term in Eq.(13) can be estimated as ~ 5 meV. This value is much smaller than the amount of uncertainty in our DFT computations and, therefore, can be safely neglected. As it is commonly practiced, we will neglect the very small vibration contributions to  $g_j$ , including contributions from zero-point oscillations to the vibrational part of the total energy. This rough estimate is usually valid, but can be broken if the studied material has soft modes. The same consideration is valid for slabs used in the present

simulations. While it might be important to check vibrational contributions in some cases, here we will neglect it. Besides, facilities in computer codes for calculations of vibrational spectra of crystals and slabs appeared only within a few last years and such calculations are still very demanding and practically possible only for relatively small unit cells. Therefore, we approximate the Gibbs free energies with the total energies obtained from DFT calculations:

$$g_j \approx E_j \quad (9)$$

Then, replacing the chemical potentials of La and Mn atoms by their deviations from chemical potentials in the most stable phases of respective elementary crystals,

$$\Delta\mu_{La} = \mu_{La} - g_{La}^{bulk} \approx \mu_{La} - E_{La}^{bulk} \quad (10)$$

and

$$\Delta\mu_{Mn} = \mu_{Mn} - g_{Mn}^{bulk} \approx \mu_{Mn} - E_{Mn}^{bulk} \quad (11)$$

and chemical potential of O atoms by its deviation from the energy of an oxygen atom in a free, isolated O<sub>2</sub> molecule (  $E_{O_2}^{total} / 2$  ),

$$\Delta\mu_O = \mu_O - \frac{E_{O_2}^{total}}{2} \quad (12)$$

we can determine the surface Gibbs free energy from

$$\Omega^i = \phi^i - \Gamma_{A,Mn}^i \Delta\mu_{Mn} - \Gamma_{A,O}^i \Delta\mu_O. \quad (13)$$

We can express the constant  $\phi^i$  in Eq. (13) as

$$\begin{aligned} \phi^i &= \frac{1}{2} \left[ G_{slab}^i - N_A^i g_{LaMnO_3}^{bulk} \right] - \Gamma_{A,Mn}^i g_{Mn}^{bulk} - \frac{1}{2} \Gamma_{A,O}^i E_{O_2} \approx \\ &\approx \frac{1}{2} \left[ E_{slab}^i - N_A^i E_{LaMnO_3}^{bulk} \right] - \Gamma_{A,Mn}^i E_{Mn}^{bulk} - \frac{1}{2} \Gamma_{A,O}^i E_{O_2}, \end{aligned} \quad (14)$$

what resembles the expression for the Gibbs free energy of surface formation. Here  $E_{slab}$  stands for the total energy of a slab and replaces the Gibbs free energy of the slab.

The equilibrium condition (5) can be rewritten as

$$\Delta\mu_{La} + \Delta\mu_{Mn} + 3\Delta\mu_O = \Delta g_f^{bulk}(LaMnO_3) \quad (15)$$

where

$$\begin{aligned} \Delta g_f^{bulk}(LaMnO_3) &= g_{LaMnO_3}^{bulk} - g_{La}^{bulk} - g_{Mn}^{bulk} - \frac{3}{2} E_{O_2} \approx \\ &\approx E_{LaMnO_3}^{bulk} - E_{La}^{bulk} - E_{Mn}^{bulk} - \frac{3}{2} E_{O_2}. \end{aligned} \quad (16)$$

Here  $\Delta g_f^{bulk}(LaMnO_3)$  has meaning of the Gibbs free energy of LaMnO<sub>3</sub> formation from La, Mn and O<sub>2</sub> in their standard states.

The range of values of the chemical potentials which consistent with existence and stability of the crystal (LMO here) itself is determined by the set of the following conditions. To prevent La and Mn metals from leaving LMO and forming precipitates, their chemical potentials must be lower in LMO than in corresponding bulk metals. These conditions mean:

$$\mu_{Mn} < g_{Mn}^{bulk} \quad (17)$$

and

$$\mu_{La} < g_{La}^{bulk} \quad (18)$$

Similarly, precipitation of oxides does not occur, if the chemical potentials of atoms in LMO are smaller than in the oxides:

$$2\mu_{La} + 3\mu_O < g_{La2O3}^{bulk} \quad (19)$$

$$\mu_{Mn} + \mu_O < g_{MnO}^{bulk} \quad (20)$$

$$3\mu_{Mn} + 4\mu_O < g_{Mn3O4}^{bulk} \quad (21)$$

$$2\mu_{Mn} + 3\mu_O < g_{Mn2O3}^{bulk} \quad (22)$$

and

$$\mu_{Mn} + 2\mu_O < g_{MnO2}^{bulk} \quad (23)$$

Exclusion of La chemical potential and expressing of these conditions through the deviations of the chemical potentials (10-12) transform the conditions to

$$\Delta\mu_{Mn} < 0 \quad (24)$$

$$\Delta\mu_{Mn} + 3\Delta\mu_O > \Delta g_f^{bulk}(LaMnO_3) \quad (25)$$

$$x\Delta\mu_{Mn} + y\Delta\mu_O < \Delta g_f^{bulk}(Mn_xO_y) \quad (26)$$

and

$$2\Delta\mu_{Mn} + 3\Delta\mu_O > 2\Delta g_f^{bulk}(LaMnO_3) - \Delta g_f^{bulk}(La_2O_3) \quad (27)$$

where the formation energies of oxides are defined by

$$\begin{aligned} \Delta g_f^{bulk}(M_xO_y) &= g_{M_xO_y}^{bulk} - xg_M^{bulk} - \frac{y}{2}E_{O_2} \approx \\ &\approx E_{M_xO_y}^{bulk} - xE_M^{bulk} - \frac{y}{2}E_{O_2} . \end{aligned} \quad (28)$$

Note, however, that sometimes compositions are fixed by bringing the multinary crystals into coexistence with less complex sub-phases.

If the SGFE becomes negative, surface formation becomes energetically favorable and the crystal will be destroyed. Therefore, the condition for sustaining a crystal structure is for SGFE to be positive for all potential surface terminations. Therefore, one more set of conditions on the chemical potentials of the crystal components can be written as

$$\Omega^i > 0 \quad (29)$$

where  $i$  corresponds to the surface with the lowest SGFE.

### 3.2 Method of analysis for LSM surfaces

In LSM we have to re-define the SGFEs, because there are now four components in this material (instead of three in LMO) with Sr atoms substituting a fraction of La atoms in the perovskite A sub-lattice. The SGFE definition for LSM can be written as

$$\Omega^i = \frac{1}{2} \left[ G_{slab}^i - N_{La}^i \mu_{La} - N_{Sr}^i \mu_{Sr} - N_{Mn}^i \mu_{Mn} - N_O^i \mu_O \right] \quad (30a)$$

Let us denote concentration of Sr atoms in the bulk of LSM as

$$x_b = \frac{N_{Sr}^{bulk}}{N_A^{bulk}} \quad (30b)$$

where  $N_{Sr}^{bulk}$  is the average number of Sr atoms per crystal unit cell in the bulk. Then for LSM

$$N_{La}^{bulk} = (1 - x_b) N_A^{bulk} \quad (31)$$

becomes the average number of La atoms per LSM unit cell in the bulk.

The chemical potential of a LSM formula unit is

$$\mu_{LSM} = (1 - x_b) \mu_{La} + x_b \mu_{Sr} + \mu_{Mn} + 3\mu_O \quad (32)$$

Equilibrium between LSM surface and its bulk means that

$$\mu_{LSM} = g_{LSM}^{bulk} \quad (33)$$

We will continue using approximation (9) in the following, replacing the Gibbs free energies of bulk and slab unit cells by their total energies. The conditions (32, 33) impose restrictions on four chemical potentials of all LSM components and reduces the number of independent components to three. We have chosen to keep the chemical potentials of O, Mn and La as independent variables. Then the chemical potential of the Sr atom can be expressed as

$$\mu_{Sr} = \frac{1}{x_b} \left( E_{LSM} - (1 - x_b) \mu_{La} - \mu_{Mn} - 3\mu_O \right) \quad (34a)$$

and its deviation (analogous to eqs. (10-12) and keeping in mind approximation (9)) as

$$\Delta\mu_{Sr} = \mu_{Sr} - g_{Sr}^{bulk} \approx \mu_{Sr} - E_{Sr}^{bulk} \quad (34b)$$

The expressions for the excesses  $\Gamma_{A,O}^i$  and  $\Gamma_{A,Mn}^i$  do not change with respect to LMO. We still have to remember only that  $N_A$  does not coincide any more with  $N_{La}$  in LSM.  $N_A$  refers only to the number of A-sites in the perovskite unit cell, but not to the number of La atoms. Since we excluded chemical potential for Sr, only the excesses for La atoms will be required. For the calculation of excesses of La atoms we have to account for mixing of La and Sr atoms in A-site of the perovskite lattice. Using eqs.(7,31), the excess of La atoms for surface  $i$  can be expressed as:

$$\Gamma_{A,La}^i = \frac{1}{2} \left( N_{La}^i - \frac{N_{La}^{bulk}}{N_A^{bulk}} N_A^i \right) = \frac{1}{2} (N_{La}^i - (1-x_b)N_A^i) \quad (35)$$

Then SGFEs for LSM reads

$$\Omega^i = \phi^i - \Gamma_{A,La}^i \Delta\mu_{La} - \Gamma_{A,Mn}^i \Delta\mu_{Mn} - \Gamma_{A,O}^i \Delta\mu_O \quad (36)$$

where

$$\phi^i = \frac{1}{2} (G_{slab}^i - N_A^i E_{LSM}^{bulk}) - \Gamma_{A,La}^i E_{La}^{bulk} - \Gamma_{A,Mn}^i E_{Mn}^{bulk} - \Gamma_{A,O}^i \frac{E_{O_2}}{2} \quad (37)$$

The conditions of LSM crystal stability include the same bounds which work for LMO. However, we have to add conditions preventing precipitations of several new materials and express all conditions through three chemical potentials for La, Mn and O atoms. Precipitation of Mn, La, and Sr metals will be avoided if

$$\Delta\mu_{Mn} < 0 \quad (38)$$

$$\Delta\mu_{La} < 0 \quad (39)$$

and

$$(1-x_b)\Delta\mu_{La} + \Delta\mu_{Mn} + 3\Delta\mu_O > \Delta g_f^{bulk}(LSM) \quad (40)$$

where Gibbs free energy of LSM formation is

$$\Delta g_f^{bulk}(LSM) = E_{LSM}^{bulk} - x_b E_{Sr}^{bulk} - (1-x_b)E_{La}^{bulk} - E_{Mn}^{bulk} - \frac{3}{2}E_{O_2} \quad (41)$$

Precipitation of oxides is avoided, if

$$x\Delta\mu_{Mn} + y\Delta\mu_O < \Delta g_f^{bulk}(Mn_xO_y) \quad (42)$$

$$2\Delta\mu_{La} + 3\Delta\mu_O < \Delta g_f^{bulk}(La_2O_3) \quad (43)$$

$$(1-x_b)\Delta\mu_{La} + \Delta\mu_{Mn} + (3-x_b)\Delta\mu_O > \Delta g_f^{bulk}(LSM) - x_b\Delta g_f^{bulk}(SrO) \quad (44)$$

where

$$\Delta g_f^{bulk}(SrO) = E_{SrO}^{bulk} - E_{Sr}^{bulk} - \frac{E_{O_2}}{2} \quad (45)$$

Similarly, precipitation of  $LaMnO_3$  and  $SrMnO_3$  perovskites will be prevented, if

$$\begin{aligned} \frac{1}{1-x_b} \Delta g_f^{bulk}(LSM) - \frac{x_b}{1-x_b} \Delta g_f^{bulk}(SrMnO_3) < \\ < \Delta \mu_{Mn} + \Delta \mu_{La} + 3\Delta \mu_O < \Delta g_f^{bulk}(LaMnO_3). \end{aligned} \quad (46)$$

Here the Gibbs free energy of  $SrMnO_3$  formation is defined as

$$\Delta g_f^{bulk}(SrMnO_3) = E_{SrMnO_3}^{bulk} - E_{Sr}^{bulk} - E_{Mn}^{bulk} - \frac{3}{2} E_{O_2}. \quad (47)$$

Lastly, spontaneous formation of surfaces does not occur, if condition (29) is satisfied as well.

### 3.3 Determination of the chemical potential of oxygen atom

As mentioned above, an exchange of O atoms between surfaces and environment occurs at all surfaces, especially at high temperatures. Moreover, such an exchange is a key factor in many electrochemical and catalytic processes. Therefore, oxygen in the studied crystal (for instance, LMO or LSM, in this Chapter) has to be considered in equilibrium with oxygen gas in atmosphere beyond the crystal surface. The equilibrium in exchange with O atoms means equality of oxygen chemical potentials in a crystal and in the atmosphere:

$$\mu_O = \frac{1}{2} \mu_{O_2}^{gas} \quad (48)$$

Chemical potentials are hardly available experimentally. It is much more convenient to operate with gas temperatures and pressures determining the oxygen chemical potential. At the same time, the Gibbs free energies of crystals are insensitive to temperature and the pressure (within approximations accepted in our present description). Therefore, we can use the dependence of oxygen gas chemical potential  $\mu_{O_2}^{gas}$  to express the Gibbs free energies for surfaces through temperature and oxygen gas partial pressure.

Oxygen gas under the considered conditions can be treated (to a very good approximation) as an ideal gas. Therefore, dependence of its chemical potential from pressure can be expressed by the standard expression (as done by Johnston et al., 2004 and Reuter & Scheffler, 2001b)

$$\mu_{O_2}^{gas}(T, p) = \mu_{O_2}^{gas}(T, p^0) + kT \ln \left( \frac{p}{p^0} \right) \quad (49)$$

where  $k$  is the Boltzmann constant. Here  $p^0$  is the reference pressure which we can take as the standard pressure (1 atm.). The temperature dependence of  $\mu_{O_2}^{gas}(T, p^0)$  includes contributions from molecular vibrations and rotations, as well as ideal-gas entropy at pressure  $p^0$ . We can evaluate the temperature dependence of  $\mu_{O_2}^{gas}(T, p^0)$  using experimental



data from the standard thermodynamic tables (Chase, 1998; Linstrom & Mallard, 2003), following Johnston et al., 2004 and Reuter & Scheffler, 2001b. These data are collected in Table 1. For this we define an isolated oxygen molecule  $E_{O_2}$  as the reference state. Changes in the chemical potential for oxygen atom can be written as

$$\begin{aligned} \Delta\mu_O(T,p) &= \mu_O(T,p) - \frac{1}{2}E_{O_2} = \\ &= \frac{1}{2} \left\{ \Delta G_{O_2}^{gas}(T,p^0) + kT \ln \left( \frac{p}{p^0} \right) \right\} + \delta\mu_O^0 \end{aligned} \quad (50)$$

Here  $\Delta G_{O_2}^{gas}(T,p^0)$  is the change in the oxygen gas Gibbs free energy at the pressure  $p^0$  and temperature  $T$  with respect to its Gibbs free energy at  $T^0=298.15$  K

$$\begin{aligned} \Delta G_{O_2}^{gas}(T,p^0) &= G_{O_2}^{gas}(T,p^0) - G_{O_2}^{gas}(T^0,p^0) = \\ &= H_{O_2}^{gas}(T,p^0) - H_{O_2}^{gas}(T^0,p^0) - TS_{O_2}^{gas}(T,p^0) + T^0 S_{O_2}^{gas}(T^0,p^0) \end{aligned} \quad (51)$$

T, K	$\Delta G_{O_2}^{gas}(T,p^0)$ , eV	T, K	$\Delta G_{O_2}^{gas}(T,p^0)$ , eV
100	-0.07	1000	-1.10
200	-0.17	1100	-1.23
250	-0.22	1200	-1.36
298.15	-0.27	1300	-1.49
300	-0.27	1400	-1.62
400	-0.38	1500	-1.75
500	-0.50	1600	-1.88
600	-0.61	1700	-2.02
700	-0.73	1800	-2.16
800	-0.85	1900	-2.29
900	-0.97	2000	-2.43

Table 1. Variations in the Gibbs free energy for oxygen gas at standard pressure ( $p^0=1$  atm.) with respect to its value at 0 K. Data are taken from thermodynamic tables (Chase, 1998; Linstrom & Mallard 2003).

The  $\delta\mu_O^0$  in Eq. (50) is a correction which matches experimental data and the results of quantum-mechanical computations. This correction can be estimated from computations of metal oxides and metals, in a way similar to Johnston et al., 2004. Enthalpy of an  $M_xO_y$  oxide can be written as

$$h_{M_xO_y}^0 = xh_M^0 + \frac{y}{2}h_{O_2}^0 + \Delta H_{f,M_xO_y}^0 \quad (52)$$

Here enthalpies of the oxide,  $h_{M_xO_y}^0$ , and of the metal,  $h_M^0$ , can be approximated by the total energies for these materials calculated at 0 K on the same grounds as for approximation

(9). The formation heat for La and Mn oxides under standard conditions can also be found in thermodynamic tables (Chase, 1998; Linstrom & Mallard, 2003). Equation (52) allows us to estimate the standard oxygen gas enthalpy. Since we define the total energy of an oxygen molecule as a zero for chemical potential and enthalpy calculations, the correction for the enthalpy could be defined as

$$\delta h_{O_2}^0 = h_{O_2}^0 - E_{O_2} \quad (53)$$

Using the experimental standard entropy for oxygen (Chase, 1998; Linstrom & Mallard, 2003) as  $S_{O_2}^0 = 205.147 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ , the correction to the oxygen chemical potential can be calculated as

$$\begin{aligned} \delta \mu_O^0 &= \frac{1}{2}(\delta h_{O_2}^0 - T^0 S_{O_2}^{gas}(T^0, p^0)) \\ &= \frac{1}{y}(h_{M_x O_y}^0 - x h_M^0 - \Delta H_{f, M_x O_y}^0) - \frac{1}{2}(E_{O_2} + T^0 S_{O_2}^{gas}(T^0, p^0)) \end{aligned} \quad (54)$$

### 3.4 Thermodynamic consideration of oxygen adsorption and vacancy formation

Let us consider formation of relevant oxygen species and point defects in the bulk and at the  $\text{LaMnO}_3$  surface. We use the same approximation as in the previous sections: we neglect the changes of vibrational entropy in the solid, thus only states comprising gaseous  $\text{O}_2$  exhibit the temperature-dependent Gibbs free energy contribution. In this approximation, differences between the Gibbs energies for bulk crystals or slabs (including defects and adsorbates) can be replaced with the differences in the total energy calculated from DFT, while variation of oxygen chemical potential for gaseous  $\text{O}_2$  is taken from experimental data. The Gibbs free energy of reaction for removal of a neutral O atom ( $1/2 \text{ O}_2$ ) from the bulk (i.e. formation of one neutral  $\text{V}_O$  and allocation of the left-behind two electrons mainly on two nearest Mn) is defined as

$$\Delta g_f^{bulk}(\text{LaMnO}_3 : \text{V}_O) = E^{bulk}(\text{LaMnO}_3 : \text{V}_O) + \mu_O(T, p_{O_2}) - E_{\text{LaMnO}_3}^{bulk} \quad (55)$$

where  $E^{bulk}(\text{LaMnO}_3 : \text{V}_O)$  is the total energy per bulk supercell with an oxygen vacancy.

This definition can be re-written as

$$\Delta g_f^{bulk}(\text{LaMnO}_3 : \text{V}_O) = \Delta E_f^{bulk}(\text{LaMnO}_3 : \text{V}_O) + \Delta \mu_O(T, p_{O_2}) \quad (56)$$

where

$$\Delta E_f^{bulk}(\text{LaMnO}_3 : \text{V}_O) = E^{bulk}(\text{LaMnO}_3 : \text{V}_O) + \frac{1}{2}E_{O_2} - E_{\text{LaMnO}_3}^{bulk} \quad (57)$$

is the formation energy of a neutral oxygen vacancy with respect to the calculated energy  $\frac{1}{2}E_{O_2}$  for oxygen atom in the molecule. The variation of oxygen chemical potential

$\Delta\mu_{\text{O}}(T, p_{\text{O}_2})$  due to  $T, p_{\text{O}_2}$  is described by Eq. (50). Similarly, the vacancy formation energy for the surface vacancy can be presented as

$$\Delta g_f^{\text{surf}}(\text{LaMnO}_3: \text{V}_{\text{O}}) = \Delta E_f^{\text{surf}}(\text{LaMnO}_3: \text{V}_{\text{O}}) + \Delta\mu_{\text{O}}(T, p_{\text{O}_2}) \quad (58)$$

where

$$\Delta E_f^{\text{surf}}(\text{LaMnO}_3: \text{V}_{\text{O}}) = \frac{1}{2} \left( E^{\text{slab}}(\text{LaMnO}_3: 2\text{V}_{\text{O}}) + E_{\text{O}_2} - E_{\text{LaMnO}_3}^{\text{slab}} \right) \quad (59)$$

Here we accounted for the fact that we use a symmetrical slab with an oxygen vacancy at each side of the slab. The total energy of such a slab is written as  $E^{\text{slab}}(\text{LaMnO}_3: 2\text{V}_{\text{O}})$ , because the slab has two vacancies.  $E_{\text{LaMnO}_3}^{\text{slab}}$  is the total energy of the slab without defects. The Gibbs free energies of adsorption can be written in a similar way:

$$\Delta g_{\text{ads}}^{\text{surf}}(\text{LaMnO}_3: \text{O}_{\text{ads}}) = \Delta E_{\text{ads}}^{\text{surf}}(\text{LaMnO}_3: \text{O}_{\text{ads}}) - \Delta\mu_{\text{O}} \quad (60)$$

$$\Delta E_{\text{ads}}^{\text{surf}}(\text{LaMnO}_3: \text{O}_{\text{ads}}) = \frac{1}{2} \left( E^{\text{slab}}(\text{LaMnO}_3: 2\text{O}_{\text{ads}}) - E_{\text{O}_2} - E_{\text{LaMnO}_3}^{\text{slab}} \right) \quad (61)$$

and

$$\Delta g_{\text{ads}}^{\text{surf}}(\text{LaMnO}_3: \text{O}_{2, \text{ads}}) = \Delta E_{\text{ads}}^{\text{surf}}(\text{LaMnO}_3: \text{O}_{2, \text{ads}}) - \Delta\mu_{\text{O}_2} \quad (62)$$

$$\Delta E_{\text{ads}}^{\text{surf}}(\text{LaMnO}_3: \text{O}_{2, \text{ads}}) = \frac{1}{2} \left( E^{\text{slab}}(\text{LaMnO}_3: 2\text{O}_{2, \text{ads}}) - 2E_{\text{O}_2} - E_{\text{LaMnO}_3}^{\text{slab}} \right) \quad (63)$$

Here  $\mu_{\text{O}_2} = 2\mu_{\text{O}}$ , and we have to take account for two adsorbed O or O<sub>2</sub> on the symmetrical slab. It is important to remember that the adsorption energy (60) for atomic O species is given relative to half an O<sub>2</sub> molecule, but not with respect to gaseous O atoms. In the present Chapter we will describe the vacancy formation energies and the adsorption energies of O atoms and O<sub>2</sub> molecules obtained with plane wave BS and PW91 functional.

## 4. Results and discussions

### 4.1 Stability of LMO surface terminations: Plane-wave DFT simulations

Based on the results of plane-wave calculations and theoretical considerations described in Section 3, the phase diagrams characterizing stability of different LMO surfaces have been drawn in Figure 1. These diagrams were built for both low-temperature orthorhombic and high-temperature cubic phases. For O-terminated [011] and LaO+O [001] surfaces it was not possible to keep the cubic structure during lattice relaxation. Therefore, we used  $\Omega^{\text{f}}$  values for the orthorhombic phase for both phase diagrams in Figure 1, as it was done, for instance, by Bottin et al., 2003. The calculated input data used for drawing this figure are collected in Tables 2 and 3. Optimized geometries for the slabs can be found in Matrikov et al., 2009. The surface stability regions in the diagrams are limited by the lines 2, 6 and 4. These lines correspond to boundaries where coexistence occurs of LMO with La<sub>2</sub>O<sub>3</sub>, MnO<sub>2</sub> and Mn<sub>3</sub>O<sub>4</sub>, respectively. Because of the DFT deficiencies in describing the relative energies for materials

surface		$\Gamma_{La,O}^i$	$\Gamma_{La,Mn}^i$	$\phi_{La}^i$ (eV/unit cell)	$\phi_{La}^i$ (J/m <sup>2</sup> )
orientation	termination				
[001]	LaO	-1	-0.5	5.38 (5.42)	5.67 (5.70)
[001]	LaO+O	-0.875	-0.5	4.90	5.16
[001]	MnO <sub>2</sub>	1	0.5	-3.02 (-2.93)	-3.18 (-3.08)
[001]	MnO <sub>2</sub> +O	1.125	0.5	-3.16 (-3.53)	-3.33 (-3.72)
[110]	LaMnO	-1	0	6.13 (6.05)	4.56 (4.51)
[110]	O <sub>2</sub>	1	0	-1.16 (-1.26)	-0.86 (-0.94)
[110]	O	0	0	1.09	0.81

Table 2. Parameters defining the surface Gibbs free energies  $\Omega^i$  (Eq. 13) as function of O and Mn chemical potentials : excesses  $\Gamma_{La,O}^i$  and  $\Gamma_{La,Mn}^i$  of O and Mn atoms in the surfaces with respect to La atoms (7), and free energy of formation  $\phi_{La}^i$  (14) at  $\Delta\mu_O = \Delta\mu_{Mn} = 0$  eV for the LaMnO<sub>3</sub> surfaces under consideration. These results produced with plane wave BS and PW91 functional. Values of  $\phi_{La}^i$  without brackets are for the orthorhombic phase, values in brackets are for the cubic phase. Reprinted with permission from Mastrikov et al., 2010 . Copyright 2010 American Chemical Society.

Crystal	Calculated $\delta\mu_O^0$	
	Plane wave BS + PW91	LCAO + B3LYP
La <sub>2</sub> O <sub>3</sub>	-0.41	-0.64
Mn <sub>2</sub> O <sub>3</sub>	-0.87	-0.53
MnO	-0.52	-0.59
MnO <sub>2</sub>	-1.14	-0.09
Mn <sub>3</sub> O <sub>4</sub>	-0.90	-0.15
average	-0.77	-0.40

Table 3. The chemical potential correction (eV), Eq.(54), calculated for different oxides for both employed modeling techniques: (i) plane wave BS and PW91 functional and (ii) LCAO approach based on Gaussian-type atom-centered BS and hybrid B3LYP functional . The last line gives the average correction used in plotting the oxygen chemical potentials of the phase diagrams in Figures 1, 3, and 6.

with different degrees of oxidation, one should treat the obtained data with some precaution. Thus, we highlighted by solid lines the boundaries where metal oxides La<sub>2</sub>O<sub>3</sub> and Mn<sub>2</sub>O<sub>3</sub> with metals in oxidation state 3+ (lines 2 and 5) begin to precipitate in the perovskite. In these oxides, metal oxidation numbers coincide with the oxidation states for the same metals in LaMnO<sub>3</sub>. Right hand side of the diagrams in Figure 1 contains a family of chemical potentials of O atoms (50) as functions of temperature and partial pressure of oxygen gas. This part of the figures allows us to translate easily-measurable external parameters (temperature and oxygen gas pressure) into oxygen chemical potential, which is one of the variables determining explicitly the SGFE. Using this part of the figures, we can relate points on the phase diagrams with the conditions under which experiments and/or industrial processes occur. To do this, one can just to draw a vertical line for a given temperature

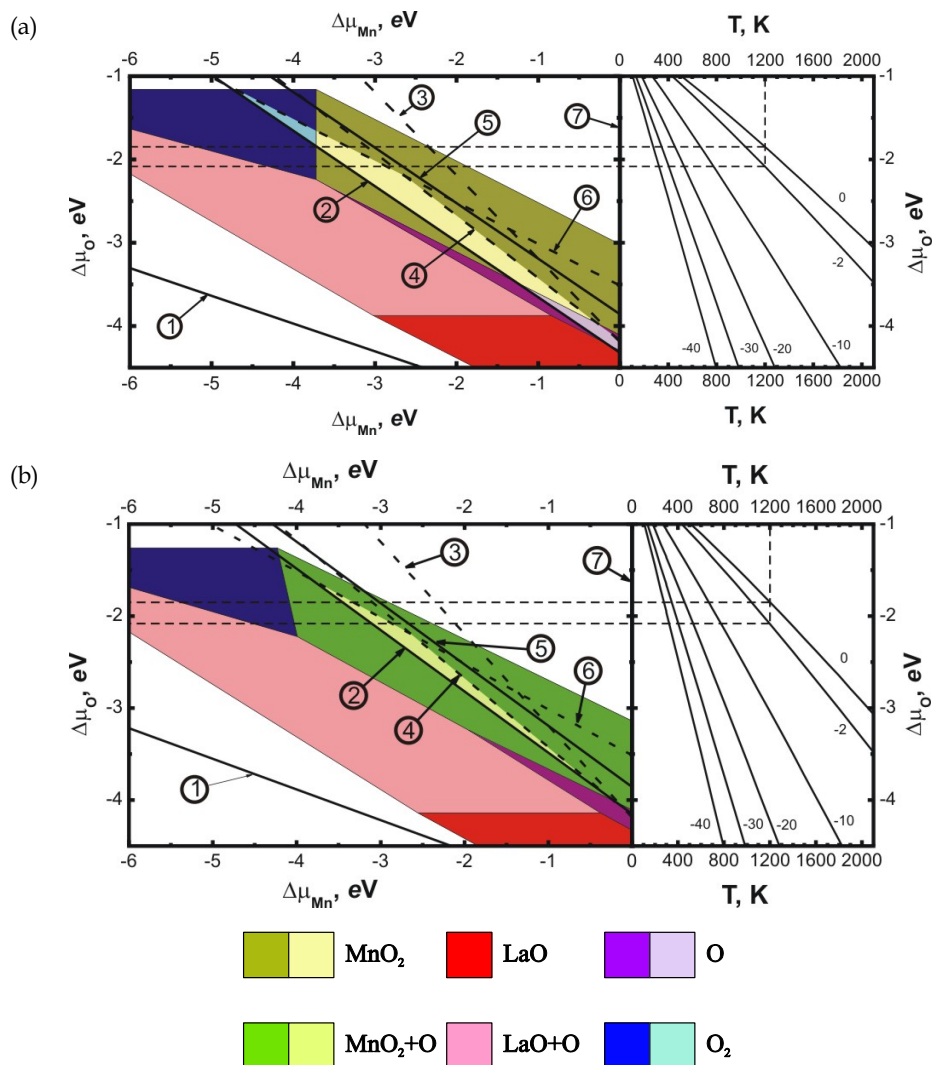


Fig. 1. Phase diagrams calculated with plane wave BS and PW91 GGA functional: The regions of stability of LaMnO<sub>3</sub> surfaces with different terminations (LaO- and MnO<sub>2</sub>-terminated [001] surfaces without and with adsorbed O atom, O<sub>2</sub>- and O-terminated [011] surfaces) for both orthorhombic (a) and cubic (b) phases as functions of manganese and oxygen atoms chemical potential variations. Parameters for all lines on the left side of the figures are collected in Table 2. The encircled numbers point to lines, where metals or their oxides begin to precipitate: (1) metal La, (2) La<sub>2</sub>O<sub>3</sub>, (3) MnO, (4) Mn<sub>3</sub>O<sub>4</sub>, (5) Mn<sub>2</sub>O<sub>3</sub>, (6) MnO<sub>2</sub>, and (7) metal Mn. The right side of the figures contains a family of  $\Delta\mu_{\text{O}}$  as functions of temperature at various oxygen gas pressures according to Eq. (50) and Table 1. The labels  $m$  on the lines specifies the pressure according to:  $p_{\text{O}_2} = 10^m$  atm. Reprinted with permission from Mastrikov et al., 2010. Copyright 2010 American Chemical Society.

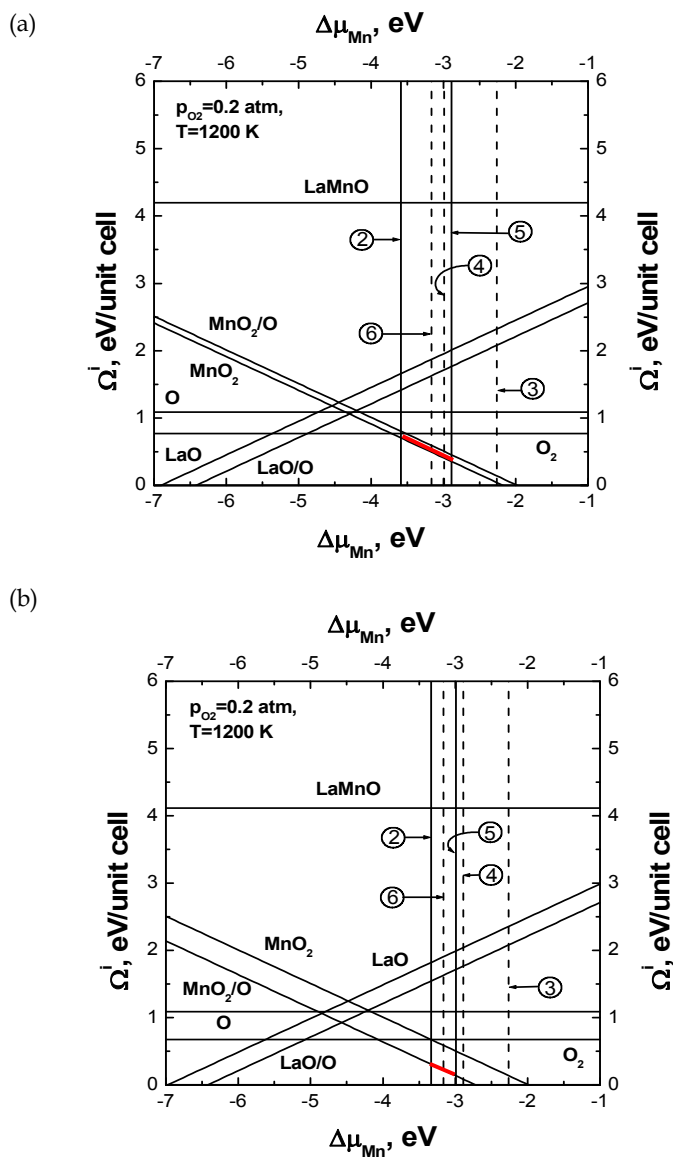


Fig. 2. Surface Gibbs Free Energies  $\Omega^i$  for  $\text{LaMnO}_3$  in (a) orthorhombic and (b) cubic phases as functions of  $\Delta\mu_{\text{Mn}}$  at  $T = 1200 \text{ K}$  and  $p_{\text{O}_2} = 0.2 \text{ atm}$ . Line numbers are the same as in Figure 4.1. The red lines indicate the most stable surface in the stability window between the precipitation lines for  $\text{La}_2\text{O}_3$  and  $\text{Mn}_2\text{O}_3$ . Reprinted with permission from Mastrov et al., 2010. Copyright 2010 American Chemical Society.

and its crossings with lines corresponding to different gas pressures creates a pressure scale for this particular temperature. This can replace the axis for oxygen chemical potential. Alternatively, moving along the lines for chemical potential at a particular gas pressure, we can study the phase behavior with temperature. Figure 1 shows such a consideration for  $T=1200$  K which is a typical condition for SOFC operations. We marked on these phase diagrams the most important range of oxygen gas partial pressures (between  $pO_2 = 0.2$  atm. and 1 atm). Oxygen-rich conditions with a larger O chemical potential correspond to higher oxygen gas partial pressures and/or lower temperatures; in turn, oxygen-pure conditions with the lower O chemical potentials correspond to smaller oxygen gas partial pressures and/or higher temperatures.

Consistent positioning of these experimental curves with respect to our computed stability diagram requires also the correction described by Eq. (54). When drawing the right side of Figure 1, we used the correction of  $-0.77$  eV (Table 3) calculated as average of a series of different oxides. It was calculated using the same set of oxides, which precipitation is considered in our plane-wave modeling. Both the values and the scattering ( $\pm 0.37$  eV) of calculated corrections are much larger than in similar studies (Heifets et al., 2007a, 2007b; Johnston et al., 2004; Reuter & Schefer, 2001a) for non-magnetic oxides (e.g. SrTiO<sub>3</sub>). Here we consider manganese oxides which are spin-polarized solids. Besides, we included several Mn oxides with various oxidation states. This is a typical situation where DFT calculations face well known problems. The scattering of the correction magnitudes provides an estimate of uncertainty in positioning of the chemical potentials for O atoms on the right side of the phase diagrams.

Figure 2 shows cross sections of the phase diagrams at  $T = 1200$  K and  $pO_2 = 0.2$  atm., i.e. in the range of typical SOFC operational conditions. Correspondingly, at the cross sections of the diagrams (Figure 2), the stability region lies between lines 2 and 6. This figure helps to clarify behavior of the SGFEs for surfaces with various terminations.

As it can be seen from Figures 1 and 2, under fuel cell operational conditions in both LMO phases the MnO<sub>2</sub>-terminated [001] surface is the most stable one. In the orthorhombic phase it is the clean MnO<sub>2</sub>-terminated surface, whereas in the high-temperature cubic phase the most stable surface contains adsorbed O atoms. This indicates that under identical conditions higher O adsorbate coverage is expected for the cubic LMO phase. Modeling with plane-wave BS and PW91 functional suggest that, when LMO crystal is heated, precipitation of La<sub>2</sub>O<sub>3</sub> or Mn<sub>3</sub>O<sub>4</sub> occurs, depending on chemical potentials variations during heating.

#### 4.2 Stability of LMO surface terminations: LCAO simulations

Calculations performed within the LCAO approach combined with hybrid B3LYP functional were also employed in order to draw the phase diagram for stability of different LMO surface terminations (Figure 3). These calculations were carried out for a cubic phase and A-AFM magnetic ordering, where spins have the same orientations in the planes parallel to the surfaces of the slabs, but have opposite directions in neighbor planes. The comparison of stability shown in this figure includes only two primary candidates for the stable surfaces: LaO- and MnO<sub>2</sub>-terminated (001) surfaces. The stability range is limited by lines 2, 3, and 5, which correspond to precipitation of La<sub>2</sub>O<sub>3</sub>, MnO, and Mn<sub>2</sub>O<sub>3</sub>. These are substantially different oxides than suggested above in computations performed with plane-wave BS and PW91 functional. Indeed, the gap between precipitation of La<sub>2</sub>O<sub>3</sub> and Mn<sub>2</sub>O<sub>3</sub> shifted down significantly. Now the boundary between stability regions for LaO- and

MnO<sub>2</sub>-terminated surfaces crosses the gap where LMO is stable, while PW91-GGA calculations described above and by Mastrikov et al., 2009, 2010 suggested that only the MnO<sub>2</sub>-terminated surface was stable. In calculations with hybrid B3LYP functional the MnO<sub>2</sub>-terminated surface seems to be stable, up to SOFC operational temperatures (1200 K) under ambient oxygen gas partial pressures ( $p_{O_2}=0.2$  atm.). Above this temperature LaO-terminated surface gradually becomes more stable in the larger range in LMO crystal stability region until at ~1900 K it becomes the only stable surface. A precipitation of MnO or La<sub>2</sub>O<sub>3</sub> has to occur while LMO crystal is heated.

Positioning the family of O atom chemical potential curves on the right side of Figure 3 was done in the same way as for Figure 1, but using LCAO calculations with hybrid B3LYP functional. The averaged correction  $\delta\mu_O^0$  (54) in this case is noticeably smaller ( $-0.40$  eV) than it was for PW91-GGA functional. However, deviations of this correction from its average value ( $\pm 0.3$  eV) is still large. This fact likely comes from the DFT difficulties, taking place even within hybrid functionals for spin-polarized systems. For diamagnetic systems, for instance SrTiO<sub>3</sub>, such deviation drops down, from  $\sim 0.25$  eV in LDA calculations (Johnston et al., 2004) to  $\sim 0.03$  eV in calculations (Heifets et al., 2007b) with the hybrid functional.

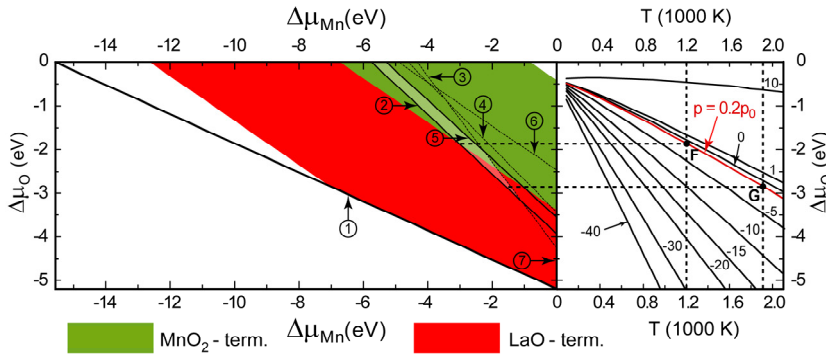


Fig. 3. Thermodynamic LaMnO<sub>3</sub> [001] surface stability diagram as a function of O and Mn chemical potentials. It compares stabilities of both LaO- and MnO<sub>2</sub>-terminated [001] surfaces and accounts for precipitation conditions for (1) metal La, (2) La<sub>2</sub>O<sub>3</sub>, (3) MnO, (4) Mn<sub>3</sub>O<sub>4</sub>, (5) Mn<sub>2</sub>O<sub>3</sub>, (6) MnO<sub>2</sub>, and (7) metal Mn, the same set as at Figures 1 and 2. The stable region is shown as lightened area between precipitation lines 2,3, and 5. The right side shows a family of oxygen chemical potentials under different conditions. The label  $m$  indicates the O<sub>2</sub> gas partial pressure:  $10^m$  atm. Red line corresponds to oxygen partial pressure  $p=0.2p_0$  as in the ambient atmosphere.

surface		$\Gamma_{La,O}^i$	$\Gamma_{La,Mn}^i$	$\phi_{La}^i$ (eV/unit cell)	$\phi_{La}^i$ (J/m <sup>2</sup> )
orientation	termination				
[001]	LaO	-1	-0.5	6.32	6.46
[001]	MnO <sub>2</sub>	1	0.5	-0.42	-0.43

Table 4. Parameters defining the surface Gibbs free energies  $\Omega^i$  (Eq. 13) and used to build diagram in Figure 3. The same as Table 2, but for the cubic phase of LMO only and produced with LCAO approach and hybrid B3LYP functional.



### 4.3 Stability of surface terminations for LSM: LCAO simulations

Since the SGFEs for LSM surfaces depend now on three variables, it is a little more complicated to draw corresponding phase diagrams. Therefore, we have drawn only several sections for the most interesting parts of the phase diagram for bulk concentration of Sr atoms  $x_b = 1/8$ . Thus, Figure 4 shows the section of surface stability phase diagram under ambient oxygen gas partial pressure  $pO_2 = 0.2$  atm. and three various temperatures: a) 300 K - room temperature (RT), b) 1100 K, which is approximately the SOFC operational temperature, and c) 1500 K, which is close to sintering temperatures. We compared several terminations of LSM (100) surfaces:  $MnO_2$ ,  $La_{1-x_s}Sr_{x_s}O$ , in the last case concentrations of Sr atoms in the surface layer were varied:  $x_s = 0.25, 0.5, 0.75$  and 1 (which simulates a segregation effect). Only three terminations appear at the shown sections:  $MnO_2$ ,  $La_{0.75}Sr_{0.25}O$ , and SrO. Here we accounted for precipitation of metals (La, Mn, Sr),  $Mn_2O_3$  and  $La_2O_3$  oxides, and  $LaMnO_3$  and  $SrMnO_3$  perovskites. These sections of the surface phase diagram indicate that the LSM crystal can be stable only within a small quadrangle region in the presented sections. At low, room temperature two of considered terminations -  $MnO_2$  and  $La_{0.75}Sr_{0.25}O$  - are stable. At the higher temperatures  $La_{0.75}Sr_{0.25}O$ -terminated surface gradually occupies a larger portion of the stability region. Already at SOFC operational temperatures ( $T \approx 1100$  K) this termination becomes stable in the entire stability region. Thus, Sr dopant atoms in LSM cause a relative stabilization of the

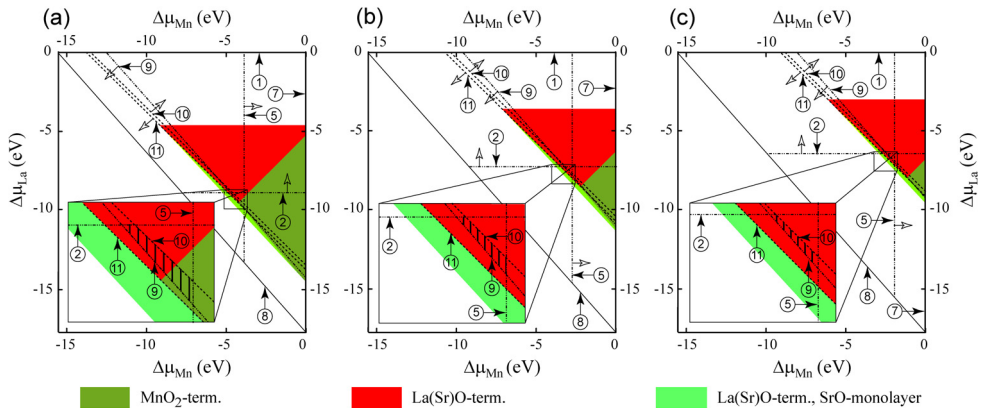


Fig. 4. Sections of surface stability diagram for LSM (001) surface structures for  $O_2$  partial pressure  $p = 0.2p_0$  and temperatures of (a) 300 K (RT), (b) 1100 K (SOFC operational temperature), and (c) 1500 K (sintering temperature) (Piskunov et al., 2008). The region, where LSM ( $x_b = 1/8$ ) is stable, is the shaded area between  $LaMnO_3$ ,  $La_2O_3$ ,  $Mn_2O_3$ , and SrO precipitation lines. The numbers from 1 to 11 in the circles indicate precipitation lines for (1) La, (2)  $La_2O_3$ , (3) MnO, (4)  $Mn_3O_4$ , (5)  $Mn_2O_3$ , (6)  $MnO_2$ , (7) Mn, (8) Sr, (9) SrO, (10)  $LaMnO_3$ , (11)  $SrMnO_3$ . (Some of the mentioned oxides are not considered in this Figure, but the numbering is designed to keep consistency of notations between figures.) Hollow arrows indicate the sides from respective precipitation lines where the precipitation occurs. Insets show magnified areas with the region of LSM stability (a shaded quadrangle). Reprinted with permission from Piskunov et al., 2008. Copyright 2008 American Physical Society.

$La_{1-x_s}Sr_{x_s}O$  - terminated surface with respect to the  $MnO_2$ -terminated surface. However, as soon as Sr concentration  $x_s$  at the  $La_{1-x_s}Sr_{x_s}O$  -terminated surface becomes 0.5 or larger due to Sr segregation, such a surface becomes unstable.

For better understanding changes in the surface stability with temperature, we have drawn two additional cross-sections along the precipitation lines for SrO and  $LaMnO_3$  at  $p_{O_2}=0.2$  atm. These cross-sections are presented in Figure 5. It can be seen here that upon heating the  $MnO_2$ -terminated surface leaves the stability region and becomes replaced by the  $La_{0.75}Sr_{0.25}O$ -terminated surface. As heating continues, precipitation of  $La_2O_3$  or MnO begins. This is consistent with experimental observations by Kuo et al., 1989. A similar degradation process without Sr doping would require stronger overheating or very strongly reducing conditions. Detailed LCAO hybrid functional calculations of oxygen atom adsorption are necessary (see preliminary results in Piskunov et al, 2011), in order to check PW91-GGA prediction (discussed in previous subsection) that the  $MnO_2$ -terminated surface is stabilized by adsorbed oxygen atoms.

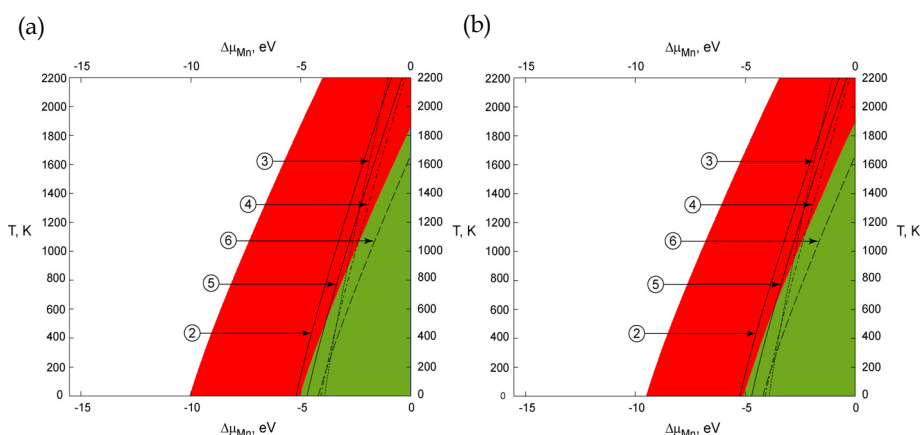


Fig. 5. LCAO calculated cross-sections of surface stability diagram for LSM (001) surface structures along (a) SrO and (b)  $LaMnO_3$  precipitation lines for  $O_2$  partial pressure  $p=0.2p_0$ . Meaning of colors (terminations) and numbers (correspond to precipitation lines) are the same as in Figure 4.

#### 4.4 Oxygen adsorption and vacancy formation in LMO

As shown above, the  $MnO_2$ -terminated (001) surface of  $LaMnO_3$  appears to be the most stable one. Therefore, we optimized the atomic structure of surface oxygen vacancies, as well as O atoms and  $O_2$  molecules adsorbed at different sites on this surface. For a comparison we also optimized the structure of oxygen vacancies in the  $LaMnO_3$  bulk and at the LaO-terminated [001] surface. These simulations were performed using plane wave BS and PW91 functional. Details of the atomic position optimization are described by Matrikov et al., 2010. In this Chapter, we limit our discussion only to the energies of different adsorbed species and vacancies and thermodynamic consideration of the relevant processes. Note that some adsorbed species have tilted geometry. For example, the lowest energy for the adsorbed  $O_2$  molecule on  $MnO_2$ -terminated surface is atop of Mn ion with the angle between O-O bond and Mn-O direction being  $\sim 50^\circ$ .

The adsorption energies for O atoms ( $\Delta E_{ads}^{surf}(LaMnO_3:O_{ads})$ ) and O<sub>2</sub> molecules ( $\Delta E_{ads}^{surf}(LaMnO_3:O_{2,ads})$ ), as well as the surface and bulk formation energies ( $\Delta E_f^{surf}(LaMnO_3:V_O)$  and  $\Delta E_f^{bulk}(LaMnO_3:V_O)$ ) for oxygen vacancies are collected in Table 5. For a classification of different molecular oxygen species we considered atomic charges and the O-O bond length. The data in Table 5 suggest that atomic adsorption of O atoms is energetically more preferable than adsorption of O<sub>2</sub> molecule. In both cases the best adsorption site for both O atom and O<sub>2</sub> molecule on MnO<sub>2</sub>-terminated surface is on top of surface Mn ion. Oxygen vacancies have smaller formation energy on MnO<sub>2</sub>-terminated surface than in the bulk suggesting vacancy segregation towards this surface. In contrary, much more energy is required to create an oxygen vacancy on LaO-terminated surface.

Label of configuration	$-E_f$ or $E_{ads}$ , eV	Charge, e <sub>0</sub>	O-O distance, Å	"chemical assignment"
		0.	1.30	gaseous O <sub>2</sub>
I	-1.1 <sup>a)</sup>	-0.42	1.36	tilted superoxide atop one Mn <sup>surf</sup>
II	-0.9 <sup>a)</sup>	-0.65	1.42	horizontal peroxide atop one Mn <sup>surf</sup>
III	-0.9 <sup>a)</sup>	-0.69	1.41	horiz. peroxide atop O <sup>surf</sup>
VII	-0.5 <sup>a)</sup>	-0.84	1.62	TS of dissociation without V <sub>O</sub> , atop O <sup>surf</sup> and bridging two Mn <sup>surf</sup>
IV	-2.8 <sup>a)</sup>	-1.19	1.50	tilted peroxide in V <sub>O</sub>
V	-2.4 <sup>a)</sup>	-1.25	1.50	"vertical peroxide" in V <sub>O</sub>
V	+0.9 <sup>b)</sup>	-1.25	1.50	TS of O <sup>-</sup> diffusion along surface
VI	-1.1 <sup>b)</sup>	-0.62		O <sup>-</sup> adsorbed atop Mn
VI'	-1.8 <sup>b)</sup>			O <sup>-</sup> adsorbed next to a surface vacancy
VIII	-3.3 <sup>b)</sup>	-1.19		O ion in MnO <sub>2</sub> [001] surface layer
IX	-4.3 <sup>b)</sup>	-1.25		bulk O ion
X	-5.1 <sup>b)</sup>	-1.32		O ion in LaO[001] surface layer

Table 5. Bond lengths, Bader charges and "chemical assignment" of the different oxygen species. Experimental O-O bond lengths (NIST, 2010) for comparison: gaseous O<sub>2</sub> 1.21 Å, hydrogen superoxide radical HO<sub>2</sub> 1.33 Å, hydrogen peroxide H<sub>2</sub>O<sub>2</sub> 1.48 Å. TS = transition state. Energies (compare Figure 4.6; for adsorbate coverage of 12.5 %): <sup>a)</sup> relative to gaseous O<sub>2</sub> in triplet state over defect-free surface, <sup>b)</sup> relative to half a gaseous O<sub>2</sub> over defect-free surface.

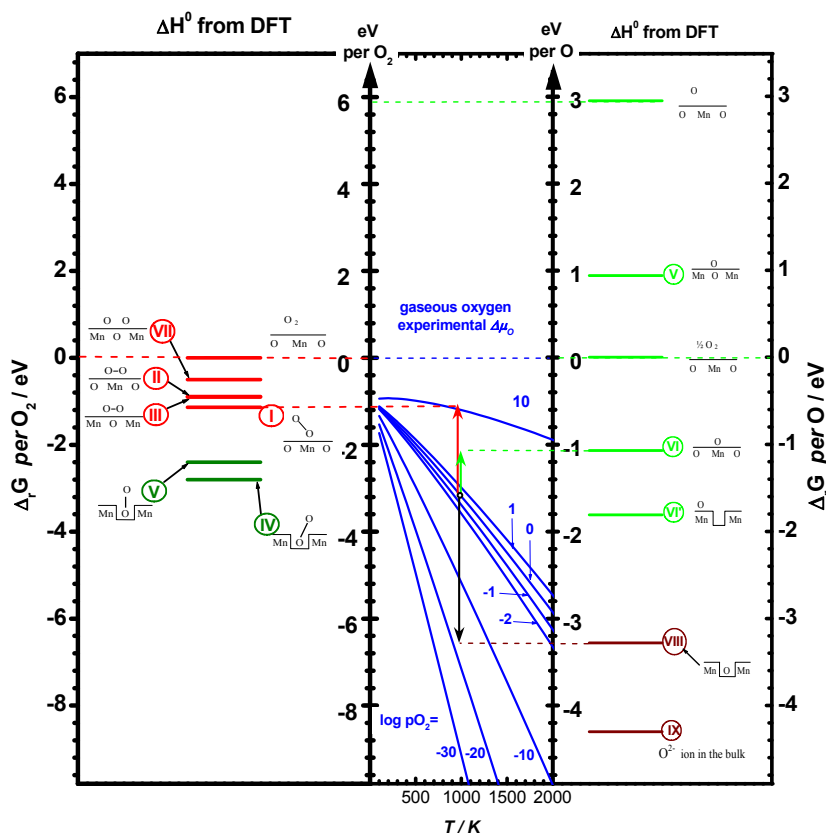


Fig. 6. Spectrum of possible “one-particle” states, where “particles” are O atoms (right panel) and O<sub>2</sub> molecules (left panel). Each level in these panels corresponds to relative energies ( $\Delta E_T$ ) of different molecular and atomic species occurring during oxygen incorporation reaction on the MnO<sub>2</sub>[001]-terminated surface of LaMnO<sub>3</sub>, cf. Table 5. The axes on the left and right give the energy  $\Delta E_T$  relative to resting O<sub>2</sub> molecule away from the surface (on the left) or an atom in such O<sub>2</sub> molecule (on the right). In the ground state of the crystal all lattice sites in crystal bulk (states X) and I surface (states VIII) are occupied (red levels) and the rest of the “one-particle” states vacant. The numbers at levels correspond to the numbers assigned to respective states in Table 5. The highest level on the right panel corresponds to a free (not in a molecule) O atom away from the crystal. The central panel shows the experimental  $T$ - and  $p_{\text{O}_2}$ -dependence of the Gibbs energy of gaseous O<sub>2</sub> (Table 1 and Eq.(50)), its energy scale refers to an O<sub>2</sub> molecule on the left and to an O atom in an O<sub>2</sub> molecule on the right. The labels  $m$  on the lines represent the pressure:  $p_{\text{O}_2}=10^m$  atm. The arrows indicate various Gibbs reaction energies due to moving of a “particle” between crystal and gas: red = formation of adsorbed superoxide O<sub>2</sub><sup>-</sup> on defect-free surface; green = formation of adsorbed O- atop Mn on defect-free surface; black = incorporation of oxygen into a surface oxygen vacancy.

The collected energies allow us to draw the diagram shown in Figure 6. This diagram is based on a standard model of “non-interacting particles”, where “particles” are O atoms and O<sub>2</sub> molecules in different positions. The energy levels drawn at the side panels represent single-particle energies corresponding to bringing a particle to a given position at the surface or in the bulk. The left hand panel refers to bringing a free gas-phase O<sub>2</sub> molecule to the crystal surface. Similarly, the right hand panel refers to taking an O atom from a free O<sub>2</sub> molecule and placing it on the crystal surface. These processes include also placing of an atom or a molecule into surface vacancies: this is a process inverse to the formation of a vacancy. Therefore, to place the corresponding energy level (at right hand panel), one has to use the vacancy formation energy with the opposite sign. A similar logic was applied in placing the energy level for bringing an oxygen atom into vacancy in the bulk. Such an O atom in a vacancy becomes actually a regular O atom in the crystal lattice (wherever, in the bulk or in the surface). Therefore, the energies of such states can be considered as those for an O atom in the bulk or on the surface. In the ground state of the crystal all lattice sites (states VIII, IX and X) are occupied and all other states vacant.

The variation of the oxygen chemical potential is drawn in the central panel as a function of temperature for several gas partial pressures. These curves are drawn in the same way as similar lines on the right hand side in Figure 1, including the offset defined by Eq. (54). Because the energy scale at the left panel is twice as large as at the right panel, the same curves represent variations either in the chemical potentials for an O<sub>2</sub> molecule, if they are referred to the left panel, or for O atom, if they are referred to the right one. In such an arrangement, the diagram in Figure 6 can be used to represent the Gibbs energies for reactions of exchange with O atoms or O<sub>2</sub> molecules between oxygen gas and both crystal bulk and surfaces. For example, red arrow represents an adsorption of an O<sub>2</sub> molecule atop surface Mn ion in the tilted position (configuration I) from oxygen gas under partial pressure  $p_{\text{O}_2}=1$  atm. and  $T=1000$  K. The Gibbs free energy of corresponding reaction can be obtained by subtracting the energy of the initial state from that of the final state. For the reaction described by the red arrow this energy indeed corresponds to the adsorption energy for O<sub>2</sub> molecule. Similarly, the green arrow describes the adsorption of O atom atop Mn ion in MnO<sub>2</sub>-terminated surface. Lastly, the black arrow describes incorporation of an O atom into a surface oxygen vacancy. In the latter case, an arrow with opposite direction corresponds to the formation of a surface oxygen vacancy, as it can be confirmed by a comparison with Eqs. (60, 61).

The diagram in Figure 6 is very suitable way of a graphical representation of the exchange between a gas and a crystal with various species and the analysis of corresponding processes. For a given temperature and oxygen partial pressure this diagram allows one to read the Gibbs reaction energy of a process and thus to obtain its mass action constant. As an example, let us discuss some processes under typical fuel cell conditions of  $T = 1000$  K and  $p_{\text{O}_2} = 1$  atm. The formation of molecular adsorbates (superoxide I = red arrow, and peroxide II) is endergonic by  $\Delta_r G \approx +2$  eV per O<sub>2</sub> since the entropy loss overcompensates the electronic energy gain. Even the formation of adsorbed atomic O<sup>-</sup> (species VI, green arrow) is still slightly endergonic, by  $\Delta_r G \approx +0.5$  eV per O, what leads the low adsorbate coverage under SOFC conditions. Only the oxygen incorporation into a surface vacancy (black arrow) is strongly exergonic, by  $\Delta_r G \approx -1.7$  eV per O (i.e. the inverse process, surface oxygen vacancy formation, is endergonic by +1.7 eV). Also, changes in temperature and/or partial pressure can change the sign of the reaction energy. To give an example: while oxygen atom adsorption is exothermic here, it changes from exergonic at low temperatures and/or high partial pressures to endergonic at higher temperatures and/or lower pressures.

## 5. Conclusions and perspectives

In this Chapter, combining *ab initio* calculations of the atomic and electronic structure with chemical thermodynamics, we have described how to predict surface stabilities with different orientations and terminations for  $ABO_3$  perovskite materials and solid solutions. We considered also adsorption and formation of vacancies in the bulk and at the surfaces. The input data for such thermodynamic analyses are available from standard DFT calculations. We neglected vibrational contributions to the Gibbs free energies in present simulations, since there are good arguments that these contributions are quite small. However, the latest versions of the computer codes used here (VASP and CRYSTAL) are capable to perform the calculations of phonon spectra, which can be used to include vibration contributions into thermodynamic potentials. Despite currently such surface phonon calculations are computationally very expensive, with development of new algorithms and faster computers such analysis in near future could become a routine practice.

We have applied the described techniques to experimentally well studied SOFC cathode materials (LMO and LSM). The energies calculated by DFT techniques allow us to draw the surface stability diagrams showing that under the SOFC operational conditions  $MnO_2$ -terminated surface of defect/dopant-free LMO is the most stable. The  $MnO_2$ -terminated surface with adsorbed O atoms becomes more stable in a cubic phase of LMO. (The analysis based on results of calculations with hybrid functional gives higher precision of the obtained diagrams). Introduction of Sr dopant into LMO leads to a significant decrease of stability of  $MnO_2$ -terminated surface with respect to stability of Sr-containing  $La_{1-x_s}Sr_{x_s}O$ -terminated

surface. At elevated temperatures only the latter one can be found within an entire region of LSM stability which definitely should affect the LSM cathode performance due to our prediction that oxygen vacancies easily segregate from LMO bulk towards  $MnO_2$ -terminated surface but not to the LaO-terminated one (Mastrikov et al, 2010).

Thermodynamic consideration of oxygen adsorption and formation of surface vacancies allowed us to move beyond the usual analysis of purely DFT electronic energy differences in these processes at zero K and to describe changes in the Gibbs free energies under realistic environmental conditions (high temperature and partial pressure of oxygen gas). In particular, we have shown that an oxygen adsorption from the gas phase is exergonic at low temperatures, but becomes endergonic at SOFC operational temperatures. Consideration of the energy differences between O atoms and  $O_2$  molecules in the gas phase and at surfaces makes it possible to determine theoretically preferred adsorption sites, adsorption energies and formation energies for vacancies (e.g. Kotomin et al., 2008, Mastrikov et al., 2010, and Piskunov et al, 2011). Thus the obtained data can be used for analysis of the kinetics of chemical reactions and investigation of their mechanisms. For example, we performed such an analysis (Mastrikov et al., 2010) in the study of oxygen reduction and incorporation into the LMO surfaces.

Note also that under realistic experimental conditions in a multinary crystal typically only one component is reversibly exchangeable so that only one chemical potential can be varied in-situ. Typically the others are varied under preparation conditions, and then so-established sublattice stoichiometry is frozen under experimental conditions. This complex interplay of in-situ and ex-situ parameters has been discussed by Maier, 2003.

Various questions, where thermodynamic approach is necessary, are still open. One example is Sr segregation towards LSM surfaces which was experimentally observed e.g. by

Fister et al., 2008 and Herger et al., 2008. In the study by Herger et al., 2008, the Sr segregation energy was estimated to be ~0.16 eV while preliminary theoretical estimate gave the segregation energy ~0.54 eV (Piskunov et al., 2008). Another question is, how vacancy formation and oxygen adsorption energies depend on surface Sr concentration. Lastly, our present simplified simulations assume that LMO and LSM have large flat surfaces. However, one can expect in reality much more rough surface structure containing many facets and steps. This is important also because SOFC cathode materials are polycrystalline. Therefore, there is a necessity to check, which kinds of steps and facets are the most stable ones and likely to exist at LMO and LSM surfaces and how their presence affects adsorption of oxygen atoms and molecules, vacancy formation and reactions occurring at the surfaces.

## 6. Acknowledgements

EH is indebted to the Max Planck Institute for financial support through the honorary contract. EK, YM thank the EC FP7 NASA-OTM project (grant agreement N 228701) for partial financial support. S.P. is thankful for the financial support through the ESF project Nr. 2009/0216/1DP/1.1.1.2.0/09/APIA/VIAA/044. This study was supported by a grant of computer time at the EMS Laboratory at PNNL (Project No 42498). Authors are greatly indebted to R. Merkle, J. Fleig, R.A. Evarestov, D. Gryaznov, C. Noguera, and M. W. Finnis for many stimulating discussions.

## 7. References

- Becke, A. D. (1993). A new mixing of Hartree-Fock and local density-functional theories. *Journal Chemical Physics*, Vol.98, No.2, (January 1993), pp. 1372-1377, ISSN: 1089-7690
- Bloch, P. E. (1994). *Physical Review B*, Vol. 50, No.24, (December 1994), pp.17953- 17979, ISSN: 1550-235X
- Bottin, F.; Finocchi, F. & Noguera, C. (2003). Stability and electronic structure of the (1×1) SrTiO<sub>3</sub>(110) polar surfaces by first-principles calculations. *Physical Review B*, Vol.68, No.3, (July 2003), a.035418, ISSN: 1550-235X
- Chase, M.W. (1998). *NIST-JANAF Thermochemical Tables*, American Chemical Society, Washington, DC, USA, ISBN: 1563968312
- Coey, J. M. D.; Viret M. & von Molnár, S. (1999). Mixed-valence manganites. *Advances in Physics*, Vol.48, No.2, (November 2010), pp. 167-293, ISSN 1460-6976
- Dolg, M.; Stoll, H.; Savin, A. & Preuss, H. (1989). Energy-adjusted pseudopotentials for the rare earth elements. *Theoretica Chimica Acta*, Vol.75, No.3, (December 1989), pp. 173-194, ISSN: 1432-2234
- Dovesi R.; Saunders, V.R.; Roetti, C.; Orlando, R.; Zicovich-Wilson, C. M.; Pascale, F.; Civalieri, B.; Doll, K.; Harrison, N.M.; Bush, I.J.; D'Arco, Ph. & M. Llunell (2007) *CRYSTAL06 User's Manual*, University of Torino, Torino, Italy, Retrieved from <http://www.crystal.unito.it/>
- Evarestov, R. A.; Kotomin, E. A.; Matrikov, Yu.; Gryaznov, D.; Heifets, E. & Maier, J. (2005). Comparative density-functional LCAO and plane-wave calculations of LaMnO<sub>3</sub> surfaces. *Physical Review B*, Vol. 72, No.21, (December 2005), a. 214411, ISSN: 1550-235X

- Fister, T.T.; Fong, D.D.; Eastman, J.A.; Baldo, P.M.; Highland, M.J.; Fuoss, P.H.; Balasubramaniam, K. R.; Meador, J. C. & Salvador, P. A. (2008). *In situ* characterization of strontium surface segregation in epitaxial  $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$  thin films as a function of oxygen partial pressure *Applied Physics Letters*, Vol. 93, No. 15, (October 2008), a. 151904, ISSN 1077-3118
- Fleig J.; Kreuer, K.D. & Maier, J. (2003). Ceramic Fuel Cells, In: *Handbook of Advanced Ceramics: Processing and Their Applications, Vol. II*, S. Somiya, (ed.), F. Aldinger (Ed.), R. M. Spriggs (Ed.), K. Uchino (Ed.), K. Koumoto (Ed.), M. Kaneno (Ed.), pp. 59–105, Elsevier, ISBN 978-0-12-654640-8, Amsterdam, Netherlands.
- Gibbs J.W. (1948). *Collected Works*, Yale University Press, New Haven, USA
- Haghiri-Gosnet, A. & Renard, J. (2003). CMR manganites: physics, thin films and devices *Journal of Physics D*, Vol.36, No.8, (April 2003), pp. R127-R150, ISSN 1361-6463
- Hay P. J. & Wadt, W. R. (1985). Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core. *Journal Chemical Physics*, Vol.82, No.1, (January 1985), pp. 299-310, ISSN: 1089-7690
- Heifets E.; Ho, J. & Merinov, B. (2007a). Density functional simulation of the  $\text{BaZrO}_3$  (011) surface structure. *Physical Review B*, Vol. 75, No.15, (April 2007) a. 155431, ISSN: 1550-235X
- Heifets, E.; Piskunov, S.; Kotomin, E. A.; Zhukovskii, Y. F. & Ellis, D., (2007b). Electronic structure and thermodynamic stability of double-layered  $\text{SrTiO}_3$  (001) surfaces: *Ab initio* simulations. *Physical Review B*, Vol.75, No.11, (March 2007), a. 115417, ISSN: 1550-235X
- Herger, R.; Willmott, P. R.; Schlepütz, C. M.; Björck, M.; Pauli, S. A.; Martoccia, D.; Patterson, B. D.; Kumah, D.; Clarke, R.; Yacoby, Y. & Döbeli, M. (2008). Structure determination of monolayer-by-monolayer grown  $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$  thin films and the onset of magnetoresistance. *Physical Review B*, Vol.77, No.8, (February 2008), a. 085401, ISSN: 1550-235X
- Johnston, K.; Castell, M. R.; Paxton, A. T. & Finnis, M. W. (2004).  $\text{SrTiO}_3(001)(2\times 1)$  First-principles calculations of surface energy and reconstructions: atomic structure compared with scanning tunneling microscopy images. *Physical Review B*, Vol.70, No.8, (August 2004), a. 085415 ISSN: 1550-235X
- Kotomin, E.A.; Mastrov, Yu.; Heifets, E. & Maier, J. (2008). Adsorption of atomic and molecular oxygen on the  $\text{LaMnO}_3$  surfaces: *ab initio* supercell calculations and thermodynamics, *Physical Chemistry Chemical Physics*, Vol. 10, No. 7, (June 2008), pp. 4644-4649, ISSN: 1463-9084
- Kresse, G. & Hafner, J. (1993). Ab initio molecular dynamics for liquid metals. *Physical Review B*, Vol. 47, No.1, (January 1993), pp. 558-561,
- Kresse, G. & Furthmüller, J. (1996). *Physical Review B*, Vol. 54, No.16, (October 1996), pp.11169-11186, ISSN: 1550-235X
- Kresse, G. & Joubert, D. (1999). From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, Vol. 59, No.3, (January 1999), pp. 1758-1775, ISSN: 1550-235X
- Kresse, G.; Marsman, M. & Furthmüller, J. (2011) *VASP, the Guide*, University of Vienna, Austria, Available from <http://cms.mpi.univie.ac.at/vasp/vasp/vasp.html>



- Kuo J. H.; Anderson, H. U. & Sparlin D. M. (1989). Oxidation-reduction behavior of undoped and Sr-doped LaMnO<sub>3</sub> nonstoichiometry and defect structure. *Journal of Solid State Chemistry*, Vol.83, No.1, (November 1989), pp. 52-60, ISSN: 0022-4596
- Linstrom, P.J.(ed.) & Mallard, W.G.(ed.) (2003). *NIST Chemistry WebBook, NIST Standard Reference Database No. 69*, National Institute of Standards and Technology, Gaithersburg, MD, USA, Available from <http://webbook.nist.gov/chemistry/>
- Maier J. (2003) Complex oxides: high temperature defect chemistry vs low temperature defect chemistry. *Physical Chemistry Chemical Physics*, Vol. 5, No.11, (April 2003), pp. 2164-2173, ISSN: 1463-9084
- Mastrikov, Yu. A.; Heifets, E.; Kotomin, E. A. & Maier, J. (2009). Atomic, electronic and thermodynamic properties of cubic and orthorhombic LaMnO<sub>3</sub> surfaces. *Surface Science*, Vol. 603, No.2, (January 2009), pp. 326-335, ISSN: 00396028
- Mastrikov, Yu. A.; Merkle, R.; Heifets, E.; Kotomin, E. A. & Maier, J. (2010). Pathways for Oxygen Incorporation in Mixed Conducting Perovskites: A DFT-Based Mechanistic Analysis for (La, Sr)MnO<sub>3-δ</sub>. *Journal Physical Chemistry C*, Vol.114, No. 7, (July 2010), pp. 3017-3027, ISSN: 1932-7455
- Monkhorst, H. J. & Pack, J. D. (1976). Special points for Brillouin-zone integrations. *Physical Review B*, Vol.13, No.12, (June 1976), pp. 5188-5192, ISSN: 1550-235X
- NIST (2010). *NIST computational chemistry comparison and benchmark database*, NIST, Gaithersburg, Maryland, USA, Available from <http://cccbdb.nist.gov/>
- Padilla J. & Vanderbilt, D. (1997). Ab initio study of BaTiO<sub>3</sub> surfaces. *Physical Review B*, Vol.56, No. 3, (July 1997), pp. 1625-1631, ISSN: 1550-235X
- Padilla J. & Vanderbilt, D. (1998). Ab initio study of SrTiO<sub>3</sub> surfaces *Surface Science*, Vol.418, No.1, (November 1998), pp. 64-70, ISSN: 00396028
- Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J. & Fiolhais, C. (1992). Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B*, Vol.46, No.11, (September 1992), pp. 6671-6687, ISSN: 1550-235X
- Piskunov, S.; Heifets, E.; Eglitis, R. I. & Borstel, G. (2004). Bulk properties and electronic structure of SrTiO<sub>3</sub>, BaTiO<sub>3</sub>, PbTiO<sub>3</sub> perovskites: an ab initio HF/DFT study. *Computational Materials Science*, Vol. 29, No.2, (February 2004), pp. 165-178, ISSN: 0927-0256
- Piskunov, S.; Spohr, E.; Jacob, T.; Kotomin, E.A. & Ellis D. (2007). Electronic and magnetic structure of La<sub>0.875</sub>Sr<sub>0.125</sub>MnO<sub>3</sub> calculated by means of hybrid density-functional theory. *Physical Review B*, Vol. 76, No.1, (July 2007), a. 012410, ISSN: 1550-235X
- Piskunov, S.; Heifets, E.; Jacob, T.; Kotomin, E. A.; Ellis, D. E. & Spohr, E. (2008). Electronic structure and thermodynamic stability of LaMnO<sub>3</sub> and La<sub>1-x</sub>Sr<sub>x</sub>MnO<sub>3</sub> (001) surfaces: Ab initio calculations. *Physical Review B*, Vol.78, No.12, (September 2008), a. 121406, ISSN: 1550-235X
- Piskunov, S.; Jacob, T. & Spohr, E (2011). Oxygen adsorption at La<sub>1-x</sub>Sr<sub>x</sub>MnO<sub>3</sub> (001) surfaces: Predictions from first principles. *Physical Review B*, Vol.83, No.7, (February 2011), a. 073402, ISSN: 1550-235X
- Piskunov, S.; Jacob, T. & Spohr E. (2011) Oxygen adsorption at (La,Sr)MnO<sub>3</sub> (001) surfaces: Predictions from first principles. *Physical Review B*, Vol. 85, No 7, (February 2011), pp. 073402 (1-4), ISSN: 1550-235X

- Pojani, A.; Finocchi, F. & Noguerra, C. (1999). Polarity on the SrTiO<sub>3</sub> (111) and (110) surfaces. *Surface Science*, Vol.442, No.2, (November 1999), pp. 179 -198, ISSN: 00396028
- Reuter, K. & Scheffler, M. (2001a). Stability and electronic structure of the (1×1) SrTiO<sub>3</sub>(110) polar surfaces by first-principles calculations. *Surface Science*, Vol. 490, No. 1-2, (September 2001), pp. 20-28, ISSN: 00396028
- Reuter, K. & Scheffler, M. (2001b). Composition, structure, and stability of RuO<sub>2</sub>(110) as a function of oxygen pressure. *Physical Review B*, Vol.65, No. 3, (December 2001), a. 035406, ISSN: 1550-235X
- Reuter K. & Scheffler, M. (2004). Oxide formation at the surface of late 4d transition metals: insights from first-principles atomistic thermodynamics. *Applied Physics A*, Vol.78, No.6, (February 2004), pp. 793-798, ISSN: 1432-0630
- Towler, M. D.; Allan, N. L.; Harrison, N. M.; Saunders, V. R.; Mackrodt, W. C. & Aprà, E. (1994). *Ab initio* study of MnO and NiO. *Physical Review B*, Vol. 50, No.8, (August 1994), pp. 5041-5054, ISSN: 1550-235X
- Zhou, W.; Ran, R. & Shao Z. (2009). Progress in understanding and development of Ba<sub>0.5</sub>Sr<sub>0.5</sub>Co<sub>0.8</sub>Fe<sub>0.2</sub>O<sub>3- $\delta$</sub> -based cathodes for intermediate-temperature solid-oxide fuel cells: A review. *Journal of Power Sources*, Vol.192, No. 2, (July 2009), pp. 231-246, ISSN: 0378-7753
- Wagner, C. & Schottky, W. (1930). *Z. Phys. Chemie B*, Vol. 11, No., (1930) pp. 163-, ISSN
- Wagner, C. (1936). *Z. Phys. Chem. B* Vol. 32, No., (1936), pp. 447-, ISSN

# Advances in Interfacial Adsorption Thermodynamics: Metastable-Equilibrium Adsorption (MEA) Theory

Gang Pan, Guangzhi He and Meiyi Zhang  
*State Key Laboratory of Environmental Aquatic Chemistry, Research  
Center for Eco-Environmental Sciences, Chinese Academy of Sciences,  
People's Republic of China*

## 1. Introduction

Interfacial processes are central to understanding many processes in environmental sciences and technologies, chemical engineering, earth sciences, ocean sciences and atmospheric sciences. Thermodynamics has been used as a classical method to describe interfacial equilibrium properties over the last century. Experimentally measurable macroscopic parameters of adsorption density and concentration are widely used as the basic parameters in many equations/models to describe the equilibrium characteristics of adsorption reactions at solid-water interfaces. For instance, methods of equilibrium adsorption constants or adsorption isotherms are commonly used to describe the equilibrium relationship between concentration in solution and adsorption density on solid surfaces.

However, thermodynamics has limitations in describing the equilibrium properties for surface adsorption reactions at solid-water interfaces. A fundamental principle has been missing in the conventional theoretical system where the microscopic structures on the solid surfaces are not taken into account in the conventional macroscopic methodology such as equilibrium adsorption constants and/or adsorption isotherms. The equilibrium properties for surface adsorption were conventionally described by macroscopic parameters such as adsorption density. Unfortunately, adsorption density is not a thermodynamic state variable and is generally affected by the microscopic metastable equilibrium surface structures, which make the equilibrium properties, such as equilibrium constants and/or adsorption isotherms, be fundamentally dependent on the kinetic paths and/or the reactant concentration conditions (e.g. the “adsorbent concentration effect” and “adsorbate concentration effect”). Failure in recognizing this theoretical gap has greatly hindered our understanding on many adsorption related issues especially in applied science and technology fields where the use of surface concentration ( $\text{mol}/\text{m}^2$ ) is common and inevitable.

With the application of spectroscopy and quantum chemical calculation techniques to solid-liquid interface systems, such as synchrotron based X-ray absorption spectroscopy, it is now possible to develop new thermodynamic methodologies to describe the real equilibrium properties of surface adsorption reactions and to reveal the relationships between macroscopic equilibrium properties and the microscopic metastable equilibrium adsorption

(MEA) structures. These studies represent advances on how microscopic surface molecule structures affect the macroscopic relationships in surface adsorption thermodynamics. Surface microstructures greatly affect the local chemical properties, long-range interaction, surface reactivity, and bioavailability of pollutants in the environment. Both experimental techniques and thermodynamic theoretical development on interfacial processes are essential for the development of molecular environmental and geological sciences.

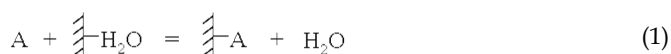
It has been a basic concept in traditional thermodynamic adsorption theories that adsorption density ( $\Gamma$ , mol/m<sup>2</sup>) is a state variable (a function that is only determined by the state and not affected by the path), so that the equilibrium adsorption constants defined by the ratio of equilibrium adsorption density on solid surfaces to the concentration in solution should be constant that is the reflection of the unique equilibrium characteristic of the reaction.<sup>1</sup> Over the last century, the macroscopic methodology (e.g. surface complexation models) of equilibrium adsorption constants and adsorption isotherms are widely used to describe the equilibrium limits of adsorption reactions and predict the theoretical yield in many fields.<sup>2,3</sup> These relationships were deemed to obey the basic properties of chemical thermodynamics, i.e. the equilibrium constant should be constant and be independent of kinetics or initial reactant concentrations under fixed thermodynamic conditions.

However, an abnormal phenomenon called particle/adsorbent concentration effect ( $C_p$  effect), i.e. the dependence of adsorption isotherms on one of the reactant concentrations  $C_p$ , has caused great confusion over the last three decades because it cannot be interpreted by the existing thermodynamic theories.<sup>2, 4-8</sup> Several hundreds of papers have been published on this issue but the underlined theoretical reason, which is far more important than the  $C_p$  effect itself, still remains not clear to most researchers. Most studies so far attribute  $C_p$  effect to various experimental artifacts.<sup>9, 10</sup> However, after these artifacts are excluded from the experiments,  $C_p$  effect may disappear in some systems,<sup>9</sup> but still exist in other systems.<sup>3, 11</sup> Thus, the problem becomes rather confused based on empirical or experimental analysis only.

Metastable-equilibrium adsorption (MEA) theory indicates that,<sup>12-14</sup> for a given adsorption reaction under fixed thermodynamic conditions, a polyhedral adsorbate molecule is generally ended in various MEA states with different energies and geometries rather than a unique equilibrium state when the reaction reaches to the apparent equilibrium. Unlike concentration in solutions, adsorption density (mol/m<sup>2</sup>) on solid surfaces no longer unambiguously corresponds to thermodynamic state variables, because adsorption density can only count for the mass but not the chemical potentials/energies of different microscopic MEA states that construct the real equilibrium adsorption state. When the adsorption density is not treated as a thermodynamic state variable, a theoretical equation known as “MEA inequality” is deduced from the fundamental thermodynamic laws.<sup>12</sup>

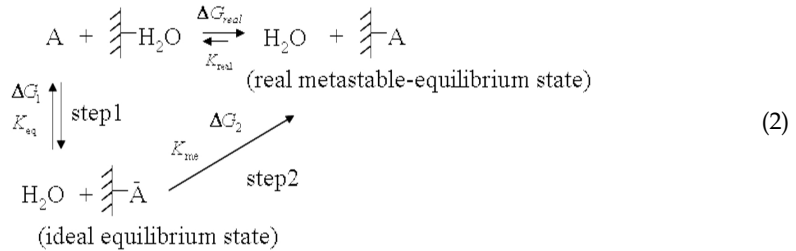
## 2. Metastable-equilibrium adsorption inequality

Suppose the adsorption of a pure solute A in a pure solvent onto a solid surface can be schematically represented by the equation (1, 2)



where A stands for solute in solution,  $\overset{\text{||}}{\text{||}}\text{-H}_2\text{O}$  for adsorbed solvent,  $\overset{\text{||}}{\text{||}}\text{-A}$  for adsorbed A, and  $\text{H}_2\text{O}$  for solvent in solution.

Since the Gibbs free energy is a state function and its change depends only on the initial and final states of the system, we can replace the real adsorption process of [1], which is generally thermodynamically irreversible, with two ideal reversible processes that lead to the same final state, in order to calculate the Gibbs free energy change,



where " $\xrightleftharpoons{\quad}$ " indicates that the real adsorption process can be irreversible. Step 1 represents an imagined reversible adsorption process where the final concentration of A on the solid surface is the same as that in the real irreversible process [1].  $\overset{\text{||}}{\text{||}}\text{-A}$  represents an ideal equilibrium stable state of adsorbed A, and  $\overset{\text{||}}{\text{||}}\text{-A}$  represents a real metastable-equilibrium adsorption state.  $\overset{\text{||}}{\text{||}}\text{-A}$  and  $\overset{\text{||}}{\text{||}}\text{-A}$  represent different thermodynamic states of adsorbed A, although they have the same value of adsorption density.  $\Delta G_1$  and  $K_{eq}$  are the change in Gibbs free energy and equilibrium constant of step 1, respectively.  $\Delta G_2$  of step 2 is the difference in Gibbs free energy between the reaction products of the real irreversible process [1] and the ideal reversible process (step 1).  $K_{me}$  is the equilibrium constant of step 2.

Thus,

$$\Delta G_{real} = \Delta G_1 + \Delta G_2 \tag{3}$$

since

$$\Delta G = -RT \ln K$$

Thus,

$$K_{real} = K_{eq} \times K_{me} \tag{4}$$

In step 2

$$\Delta G_2 = \left( G_{\text{H}_2\text{O}} + G_A^{solid} \right)_{real} - \left( G_{\text{H}_2\text{O}} + G_A^{solid} \right)_{ideal} \tag{5}$$

assuming

$$\begin{aligned}
 \left( G_{\text{H}_2\text{O}} \right)_{real} &= \left( G_{\text{H}_2\text{O}} \right)_{ideal} \\
 \Delta G_2 &= \left( G_A^{solid} \right)_{real} - \left( G_A^{solid} \right)_{ideal} \geq 0
 \end{aligned} \tag{6}$$

where “=” represents a reversible process and “>” corresponds to an irreversible process. Equation [6] indicates that if the process is not thermodynamically reversible, then for the same amount of adsorbed  $A$ , the real state of  $\frac{1}{2}A$ , which is of metastable equilibrium in nature, will have a higher Gibbs free energy  $(G_A^{solid})_{real}$  than the ideal equilibrium state  $(G_A^{solid})_{ideal}$ . Step 2 is therefore not a thermodynamically spontaneous process.

Since  $K_{me} = e^{-\Delta G_2/RT}$  and  $\Delta G_2 \geq 0$ ,

$$0 < K_{me} \leq 1 \quad (< \text{ for irreversible process, } = \text{ for reversible process}). \quad (7)$$

We call  $K_{me}$  the metastable-equilibrium coefficient. It measures the deviation of a metastable-equilibrium state from the ideal equilibrium state. Combining [4] and [7], we get the *MEA inequality*:

$$K_{real} \leq K_{eq} \quad (< \text{ for irreversible process, } = \text{ for reversible process}). \quad (8)$$

$K_{eq}$  is the ideal equilibrium constant for an ideal reversible process which has a unique value under constant temperature, pressure, and composition of the solution.  $K_{real}$  is the experimentally measured equilibrium constant for a real adsorption process and is not necessarily constant under fixed temperature, pressure, and composition of solution, but decreases as  $K_{me}$  decreases.

MEA inequality indicates that equilibrium constants or adsorption isotherms are fundamentally affected by the kinetic factor of thermodynamic irreversibility (including both mass and energetic irreversibility for a forward-backward reaction), because when the surface reaction is processed through different irreversible kinetic pathways it may reach to different MEA states under the same thermodynamic conditions. By using the MEA inequality to reformulate the existing equilibrium adsorption theories, it is possible to modify some of the existing isotherm equations into metastable-equilibrium equations.

### 2.1 Langmuir-type metastable equilibrium adsorption isotherm

The equilibrium constant for the adsorption process [1] is

$$K_{real} = \frac{(a_A)_s \times (a_{H_2O})_l}{(a_{H_2O})_s \times (a_A)_l} = \frac{(f_A)_s \times (x_A)_s \times (a_{H_2O})_l}{(f_{H_2O})_s \times (x_{H_2O})_s \times (a_A)_l} \quad (9)$$

where  $a_i$  stands for the activity of a given component in [1], the subscripts  $s$  and  $l$  refer to surface and bulk values, respectively,  $(f_i)_s$  is the surface activity coefficient, and  $(x_i)_s$  is the mole fraction surface concentration.

In dilute solution, the surface activity coefficient in the solid may be set equal to unity (1), so that

$$K_{real} = \frac{(x_A)_s \times (a_{H_2O})_l}{(x_{H_2O})_s \times (a_A)_l} \quad (10)$$

According to Eq. [4], we have

$$K_{eq} \times K_{me} = \frac{(x_A)_s \times (a_{H_2O})_l}{(x_{H_2O})_s \times (a_A)_l} \quad (11)$$

By multiplying both  $(x_A)_s$  and  $(x_{H_2O})_s$  by the total surface area  $A_T$ , and assuming that the molecular size of solute and solvent are similar, we have  $\theta_A = (x_A)_s \times A_T$ , and  $\theta_{H_2O} = (x_{H_2O})_s \times A_T$ , where  $\theta_i$  is the fraction of the surface occupied by component  $i$ . Because  $\theta_A + \theta_{H_2O} = 1$ , [11] becomes

$$\theta_A = \frac{K_{eq} \times K_{me} \times (a_A)_l}{(a_{H_2O})_l} \bigg/ \left( 1 + \frac{K_{eq} \times K_{me} \times (a_A)_l}{(a_{H_2O})_l} \right) \quad (12)$$

Since the activity of solvent  $(a_{H_2O})_l$  can be considered constant,  $K_{eq}/(a_{H_2O})_l$  may be defined as a constant  $b$ ,

$$\theta_A = \frac{b \times K_{me} \times (a_A)_l}{1 + b \times K_{me} \times (a_A)_l} \quad (13)$$

Practically, the adsorption amount is often expressed in terms of the adsorption density  $\Gamma$ ; thus,

$$\theta = \frac{\Gamma}{\Gamma_{max}} \quad (14)$$

where  $\Gamma_{max}$  is the characteristic saturation adsorption capacity for a given reaction, which is the maximum value of the equilibrium  $\Gamma$  as the equilibrium concentration of the solute increases. In dilute solution, activity  $(a_A)_l$  is approximately equal to concentration  $C_A$ , so [13] becomes

$$\Gamma = \frac{b' \times K_{me} \times C_{eq}}{1 + b' \times K_{me} \times C_{eq}} \quad (15)$$

where  $b' = b \times \Gamma_{max}$ ;  $b$  and  $b'$  are constant under fixed temperature and pressure, and are truly independent of the kinetics of the process.

Equations [13] and [15] are called *Langmuir-type metastable-equilibrium isotherm equations*, since when  $K_{me} = 1$ , i.e., under the ideal equilibrium condition, they are reduced to the conventional Langmuir equation. Only under this ideal condition ( $K_{me} = 1$ ) can the isotherm be independent of the kinetic process. Generally, the equilibrium relationship between  $\Gamma$  and  $C_{eq}$  would be influenced by the metastability of the adsorption state.

## 2.2 Freundlich-type metastable-equilibrium adsorption isotherm

By assuming an exponential distribution of adsorption energy, and assuming that for each energy level the adsorbate coverage  $\theta$  follows the Langmuir-type metastable-equilibrium isotherm [13], a Freundlich-type metastable isotherm equation can be obtained,<sup>12</sup>

$$\Gamma = \alpha \times K_{me} \times C_{eq}^{\beta} \quad (16)$$

where  $\alpha$  is a constant under isothermal conditions. Under ideal equilibrium conditions ( $K_{me} = 1$ ), [16] is reduced to the conventional Freundlich equation. Equation [16] indicates that the adsorption isotherm is shifted to the lower  $\Gamma$  as  $K_{me}$  decreases.

### 2.3 Particle concentration ( $C_p$ ) effect isotherm equations

According to reaction rate theory, adsorption speed should increase as particle concentration (i.e., reactant concentration) increases.<sup>15, 16</sup> Since the reversibility for a physical adsorption process on a plain solid surface generally declines as the speed of the process increases, the adsorption reversibility could decline as the particle concentration increases. Here, we assume that changes in  $C_p$  can affect the metastable-equilibrium adsorption state,

$$K_{me} = \gamma \times C_p^{-n} \quad (17)$$

where  $\gamma$  is a constant and  $n$  is an empirical parameter,  $n \geq 0$ .

Substituting [17] into [15], we obtain a semi-empirical Langmuir-type  $C_p$  effect isotherm equation

$$\Gamma = \frac{k' \times C_p^{-n} \times C_{eq}}{1 + k \times C_p^{-n} \times C_{eq}} \quad (18)$$

where  $k' = b' \times \gamma$  and  $k = b \times \gamma$ . For a given adsorption reaction,  $k'$  and  $k$  are equilibrium adsorption constants which are independent of the  $C_{eq}$  and  $C_p$  conditions.

Substituting [17] into [16], we obtain a Freundlich-type  $C_p$  effect isotherm equation,

$$\Gamma = k_{sp} \times C_p^{-n} \times C_{eq}^{\beta} \quad (19)$$

where  $k_{sp} = \alpha \times \gamma$ . For a given adsorption reaction,  $k_{sp}$  is an equilibrium adsorption constant which is independent of the  $C_{eq}$  and  $C_p$  conditions.

The  $C_p$  effect isotherm equations [18] and [19] predict that, by affecting the metastable-equilibrium adsorption state (or the adsorption reversibility), particle concentration can fundamentally influence the equilibrium constants or adsorption isotherms.

### 3. Macroscopic thermodynamic evidences of metastable-equilibrium adsorption

After Screening study of many adsorption systems, we found that there are obvious  $C_p$  effect in many irreversible adsorption systems (e.g., Zn-goethite, Zn-manganite, Zn-anatase, As(V)-anatase), and no  $C_p$  effect in reversible adsorption systems (e.g., Cd-goethite and Zn- $\delta$ -MnO<sub>2</sub>). Taking the adsorption of Zn and Cd on goethite as a typical example, the detailed interpretation of the existence and disappearance of the  $C_p$  effect using the metastable-equilibrium adsorption (MEA) theory are presented below.<sup>11, 17</sup>



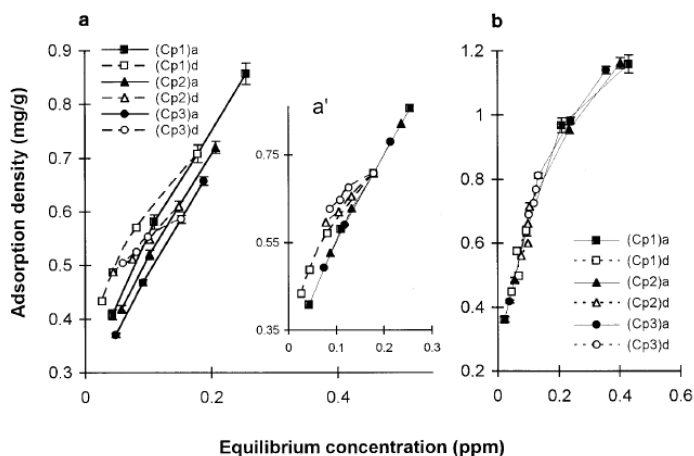


Fig. 1. Adsorption (solid lines, closed symbols) and desorption (dotted lines, open symbols) isotherms under different  $C_p$  conditions in Zn-goethite (a) and Cd-goethite (b) systems. (a)  $C_{p1}=0.38$  g/L,  $C_{p2}=1.53$  g/L,  $C_{p3}=2.3$  g/L, pH=6.4, equilibration time for adsorption and desorption are 12 days and 10 days, respectively.  $a'$ , a comparison of the sizes of hysteresis in Figure 1a, when the first points of the desorption isotherms are translationally moved to the same point. (b) pH=7.1, equilibration time for adsorption and desorption are 20 days and 14 days, respectively.

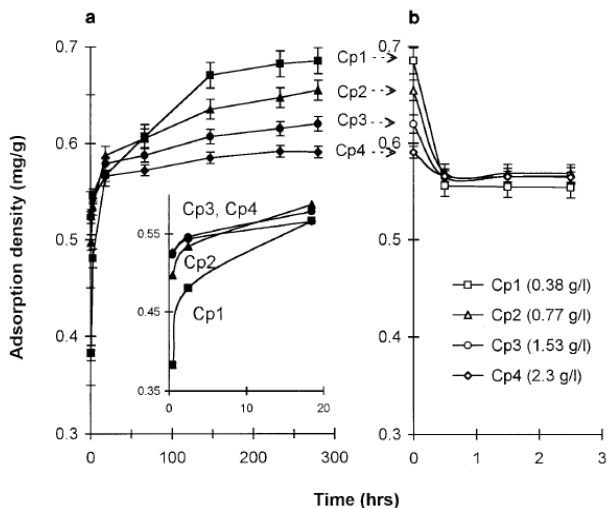


Fig. 2. Adsorption (a) and desorption (b) kinetic curves under different  $C_p$  conditions in Zn-goethite system. pH=6.4. The inset chart in (a) shows the initial stage of the adsorption. The final Zn concentrations of the four experiments in (a) are  $(C_{eq})_{Cp1}=0.18$  ppm,  $(C_{eq})_{Cp2}=0.17$  ppm,  $(C_{eq})_{Cp3}=0.16$  ppm,  $(C_{eq})_{Cp4}=0.15$  ppm. In order to examine the desorption rate effectively, only the initial stage of desorption is presented in (b).

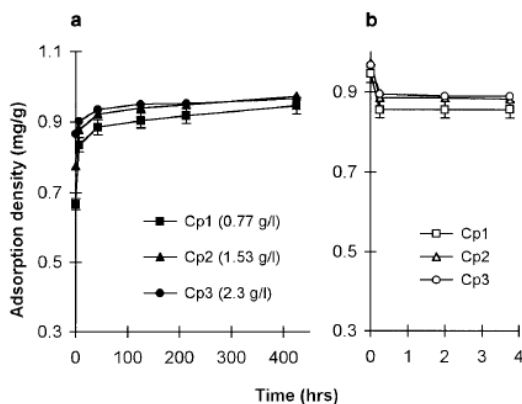


Fig. 3. Adsorption (a) and desorption (b) kinetic curves under different  $C_p$  conditions in Cd-goethite system. pH=7.1. The final Cd concentrations of the three experiments in (a) are  $(C_{eq})_{Cp1}=0.235$  ppm,  $(C_{eq})_{Cp2}=0.211$  ppm,  $(C_{eq})_{Cp3}=0.210$  ppm.

In a controlled simple aqueous system containing Zn-goethite, where a clear  $C_p$  effect is observed, an increase in particle concentration causes a simultaneous decrease in adsorption reversibility and in the adsorption isotherm (Figure 1a). At the same time, Zn adsorbed under a lower  $C_p$  condition desorbs faster (indicating more adsorption reversibility) than that under a higher  $C_p$  condition (Figure 2). In another controlled simple aqueous system of Cd-goethite, where no  $C_p$  effect is observed, changes in  $C_p$  does not cause discernible changes in adsorption hysteresis and in the adsorption isotherm (Figure 1b). Little difference in desorption rate is observed for the Cd adsorbed under different  $C_p$  conditions (Figure 3). Both the  $C_p$  effect and the non- $C_p$  effect results can be qualitatively explained and quantitatively described by the MEA theory.

In the Freundlich-type  $C_p$  effect isotherm equation [19], we called  $k_{sp}$  the specific adsorption constant and  $n$  the  $C_p$  effect index.  $k_{sp}$  is a measure of adsorption capacity and  $n$  is a measure of the degree of the  $C_p$  effect. Like b,  $k_{sp}$  and  $n$  can be calculated from adsorption isotherm data. The method is described below.

Take the logarithm of both sides of Eq. [19],

$$\log \Gamma = \log k_{sp} - n \log C_p + \beta \log C_{eq} \quad (20)$$

For a given adsorption isotherm ( e.g., isotherm a, b, or c in Figure 4 ),  $C_p$  is a constant, and Eq. [20] becomes:

$$\log \Gamma = A + \beta \log C_{eq} \quad (21)$$

where  $A = \log k_{sp} - n \log C_p$ . It can be seen from [21] that, if the relationship between  $\Gamma$  and  $C_{eq}$  under a specified  $C_p$  condition can be described by Eq. [19], then the plot of  $\log \Gamma$  vs  $\log C_{eq}$  should be a straight line. From the slope of the straight line,  $\beta$  can be obtained. Under a given  $C_{eq}$  ( e.g., data for ' $\Gamma_1, C_{p1}$ ', ' $\Gamma_2, C_{p2}$ ', ' $\Gamma_3, C_{p3}$ ' in Figure 4), Eq. [20] becomes

$$\log \Gamma = B - n \log C_p \quad (22)$$

where  $B = \log k_{sp} + \beta \log C_{eq}$ . Under this condition, if the plot of  $\log \Gamma$  vs  $\log C_p$  is a straight line, then the influence of  $C_p$  on the isotherm can be described by Eq. [19]. From the slope of the straight line,  $n$  is obtained.

From the intercepts of either Eq. [21] or Eq. [22],  $k_{sp}$  can be calculated.

Based on the adsorption isotherm data of Figure 1, the plots of  $\log \Gamma$  vs  $\log C_{eq}$  and the plots of  $\log \Gamma$  vs  $\log C_p$  for Zn-goethite and Cd-goethite systems are presented in Figure 5 and 6, respectively. Good linear relationships were obtained. For the Zn-goethite system,  $k_{sp}=1.381$ ,  $\beta=0.4136$ ,  $n=0.0819$ . For the Cd-goethite system,  $k_{sp}=1.778$ ,  $\beta=0.435$ ,  $n \approx 0$ .

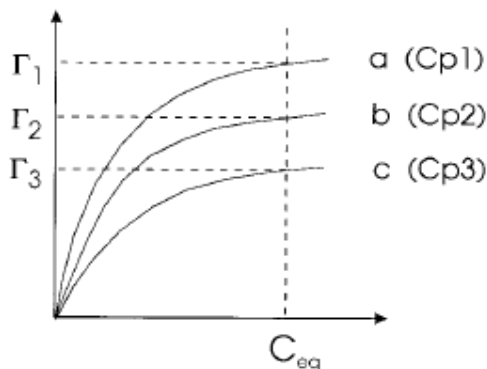


Fig. 4. Data of  $(\Gamma, C_{eq})$  under constant  $C_p$  are used to calculate  $b$ . Data of  $(\Gamma, C_p)$  under constant  $C_{eq}$  are used to calculate  $n$ .

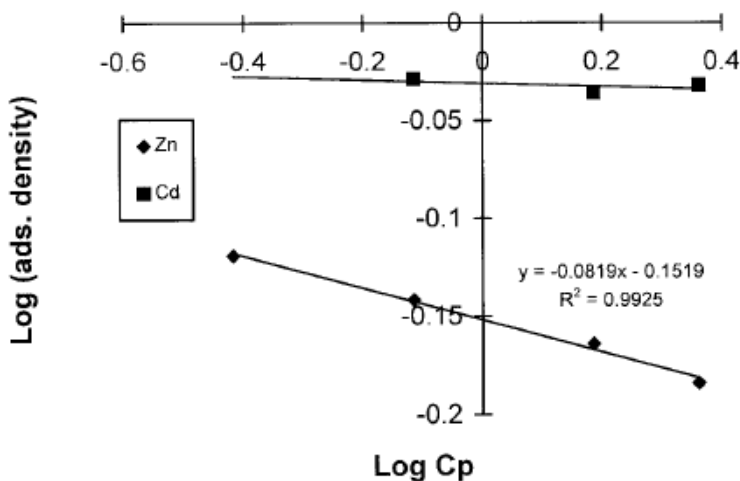


Fig. 5. Plot of  $\log C_p$  vs  $\log \Gamma$  under the condition of  $C_{eq}=0.2$  ppm for Zn-goethite and Cd-goethite systems, respectively. The different slopes of the two lines indicate that the size of the  $C_p$  effect is different in these two systems.

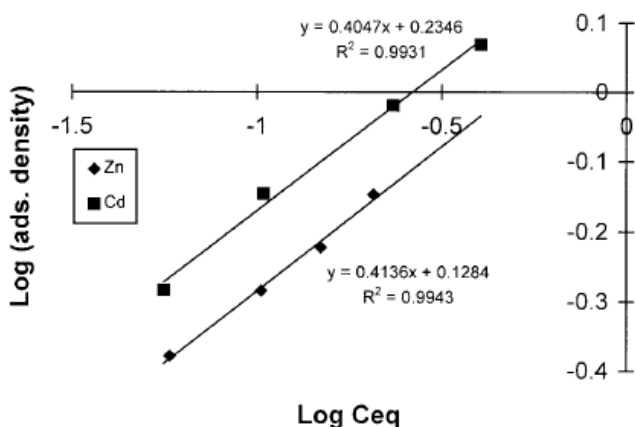


Fig. 6. Plot of  $\log C_{eq}$  vs  $\log \Gamma$  under the condition of  $C_p=1.534$  g/L for Zn-goethite and Cd-goethite systems, respectively. The similar slope of the two lines indicates that both the adsorption of Zn and Cd on goethite have similar  $\beta$  values.

After the specific adsorption constant ( $k_{sp}$ ), the  $C_p$  effect index ( $n$ ), and  $\beta$  are calculated, the  $C_p$  effect adsorption isotherm equations for Zn-goethite and Cd-goethite systems can be expressed as  $\Gamma = 1.381 \times C_p^{-0.0819} \times C_{eq}^{0.4136}$  and  $\Gamma = 1.778 \times C_{eq}^{0.435}$ , respectively. Figures 7 and 8 show that the calculated isotherms fit the experimental data well.

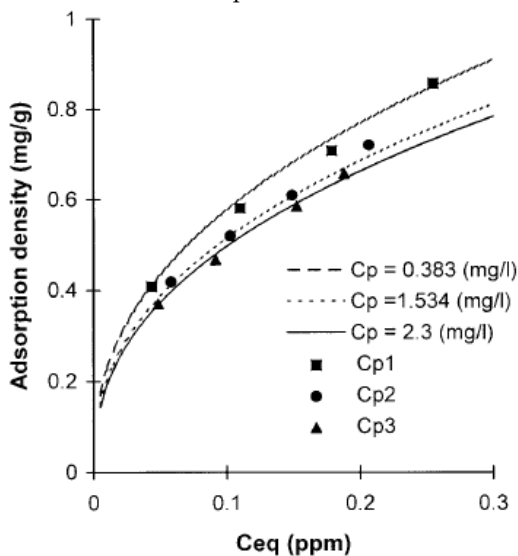


Fig. 7. Comparison between calculated and measured isotherms under different  $C_p$  conditions in the Zn-goethite system. Lines are calculated from the  $C_p$  effect isotherm equation  $\Gamma = 1.381 \times C_p^{-0.0819} \times C_{eq}^{0.4136}$ . Points are adsorption data from Figure 1a.

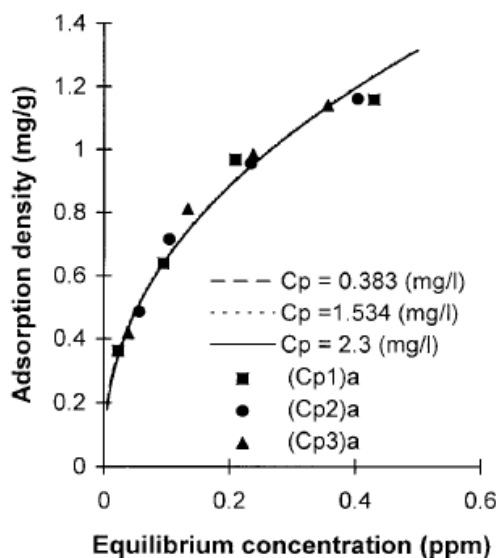


Fig. 8. Comparison between calculated and measured isotherms under different  $C_p$  conditions in Cd-goethite system. Lines are calculated from the  $C_p$  effect isotherm equation  $\Gamma = 1.778 \times C_{eq}^{0.435}$ . Points are adsorption data from Figure 1b.

According to MEA theory, for the ideal reversible adsorption reactions, changes in  $C_p$  have no influence on the reversibility of MEA states, and it should have no  $C_p$  effect in such systems when experimental artifacts are excluded.<sup>11, 18</sup> For partially irreversible adsorption reactions, changes in  $C_p$  may significantly affect the irreversibility and the microscopic MEA structures, and a  $C_p$  effect should fundamentally exist in irreversible adsorption systems.<sup>11, 17</sup> Therefore, the MEA theory provided a rational explanation for the phenomena of  $C_p$  effect and non- $C_p$  effect from the fundamental thermodynamic principle.

#### 4. Microscopic measurement of metastable-equilibrium adsorption state

It should be noted that, when the  $C_p$  effect isotherm equations are used in the modeling of practical adsorption processes, they may be totally empirical and does not imply particular physical mechanism. The macroscopic adsorption behavior is fundamentally controlled by the microscopic reaction mechanism of adsorbed molecules on solid surfaces. Therefore, the direct Measurement on the microstructures at solid-water interfaces is crucial to verifying the MEA principle.

Macroscopic thermodynamic results<sup>19, 20</sup> showed that Zn(II) adsorbed on manganite was largely irreversible (adsorption and desorption isotherms corresponding to the forward and backward reactions did not coincide, see Figure 9), but the adsorption of Zn (II) on  $\delta$ -MnO<sub>2</sub> was highly reversible (there was no apparent hysteresis between the adsorption and desorption isotherms, see Figure 10). This contrast adsorption behavior between the two forms of manganese oxides could be explained from the different microscopic structures

between  $\delta$ -MnO<sub>2</sub> and manganite, as well as the linkage modes of adsorbed Zn(II) on  $\delta$ -MnO<sub>2</sub> and manganite.<sup>19</sup>

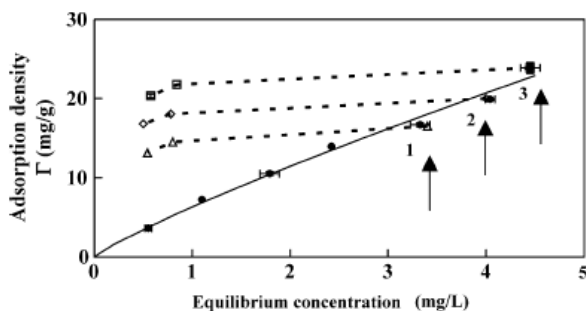


Fig. 9. Adsorption (closed symbols) and desorption (open symbols) isotherms of Zn(II) on manganite. EXAFS samples were indicated by arrows.

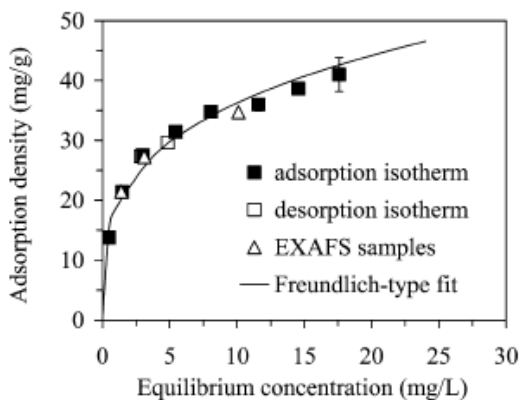


Fig. 10. Adsorption (■) and desorption (□) isotherms of Zn(II) on  $\delta$ -MnO<sub>2</sub>. EXAFS samples were symbolized with blank triangles (△).

Manganite had a structure with rows of edge-sharing Mn(II)O<sub>6</sub> octahedra linked to adjacent rows through corners. Due to the Jahn-Teller effect of Mn(II) ions and to the presence of both O and OH groups, the MnO<sub>6</sub> octahedra were highly distorted: each Mn is bound to four equatorial oxygen and two axial oxygen atoms.<sup>21, 22</sup> This distortion gave rise to a mild layered structure. Hydrolyzable Zn could be bonded on MnO<sub>6</sub> octahedra of manganite surface via edge and corner-sharing coordination modes.<sup>21, 22</sup> The basic structure of  $\delta$ -MnO<sub>2</sub> consisted of layers of edge-sharing MnO<sub>6</sub> octahedra alternating with a layer of water molecules. One-sixth of Mn<sup>4+</sup> positions were empty, which gave a layer charge that was compensated by two Zn atoms located above and below the vacancy.<sup>23, 24</sup> Hydrolyzable Zn could be taken up in the interlayer to form tridentate corner-sharing complexes.<sup>25, 26</sup> These differences in crystallographic structure resulted in different linkage modes for the adsorption of Zn on manganite and  $\delta$ -MnO<sub>2</sub>.

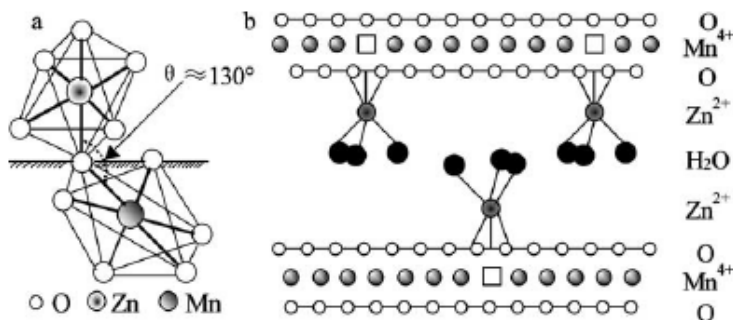


Fig. 11. Corner-sharing linkage (a) and interlayer structures of Zn(II) adsorbed on  $\delta$ -MnO<sub>2</sub> (b). (a)  $R_{Zn-O} = 2.07 \text{ \AA}$ ,  $R_{Mn-O} = 1.92 \text{ \AA}$ ,  $R_{Zn-Mn} = 3.52 \text{ \AA}$ . (b) Squares were vacant sites, illustration diagram adapted from Wadsley,<sup>27</sup> Post and Appleman,<sup>28</sup> and Manceau et al..<sup>25</sup>

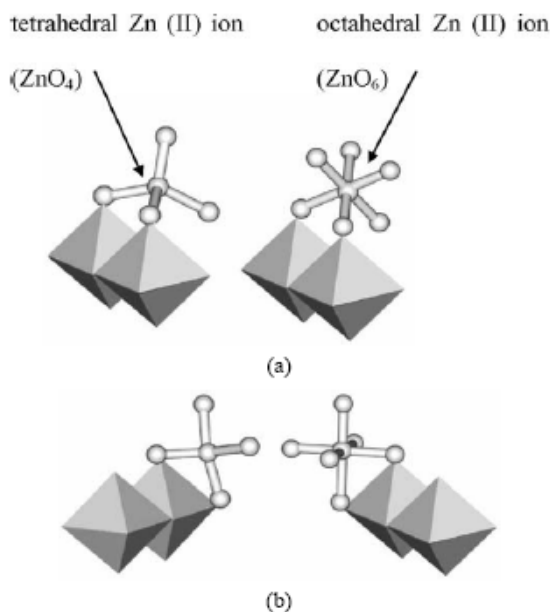


Fig. 12. Two types of linkage between adsorbed Zn(II) (octahedron and tetrahedron) and MnO<sub>6</sub> octahedra on the  $\gamma$ -MnOOH surfaces. (a) Double-corner linkage mode; (b) edge-linkage mode.

Extended X-ray absorption fine structure (EXAFS) analysis showed that Zn(II) was adsorbed onto  $\delta$ -MnO<sub>2</sub> in a mode of corner-sharing linkage, which corresponded to only one Zn-Mn distance of  $3.52 \text{ \AA}$  (Figure 11). However, there were two linkage modes for adsorbed Zn(II) on manganite surface as inner-sphere complexes, edge-sharing linkage and corner-sharing linkage, which corresponded to two Zn-Mn distances of  $3.07$  and  $3.52 \text{ \AA}$  (Figure 12). The

edge-sharing linkage was a stronger adsorption mode than that of the corner-sharing linkage, which would make it more difficult for the edge linkage to be desorbed from the solid surfaces than the corner linkage.<sup>20</sup> So adsorption of Zn(II) onto manganite was more irreversible than that on  $\delta$ -MnO<sub>2</sub>. This implied that the adsorption reversibility was influenced by the proportion of different bonding modes between adsorbate and adsorbent in nature.

Due to the contrast adsorption linkage mode, Zn(II) adsorbed on  $\delta$ -MnO<sub>2</sub> and manganite can be in very different metastable-equilibrium adsorption (MEA) states, which result in the different macroscopic adsorption-desorption behavior. For example, the extents of inconstancy of the equilibrium adsorption constant and the particle concentration effect are very different in the two systems. Adsorption of metals on  $\delta$ -MnO<sub>2</sub> and manganite may therefore be used as a pair of model systems for comparative studies of metastable-equilibrium adsorption.

## 5. Temperature dependence of metastable-equilibrium adsorption

Since temperature ( $T$ ) is expected to affect both adsorption thermodynamics and kinetics, the adsorption-desorption behavior may be  $T$ -dependent. The adsorption irreversibility of Zn(II) on anatase at various temperatures was studied using a combination of macroscopic thermodynamic methods and microscopic spectral measurement.

Adsorption isotherm results<sup>29</sup> showed that, when the temperature increased from 5 to 40 °C, the Zn(II) adsorption capacity increased by 130% (Figure 13). The desorption isotherms significantly deviate from the corresponding adsorption isotherms, indicating that the adsorption of zinc onto anatase was not fully reversible. The thermodynamic index of irreversibility (TII) proposed by Sander et al.<sup>30</sup> was used to quantify the adsorption irreversibility. The TII was defined as the ratio of the observed free energy loss to the maximum possible free energy loss due to adsorption hysteresis, which was given by

$$\text{TII} = \frac{\ln C_{\text{eq}}^{\gamma} - \ln C_{\text{eq}}^{\text{D}}}{\ln C_{\text{eq}}^{\text{S}} - \ln C_{\text{eq}}^{\text{D}}} \quad (23)$$

where  $C_{\text{eq}}^{\text{S}}$  is the solution concentration of the adsorption state S ( $C_{\text{eq}}^{\text{S}}, q_{\text{eq}}^{\text{S}}$ ) from which desorption is initiated;  $C_{\text{eq}}^{\text{D}}$  is the solution concentration of the desorption state D ( $C_{\text{eq}}^{\text{D}}, q_{\text{eq}}^{\text{D}}$ );  $C_{\text{eq}}^{\gamma}$  is the solution concentration of hypothetical reversible desorption state  $\gamma$  ( $C_{\text{eq}}^{\gamma}, q_{\text{eq}}^{\gamma}$ ).  $C_{\text{eq}}^{\text{S}}$  and  $C_{\text{eq}}^{\text{D}}$  are determined based on the experimental adsorption and desorption isotherms, and are easily obtained from the adsorption branch where the solid-phase concentration is equal to  $q_{\text{eq}}^{\text{D}}$ .

Based on the definition, the TII value lies in the range of 0 to 1, with 1 indicating the maximum irreversibility. The TII value (0.63, 0.34, 0.20) decreased by a factor of >3 when the temperature increased from 5 to 40 °C. This result indicated that the adsorption of Zn(II) on the TiO<sub>2</sub> surfaces became more reversible with increasing temperature.<sup>29</sup>

EXAFS spectra results showed that the hydrated Zn(II) was adsorbed on anatase through edge-sharing linkage mode (strong adsorption) and corner-sharing linkage mode (weak adsorption), which corresponded to two average Zn-Ti atomic distances of  $3.25 \pm 0.02$  and  $3.69 \pm 0.03$  Å, respectively.<sup>29</sup> According to the DFT results (Figure 14),<sup>13</sup> EXAFS measured the



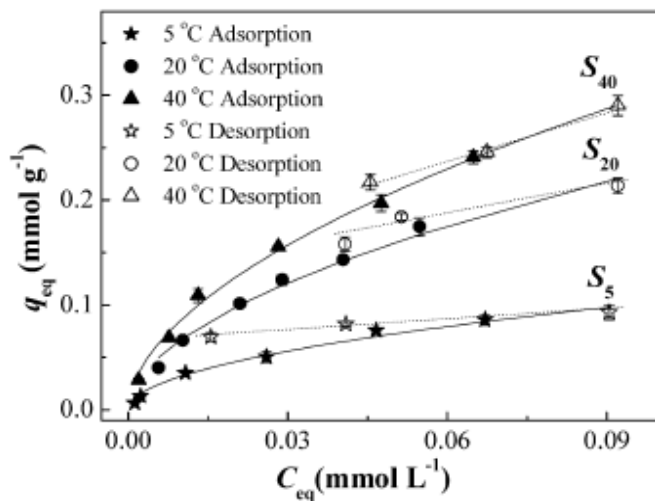


Fig. 13. Adsorption and desorption isotherms of Zn(II) on anatase at various temperatures. Symbols, experimental data; solid lines, model-fitted adsorption isotherms; dashed lines, model-fitted desorption isotherms.  $S_5$ ,  $S_{20}$ , and  $S_{40}$  indicate where desorption was initiated and samples selected for subsequent EXAFS analysis. Data given as mean of duplicates and errors refer to the difference between the duplicated samples.

corner-sharing linkage mode at the Zn-Ti distance of 3.69 Å may be a mixture of 4-coordinated bidentate binuclear (BB, 3.48 Å) and 6-coordinated monodentate mononuclear (MM, 4.01 Å) MEA states. DFT calculated energies showed that the MM complex was an energetically unstable MEA state compared with the BB (-8.58 kcal/mol) and BM (edge-sharing bidentate mononuclear, -15.15 kcal/mol) adsorption modes,<sup>13</sup> indicating that the MM linkage mode would be a minor MEA state, compared to the BB and BM MEA state. In the X-ray absorption near-edge structure analysis (XANES), the calculated XANES of BB and BM complexes reproduced all absorption characteristics (absorption edge, post-edge absorption oscillation and shape resonances) from the experimental XANES spectra (Figure 15).<sup>13</sup> Therefore, the overall spectral and computational evidence indicated that the corner-sharing BB and edge-sharing BM complexation mode coexisted in the adsorption of Zn(II) on anatase.

As the temperature increased from 5 to 40 °C, the number of strong adsorption sites (edge linkage) remained relatively constant while the number of the weak adsorption sites (corner linkage) increased by 31%.<sup>29</sup> These results indicate that the net gain in adsorption capacity and the decreased adsorption irreversibility at elevated temperatures were due to the increase in available weak adsorption sites or the decrease in the ratio of edge linkage to corner linkage. Both the macroscopic adsorption/desorption equilibrium data and the molecular level evidence indicated a strong temperature dependence for the metastable-equilibrium adsorption of Zn(II) on anatase.

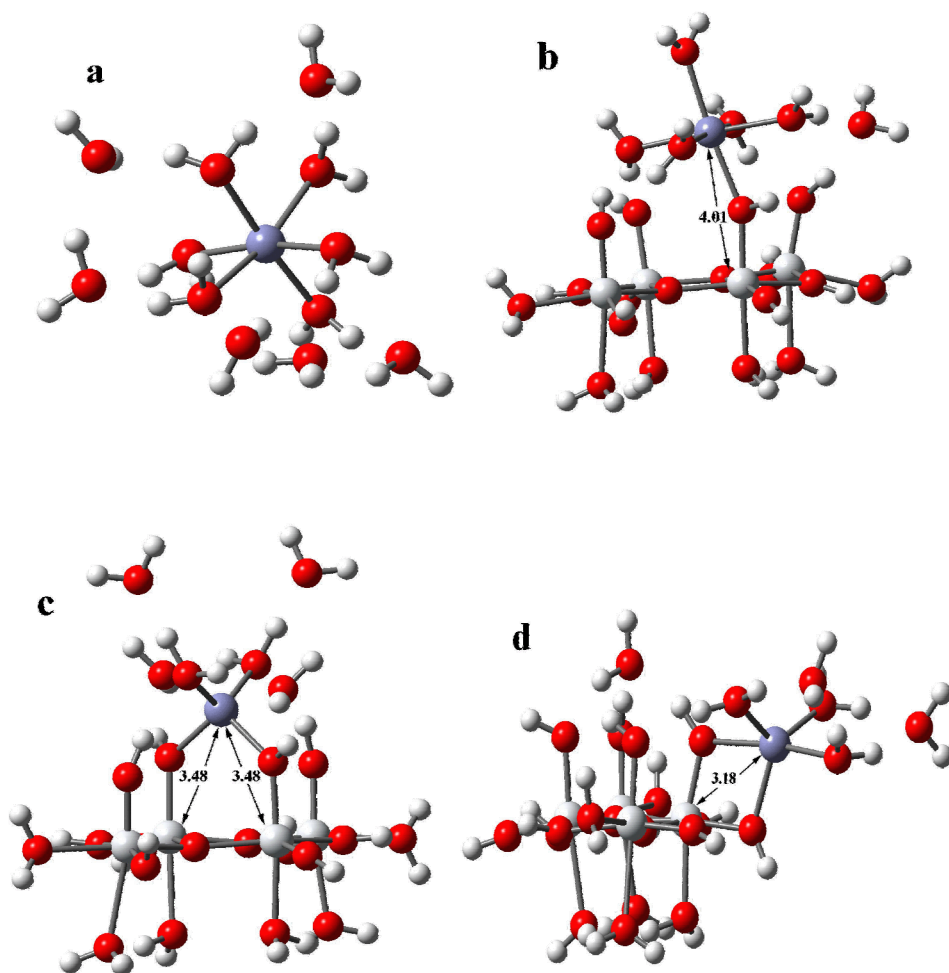


Fig. 14. Calculated Zn(II)-TiO<sub>2</sub> surface complexes using density functional theory: (a) dissolved Zn(II) with six outer-sphere water molecules; (b) monodentate mononuclear (MM); (c) bidentate binuclear (BB); (d) bidentate mononuclear (BM). Purple, red, big gray, small gray circles denote Zn, O, Ti, H atoms, respectively. Distances are shown in angstroms.

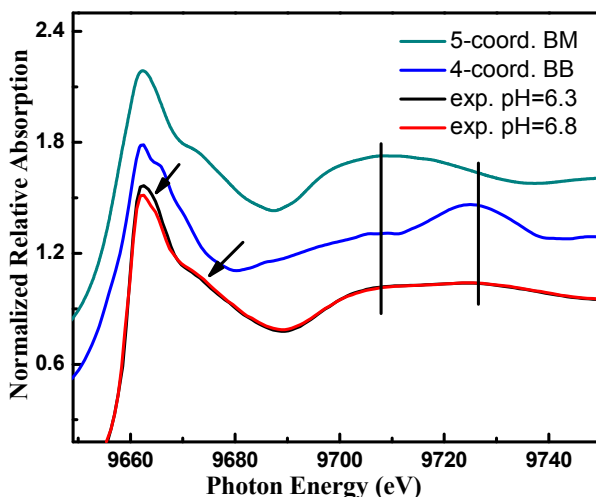


Fig. 15. Calculated XANES spectra of 4-oxygen coordinated BB and 5-oxygen coordinated BM complex and experimental XANES spectra.

## 6. pH dependence of metastable-equilibrium adsorption

According to MEA theory, both adsorbent/particle concentration (i.e.,  $C_p$ ) and adsorbate concentration could fundamentally affect equilibrium adsorption constants or isotherms when a change in the concentration of reactants (adsorbent or adsorbate) alters the reaction irreversibility or the MEA states of the apparent equilibrium. On the other hand, a general theory should be able to predict and interpret more phenomena. To test new phenomenon predicted by MEA theory can not only cross-confirm the theory itself but also provide new insights/applications in broadly related fields. The influence of adsorbate concentration on adsorption isotherms and equilibrium constants at different pH conditions was therefore studied in As(V)-anatase system using macroscopic thermodynamics and microscopic spectral and computational methods.<sup>14, 31, 32</sup>

The thermodynamic results<sup>14</sup> showed that, when the total mass of arsenate was added to the  $\text{TiO}_2$  suspension by multiple batches, the adsorption isotherms declined as the multi-batch increased, and the extent of the decline decreased gradually as pH decreased from 7.0 to 5.5 (Figure 16). This result provided a direct evidence for the influence of adsorption kinetics (1-batch/multi-batch) on adsorption isotherm and equilibrium constant, and indicated that the influence varied with pH.

According to MEA theory, for a given batch adsorption reaction under the same thermodynamic conditions, when the reaction is conducted through different kinetic pathways (1-batch/multi-batch), different MEA states (rather than a unique ideal equilibrium state) could be reached when the reaction reaches an apparent equilibrium (within the experimental time such as days).<sup>14</sup> Equilibrium constants or adsorption isotherms, which are defined by adsorption density, are inevitably affected by the reactant concentration when they alter the final MEA states.<sup>11, 12</sup>

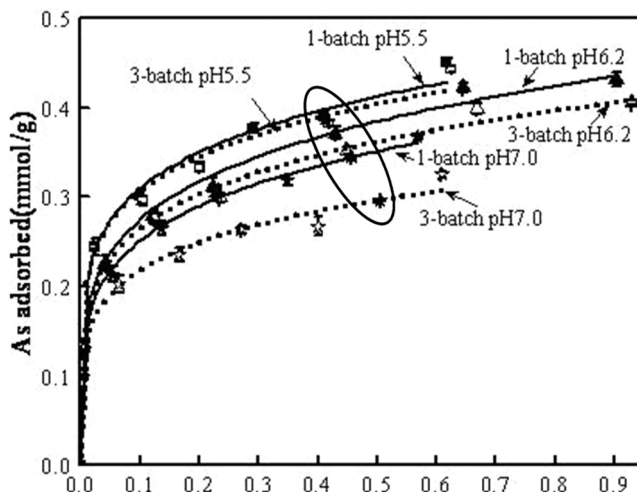


Fig. 16. Adsorption isotherms of As (V) on  $\text{TiO}_2$  in 0.01 mol/L  $\text{NaNO}_3$  solution at 25 °C under different pH.  $\text{TiO}_2$  particle concentration is 1 g/L. 1-batch stands for a series of total arsenate being added to  $\text{TiO}_2$  suspension in one time, and 3-batch stands for the total arsenate being added averagely to  $\text{TiO}_2$  suspension in 3 times every 4 hours. EXAFS samples were marked by ellipse, in which the initial total As (V) concentration is 0.80 mmol/L.

Sample	As-O			As-Ti						Res.	$\text{CN}_1/\text{CN}_2$
	CN	As-O		BB			MM				
		R(Å)	$\sigma^2$	$\text{CN}_1$	$R_1(\text{Å})$	$\sigma^2$	$\text{CN}_2$	$R_2(\text{Å})$	$\sigma^2$		
1-batch pH5.5	3.9	1.68	0.002	1.9	3.17	0.008	1.1	3.60	0.01	8.6	1.8
3-batch pH5.5	4.0	1.68	0.002	2.2	3.26	0.01	0.9	3.61	0.008	14.2	2.4
1-batch pH6.2	4.0	1.68	0.002	1.8	3.16	0.007	1.0	3.59	0.006	11.0	1.7
3-batch pH6.2	3.9	1.68	0.002	2.1	3.19	0.008	0.8	3.59	0.01	9.0	2.5
1-batch pH7.0	4.1	1.69	0.002	1.8	3.17	0.007	1.1	3.59	0.001	13.2	1.6
3-batch pH7.0	4.1	1.68	0.002	2.2	3.22	0.004	1.0	3.60	0.001	10.9	2.2
As(V)-pH5.5	4.1	1.68	0.004							6.7	
As(V)-pH7.0	4.1	1.69	0.003							5.3	
Calculated values	4.0	1.70		2.0	3.25		1.0	3.52			

Table 1. Summary of As(V) K-edge EXAFS results for 1-batch and 3-batch adsorption samples at pH 5.5, 6.2 and 7.0.

The comparison of EXAFS measured and DFT calculated results indicated that arsenate mainly formed inner-sphere bidentate binuclear (BB) and monodentate mononuclear (MM) surface complexes on  $\text{TiO}_2$ , where EXAFS measured two As-Ti distances of  $3.20 \pm 0.05$  and  $3.60 \pm 0.02$  Å (Table 1) corresponded to the DFT calculated values of BB (3.25 Å) and MM (3.52 Å) complexes (Figure 17), respectively.<sup>14</sup>

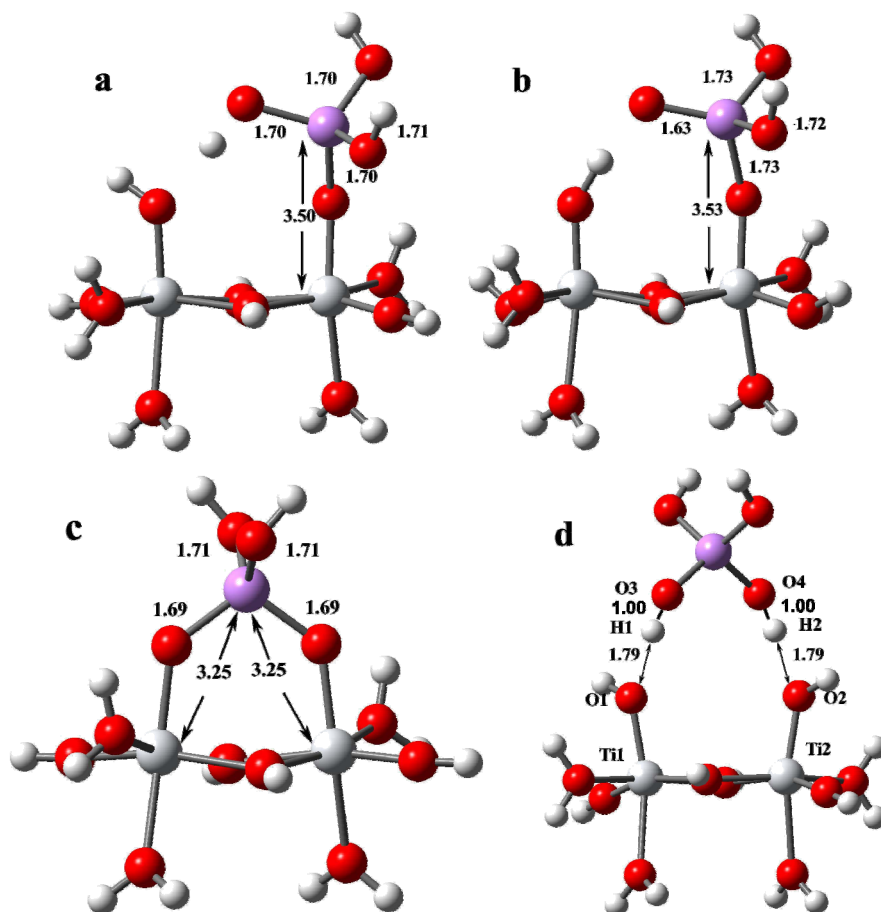


Fig. 17. DFT calculated structure of inner-sphere and H-bond adsorption products of arsenate on TiO<sub>2</sub>: (a) monodentate mononuclear arsenate H-bonded to a H<sub>2</sub>O surface functional group occupying the adjacent surface site (MM<sub>1</sub>); (b) monodentate mononuclear arsenate H-bonded to a -OH surface functional group occupying the adjacent surface site (MM<sub>2</sub>); (c) bidentate binuclear (BB) complex; (d) H-bonded complex. Red, big gray, small gray, purple circles denote O, Ti, H, As atoms, respectively. Distances are shown in angstroms.

The EXAFS coordination number of CN<sub>1</sub> and CN<sub>2</sub> represented statistically the average number of nearest Ti atoms around the As atom corresponding to a specific interatomic distance. We used the coordination number ratio of CN<sub>1</sub>/CN<sub>2</sub> to describe the relative proportion of BB mode to MM mode in adsorption samples. The CN<sub>1</sub>/CN<sub>2</sub> was 1.6 and 2.2 for 1-batch and 3-batch adsorption samples at pH 7.0, respectively (Table 1),<sup>14</sup> indicating that 3-batch adsorption samples contained more BB adsorbed arsenate than that of 1-batch adsorption samples. This result was cross-confirmed by measuring the spectral shift of X-ray absorption near edge structure (XANES) and Fourier transform infrared spectroscopy (FTIR).

DFT calculation showed that the theoretical XANES transition energy of BB complex was 0.62eV higher than that of MM complex. Therefore, the blue-shift of As (V) K-absorption edge observed from 1-batch to 3-batch adsorption samples suggested a structural evolution from MM to BB adsorption as the multi-batch increased (Figure 18).<sup>31</sup>

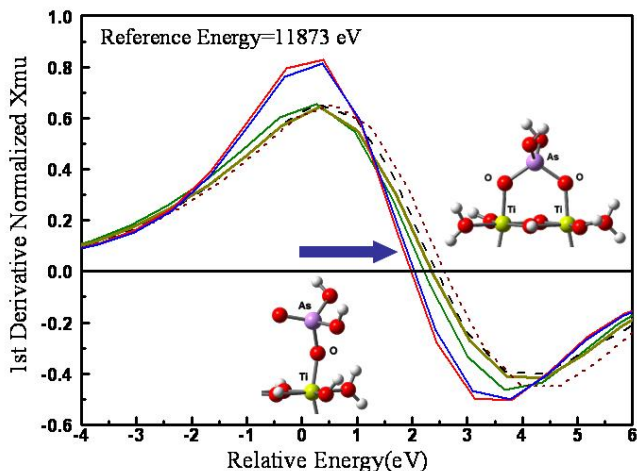


Fig. 18. The first derivative K-edge XANES spectra of As (V) adsorption on anatase.

The DFT calculated frequency analysis showed that the As-OTi asymmetric stretching vibration ( $\nu_{as}$ ) of MM and BB complexes located at 855 and 835  $\text{cm}^{-1}$ , respectively. On the basis of this theoretical analysis, the FTIR measured red-shift of As-OTi  $\nu_{as}$  vibration from 1-batch sample (849  $\text{cm}^{-1}$ ) to 3-batch sample (835  $\text{cm}^{-1}$ ) suggested that the ratio of BB/MM in 3-batch sample was higher than that in 1-batch sample (Figure 19).<sup>32</sup>

The good agreement of EXAFS results of  $CN_1/CN_2$  with XANES and FTIR analysis also validated the reliability of the CN ratio used as an index to approximate the proportion change of surface complexation modes. BB complex occupies two active sites on adsorbent surface whereas MM occupies only one. For monolayer chemisorption, a unit surface area of a given adsorbent can contain more arsenate molecules adsorbed in MM mode than that in BB mode. Therefore, the increase of the proportion of BB complex from 1-batch to 3-batch addition mode was shown as the decrease of adsorption density in 3-batch isotherm (Figure 16).

Table 1 showed that the relative proportion of BB and MM complex was rarely affected by pH change from 5.5 to 7.0, indicating that the pH dependence for the influence of adsorption kinetics (1-batch/multi-batch) on adsorption isotherm was not due to inner-sphere chemisorption.<sup>14</sup> The influence of pH on adsorption was simulated by DFT theory through changing the number of  $\text{H}^+$  in model clusters. Calculation of adsorption energy showed that the thermodynamic favorability of inner-sphere and outer-sphere adsorption was directly related to pH (Table 2).<sup>14</sup> As pH decreased, the thermodynamic favorability of inner-sphere and outer-sphere arsenate adsorption on Ti-(hydr)oxides increased. This DFT result explained why the adsorption densities of arsenate (Figure 16) and equilibrium adsorption constant (Table 2) increased with the decrease of pH.

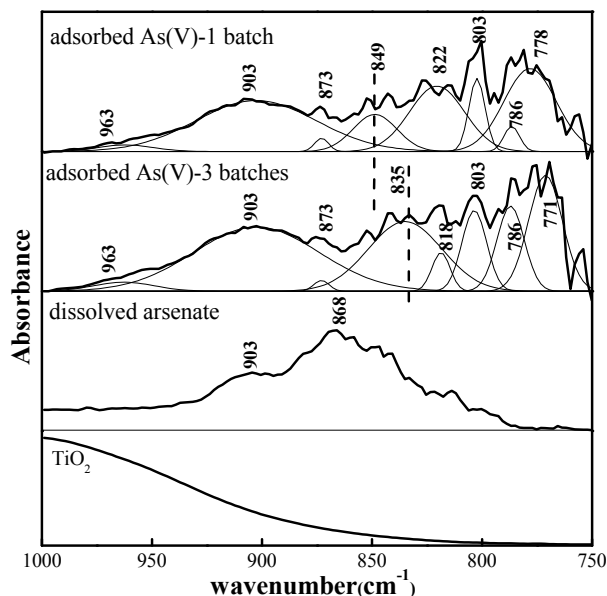


Fig. 19. ATR-FTIR spectra of adsorbed As(V) of 1-batch and 3-batch adsorption samples, dissolved arsenate, and TiO<sub>2</sub> at pH 7.0.

Theoretical equilibrium adsorption constant ( $K$ ) of calculated surface complexes (BB, MM and H-bonded complexes in this adsorption system) that constructed real equilibrium adsorption constant were significantly different in the order of magnitude under the same thermodynamic conditions (Table 2). The theoretical  $K$  were in the order of BB ( $6.80 \times 10^{42}$ ) > MM ( $3.13 \times 10^{39}$ ) > H-bonded complex ( $3.91 \times 10^{35}$ ) under low pH condition, and in the order of MM ( $1.54 \times 10^{-5}$ ) > BB ( $8.72 \times 10^{-38}$ ) > H-bonded complex ( $5.01 \times 10^{-45}$ ) under high pH condition. Therefore, even under the same thermodynamic conditions, the real equilibrium adsorption constant would vary with the change of the proportion of different surface complexes in real equilibrium adsorption.

DFT results (Table 2) showed that H-bond adsorption became thermodynamically favorable ( $-203.1$  kJ/mol) as pH decreased. H-bonded adsorption is an outer-sphere electrostatic attraction essentially (see Figure 17d), so it was hardly influenced by reactant concentration (multi-batch addition mode).<sup>14</sup> Therefore, as the proportion of outer-sphere adsorption complex increased under low pH condition, the influence of adsorption kinetics (1-batch/multi-batch) on adsorption isotherm would weaken (Figure 16).

Both the macroscopic adsorption data and the microscopic spectral and computational results indicated that the real equilibrium adsorption state of As(V) on anatase surfaces is generally a mixture of various outer-sphere and inner-sphere metastable-equilibrium states. The coexistence and interaction of outer-sphere and inner-sphere adsorptions caused the extreme complicity of real adsorption reaction at solid-liquid interface, which was not taken into account in traditional thermodynamic adsorption theories for describing the macroscopic relationship between equilibrium concentrations in solution and on solid surfaces. The reasoning behind the adsorbent and adsorbate concentration effects is that the conventional adsorption thermodynamic methods such as adsorption isotherms, which are

defined by the macroscopic parameter of adsorption density (mol/m<sup>2</sup>), can be inevitably ambiguous, because the chemical potential of mixed microscopic MEA states cannot be unambiguously described by the macroscopic parameter of adsorption density. Failure in recognizing this theoretical gap has greatly hindered our understanding on many adsorption related issues especially in applied science and technology fields where the use of surface concentration (mol/m<sup>2</sup>) is common or inevitable.

HO/AsO <sub>4</sub>	Adsorption reaction equations	ΔG	K
<b>Bidentate binuclear complexes</b>			
0	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_4(H_2O)_6]^{4+} \rightarrow [Ti_2(OH)_4(H_2O)_4AsO_2(OH)_2]^{3+} (H_2O)_2 + 12H_2O$	-244.5	6.80×10 <sup>42</sup>
1	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_5(H_2O)_5]^{3+} \rightarrow [Ti_2(OH)_4(H_2O)_4AsO_2(OH)_2]^{3+} (H_2O)_2 + OH^- (H_2O)_{11}$	13.1	5.15×10 <sup>-3</sup>
2	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_6(H_2O)_4]^{2+} \rightarrow [Ti_2(OH)_4(H_2O)_4AsO_2(OH)_2]^{3+} (H_2O)_2 + 2OH^- (H_2O)_{10}$	211.5	8.72×10 <sup>-38</sup>
<b>Monodentate mononuclear complexes</b>			
0	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_4(H_2O)_6]^{4+} \rightarrow [Ti_2(OH)_4(H_2O)_5AsO_2(OH)_2]^{3+} H_2O + 12H_2O$	-225.4	3.13×10 <sup>39</sup>
1-1	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_5(H_2O)_5]^{3+} \rightarrow [Ti_2(OH)_4(H_2O)_5AsO_2(OH)_2]^{3+} H_2O + OH^- (H_2O)_{11}$	32.1	2.37×10 <sup>-6</sup>
1-2	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_5(H_2O)_5]^{3+} \rightarrow [Ti_2(OH)_5(H_2O)_4AsO_2(OH)_2]^{2+} H_2O + 12H_2O$	-135.6	5.72×10 <sup>23</sup>
2	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_6(H_2O)_4]^{2+} \rightarrow [Ti_2(OH)_5(H_2O)_4AsO_2(OH)_2]^{2+} H_2O + OH^- (H_2O)_{11}$	27.5	1.54×10 <sup>-5</sup>
<b>H-bond complexes</b>			
0	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_4(H_2O)_6]^{4+} \rightarrow [Ti_2(OH)_4(H_2O)_6AsO_2(OH)_2]^{3+} + 12H_2O$	-203.1	3.91×10 <sup>35</sup>
1	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_5(H_2O)_5]^{3+} \rightarrow [Ti_2(OH)_4(H_2O)_6AsO_2(OH)_2]^{3+} + OH^- (H_2O)_{11}$	54.4	2.96×10 <sup>-10</sup>
2	$H_2AsO_4^- (H_2O)_{12} + [Ti_2(OH)_6(H_2O)_4]^{2+} \rightarrow [Ti_2(OH)_4(H_2O)_6AsO_2(OH)_2]^{3+} + 2OH^- (H_2O)_{10}$	252.9	5.01×10 <sup>-45</sup>

Table 2. Calculated ΔG<sub>ads</sub> (kJ/mol) and equilibrium adsorption constant *K* at 25 °C of arsenate on various protonated Ti-(hydr)oxide surfaces.

Metastable-equilibrium adsorption (MEA) theory pointed out that adsorbate would exist on solid surfaces in different forms (i.e. MEA states) and recognized the influence of adsorption reaction kinetics and reactant concentrations on the final MEA states (various outer-sphere and inner-sphere complexes) that construct real adsorption equilibrium state. Therefore, traditional thermodynamic adsorption theories need to be further developed by taking metastable-equilibrium adsorption into account in order to accurately describe real equilibrium properties of surface adsorption.



## 7. Acknowledgment

The study was supported by NNSF of China (20073060, 20777090, 20921063) and the Hundred Talent Program of the Chinese Academy of Science. We thank BSRF (Beijing), SSRF (Shanghai), and KEK (Japan) for supplying synchrotron beam time.

## 8. References

- [1] Atkins, P. W.; Paula, J. d., *Physical Chemistry, 8th edition*. Oxford University Press: Oxford, 2006.
- [2] Sverjensky, D. A., *Nature* 1993, 364 (6440), 776-780.
- [3] O'Connor, D. J.; Connolly, J. P., *Water Res.* 1980, 14 (10), 1517-1523.
- [4] Voice, T. C.; Weber, W. J., *Environ. Sci. Technol.* 1985, 19 (9), 789-796.
- [5] Honeyman, B. D.; Santschi, P. H., *Environ. Sci. Technol.* 1988, 22 (8), 862-871.
- [6] Benoit, G., *Geochim. Cosmochim. Acta* 1995, 59 (13), 2677-2687.
- [7] Benoit, G.; Rozan, T. F., *Geochim. Cosmochim. Acta* 1999, 63 (1), 113-127.
- [8] Cheng, T.; Barnett, M. O.; Roden, E. E.; Zhuang, J. L., *Environ. Sci. Technol.* 2006, 40, 3243-3247.
- [9] McKinley, J. P.; Jenne, E. A., *Environ. Sci. Technol.* 1991, 25 (12), 2082-2087.
- [10] Higgo, J. J. W.; Rees, L. V. C., *Environ. Sci. Technol.* 1986, 20 (5), 483-490.
- [11] Pan, G.; Liss, P. S., *J. Colloid Interface Sci.* 1998, 201 (1), 77-85.
- [12] Pan, G.; Liss, P. S., *J. Colloid Interface Sci.* 1998, 201 (1), 71-76.
- [13] He, G. Z.; Pan, G.; Zhang, M. Y.; Waychunas, G. A., *Environ. Sci. Technol.* 2011, 45 (5), 1873-1879.
- [14] He, G. Z.; Zhang, M. Y.; Pan, G., *J. Phys. Chem. C* 2009, 113, 21679-21686.
- [15] Nyffeler, U. P.; Li, Y. H.; Santschi, P. H., *Geochim. Cosmochim. Acta* 1984, 48 (7), 1513-1522.
- [16] Dzombak, D. A.; Morel, F. M. M., *J. Colloid Interface Sci.* 1986, 112 (2), 588-598.
- [17] Pan, G.; Liss, P. S.; Krom, M. D., *Colloids Surf., A* 1999, 151 (1-2), 127-133.
- [18] Pan, G., *Acta Scientiae Circumstantia* 2003, 23 (2), 156-173(in Chinese).
- [19] Li, X. L.; Pan, G.; Qin, Y. W.; Hu, T. D.; Wu, Z. Y.; Xie, Y. N., *J. Colloid Interface Sci.* 2004, 271 (1), 35-40.
- [20] Pan, G.; Qin, Y. W.; Li, X. L.; Hu, T. D.; Wu, Z. Y.; Xie, Y. N., *J. Colloid Interface Sci.* 2004, 271 (1), 28-34.
- [21] Bochatay, L.; Persson, P., *J. Colloid Interface Sci.* 2000, 229 (2), 593-599.
- [22] Bochatay, L.; Persson, P.; Sjoberg, S., *J. Colloid Interface Sci.* 2000, 229 (2), 584-592.
- [23] Drits, V. A.; Silvester, E.; Gorshkov, A. I.; Manceau, A., *Am. Mineral.* 1997, 82 (9-10), 946-961.
- [24] Post, J. E.; Veblen, D. R., *Am. Mineral.* 1990, 75 (5-6), 477-489.
- [25] Manceau, A.; Lanson, B.; Drits, V. A., *Geochim. Cosmochim. Acta* 2002, 66 (15), 2639-2663.
- [26] Silvester, E.; Manceau, A.; Drits, V. A., *Am. Mineral.* 1997, 82 (9-10), 962-978.
- [27] Wadsley, A. D., *Acta Crystallographica* 1955, 8 (3), 165-172.
- [28] Post, J. E.; Appleman, D. E., *Am. Mineral.* 1988, 73 (11-12), 1401-1404.
- [29] Li, W.; Pan, G.; Zhang, M. Y.; Zhao, D. Y.; Yang, Y. H.; Chen, H.; He, G. Z., *J. Colloid Interface Sci.* 2008, 319 (2), 385-391.
- [30] Sander, M.; Lu, Y.; Pignatello, J. J. A thermodynamically based method to quantify true sorption hysteresis; *Am Soc Agronom*: 2005; pp 1063-1072.

- [31] He, G. Z.; Pan, G.; Zhang, M. Y.; Wu, Z. Y., *J. Phys. Chem. C* 2009, 113 (39), 17076-17081.
- [32] Zhang, M. Y.; He, G. Z.; Pan, G., *J. Colloid Interface Sci.* 2009, 338 (1), 284-286.

# Towards the Authentic *Ab Initio* Thermodynamics

In Gee Kim

Graduate Institute of Ferrous Technology,  
Pohang University of Science and Technology, Pohang  
Republic of Korea

## 1. Introduction

A phase diagram is considered as a starting point to design new materials. Let us quote the statements by DeHoff (1993):

A phase diagram is a map that presents the domains of stability of phases and their combinations. A point in this space, which represents a state of the system that is of interest in a particular application, lies within a specific domain on the map.

In practice, for example to calculate the lattice stability, the construction of the phase diagram is to find the phase equilibria based on the comparison of the Gibbs free energies among the possible phases. Hence, the most important factor is the accuracy and precision of the given Gibbs free energy values, which are usually acquired by the experimental assessments. Once the required thermodynamic data are obtained, the phase diagram construction becomes rather straightforward with modern computation techniques, so called CALPHAD (CALculation of PHASE Diagrams) (Spencer, 2007). Hence, the required information for constructing a phase diagram is the reliable Gibbs free energy information. The Gibbs free energy  $G$  is defined by

$$G = E + PV - TS, \quad (1)$$

where  $E$  is the internal energy,  $P$  is the pressure,  $V$  is the volume of the system,  $T$  is the temperature and  $S$  is the entropy. The state which provides the minimum of the free energy under given external conditions at constant  $P$  and  $T$  is the equilibrium state. However, there is a critical issue to apply the conventional CALPHAD method in general materials design. Most thermodynamic information is relied on the experimental assessments, which do not available occasionally to be obtained, but necessary. For example, the direct thermodynamic information of silicon solubility in cementite had not been available for long time (Ghosh & Olson, 2002; Kozeschnik & Bhadeshia, 2008), because the extremely low silicon solubility which requires the information at very high temperature over the melting point of cementite. The direct thermodynamic information was available recently by an *ab initio* method (Jang et al., 2009). However, the current technology of *ab initio* approaches is usually limited to zero temperature, due to the theoretical foundation; the density functional theory (Hohenberg & Kohn, 1964) guarantees the unique total energy of the ground states only. The example demonstrates the necessity of a systematic assessment method from first principles. In order to obtain the Gibbs free energy from first principles, it is convenient to use the equilibrium statistical mechanics for grand canonical ensemble by introducing the *grand*

partition function

$$\Xi(T, V, \{\mu_i\}) = \sum_{N_i} \sum_{\zeta} \exp \left( -\beta \left( E_{\zeta}(V) - \sum_i \mu_i N_i \right) \right), \quad (2)$$

where  $\beta$  is the inverse temperature  $(k_B T)^{-1}$  with the Boltzmann's constant  $k_B$ ,  $\mu_i$  is the chemical potential of the  $i$ th component,  $N_i$  is the number of atoms. The sum of  $\zeta$  runs over all accessible microstates of the system; the microstates include the electronic, magnetic, vibrational and configurational degrees of freedom. The corresponding *grand potential*  $\Omega$  is found by

$$\Omega(T, V, \{\mu_i\}) = -\beta^{-1} \ln \Xi. \quad (3)$$

The Legendre transformation relates the grand potential  $\Omega$  and the *Helmholtz free energy*  $F$  as

$$\Omega(T, V, \{\mu_i\}) = F - \sum_i \mu_i N_i = E - TS - \sum_i \mu_i N_i. \quad (4)$$

It is noticeable to find that the Helmholtz free energy  $F$  is able to be obtained by the relation

$$F(T, V, N) = -\beta^{-1} \ln Z, \quad (5)$$

where  $Z$  is the *partition function* of the canonical ensemble defined as

$$Z(T, V, N) = \sum_{\zeta} \exp(-\beta E_{\zeta}(V, N)). \quad (6)$$

Finally, there is a further Legendre transformation relationship between the Helmholtz free energy and the Gibbs free energy as

$$G = F + PV. \quad (7)$$

Let us go back to the grand potential in Eq. (4). The total differential of the grand potential is

$$d\Omega = -SdT - PdV - \sum_i N_i d\mu_i, \quad (8)$$

with the coefficients

$$S = - \left( \frac{\partial \Omega}{\partial T} \right)_{V\mu}, \quad P = - \left( \frac{\partial \Omega}{\partial V} \right)_{T\mu}, \quad N_i = - \left( \frac{\partial \Omega}{\partial \mu_i} \right)_{TV}. \quad (9)$$

The Gibbs-Duhem relation,

$$E = TS - PV + \sum_i \mu_i N_i, \quad (10)$$

yields the thermodynamic functions as

$$F = -PV + \sum_i \mu_i N_i, \quad G = \sum_i \mu_i N_i, \quad \Omega = -PV. \quad (11)$$

Since the thermodynamic properties of a system at equilibrium are specified by  $\Omega$  and derivatives thereof, one of the tasks will be to develop methods to calculate the grand potential  $\Omega$ .

In principle, we can calculate any macroscopic thermodynamic states if we have the complete knowledge of the (grand) partition function, which is able to be constructed from first principles. However, it is impractical to calculate the partition function of a given system because the number of all accessible microstates, indexed by  $\zeta$ , is enormously large.

Struggles have been devoted to calculate the summation of all accessible states. The number of all accessible states is evaluated by the constituents of the system and the types of interaction among the constituents. The general procedure in statistical mechanics is nothing more than the calculation of the probability of a specific number of dice with the enormous number of repetitions of the dice tosses. The fundamental principles of statistical mechanics of a mechanical system of the degrees of freedom  $s$  is well summarized by Landau & Lifshitz (1980). The state of a mechanical system is described a point of the *phase space* represented by the generalized coordinates  $q_i$  and the corresponding generalized momenta  $p_i$ , where the index  $i$  runs from 1 to  $s$ . The time evolution of the system is represented by the trajectory in the phase space. Let us consider a closed large mechanical system and a part of the entire system, called *subsystem*, which is also large enough, and is interacting with the rest part of the closed system. An exact solution for the behavior of the subsystem can be obtained only by solving the mechanical problem for the entire closed system.

Let us assume that the subsystem is in the small phase volume  $\Delta p \Delta q$  for short intervals. The probability  $w$  for the subsystem stays in the  $\Delta p \Delta q$  during the short interval  $\Delta t$  is

$$w = \lim_{D \rightarrow \infty} \frac{\Delta t}{D}, \quad (12)$$

where  $D$  is the long time interval in which the short interval  $\Delta t$  is included. Defining the probability  $dw$  of states represented in the phase volume,

$$dpdq = dp_1 dp_2 \dots dp_s dq_1 dq_2 \dots dq_s,$$

may be written

$$dw = \rho(p_1, p_2, \dots, p_s, q_1, q_2, \dots, q_s) dpdq, \quad (13)$$

where  $\rho$  is a function of all coordinates and momenta in writing for brevity  $\rho(p, q)$ . This function  $\rho$  represents the density of the probability distribution in phase space, called (*statistical*) *distribution function*. Obviously, the distribution function is normalized as

$$\int \rho(p, q) dpdq = 1. \quad (14)$$

One should note that the statistical distribution of a given subsystem does not depend on the initial state of any other subsystems of the entire system, due to the entirely outweighed effects of the initial state over a sufficiently long time.

A physical quantity  $f = f(p, q)$  depending on the states of the subsystem of the solved entire system is able to be evaluated, in the sense of the statistical average, by the distribution function as

$$\bar{f} = \int f(p, q) \rho(p, q) dpdq. \quad (15)$$

By definition Eq. (12) of the probability, the statistical averaging is exactly equivalent to a time averaging, which is established as

$$\bar{f} = \lim_{D \rightarrow \infty} \frac{1}{D} \int_0^D f(t) dt. \quad (16)$$

In addition, the Liouville's theorem

$$\frac{d\rho}{dt} = \sum_{i=1}^s \left( \frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right) = 0 \quad (17)$$

tells us that the distribution function is constant along the phase trajectories of the subsystem. Our interesting systems are (quantum) mechanical objects, so that the counting the number of accessible states is equivalent to the estimation of the relevant phase space volume.

## 2. Phenomenological Landau theory

A ferromagnet in which the magnetization is the order parameter is served for illustrative purpose. Landau & Lifshitz (1980) suggested a phenomenological description of phase transitions by introducing a concept of *order parameter*. Suppose that the interaction Hamiltonian of the magnetic system to be

$$\sum_{i,j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j, \quad (18)$$

where  $\mathbf{S}_i$  is a localized Heisenberg-type spin at an atomic site  $i$  and  $J_{ij}$  is the interaction parameter between the spins  $\mathbf{S}_i$  and  $\mathbf{S}_j$ .

In the ferromagnet, the total magnetization  $\mathbf{M}$  is defined as the thermodynamic average of the spins

$$\mathbf{M} = \left\langle \sum_i \mathbf{S}_i \right\rangle, \quad (19)$$

and the magnetization  $\mathbf{m}$  denotes the magnetization per spin

$$\mathbf{m} = \left\langle \frac{1}{N} \sum_i \mathbf{S}_i \right\rangle, \quad (20)$$

where  $N$  is the number of atomic sites. The physical order is the alignment of the microscopic spins.

Let us consider a situation that an external magnetic field  $\mathbf{H}$  is applied to the system. Landau's idea<sup>1</sup> is to introduce a function,  $\mathcal{L}(\mathbf{m}, \mathbf{H}, T)$ , known as the Landau function, which describes the "thermodynamics" of the system as function of  $\mathbf{m}$ ,  $\mathbf{H}$ , and  $T$ . The minimum of  $\mathcal{L}$  indicates the system phase at the given variable values. To see more details, let us expand the Landau function with respect to the order parameter  $\mathbf{m}$ :

$$\mathcal{L}(m, H, T) = \sum_n^4 a_n(H, T) m^n, \quad (21)$$

where we assumed that both the magnetization  $\mathbf{m}$  and the external magnetic field  $\mathbf{H}$  are aligned in a specific direction, say  $\hat{\mathbf{z}}$ . When the system undergoes a *first-order phase transition*, the Landau function should have the properties

$$\left. \frac{\partial \mathcal{L}}{\partial m} \right|_{m_A} = \left. \frac{\partial \mathcal{L}}{\partial m} \right|_{m_B} = 0, \quad \mathcal{L}(m_A) = \mathcal{L}(m_B), \quad (22)$$

<sup>1</sup> The description in this section is following Negele & Orland (1988) and Goldenfeld (1992).

for the minima points  $A$  and  $B$ . For the case of the *second-order phase transition*, it is required that

$$\frac{\partial \mathcal{L}}{\partial m} = \frac{\partial^2 \mathcal{L}}{\partial m^2} = \frac{\partial^3 \mathcal{L}}{\partial m^3} = 0, \quad \frac{\partial^4 \mathcal{L}}{\partial m^4} > 0. \quad (23)$$

The second derivative must vanish because the curve changes from concave to convex and the third derivative must vanish to ensure that the critical point is a minimum. It is convenient to reduce the variables in the vicinity of the critical point  $t \equiv T - T_C$  and  $h \equiv H - H_c = H$ , where  $T_C$  is the Curie temperature and  $H_c$  is the critical external field, yielding the Landau coefficient

$$a_n(H, T) \mapsto a_n(h, t) = b_n + c_n h + d_n t, \quad (24)$$

and then the Landau function near the critical point is

$$\mathcal{L}(m, h, t) = c_1 h m + d_2 t m^2 + c_3 h m^3 + b_4 m^4, \quad d_2 > 0, \quad b_4 > 0. \quad (25)$$

Enforcing the inversion symmetry,  $\mathcal{L}(m, H, T) = \mathcal{L}(-m, -H, T)$ , the Landau function will be

$$\mathcal{L}(m, h, t) = d_2 t m^2 + b_4 m^4.$$

In order to see the dependency to the external field  $H$ , we add an arbitrary  $H$  field coupling term and change the symbols of the coefficients  $d_2$  to  $a$  and  $b_4$  to  $\frac{1}{2}b$ :

$$\mathcal{L} = a t m^2 + \frac{1}{2} b m^4 - H m. \quad (26)$$

Let us consider the second-order phase transition with  $H = 0$ . For  $T > T_C$ , the minimum of  $\mathcal{L}$  is at  $m = 0$ . For  $T = T_C$ , the Landau function has zero curvature at  $m = 0$ , where the point is still the global minimum. For  $T < T_C$ , the Landau function Eq. (26) has two degenerate minima at  $m_s = m_s(T)$ , which is explicitly

$$m_s(t) = \pm \sqrt{\frac{-at}{b}}, \quad \text{for } t < 0. \quad (27)$$

When  $H \neq 0$ , the differentiation of  $\mathcal{L}$  with respect to  $m$  gives the magnetic equation of state for small  $m$  as

$$a t m + b m^3 = \frac{1}{2} H. \quad (28)$$

The isothermal magnetic susceptibility is obtained by differentiating Eq. (28) with respect to  $H$ :

$$\chi_T(H) \equiv \left. \frac{\partial m(H)}{\partial H} \right|_T = \frac{1}{2 \left\{ a t + 3 b (m(H))^2 \right\}}, \quad (29)$$

where  $m(H)$  is the solution of Eq. (28). Let us consider the case of  $H = 0$ . For  $t > 0$ ,  $m = 0$  and  $\chi_T = 1/(2at)$ , while  $m^2 = -at/b$  and  $\chi_T = -1/(4at)$ . As the system is cooled down, the nonmagnetized system,  $m = 0$  for  $t > 0$ , occurs a spontaneous magnetization of  $(-at/b)^{\frac{1}{2}}$  below the critical temperature  $t < 0$ , while the isothermal magnetic susceptibility  $\chi_T$  diverges as  $1/t$  for  $t \rightarrow 0$  both for the regions of  $t > 0$  and  $t < 0$ .

For the first-order phase transition, we need to consider Eq. (25) with  $c_1 = 0$  and changing the coefficient symbols to yield

$$\mathcal{L} = a t m^2 + \frac{1}{2} m^4 + C m^3 - H m. \quad (30)$$

For  $H = 0$ , the equilibrium value of  $m$  is obtained as

$$m = 0, \quad m = -c \pm \sqrt{c^2 - at/b}, \quad (31)$$

where  $c = 3C/4b$ . The nonzero solution is valid only for  $t < t^*$ , by defining  $t^* \equiv bc^2/a$ . Let  $T_c$  is the temperature where the coefficient of the term quadratic in  $m$  vanishes. Suppose  $t_1$  is the temperature where the value of  $\mathcal{L}$  at the secondary minimum is equal to the value at  $m = 0$ . Since  $t^*$  is positive, this occurs at a temperature greater than  $T_c$ . For  $t < t^*$ , a secondary minimum and maximum have developed, in addition to the minimum at  $m = 0$ . For  $t < t_1$ , the secondary minimum is now the global minimum, and the value of the order parameter which minimizes  $\mathcal{L}$  jumps *discontinuously* from  $m = 0$  to a non-zero value. This is a first-order transition. Note that at the first-order transition,  $m(t_1)$  is not arbitrarily small as  $t \rightarrow t_1^-$ . In other words, the Landau theory is *not* valid. Hence, the first-order phase transition is arisen by introducing the cubic term in  $m$ .

Since the Landau theory is fully phenomenological, there is no strong limit in selecting order parameter and the corresponding conjugate field. For example, the magnetization is the order parameter of a ferromagnet with the external magnetic field as the conjugate coupling field, the polarization is the order parameter of a ferroelectric with the external electric field as the conjugate coupling field, and the electron pair amplitude is the order parameter of a superconductor with the electron pair source as the conjugate coupling field. When a system undergoes a phase transition, the Landau theory is usually utilized to understand the phase transition.

The Landau theory is motivated by the observation that we could replace the interaction Hamiltonian Eq. (18)

$$\sum_{i,j} J_{ij} S_i S_j = \sum_i S_i \sum_j J_{ij} \left[ \langle S_j \rangle + \{S_j - \langle S_j \rangle\} \right] \quad (32)$$

by  $\sum_{ij} S_i J_{ij} \langle S_j \rangle$ . If we can replace  $S_i S_j$  by  $S_i \langle S_j \rangle$ , it is also possible to replace  $\langle S_i S_j \rangle$  by  $\langle S_i \rangle \langle S_j \rangle$  on average if we assume the translational invariance. The fractional error implicit in this replacement can be evaluated by

$$\varepsilon_{ij} = \frac{|\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle|}{\langle S_i \rangle \langle S_j \rangle}, \quad (33)$$

where all quantities are measured for  $T < T_c$  under the Landau theory. The numerator is just a correlation function  $C$  and the interaction range  $|\mathbf{r}_i - \mathbf{r}_j| \sim R$  will allow us to rewrite  $\varepsilon_{ij}$  as

$$\varepsilon_R = \frac{|C(R)|}{m_s^2}, \quad (34)$$

where we assume the correlation function being written as

$$C(R) = g f\left(\frac{R}{\xi}\right), \quad (35)$$

where  $f$  is a function of the correlation length  $\xi$ . For  $T \ll T_c$ , the correlation length  $\xi \sim R$ , and the order parameter  $m$  is saturated at the low temperature value. The error is roughly



estimated as

$$\varepsilon_R \simeq \left( \frac{T}{T_C} \right) \left( \frac{a}{R} \right)^d, \quad (36)$$

where  $a$  is the lattice constant and  $d$  is the dimensionality of the interaction. In Eq. (36),  $(a/R)^{-d}$  is essentially the coordination number  $z > 1$ , so that  $\varepsilon_R < 1$  and the mean field theory is self-consistent.

On the other hand, the correlation length grows toward infinity near the critical point;  $R \ll \xi$  for  $t \rightarrow 0$ . A simple arithmetics yields  $m \sim |t|^\beta$ , where a *critical exponent*  $\beta$  is  $\frac{1}{2}$  for a ferromagnet. This result leaves us the error

$$\varepsilon_R \sim \frac{1}{|t|^{2\beta}} \left( \frac{a}{R} \right)^d, \quad (37)$$

which tends to infinity as  $t \rightarrow 0$ . Hence, the Landau theory based on the mean-field approximation has error which diverges as the system approaches to the critical point. Mathematically, the Landau theory expands the Landau function in terms of the order parameter. The Landau expansion itself is mathematically non-sense *near* the critical point for dimensions less than four. Therefore, the Landau theory is not a good tool to investigate significantly the phase transitions of the system.

### 3. Matters as noninteracting gases

Materials are basically made of atoms; an atom is composed of a nucleus and the surrounding electrons. However, it is convenient to distinguish two types of electrons; the *valence electrons* are responsible for chemical reactions and the *core electrons* are tightly bound around the nucleus to form an *ion* for screening the strongly divergent Coulomb potential from the nucleus. It is customary to call valence electrons as electrons.

The decomposition into electrons and ions provides us at least two advantages in treating materials with first-principles. First of all, the motions of electrons can be decoupled adiabatically from the those of ions, since electrons reach their equilibrium almost immediately by their light mass compared to those factors of ions. The decoupling of the motions of electrons from those of ions is accomplished by the Born-Oppenheimer adiabatic approximation (Born & Oppenheimer, 1927), which decouples the motions of electrons approximately begin independent adiabatically from those of ions. In practice, the motions of electrons are computed under the external potential influenced by the ions at their *static* equilibrium positions, before the motions of ions are computed under the external potential influenced by the electronic distribution. Hence, the fundamental information for thermodynamics of a material is its electronic structures. Secondly, the decoupled electrons of spin half are identical particles following the Fermi-Dirac statistics (Dirac, 1926; Fermi, 1926). Hence, the statistical distribution function of electrons is a closed fixed form. This feature reduces the burdens of calculation of the distribution function of electrons.

#### 3.1 Electronic subsystem as Fermi gas

The consequence of the decoupling electrons from ions allows us to treat the distribution functions of distinguishable atoms, for example, an iron atom is distinguished from a carbon atom, can be treated as the source of external potential to the electronic subsystem. Modelling of electronic subsystem was suggested firstly by Drude (1900), before the birth of quantum mechanics. He assumes that a metal is composed of electrons wandering on the positive homogeneous ionic background. The interaction between electrons are cancelled to allow us

for treating the electrons as a noninteracting gas. Albeit the Drude model oversimplifies the real situation, it contains many useful features of the fundamental properties of the electronic subsystem (Ashcroft & Mermin, 1976; Fetter & Walecka, 2003; Giuliani & Vignale, 2005). As microstates is indexed as  $i$  of the electron subsystem, the Fermi-Dirac distribution function is written in terms of occupation number of the state  $i$ ,

$$n_i^0 = \frac{1}{e^{\beta(\epsilon_i - \mu)} + 1}, \quad (38)$$

where  $\epsilon_i$  is the energy of the electronic microstate  $i$  and  $\mu$  is the chemical potential of the electron gas. At zero-temperature, the Fermi-Dirac distribution function becomes

$$\frac{1}{e^{\beta(\epsilon - \mu)} + 1} = \theta(\mu - \epsilon) \quad (39)$$

and the chemical potential becomes the Fermi energy  $\epsilon_F$ . In the high-temperature limit, the Fermi-Dirac distribution function recedes to

$$n^0 = e^{\beta(\epsilon - \mu)}, \quad (40)$$

the Maxwell-Boltzmann distribution function. With the nonrelativistic energy spectrum

$$\epsilon_{\mathbf{p}} = \frac{\mathbf{p}^2}{2m} = \frac{\hbar^2 k^2}{2m} = \epsilon_{\mathbf{k}}, \quad (41)$$

where  $\mathbf{p}$  is the single-particle momentum,  $\mathbf{k}$  is the corresponding wave vector, the grand potential in Eq. (3) is calculated in a continuum limit<sup>2</sup> as

$$-\beta\Omega_0 = \beta PV = \frac{2}{3} \frac{gV}{3\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_0^\infty d\epsilon \frac{\epsilon^{\frac{3}{2}}}{e^{\beta(\epsilon - \mu)} + 1} \quad (42)$$

and the number density is written as

$$\frac{N}{V} = \frac{g}{4\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_0^\infty d\epsilon \frac{\epsilon^{\frac{1}{2}}}{e^{\beta(\epsilon - \mu)} + 1}, \quad (43)$$

where  $g$  is 2, the degeneracy factor of an electron. After math (Fetter & Walecka, 2003), we can obtain the low-temperature limit ( $T \rightarrow 0$ ) of the grand potential of the noninteracting homogeneous electron gas as

$$PV = \frac{2}{3} \frac{gV}{4\pi^2} \left( \frac{2m}{\hbar} \right)^{\frac{3}{2}} \left[ \frac{2}{5} \mu^{\frac{5}{2}} + \beta^{-2} \frac{\pi^2}{4} \mu^{\frac{1}{2}} + \dots \right] \quad (44)$$

<sup>2</sup> It is convenient to convert a summation over single-particle spectra to an integral over wavenumbers according to  $\sum_i \rightarrow g \int d^3n = gV (2\pi)^{-3} \int d^3k$  for a very large periodic system, hence a continuum case. If we have knowledge of the single-particle energy dispersion relation, the wavenumber integral is also replaced by an integral over energy as  $gV (2\pi)^{-3} \int d^3k F(\epsilon_{\mathbf{k}}) \rightarrow g \int_{-\infty}^{\infty} d\epsilon \mathcal{D}(\epsilon) F(\epsilon)$ , where  $\mathcal{D}(\epsilon)$  is the density of states.

and the chemical potential from the relation  $N = (\partial(PV) / \partial\mu)_{TV}$  as

$$\mu = \epsilon_F \left[ 1 - \frac{\pi^2}{12} \left( \frac{1}{\beta\epsilon_F} \right)^2 + \dots \right]. \quad (45)$$

The low temperature limit entropy  $S$  is calculated as

$$S(\beta, V, \mu) = \left( \frac{\partial(PV)}{\partial T} \right)_{V\mu} = \frac{gV}{4\pi^2} \left( \frac{2m}{\hbar} \right)^{\frac{3}{2}} \frac{2}{3} \left[ \frac{2\pi^2}{4} \frac{k_B}{\beta} + \dots \right]. \quad (46)$$

It is thus the heat capacity of the noninteracting homogeneous electron gas to be

$$C_V = T \left( \frac{\partial S}{\partial T} \right)_{VN} = \frac{\pi^2}{2} N k_B \frac{1}{\epsilon_F \beta}. \quad (47)$$

The internal energy is simply calculated by a summation of the microstate energy of all the occupied states to yield

$$\frac{E}{V} = \frac{g}{4\pi^2} \int_0^\mu \epsilon^{\frac{3}{2}} d\epsilon = \frac{g}{4\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \frac{2}{5} \mu^{\frac{5}{2}}. \quad (48)$$

All the necessary thermodynamic information of the homogeneous noninteracting electron subsystem is acquired.

### 3.2 Elementary excitation as massive boson gas

For the case of ions, the treatment is rather complex. One can immediately raise the same treatment of the homogeneous noninteracting ionic gas model as we did for the electronic subsystem. Ignoring the nuclear spins, any kinds of ions are composed of fully occupied electronic shells to yield the effective zero spin; ions are massive bosons. It seems, if the system has single elemental atoms, that the ionic subsystem can be treated as an indistinguishable homogeneous noninteracting bosonic gas, following the Bose-Einstein statistics (Bose, 1926; Einstein, 1924; 1925). However, the ionic subsystem is hardly treated as a boson gas. Real materials are not elemental ones, but they are composed of many different kinds of elements; it is possible to distinguish the atoms. They are partially distinguishable each other, so that a combinatorial analysis is required for calculating thermodynamic properties (Ruban & Abrikosov, 2008; Turchi et al., 2007). It is obvious that the ions in a material are approximately distributed in the space isotropically and homogeneously. Such phases are usually called *fluids*. As temperature goes down, the material in our interests usually crystallizes where the homogeneous and isotropic symmetries are broken spontaneously and individual atoms all occupy nearly fixed positions.

In quantum field theoretical language, there is a massless boson, called Goldstone boson, if the Lagrangian of the system possesses a continuous symmetry group under which the the ground or vacuum state is not invariant (Goldstone, 1961; Goldstone et al., 1962). For example, phonons are emerged by the violation of translational and rotational symmetry of the solid crystal; a longitudinal phonon is emerged by the violation of the gauge invariance in liquid helium; spin waves, or magnons, are emerged by the violation of spin rotation symmetry (Anderson, 1963). These quasi-particles, or elementary excitations, have known in many-body theory for solids (Madelung, 1978; Pines, 1962; 1999). One has to note two facts: (i) the elementary exciations are not necessarily to be a Goldstone boson and (ii) they are not

necessarily limited to the ionic subsystem, but also electronic one. If the elementary excitations are fermionic, thermodynamics are basically calculable as we did for the non-interacting electrons gas model, in the beginning of this section. If the elementary excitations are (Goldstone) bosonic, such as phonons or magnons, a thermodynamics calculation requires special care. In order to illustrative purpose, let us see the thermodynamic information of a system of homogeneous noninteracting massive bosons.

The Bose-Einstein distribution function gives the mean occupation number in the  $i$ th state as

$$n_i^0 = \frac{1}{e^{\beta(\epsilon_i - \mu)} - 1}. \quad (49)$$

Since the chemical potential of a bosonic system vanishes at a certain temperature  $T_0$ , a special care is necessary during the thermodynamic property calculations (Cornell & Wieman, 2002; Einstein, 1925; Fetter & Walecka, 2003). The grand potential of an ideal massive boson gas, where the energy spectrum is also calculated as in Eq. (41), is

$$-\beta\Omega_0 = \beta PV = -\frac{gV}{4\pi^2} \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \int_0^\infty d\epsilon \epsilon^{\frac{1}{2}} \ln(1 - e^{\beta(\mu - \epsilon)}). \quad (50)$$

The integration by part yields

$$PV = \frac{gV}{4\pi^2} \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \frac{2}{3} \int_0^\infty d\epsilon \frac{\epsilon^{\frac{3}{2}}}{e^{\beta(\epsilon - \mu)} - 1}. \quad (51)$$

The internal energy is calculated to be

$$E = \sum_i n_i^0 \epsilon_i = \frac{3}{2} PV = \frac{gV}{4\pi^2} \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \int_0^\infty d\epsilon \frac{\epsilon^{\frac{3}{2}}}{e^{\beta(\epsilon - \mu)} - 1}, \quad (52)$$

and the number density is calculated to be

$$\frac{N}{V} = \frac{g}{4\pi^2} \left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}} \int_0^\infty d\epsilon \frac{\epsilon^{\frac{1}{2}}}{e^{\beta(\mu - \epsilon)} - 1}. \quad (53)$$

A care is necessary in treating Eq. (53), because it is meaningful only if  $\epsilon - \mu \geq 0$ , or

$$\mu \leq 0 \quad (54)$$

with the consideration of the fact  $\epsilon \geq 0$ .

In the classical limit  $T \rightarrow \infty$ , or  $\beta \rightarrow 0$ , for fixed  $N$ , we have

$$\beta\mu \rightarrow -\infty. \quad (55)$$

Recall that the classical limit yields the Maxwell-Boltzmann distribution

$$n_i^0 = e^{-\beta(\epsilon_i - \mu)} \quad (56)$$

for both fermions and bosons, and the corresponding grand potential becomes

$$\Omega_0 = -PV = -\frac{1}{\beta} \sum_i e^{\beta(\mu - \epsilon_i)}. \quad (57)$$

The classical chemical potential  $\mu_c$  is now calculated as

$$\beta\mu_c = \ln \left[ \frac{N}{gV} \left( \frac{2\pi\hbar^2}{m} \right)^{\frac{3}{2}} \beta^{\frac{3}{2}} \right]. \quad (58)$$

As  $\beta$  increases at fixed density,  $\beta\mu_c$  passes through zero and becomes positive, diverging to infinity at  $\beta \rightarrow \infty$ . This contradicts to the requirement Eq. (54). The critical temperature  $\beta_0$ , where the chemical potential of an ideal boson gas vanishes, is calculated by using Eq. (53) with  $\mu = 0$  to be

$$\frac{1}{\beta_0} = \frac{\hbar^2}{2m} \left( \frac{4\pi^2}{g\Gamma\left(\frac{3}{2}\right)\zeta\left(\frac{3}{2}\right)} \right)^{\frac{2}{3}} \left( \frac{N}{V} \right)^{\frac{2}{3}}, \quad (59)$$

where  $\Gamma$  and  $\zeta$  are Gamma function and zeta function, respectively. For  $\mu = 0$  and  $\beta > \beta_0$ , the integral in Eq. (53) is less than  $N/V$  because these conditions increase the denominator of the integrand relative to its value at  $\beta_0$ .

The breakdown of the theory was noticed by Einstein (1925) and was traced origin of the breakdown was the converting the conversion of the summation to the integral of the occupation number counting in Eq. (53). The total number of the ideal massive Bose gas is counted, using the Bose-Einstein distribution function Eq. (49), by

$$N = \sum_i \frac{1}{e^{\beta(\epsilon_i - \mu)} - 1}. \quad (60)$$

It is obvious that the bosons tends to occupy the ground state for the low temperature range  $\beta > \beta_0$ , due to the limitation of the occupation number of bosons. As temperature goes down, the contribution of the ground state occupation to the number summation increases. However, the first term of Eq. (60) is omitted, in the Bose-Einstein distribution, during the conversion to the integral Eq. (53) as  $\mu \rightarrow 0^-$  for  $\beta > \beta_0$ , because the fact that  $\epsilon_i = 0$  vanishes the denominator  $e^{\frac{1}{2}}$  of the integrand in Eq. (53). The number density of the Bose particles with energies  $\epsilon > 0$  is computed by Eq. (53) to be

$$\frac{N_{\epsilon>0}}{V} = \frac{N}{V} \left( \frac{\beta}{\beta_0} \right)^{-\frac{3}{2}}, \quad (61)$$

while the number density at the ground state is evaluated to be

$$\frac{N_{\epsilon=0}}{V} = \frac{N}{V} \left[ 1 - \left( \frac{\beta}{\beta_0} \right)^{-\frac{3}{2}} \right], \quad (62)$$

with the chemical potential  $\mu = 0^-$  for  $\beta > \beta_0$ . The internal energy density of the degenerate massive boson gas for  $\beta > \beta_0$  is then computed (Fowler & Jones, 1938) as

$$\frac{E}{V} = \frac{g}{4\pi^2} \left( \frac{2m}{\beta\hbar^2} \right)^{\frac{3}{2}} \frac{1}{\beta} \Gamma\left(\frac{5}{2}\right) \zeta\left(\frac{5}{2}\right). \quad (63)$$

The constant-volume heat capacity for  $\beta > \beta_0$  becomes

$$C_V = \frac{5}{2} \left[ \frac{\Gamma\left(\frac{5}{2}\right) \zeta\left(\frac{5}{2}\right)}{\Gamma\left(\frac{3}{2}\right) \zeta\left(\frac{3}{2}\right)} N k_B \left(\frac{\beta}{\beta_0}\right)^{-\frac{3}{2}} \right], \quad (64)$$

and the pressure for  $\beta > \beta_0$  becomes

$$P = \frac{2}{3} \frac{2\sqrt{2}}{4\pi^2} \Gamma\left(\frac{5}{2}\right) \zeta\left(\frac{5}{2}\right) \frac{m^{\frac{3}{2}} g}{\hbar^3} \beta^{-\frac{5}{2}}. \quad (65)$$

It is interesting to find that the pressure approaches to zero as temperature goes to zero, *i.e.*  $\beta \rightarrow \infty$ . In other words, the Bose gas exerts no force on the walls of the container at  $T = 0$ , because all of the particles condensate in the zero-momentum state. The pressure is independent of the number density  $N/V$ , depending only on temperature  $\beta$ . The two different summations above and below the temperature  $\beta_0$  lead us that the heat capacity as a function of temperature has discontinuity in its slope (Landau & Lifshitz, 1980) as

$$\Delta \left( \frac{\partial C_V}{\partial V} \right)_{\beta_0} = -\frac{27}{4} \left[ \frac{\Gamma\left(\frac{3}{2}\right) \zeta\left(\frac{3}{2}\right)}{\pi} \right]^2 N k_B^2 \beta_0^{-1}. \quad (66)$$

This implies that a homogeneous massive ideal Bose gas system exhibits a *phase transition* at  $\beta_0$  *without* interaction. This phenomenon is known as the *Bose-Einstein condensation*. A good review for the realization of the Bose-Einstein condensation is provided by Cornell & Wieman (2002).

### 3.3 Elementary excitations as massless boson gas

As we stated previously, some elementary excitations emerged by the spontaneous symmetry breaking are *massless* bosons (Goldstone, 1961; Goldstone et al., 1962) as well as gauge bosons, which are elemental particles, *e.g.* photons, arisen from the fundamental interactions, *electromagnetic fields* for photons, with gauge degrees of freedom. Whether a boson is a Goldstone boson or a gauge one, the procedure described above is not applicable to the massless character of the boson, because its energy spectrum is not in the form of Eq. (41).

One has to remind that a massless boson does not carry mass, but it carries momentum and energy. The energy spectrum of a massless boson is given by

$$\epsilon = \hbar\omega \quad (67)$$

and the frequency  $\omega$  is obtained by the corresponding momentum  $\mathbf{p} = \hbar\mathbf{k}$  through a dispersion relation

$$\omega = \omega(\mathbf{k}). \quad (68)$$

The number of bosons  $N$  in the massless boson gas is a variable, and not a given constants as in an ordinary gas. Therefore,  $N$  itself must be determined from the thermal equilibrium condition, the (Helmholtz) free energy minimum  $(\partial F/\partial N)_{T,V} = 0$ . Since  $(\partial F/\partial N)_{T,V} = \mu$ , this gives

$$\mu = 0. \quad (69)$$

In these conditions, the mean occupation number is following the Planck distribution function (Planck, 1901)

$$n_{\mathbf{k}}^0 = \frac{1}{e^{\beta\hbar\omega_{\mathbf{k}}} - 1}, \quad (70)$$

which was originally suggested for describing the distribution function problem of black-body radiations. Note that the Planck distribution function is a special case of the Bose-Einstein distribution function with zero chemical potential.

Considering the relation Eq. (4) and the condition Eq. (69), the grand potential of a massless Bose gas subsystem becomes the same as the Helmholtz free energy. The details of the thermodynamic properties are depending on the dispersion relation of the bosons. Photons are quantized radiations based on the fact of the *linearity* of electrodynamics,<sup>3</sup> so that photons do not interact with one another. The photon dispersion relation is linear,

$$\omega = ck, \quad (71)$$

where  $k$  includes the definite polarizations. The photon Helmholtz free energy is calculated as

$$F_0 = T \frac{V}{\pi^2 c^3} \int_0^\infty d\omega \omega^2 \ln(1 - e^{-\beta\hbar\omega}). \quad (72)$$

The standard integration method yields

$$F = -\frac{4}{3} \frac{\sigma V}{c} T^4, \quad (73)$$

where  $\sigma$  is called the *Stefan-Boltzmann constant* defined as

$$\sigma = \frac{\pi^2}{60} \frac{k_{\text{B}}^4}{\hbar^3 c^2}. \quad (74)$$

The entropy is

$$S = -\frac{\partial F}{\partial T} = \frac{16}{3} \frac{\sigma V}{c} T^3. \quad (75)$$

The total radiation energy  $E = F + TS$  is

$$E = \frac{4\sigma V}{c} T^4 = -3F, \quad (76)$$

which proportional to the fourth power of the temperature; *Boltzmann's law*. The constant volume heat capacity of the radiation is

$$C_V = \left( \frac{\partial E}{\partial T} \right)_V = \frac{16\sigma V}{c} T^3, \quad (77)$$

and the pressure is

$$P = -\left( \frac{\partial F}{\partial V} \right)_T = \frac{4\sigma}{3c} T^4. \quad (78)$$

Hence, the equation of states of the photon gas is

$$PV = \frac{1}{3}E. \quad (79)$$

<sup>3</sup> The nonlinear character appeared in the quantum electrodynamics will not be discussed here.

The procedure for the photonic subsystem is quite useful in computing thermodynamics for many kinds of elementary excitations, which are usually massless (Goldstone) bosons in a condensed matter system. It is predicted that any crystal must be completely ordered, and the atoms of each kind must occupy entirely definite positions, in a state of complete thermodynamic equilibrium (Andreev & Lifshitz, 1969; Leggett, 1970). It is well known that the ions vibrate even in the zero temperature with several vibration modes (Aschcroft & Mermin, 1976; Callaway, 1974; Jones & March, 1973a; Kittel, 2005; Landau & Lifshitz, 1980; Madelung, 1978; Pines, 1999). The energy spectrum of a phonon in a mode  $j$  and a wavevector  $\mathbf{q}$  contains the zero vibration term

$$\epsilon_{i\mathbf{q}} = \left( n_{j\mathbf{q}} + \frac{1}{2} \right) \hbar\omega_j(\mathbf{q}), \quad (80)$$

where  $n_{j\mathbf{q}}$  is the occupation number of the single-particle modes of  $j$  and  $\mathbf{q}$ . The corresponding partition function is then written as

$$\begin{aligned} Z &= \sum_{n_{j,\mathbf{q}}} \exp \left[ -\beta\epsilon_j(\mathbf{q}) \right] \\ &= \prod_{j,\mathbf{q}} \frac{\exp \left[ -\beta\hbar\omega_j(\mathbf{q}) \right]}{1 - \exp \left[ -\beta\hbar\omega_j(\mathbf{q}) \right]}. \end{aligned} \quad (81)$$

The Helmholtz free energy of the phonon subsystem is

$$F = -\beta \ln Z = \beta \sum_{j,\mathbf{q}} \ln \left\{ 2 \sinh \left[ \frac{\beta\hbar\omega_j(\mathbf{q})}{2} \right] \right\}. \quad (82)$$

In the noninteracting phonon gas condition, the entropy  $S$ , internal energy  $E$ , and the volume constant specific heat  $C$  becomes

$$\begin{aligned} S &= - \left( \frac{\partial F}{\partial T} \right)_V \\ &= k_B \sum_{j,\mathbf{q}} \left[ \frac{\beta}{2} \hbar\omega_j(\mathbf{q}) \coth \left( \frac{\beta}{2} \hbar\omega_j(\mathbf{q}) \right) - \ln \left\{ 2 \sinh \left( \frac{\beta}{2} \hbar\omega_j(\mathbf{q}) \right) \right\} \right]. \end{aligned} \quad (83)$$

$$\begin{aligned} E &= F - TS - \left( \frac{\partial F}{\partial T} \right)_V \\ &= \sum_{j,\mathbf{q}} \left\{ \frac{\hbar\omega_j(\mathbf{q})}{2} + \frac{\hbar\omega_j(\mathbf{q})}{e^{\beta\hbar\omega_j(\mathbf{q})} - 1} \right\}, \end{aligned} \quad (84)$$

$$C_V = \left( \frac{\partial E}{\partial T} \right)_V = k_B \frac{\sum_{j,\mathbf{q}} \left[ \frac{\hbar\omega_j(\mathbf{q})}{2} \right]^2}{\sinh \left[ \frac{\hbar\omega_j(\mathbf{q})}{2} \right]}. \quad (85)$$

Although the realistic phonon dispersion relations are complicated (Aschcroft & Mermin, 1976; Kittel, 2005), there are two useful model dispersion relations of phonons. Einstein (1907; 1911) modelled the density of states as constant frequencies in each vibrational mode as  $\mathcal{D}(\omega) = N\delta(\omega - \omega_0)$ , where delta function is centered at  $\omega_0$ . This model is, in turn, useful



to treat high temperature phonon thermodynamics. The thermal energy of the noninteracting phonon system is

$$E = Nn^0\hbar\omega = \frac{N\hbar\omega}{e^{\beta\hbar\omega} - 1}, \quad (86)$$

and so the heat capacity of the system is

$$C_V = \left( \frac{\partial E}{\partial T} \right)_V = Nk_B (\beta\hbar\omega)^2 \frac{e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega} - 1)^2}. \quad (87)$$

On the other hand, Debye (1912) approximated that the velocity of sound is taken as constant for each phonon mode with the dispersion relation  $\omega \sim q$ . In this case, the method used for the photon gas is directly applied. At low temperature limit, the Debye extracted the  $T^3$  law of the heat capacity as

$$C_V \simeq \frac{12\pi^4}{5} nk_B \left( \frac{T}{\Theta_D} \right)^3, \quad (88)$$

where  $\Theta_D$  is known as the Debye temperature.

#### 4. Matters as interacting liquids

The discussions in Sec. 3 has succeed to describe the thermodynamic properties of materials in many aspects, and hence such descriptions were treated in many textbooks. However, the oversimplified model fails the many important features on the material properties. One of the important origin of such failures is due to the ignorance of the *electromagnetic* interaction among the constituent particles; electrons and ions, which carry electric charges. However, the inclusion of interactions among the particles is enormously difficult to treat. To date the quantum field theory (QFT) is known as the standard method in dealing with the interacting particles. There are many good textbooks on the QFT (Berestetskii et al., 1994; Bjorken & Drell, 1965; Doniach & Sondheimer, 1982; Fetter & Walecka, 2003; Fradkin, 1991; Gross, 1999; Itzykson & Zuber, 1980; Mahan, 2000; Negele & Orland, 1988; Parisi, 1988; Peskin & Schroeder, 1995; Zinn-Justin, 1997) in treating the interacting particles systematically in various aspects. In this article, the idea of the treatments will be reviewed briefly, instead of dealing with the full details.

The idea of noninteracting particles inspires an idea to deal with the electronic subsystem as a sum of independent particles under a given potential field (Hartree, 1928), with the consideration of the effect of Pauli exclusion principle (Fock, 1930), which it is known as the exchange effect. This idea, known as the Hartree-Fock method, was mathematically formulated by introducing the Slater determinant (Slater, 1951) for the many-body electronic wave function. The individual wave function of an electron can be obtained by solving either Schrödinger equation (Schrödinger, 1926a;b;c;d) for the nonrelativistic cases or Dirac equation (Dirac, 1928a;b) for the relativistic ones.<sup>4</sup>

Since an electron carries a fundamental electric charge  $e$  in its motion, it is necessary to deal with electromagnetic waves or their quanta photons. Immediate necessity was arisen in order to deal with both electrons and photons in a single quantum theoretical framework in consideration of the Einstein's special theory of relativity (Einstein, 1905). Jordan & Pauli (1928) and Heisenberg & Pauli (1929) suggested that a new formalism to treat both the

<sup>4</sup> The immediate relativistic version of the Schrödinger equation was derived by Gordon (1926) and Klein (1927), known as the Klein-Gordon equation. The Klein-Gordon equation is valid for the Bose-Einstein particles, while the Dirac equation is valid for the Fermi-Dirac particles.

electrons and the radiations as quantized objects in such a way of a canonical transformation to the normal modes of their fields; this method is called the *second quantization* or the *field quantization*. The canonical transformation technique to the normal mode is a well established classical method for continuous media (Goldstein, 1980). The idea treats both the electrons and radiations as continuous fields and quantized them for their own normal modes (Heisenberg & Pauli, 1929; Jordan & Pauli, 1928). The practically available solutions was suggested for the nonrelativistic case by Bethe (1947) followed by the fully relativistic case by Dyson (1949a;b); Feynman (1949a;b); Schwinger (1948; 1949a;b); Tomonaga (1946), so called the *renormalization* for cancelling the unavoidable divergencies appeared in the quantum field theory. This kind of theory on the electrons and radiations is called as the quantum electrodynamics (QED), which is known to be the most precise theory ever achieved (Peskin & Schroeder, 1995) with the error between the theory and experiment to be less than a part per billion (ppb) (Gabrielse et al., 2006; 2007; Odom et al., 2006).

#### 4.1 The concepts of quantum field theory

Feynman (1949b) visualized the underlying concept of the quantum field theory by reinterpreting the nonrelativistic Schrödinger equation with the Green's function concept. As in classical mechanics, a Hamiltonian operator  $\hat{H}$  contains all the mechanical interactions of the system. The necessary physical information of the system is contained in the wave function  $\psi$ . The Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi, \quad (89)$$

describes the change in the wave function  $\psi$  in an infinitesimally time interval  $\Delta t$  as due to the operation if an operator is  $e^{-i\frac{\hat{H}}{\hbar}\Delta t}$ . This description is equivalent to the description that the wave function  $\psi(\mathbf{x}_2, t_2)$  at  $\mathbf{x}_2$  and  $t_2$  is evolved one from the wave function  $\psi(\mathbf{x}_1, t_1)$  at  $\mathbf{x}_1$  and  $t_1$  through the equation

$$\psi(\mathbf{x}_2, t_2) = \int K(\mathbf{x}_2, t_2; \mathbf{x}_1, t_1) \psi(\mathbf{x}_1, t_1) d^3\mathbf{x}_1, \quad (90)$$

where  $K$  is the kernel of the evolution and  $t_2 > t_1$ . If  $\psi(\mathbf{x}_1, t_1)$  is expanded in terms of the eigen function  $\phi_n$  with the eigenvalue  $E_n$  of the constant Hamiltonian operator  $\hat{H}$  as  $\sum_n c_n \phi_n(\mathbf{x})$ , one can find for  $t_2 > t_1$

$$K(2, 1) = \sum_n \phi_n(\mathbf{x}_2) \phi_n^*(\mathbf{x}_1) e^{-i\frac{E_n}{\hbar}(t_2-t_1)}, \quad (91)$$

where we abbreviated 1 for  $\mathbf{x}_1, t_1$  and 2 for  $\mathbf{x}_2, t_2$  and define  $K(2, 1) = 0$  for  $t_2 < t_1$ . It is straightforward to show that  $K$  can be defined by that solution of

$$\left( i\hbar \frac{\partial}{\partial t_2} - \hat{H}_2 \right) K(2, 1) = i\hbar \delta(2, 1), \quad (92)$$

where  $\delta(2, 1) = \delta(t_2 - t_1) \delta(x_2 - x_1) \delta(y_2 - y_1) \delta(z_2 - z_1)$  and the subscript 2 on  $\hat{H}_2$  means that the operator acts on the variables of 2 of  $K(2, 1)$ . The kernel  $K$  is now called as the *Green's function* and it is the total amplitude for arrival at  $\mathbf{x}_2, t_2$  starting from  $\mathbf{x}_1, t_1$ . The transition amplitude for finding a particle in state  $\chi(2)$ , if it was in  $\psi(1)$ , is

$$\int \chi^*(2) K(2, 1) \psi(1) d^3\mathbf{x}_1 d^3\mathbf{x}_2. \quad (93)$$

A quantum mechanical system is described equally well by specifying the function  $K$ , or by specifying the Hamiltonian operator  $\hat{H}$  from which it results.

Let us consider a situation that a particle propagates from 1 to 2 through 3 in a weak potential operator  $\hat{U}(3)$ , which differs from zero only for  $t$  between  $t_1$  and  $t_2$ . The kernel is expanded in powers of  $\hat{U}$  that

$$K(2, 1) = K_0(2, 1) + K^{(1)}(2, 1) + K^{(2)}(2, 1) + \dots \quad (94)$$

To zeroth order in  $\hat{U}$ ,  $K$  is that for a free particle,  $K_0(2, 1)$ . Let us consider the situation if  $U$  differs from zero only for the infinitesimal time interval  $\Delta t_3$  between some time  $t_3$  and  $t_3 + \Delta t_3$  for  $t_1 < t_3 < t_2$ . The particle will propagate from 1 to 3 as a *free* particle,

$$\psi(3) = \int K_0(3, 1) \psi(1) d^3x_1. \quad (95)$$

For the short time interval  $\Delta t_3$ , the wave function will change to

$$\psi(\mathbf{x}, t_3 + \Delta t_3) = e^{-i\frac{\hat{H}}{\hbar}\Delta t_3} \psi(\mathbf{x}, t_3), \quad (96)$$

after solving the Schrödinger equation in Eq. (89). The particle at 2 then propagates *freely* from  $\mathbf{x}_3, t_3 + \Delta t_3$  as

$$\psi(\mathbf{x}_2, t_2) = \int K_0(\mathbf{x}_2, t_2; \mathbf{x}_3, t_3 + \Delta t_3) \psi(\mathbf{x}_3, t_3 + \Delta t_3). \quad (97)$$

We can decompose the Hamiltonian operator  $\hat{H}$  by  $\hat{H}_0 + \hat{U}$ , where  $\hat{H}_0$  is the Hamiltonian operator of the free particle. The change in wave function by  $\hat{U}$  will be

$$\Delta\psi = -\frac{i}{\hbar} \hat{U}(3) \psi(3) \Delta t_3. \quad (98)$$

The wave function at 2 is that of the propagated particle from  $t_3 + \Delta t_3$  to be

$$\psi(2) = \int K_0(2, 3') \psi(3') d^3x_3, \quad (99)$$

where  $3'$  abbreviates  $\mathbf{x}_3, t_3 + \Delta t_3$ , by a free propagation. The difference of the wave function at 2 is obtained by

$$\Delta\psi(2) = -\Delta t_3 \frac{i}{\hbar} \int K_0(2, 3) \hat{U}(3) K_0(3, 1) \psi(1) d^3x_3 d^3x_1. \quad (100)$$

Therefore, the first order expansion of the kernel  $K$  is then

$$K^{(1)}(2, 1) = -\frac{i}{\hbar} \int K_0(2, 3) \hat{U}(3) K_0(3, 1) d3, \quad (101)$$

where  $d3 = d^3x_3 dt_3$ . We can imagine that a particle travels as a free particle from point to point, but is scattered by the potential operator  $\hat{U}$  at 3. The higher order terms are also analyzed in a similar way.

The analysis for the charged free Dirac particle gives a new interpretation of the antiparticle, which has the reversed charge of the particle; for example, a positron is the antiparticle of an electron. The Dirac equation (Dirac, 1928a;b) has negative energy states of an electron. Dirac interpreted himself that the negative energy states are fully occupied in vacuum, and an

elimination of one electron from the vacuum will carry a positive charge; the unoccupied state was interpreted as a *hole*. Feynman (1949b) reinterpreted that the *hole* is a positron, which is an electron propagating backward in time. The interpretation has the corresponding classical electrodynamic picture. If we record the trajectory of an electron moving in a magnetic field, the trajectory of the electron will be bent by the Lorentz force exerting on the electron. When we reverse the record in time of the electron in the magnetic field, the bending direction of the trajectory is that of the positively charged particle with the same mass to the electron. Therefore, we understand that a particle is propagating forward in time, while the corresponding antiparticle or the hole is propagating backward in time. Due to the negative energy nature of the hole or antiparticle, a particle-hole pair will be annihilated when the particle meet the hole at a position during their propagations in space-time coincidentally. Reversely, vacuum can create the particle-hole pair from the vacuum fluctuations.

Now consider a system of two particles  $a$  and  $b$  propagate from 1 to 3 interacting at 5 for the particle  $a$  and from 2 to 4 interacting at 6 for the particle  $b$ . In the case of free particles, the kernel  $K_0$  is a simple multiple of two free particle kernels  $K_{0a}$  and  $K_{0b}$  as

$$K_0(3,4;1,2) = K_{0a}(3,1)K_{0b}(4,2). \quad (102)$$

When two particles are interacting through a two particle potential  $\hat{U}$ , the first-order expansion term of  $K$  may be written (Feynman, 1949a) as

$$K^{(1)}(3,4;1,2) = -\frac{i}{\hbar} \iint K_{0a}(3,5)K_{0b}(4,6)\hat{U}(5,6)K_{0a}(5,1)K_{0b}(6,2)d5d6. \quad (103)$$

One important difference from Eq. (101) is that the interaction  $\hat{U}$  at a specific space-time position 3 is replaced by the interaction field  $\hat{U}(5,6)$ . The interaction field is also able to be quantized and the interaction field  $\hat{U}(5,6)$  is interpreted as an interaction field quantum propagating *freely* between 5 and 6. For the case of two-electron interaction, the particles are electrons interacting through the electromagnetic interaction. One electron  $a$  propagates from 1 to 3 and the other  $b$  propagates from 2 to 4. During their propagations, the electron  $a$  emits (absorbs) a photon at point 5, while the electron  $b$  absorbs (emits) the photon at 6. The wave function of each electron differs by the emission of the photon at 5 or 6 from its wave function at the origin of the propagation. The process of emission and absorption of the photons during the electrons propagations change the energy of the electronic subsystem. The process to compute the energy of the Fermi liquid in the perturbative treatment of the interaction requires the consideration of the essential many-body treatment available by the procedures suggested by Dyson (1949a;b); Feynman (1949a;b); Schwinger (1948; 1949a;b); Tomonaga (1946).

For the future reasons, it is useful to see the consequence of the step function behavior of the kernel  $K$ . As described above,  $K(2,1)$  has its meaning as the solution of the Green's function Eq. (92) only if  $t_2 > t_1$ . It is convenient to use multiply the step function  $\theta(t_2 - t_1)$  to the kernel  $K$  for implying the physical meaning. The step function has an integral representation

$$\theta(t - t') = -\int_{-\infty}^{\infty} \frac{d\omega}{2\pi i} \frac{e^{-i\omega(t-t')}}{\omega + i\eta}, \quad (104)$$

where  $\eta$  is an positive real number. If  $t > t'$ , then the contour must be closed in the lower-half  $\omega$  plane, including the simple pole at  $\omega = -i\eta$  with residue  $-1$ . If  $t < t'$ , then the contour must be closed in the upper-half  $\omega$  plane and gives zero, because the integrand has no singularity for  $\Im\omega > 0$ .

For a noninteracting particle, the eigen function will be a plane wave

$$\phi_{\mathbf{k}}(\mathbf{x}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (105)$$

and the eigenvalue will be  $\epsilon_{\mathbf{k}}^0 = \hbar\omega_{\mathbf{k}}^0$ . In the limit of an infinite volume, the summation over  $n$ , to be over  $\mathbf{k}$ , in Eq. (91) becomes an integration and then the consideration of the identities given in Eqs. (104) and (105) yields

$$K_0(\mathbf{x}, t; \mathbf{x}', t') = \frac{1}{2\pi^4} \int d^3\mathbf{k} \int_{-\infty}^{\infty} d\omega e^{i\mathbf{k}\cdot\mathbf{x}} e^{-i\omega(t-t')} \left[ \frac{\theta(k - k_F)}{\omega - \omega_{\mathbf{k}}^0 + i\eta} + \frac{\theta(k_F - k)}{\omega - \omega_{\mathbf{k}}^0 - i\eta} \right],$$

which immediately yields

$$K_0(\mathbf{k}, \omega) = \left[ \frac{\theta(k - k_F)}{\omega - \omega_{\mathbf{k}}^0 + i\eta} + \frac{\theta(k_F - k)}{\omega - \omega_{\mathbf{k}}^0 - i\eta} \right], \quad (106)$$

where  $k_F$  is the Fermi wave vector. Since the  $\pm i\eta$  were introduced to render the time integral in Eq. (104) convergent, it is convenient to take the limit  $\eta \rightarrow 0^+$ . Eq. (106) diverges at  $\omega = \omega_{\mathbf{k}}^0 \pm i\eta$ . If we compute the transition amplitude in Eq. (93) from the state  $\psi(1)$  to the state  $\psi(2)$  in the Fourier transformed space, the kernel function form of Eq. (106) implies that the transition amplitude will be maximum around  $\omega = \omega_{\mathbf{k}}^0 \pm i\eta$ . In the limit  $\eta \rightarrow 0^+$ , the free particle remains in the state  $\mathbf{k}$ : the particle will keep its momentum and hence its (kinetic) energy. This is nothing more than the celebrating statement of *inertial motion* by Galileo.

#### 4.2 Self-energy

This idea that a particle propagates freely until it faced with the a scattering center, where the particles emit or absorb the interacting quanta, is nothing more than an extension of the model introduced by Drude (1900) for electrons in metals. We already obtained the thermodynamic information of noninteracting gases in Sec. 3. Hence, the remaining task is to see the effects of the interaction from the noninteracting gas.

Let us come back to the case of a particle propagating from 1 to 2 in the way given in Eq. (101) by considering the interaction process demonstrated in Eq. (103). The perturbation procedure for the interacting fermions includes, in its first-order expansion, two fundamental processes (Fetter & Walecka, 2003), which are the prototypes of the interactions of all order perturbation expansion. For the first case, the particle  $a$  propagates from 1 to 3, emits (absorbs) a boson propagating to 4, where the other particle  $b$  absorbs (emits) the boson. The particle  $b$  propagates after the absorption (emission) at 4 to the position 4 again, just before it absorbing (emitting) the boson. This process is known as the *vacuum polarization* and it is equivalent, for an electronic system, to the method of Hartree (1928). In terms of the classical electrodynamics, the process is nothing more than that an electron moves in a Coulomb potential generated by the neighboring charge density. Secondly, the particle propagates firstly from 1 to 3 and it emits (absorbs) a boson at 3 to change its state. The particle in the new state then propagates from 3 to 4, where the new state particle absorbs (emits) the boson propagated from 3, and change its states to the original one in propagating to 2. This process is known as the *exchange* and it is equivalent, for the electronic system, to the Fock (1930) consideration of the Pauli's exclusion principle.

The energy spectrum of the particle itself will change during both the processes. Dyson (1949a,b) discussed that the changing the particle energy itself by the perturbative treatment of

the fermion interaction when we consider single particle propagation from 1 to 2. The energy of the particle differs from the noninteracting particle propagation, and it can be systematically included to the particle propagation kernel as

$$K(2,1) = K_0(2,1) + \iint K(2,4)\Sigma(4,3)K(3,1)d3d4, \quad (107)$$

where  $\Sigma$  is known as the *self-energy*. When we consider the all order perturbation, the exact single-particle propagation can be obtained by using the successive self-energy inclusion as

$$K(2,1) = K_0(2,1) + \iint K(2,4)\Sigma(4,3)K(3,1)d3d4 \\ + \iiint K(2,6)\Sigma(6,5)K(5,4)\Sigma(4,3)K(3,1)d3d4d5d6 + \dots, \quad (108)$$

with the special care of the self-energy to be proper.<sup>5</sup> This equation is known as the *Dyson equation*.

The consequence of the Dyson equation can be seen easily if we perform a four dimensional Fourier transform on Eq. (108) with respect to the difference  $\mathbf{y} - \mathbf{x}$ ,  $t_2 - t_1$  into the momentum space to an algebraic form

$$K(k) = K_0(k) + K_0(k)\Sigma(k)K(k), \quad (109)$$

where  $k$  abbreviates  $(\mathbf{k}, \omega)$ . We can solve Eq. (109) as

$$K(k) = \frac{1}{K_0^{-1}(k) - \Sigma(k)}. \quad (110)$$

Considering the self-energy  $\Sigma$  is complex, the  $i\eta$  in the free particle kernel Eq. (106) is no more relevant. Therefore, we obtain the solution of the Dyson equation as

$$K(k) = K(\mathbf{k}, \omega) = \frac{1}{\omega - \hbar^{-1}\epsilon_{\mathbf{k}}^0 - \Sigma(\mathbf{k}, \omega)}. \quad (111)$$

The physical meaning of Eq. (111) is straightforward: an interacting particle propagates as the free particle does, but its excitation energy differs by a *dressing* term  $\Sigma(\mathbf{k}, \omega)$ .

Lehmann (1954) and Galitskii & Migdal (1958) discussed the usefulness of Eq. (111) in the applications for many-body systems. In the Lehmann representation, the frequency  $\omega$  is a complex number to be

$$\hbar\omega = \epsilon_{\mathbf{k}} - i\gamma_{\mathbf{k}}, \quad (112)$$

where  $\gamma_{\mathbf{k}}$  is the damping of the particle. The singularity of the exact Green's function  $K(\mathbf{k}, \omega)$ , considered as a function of  $\omega$ , determine both the excitation energy  $\epsilon_{\mathbf{k}}$  of the system and its damping  $\gamma_{\mathbf{k}}$ . Furthermore, the chemical potential can be determined as the point where  $\Im\Sigma(\mathbf{k}, \omega)$  changes the sign, because

$$\Im\Sigma(\mathbf{k}, \omega) \geq 0, \quad \omega < \mu/\hbar \\ \Im\Sigma(\mathbf{k}, \omega) \leq 0, \quad \omega > \mu/\hbar. \quad (113)$$

<sup>5</sup> The proper implies the terms that cannot be disintegrated into the lower order expansion terms during the perturbation expansion.

A similar analysis can be carried out for the interaction between two particles, which always consists of the lowest-order interaction plus a series of proper expansion. The four dimensional Fourier transformation to the  $q$  coordinates yields

$$U(q) = \frac{U_0(q)}{1 - \Pi(q)U_0(q)}. \quad (114)$$

Introducing a generalized dielectric function

$$\kappa(q) = 1 - U_0(q)\Pi(q), \quad (115)$$

the *screening* of the lowest-order interaction by the polarization of the medium is obtained as

$$U(q) = \frac{U_0(q)}{\kappa(q)}. \quad (116)$$

### 4.3 Goldstone's theorem: the many-body formalism

Goldstone (1957) provided a new picture of the many-body systems with the quantum field theoretic point of view, presented above. Let us the free particle Hamiltonian  $\hat{H}$  has a many-body eigenstate  $\Phi$ , which is a determinant formed from  $N$  particles of the  $\psi_n$ , and which is able to be described by enumerating these  $N$  one-particle states. Suppose that  $\hat{H}_0$  has a non-degenerate ground state  $\Phi_0$  formed from the lowest  $N$  of the  $\psi_n$ . The states  $\psi_n$  occupied in  $\Phi_0$  will be called unexcited states, and all the higher states  $\psi_n$  will be called excited states. An eigenstate  $\Phi$  of  $\hat{H}_0$  can be described by enumerating all the excited states which are occupied, and all the unexcited states which are not occupied. An unoccupied unexcited state is regarded as a *hole*, and the theory will deal with particles in excited states and holes in the unexcited states. In this treatment, the ground state  $\Phi_0$  is considered as a new *vacuum*, a particle is considered as an occupied states in the excited states, and the hole is essentially different from the positrons, in which are the symmetric counterpart of the electrons. Goldstone (1957) derived the energy difference between the system with and without interactions,  $\hat{H}_1$ , in the Dirac notation,<sup>6</sup> as

$$E - E_0 = \langle \Phi_0 | \hat{H}_1 \sum_n \left( \frac{1}{E_0 - \hat{H}_0} \hat{H}_1 \right)^n | \Phi_0 \rangle, \quad (117)$$

where the summation should do on the linked<sup>7</sup> terms of the perturbation. The noninteracting Hamiltonian  $\hat{H}_0$  in the denominator can be replaced by the corresponding eigenvalues, because Eq. (117) is interpreted by inserting a complete set of eigenstates of  $\hat{H}_0$  between each interaction  $\hat{H}_1$ . The physical situation can be visualized as follows: (1) The interaction Hamiltonian  $\hat{H}_1$  operate on  $|\Phi_0\rangle$  creates a state with two particles and two holes. This state propagates with  $(E_0 - \hat{H}_0)^{-1}$ . (2) The next  $\hat{H}_1$  can create more particle-hole pairs or scatter the existing particle-hole pair and so on. (3) The final  $\hat{H}_1$  must then return the system to the ground state  $|\Phi_0\rangle$ . This process gives the difference in energy of the interacting many-body system from the noninteracting one. By choosing the first-order perturbation in

<sup>6</sup> In the Dirac notation, a quantum state  $n$  is written in the Hilbert space of form  $|n\rangle$  and the corresponding conjugate state is written as  $\langle n|$ . The wave function is the projection to the position space, such that  $\psi_n(\mathbf{x}) = \langle \mathbf{x} | n \rangle$ . For an operator  $\hat{A}$ ,  $\langle m | \hat{A} | n \rangle$  is called as a *matrix element* to represent the probability for transtion from the state  $|n\rangle$  by the operation  $\hat{A}$  to the state  $\langle m|$ . Readers can see the details in Dirac (1998); Sakurai (1994).

<sup>7</sup> During the expansion, there are terms describing pair creation and annihilations corresponding to the free particle Green's function.

the Goldstone scheme, it is successfully reproduced the Hartree-Fock equation. The method for the perturbation expansion in the Goldstone scheme is usually called as coupled cluster expansion.

#### 4.4 Matsubara method for finite temperatures

For simplicity, consider a system with one kind of particles in dealing with the grand partition function defined in Eq. (2), *i.e.*,

$$\Xi = \sum_i e^{-\beta(\epsilon_i - \mu N)}. \quad (118)$$

It is convenient to introduce a *grand canonical Hamiltonian* operator

$$\hat{\mathcal{K}} = \hat{H} - \mu \hat{N}. \quad (119)$$

The operator form forces the equivalent form of Eq. (118) to be a trace

$$\Xi = e^{-\beta\Omega} = \text{Tre}^{-\beta\hat{\mathcal{K}}}, \quad (120)$$

and its corresponding statistical density matrix in the form

$$\hat{\rho} = \frac{e^{-\beta\hat{\mathcal{K}}}}{\Xi} = e^{\beta(\Omega - \hat{\mathcal{K}})}. \quad (121)$$

Armed with Eqs. (120) and (121), all the thermodynamic information are calculable. However, difficulty is in its thermally evolving nature of those equations, instead of their time evolving nature of the Schrödinger equation.

In order to compute thermodynamic properties of ferromagnetic materials, Bloch (1932) suggested a statistical equation, resembling the ordinary Schrödinger equation, by treating the evolution of the statistical density with respect to temperature as

$$-\frac{\partial \rho}{\partial \beta} = \hat{H}\rho. \quad (122)$$

Matsubara (1955) and Gaudin (1960) proved that the Green's function formalism in quantum field theory also satisfies the Wick's theorem (Wick, 1950), hence the perturbation expansion in all order satisfies, when the "single-particle" field operator follows the Bloch equation by switching the real-time  $t$  to the *imaginary-time*,  $\tau = it$ . Note that the term "single-particle" does not mean a real single-particle.

The procedure in Sec. 4.1 is applied also. The single-particle temperature Green's function  $\mathcal{G}$  may follow the Green's function equation

$$\left[ \hbar \frac{\partial}{\partial \tau_2} - \hat{\mathcal{K}} \right] \mathcal{G}(\mathbf{x}_1 \tau_1; \mathbf{x}_2 \tau_2) = \pm \hbar \delta(\mathbf{x}_2 - \mathbf{x}_1) \delta(\tau_2 - \tau_1), \quad (123)$$

where  $\pm$  indicates the temperature Green's function for fermions and bosons, respectively. However, the temperature Green's function does not depend on  $\tau_1$  and  $\tau_2$  separately, but it depends on the difference  $\tau \equiv \tau_1 - \tau_2$ . Each imaginary time  $\tau$  has  $\beta\hbar$  periodicity. The temperature Green's function is then in the form

$$\mathcal{G}(\mathbf{x}_1, \tau_1; \mathbf{x}_2, \tau_2) = \mathcal{G}(\mathbf{x}_1, \mathbf{x}_2; \tau). \quad (124)$$



For both fermionic and bosonic *statistics*,  $\mathcal{G}$  is *periodic* over the range  $2\beta\hbar$  and may expanded in a Fourier serie

$$\mathcal{G}(\mathbf{x}_1, \mathbf{x}_2, \tau) = \frac{1}{\beta\hbar} \sum_n e^{i\omega_n \tau} \mathcal{G}(\mathbf{x}_1, \mathbf{x}_2, \omega_n), \quad (125)$$

where

$$\omega_n = \frac{n\pi}{\beta\hbar}. \quad (126)$$

This shows that  $e^{-i\omega_n \beta\hbar}$  is equal to  $e^{-in\pi} = (-1)^n$ , and the factor  $\frac{1}{2} (1 \pm e^{-i\omega_n \beta\hbar})$  reduces to  $\frac{1}{2} [1 \pm (-1)^n]$ . The Fourier coefficient in Eq. (125) is then

$$\mathcal{G}(\mathbf{x}_1, \mathbf{x}_2, \omega_n) = \int_0^{\beta\hbar} d\tau e^{i\omega_n \tau} \mathcal{G}(\mathbf{x}_1, \mathbf{x}_2, \tau), \quad (127)$$

where

$$\omega_n = \begin{cases} \frac{2n\pi}{\beta\hbar}, & \text{boson} \\ \frac{(2n+1)\pi}{\beta\hbar}, & \text{fermion} \end{cases}. \quad (128)$$

This is known as the *Matsubara frequency*.

The coordinates in  $\mathbf{x}$  can be transformed to the momentum space  $\mathbf{k}$ . There is a corresponding solution of the Dyson equation in Eq. (110) for the temperature Green's function as

$$\mathcal{G}(\mathbf{k}, \omega_n) = \frac{1}{\mathcal{G}_0^{-1} - \Sigma(\mathbf{k}, \omega_n)}, \quad (129)$$

with the noninteracting temperature Green's function of the form

$$\mathcal{G}_0(\mathbf{k}, \omega_n) = \frac{1}{i\omega_n - \hbar^{-1}(\epsilon_{\mathbf{k}}^0 - \mu)}. \quad (130)$$

The thermodynamic variables  $N$ ,  $E$ , and  $\Omega$  can by obtained by the Eq. (129): The grand potential is

$$\begin{aligned} \Omega(T, V, \mu) &= \Omega_0(T, V, \mu) \\ &\mp V(2s+1) \int_0^1 \frac{d\lambda}{\lambda} \int \frac{d^3k}{(2\pi)^2} \frac{1}{\beta\hbar} \sum_n e^{i\omega_n \eta} \\ &\quad \times \left[ \frac{1}{2}\hbar + \frac{\frac{1}{2}\Sigma^\lambda(\mathbf{k}, \omega_n)}{i\omega_n - \hbar^{-1}(\epsilon_{\mathbf{k}}^0 - \mu) - \Sigma^\lambda(\mathbf{k}, \omega_n)} \right], \end{aligned} \quad (131)$$

where  $s$  is the spin of the particles,  $\eta$  is the infinitesimally small positive number, and  $\lambda$  is the interaction strength control parameter.<sup>8</sup> The internal energy is

$$\begin{aligned} E(T, V, \mu) &= \mp V(2s+1) \int \frac{d^3k}{(2\pi)^2} \frac{1}{\beta\hbar} \sum_n e^{i\omega_n \eta} \\ &\quad \times \left[ \frac{1}{2}\hbar + \frac{\epsilon_{\mathbf{k}}^0 + \frac{1}{2}\Sigma^\lambda(\mathbf{k}, \omega_n)}{i\omega_n - \hbar^{-1}(\epsilon_{\mathbf{k}}^0 - \mu) - \Sigma^\lambda(\mathbf{k}, \omega_n)} \right], \end{aligned} \quad (132)$$

<sup>8</sup> When an interacting Hamiltonian  $\hat{H}_I$  is applied to the noninteracting system with the Hamiltonian  $\hat{H}_0$ , the interaction strength is controlled by the parameter  $\lambda$  to yield Hamiltonian  $\hat{H} = \hat{H}_0 + \lambda\hat{H}_I$  with its  $\lambda$  parameterized eigenstate  $|\Psi_0(\lambda)\rangle$ . The energy difference  $E - E_0$  is calculated by  $\int_0^1 \frac{d\lambda}{\lambda} \langle \Psi_0(\lambda) | \lambda \hat{H}_I | \Psi_0(\lambda) \rangle$ .

and the number of particles is

$$N(T, V, \mu) = \mp V (2s + 1) \int \frac{d^3k}{(2\pi)^2} \frac{1}{\beta \hbar} \sum_n \frac{e^{i\omega_n \eta}}{i\omega_n - \hbar^{-1}(\epsilon_{\mathbf{k}}^0 - \mu) - \Sigma^\lambda(\mathbf{k}, \omega_n)}. \quad (133)$$

The temperature Green's function in Eq. (129) is very similar to Eq. (110), which includes the real single-particle energy spectrum. However, there is one important difference that the frequency is not a real frequency or energy, because  $\omega_n$  is a *discrete* variable. The computation of the energy spectra is available by using the temperature Green's function.

It is convenient to introduce a real-time Green's function corresponding to the real-time trace in Eq. (120) multiplied by the statistical grand distribution Eq. (121). The real-time Green's function is then written (Fetter & Walecka, 2003) in the momentum space representation as

$$\bar{G}(\mathbf{k}, \omega) = \frac{1}{1 \mp e^{-\beta \hbar \omega}} \bar{G}^R(\mathbf{k}, \omega) + \frac{1}{1 \mp e^{\beta \hbar \omega}} \bar{G}^A(\mathbf{k}, \omega). \quad (134)$$

Let us introduce a generic real-time Green's function of a complex variable  $z$  which satisfies

$$\bar{\Gamma}(\mathbf{k}, z) = \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{A(\mathbf{k}, \omega')}{z - \omega'}, \quad (135)$$

where  $A(\mathbf{k}, \omega')$  is the spectral weight function (Galitskii & Migdal, 1958), which yields the integral representations of the *retarded* Green's function

$$\bar{G}^R(\mathbf{k}, \omega) = \bar{\Gamma}(\mathbf{k}, \omega + i\eta) = \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{A(\mathbf{k}, \omega')}{\omega - \omega' + i\eta}, \quad (136)$$

and the *advanced* Green's function

$$\bar{G}^A(\mathbf{k}, \omega) = \bar{\Gamma}(\mathbf{k}, \omega - i\eta) = \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{A(\mathbf{k}, \omega')}{\omega - \omega' - i\eta}. \quad (137)$$

The spectral weight function should satisfy the sum rule

$$\int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} A(\mathbf{k}, \omega') = 1 \quad (138)$$

for both bosons and fermions. With this sum rule, we have an asymptotic behavior of the Green's function

$$\bar{\Gamma}(\mathbf{k}, \omega) \sim \frac{1}{\omega} \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} A(\mathbf{k}, \omega') \sim \frac{1}{\omega} \quad \text{for } |\omega| \rightarrow \infty. \quad (139)$$

The spectral function constructs also the temperature Green's function

$$\mathcal{G}(\mathbf{k}, \omega_n) = \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{A(\mathbf{k}, \omega')}{i\omega_n - \omega'}. \quad (140)$$

In other words, the function  $\bar{\Gamma}(\mathbf{k}, z)$  in Eq. (135) determines the temperature Green's function  $\mathcal{G}$  as well as the real-time Green's functions  $\bar{G}^R$  and  $\bar{G}^A$  by analytic continuation technique in the complex  $\omega$  plane. The sum rule ensures the uniqueness of the analytic continuation (Baym & Mermin, 1947). Once we evaluate the temperature Green's function  $\mathcal{G}$  only at the

discrete set of points  $\{i\omega_n\}$ , the thermodynamic variables  $\Omega$ ,  $E$ , and  $N$  are obtained. The spectral weight function is able to be derived from the temperature Green's function  $\mathcal{G}$ . Then we can have the real-time Green's functions  $\bar{G}^R$  and  $\bar{G}^A$ , which give us the quasiparticle energy spectrum.

Hence, we complete the equilibrium thermodynamics; information on macroscopic thermodynamic variables as well as microstates.

## 5. Irreversible processes

The Green's function method reviewed above describes the equilibrium thermodynamics of a matter composed of fermions with the interactions mediated by the corresponding bosons. The power of the Green's function method does not limited to the equilibrium thermodynamics, but does solve many irreversible thermodynamic problems for the situation not far from the equilibrium states. The information from the irreversible processes of not-so-far from the equilibrium is extremely important for building an equilibrium phase diagram.

The procedure for acquiring the thermodynamic information includes a certain perturbation to the system for equilibrium process. For example, one should apply the external pressure to the system for changing its volume, one should apply a magnetic field  $\mathbf{H}$  to the system for measuring its magnetization, one should add heat to the system for increasing its temperature, etc. Such realistic experimental procedure should be considered for comparing the theoretical and experimental thermodynamic data assessments. However, a certain kind of phases in the system does not stable under the given condition. Even the Gibbs free energies for some metastable/unstable structures are undefinable because they are dynamically unstable. This kind of problems has been studied extensively in the CALPHAD society, see Kaufman (2001) for example.

Many difficulties in practical CALPHAD computations arise when the system has magnetic instabilities. The inclusion of magnetism in the CALPHAD method has been devoted for long time (Hillert & Jarl, 1978; Midownik, 1977). However, the current methods to build any phase diagram with magnetism are hardly thought to be completed, in terms of both the first-principles and the empirical one. The apparent reason for this failure is due to the nature of magnetism, which we have not understood its underlying physics yet.

Once magnetism involves, we are facing Invar effects (Kim, 1988; 1999), magnetostrictions (Lines, 1979), and magnetocaloric effects (de Oliveira & von Ranke, 2010). As stated in Sec. 3.2, there are spontaneous symmetry breaking related with elementary excitations, for example magnons or spin waves emerge when the spin rotation symmetry is broken. In a ferromagnetic system, the magnetic *susceptibility* which is defined by the magnetization with respect to the external magnetic field diverges below the Curie temperature  $T_C$  without external magnetic fields. In general, the computation of the magnetic susceptibility should consider the correct underlying physical mechanism. The experimental procedure for measuring any susceptibilities is the irreversible processes. This is also in great importance to understand the phase transitions.

### 5.1 Linear response theory

A theory for phase transition should be developed; the theory leaves the non-diverging error near the critical region, where a set of physical observables diverges. Kubo (1957) and Kubo et al. (1957) developed such a theory, called as the *linear response theory*, for the case of not-so-far from equilibrium, by the quantum mechanical interpretations on the Liouville' theorem in Eq. (17) and the Bloch equation in Eq. (122).

Let the Hamiltonian of the *natural motion* of a many-body system to be  $\hat{H}$ . We would like to apply an external perturbation  $\hat{H}'$  to the system, which yields now the total Hamiltonian

$$\hat{H}_{\text{total}} = \hat{H} + \hat{H}'. \quad (141)$$

We consider the external perturbation being a time dependent driving force

$$\hat{H}'(t) \sim -F(t). \quad (142)$$

When the system is in its natural motion, the statistical distribution function, or the *density matrix*,  $\hat{\rho}$ , is calculated by Eq. (121). With the external perturbation  $\hat{H}'$ , the *total* density matrix will be differ from its natural density matrix as

$$\hat{\rho}_{\text{total}} = \hat{\rho} + \hat{\rho}'. \quad (143)$$

For any physical observables  $\hat{A}$ , the thermal expectation value of  $\hat{A}$  is given by

$$\langle \hat{A} \rangle = \text{Tr}(\hat{\rho}'(t) \hat{A}) \equiv \langle A(t) \rangle. \quad (144)$$

The initial ensemble which represents statistically the initial state of the system is specified by the density matrix  $\hat{\rho}$  satisfying

$$[\hat{H}, \hat{\rho}]_{\mp} = [\hat{\mathcal{K}}, \hat{\rho}]_{\mp} = 0, \quad (145)$$

where  $[\hat{A}, \hat{B}]_{\mp}$  is the Poisson bracket to be  $\hat{A}\hat{B} \mp \hat{B}\hat{A}$ , for any quantum operators  $\hat{A}$  and  $\hat{B}$ . The symbol  $\mp$  indicates the Poisson bracket for bosons and fermions, respectively. On the other hand, the motion of the ensemble under the perturbation Eq. (142) is represented by  $\hat{\rho}'(t)$ , which obeys the equation

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} (\hat{\rho} + \hat{\rho}') &= [\hat{H} + \hat{H}', \hat{\rho} + \hat{\rho}]_{\mp} \\ &= [\hat{\mathcal{K}} + \hat{H}', \hat{\rho} + \hat{\rho}']_{\mp}. \end{aligned} \quad (146)$$

By assuming that the external perturbation is sufficiently weak, we have a *linear* equation

$$i\hbar \frac{\partial}{\partial t} \hat{\rho}' = [\hat{\mathcal{K}}, \hat{\rho}]_{\mp} + [\hat{H}', \hat{\rho}]_{\mp}, \quad (147)$$

with Eq. (145) and the neglecting the second order term,  $[\hat{H}', \hat{\rho}']_{\mp}$ .

Let us solve Eq. (147) by introducing the interaction representation of the density matrix as

$$\hat{\rho}'_I(t) = e^{i\hat{\mathcal{K}}t/\hbar} \hat{\rho}' e^{-i\hat{\mathcal{K}}t/\hbar}. \quad (148)$$

Differentiation both sides of Eq. (148) with respect to  $t$  yields

$$i\hbar \frac{\partial}{\partial t} \hat{\rho}'_I(t) = -[\hat{\mathcal{K}}, \hat{\rho}'(t)]_{\mp} + e^{i\hat{\mathcal{K}}t/\hbar} \left( i\hbar \frac{\partial}{\partial t} \hat{\rho}' \right) e^{-i\hat{\mathcal{K}}t/\hbar},$$

where the second term on the right hand side is replaced by Eq. (148) to obtain

$$i\hbar \frac{\partial}{\partial t} \hat{\rho}'_I(t) = e^{i\hat{\mathcal{K}}t/\hbar} [\hat{H}', \hat{\rho}]_{\mp} e^{-i\hat{\mathcal{K}}t/\hbar} \equiv k(t).$$

This differential equation is readily solved as

$$\hat{\rho}'_I(t) = \frac{1}{i\hbar} \int_{-\infty}^t k(t') dt' + c, \quad (149)$$

where  $c$  is a constant. Let us assume that the system is in its equilibrium at  $t = -\infty$ , i.e.,  $\hat{H}'(t = -\infty) = 0$ , and consequently  $\hat{\rho} = \hat{\rho}' = 0$  for  $t = -\infty$ . In this case, the constant  $c = 0$ . Considering the interaction representation in Eq. (148), we arrive at the solution

$$\begin{aligned} \hat{\rho}'(t) &= \frac{1}{i\hbar} e^{-i\hat{K}t/\hbar} \int_{-\infty}^t dt' k(t') e^{i\hat{K}t'/\hbar} \\ &= \frac{1}{i\hbar} \int_{-\infty}^t dt' e^{-i\hat{K}(t-t')/\hbar} [\hat{H}'(t'), \hat{\rho}]_{\mp} e^{i\hat{K}(t-t')/\hbar}. \end{aligned} \quad (150)$$

Therefore, the linear response  $\hat{A}(t)$  by the external perturbation  $\hat{H}'$  is able to be computed by using Eqs. (144) and (150) to be

$$\langle \hat{A}(t) \rangle = \frac{1}{i\hbar} \int_{-\infty}^t dt' \text{Tr} \left( e^{-i\hat{K}(t-t')/\hbar} [\hat{H}'(t'), \hat{\rho}]_{\mp} e^{i\hat{K}(t-t')/\hbar} \hat{A} \right). \quad (151)$$

When the external perturbation Hamiltonian in Eq. (142) can be written in an oscillatory form

$$\hat{H}' = -\hat{B}e^{i\omega t}, \quad (152)$$

we have the linear response as

$$\langle \hat{A}(\omega) \rangle = \pm \frac{i}{\hbar} \int_0^{\infty} e^{i\omega(t-t')} \text{Tr} \left( \hat{\rho} [\hat{A}_I((t-t')), \hat{B}]_{\mp} \right) dt(t-t'), \quad (153)$$

where  $A_I(t-t')$  is in the interaction representation as in Eq. (148) and we use the cyclic property of trace.

Since the external perturbation  $\hat{H}'$  does not appear in Eq. (153) at all, we can treat  $\hat{A}_I$  as the Heisenberg representation with respect to the natural motion described by  $\hat{K}$ . Consequently, the linear response is determined from a correlation function of the *fluctuations* of the relevant quantities in the absence of the external perturbation.

Computation of the dynamic magnetic susceptibility illustrates the usefulness of the linear response theory. When the oscillatory external field in Eq. (152) is applied in the  $\nu$ -direction, the thermal expectation value of the magnetization in the  $\mu$ -direction is obtained as

$$\langle \hat{M}_{\mu}(\omega) \rangle = \frac{i}{\hbar} \int_0^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{M}_{\mu}(\tau), \hat{M}_{\nu}]_{-} \right) H_{\nu} d\tau, \quad (154)$$

where we use the abbreviation  $\tau = t - t'$  and replaces  $\hat{B}$  by  $\hat{M}_{\nu}H_{\nu}$ , where  $H_{\nu}$  is the  $\nu$ -the component of the external magnetic field  $\mathbf{H}$ . Dynamic magnetic susceptibility, defined as

$$\chi_{\mu\nu}(\omega) \equiv \frac{\langle \hat{M}_{\mu}(\omega) \rangle}{H_{\nu}}, \quad (155)$$

is then calculated as

$$\chi_{\mu\nu}(\omega) = \frac{i}{\hbar} \int_0^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{M}_{\mu}(\tau), \hat{M}_{\nu}]_{-} \right) d\tau. \quad (156)$$

The extension to the case of the response to an magnetic field oscillating in space with the wave vector  $\mathbf{q}$  is straightforward to be

$$\chi_{\mu\nu}(\mathbf{q}, \omega) = \frac{i}{\hbar} \int_0^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{M}_\mu(\mathbf{q}, \tau), \hat{M}_\nu(-\mathbf{q})]_- \right) d\tau. \quad (157)$$

In general, a linear response function is given in the form

$$\chi_{AB}(\mathbf{q}, \omega) = \pm \frac{i}{\hbar} \int_0^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{A}(\mathbf{q}, \tau), \hat{B}(-\mathbf{q})]_{\mp} \right) d\tau. \quad (158)$$

This is the virtue of the linear response theory.

### 5.2 Fluctuation-dissipation theorem

The Kubo formula of the linear response function, or the generalized susceptibility, is given in Eq. (158). By extending the interval of integration to  $(-\infty, \infty)$  in Eq. (158), we define a function  $f$  as

$$f_{AB}(\mathbf{q}, \omega) = \pm \frac{i}{\hbar} \int_{-\infty}^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{A}(\mathbf{q}, \tau), \hat{B}(-\mathbf{q})]_{\mp} \right) d\tau. \quad (159)$$

The integration is now divided into the negative and positive ranges as

$$\begin{aligned} f_{AB}(\mathbf{q}, \omega) &= \pm \frac{i}{\hbar} \int_0^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{A}(\mathbf{q}, \tau), \hat{B}(-\mathbf{q})]_{\mp} \right) d\tau \\ &\mp \frac{i}{\hbar} \int_0^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{A}(\mathbf{q}), \hat{B}(-\mathbf{q}, -\tau)]_{\mp} \right) d\tau, \end{aligned} \quad (160)$$

where we changed  $\tau$  to  $-\tau$  in the second term of the right hand side, and moved the factors  $\exp(\pm i\mathcal{K}\tau/\hbar)$  around  $A(\mathbf{q})$  to around  $B(-\mathbf{q})$  using the cyclic property of trace. Since Eq. (158) has a symmetry

$$\chi_{AB}^*(\mathbf{q}, \omega) = \chi_{AB}(-\mathbf{q}, -\omega), \quad (161)$$

it is able to denote  $f_{AB}$  as

$$f_{AB} = \chi_{AB}(\mathbf{q}, \omega) \mp \chi_{BA}^*(\mathbf{q}, \omega). \quad (162)$$

Eq. (159) is written as

$$f_{AB}(\mathbf{q}, \omega) = \left( 1 \mp e^{-\beta\hbar\omega} \right) \frac{i}{\hbar} \int_{-\infty}^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{A}(\mathbf{q}, \tau), \hat{B}(-\mathbf{q})]_{\mp} \right) d\tau, \quad (163)$$

because of the identity

$$\int_{-\infty}^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{A}(\tau), \hat{B}]_{\mp} \right) d\tau = \pm e^{\beta\hbar\omega} \int_{-\infty}^{\infty} e^{i\omega\tau} \text{Tr} \left( \hat{\rho} [\hat{B}, \hat{A}(\tau)]_{\mp} \right) d\tau. \quad (164)$$

The two representations of  $f_{AB}$  in Eqs. (162) and (163) lead us a result

$$\int_{-\infty}^{\infty} e^{i\omega t} \text{Tr} \left( \hat{\rho} [\hat{A}(\mathbf{q}, t), \hat{B}(-\mathbf{q})]_{\mp} \right) dt = \frac{i\hbar}{e^{-\beta\hbar\omega} \mp 1} \{ \chi_{AB}(\mathbf{q}, \omega) \mp \chi_{BA}^*(\mathbf{q}, \omega) \}, \quad (165)$$

where we changed the variable  $\tau$  by  $t$ . Eq. (165) relates the linear response function and the ordinary correlation function. For a special case of  $A = B$ , Eq. (165) is reduced to

$$\int_{-\infty}^{\infty} e^{i\omega t} \text{Tr} \left( \hat{\rho} [\hat{A}(\mathbf{q}, t), \hat{A}(-\mathbf{q})]_{\mp} \right) dt = -\frac{\hbar}{e^{-\beta\hbar\omega} \mp 1} 2\Im \chi_{AA}(\mathbf{q}, \omega). \quad (166)$$

The physical meaning of Eq. (166) is now clear: the correlation function of the thermal fluctuation determines the imaginary part of the linear response function, *i.e.*, the dissipation of the system. Therefore, we call Eqs. (165) and (166) as the *fluctuation-dissipation theorem*.

One can easily find a similarity between the fluctuation-dissipation theorem in Eqs. (165)–(166) and Eq. (134). The spectral representation of the retarded real-time Green's function Eq. (136) is the momentum space Lehmann representation of the two-time retarded Green's function defined by

$$i\bar{G}_{\alpha\beta}^R(\mathbf{x}t, \mathbf{x}'t') = \theta(t - t') \text{Tr} \left( \hat{\rho} \left[ \hat{\psi}_\alpha(\mathbf{x}t), \hat{\psi}_\beta^\dagger(\mathbf{x}'t') \right]_{\mp} \right). \quad (167)$$

The corresponding advanced Green's function is defined by

$$i\bar{G}_{\alpha\beta}^A(\mathbf{x}t, \mathbf{x}'t') = -\theta(t - t') \text{Tr} \left( \hat{\rho} \left[ \hat{\psi}_\alpha(\mathbf{x}t), \hat{\psi}_\beta^\dagger(\mathbf{x}'t') \right]_{\mp} \right). \quad (168)$$

In these definitions of the two-time Green's functions,  $\hat{\psi}$  and  $\hat{\psi}^\dagger$  are respectively the field annihilation and creation operators in the Heisenberg representation

$$\hat{A}(\mathbf{x}t) \equiv e^{i\hat{\mathcal{K}}t/\hbar} \hat{A} e^{-i\hat{\mathcal{K}}t/\hbar}. \quad (169)$$

A two-time Green's function represents the correlation of thermal fluctuation of the process that a particle created at  $\mathbf{x}'t'$  and it annihilated at  $\mathbf{x}t$ . In other words, the two-time Green's functions are nothing more than the correlation functions of two operators  $\hat{\psi}_\alpha$  and  $\hat{\psi}_\beta^\dagger$ . Note that all the quantum operators are composed of the field operators  $\hat{\psi}_\alpha$  and  $\hat{\psi}_\beta^\dagger$ , so that the two-time Green's function can be used to describe all the physical correlations of the system. In the linear response theory, the interaction representation of an operator is not different from the Heisenberg representation in the two-time Green's functions. Therefore, the linear response is nothing more than the retarded two-time Green's function. Note that the *causality* of any physical process stresses that the dissipation  $\chi$  should be the imaginary part of the retarded Green's function.

Now, we have completed the formal theory for computing thermodynamic properties in terms of the quantum field theory. An experimental observation is indeed a measurement of a response of the system by a corresponding external perturbation. The response function is related to the retarded two-time Green's function in terms of the fluctuation-dissipation theory. The two-time Green's function provides the quasiparticle energy spectra. The two-time Green's function can be transformed to a temperature Green's function by introducing the spectral weight function. Once we have the temperature Green's function, we can calculate the thermodynamic information, such as the number of particles  $N$ , the internal energy  $E$ , and the grand potential  $\Omega$ , directly. By performing derivatives on the grand potential, the remaining thermodynamic functions can be evaluated. When there is a phase transition, the response function diverges near the critical region, but the grand potential does not, because the grand potential is a integrated value of the temperature Green's function in the complex  $\omega$  plane.

Kubo (1966) himself demonstrated the applications of the fluctuation-dissipation theorem to compute the density response, conduction and diffusion. The fluctuation-dissipation theorem (Kubo, 1957) is able to explain the irreversible processes related with fluctuations (Callen & Welton, 1951; Casimir, 1945; Nyquist, 1928; Onsager, 1931a,b). Further applications are well summarized by Gammaitoni et al. (1998).

The remained work is mainly on the applications to the realistic matter.

## 6. Inhomogeneity is reality

We have a brief review on the formalism of the quantum field theoretic approach for thermodynamics properties of the system in homogeneous (and isotropic) distributions. The homogeneous approximation of a matter has served as a model system of matter for investigating the nature of the complex phenomena occurring in the matter. The study on electron liquid is an example (Fetter & Walecka, 2003; Giuliani & Vignale, 2005; Kim, 1999; Pines, 1962; 1999). Phonon-mediated superconductivity can be solvable under the method after performing a suitable transformation for treating Cooper pairs (Schrieffer, 1988). A relativistic version of the method has been applied to plasmas (Melrose, 2008).

However, the realistic materials has internal (micro-)structures and even pure elemental materials are not fully homogeneous. Steels as representative structural materials have large variety of microstructures, which influences on the properties, *e.g.*, strength and elongations (Bhadeshia & Honeycombe, 2006). In metals and alloys, the internal structures of materials are governed by the equilibrium thermodynamics and the irreversible processes (Christian, 2002). It is necessary to deal with the inhomogeneity when we are dealing with thermodynamics and irreversible processes for materials sciences.

### 6.1 Density functional theory

The applications of the formal quantum theories, developed in a homogeneous model system, to the realistic inhomogeneous systems have long time history (Fermi, 1927; Slater, 1951; Thomas, 1938). An innovative formalism to treat the inhomogeneous electronic system was suggested by Hohenberg & Kohn (1964). They firstly proved that the ground state energy is a *unique* functional of density,  $E[n(\mathbf{x})]$ . Due to the uniqueness of the ground state energy, the discussions in the previous sections can be reformulated with density. This is the reason why the Hohenberg-Kohn theory is called the density functional theory (DFT). For a slowly vaying density profile, the density functional theory is a generalization of the Thomas-Fermi theory. They proved also that the essential density gradient in the inhomogeneous electron gas can be used for describing the energy functional.

An immediate extension of the Hohenberg-Kohn theory to the finite temperature was done by Mermin (1965). At a given temperature and chemical potential, no two local potential  $v(\mathbf{x})$  lead to the same equilibrium density. In other words, one can define a density functional  $F[n(\mathbf{x})]$  independent of  $v(\mathbf{x})$ , such that the quantity

$$\Omega \equiv \int v(\mathbf{x}) n(\mathbf{x}) d^3\mathbf{x} + F[n(\mathbf{x})], \quad (170)$$

is at a minimum and equal to the grand potential when  $n(\mathbf{x})$  is the equilibrium density in the grand canonical ensemble in the presence of  $v(\mathbf{x})$ . Kohn & Sham (1965) themselves were able to derive the forms for the grand potential, specific heat, and spin susceptibility.

Schneider (1971) provided a general theory of the liquid-solid phase transition by employing the finite temperature density functional theory by virtue of the fluctuation-dissipation theorem on the density response. During the construction of the functional  $F[n]$ , the external perturbation potential is assumed to be of form

$$v^{\text{ext}}(\mathbf{x}) = \frac{1}{2} \sum_{\mathbf{q} \neq 0} \left( v^{\text{ext}}(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{x}} + c.c. \right). \quad (171)$$



The density of the system will contain a linear response to be

$$n(\mathbf{x}) = \bar{n} + \frac{1}{2} \sum_{\mathbf{q} \neq 0} \left( n(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{x}} + c.c. \right), \quad (172)$$

where  $\bar{n}$  is the mean density and  $n(\mathbf{q})$  is the density fluctuations. The Mermin condition (Mermin, 1965) yields

$$\frac{n(\mathbf{q})}{\bar{n}} = \chi(q, [\bar{n}]) v^{\text{ext}}(\mathbf{q}), \quad (173)$$

where the density response function is

$$-\chi(q, [\bar{n}]) = \frac{1}{\beta^{-1} + v^{\text{eff}}(q, [\bar{n}])}, \quad (174)$$

and  $v^{\text{eff}}$  is the corresponding effective potential. For free energy minimum, the density response function should satisfy the condition

$$-\frac{1}{\chi(q, [\bar{n}])} > 0 \quad (175)$$

for the system is in an equilibrium state either absolutely stable or metastable. Therefore,  $-\chi^{-1}(q, [\bar{n}]) = 0$  determines the stability limit of the fluid  $(q_L, P_L, T_L)$ . The static stability condition can be written as

$$-\frac{1}{\chi^{-1}(q, [\bar{n}])} = \frac{1}{\beta S(q)} = \frac{1}{\beta} + v^{\text{eff}}(q, [\bar{n}]) = 0, \quad (176)$$

where  $S(q)$  is the static mean density fluctuation. The condition Eq. (176) implies that the mean density fluctuation  $S(q)$  with the particular wave number  $q_L$  increases as this limit is approached. Near the stability limit only the region  $q \simeq q_L$  is important. We may set the effective potential dependence on  $T$  and  $q$  as

$$\beta v^{\text{eff}}(q) = -1 + A(T - T_L) + B(q - q_L)^2. \quad (177)$$

Then, the density fluctuation for constant  $\bar{n}$  is written as

$$S(q) = \frac{1}{A(T - T_L)} \frac{1/\zeta^2}{1/\zeta^2 + (q - q_L)^2}, \quad (178)$$

where  $\zeta^2 = B/A(T - T_L)$ . As  $q$  approaches to  $q_L$ ,  $S(q)$  increases and the peak becomes sharper. The same effect is also seen in the dynamic mean density fluctuation  $S(q, \omega)$  at the static case  $\omega = 0$ .

This result is consistent with our everyday experiences. In a fluid state, there are random density fluctuations, which can be seen if there is an interface; a random wave distribution on a smooth lake surface is an example. Such density fluctuation is represented in the Fourier terms, by a linear combination of the wave vectors and frequencies. When the temperature is cooled down with a constant pressure, the fluid transforms to crystals, *e.g.*, ice. A crystal is by definition a regular arrangement of density, with a spatial period of a corresponding wave vector. Eq. (178) tells us that the static mean density fluctuation at the critical wave vector  $q_L$  with the corresponding regularity will be sustained, even if there is no external

density perturbation, at the given pressure and temperature. In the mean time, the remained fluctuation modes will be suppressed. This state with the density distribution of a specific spatial regularity, represented by  $q_L$ , is the *crystal*. Hence, the denominator of Eq. (178) explains the critical properties of the solidification.

Once the liquid transformed to a solid, the description of the system state should be switched to a new language for solid states. Every crystalline solid is categorized by its three dimensional translation and rotation symmetries to be the one of the 230 space groups (Bravais, 1845). Bloch (1928) proved, by using the periodic boundary condition (Born & von Kármán, 1912; 1913) and the group representation theory (Wigner, 1927), that the electronic wave function in a crystalline solid should be in the form

$$\psi_{n\mathbf{k}}(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{R}} u_{n\mathbf{k}}(\mathbf{x}), \quad (179)$$

where  $n$  is the band index and  $\mathbf{R}$  is the Bravais lattice vector. The Bloch function has three consequences (Ashcroft & Mermin, 1976): (1) The momentum eigenvalue  $\hbar\mathbf{k}$  of the momentum operator  $\mathbf{p} = (\hbar/i)\nabla$  does not conserve, instead the crystal momentum  $\hbar\mathbf{k} + e^{i\mathbf{k}\cdot\mathbf{x}}\frac{\hbar}{i}\nabla$  conserves. (2) The wave vector  $\mathbf{k}$  is confined to the first Brillouin zone, so that the index  $n$  appear in Eq. (179) represent the multiplicity of the  $\mathbf{k}$  confinements. (3) The nonvanishing velocity of a particle in a level specified by band index  $n$  and wave vector  $\mathbf{k}$  is given by

$$\mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\epsilon_n(\mathbf{k}), \quad (180)$$

where  $\nabla_{\mathbf{k}}$  is a gradient operator with respect to the wave vector  $\mathbf{k}$ . Further consequences can be found in many solid-state physics text books (Altmann, 1995; Ashcroft & Mermin, 1976; Callaway, 1974; Jones & March, 1973a;b; Kittel, 2005; Madelung, 1978).

On the other hand, the numerator,  $1/\zeta^2$  in Eq. (178) explains another aspect of the solidification. The numerator describes the width in the wave vector space of the diverging static mean density fluctuation  $S(q)$ . As  $q \rightarrow q_L$ , the wave vectors in a narrow range around  $q_L$  also contributes to  $S(q)$ . When two waves of near wave vector (or frequency) superposes in the same position, we observe a very well known phenomenon, *beats* (Landau & Lifshitz, 1978). The broadening in  $S(q)$  implies that there are many wave vectors around superposing in the space to exhibits a complex pattern of spatial beats during the crystallization process. In other words, the crystallization will not happen in whole the liquid simultaneously, but the crystal nucleates.

Note that the numerator  $1/\zeta^2$  contains the details of the interaction of the liquid. As we seen in Sec. 5.2, the dissipation of the external perturbation or the system response is determined by the internal fluctuation of the system. The internal fluctuation is determined by the details of the interaction of the system. Hence, we need to have the detailed interaction information of the system also for understanding phase transition phenomena completely.

## 6.2 Kohn-Sham equations

Kohn & Sham (1965) suggested two approximation methods for treating an inhomogeneous system of interacting electrons. Let us describe the first approximation first. The ground-state energy of the system in a static potential  $v(\mathbf{x})$  can be written in the form, in Hartree atomic unit  $\hbar = e^2 = m_e = 1$ ,

$$E = T_0[n] + \int v(\mathbf{x})n(\mathbf{x})d^3\mathbf{x} + \frac{1}{2}\iint\frac{n(\mathbf{x})n(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}d^3\mathbf{x}d^3\mathbf{x}' + E_{xc}[n], \quad (181)$$

where  $T_0$  is the kinetic energy functional of a system of noninteracting electrons and  $E_{xc}[n]$  is the exchange and correlation energy functional of an interacting system with density  $n(\mathbf{x})$ . By assuming

$$E_{xc}[n] = \int n(\mathbf{x}) \epsilon_{xc}(n(\mathbf{x})) d^3\mathbf{x}, \quad (182)$$

the stationary condition of the energy functional Eq. (181) with respect to the density variation  $\delta n$  leads us to have an equation

$$\frac{\delta T_0[n]}{\delta n(\mathbf{x})} + v(\mathbf{x}) + \mu_{xc}(n(\mathbf{x})) = 0, \quad (183)$$

where  $v(\mathbf{x})$  is the result of the functional derivatives of the static integration terms of Eq. (181) and  $\mu_{xc}$  is the result of the functional derivatives of the exchange-correlation functional  $E_{xc}[n]$ . By assuming  $\mu_{xc}(n(\mathbf{x}))$  to be local, we can imagine a “fictitious” particle that moves by the single-particle Schrödinger-like first *Kohn-Sham* equation

$$\left( -\frac{1}{2}\nabla^2 + \varphi(\mathbf{x}) + \mu_{xc}(n(\mathbf{x})) \right) \psi_i^{\text{KS}}(\mathbf{x}) = \epsilon_i^{\text{KS}} \psi_i^{\text{KS}}(\mathbf{x}), \quad (184)$$

and setting

$$n(\mathbf{x}) = \sum_{i=1}^N \left| \psi_i^{\text{KS}}(\mathbf{x}) \right|^2, \quad (185)$$

where  $N$  is the number of electrons. Since we have no knowledge on the density  $n(\mathbf{x})$ , it should be obtained self-consistently starting from a guessed density profile  $n(\mathbf{x})$ . We compute the Kohn-Sham orbitals  $\psi^{\text{KS}}$  under the static potential  $\varphi(\mathbf{x})$  as well as the exchange-correlation potential  $\mu_{xc}(n(\mathbf{x}))$  and then evaluate the density from Eq. (185). Once we obtain the self-consistent density profile  $n(\mathbf{x})$ , the energy of the system is obtained by Eq. (181), if we have an *explicit* functional form of  $E_{xc}[n]$ . Indeed the knowledge of the exchange-correlation functional form is important to obtain the exchange-correlation potential during solving Eq. (184). The cost of computation Eq. (184) is only that of the Hartree equation times the number of self-consistent iterations. The cheap Kohn-Sham equation have get superiority in solving many-body problems as developing computer technology (Argaman & Makov, 2000).

The second approximation by Kohn & Sham (1965) is based on the assumption that

$$E_{xc}[n] = E_x[n] + \int n(\mathbf{x}) \epsilon_c(n(\mathbf{x})) d^3\mathbf{x}. \quad (186)$$

This assumption yields the Hartree-Fock-like second Kohn-Sham equation

$$\left( -\frac{1}{2}\nabla^2 + \varphi(\mathbf{x}) + \mu_c(n(\mathbf{x})) \right) \psi_i^{\text{KS}}(\mathbf{x}) - \int \frac{n_1(\mathbf{x}, \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \psi_i^{\text{KS}}(\mathbf{x}') d^3\mathbf{x}' = \epsilon_i^{\text{KS}} \psi_i^{\text{KS}}(\mathbf{x}), \quad (187)$$

where  $\mu_c(\mathbf{x})$  is the correlation potential and

$$n_1(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^N \psi_j^{\text{KS}}(\mathbf{x}) \psi_j^{*\text{KS}}(\mathbf{x}'). \quad (188)$$

This procedure is similar to the Hartree-Fock method with the local correlation effect, and its cost of computation is comparable to the computational labor intensive Hartree-Fock method due to its nonlocal density matrix operation on the Kohn-Sham orbitals  $\psi_i(\mathbf{x})$  appeared in

Eq. (187). Hence, the application of the Kohn-Sham equation of the type in Eq. (187) has been hesitated. Although the practical success of the first Kohn-Sham equation in Eq. (184), the approximations are remained uncontrollable, because there is no known explicit exchange-correlation functional. Recent developments of computer technology, many approaches has been tried to utilize the second type Kohn-Sham equation (Kümmel & Kronik, 2008). This kind of approaches has solved many problems in the first-type Kohn-Sham equation.

However, there is a critical issue on the solutions of the Kohn-Sham equations. Koopmans (1934) proved that the  $i$ th occupied eigenvalues of the Hartree-Fock equation is identified as the energy required to remove an electron from that orbital without perturbing the rest of the system. This theorem is not valid in the density functional theory, in general. Only the highest occupied eigenvalue of the Kohn-Sham equations has a rigorous physical meaning that the partial derivative is exactly the chemical potential of the system (Almbladh & von Barth, 1985; Levy et al., 1984). In other words, the Kohn-Sham orbitals  $\psi_i^{\text{KS}}(\mathbf{x})$  except the highest occupied orbital are physically meaningless, but they are useful only for building the density in Eq. (185). Indeed, the Kohn-Sham charge density generated by Eq. (185) differs hardly to that of the Hartree-Fock charge density (Stowasser & Hoffmann, 1999). A quantitative linear scaling interpretation of Kohn-Sham eigenvalues with respect to corresponding Hartree-Fock eigenvalues was suggested (Stowasser & Hoffmann, 1999).

### 6.3 Spectral weight function and quasiparticle spectra

As seen in Sec. 6.1, the detailed interaction information of the system is equally important to understand the phase transition as well as the equilibrium properties of the system. In the density functional theoretical framework, the Kohn-Sham equations have a critical problem of the unknown physical meanings of their eigenvalues. The resultant Kohn-Sham orbitals ensure the ground state energy, the density  $n(\mathbf{x})$ , and the chemical potential. The other information obtained by using the Kohn-Sham orbitals are *fictitious*. It is fortunate or unfortunate that the Kohn-Sham orbitals are close to the quasiparticle orbitals, which have the concrete physical meaning through the Koopmans' theorem (Koopmans, 1934). Since there is no further proof on the physical meaning of the Kohn-Sham orbitals, a promising way to build the physically seamless theoretical framework for phase diagrams is starting from the quantum field theory described in Sec. 4, armed with the Kubo's framework in Sec. 5.1–5.2.

The Hamiltonian operator  $\hat{H}$ , the momentum operator  $\hat{\mathbf{P}}$ , and the number operator  $\hat{N}$  commute each other. By symmetry, those operators have a set of eigenstates  $\{|m\rangle\}$  simultaneously (Tinkam, 1964), such that

$$\hat{\mathbf{P}} |m\rangle = \mathbf{P}_m |m\rangle, \quad (189)$$

$$(\hat{H} - \mu\hat{N}) |m\rangle = (E_m - \mu N_m) |m\rangle. \quad (190)$$

Note that Eq. (190) is understood as

$$\hat{\mathcal{K}} |m\rangle = K_m |m\rangle. \quad (191)$$

With the aid of the Sokhotsky-Weierstrass theorem (Byron & Fuller, 1992),

$$\frac{1}{\omega \pm i\eta} = \wp \frac{1}{\omega} \mp i\pi\delta(\omega), \quad (192)$$

we will have a Lehmann representation of the real-time Green's function for real  $\omega$  such as

$$\begin{aligned} \bar{G}(\mathbf{k}, \omega) = & \frac{e^{\beta\Omega}}{2s+1} \sum_{m,n} e^{-\beta K_m} (2\pi)^3 \delta \left[ \mathbf{k} - \hbar^{-1} (\mathbf{P}_n - \mathbf{P}_m) \right] |\langle m | \hat{\psi}_\alpha | n \rangle|^2 \\ & \times \left\{ \wp \frac{1}{\omega - \hbar^{-1} (K_n - K_m)} \left( 1 \mp e^{-\beta(K_n - K_m)} \right) \right. \\ & \left. - i\pi\delta \left[ \omega - \hbar^{-1} (K_n - K_m) \right] \left( 1 \pm e^{-\beta(K_n - K_m)} \right) \right\}, \end{aligned} \quad (193)$$

where  $s$  is the spin degeneracy and  $\wp$  denotes a principal value. The imaginary part of Eq. (193) is written as

$$\begin{aligned} \Im \bar{G}(\mathbf{k}, \omega) = & -\frac{\pi e^{\beta\Omega}}{2s+1} \sum_{m,n} e^{-\beta K_m} (2\pi)^3 \delta \left[ \mathbf{k} - \hbar^{-1} (\mathbf{P}_n - \mathbf{P}_m) \right] |\langle m | \hat{\psi}_\alpha | n \rangle|^2 \\ & \times \delta \left[ \omega - \hbar^{-1} (K_n - K_m) \right] \left( 1 \pm e^{-\beta(K_n - K_m)} \right). \end{aligned} \quad (194)$$

Landau (1958) proved a relation

$$\Re \bar{G}(\mathbf{k}, \omega) = -\wp \int_{-\infty}^{\infty} \frac{d\omega'}{\pi} \frac{\Im \bar{G}(\mathbf{k}, \omega')}{\omega - \omega'} \frac{1 \mp e^{-\beta\omega'}}{1 \pm e^{-\beta\omega'}}. \quad (195)$$

Considering the spectral representation form in Eq. (135) we have the detailed spectral weight function of form

$$\begin{aligned} A(\mathbf{k}, \omega) = & \frac{e^{\beta\Omega}}{2s+1} \sum_{m,n} e^{-\beta K_m} (2\pi)^3 \delta \left[ \mathbf{k} - \hbar^{-1} (\mathbf{P}_n - \mathbf{P}_m) \right] \\ & \times 2\pi\delta \left[ \omega - \hbar^{-1} (K_n - K_m) \right] \left( 1 \mp e^{-\beta(K_n - K_m)} \right) |\langle m | \hat{\psi}_\alpha | n \rangle|^2. \end{aligned} \quad (196)$$

Since the spectral weight function  $A(\mathbf{k}, \omega)$  relates the real-time Green's function and the temperature Green's function, we can, in principle, solve all the thermodynamic problems with detailed interaction information of the system: The thermal Green's function describes the equation of states, while the imaginary part of the retarded Green's function from the real-time Green's function yields the fluctuations to describe the phase transition, including the transport properties. The remained difficulty is to calculate the eigenvalues  $K_m$  and  $\mathbf{P}_m$  in Eqs. (189)–(190) in the case of interacting systems.

Is it able to calculate the eigenvalues of the interacting system at a finite temperature?

Let us come back to the idea of Born & Oppenheimer (1927) adiabatic separation of the motions of electrons and ions. In the Born-Oppenheimer scheme, we first calculate the motions of electrons in a given ionic background and then calculate the motions of ions at the given electronic one. It is well known that the given symmetry and many-body effects will emerge as elementary excitations (Pines, 1999), which can be extended even to a *macroscopic* levels (Anderson, 1997). Some of the elementary excitations may originate from either electrons or ions. For example, plasmons are the elementary excitations of collective motions of electrons, while phonons are the elementary excitation of the collective motions of ions. One can also interesting to more complex excitations, which are composites of those elementary excitations. Anyhow, all such properties are computed from the electronic structures of the system.

It is instructive to consider the formal form in Eq. (129) of the Dyson's equation. Since the Dyson's equation expresses the difference between the noninteracting and interacting temperature Green's functions in terms of the self-energy,  $\Sigma(\mathbf{k}, \omega_n)$ ; all the many-body effects are contained in the self-energy. Considering the real space Dyson's equation form in Eq. (107), we may have a Hartree-Fock-like self-consistent equation (Fetter & Walecka, 2003)

$$\left[ -\frac{\hbar}{2m} \nabla_1^2 + v(\mathbf{x}_1) \right] \psi_j(\mathbf{x}_1) + \hbar \int d^3\mathbf{x}_2 \Sigma(\mathbf{x}_1, \mathbf{x}_2) \psi_j(\mathbf{x}_2) = \epsilon_j \psi_j(\mathbf{x}_1), \quad (197)$$

where  $v(\mathbf{x}_1)$  is the static one-body potential to describe the *inhomogeneity* as in the density functional theory,  $\nabla_1^2$  is the Laplacian operator with respect to the coordinate  $\mathbf{x}_1$ , and  $\mu$  is the chemical potential obtained by

$$N(T, V, \mu) = (2s + 1) \sum_j n_j, \quad (198)$$

with the number distribution function

$$n_j = \frac{1}{e^{\beta(\epsilon_j - \mu)} \mp 1}. \quad (199)$$

The chemical potential  $\mu(T, V, N)$  is, in principle, able to be inverted from Eq. (198) if  $N$  is fixed. In this derivation, we assumed that all the necessary frequency summed are already done for making  $\Sigma$  being depending explicitly on  $T$  and  $\mu$ , but being independent to the frequency  $\omega_n$  of the Hartree-Fock form and the temperature Green's function

$$\mathcal{G}(\mathbf{x}_1, \mathbf{x}_2, \omega_n) = \sum_j \frac{\psi_j(\mathbf{x}_1) \psi_j^*(\mathbf{x}_2)}{i\omega_n - \hbar^{-1}(\epsilon_j - \mu)}, \quad (200)$$

where  $\psi_j$  is a single-particle wave function with energy  $\epsilon_j$  for forming a complete orthonormal set, forms the same form of the noninteracting temperature Green's function  $\mathcal{G}^0(\mathbf{x}_1, \mathbf{x}_2, \omega_n)$  with  $\psi_j^0$ , the solution of a differential equation

$$\left[ i\hbar\omega_n + \frac{\hbar}{2m} \nabla_1^2 + \mu - v(\mathbf{x}_1) \right] \psi_j^0(\mathbf{x}_1) = \epsilon_j^0 \psi_j^0(\mathbf{x}_1). \quad (201)$$

In this self-consistent formalism, the quasiparticle energy spectrum becomes

$$\epsilon_{\mathbf{k}} = \epsilon_{\mathbf{k}}^0 + \hbar \Sigma(\mathbf{k}), \quad (202)$$

where we recover the momentum space representation.

This equation tells us that the noninteracting quasiparticle moving with the energy spectrum  $\epsilon_{\mathbf{k}}^0$  is *dressed* by the self-energy  $\Sigma(\mathbf{k})$  by the sophisticated many-body interactions. The energy dispersion of a free particle of mass  $m^*$  will have a dispersion

$$\nabla_{\mathbf{k}} \epsilon_{\mathbf{k}} = \frac{\mathbf{k}}{m^*}.$$

Equating this relation to result of the operation of the gradient operator on Eq. (202) yields

$$\frac{m}{m^*} = 1 + \frac{\hbar \nabla_{\mathbf{k}} \Sigma(\mathbf{k})}{\epsilon_{\mathbf{k}}^0}. \quad (203)$$

This means that the motion of the quasiparticle of mass  $m$  may be thought as that of a free particle of mass  $m^*$ , of which all the many-body effects are corrected. Now we can call  $m^*$  as the *effective mass* of the quasiparticle. The noninteracting quasiparticle Green's function may be in the form

$$\bar{G}(\mathbf{k}, \omega) = \frac{Z}{\omega - \hbar^{-1}(\epsilon_{\mathbf{k}} - \mu)}, \quad (204)$$

where  $Z$  is called the *renormalization factor*, which contains the effective mass effect in Eq. (203). Considering the spectral function relation with the Green's function, we have

$$A(\mathbf{k}, \omega) = Z\delta(\omega + \mu - \hbar\mathbf{k}). \quad (205)$$

In other words, the results of a photoemission spectroscopy are just the renormalization factor of the system. The photoemission experiments gives the quasiparticle spectra.

The idea of renormalization was utilized by Landau (1957a;b; 1959), known as the phenomenological *Landau Fermi-liquid theory*. Landau classified the quasiparticles as the levels of a Fermi liquid corresponding to the classification of the levels of noninteracting particles. The interaction is slowly *turned on* to the particles gradually become quasiparticles. The methods described above can be directly applied for the quasiparticles. Using this concept, Landau (1957a;b; 1959) was able to calculate the macroscopic quantities, such as specific heat, entropy, magnetic susceptibility, kinetic motions, thermal conductivity, viscosity, dispersion and absorption of sound, and fluctuations in the distribution function. The relation with the quantum field theory of the Landau Fermi liquid theory was investigated by Nozières & Luttinger (1962) and Luttinger & Nozières (1962), but it is not clear yet. Good introductory reviews on the Ladau Fermi-liquid theory are given by Abrikosov & Khalatnikov (1959); Giuliani & Vignale (2005); v. Löhneysen et al. (2006).

Bohm & Pines (1952) suggested a different point of view with a canonical transformation to treat a particle and its consequent collective excitations simultaneously. In this formalism, the canonical transformation stresses the elementary ingredients of the Hamiltonian are the particle-hole pairs, instead of individual particles. Bohm & Pines (1952) approximated the coupling between the excitations corresponding to different momentum transfers  $q$  and  $q'$ . This ignorance is originated from an intuition that the momentum transfer phase  $q - q'$  will be random to be enough for cancelling themselves. Hence, the method is known as the *random phase approximation (RPA)*. The RPA was applied to the electronic system with plasmons and refined (Sawada, 1957; Sawada et al., 1957). The RPA picture as the similarity with that of (Goldstone, 1957) and it deals with the correlation function corresponding to the collective excitations.

These features of the RPA have been utilized in susceptibility investigations. For instance, Kim et al. (1973) calculated a spin polarized susceptibility<sup>9</sup> based on the fluctuation-dissipation theorem and the free energy by considering higher order self-energy correction (Kim, 1974) within the RPA. Kim (1982; 1989; 1999) used the RPA technique to demonstrate, within the homogeneous jellium model, that all the thermodynamic properties, including the ferromagnetic phase transition, of the metallic ferromagnetism can be obtained by considering the spin-polarized electrons with the collective excitations of phonons and

<sup>9</sup> A practical self-consistent version was given by Lee et al. (1989).

magnons. All those quasiparticles should interact each other, in contrast to the typical considerations that quasiparticles hardly interacts each other. Lee et al. (2006) confirmed that there are interactions between phonon and magnetism, and Bodraykov (2007) in his phenomenological model reproduced the invar effect of the temperature dependent thermal expansion coefficient by considering the results of the interacting phonon-magnon (Kim, 1982).

## 7. Defects

Although there is criticism on the validaties of the Landau Fermi-liquid concept (v. Löhneysen et al., 2006), a new insight for dealing macroscopic physical observables, *e.g.*, even domain (grain) boundaries and dislocations, combined with the concept of symmetry breaking (Anderson, 1997). Since the semi-phenomenological construction of the Ladau Fermi-liquid theory, it is possible to switch our intersts from the quasiparticle space to the order parameter space. This transformation to the order parameter space is familiar with that of Landau theory of phase transition. Toulouse & Kléman (1976) suggested an innovative idea that the defects in ordered media can be classified in terms of the *topological defects* of the corresponding order parameter. This topological concept was immediately applied to the vortex textures in superfluid  $^3\text{He}$  for showing that the textures with vorticity without vortex core can be constructed (Anderson & Toulouse, 1977). One of the most striking application of the topological defects theory may be the explanation of the entanglement free dislocation crossings (Poénaru & Toulouse, 1977); dislocations can be entangled if and only if they are members of separate class of  $\pi_1(R)$ . We will see the meaning of  $\pi_1(R)$  later. A pedagogical review on the topological defects was given by Mermin (1979). Recently, Kleman & Friedel (2008) reviewed the practical interpretations of the homotopy group theory of the topological disclination, dislocations, and continuous defects.

### 7.1 Topological defects

In a periodic solid, the elastic variable is the vector displacement field  $\mathbf{u}(\mathbf{x})$ . Changes of  $\mathbf{u}$  by a Bravais lattice vector  $\mathbf{R}$  leave the lattice unchanged. Thus the order parameter space for the solid is the three-dimensional space of displacements  $\mathbf{u}$  with lattice points. A dislocation with a core along  $\mathbf{l}$  is characterized by<sup>10</sup>

$$\oint_C d\mathbf{u} = \oint_C \frac{d\mathbf{u}}{ds} ds = \mathbf{R} \equiv \mathbf{b}, \quad (206)$$

where  $C$  is a curve enclosing the core and  $\mathbf{b}$  is called the *Burgers vector*. The construction of the Burgers vector is fully topological, but there are two nontopological descriptions; screw dislocations if  $\mathbf{b} \parallel \mathbf{l}$  and edge dislocations if  $\mathbf{b} \perp \mathbf{l}$ . Since the Bravais lattice vector, which leaves after a single winding curve  $C$  passed to enclose the core, is discrete, the dislocation is, from homotopy group theory, *line*.

Let us look at the characters of dislocations with little rigor (Mermin, 1979). Suppose that  $R$  is an order-parameter space. There is a map  $f(z)$  of interval  $0 \leq z \leq 1$  into  $R$  defining the real-space loop or circles into order-parameter space to give closed curves in order-parameter space itself. A continuous directed curve described by maps  $f$  passing through a base point  $x$  in  $R$ . The close curve is decribed by  $f(0) = f(1) = x$ . Two loops  $f$  and  $g$  are *homotopic at  $x$* , if there is a continous family of loops, all passing through  $x$ . The classes of homotopic loops at  $x$  form a group, called  $\pi_1(R, x)$ , and is known as the *fundamental group* of  $R$  at  $x$ . There are other groups associated with the mappings of the surface of an  $n$  sphere  $S_n$  in Euclidean  $n + 1$  space

<sup>10</sup> Here, we follow the definition by Chaikin & Lubensky (1995).



called  $\pi_n(R, x)$ . The fundamental group is called the *first homotopy group* and  $\pi_n$  is then called the *n*th homotopy group. For example,  $\pi_2(S_2, x)$  is called when the maps  $f$  at the base point  $x$  is on the sphere  $S$ , *i.e.*, say,  $u$  and  $v$  into  $R$ , of three dimensional order-parameter space. For an integer  $Z$ , if  $\pi_1(S_1) = Z$ , the group structure does not depend on the choice of base point  $x$  and the map is topologically the same as the circle. Any mapping of a loop into a sphere can be shrunk to a point, so that the fundamental group of the surface of a sphere consists only of the identity,  $\pi_1(S_1) = 0$ . Trivial point defects are associated with mappings that can be deformed to the constant map.

Let us say the group  $G$  containing translations as well as proper rotations, *i.e.* the proper part of the full Euclidean group. Let  $H$  be the subgroup of  $G$  containing those rigid body rotations that leave the reference system invariant. Naively the order-parameter space  $R$  is the coset space  $G/H$ . If we ignore of the rotational symmetries of crystals, the full proper Euclidean group can be replaced by the subgroup  $T(3)$  of translations. On the other hand the isotropy subgroup  $H$  becomes the subgroup of  $T(3)$  consisting of translations through Bravais lattice vectors. Since  $T(3)$  is parameterized by all of Euclidean three-space, it is connected and simply connected. Since  $H$  is discrete, we identifies  $\pi_1(G/H)$  with  $H$  itself. Thus the line defects, in our case dislocations, are characterized by Bravais lattice vectors. The vector is usually known as the Burgers vectors. Because  $H$  is discrete  $\pi_2(G/H) = 0$ . In this case, the homotopy theory says notopologically stable *point* defects. However, we know thermodynamically stable point defects, *e.g.*, vacancies or voids. The physical means of removing a void are reminiscent of the nonlocal means available for the elimination of topologically stable defects.

Now let us recover the operations of the rigid body rotations. The fundamental group is identified with the double group. The three dimensional rotational parts of the group operations are lifted from  $SO(3)$  rotation to  $SU(2)$ . Discrete operations in the space group with no translational part correspond to line defects, in which the local crystal structure rotates through an angle associated with a point group operation as the line is encircled. Such defects are known as disclinations. Although the  $\pi_3(R)$  has the classification problem (Mermin, 1979), the domain walls (grain boundaries) are thought as the string bounded topological defects by the broken discrete symmetry (Vilenkin, 1985), hence so that textures evolve.

## 7.2 Renormalization group and gauge theory

Once the order parameter class is identified, the partition functions are divided by the corresponding topologically invariant number. Indeed, Burakovski et al. (2000) rederived the free energy density as a function of dislocation density depending on temperature, starting from the partition function computation for the indistinguishable dislocation loops. The resultant free energy density function is very similar to that out of the constitutional approach (Cotterill, 1980). The free-energy is in the form of the two-dimensional  $xy$  model, which should undergoes a two-dimensional Kosterlitz-Thouless phase transtion (Kosterlitz & Thouless, 1973). The theory break down point, *i.e.*, the melting temperature is also derived. They also showed that the dislocation density  $\rho$  cannot be a continuous function of temperature, but it is a gap function

$$\rho(T) = \begin{cases} 0, & T < T_m, \\ \rho(T_m) \sim b^{-2}, & T = T_m, \end{cases} \quad (207)$$

where  $T_m$  is the melting temperature. The dislocation-mediated melting mechanism was observed in a bulk colloidal crystals (Alsayed et al., 2005). A good review on two-dimensional melting phenomena is given by Strandburg (1986; 1989).

Since the inclusion of the symmetry classification, the *renormalization group* approach is appropriate to order parameters, or phase fields. The renormalization group theory has been served as the most appropriate tool for investigating phase transitions and critical phenomena (Anderson, 1997; Fradkin, 1991; Goldenfeld, 1992; Zinn-Justin, 1997). The idea of the renormalization group approach is a successive scaling transformation of the system interaction in real space, known as coarse-graining, or cutoff the high momentum fluctuations larger than arbitrary value  $\Lambda$  to eliminate the ultraviolet catastrophe of the perturbation expansion (Wilson, 1975; 1983). During the successive scaling, the interaction Hamiltonian of the phase field should keep its symmetry. The resultant Hamiltonian is arrived at a *fixed point* independent on the scale of the system, so it gives the *universal* behavior of the critical phenomena. The singularity in the order parameter equation gives the dissipation of the order parameter. The good examples of the RG procedure with the topological defects were given by Fradkin (1991).

Elegance is in the renormalization group theory, because it does not depend on the system scale but on the symmetry. For example, the dislocation-mediated melting transition is categorized as the problems of the vortex formation in superfluid  $^3\text{He}$  (Anderson & Toulouse, 1977). In this sense, transition phenomena in the world can be categorized by *universality class* defined by the critical exponents.

However, is it enough?

Teichler (1981) criticised the above approach, because it is unsolved how we see the deviation of the equations of motion of elementary excitations in an ordinary medium with topological defects to the corresponding equations in the ideal configuration. He also showed that the transformed single electron Hamiltonian for describing the distorted crystal by dislocation (Brown, 1845) is described in terms of *gauge fields* in addition to the well-known deformation potentials. Gauge means the redundant degrees of freedom in Hamiltonian or Lagrangian, but its transformation behavior characterizes the interaction. The quantization of gauge field is gauge boson; *e.g.*, photons are a kind of gauge bosons out of the gauge degree of freedom of electromagnetic fields.

The quantum motions of electrons in a gauge field due to the topological dislocation defects have been investigated (Bausch, Schmitz & Truski, 1999; Bausch, Schmitz & Turski, 1999; Turski & Mińkowski, 2009). The main conclusions of the gauge field theory (Bausch et al., 1998) are that the motion of a quantum particle is described in a curved-twisted space, as in gravitational field, called Riemann-Cartan manifold and that the Hamiltonian is separated in two parts that the covariant part and noncovariant parts (Larzar, 2010). The gauge field method has been applied to the elastic properties of solids (Dietel & Kleinert, 2006; Kleinert & Jiang, 2003; Turski et al., 2007) by modelling the elastic energy of the lattices by harmonic displacement fields. The computed melting temperatures of selected bcc and fcc elemental crystals agrees acceptably with the experimental values (Kleinert & Jiang, 2003). The most striking result of the defect gauge theory is the Aharonov & Bohm (1949) type interferences near the dislocation. This fact indicates that the dislocations are not be classical objects anymore, but are quantum mechanical ones.

## 8. Outlook and proposal

We have arrived at the stage to consider the applications of the theoretical framework in materials science.

Given a problem, we have to identify the order parameters governed by a certain set of symmetries. Based on the symmetry we need to classify the possible topological defects. The problem is separated into two parts: the scale independent universal properties and the elementary excitations. The tools for the universal properties are the homotopy theory and the

renormalization group, while those for elementary excitations are conventional perturbative renormalization technique with the finite temperature Green's function. The effects of the identified topological defects on the elementary excitations can be dealt with the gauge field theory. The phase transition mechanism of each object can be understood by utilizing the fluctuation-dissipation theorem. Note that all the methods are based on the computation of partition functions by quantum field theoretic manner.

A reversed approach, seems more practical, is also possible. One can firstly identify the ideal structure and compute the equilibrium thermodynamics using the Green's function technique. The fundamental irreversible processes are also available from this technique. When one is faced a problem related with (topological) defects, the renormalization group and the defect gauge theory should be considered.

They usually raise a question on the practicality of the quantum field theoretic approach. Theories introduced here are nothing new, but tried to describe them in a unified way. Those theories have been developed and published from individual schools. This fact implies that the methods can be applied to materials engineering with a suitable software package and computers. There are the other approach for materials design. For example, Olson (2000) viewed a material as a system rather than a set of matters. Once an engineering target is given, one has to identifies the required properties in terms of materials properties and the corresponding structure with a suitable set of processes. The basic assumption is that the necessary theoretical and experimental tools are already prepared in a database and identified for the purpose. The length and time scales are covered from electronic scale to devices. A well organized spin-off company indeed does this business. The practicality awareness can be solved if we have a unified theoretical tool. This approach is an *ad hoc* combination of the current methods. The quantum field theoretic approach proposed in this article removes such *ad hoc* awareness.

In view point of computer resources, the required computer performance of Olson's multiscase approach may be of order from peta-floating point operations per second (peta-FLOPS) to exa-FLOPS. The individual sectors of the theories introduced here utilize from a desktop personal computer to one of order tera-FLOPS. If there is a computer software package in a unified theoretical framework, the required computer performance may sit on the performance of order peta-FLOPS. Hence, the practicality will not be a major problem.

In order to utilize a theoretical framework in engineering, a unified software package development is necessary. Unfortunately, such unified theoretical scheme has not been established yet. The reason can be thought in two-folds. Firstly, the theoretical maturity may not been achieved satisfactory by means of both theory itself and its implementation into a computer software package. When the author (hereafter referred as I) experienced first density functional calculation in 1994, I could remember the names of most researchers working in the density functional world. Now the number of citation of Kohn & Sham (1965) paper is approaching to 20,000 at the time of writing this manuscript. Most of them are using the software package as a black-box, due to the maturity of the software packages and the developments of computer technology. This example demonstrates the power of the matured theory and the corresponding computer softwares.

Secondly, the quantum field theory itself is too difficult to understand for beginners. Cardy (2000) explained, at the preface of his book, this situation as

... and the the student, if he or she is lucky, has just about time to learn how to calculate the critical exponents of the Ising model in  $4 - \epsilon$  dimensions before the course comes to an end.

The Ising model is considered as one of the simplest solvable "toy" model. In addition, many researchers in quantum field theory are working with their own "toys." The maturity of such

toys seems enough to evolve to as tools. Covering many topics in the renormalization group and quantum field theory is not so practical to learn in depth for typical material scientists. Such study will be done by the experts. The pedagogy for understanding the quantum field theory is also equally important.

To date there is no further fundamental theoretical tool in describing our physical world rather than the quantum field theory. It is hardly necessary to go beyond the quantum field theory for materials scientists. I would like to emphasize that we are standing at the starting point for a long journey to utilize the unified method with the reinterpreted terminology, *ab initio* thermodynamics authentically.

## 9. Acknowledgements

The author deeply appreciates Profs. Jae Il Lee and Se-Kyeun Kwon for their decades-long discussions and critics on this subject. A special acknowledgement is dedicated to Prof. Byeong-Joo Lee for his mentoring advises on the CALPHAD methods. Prof. H. K. D. H. Bhadeshia should receive thanks of the author for his comments from a metallurgist's view point.

This work was supported in part by POSCO Steel Innovation Program and in part by the Basic Science Research Program (Grant No. 2009-0088216) and the World Class University (WCU) Program (Project No. R32-2008-000-10147-0) through the National Research Foundation funded by the Ministry of Education, Science and Technology of Republic of Korea.

This article is dedicated for the memory of Late Prof. Duk-Joo Kim.

## List of Symbols

- $A$ : spectral weight function
- $C(R)$ : correlation function of interaction range  $R$
- $C_V$ : volume constant heat capacity
- $D$ : long time interval
- $\mathcal{D}$ : density of states
- $E$ : internal energy
- $E_{xc}[n]$ : exchange-correlation energy functional of density  $n$
- $F$ : Helmholtz free energy
- $F(t)$ : time dependent driving force
- $F[n(\mathbf{x})]$ : energy functional of density  $n(\mathbf{x})$
- $G$ : Gibbs free energy; full group in the order-parameter space  $R$  (depending on context)
- $G/H$ : coset space
- $\bar{G}$ : real-time Green's function
- $\bar{G}^A$ : advanced Green's function
- $\bar{G}^R$ : retarded Green's function
- $\mathcal{G}$ : temperature Green's function
- $H$ : subgroup of  $G$
- $\hat{H}$ : Hamiltonian operator
- $\hat{H}'$ : external perturbation Hamiltonian operator

- $\mathbf{H}$ : external magnetic field
- $\Im$ : imaginary part extraction operator
- $J_{ij}$ : interaction of spins between  $i$ th and  $j$ th sites
- $K$ : propagator, Green's function
- $\hat{\mathcal{K}}$ : grand canonical Hamiltonian operator
- $\mathcal{L}$ : Landau function
- $\mathbf{M}$ : total magnetization
- $\hat{M}_\mu$ : magnetization in  $\mu$ -direction
- $N$ : number of particles of the system
- $N_i$ : number of particles of  $i$ th component
- $P$ : pressure
- $\wp$ : Cauchy principal value
- $\hat{\mathbf{P}}$ : momentum operator
- $\mathbf{P}_m$ : eigenvalue of the momentum operator
- $R$ : interaction range; order-parameter space (depending on context)
- $\mathbf{R}$ : Bravais lattice vector
- $\Re$ : real part extraction operator
- $S$ : entropy
- $S(q)$ : mean density fluctuation as function of wavevector  $q$
- $\mathbf{S}_i$ : localized atomic spin at the site  $i$
- $SO(3)$ : special orthogonal rotation group at three dimension
- $SU(2)$ : special unitary rotation group at two dimension
- $T$ : temperature
- $T_0$ : Bose-Einstein condensation temperature
- $T_0[n]$ : kinetic energy functional of the noninteracting system of density  $n$
- $T_C$ : Curie temperature
- $T_L$ : liquifying temperature
- $T_m$ : melting temperature
- $T(3)$ : Euler translational subgroup
- $\hat{U}$ : unary external potential operator
- $V$ : volume of the system
- $Z$ : partition function; renormalization factor; an integer (depending on context)
- $a$ : lattice constant
- $a_n(H, T)$ : expansion coefficient of the Landau function as function of  $H$  and  $T$
- $b$ : norm of Burgers vector
- $\mathbf{b}$ : Burgers vector
- $b_n, c_n, d_n$ : expansion coefficients of the Landau function

- *c.c.*: complex conjugate
- *d*: dimensionality of the interaction
- $d_1$ : differential operator for  $x_1, t_1$
- $dw$ : probability of states represented in the phase volume
- $f_{AB}$ : generalized response function
- $f(p, q)$ : any physical quantity as a function of  $p$  and  $q$
- $\bar{f}$ : average of  $f(p, q)$
- $f(z)$ : order parameter map
- $g$ : spin degeneracy factor
- $h$ : reduced external magnetic field ( $H - H_c$ )
- $\hbar$ : Planck's constant
- $i$ : index for the microstates
- $\mathbf{k}$ : wavevector
- $k_B$ : Boltzmann's constant
- $k_F$ : norm of the Fermi wavevector
- $k(t)$ : solution of the equation of motion
- $\mathbf{l}$ : dislocation core axis vector
- $m$ : mass of the particle
- $m^*$ : effective mass
- $m_s$ : magnetization of the Landau function at the global minimum
- $\mathbf{m}$ : magnetization per spin
- $n$ : band index
- $n(\mathbf{x})$ : number density as function of  $\mathbf{x}$
- $n_i^0$ : occupation number of the state  $i$  of the noninteracting system
- $p$ : norm of the momentum  $\mathbf{p}$
- $\mathbf{p}$ : single-particle momentum
- $p_i$ : generalized momentum of the  $i$ th degree of freedom
- $q$ : norm of wavevector  $\mathbf{q}$
- $q_L$ : liquifying wavevector
- $\mathbf{q}$ : wavevector
- $q_i$ : generalized coordinate of the  $i$ th degree of freedom
- $s$ : degree of freedom of the system; number of spins (depending on context)
- $t$ : time; reduced temperature  $T - T_C$  (depending on context)
- $t^*$ : critical temperature
- $\mathbf{u}(\mathbf{x})$ : vector displacement field at position  $\mathbf{x}$
- $u_{n\mathbf{k}}$ : Bloch function of band  $n$  and wavevector  $\mathbf{k}$  at position  $\mathbf{x}$
- $v$ : static potential

- $v^{\text{ext}}$ : static external potential
- $\mathbf{v}_n(\mathbf{k})$ : nonvanishing velocity of a Bloch wavefunction of the band  $n$  and the wavevector  $\mathbf{k}$
- $x$ : base point of order parameter space  $R$
- $\mathbf{x}$ : position
- $\Delta p \Delta q$ : small phase interval
- $z$ : coordinate of the order parameter space  $R$
- $\Gamma$ : gamma function
- $\bar{\Gamma}$ : generic real-time Green's function in Lehmann representation
- $\Pi$ : generalized polarization
- $\Sigma$ : self-energy
- $\Sigma^\lambda$ : self-energy for the interaction strength of  $\lambda$
- $\Theta_D$ : Debye temperature
- $|\Phi_0\rangle$ : ground state of the system
- $\Xi$ : grand partition function
- $\Omega$ : grand (thermodynamic) potential
- $\Omega_0$ : grand potential of the noninteracting system
- $\beta$ : inverse temperature  $1/k_B T$
- $\gamma_{\mathbf{k}}$ : damping of the quasiparticle of wavevector  $\mathbf{k}$
- $\delta$ : Dirac delta function
- $\epsilon$ : single-particle energy
- $\epsilon_F$ : Fermi energy
- $\epsilon_{\mathbf{k}}$ : energy spectrum of the particle at the wavevector  $\mathbf{k}$
- $\epsilon_{\mathbf{p}}$ : energy spectrum of the particle at the momentum  $\mathbf{p}$
- $\epsilon_{\mathbf{k}}^0$ : energy spectrum of the noninteracting single-particle at wavevector  $\mathbf{k}$
- $\epsilon_c$ : correlation energy density
- $\epsilon_{xc}$ : exchange-correlation energy density
- $\epsilon_{ij}$ : fractional error of the Landau theory as function of the  $i$ th and  $j$ th sites
- $\epsilon_R$ : fractional error of the Landau theory as function of  $R$
- $\eta$ : infinitesimal real positive convergence factor
- $\kappa$ : generalized dielectric function
- $\lambda$ : interaction control parameter
- $\mu$ : chemical potential
- $\mu_i$ : chemical potential of  $i$ th component
- $\mu_c$ : correlation energy potential
- $\mu_{xc}$ : exchange-correlation potential
- $\psi$ : wavefunction

- $\hat{\psi}_\alpha$ : field annihilation operator at the state  $\alpha$
- $\hat{\psi}_\alpha^\dagger$ : field creation operator at the state  $\alpha$
- $\psi_i^{\text{KS}}$ :  $i$ th Kohn-Sham orbital
- $\phi_n$ :  $n$ th eigenfunction of  $\hat{H}$
- $\rho$ : statistical distribution function
- $\hat{\rho}$ : statistical distribution function operator of unperturbed system
- $\hat{\rho}'$ : statistical distribution function operator by perturbation
- $\theta$ : step function
- $\chi_T(H)$ : isothermal magnetic susceptibility as function of  $H$
- $\chi_{\mu\nu}$ :  $\mu$ -direction magnetic susceptibility with respect to  $\nu$ -direction external field
- $\chi^*$ : final state wavefunction
- $\sigma$ : Stefan-Boltzmann constant
- $\tau$ : imaginary time; time difference (depending on context)
- $\pi_n(R, x)$ :  $n$ th homotopy group of  $R$  at base point  $x$
- $\zeta$ : correlation length
- $\zeta$ : index for microstates of the system; zeta function (depending on context)
- $\omega$ : frequency of a single-particle
- $\omega_n$ : Matsubara frequency
- $[\hat{A}, \hat{B}]$ : Poisson bracket of operators  $\hat{A}$  and  $\hat{B}$
- $4 - \epsilon$ : near four dimension

## 10. References

- Abrikosov, A. A. & Khalatnikov, I. M. (1959). The theory of a Fermi liquid (the properties of liquid  $^3\text{He}$  at low temperatures), *Rep. Prog. Phys.* 22(1): 329–367.
- Aharonov, Y. & Bohm, D. (1949). Significance of electromagnetic potentials in the quantum theory, *Phys. Rev.* 115(3): 485–491.
- Almbladh, C. O. & von Barth, U. (1985). Exact results for the charge and spin densities, exchange-correlation potentials, and density-functional eigenvalues, *Phys. Rev. B* 31(6): 3231–3244.
- Alsayed, A. M., Islam, M. F., Zhang, J., Collins, P. J. & Yodh, A. G. (2005). Premelting at defects within bulk colloidal crystals, *Science* 309(5738): 1207–1210.
- Altmann, S. L. (1995). *Band Theory of Solids: An Introduction from the Point of View of Symmetry*, Oxford University Press, Oxford.
- Anderson, P. A. & Toulouse, G. (1977). Phase slippage without vortex cores: Vortex textures in superfluid  $^3\text{He}$ , *Phys. Rev. Lett.* 38(9): 508–511.
- Anderson, P. W. (1963). Plasmons, gauge invariance, and mass, *Phys. Rev.* 130(1): 439–442.
- Anderson, P. W. (1997). *Basic Notions of Condensed Matter Physics*, Westview Press, Reading.
- Andreev, A. F. & Lifshitz, I. M. (1969). Quantum theory of defects in crystals, *Soviet Phys. JETP* 29: 1107–.
- Argaman, N. & Makov, G. (2000). Density functional theory: An introduction, *Am. J. Phys.* 68(1): 69–79.
- Aschcroft, N. W. & Mermin, N. D. (1976). *Solid State Physics*, Thomson Learning, Singapore.



- Bausch, R., Schmitz, R. & Truski, L. A. (1999). Quantum motion of electrons in topologically distorted crystals, *Ann. Phys. (Leipzig)* 8(3): 181–189.
- Bausch, R., Schmitz, R. & Turski, L. A. (1999). Scattering of electrons on screw dislocations, *Phys. Rev. B* 59(21): 13491–13493.
- Bausch, R., Schmitz, R. & Turski, L. A. (1998). Single-particle quantum states in a crystal with topological defects, *Phys. Rev. Lett.* 80(11): 2257–2260.
- Baym, G. & Mermin, N. D. (1947). Determination of thermodynamic Green's functions, *J. Math. Phys.* 2(2): 232–234.
- Berestetskii, V. B., Lifshitz, E. M. & Pitaevskii, L. P. (1994). *Quantum Electrodynamics*, Pergamon, Exeter.
- Bethe, H. A. (1947). The electromagnetic shift of energy levels, *Phys. Rev.* 72(4): 339–341.
- Bhadeshia, H. K. D. H. & Honeycombe, R. W. K. (2006). *Steels: Microstructure and Properties (3rd Ed.)*, Elsevier, Amsterdam.
- Bjorken, J. D. & Drell, S. D. (1965). *Relativistic Quantum Fields*, McGraw-Hill, New York.
- Bloch, F. (1928). Über die Quantenmechanik der Elektronen in Kristallgittern, *Z. Physik* 52(7-8): 555–600.
- Bloch, F. (1932). Zur Theorie des Austauschproblems und der Remanenzerscheinung der Ferromagnetika, *Z. Physik* 74(5-6): 295–335.
- Bodraykov, V. Y. (2007). Thermodynamics simulation of the kovar and invar behavior of ferromagnetics, *Phys. Metal. Metallogr.* 104(1): 19–28.
- Bohm, D. & Pines, D. (1952). A collective description of electron interactions: III. Coulomb interactions in a degenerate electron gas, *Phys. Rev.* 92(3): 625.
- Born, M. & Oppenheimer, J. R. (1927). Zur Quantentheorie der Molekeln, *Ann. Phys.* 389(20): 457–484.
- Born, M. & von Kármán, T. (1912). Über Schwingungen in Raumgittern, *Phys. ZS.* 13: 297–309.
- Born, M. & von Kármán, T. (1913). Zur Theorie des spezifische Wärme fester Körper, *Phys. ZS.* 14: 15–19.
- Bose, S. N. (1926). Plancks Gesetz und Lichtquantenhypothese, *Z. Physik* 26(1): 178–181.
- Bravais, A. (1845). Mémoire sur les systèmes formés par les points distribués régulièrement sur un plan ou dans l'espace, *J. Ecole Polytech.* 19(1): 1–128.
- Brown, R. A. (1845). Dislocation scattering: a new Hamiltonian for describing multivalued elastic displacement fields, *J. Phys. F: Metal Phys.* 19(12): L241–L245.
- Burakovski, L., Preston, D. L. & Silbar, R. R. (2000). Melting as a dislocation-mediated phase transition, *Phys. Rev. B* 61(22): 15011.
- Byron, Jr., F. W. & Fuller, R. W. (1992). *Mathematics of Classical and Quantum Physics*, Dover Publications, Mineola.
- Callaway, J. (1974). *Quantum Theory of the Solid State Physics*, Academic Press, New York.
- Callen, H. B. & Welton, T. A. (1951). Irreversibility and generalized noise, *Phys. Rev.* 53(1): 34–40.
- Cardy, J. (2000). *Scaling and Renormalization in Statistical Physics*, Cambridge University Press, Cambridge.
- Casimir, H. B. G. (1945). On Onsager's principle of microscopic reversibility, *Rev. Mod. Phys.* 17(2-3): 343–350.
- Chaikin, P. M. & Lubensky, T. C. (1995). *Principles of condensed matter physics*, Cambridge University Press, Cambridge.
- Christian, J. W. (2002). *The Theory of Transformations in Metals and Alloys*, Pergamon, Amsterdam.
- Cornell, E. A. & Wieman, C. E. (2002). On the theory of quantum mechanics, *Rev. Mod. Phys.* 74(3): 875–893.

- Cotterill, R. M. J. (1980). The physics of melting, *J. Cryst. Growth* 48(4): 582–588.
- de Oliveira, N. A. & von Ranke, P. J. (2010). Theoretical aspects of the magnetocaloric effect, *Phys. Rep.* 489(4-5): 89–159.
- Debye, P. (1912). Zur Théorie der spezifischen Wärmen, *Ann. Phys.* 344(14): 789–839.
- DeHoff, R. T. (1993). *Thermodynamics in Materials Science*, McGraw-Hill, New York.
- Dietel, J. & Kleinert, H. (2006). Triangular lattice model of two-dimensional defect melting, *Phys. Rev. B* 73(2): 024113.
- Dirac, P. A. M. (1926). On the theory of quantum mechanics, *Proc. R. Soc. (London)* A112(112): 281–305.
- Dirac, P. A. M. (1928a). The quantum theory of the electron, *Proc. R. Soc. (London)* A117(778): 610–624.
- Dirac, P. A. M. (1928b). The quantum theory of the electron. Part II, *Proc. R. Soc. (London)* A118(779): 351–361.
- Dirac, P. A. M. (1998). *The Principles of Quantum Mechanics (4th Ed.)*, Clarendon Press/Oxford, Oxford.
- Doniach, S. & Sondheimer, E. H. (1982). *Green's Functions for Solid State Physicists*, Addison-Wesley, Redwood City.
- Drude, P. (1900). Zur electronentheorie der Metalle, *Ann. Phys.* 306(3): 566–613.
- Dyson, F. J. (1949a). The radiation theories of Tomonaga, Schwinger, and Feynman, *Phys. Rev.* 75(3): 486–502.
- Dyson, F. J. (1949b). The S matrix in quantum electrodynamics, *Phys. Rev.* 75(11): 1736–1755.
- Einstein, A. (1905). Zur Elektrodynamik bewegter Körper, *Ann. Phys.* 322(10): 891–921.
- Einstein, A. (1907). Die Plancksche Theorie der Strahlung und die Theorie der spezifischen Wärme, *Ann. Phys.* 327(1): 180–190.
- Einstein, A. (1911). Eine Beziehung Zwischen dem elastischen Verhalten und der spezifischen Wärme bei festen Körpern mit einatomigen Molekül, *Ann. Phys.* 34: 170–174.
- Einstein, A. (1924). Quantentheorie des einatomigen idealen Gases, *Sitzungsber. K. Preuss. Akad. Wiss., Phys. Math. Kl.* pp. 261–267.
- Einstein, A. (1925). Quantentheorie des einatomigen idealen Gases. 2. Abhandlung, *Sitzungsber. K. Preuss. Akad. Wiss., Phys. Math. Kl.* pp. 3–14.
- Fermi, E. (1926). Zur quantelung des idealen einatomigen Gases, *Z. Physik* 36(11-12): 902–912.
- Fermi, E. (1927). Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente., *Z. Physik* 48(1-2): 73–79.
- Fetter, A. L. & Walecka, J. D. (2003). *Quantum Theory of Many-Particle Systems*, Dover, Mineola.
- Feynman, R. P. (1949a). Space-time approach to quantum electrodynamics, *Phys. Rev.* 76(6): 769–789.
- Feynman, R. P. (1949b). The theory of positrons, *Phys. Rev.* 76(6): 749–759.
- Fock, V. (1930). Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems, *Z. Physik* 61(1-2): 126–148.
- Fowler, R. H. & Jones, H. (1938). The properties of a perfect Einstein-Bose gas at low temperatures, *Proc. Cambridge Phil. Soc.* 34(4): 573–576.
- Fradkin, E. (1991). *Field Theories of Condensed Matter Systems*, Addison-Wesley, Redwood City.
- Gabrielse, G., Hanneke, D., Kinoshita, T., Nio, M. & Odom, B. (2006). New determination of the fine structure constant from the electron g value and QED, *Phys. Rev. Lett.* 97(3): 030802.
- Gabrielse, G., Hanneke, D., Kinoshita, T., Nio, M. & Odom, B. (2007). Erratum: New determination of the fine structure constant from the electron g value and QED, *Phys. Rev. Lett.* 99(3): 039902.

- Galitskii, V. M. & Migdal, A. B. (1958). Application of quantum field theory methods to the many body problem, *Soviet Phys. JETP* 7: 96–104.
- Gammaitoni, L., Häggi, P., Jung, P. & Marchesoni, F. (1998). Stochastic resonance, *Rev. Mod. Phys.* 70(1): 223–287.
- Gaudin, M. (1960). Une Démonstration Simplifiée du Théorème de Wick en Mécanique Statistique, *Nucl. Phys.* 15(1): 89–91.
- Ghosh, G. & Olson, G. B. (2002). Precipitation of paraequilibrium cementite: Experiments, and thermodynamic and kinetic modeling, *Acta Mater.* 50(8): 2099–2119.
- Giuliani, G. F. & Vignale, G. (2005). *Quantum Theory of the Electron Liquid*, Cambridge University Press, Cambridge.
- Goldenfeld, N. (1992). *Lectures on Phase Transitions and the Renormalization Group*, Perseus, Reading.
- Goldstein, H. (1980). *Classical Mechanics (2nd Ed.)*, Addison-Wesley, Reading.
- Goldstone, J. (1957). Derivation of Brueckner many-body theory, *Proc. R. Soc. (London)* 239(1217): 267–279.
- Goldstone, J. (1961). Field theories with «superconductor» solutions., *Nuovo Cim.* 19(1): 154–164.
- Goldstone, J., Salam, A. & Weinberg, S. (1962). Broken symmetries, *Phys. Rev.* 127(3): 965–970.
- Gordon, W. (1926). Der Comptoneffekt der Schrödingerschen Theorie, *Z. Physik* 40(1-2): 117–133.
- Gross, F. (1999). *Relativistic Quantum Mechanics and Field Theory*, John Wiley & Sons, New York.
- Hartree, D. R. (1928). The wave mechanics of an atom with a non-Coulomb central field. Part II. Some results and discussion, *Proc. Cambridge Phil. Soc.* 24(1): 111–132.
- Heisenberg, W. & Pauli, W. (1929). Zur quantendynamik der Wellenfelder, *Z. Physik* 56(1-2): 1–61.
- Hillert, M. & Jarl, M. (1978). A model for alloying effects in ferromagnetic metals, *CALPHAD* 2(3): 227–238.
- Hohenberg, P. & Kohn, W. (1964). Inhomogeneous electron gas, *Phys. Rev.* 136(3B): B864–B871.
- Itzykson, C. & Zuber, J. B. (1980). *Quantum Field Theory*, McGraw-Hill, New York.
- Jang, J. H., Kim, I. G. & Bhadeshia, H. K. D. H. (2009). Substitutional solution of silicon in cementite: A first-principles study, *Comp. Mater. Sci.* 44(4): 1319–1326.
- Jones, W. & March, N. H. (1973a). *Theoretical Solid State Physics Volume 1*, John Wiley & Sons, London.
- Jones, W. & March, N. H. (1973b). *Theoretical Solid State Physics Volume 2*, John Wiley & Sons, London.
- Jordan, P. & Pauli, W. (1928). Über Paulische Äquivalenzverbot, *Z. Physik* 47(3-4): 151–173.
- Kaufman, L. (2001). Computational thermodynamics and materials design, *CALPHAD* 25(2): 141–161.
- Kim, D. J. (1974). Free energy of the interacting electron gas including higher-order exchange effects, *Phys. Rev. B* 9(8): 3307–3312.
- Kim, D. J. (1982). Electron-phonon interactions and itinerant-electron ferromagnetism, *Phys. Rev. B* 25(11): 6919–6938.
- Kim, D. J. (1988). The electron-phonon interaction and itinerant electron magnetism, *Phys. Rep.* 171(4): 129–229.
- Kim, D. J. (1989). Electron-phonon interaction mechanism of magnetovolume and magnetoelasticity effects in itinerant electron ferromagnets, *Phys. Rev. B* 39(10): 6844–6856.
- Kim, D. J. (1999). *New Perspectives in Magnetism of Metals*, Kluwer Academic/Plenum, New York.

- Kim, D. J., Schwartz, B. B. & Praddaude, H. C. (1973). Spin and charge susceptibility of a ferromagnetic electrons gas, *Phys. Rev. B* 7(1): 205–214.
- Kittel, C. (2005). *Introduction to Solid State Physics*, John Wiley & Sons, Danvers.
- Klein, O. (1927). Elektrodynamik und Wellenmechanik von Standpunkt des Korrespondenzprinzips, *Z. Physik* 41(10): 407–442.
- Kleinert, H. & Jiang, Y. (2003). Defect melting models for cubic lattices and universal laws for melting temperatures, *Phys. Lett. A* 313(1-2): 152–157.
- Kleman, M. & Friedel, J. (2008). Disclinations, dislocations, and continuous defects: A reappraisal, *Rev. Mod. Phys.* 80(1): 61–115.
- Kohn, W. & Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects, *Phys. Rev.* 140(4A): A1133–A1138.
- Koopmans, T. (1934). Über die Zuordnung von Wellenfunktionen und Wigenwerten zu den Einzelnen Elektronen Eines Atoms, *Physica* 1(1-6): 104–113.
- Kosterlitz, J. M. & Thouless, D. J. (1973). Ordering, metastability and phase transitions in two-dimensional systems, *J. Phys. C: Solid State Phys.* 6: 1181–1203.
- Kozeschnik, E. & Bhadeshia, H. K. D. H. (2008). Influence of silicon on cementite precipitation in steels, *Mater. Sci. Technol.* 24(3): 343–347.
- Kubo, R. (1957). Statistical-mechanical theory of irreversible process. I. General theory and simple application to magnetic and conduction problems, *J. Phys. Soc. Japan* 12(6): 570–586.
- Kubo, R. (1966). The fluctuation-dissipation theorem, *Rep. Prog. Phys.* 29(1): 255–284.
- Kubo, R., Yokota, M. & Nakajima, S. (1957). Statistical-mechanical theory of irreversible process. I. Response to thermal disturbance, *J. Phys. Soc. Japan* 12(11): 1203–1211.
- Kümmel, S. & Kronik, L. (2008). Orbital dependent density functionals: Theory and applications, *Rev. Mod. Phys.* 80(1): 3–60.
- Landau, L. D. (1957a). Oscillations in a Fermi liquid, *Soviet Phys. JETP* 7: 101–108.
- Landau, L. D. (1957b). The theory of a Fermi liquid, *Soviet Phys. JETP* 3: 920–925.
- Landau, L. D. (1958). The properties of the Green function for particles in statistics, *Soviet Phys. JETP* 7: 182–.
- Landau, L. D. (1959). On the theory of the Fermi liquid, *Soviet Phys. JETP* 8: 70–74.
- Landau, L. D. & Lifshitz, E. M. (1978). *Mechanics*, Pergamon Press, Oxford.
- Landau, L. D. & Lifshitz, E. M. (1980). *Statistical Physics*, Elsevier, Amsterdam.
- Larzar, M. (2010). The gauge theory of dislocations: A nonuniformly moving screw dislocations, *Phys. Lett. A* 374: 3092–3098.
- Lee, J. H., Hsue, Y. C. & Freeman, A. J. (2006). Free energy of the interacting electron gas including higher-order exchange effects, *Phys. Rev. B* 73(17): 172405.
- Lee, J. I., Zhang, H. I. & Choe, A. S. (1989). Self-consistent derivation of electric and magnetic susceptibilities of spin-polarized metallic electron system, *J. Korean Phys. Soc* 22(1): 38–42.
- Leggett, A. J. (1970). Broken symmetries, *Phys. Rev. Lett.* 25(22): 1543–1546.
- Lehmann, H. (1954). Über Eigenschaften von Ausbreitungsfunktionen und Renormierungskonstanten quantisierter Felder, *Nuovo Cim.* 11(11): 342–357.
- Levy, M., Perdew, J. P. & Sahn, V. (1984). Exact differential equation for the density and ionization energy of a many-particle system, *Phys. Rev. A* 30(5): 2745–2748.
- Lines, M. E. (1979). Elastic properties of magnetic materials, *Phys. Rep.* 55(2): 133–181.
- Luttinger, J. M. & Nozières, P. (1962). Derivation of the Landau theory of Fermi liquids. I. Equilibrium properties and transport equation, *Phys. Rev.* 127(5): 1431–1440.
- Madelung, O. (1978). *Introduction to Statistical Physics*, Springer, Berlin.
- Mahan, G. (2000). *Many-Particle Physics (3rd Ed.)*, Kluwer Academic/Plenum, New York.

- Matsubara, T. (1955). A new approach to quantum-statistical mechanics, *Prog. Theor. Phys.* 14(4): 351–378.
- Melrose, D. B. (2008). *Quantum Plasmadynamics: Unmagnetized Plasmas*, Springer, New York.
- Mermin, N. D. (1965). Thermal properties of the inhomogeneous electron gas, *Phys. Rev.* 137(5A): A1441–A1443.
- Mermin, N. D. (1979). The topological theory of defects in ordered media, *Rev. Mod. Phys.* 51(3): 591–648.
- Midownik, A. P. (1977). The calculation of magnetic contributions to phase stability, *CALPHAD* 1(2): 133–158.
- Negele, J. W. & Orland, H. (1988). *Quantum Many-Particle Systems*, Addison-Wesley, Redwood City.
- Nozières, P. & Luttinger, J. M. (1962). Derivation of the Landau theory of Fermi liquids. I. Formal preliminaries, *Phys. Rev.* 127(5): 1423–1431.
- Nyquist, H. (1928). Thermal agitation of electric charge in conductors, *Phys. Rev.* 32(1): 110–113.
- Odom, B., Hanneke, D., D’Urso, B. & Gabrielse, G. (2006). New measurement of the electron magnetic moment using a one-electron quantum cyclotron, *Phys. Rev. Lett.* 97(3): 030801.
- Olson, G. B. (2000). Designing a new material world, *Science* 288(5468): 993–998.
- Onsager, L. (1931a). Reciprocal relations in irreversible processes. I., *Phys. Rev.* 37(4): 405–426.
- Onsager, L. (1931b). Reciprocal relations in irreversible processes. II., *Phys. Rev.* 38(12): 2265–2279.
- Parisi, G. (1988). *Statistical Field Theory*, Addison-Wesley, Redwood City.
- Peskin, M. E. & Schroeder, D. V. (1995). *An Introduction to Quantum Field Theory*, Addison-Wesley, Reading.
- Pines, D. (1962). *The Many-Body Problem*, Benjamin/Cummings, Reading.
- Pines, D. (1999). *Elementary excitations in Solids*, Perseus Books, Reading.
- Planck, M. (1901). Ueber das Gesetz der Energieverteilung im Normalspectrum, *Ann. Phys.* 309(3): 553–563.
- Poénaru, V. & Toulouse, G. (1977). The crossing of defects in ordered media and the topology of 3-manifolds, *J. Phys. France* 8(8): 887–895.
- Ruban, A. V. & Abrikosov, I. A. (2008). Configurational thermodynamics of alloys from first principles: effective cluster interactions, *Rep. Prog. Phys.* 71(4): 046501.
- Sakurai, J. J. (1994). *Modern Quantum Mechanics (Revised Ed.)*, Addison-Wesley, Reading.
- Sawada, K. (1957). Correlation energy of an electron gas at high density, *Phys. Rev.* 106(2): 372–383.
- Sawada, K., Brueckner, K. A., Fukada, N. & Brout, R. (1957). Correlation energy of an electron gas at high density: Plasma oscillations, *Phys. Rev.* 108(3): 507–514.
- Schneider, T. (1971). Theory of the liquid-solid phase transition, *Phys. Rev. A* 3(6): 2145–2148.
- Schrieffer, J. R. (1988). *Theory of Superconductivity*, Addison-Wesley, Reading.
- Schrödinger, E. (1926a). Quantisierung als Eigenwertproblem, *Ann. Phys.* 384(4): 361–376.
- Schrödinger, E. (1926b). Quantisierung als Eigenwertproblem, *Ann. Phys.* 384(6): 489–527.
- Schrödinger, E. (1926c). Quantisierung als Eigenwertproblem, *Ann. Phys.* 385(13): 437–490.
- Schrödinger, E. (1926d). Quantisierung als Eigenwertproblem, *Ann. Phys.* 386(18): 109–139.
- Schwinger, J. (1948). Quantum electrodynamics. I. A covariant formulation, *Phys. Rev.* 74(10): 1439–1461.
- Schwinger, J. (1949a). Quantum electrodynamics. II. Vacuum polarization and self-energy, *Phys. Rev.* 75(4): 651–679.

- Schwinger, J. (1949b). Quantum electrodynamics. III. The electromagnetic properties of the electron—radiative corrections to scattering, *Phys. Rev.* 76(6): 790–817.
- Slater, J. C. (1951). A simplification of the Hartree-Fock method, *Phys. Rev.* 81(3): 385–390.
- Spencer, P. J. (2007). A brief history of calphad, *CALPHAD* 31(4): 4–27.
- Stowasser, R. & Hoffmann, R. (1999). What do the Kohn-Sham orbitals and eigenvalues mean?, *J. Am. Chem. Soc.* 121(14): 3414–3420.
- Strandburg, K. J. (1986). Two-dimensional melting, *Rev. Mod. Phys.* 60(1): 161–207.
- Strandburg, K. J. (1989). Erratum: Two-dimensional melting, *Rev. Mod. Phys.* 61(3): 747.
- Teichler, H. (1981). Gauge fields for electrons in crystals with topological defects, *Phys. Lett. A* 87(3): 113–115.
- Thomas, L. H. (1938). The calculation of atomic fields, *Proc. Cambridge Phil. Soc.* 23(5): 542–548.
- Tinkam, M. (1964). *Group Theory and Quantum Mechanics*, McGraw-Hill, New York.
- Tomonaga, S. (1946). On a relativistically invariant formulation of the quantum theory of wave fields, *Prog. Theor. Phys.* 1(2): 27–42.
- Toulouse, G. & Kléman, M. (1976). Principles of a classification of defects in ordered media, *J. Physique Lett.* 37(6): 149–151.
- Turchi, P. E. A., Abrikosov, I. A., Burton, B., Fries, S. G., Grimvall, G., Kaufman, L., Korzhavyi, P., Manga, V. R., Ohno, M., Pisch, A., Scott, A. & Zhang, W. (2007). Interface between quantum-mechanical-based approaches, experiments, and calphad methodology, *CALPHAD* 31(1): 4–27.
- Turski, L. A., Bausch, R. & Schmitz, R. (2007). Gauge theory of sound propagation in crystals with dislocations, *J. Phys.: Condens. Matter* 19(9): 096211.
- Turski, L. A. & Mińkowski, M. (2009). Spin wave interaction with topological defects, *J. Phys.: Condens. Matter* 21(37): 376001.
- v. Löhneysen, H., Rosch, A., Vojta, M. & Wölfle, P. (2006). Fermi-liquid instabilities at magnetic quantum phase transitions, *Rev. Mod. Phys.* 79(3): 1015–1075.
- Vilenkin, A. (1985). Cosmic strings and domain walls, *Phys. Rep.* 121(5): 263–315.
- Wick, G. C. (1950). The evaluation of the collision matrix, *Phys. Rev.* 80(2): 268–272.
- Wigner, E. (1927). Einige Folgerungen aus der Schrödingerschen Theorie für die Termstrukturen, *Z. Physik* 43(9-10): 624–652.
- Wilson, K. G. (1975). The renormalization group: Critical phenomena and the Kondo problem, *Rev. Mod. Phys.* 47(4): 591–648.
- Wilson, K. G. (1983). The renormalization group and critical phenomena, *Rev. Mod. Phys.* 55(3): 583–600.
- Zinn-Justin, J. (1997). *Quantum Field Theory and Critical Phenomena*, Clarendon Press/Oxford, Oxford.

# Thermodynamics of the Phase Equilibriums of Some Organic Compounds




Raisa Varushchenko and Anna Druzhinina  
*Lomonosov Moscow State University  
Russia*

## 1. Introduction

A comprehensive investigation of the phase equilibriums and determination of thermodynamic properties of pure substances is a significant object of the chemical thermodynamics. Data on the phase transitions, heat capacities, and saturation vapor pressure over the solid and liquid phases are used in many fields of science and technology, including calculations on the basis of the third law of thermodynamics. Theoretical and practical applications of thermodynamic data require verification of their reliability. The Clapeyron equation combines different properties of coexisting phases: temperature, vapor pressure, volume, enthalpy of the phase transitions, and caloric values  $C_p$  and  $C_v$ . Using this equation allows one to verify numerical data for thermodynamic concordance, to reveal unreliable quantities, and to predict failing thermodynamic properties. Mutual concordance and reliability of the calorimetric data on the heat capacity, the saturated vapor pressures, and the properties of phase transition can be verified by comparison of the absolute entropies determined from the experimental data by the third thermodynamic law,  $S_m^0(g)(expt)$  with those ones calculated by statistical thermodynamics,  $S_m^0(stat)$ . A congruence of these values within errors limits justifies their reliability. Critical analyses of the recent data on thermodynamic properties of some organic compounds are published by the National Institute of the Standards and Technology [NIST], USA. Literature data on the vapor pressures and the enthalpies of vaporization for *n*-alkanes  $C_5 - C_{20}$  were reviewed and critically analyzed in the reference (Ruzicka & Majer, 1994). Thermodynamic properties of many classes of organic compounds were considered in monograph (Domalski & Hearing, 1993; Poling et al., 2001) that favoured the development of the Benson's calculation method. This chapter deals with reviewing and summarizing the data on the phase equilibriums carried out for some functional organic compounds by the low temperature adiabatic calorimetry, comparative ebulliometry, and vaporization calorimetry in the Luginin's Laboratory of Thermochemistry [LLT] of the Moscow State University [MSU] and other research centres. The numerous data on the heat capacity, the vapor pressure, enthalpies of the phase transitions, and derived thermodynamic functions were obtained for series of freons, cyclic hydrocarbons and fluorocarbons, and derivatives of ferrocene. A sufficient attention was given to the critical analyses of the thermodynamic data, their reliability, and to interconnections between the properties and some structural parameters of the

compounds. Experimental and calculation methods for determination of the properties rely mostly on the LLT-school.

Freons are halogen derivatives of ethane and propane which possess a unique combination of the useful properties: high volatility, high enthalpy of vaporization, no combustibility, biological inertness, etc. Due to these properties, freons have found a wide application in many areas of science, technology, and medicine (Varushchenko et. al., 2007).

Alkyl derivatives of adamantane,  $C_{10}H_{16}$ , , are of an interest due to tendency to form plastic crystals. Bicyclic *cis*- and *trans*- isomers of decaline,  $C_{10}H_{18}$ , , and hydrindane,  $C_9H_{16}$ , , have poor intermolecular interactions and also form plastic crystals. Their perfluoride counterparts exhibit high chemical stability, absolute biological inertness, and capacity for dissolving and transferring large amounts of gases, in particular, oxygen and carbon dioxide. Due to these properties, perfluorocarbons have found wide application in biology and medicine as effective gas-transferring media and artificial blood substitutes. A mixture of perfluorodecaline,  $C_{10}F_{18}$ , and perfluoro-N-(4-methylcyclohexyl)piperidine,  $C_{12}F_{23}N$ , forms of the "Ftorosan" blood substitute (Russia) (Ries, 1991). Bicyclic *cis*- and *trans*- isomers of decaline and hydrindane are of the interest in study of an interconnection between thermodynamic properties and the structure of the compounds when passing from perfluorocarbons to their hydrocarbon counterparts.

Alkyl- and acyl- ferrocene derivatives [FD] are the sandwich-type organometallic compounds discovered in the 50<sup>th</sup> years of the XX century. Owing to favourable conjunction of the chemical and physical properties, namely low toxicity, high thermal stability, and volatility, some FD has found ever-increasing application in technology (electric materials, regulators of fuel combustion etc.) and medicine (anti-cancer and blood-creating drugs).

This chapter is intended for researchers with an interest in measuring characteristics of the phase transitions and in determination of the equilibrium properties by experimental and theoretical methods. A number of relationships for practical use are represented with illustrative examples and necessary recommendations. The chapter contains main references to the literature used in reviewing and summarizing the numerous data on the properties of some functional organic compounds.

Part 2 deals with the ebulliometric and transpiration methods for determination of the saturation vapor pressure in dependence on the temperature. Design of devices and experimental techniques and mathematical processing of the vapor pressures are given. A modified ebulliometer of an original construction was given for determination of the  $pT$  parameters in moderate ("atmospheric") pressure region. The enthalpies of vaporization obtained by direct calorimetric method and those ones calculated from the vapor pressure are compared for justifying their reliability. An interconnection between the properties derived from the vapor pressure and some structural parameters of the substances are analyzed.

Part 3 considers the low-temperature adiabatic calorimetry for measuring the heat capacity and studying the properties of the phase transitions. Experimental technique has been presented by modern completely automated adiabatic calorimeter used in LLT. Experimental determination and mathematical processing of the phase transitions were given including an X-ray analysis of crystal structure and the infrared and Raman spectroscopy for interpretation of the processes occurring during the solid-phase transitions. Main thermodynamic functions (changes of the entropy, enthalpy, and Gibb's energy) in



condensed states were calculated on the basis of the heat capacities and the properties of the solid-to-solid transitions and fusion.

Part 4 deals with 1) determination of the ideal gas thermodynamic functions by experimental and theoretical methods, 2) verification of the thermodynamic functions by comparing the absolute entropies calculated on the basis of the third thermodynamic law and by statistical thermodynamics, and 3) the methods of extending the saturated vapor pressure of the "atmospheric" range of pressure to entire region of liquids from the triple to the critical temperatures.

Parts 5, 6, 7, and 8 present Conclusion, Acknowledgments, References, and Appendix, respectively.

## 2. Temperature dependence of saturated vapor pressure

The values of the vapor pressure of liquid substances are mostly determined by the static and dynamic (mainly ebulliometric) methods. A comparative ebulliometry is frequently employed due to its simpler technique and suitability for the series of determinations. The greatest number of saturated vapor pressure of organic compounds was obtained by this method in the moderate ("atmospheric") range of pressure  $2-3 \leq (\Delta p, \text{kPa}) \leq 100-150$ . The highest accurate of vapor pressure is usually attained in this range that makes it possible to obtain reliable derivative values, in particular, the enthalpies of vaporization. Few  $pT$  data are available in the literature for the entire region of liquid phase because of methodical difficulties and high errors of determination at low ( $<1$  kPa) and high ( $>200$  kPa) pressures.

### 2.1 Experimental and mathematical processing

Fig. 1 presents a schematic view of a setup designed for determinations of the temperature dependence of saturation vapor pressure by comparative ebulliometry (Varouchtchenko & Droujinina, 1995).

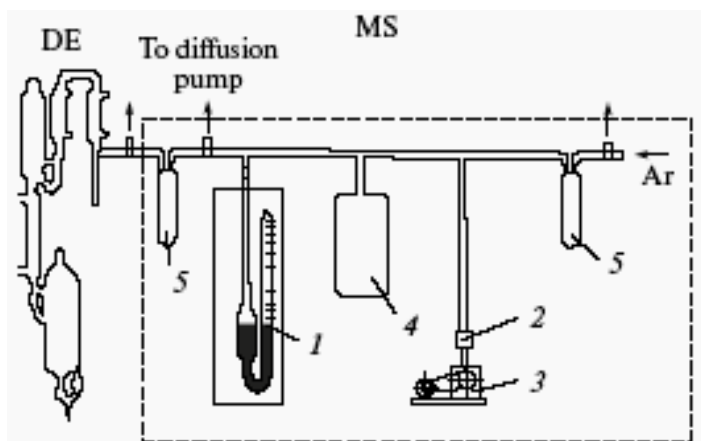


Fig. 1. The setup for determination of the  $pT$  parameters: DE, differential ebullimeter; MS, manometer system; (1) mercury-contact manometer; (2) electromagnetic valve; (3) roughing pump; (4) ballast reservoir; (5) traps

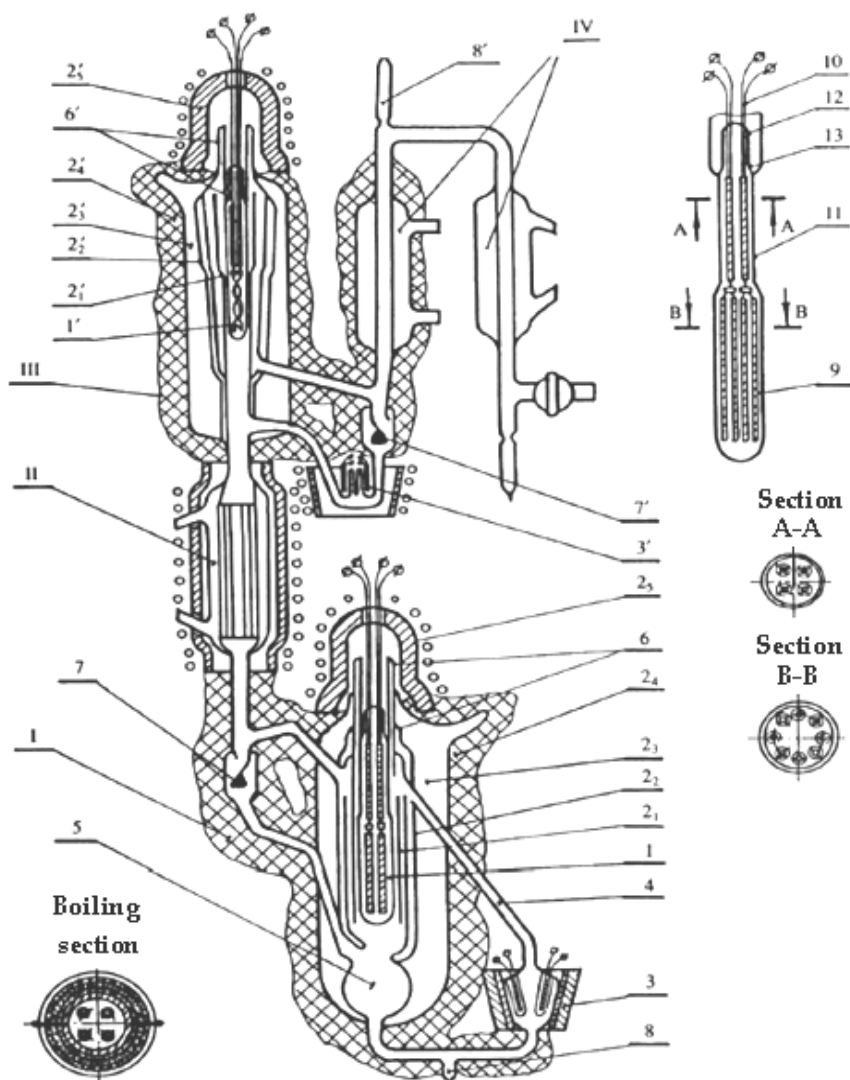


Fig. 2. The differential ebullimeter: I, boiling section; II, rectification column; III, condensation section; IV, system of coolers for returning and collecting a condensate; (1, 1') platinum resistance thermometers; (2<sub>1</sub>, 2<sub>1</sub>') glass screens of thermometers; (2<sub>2</sub>, 2<sub>2</sub>') silver radiation screens; (2<sub>3</sub>, 2<sub>3</sub>') vacuum shells; (2<sub>4</sub>, 2<sub>4</sub>') heat-insulating layer (asbestos); (2<sub>5</sub>, 2<sub>5</sub>') shells for heating the thermometer parts extending from the ebullimeter; (3) boiler; (3') U-shaped liquid valve; (4) Cottrell pump; (5) spherical reservoir; (6 (13), 6') differential Chromel-Alumel thermocouples; (7, 7') droplet counters; (8, 8') branches for outlet and inlet of liquid; (9) sensing element of platinum resistance thermometer; (10) platinum wires; (11) protective glass tube; (12) Pyrex-tungsten glass-molybdenum glass transition.

The setup consists of a differential ebulliometer used for measuring the boiling and condensation temperatures and manometer system operating in the manostat mode. The main part of MS system is a mercury-contact (tungsten) manometer that serves for automatic control and determination of the pressure inside the ebulliometer. Argon was introduced into the system to maintain the constant pressure equal to that of the saturation vapors of the substance under study. The temperature of the (liquid + vapor) equilibrium was measured at 20 fixed pressures controlled by manometer system.

A schematic view of modified Swietoslowski -type ebulliometer is given in Fig. 2. The differential ebulliometer was used for determination of the temperature dependence of the vapor pressure by measuring the boiling,  $T_{boil}$ , or (rarely) condensation,  $T_{cond}$ , temperatures and for estimation of an ebulliometric degree of purity for the samples by the difference ( $T_{boil} - T_{cond}$ ). The latter is below 0.005 K for the pure substances. The boiling and condensation sections and other parts of differential ebulliometer were made of "Pyrex" glass and were sealed together. The modification of the ebulliometer was directed for solving three basic problems: 1) increasing the thermometric sensitivity of a system used for temperature measurements of the (liquid - vapor) equilibrium; 2) decreasing a heat exchange of the temperature sensors with the surrounding, and 3) reducing the superheating of the boiling liquid, that leads to increasing the accuracy of the temperature measurements.

For increasing the sensitivity of the thermometers, their protecting tubes were soldered in the boiling and condensation sections of the ebulliometer. Sensing elements of vibration - resistant thermometers ( $R \approx 100\Omega$ ) consisted of a few platinum spirals wound around glass capillaries. The latter had coefficients of linear expansion close to that of platinum.

Connecting wires (current and potential) of the thermometer were vacuum - tight sealed through the glass-molybdenum part of a passage (12, Fig. 2) of the protecting tube. The thermometers were graduated in Mendeleev's Institute of Metrology (S. Petersburg) at the triple - point temperature of water (273.16 K) and melting temperatures of tin (505.118 K) and gallium (302.920 K). The summary error of graduation is  $\pm 3 \cdot 10^{-3} K$ . A special system of heat insulation of the thermometers was employed. It consisted of: glass screens washed by boiling liquid or by condensate ( $2_1$  or  $2_1'$ ), respectively (Fig. 2); silver radiation screens; vacuum jackets ( $p \approx 1.3 \cdot 10^{-3} Pa$ ); heat - insulating layers (asbestos), and electrically heated screens ( $2_5$  and  $2_5'$ ) for upper parts of thermometers overhanging from ebulliometer. Application of such heat - insulating system made it possible to conduct precision temperature measurements without heating the main part of the ebulliometer. The error of temperature measurements caused by heat exchange of the thermometers with the surroundings, were estimated on the basis of heat - exchange laws to be  $\approx 1 \cdot 10^{-3} K$ . The superheating of the liquid was reduced by using several internal and two external boiler heaters which promoted to smooth boiling of the liquid. Performed modification of the ebulliometer allowed cutting down substantially the amount of liquid which was spent for heating the inner surfaces of instrument up to working temperature.

Thus, the necessary volume of liquid was reduced several times: down to  $\approx 5cm^3$  when measuring only the boiling temperature and to  $\approx 8cm^3$  when both boiling and condensation temperatures were measured. The use of the comparison method makes it possible to

reduce the  $pT$  parameters determination to the precision temperature measurements. The temperature was automatically measured by potentiometer method and the results were displayed on a personal computer [PC] screen with the aid of the AK-6.25 computer-measurement system designed at All-Russia Research Institute of Physico-technical and Radio-technical Measurements [VNIIFTRI].

An automatic maintenance of the constant pressure was attained by a mercury-contact manometer which was controlled by vacuum pump via an electromagnetic valve (Fig. 1). The pressure of argon fluctuated in the limits from ( $\pm 20$  to  $\pm 40$ ) Pa. The boiling temperature was measured at the highest pressure in the cycle at the moment of mercury-to-tungsten contact. The manometer was thermostated at the temperature ( $300.00 \pm 0.02$ ) K. The measurements of the boiling and condensation temperatures were conducted after attaining thermodynamic equilibrium in the ebulliometer. To be assured that the liquid under study had not decomposed, the boiling temperature at one of initial points of the  $pT$  curve was measured several times during the ebulliometric experiments.

Errors of temperature  $S_T$  and vapor pressure  $S_p$  measurements were calculated as:

$$|S_T| = \left\{ (t \cdot S_{T1})^2 + (S_{T2})^2 \right\}^{1/2}$$

$$|S_p| = \left\{ (dp/dT)_1^2 \cdot (t \cdot S_{T1})^2 + (dp/dT)_3^2 \cdot (t \cdot S_{T3})^2 \right\}^{1/2}$$

where  $|t \cdot S_{T1}|$  and  $|t \cdot S_{T3}|$  denote the instrumental errors of the temperature measurements ( $\leq 5 \cdot 10^{-3} K$ ) at the substance research and at graduation of the mercury-contact manometer;  $t$  is Student's criterion;  $|S_{T2}| = 3 \cdot 10^{-3} K$  denotes the error of graduation of the thermometer; and  $(dp/dT)_1$  and  $(dp/dT)_3$  are temperature coefficients of the pressure for standard and studied substances, respectively. The total uncertainty of temperature measurement was  $|S_T| \leq 6 \cdot 10^{-3} K$ . The error of graduation of the mercury-contact manometer by means of water and *n*-decane and the error of determination of the vapor pressure of the substance under study were equal to  $|S_p(\text{grad.})| = (13 \text{ to } 20) \text{ Pa}$  and  $|S_p| = (20 \text{ to } 26) \text{ Pa}$ , respectively.

The accuracy of ebulliometric measurements was checked by determinations of the saturation vapor pressures of substances having significantly different boiling temperatures, namely benzene and undecane. The normal boiling temperatures of the standard substances obtained in this work agree within errors limits  $\pm \leq 0.01 K$  with precise values of reference (Boublik et al., 1984).

Comparative ebulliometry was employed for determination a series of saturation vapor pressures in dependence on temperature for some freons; halogen - ethanes and -propanes; alkyladamantanes; *cis*- and *trans*-hydrindanes, *cis*- and *trans*-decalines, and their fluoridated counterparts.

The mathematical processing of the observed boiling temperatures and vapor pressures were conducted by the semi-empirical equation:

$$\begin{aligned}
 -R \cdot T \cdot \ln(p) = \Delta H_m(\langle T \rangle) - a_1 \cdot T + a_2 \cdot \{T - \langle T \rangle - T \cdot \ln(T / \langle T \rangle)\} \\
 - a_3 \cdot \left\{ (1/2) (T^2 - \langle T \rangle^2) - T \cdot \langle T \rangle \cdot \ln(T / \langle T \rangle) \right\}
 \end{aligned} \quad (1)$$

where  $\langle T \rangle$  denotes the mean temperature and  $\Delta H_m(\langle T \rangle)$ ,  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  are parameters. Equation (1) was derived by integration of the Clapeyron equation:

$$d \ln(p) / dT = \Delta_{vap} H_m / (\Delta Z \cdot R \cdot T^2) \quad (2)$$

with the approximation for  $\Delta_{vap} H_m / \Delta Z$ :

$$\Delta_{vap} H_m / \Delta Z = \Delta H'_m(T) = \Delta H_m(\langle T \rangle) + \alpha_2 \cdot (T - \langle T \rangle) + (1/2) \cdot \alpha_3 \cdot (T - \langle T \rangle)^2, \quad (3)$$

where  $\Delta Z$  denotes the difference of compression factors of gas and liquid. Equation (3) in turn was developed by integration of the approximation for  $\Delta C_{p,m} = C_{p,m}^0(g) - C_{p,m}(liq) \approx \alpha_2 + \alpha_3 \cdot (T - \langle T \rangle)$ , as a linear function of the temperature. The treatment of the  $pT$  parameters was carried out by the least-squares method [LSM] using orthogonal functions (Kornilov & Vidavski, 1969). Mathematical processing of the saturation vapor pressures is given in Appendix. A system of normal equations of LSM is a diagonal matrix relative to the orthogonal functions. The latter are mutually independent that allows to evaluate their uncertainties and those ones for the  $\ln\{p(T)/Pa\}$  and  $\Delta_{vap} H_m(T)$  functions and, as a result, to choice of an adequate number of terms of relations (1) and (3) by curtailing or expanding terms to suit the accuracy of the parameters of these relations without a new treatment of  $pT$  data. Final equations for these functions are set out for compactness, as:

$$\ln(p / Pa) = A + B / T + C \cdot \ln(T / K) + D \cdot T \quad (4)$$

$$\Delta_{vap} H_m = R \cdot (-B + C \cdot T + D \cdot T^2) \cdot \Delta Z \pm [s\{\Delta H'_m(T)\} + \Delta_{vap} H_m \cdot \Delta(\Delta Z)] \quad (5)$$

where  $A$ ,  $B$ ,  $C$ , and  $D$  are constants related to the parameters of equation (1) by linear correlations;  $s\{\Delta H'_m(T)\}$  is the uncertainty of  $\Delta_{vap} H_m$  value resulting from errors of  $(p, T)$  parameters; and  $\Delta(\Delta Z)$  is the error of the  $\Delta Z$  difference estimation. The  $s\{\Delta H'_m(T)\}$  values are evaluated by the law of random errors accumulation on the basis of dispersions of the orthogonal functions (Appendix).

Because the coefficients of equations (4) and (5) are correlated, the numbers of digits in  $A$ ,  $B$ ,  $C$ , and  $D$  coefficients were selected so that the calculated  $p$  values would not exceed the experimental errors of the vapor pressure determination. (Appendix). Statistical analysis of the error of the smallest parameter  $\alpha_3$  of the equation (1) (accordingly,  $D$  of equations (4) and (5)) was evaluated by the Fisher criterion,  $F$ . If the inequality:

$$F = \alpha_3^2 / s^2(\alpha_3) \geq F_{0.05}(1, f), \quad (6)$$

Compounds	Purity, mol. %	$\Delta T$ ( $pT$ ), K	$n$	$A$	$-B$	$-C$	$D \cdot 10^3$	$S_p$ , Pa
Freons and halogen -ethanes and -propanes								
CFCl <sub>2</sub> CFCl <sub>2</sub>	99.30 <sup>a</sup>	313–361	12	59.2013	5922.1	6.71105	3.3075	2.1
CF <sub>2</sub> ClCFCl <sub>2</sub>	99.80 <sup>a</sup>	298–316	7	42.1123	4680.0	3.96897	-	19.0
CF <sub>2</sub> ClCF <sub>2</sub> Cl	99.79 <sup>a</sup>	178–277	10	320.3557	10029.4	55.11366	110.0486	34.5
CF <sub>2</sub> BrCF <sub>2</sub> Br	99.50 <sup>c</sup>	298–320	8	43.4680	4732.1	4.17343	-	10.5
CF <sub>2</sub> ClCHCl <sub>2</sub>	99.64 <sup>b</sup>	297–345	14	141.7579	7982.7	21.01696	25.5415	15.0
CFCl <sub>2</sub> CHFCl	99.38 <sup>b</sup>	289–346	16	137.9719	7912.3	20.33885	24.3167	11.0
CF <sub>3</sub> CHCl <sub>2</sub>	99.83 <sup>a</sup>	256–454	45	90.4773	5766.1	12.41706	13.8312	3.8
CF <sub>2</sub> ClCHFCl	99.51 <sup>a</sup>	278–303	9	655.6005	20714.9	111.4046	177.9514	10.0
CF <sub>3</sub> CHClBr	99.7 <sup>a</sup>	297–323	8	45.2236	4950.1	4.37789	-	5.6
CF <sub>3</sub> CH <sub>2</sub> CH <sub>2</sub> Cl	99.90 <sup>a</sup>	297–315	8	47.6041	5040.0	4.71465	-	11.9
CF <sub>3</sub> CH <sub>2</sub> CHCl <sub>2</sub>	99.30 <sup>c</sup>	302–341	12	133.0670	8026.1	19.24539	21.0475	13.9
CF <sub>3</sub> CH <sub>2</sub> CCl <sub>3</sub>	99.59 <sup>a</sup>	321–364	12	243.2798	11578.6	38.06825	48.1087	20.3
CF <sub>3</sub> CH <sub>2</sub> CFCl <sub>2</sub>	99.9 <sup>b</sup>	297–333	9	1415.3662	43459.2	241.45741	366.8286	36.8
CHCl <sub>2</sub> CH <sub>3</sub>	99.9 <sup>b</sup>	294–330	11	135.6402	7511.9	20.10746	25.2571	10.1
CH <sub>2</sub> ClCH <sub>2</sub> Cl	99.9 <sup>b</sup>	299–356	15	83.3156	6652.5	10.77546	9.2008	6.2
CH <sub>2</sub> BrCH <sub>2</sub> Br	99.9 <sup>b</sup>	331–426	12	127.1428	8810.4	17.99911	18.0408	23.1
CHCl <sub>2</sub> CH <sub>2</sub> Cl	99.9 <sup>b</sup>	316–384	11	90.36301	7530.3	11.71063	9.0204	7.1
CHCl <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	99.9 <sup>b</sup>	312–362	14	89.3533	6827.0	11.8439	10.8245	4.5
CH <sub>2</sub> ClCHClCH <sub>3</sub>	99.9 <sup>b</sup>	303–368	15	98.1061	7219.8	13.2971	12.6285	14.4
CH <sub>2</sub> ClCH <sub>2</sub> CH <sub>2</sub> Cl	99.8 <sup>b</sup>	330–393	14	140.9922	9134.9	20.33477	21.1678	5.1
CH <sub>3</sub> CCl <sub>2</sub> CH <sub>3</sub>	99.72 <sup>a</sup>	295–341	12	176.6209	8828.3	27.13786	35.4801	11.6
CH <sub>3</sub> CCl <sub>3</sub>	99.99 <sup>a</sup>	296–371	18	44.7407	5209.8	4.29370	-	8.0
CH <sub>3</sub> CCl <sub>3</sub>	99.95 <sup>b</sup>	174–223	8	117.9294	4801.5	18.43752	34.8788	6.3
Alkylderivatives of adamantane								
1,3,5-TMA	99.98 <sup>a</sup>	385–482	16	130.7579	10220.9	18.22485	15.8278	10.0
1,3-DMA	99.9 <sup>a</sup>	352–526	24	101.7980	9034.5	13.50183	10.5779	2.0
1-EA	99.93 <sup>a</sup>	387–498	14	115.7566	10147.8	15.57166	11.9069	10.6
Bicyclic hydrocarbons								
cis-C <sub>9</sub> H <sub>16</sub>	99.99 <sup>a</sup>	351–442	18	118.1951	9158.1	16.27432	14.2522	9.2
trans-C <sub>9</sub> H <sub>16</sub>	99.98 <sup>a</sup>	345–435	18	107.7679	8639.8	14.60163	12.5083	11.3
cis-C <sub>10</sub> H <sub>18</sub>	99.87 <sup>a</sup>	373–470	19	129.3296	10107.8	17.94129	15.3346	7.5
trans-C <sub>10</sub> H <sub>18</sub>	99.98 <sup>a</sup>	366–461	19	105.6064	9031.6	14.11863	11.3055	5.1
Bicyclic perfluorocarbons								
cis-C <sub>9</sub> F <sub>16</sub>	99.69 <sup>a</sup>	316–392	17	160.6582	9773.9	22.20257	21.4086	8.0
trans-C <sub>9</sub> F <sub>16</sub>	99.40 <sup>a</sup>	314–389	17	153.5136	9444.2	21.06404	20.3262	6.0
cis-C <sub>10</sub> F <sub>18</sub>	99.57 <sup>a</sup>	315–416	19	218.3225	12353.8	32.91383	34.7585	6.4
trans-C <sub>10</sub> F <sub>18</sub>	99.46 <sup>a</sup>	313–414	18	195.0918	11539.7	29.04017	29.8875	10.9
C <sub>5</sub> F <sub>10</sub> N-C <sub>6</sub> F <sub>10</sub> -CF <sub>3</sub>	99.66 <sup>a</sup>	374–461	18	210.0577	13500.8	30.83011	28.0834	5.0

<sup>a</sup>Adiabatic calorimetric; <sup>b</sup> DSC.

Table 1. Thermodynamic parameters of comparative ebulliometry for compounds studied: freons; halogen -ethanes and -propanes; 1,3-dimethyladamahtane [1,3-DMA], 1,3,5-trimethyladamahtane [1,3,5-TMA] and 1-ethyladamahtane [1-EA]; perfluorobicyclo(4,3,0)-nonanes [*cis*- and *trans*- C<sub>9</sub>F<sub>16</sub>], bicyclo(4,3,0)nonanes, [*cis*- and *trans*- C<sub>9</sub>H<sub>16</sub>], perfluoro-bicyclo(4,4,0)decane, [*cis*- and *trans*- C<sub>10</sub>F<sub>18</sub>], bicyclo(4,4,0)decanes [*cis*- and *trans*- C<sub>10</sub>H<sub>18</sub>]; perfluoro-N-(4-methyl-cyclohexyl)piperidine [C<sub>5</sub>F<sub>10</sub>N-C<sub>6</sub>F<sub>10</sub>-CF<sub>3</sub>] (Varushchenko et al., 2007; Boublik et al., 1984)

is satisfied, the parameter  $\alpha_3$  (D) may be accepted as a reliable one. Here  $F$  and  $F_{0.05}(1, f)$  denote evaluated and tabulated values of the  $F$ -criterion, and  $f$  is a number of degrees of freedom. Comparing the criteria  $F$  and  $F_{0.05}(1, f)$  according to (6) showed an adequate fit of the  $pT$  parameters.

Table 1 summarizes the purity of the compounds determined by gas - liquid chromatography [g.l.c.] and adiabatic calorimetry, the temperature interval,  $\Delta T$  ( $pT$ ), and number,  $n$ , of  $pT$ -parameters, the coefficients of equations (4) and (5) and mean-square deviation [MSD] of calculated  $p_{calc}$ -values from experimental ones,  $p$ ,

$$S_p = \pm \{ \sum (p - p_{calc})^2 / (n - 4) \}^{1/2}$$

## 2.2 The enthalpy of vaporization

Experimental determinations of the enthalpies of vaporization were carried out by direct calorimetric methods and by indirect ones, on the basis of the temperature dependences of saturation vapor pressures. The first method is more precise but the second one is more often used because of its applicability for wider series of the substances.

The enthalpies of vaporization of some compounds under study were determined at  $T = 298.15$  K by calorimetric method using a carrier gas (nitrogen) (Wadsö, 1966). The method is based on measuring the energy dissipated in calorimeter for compensation of the endothermic vaporization effect. The carrier gas was employed for hastening an evaporation process and, thus, for increasing an accuracy. A modified LKB 8721-3 setup consists of some commercial parts, namely calorimetric vessel with an air brass jacket and a carrier gas system and three missing parts designed in (Varushchenko et. al., 1977): precise water thermostat, electrical scheme, and an air thermostat. The latter replaced a thermostated room that was provided for operating by this method. The calorimeter is intended for the substances with vapor pressures from 0.066 kPa to 26.6 kPa at 298 K (or normal boiling temperatures from (335 to 470) K). A mass (0.5 to 1.0) g of substance was required for a series from 6 to 8 experiments.

The calorimetric experiment was conducted at an adiabatic and, at the same time, at isothermal conditions. The temperature of the calorimetric vessel measured by a thermistor was maintained constant and equal to that of the thermostat ( $298.15 \pm 0.02$ ) K. Electrical energy used for compensation of the energy of vaporization (20 to 40) J was measured by a potentiometer method with accuracy 0.01 per cent. The mass,  $m$ , of a substance evaporated (0.07 to 0.3) g was determined to  $\pm 1 \cdot 10^{-4}$  g as the difference between masses of calorimetric vessel before and after an experiment. As the calorimeter was non-hermetic, the main error in mass determination arose from a loss of substance in weighing the vessel due to connecting and disconnecting it with the calorimetric system. All preliminary procedures such as filling the vessel with liquid, weighing it, and placing into its air jacket were made inside an air thermostat at  $T \approx 298$  K. In so doing, we reduced a loss of the substance from the vessel and the temperature over fall of the latter.

The value of  $\Delta_{vap}H$  was corrected for a small quantity of energy absorbed during the passage of nitrogen through the calorimeter under low pressure. The calorimeter was tested by measuring the enthalpies of vaporization of  $n$ -alkanes from  $C_6$  to  $C_{10}$ . Obtained values of

$\Delta_{vap}H$  at  $T = 298.15$  K agree with well established literature values (Majer & Svoboda, 1985) within (0.2 to 0.5) per cent.

A main method of determination of the enthalpies of vaporization is until now an indirect one based on the temperature dependence of the vapor pressure. This is caused by a less complicated technique for precise vapor pressure determinations than direct calorimetric measurements of  $\Delta_{vap}H_m$ . The best-accuracy estimations of  $\Delta_{vap}H_m$  values are attained for a moderate range of vapor pressure (5 to 150) kPa. The literature data on the enthalpies of vaporization obtained by indirect method are usually published without uncertainties, that can be explained by fitting the  $pT$ -parameters with  $\ln(p) = f(T)$  equations, coefficients of which were correlated. An accuracy determination of the enthalpies of vaporization in indirect method is given in Appendix. The  $\Delta_{vap}H_m$  values obtained by indirect method were computed by equation (5) using the  $\Delta Z$  difference which took into account the vapor deviation from ideality and volume changes of both phases. The  $\Delta Z$  values were calculated from formula:

$$\Delta Z = \{p / (R \cdot T)\} \cdot \{V_m(g) - V_m(liq)\}. \quad (7)$$

The molar volume  $V_m(liq)$  of liquid was evaluated on the basis of density; an adequate value for the volume of vapor,  $V_m(g)$ , was calculated from the volume-explicit virial expansion truncated after the second virial coefficient  $B_v$ . The values of  $B_v$  were evaluated on the basis of critical quantities (part 4.2) by the Tsonopolous extension of Pitzer and Curl's method (Poling et al., 2001). Comparing two series of  $\Delta Z$  values estimated from experimental and calculated values of  $V_m(g)$  of hydrocarbons enable us to accept the errors of  $\Delta Z$  evaluation  $\leq 1$  per cent.

**Freons and halogenalkanes.** Table 2 presents the normal boiling temperatures,  $T_{n.b.}$ , and the enthalpies of vaporization at  $T = 298.15$  K and  $T_{n.b.}$  for freons and hlogenalkanes, calculated from equation (4) and (5), respectively and calorimetric  $\Delta_{vap}H_m$  values.

The enthalpies of vaporization obtained both by direct and indirect methods at the saturated vapor pressure, were recalculated to the standard values by means of correction  $\delta(\Delta H_m) = p \cdot \{T \cdot (dB_v / dT) - B_v\}$ . The reliability of the calculated  $\Delta_{vap}H_m$  values were proved by their agreement with the calorimetric ones within the error limits (Table 2). Due to smaller extrapolation intervals, the errors of the enthalpies of vaporization at the normal boiling temperatures are less than  $\Delta_{vap}H_m(298.15K)$  values. Extrapolation capabilities of equations (3) and (5) were verified by comparison of calculated  $\Delta_{vap}H_m$  values at  $T = 298.15$  K with experimental ones for some well studied alkanes and alkanethiols (Boublik et al., 1984). It has been shown that these equations allowed us to estimate the  $\Delta_{vap}H_m$  values with uncertainties  $\leq 2$  per cent in extrapolation intervals  $\Delta T \leq 50$  K.

Mutual congruence of some thermodynamic properties in set of related compounds (Table 2) can be drawn from comparison of these properties in dependence on some physico-chemical characteristics having influence upon intermolecular interactions in liquid state.

Fig. 3 represent critical temperatures,  $T_c$ , normal boiling temperatures,  $T_{n.b.}$ , and enthalpies of vaporization,  $\Delta_{vap}H_m(298.15K)$ , for freons and chloroalkanes  $C_2$ ,  $C_3$  depending on the dipole moments,  $\mu(liq)$ , and coefficients of molecular packing,  $K_m$ , in the liquids. The



Compounds	$T_{n.b.}$ K	$\Delta_{vap}H_m^0(298.15K)$ (calor) kJ·mol <sup>-1</sup>	$\Delta_{vap}H_m^0(298.15K)$ ( <i>p-T</i> ) kJ·mol <sup>-1</sup>	$\Delta_{vap}H_m^0(T_{n.b.})$ ( <i>p-T</i> ) kJ·mol <sup>-1</sup>	
Freons and halogen -ethanes and -propanes					
CFCl <sub>2</sub> CFCl <sub>2</sub>	R-112	366.00±0.01	-	34.98±0.39	31.26±0.33
CF <sub>2</sub> ClCFCl <sub>2</sub>	R-113	320.76±0.01	28.61±0.09	28.71±0.41	27.21±0.44
CF <sub>2</sub> ClCF <sub>2</sub> Cl	R-114	276.65	-	-	24.9±±1.1
CF <sub>2</sub> BrCF <sub>2</sub> Br	R-114b2	320.36±0.01	28.61±0.09	28.63±0.34	27.11±0.32
CF <sub>2</sub> ClCHCl <sub>2</sub>	R-122	345.01±0.01	32.91±0.09	32.84±0.40	29.84±0.36
CFCl <sub>2</sub> CHFCI	R-122a	346.34±0.01	33.10±0.06	33.04±0.36	29.95±0.35
CF <sub>3</sub> CHCl <sub>2</sub>	R-123	300.981±0.008	-	26.35±0.33	26.14±0.34
CF <sub>2</sub> ClCHFCI	R-123a	303.02±0.01	-	26.82±0.28	26.45±0.32
CF <sub>3</sub> CHClBr	R-123b1	323.41±0.01	29.80±0.09	30.01±0.33	28.36±0.31
CF <sub>3</sub> CH <sub>2</sub> CH <sub>2</sub> Cl	R-253fa	318.84±0.01	-	29.86±0.39	28.39±0.40
CF <sub>3</sub> CH <sub>2</sub> CHCl <sub>2</sub>	R-243	345.47±0.01	34.05±0.04	34.40±0.57	31.14±0.47
CF <sub>3</sub> CH <sub>2</sub> CCl <sub>3</sub>	R-233	368.28±0.01	36.76±0.08	37.4±1.4	32.62±0.57
CF <sub>3</sub> CH <sub>2</sub> CFCl <sub>2</sub>	R-234fb	333.74±0.01	-	33.59±0.65	29.80±0.62
CHCl <sub>2</sub> CH <sub>3</sub>	1,1-DCIE	330.35±0.01	30.62±0.14	31.12±0.34	29.24±0.35
CH <sub>2</sub> ClCH <sub>2</sub> Cl	1,2-DCIE	356.61±0.01	32.15±0.01	35.36±0.39	32.18±0.35
CH <sub>2</sub> BrCH <sub>2</sub> Br	1,2-DBrE	404.55±0.01	41.73±0.02	41.97±0.95	36.23±0.51
CHCl <sub>2</sub> CH <sub>2</sub> Cl	1,1,2-TCIE	386.98±0.01	40.28±0.10	40.23±0.56	35.09±0.42
CH <sub>3</sub> CCl <sub>3</sub>	1,1,1-TCIE	347.21±0.01	32.62±0.09	32.58±0.35	29.89±0.34
CHCl <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	1,1-DCIP	361.53±0.01	35.10±0.11	35.34±0.45	31.85±0.35
CH <sub>2</sub> ClCHClCH <sub>3</sub>	1,2-DCIP	369.50±0.01	36.20±0.08	36.37±0.49	32.45±0.41
CH <sub>2</sub> ClCH <sub>2</sub> CH <sub>2</sub> Cl	1,3-DCIP	393.95±0.01	40.75±0.04	41.18±0.55	35.56±0.38
CH <sub>3</sub> CCl <sub>2</sub> CH <sub>3</sub>	2,2-DCIP	342.67±0.01	-	32.22±0.38	29.66±0.37
CH <sub>3</sub> CFI <sub>3</sub>	R-143a	225.85±0.01	-	-	19.40±0.24

Table 2. Normal boiling temperatures,  $T_{n.b.}$ , molar enthalpies of vaporization,  $\Delta_{vap}H_m(298.15K)$ , measured calorimetrically and calculated from  $pT$  data at  $T = 298.15$  K and  $T_{n.b.}$  for some freons and halogen-alkanes (Varushchenko et al., 2007; Majer & Svoboda, 1985; Boublik et al., 1984)

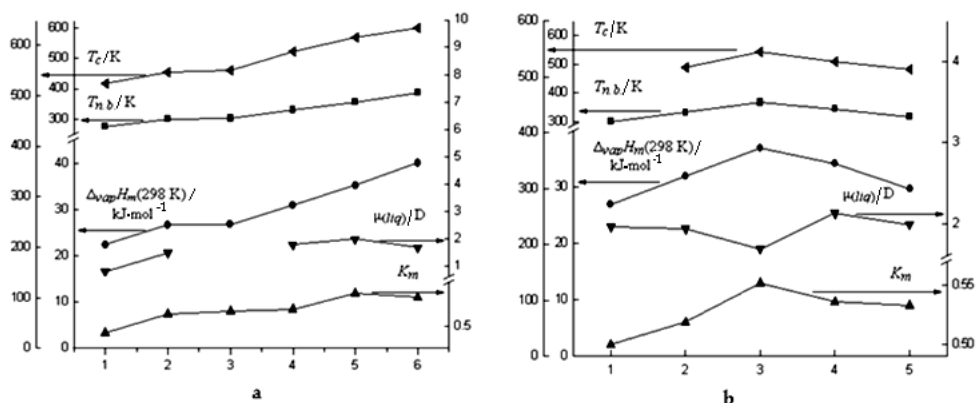


Fig. 3. Variations of thermodynamic properties  $T_c$ ,  $T_{n.b.}$  and  $\Delta_{vap}H_m(298.15K)$  in dependence on the dipole moments,  $\mu_{(liq)}$ , and coefficients of molecular packing,  $K_m$ , in series of liquid halogenated ethanes (a):  $CF_3CHCl_2$  [1],  $CF_2ClCH_2Cl$  [2],  $CF_2ClCFCl_2$  [3],  $CH_3CHCl_2$  [4],  $CH_2ClCH_2Cl$  [5],  $CHCl_2CH_2Cl$  [6]; and propanes (b):  $CF_3CH_2CF_2Cl$  [1],  $CF_3CH_2CFCl_2$  [2],  $CF_3CH_2CCl_3$  [3],  $CF_3CH_2CHCl_2$  [4],  $CF_3CH_2CH_2Cl$  [5]

coefficients  $K_m$  were calculated by analogy with (Varushchenko et al., 2007). In spite of the large atomic weight of fluorine in comparison with hydrogen, thermodynamic values of compounds decrease when hydrogen is substituted for fluorine that can be explained by decreasing of the  $\mu_{(liq)}$  and  $K_m$  parameters. Minimum  $T_c$ ,  $T_{n.b.}$ , and  $\Delta_{vap}H_m(298.15K)$  values are inherent to completely halogenated 1,1,1-trifluoro-2,2-dichloroethane, which has the lowest values of dipole moment and  $K_m$  coefficient. Maximum values of corresponding properties are observed for the most polar compounds, 1,1,2-trichloroethane, the *gauche* conformer of which is stabilized by the dipole interaction in the liquid phase.

Analysis of the data shown in Fig. 3 allows to conclude that the values of critical and normal boiling temperatures and enthalpy of vaporization vary in a series of compounds according to the combined action of the parameters responsible for intermolecular interactions and short range order of the liquid phase, thus proving the mutual consistency of the thermodynamic data in the series of halogenated ethane and propane.

**Cyclic perfluorocarbons and hydrocarbons.** A thermodynamic study of perfluorated cyclic organic compounds has scientific and practical importance. Perfluorocarbons [PFC] have high chemical and thermal stability, absolute biological inertness, and weak intermolecular interactions [IMI]. The combination of these properties can be assigned to high C-F bond strength and the shielding effect of fluorine atoms towards the carbon framework. The weakness of IMI is responsible for the ability of PFC to dissolve and transfer considerable amounts of gases, in particular, oxygen and carbon dioxide. On account of these properties, PFC have found wide application in biology and medicine as efficient gas-transfer media (blood substitutes).

Compounds	$T_{n.b.}$ K	$\Delta_{vap}H_m^0(298.15K)$ (calor), kJ·mol <sup>-1</sup>	$\Delta_{vap}H_m^0(298.15K)$ (p-T), kJ·mol <sup>-1</sup>	$\Delta_{vap}H_m^0(T_{n.b.})$ (p-T), kJ·mol <sup>-1</sup>	$\varphi(O_2)$ (298.15), cm <sup>3</sup> /100 ml
Bicyclic perfluorocarbons					
<i>cis</i> -C <sub>9</sub> F <sub>16</sub>	391.52±0.01	41.98±0.14	41.95±0.52	34.43±0.37	44.91
<i>trans</i> -C <sub>9</sub> F <sub>16</sub>	389.02±0.01	41.34±0.05	41.22±0.51	34.11±0.36	46.11
<i>cis</i> -C <sub>10</sub> F <sub>18</sub>	416.96±0.01	46.19±0.12	46.79±0.62	36.57±0.43	40.30
<i>trans</i> -C <sub>10</sub> F <sub>18</sub>	414.70±0.01	45.40±0.08	46.02±0.60	36.26±0.43	41.10
C <sub>5</sub> F <sub>10</sub> N-C <sub>6</sub> F <sub>10</sub> - CF <sub>3</sub>	460.74±0.01	56.56±0.24	56.58±0.88	40.68±0.44	34.30
Bicyclic hydrocarbons					
<i>cis</i> -C <sub>9</sub> H <sub>16</sub> <sup>a</sup>	440.99±0.01	-	46.34±0.82	38.01±0.44	25.22
<i>trans</i> -C <sub>9</sub> H <sub>16</sub> <sup>a</sup>	434.22±0.01	-	44.88±0.78	37.21±0.43	28.84
<i>cis</i> -C <sub>10</sub> H <sub>18</sub> <sup>a</sup>	468.93±0.01	-	50.90±0.94	40.46±0.45	21.25
<i>trans</i> -C <sub>10</sub> H <sub>18</sub> <sup>a</sup>	460.43±0.01	-	48.45±0.78	39.29±0.43	25.05
Alkylderivatives of adamantane					
1,3,5-TMA	483.31±0.01	51.74±0.20	51.50±0.52	40.52±0.50	-
1,3-DMA <sup>a</sup>	476.441	49.71±0.20	49.47±0.54	39.71±0.41	-
1-EA	498.86±0.01	54.96±0.28	54.6±1.3	42.56±0.51	-

<sup>a</sup>  $T_{n.b.}$  and  $\Delta_{vap}H_m(298.15K)$  values were calculated from literature data of (Boublik et. al., 1984).

Table 3. Normal boiling temperatures,  $T_{n.b.}$ , molar enthalpies of vaporization,  $\Delta_{vap}H_m(298.15K)$ , obtained by direct and indirect methods, and oxygen capacities,  $\varphi(O_2)$  (298.15K), for bicyclic hydrocarbons, perfluorocarbons, and derivatives of adamantane

The saturated vapor pressure of bicyclic PFC at temperature (310 K) of the human body,  $p_s^{310}$ , is one of the key properties of the blood substitute, which ranges from 0.16 to 2.66 kPa. A stability of an aqueous emulsion of fluorocarbon and its delivery rate from the body depends on  $p_s^{310}$  value. Medicine employs bicyclic perfluorocarbon composition with high and low vapor pressures. Perfluoro-N-(4-methylcyclohexyl)piperidine, C<sub>5</sub>F<sub>10</sub>N-C<sub>6</sub>F<sub>10</sub>-CF<sub>3</sub>, having the low value of  $p_s^{310}$  = 0.157 kPa, is a component of "Ftorosan" (Russia) blood substitute in mixture with *cis*- and *trans*- perfluorodecalines, which have higher (1.54 and 1.72) kPa values of  $p_s^{310}$ , respectively. Another key property of the blood substitutes is an oxygen capacity,  $\varphi(O_2)$  (cm<sup>3</sup>/100 ml), which is defined as a volume of oxygen, dissolved in 100 ml of the liquid. The  $\varphi(O_2)$  values were evaluated on the basis of the enthalpies of vaporization by empirical method developed within a theory of regular solutions (Lawson et al., 1978).

Table 3 presents derived thermodynamic values of cyclic compounds. The values of the normal boiling temperatures and the enthalpies of vaporization of *cis*-isomers are more than

those of *trans*-isomers in the series of perfluorobicyclo-nonanes and -decanes and their hydrocarbons analogues. Despite the more molecular mass, the normal boiling temperatures  $T_{n.b.}$  and the  $\Delta_{vap}H_m$  values of the perfluorocarbons are less than those of the hydrocarbons. On the contrary, the oxygen capacities are two times more in the series of perfluorocarbons which can be explained by more poor intermolecular interactions of PFC. Fig. 4 presents the critical temperatures, enthalpies of vaporization, and oxygen capacities,  $\varphi(O_2)$ , for *cis*- and *trans*- perfluorobicyclo(4,3,0)nonanes (1 and 2), for components of Ftorosan blood substitute (Ries, 1991), namely perfluorobicyclo(4,4,0)-decanes (3 and 4), perfluoro-N-(4-methylcyclohexyl)piperidine (5), and for some of their hydrocarbon analogues (6-9), respectively.

Due to smaller energies of intermolecular interactions, the critical temperatures and enthalpies of vaporization of perfluorocarbons are less, but oxygen capacities are more, than appropriate properties of appropriate hydrocarbons.

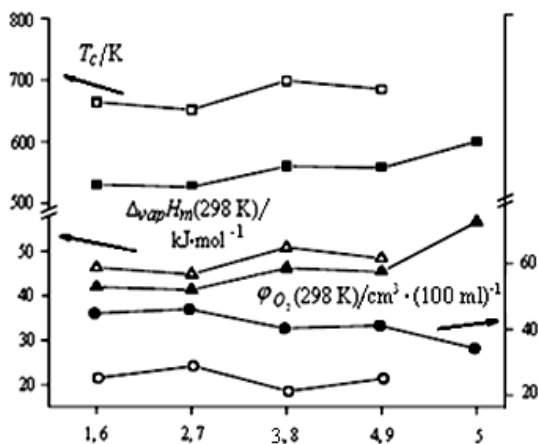


Fig. 4. Critical temperatures,  $T_c$ , enthalpies of vaporization,  $\Delta_{vap}H_m(298.15K)$ , and oxygen capacities,  $\varphi(O_2)(298.15K)$ , for perfluorinated compounds 1-5 (■ ▲ ●) and their hydrocarbon analogues 6-9 (□ △ ○)

Despite the more molecular mass, the normal boiling temperatures and enthalpies of vaporization of perfluorocarbons are less than those of appropriate hydrocarbon. This can be explained less coefficients of molecular packing,  $K_m$ , and therefore by more intermolecular distances, and as a consequence less intermolecular interactions of perfluorocarbons in comparison with their hydrogen - containing counterparts.

### 2.3 The vapor pressure and enthalpies of vaporization of the hard-volatile compounds

The saturation vapor pressures of the solid and liquid substances having  $p < 1$  kPa were determined by a dynamic method of evaporating the sample in a stream of the carrier inert gas. In calculation of the enthalpy of vaporization, the volume of vapor is well described by the ideal gas law and the volume of liquid can be easily neglected without introducing

essential error into the  $\Delta_{vap}H_m$  value. But the  $dP/dT$  or  $d\ln(p)/dT$  derivatives are determined not enough reliably because the saturation vapor pressure is a weak function of the temperature. Thus, an accuracy of determination of the enthalpy of vaporization is restricted for the compounds with low vapor pressures at about 298 K temperature.

The temperature dependences of the vapor pressures for the ferrocene derivatives [FD] were determined by a transpiration method elaborated and fully described by Verevkin S.P. and coauthors (Emel'yanenko et al., 2007). Here, only the main features of the method are given. The determination of the vapor pressure is based on the measurements of the mass of substance transpired in the stream of carrier gas (nitrogen) and the volume of the gas flowing. The vapor pressure of the substances was obtained by Dalton law for partial vapor pressures of the ideal gas mixture. A sample of the substance (~0.5 g) was placed into the U-tube, temperature of which was controlled with accuracy  $\pm 0.1$  K. A nitrogen flow, controlled by a precision Hoke valve and measured with a bubble gauge, was passed through the tube. The transferred substance was condensed in a cooled trap and was analyzed chromatographically using the external standard (hydrocarbons). The rate of the nitrogen flow was adjusted to ensure that the condensed and vapor phases were in stable equilibrium. The saturation vapor pressure  $p_{sat}$  was calculated by the formula:

$$p_{sat} = mRT / VM \quad (8)$$

where  $V = V(N_2) + V(DF)$ ;  $R = 8.314472$ ;  $m$  and  $M$  are the mass of the sample under study and molecular weight of FD, respectively;  $V(N_2)$  and  $V(DF)$  are the nitrogen and FD volumes, respectively,  $V(N_2) > V(DF)$ ; and  $T$  is the U-tube temperature. The  $V(N_2)$  value was determined from the flow rate and the measurement time.

The  $pT$  parameters of the solid FD were measured in the pressure and temperature intervals from (0.01/0.11 to 0.44/4.9) Pa and from (311/342 to 341/379) K, respectively. Appropriate pressure and temperature intervals for the liquid FD were from (0.3/1.87 to 7.88/130) Pa and from (298/384 to 358/430) K, respectively. The vapor pressures of FD were approximated by equation:

$$R \cdot \ln(p) = a + b/T + \Delta C_{p,m} \cdot \ln(T/T_{st}) \quad (9)$$

where  $a$  and  $b$  are coefficients,  $\Delta C_{p,m} = C_{p,m}(g) - C_{p,m}(cr/liq)$  is the difference between the heat capacities of the vapor and condensed phases, and  $T_{st} = 298.15$  K is the standard temperature (arbitrarily chosen). Equation (9) was deduced by integration of the correlation  $-R \cdot [(d\ln(p))/d(1/T)] = \Delta_{vap}H_{m,T} + \Delta C_{p,m}(T - T_{st})$  (Kulikov et al., 2001). The latter was obtained on the basis of Clausius-Clapeyron equation  $R \cdot [(d\ln(p))/d(1/T)] = \Delta_{vap}H_m$  in approximation of  $\Delta_{liq}^{\infty} V_m \approx R \cdot T/p$  and derivative  $d(\Delta_{vap}H_m)/dT = \Delta C_{p,m}$  by assuming that  $\Delta_{liq}^{\infty} C_{p,m}$  value is independent on the temperature in the  $pT$  interval under study. The enthalpy of vaporization was calculated by the formula:

$$\Delta_{vap}H_m(\Delta_{sub}H_m) = -b + \Delta C_{p,m} \cdot T \quad (10)$$

obtained by differentiation of equation (9) with respect to  $1/T$ . The ideal gas heat capacities of the ferrocene derivatives [FD] were obtained by additive Chickos and Acree method (Chikos & Acree Jr., 2003) that is defined as "an atom together with all of its ligands".

Table 4 lists the purity of ferrocene derivatives determined by adiabatic calorimetry (part 3.2), coefficients  $a$  and  $b$  of equations (9) and (10), and enthalpies and entropies of vaporization and sublimation of FD at  $T = 298.15$  K.

Compounds	Purity, mol. %	$a$	$-b$	$\Delta_{vap}H_m^O(T)$	$\Delta_{sub}H_m^O(T)$	$\Delta_{vap}S_m^O(T)$	$\Delta_{sub}S_m^O(T)$
				kJ·mol <sup>-1</sup>		J·K <sup>-1</sup> ·mol <sup>-1</sup>	
FM	97.56 <sup>a</sup> ;	339.7	111826.0		102.8±2.0		344.8±6.4
	99.0 <sup>b</sup>	359.5	115237.0	86.97±1.7		291.7±5.7	
BF	99.47 <sup>a</sup> ;	359.9	121314.1		109.3±2.0		366.6±6.0
	99.7 <sup>b</sup>	373.9	124422.2	90.64±1.8		304.0±5.8	
BOF	99.6 <sup>b</sup>	364.8	133682.1		119.9±2.4		402.1±6.3
		382.9	133533.0	98.2±2.0		329.4±5.2	
POF	99.24 <sup>a</sup> ;	353.7	112812.1	80.8±1.6	99.1±2.8 <sup>c</sup>	271.0±5.0	332.4±6.0
	99.0 <sup>b</sup>						
<i>n</i> -PF	98.93 <sup>a</sup> ;	320.5	95312.0	69.2±1.4		232.1±4.7	
	99.0 <sup>b</sup>						
<i>i</i> -BF	99.41 <sup>a</sup> ;	326.6	98138.8	70.7±1.5		237.1±4.9	
	99.0 <sup>b</sup>						

<sup>a</sup> Adiabatic calorimetry; <sup>b</sup> DSC; <sup>c</sup> the value calculated on the basis of correlation  $\Delta_{sub}H_m = \Delta_{vap}H_m + \Delta_{fus}H_m$ .

Table 4. The purity, coefficients of equations (9) and (10), enthalpies and entropies of vaporization,  $\Delta_{vap}H_m^O(T)$  and  $\Delta_{vap}S_m^O(T)$ , and sublimation,  $\Delta_{sub}H_m^O(T)$  and  $\Delta_{sub}S_m^O(T)$  of ferrocenylmethanol [FM], benzylferrocene [BF], benzoylferrocene [BOF], propionylferrocene [POF], *n*-propylferrocene [*n*-PF], *iso*-butylferrocene [*i*-BF] at  $T = 298.15$  K

For testing the uncertainties of transpiration method in applying to FD, a compilation of the literature data on the enthalpy of sublimation of ferrocene were carried in reference (Emel'yanenko et al., 2007). A series of 21  $\Delta_{sub}H_m^O(298.15K)$  values ranged from (70.3±1.0 to 76.78±0.85) kJ·mol<sup>-1</sup> was obtained, the most part of the data being focused in the range between (72 and 74) kJ·mol<sup>-1</sup>. Uncertainties of these quantities were probably the random errors. Taking into account the uncertainties of the initial vapor pressure data making up from (1.5 to 2) per cent, a total value of the random and systematic errors could be  $\geq 2$  %. Therefore, the errors of the enthalpies of vaporization and sublimation as derivative values of the vapor pressure in the transpiration method were evaluated as  $\pm \geq 2$  %.

### 3. The heat capacity and thermodynamic properties of the phase transitions

A heat capacity is a capability of the substance for absorbing some quantity of the energy that increases its temperature by 1 degree K. A measurement of the heat capacity is performed by adiabatic and isothermal methods. The first one allows attaining the most complete thermodynamic equilibrium or, in any case, the thermal balance in the calorimetric system. The adiabatic method is used for exploring the thermal processes with different times of relaxation and the metastable phases which can exist in wide temperature ranges. The heat capacities and thermodynamic properties of the phase transitions were

investigated in this work by low-temperature adiabatic calorimetry (Varushchenko et al., 1997a).

### 3.1 Experimental

The measurements of the heat capacities were conducted in a fully automated setup, consisted of a vacuum adiabatic calorimeter, a data acquisition and control system, AK-9.02, and a personal computer, PC (Fig. 5). The setup was produced in the National Scientific and

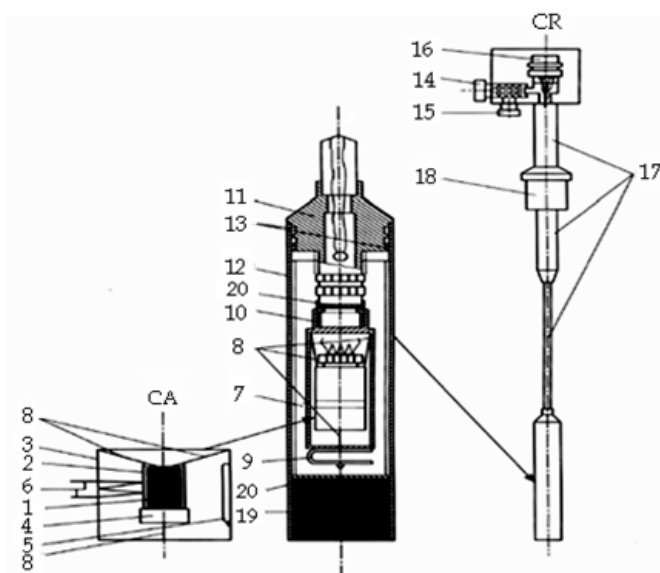


Fig. 5. The vacuum adiabatic calorimeter (CA) and cryostat (CR): (1) titanium container ( $V \sim 1 \text{ cm}^3$ ); (2) copper sleeve with the heater of calorimeter; (3) adiabatic shield; (4) bronze lid of the container; (5) the rhodium-iron resistance thermometer; (6) four-junction battery of Cu/Fe - Chromel thermocouples; (7) radiation screen - aluminium-coated Dacron-like film; (8) nylon threads; (9) spring; (10) Teflon tube; (11) plug; (12) vacuum jacket; (13) grooves of the plug; (14) valve; (15) and (16) detachable vacuum and cable joists, respectively; (17) steel tubes; (18) coupling nut; (19) charcoal getter; (20) radiation screens.

Research Institute of Physical Technical and Radio-Technical Measurements (Mendeleev, Moscow Region). The main principles of its construction were published in (Pavese & Malyshev, 1994).

The calorimetric cell consists of a container, 1, a copper sleeve, 2, in which the container is tightly held, and an adiabatic shield, 3. A bronze brass lid, 4, serves for vacuum-tight sealing the container by means of indium gasket and a simple manifold. To decrease the heat capacity of the empty calorimeter, the miniature rhodium-iron resistance thermometer, 5, ( $R_0 = 50 \Omega$ ) was mounted on the inner surface of the adiabatic shield. The thermometer, which was calibrated on ITS-90, is destined for temperature measurements from (0.5 to 373) K with accuracy  $\pm 3 \cdot 10^{-3}$  K. The temperature difference between the calorimeter and the

adiabatic shield is measured by a four-junction thermocouple, 6, (Cu + 0.1 per cent Fe alloy against Chromel), one end of which was mounted on the copper sleeve 2 and the other one was placed on the inner surface of the adiabatic shield, 3. A manganin calorimeter heater ( $R = 300 \Omega$ ) was wound non-inductively on the sleeve, 2. A well-known three-lead circuit diagram was employed for wiring the current and potential leads of the heater. Since the resistances of the current leads are equal, this diagram enables us to account for the heat generated in the leads between the calorimeter and the shield. To reduce the level of heat radiation, the shield was wrapped with several layers of aluminium-coated Lavsan film, 7, (ACLF, an analog of Mylar). The container sleeve, 2, is suspended inside the adiabatic shield on three nylon threads, 8, which are stretched by a spring, 9 (Fig. 5). The calorimeter cell has been fixed on an epoxy/fibre-glass tube, 10, of the cryostat, CR. The tube, 10, is fastened to a copper plug, 11, by means of a bayonet joint. The only removable part of the calorimeter cell is the container for the specimen.

The vacuum jacket, 12, is made from oxygen-free copper. The vacuum seal of the cryostat is provided by a KPT-8 silicon/boron nitride paste, which has high thermal conductivity value and gives a stable vacuum junction after freezing. The paste is put between the upper part of the jacket, 12, and the plug, 11, in its grooves, 13.

The top part of the cryostat (CR) has a valve, 14, detachable vacuum, 15, and cable, 16, joints; the latter connects the electrical leads of the calorimeter cell to AK-9.02 and PC. Both parts of the cryostat are jointed by the stainless steel tubes, 17. Due to small size ( $l = 120$  mm,  $d = 22.5$  mm), the cryostat is immersed directly into a commercial transportation Dewar vessels. This allows us to exclude an intermediate Dewar vessel and, thus, to reserve the coolants. A coupling nut, 18, with a Teflon shell and a rubber ring is used to fasten the cryostat airtight inside the neck of the Dewar vessel. A T-connection, fitted on the neck of the nitrogen Dewar vessel, enables us to pump out nitrogen vapors to lower the bath temperature if necessary.

The calorimeter cell is cooled down by thermal conductivity via electrical leads and by radiation heat transfer. The leads of the thermometer, heaters, and differential thermocouple form a heat shunt with the preset thermal resistance and they provide cooling of the calorimeter from room temperature to approximately  $T = 78$  K, and from  $T = 78$  K down to  $T = 5$  K for about 7 h in each Dewar vessel. The helium heat-exchange gas is not used for this purpose in order to avoid problems, connected with its desorption. To reduce the heat losses by radiation, the additional radiation screens, 20, are used (Fig. 5).

The data acquisition system AK-9.02 is a single unit, connected with a personal computer [PC]. The system AK-9.02 and the PC perform the measurements of all values that are necessary for the determination of the heat capacity, as well as the control of the measurement process and data processing.

The thermometer resistance and the calorimeter power heating are measured by a potentiometer method with cyclic inversion of the direction of thermometer current for excluding the thermal electromotive forces. All the procedures that control the measurement process are carried out by the PC, which has a simple and user-friendly interface. The results of the measurements are printed and displayed on the screen for visual monitoring.

An adiabatic condition in calorimeter is maintained by the AK-9.02 system, which allows keeping the temperature drop between the container and shield on the average within  $\pm(1-3) \cdot 10^{-3}$  K. Owing to modification of the calorimeter, the drop of temperature was reduced to  $\sim 0.5$  mK at the expense of using an eleven-junctions thermocouple instead of



four - junction one and employing an additional heater ( $R \sim 133 \Omega$ ) mounted in the upper part of the shield, to which electrical wires of the thermometer and the main heater were connected. Additional heater allows making up a lack of the second adiabatic shield that is usually employed in the adiabatic calorimeters, but cannot be placed in our miniature device. Due to small size, the cryostat with the calorimeter was placed in the transport Dewar vessels with refrigerants (liquid helium or nitrogen), that allows us to exclude an intermediate Dewar vessel and, thus, to keep the coolants. There is no constant pumping of the cryostat during the operation, since high vacuum inside the cryostat was kept by means of cry-sorption provided with an efficient charcoal getter. The degree of vacuum in cryostat is controlled by the value of the heater current in the adiabatic shield. This value was determined in a process of the calorimeter production using nitrogen and helium baths. The automatic procedure of the heat capacity measurements is performed by AK - 9.02 system running under PC control (Pvsev & Malishev, 1994). The program realizes a method of the discrete input of the energy in two modes: constant increments of temperature,  $\Delta T$  (from 1 to 2) K during measurement of the heat capacity and constant impulses of energy in studying the phase transitions.

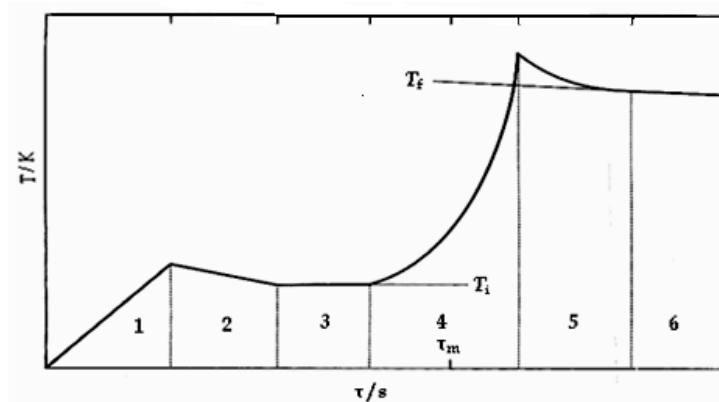


Fig. 6. Temperature,  $T$ , against time,  $\tau$ , curve in a heat capacity measurement. 1 to 6 are periods of the calorimetric measurement;  $T_i$  and  $T_f$  are an initial and final temperatures of the calorimeter in 4<sup>th</sup> main (heating) period;  $\tau_m$  is the midpoint.

The calorimetric experiment consists of six periods (Fig. 6). In the first period the calorimeter is heated to a desired temperature. A steady temperature equilibration is attained in the second period. In the third period the temperature of the calorimeter is monitored over a chosen time interval to acquire information about the temperature drift rate,  $V_i$  and to obtain the linear relation between the values  $V_i$  and the time by the least-squares method. During the fourth (heating) period the electrical energy is supplied to the calorimeter, and the heating-up time is observed. The fifth period is the same as the second one. In the sixth period the linear relation between the temperature drift rate of the calorimeter  $V_f$  and the time is established exactly in a similar manner to that in the third period. The initial and the final temperatures of the calorimeter in the main (heating) period are calculated by

extrapolating the linear dependencies of the drift rates  $V_i$  and  $V_f$  on time to the midpoint ( $\tau_m$ ) temperature (Fig. 6). This method permits the heat interchange between the calorimeter and surroundings to be taken into account (Varushchenko et al., 1997a). The reliability of this method was proved by a congruence within (0.1 to 0.2) per cent of the heat capacity values of an empty calorimeter, measured in the temperature interval (90 to 110) K using different refrigerants: liquid helium and nitrogen.

The values of the heat capacities,  $C_{s,m}$ , are fitted with polynomials:

$$C_{s,m} = \sum_i A_i \{(T - A_k) / B_k\}^i \quad (11)$$

$$C_{s,m} = \sum_i A'_i \ln\{(T - A'_k) / B'_k\}^i \quad (12)$$

where  $\{(T - A_k) / B_k\}$  and  $\{(T - A'_k) / B'_k\}$  are normalizing term,  $i$  is degree of polynomials.

The coefficients of the polynomials  $A_i$  and  $A'_i$  were estimated by the LSM.

The metrological characteristics of the calorimeter were tested by measuring the heat capacity of pure copper having a mass fraction of 0.99995 and *n*-heptane in the temperature intervals (from 8 to 372) K and (from 6 to 354) K, respectively. Obtained  $C_{s,m}$  values of copper and *n*-heptane came to an agreement with the precise heat capacities of standard substances within (0.2 to 1.4) % below the temperature 70 K and decrease to (0.01 and 0.3) % above  $T=70$  K.

### 3.2 Determination of thermodynamic properties of the phase transitions

The important characteristics of the substances: a triple point temperature,  $T_{tp}$ , and a mole fraction of impurities,  $N_2$ , were determined by calorimetric method of the fractional melting study, developed by Mair, Glasgow and Rossini. A linear dependence between the reciprocal fractions of the sample melted,  $1 / F_1$ , and the equilibrium fusion temperatures,  $T_1$ , makes it possible to calculate both the  $T_{tp}$  value and mole fraction of impurities,  $N_2$ , by equations:

$$T_1 = d + f \cdot (1 / F_1), \quad (13)$$

$$N_2 = (\Delta_{fus} H_m \cdot f) / (R \cdot T_{tp}^2) \quad (14)$$

Here  $d = T_0$  is the triple point temperature,  $T_{tp}$ , of the pure compound,  $f = (T_1 - T_0)$ , denotes a depression of the  $T_{tp}$  value caused by impurities,  $T_1$  is the triple point temperature of the completely melted substance (for  $F_1 = 1$ ),  $R = 8.314472$  J K<sup>-1</sup> mol<sup>-1</sup>, and  $\Delta_{fus} H_m$  is the enthalpy of fusion, determined by independent method.

Concave curve of the dependence  $T_i = f(1/F_i)$  can be explained by formation of solid solution of impurities with main substance. The efficient coefficient of impurities distribution between the crystal and liquid phases  $k$  is calculated by Mastrangelo's and Dornte's method. The  $k$  value close to zero, proves an absence of the solid solution. According to (Alexandrov, 1975), melting curves can be concave not only in the case of solid solutions, but also in the absence of equilibrium in the calorimeter at the onset of fusion, when the amount of the liquid phase is small and impurities can therefore be distributed no uniformly, and at the final stage of melting, when sedimentation of crystals to the bottom of container interferes with slow attainment of temperature equilibrium. In conformity with Alexandrov recommendation (Alexandrov, 1975), the  $T_{tp}$  and  $N_2$  values ought to be estimated on the basis of the linear dependence for the part of melting curve in the range of  $1/F_i$  values from 1.2 to about 8-10.

In the case of solid solution formation, the mole fraction of impurity,  $N_2$ , can be determined by the Smit and Alexandrov method using  $T_i = f(1/F_i)$  experimental data and the equation for  $N_2$  in a binary system with solid solution formation (Van Wijk & Smit, 1960) and (Alexandrov et al., 1983):

$$T_i = T_0 - (N_2 / A_k) \cdot [(1-k) / F_i^{1-k}] \quad (15)$$

Here,  $A_k$  is cryoscopic constant for the major substance; and  $k$  is a distribution coefficient of impurities between the solid and liquid phases. An insufficiency of this equation for calculating  $N_2$  consists in the need to determine  $k$  coefficient by an independent method. By differentiating and finding the logarithm, equation (15) was transformed by (Alexandrov et al., 1983) to the form:

$$\ln\{-dT_i / d(1/F_i)\} = \ln\{N_2 \cdot (1-k)^2 / A_k\} - k \cdot \ln(1/F_i) \quad (16)$$

Equation (16) is used in this work for determination of the  $k$  coefficient directly from experimental dependence of  $T_i = f(1/F_i)$  and for calculation of the mole fraction,  $N_2$ , of impurity. The  $k$  and  $N_2$  values were calculated by least-squares fits of the  $T_i$  and  $F_i$  experimental data using the linear equation (16). Mole fraction of impurity,  $N_2$ , was computed from the  $\ln\{N_2 \cdot (1-k)^2 / A_k\}$  term using the  $k$  value and the cryoscopic constant  $A_k = \Delta_{fus}H_m^0 / RT_0^2$  determined by above mentioned Rossini method. The enthalpy of fusion  $\Delta_{fus}H_m$  is determined by calorimetric method using the total enthalpy absorbed during the fusion with following subtraction of the normal heat capacities of the crystal and liquid which the substance has in the fusion region. The  $\Delta_{fus}H_m$  value is obtained from the equation:

$$\Delta_{fus}H_m = \Delta_{tot}H - \Delta H_1 - \Delta H_2 - \Delta H_3 \quad (17)$$

where  $\Delta_{tot}H$  is the total enthalpy absorbed in heating the calorimeter from initial temperature  $T_1 < T_{tp}$  to final one  $T_2 > T_{tp}$ ;  $\Delta H_1$  and  $\Delta H_2$  are the changes of enthalpy

calculated from the normal heat capacities of the crystal and liquid in the temperature intervals from  $T_1$  to  $T_{tp}$  and from  $T_{tp}$  to  $T_2$ , respectively;  $\Delta H_{emp}$  is the enthalpy increment need for heating the empty calorimeter from  $T_1$  to  $T_2$ .

### 3.3 Crystal phase transitions and molecular dynamics

The solid state transitions revealed in the molecular crystals can be explained by different polymorphous transformations, caused by changing the crystal structure, different location of the molecules and their orientational and conformational disorder in the crystal lattice. In this chapter, some thermodynamic properties of solid state transitions and fusion are reviewed for some compounds, which were studied in the Luginin's Thermochemistry Laboratory of the Moscow State University and in some other thermodynamic Laboratories. An interpretation of the solid-state transitions in organic crystals was successfully fulfilled in a set of outstanding researcher's works (Westrum & McCullough, 1965; Kolesov, 1995; Adachi, et al., 1968) and the others. An interpretation of calorimetric measurements was carried out very often on the basis of the order – disorder concept. Understanding these processes requires sometimes exploring the molecular crystals by X-ray crystallography and IR and Raman spectroscopy. In this work, the solid state transitions will be discussed including some additional physico-chemical properties of the compounds.

The values of thermodynamic properties of the phase transitions are given in Table 5. It was found by exploring IR-spectra of  $\text{CF}_2\text{ClCFCl}_2$ , that this substance comprises a mixture of *trans*- and *gauche*- conformers in solid (crystal I) and liquid states (Fig. 7). The sum  $\Delta_{trs}S_m^0 + \Delta_{fus}S_m^0 = 20.52 \text{ J}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$  for  $\text{CF}_2\text{ClCFCl}_2$  is small, while such sum used to be from (42 to 50)  $\text{J}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$  for organic crystals. By comparing the calorimetric and spectroscopic  $S_m^0$  values, it was found that  $\text{CF}_2\text{ClCFCl}_2$  has residual entropy,  $S(0) = 10.1 \text{ J}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$  at  $T = 198.15 \text{ K}$  (Higgins & Lielmers, 1965; Kolesov, 1995).

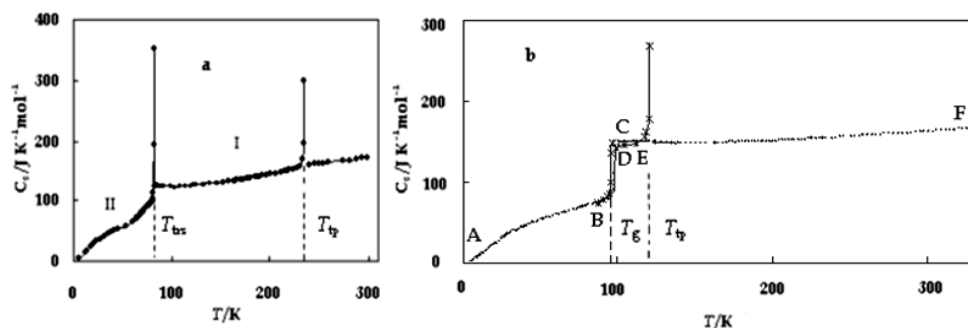


Fig. 7. Heat capacities and phase transitions of  $\text{CF}_2\text{ClCFCl}_2$  (a) and  $\text{CF}_2\text{ClCHCl}_2$  (b). Freon  $\text{CF}_2\text{ClCHCl}_2$  was in following states: glass (AB) and supercooled liquid (CE) ( $\alpha = 0$ ); partially crystalline state (BD) and supercooled liquid (DE) ( $\alpha = 0.076$ ); and liquid (EF)

Compounds	$T_{trs}$	$T_{tp}$	$\Delta_{trs}H_m^0$	$\Delta_{fus}H_m^0$	$\Delta_{trs}S_m^0$	$\Delta_{fus}S_m^0$
	K		kJ·mol <sup>-1</sup>		J·K <sup>-1</sup> ·mol <sup>-1</sup>	
Freons						
CF <sub>2</sub> ClCH <sub>2</sub> Cl	86.5±0.1 <sup>a</sup>	135.74±0.19	-	7.05±0.13	-	51.94±0.71
CF <sub>2</sub> ClCF <sub>2</sub> Cl <sup>b</sup>	109.3±0.2 134.6±0.1	180.62±0.02	1.21±0.01 2.63±0.01	1.51±0.01	11.1±0.1 19.54±0.04	8.36±0.04
CF <sub>2</sub> ClCFCl <sub>2</sub> <sup>b</sup>	82.5±0.5	236.92±0.02	0.83±0.01	2.47±0.01	10.1±0.2	10.42±0.04
Derivatives of ferrocene <sup>c</sup>						
<i>n</i> -PF	186.7±0.3	276.83±0.50	1.23±0.01	12.74±0.17	6.59±0.06	46.02±0.61
<i>i</i> -BTF	207.4±0.3 <sup>a</sup>	294.68±0.40	-	15.05±0.08	-	51.07±0.27
BeF	44.3±0.2 51.1±0.3	349.57±0.10	23.8±0.2 46.1±0.4	25.54±0.94	0.54±0.01 0.83±0.01	73.06±2.70
POF	-	311.62±0.51	-	19.16±0.12	-	61.48±0.39
<i>i</i> -BF	-	279.96±0.10	-	15.33±0.04	-	54.76±0.14
BeOF <sup>d</sup>	-	380.7	-	29.9	-	78.5
FM	-	347.80±1.0	-	22.91±0.53	-	65.87±1.52
Alkylderivatives of adamantane						
1,3,5-TMA	234.4±0.1	255.61±0.05	8.19±0.03	2.06±0.03	34.94±0.13	8.06±0.11
1,3-DMA	223.38±0.01	247.79±0.01	9.31±0.01	1.54±0.01	41.68±0.06	6.21±0.02
1-EA	-	225.56±0.02	-	11.22±0.03	-	49.74±0.08
Bicyclic hydrocarbons						
<i>cis</i> -C <sub>9</sub> H <sub>16</sub>	182.28 184.9	236.48	8.26±0.004 0.40±0.002	1.40±0.004	45.3 2.2	5.92
<i>trans</i> -C <sub>9</sub> H <sub>16</sub>	-	213.86	-	10.90±0.02	-	50.97
<i>cis</i> -C <sub>10</sub> H <sub>18</sub>	216.1	230.18±0.05	2.136	9.489±0.006	9.88	41.22
<i>trans</i> -C <sub>10</sub> H <sub>18</sub>	-	242.78±0.05	-	14.414±0.001	-	59.37
Bicyclic perfluorocarbons						
<i>cis</i> -C <sub>9</sub> F <sub>16</sub>	200.64±0.03 245.63±0.04	291.27±0.09	8.76±0.03 1.27	2.72±0.03	43.66±0.04 5.17	9.34±0.09
<i>trans</i> -C <sub>9</sub> F <sub>16</sub>	236.63±0.02	248.05±0.04	8.90	2.63±0.02	37.6	10.60±0.10
<i>cis</i> -C <sub>10</sub> F <sub>18</sub>	232.5±0.2	266.95±0.02	4.24±0.01	10.30±0.01	18.24±0.06	38.58±0.15
<i>trans</i> -C <sub>10</sub> F <sub>18</sub>	-	294.83±0.02	-	17.96±0.04	-	60.92±0.14
C <sub>5</sub> F <sub>10</sub> N-C <sub>6</sub> F <sub>10</sub> -CF <sub>3</sub>	-	293.26±0.20	-	8.32±0.02	-	28.37±0.06

<sup>a</sup> Glass like transition,  $T_g$ ; <sup>b</sup>(Kolesov, 1995) <sup>c</sup> *n*-propylferrocene [*n*-PF], *iso*-butyrylferrocene [*i*-BTF], benzylferrocene [BeF], propionylferrocene [POF], *iso*-butylferrocene [*i*-BF], benzoylferrocene [BeOF], ferrocenylmethanol [FM]; <sup>d</sup> values were measured by DSC.

Table 5. The temperatures,  $T_{trs}$ ,  $T_{tp}$ , enthalpies,  $\Delta_{trs}H_m^0$ ,  $\Delta_{fus}H_m^0$ , and entropies,

$\Delta_{trs}S_m^0$ ,  $\Delta_{fus}S_m^0$  of the solid-to-solid transitions and fusions, respectively, for some freons, ferrocene derivatives, cyclic hydrocarbons, and perfluorocarbons.

A characteristic feature of solid state transition of organic crystals is a slow thermal equilibrium between co-existing phases which very often promote to formation of metastable phases existing in a wide temperature range. In Fig. 7(b), the heat capacity  $C_{S,m}$  of 1,1-difluoro-1,2,2-trichloroethane,  $\text{CF}_2\text{ClCHCl}_2$ , is shown in the temperature interval studied. Similar  $C_{S,m} = f(T)$  dependence has been obtained for isomer of 1,1-difluoro-1,2,2-trichloroethane:  $\text{CFCl}_2\text{CHFCl}$ . Both isomers were in the forms of glasses, supercooled liquids, and partially crystalline states. The latter was attained after annealing the specimen at temperatures from (110 to 114) K during 3 days, followed by quenching it at  $T \approx 78$  K over a period of 12 h. The heat capacity jumps, accompanying G-transitions, are observed on the  $C_S - T$  curves of both freons. Taking into account this typical transitions for the glasses, the authors of reference (Adachi, et al., 1968) proposed a term “glassy crystal” for the frozen - in disordered states (AB) (Fig.7, (b)). The temperatures of the glass transition  $T_g = 95.7$  K and fusion,  $T_{fus} = 123.1 \pm 0.4$  K have been obtained. The degrees of crystallinity,  $\alpha$ , appropriated to the mole fraction of crystalline samples, equal to 0.076 and 0.116 for  $\text{CF}_2\text{ClCHCl}_2$  and  $\text{CFCl}_2\text{CHFCl}$ , respectively, were calculated on the basis of calorimetric data on  $\Delta C_S$  jumps by studying the G-transitions (Varushchenko et al., 1997b).

Fig. 8 presents the  $C_S - T$  curve of ferrocenyl-*n*-propane, which exhibits a fusion and a gradual solid-to-solid transition in the temperature range from (156 to 204) K. The temperature of the gradual transition of crystal II to crystal I of ferrocenyl-*n*-propane was ascribed to that of the maximum  $C_S$ -value in the peak of solid state transition. A test of the calorimetric experiment showed that  $\lambda$ -anomaly was accompanied by decreasing the heat capacity,  $\Delta C_S = -2.7 \text{ J}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$  and continuance changing the enthalpy and entropy. Thus, the solid-state anomaly is the phase transition of the second order and can be interpreted as the “order-disorder” transformation. The changes of the enthalpy,  $\Delta_{trs}H_m^0$ , and entropy,  $\Delta_{trs}S_m^0$  of the thermal anomaly were evaluated by summing up these values in each experimental  $C_S$  point with subtracting changes of appropriate functions for the empty calorimeter and those ones for the hypothetical normal parts of the crystals I and II.

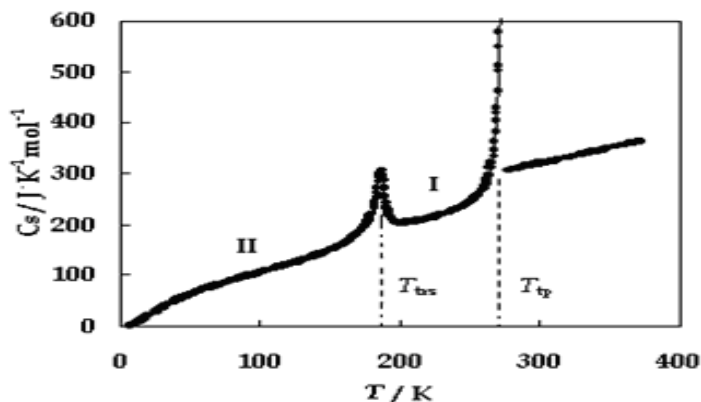


Fig. 8. Molar heat capacity,  $C_S$ , of ferrocenyl-*n*-propane as a function of temperature,  $T$ , where  $T_{tp}$ , and  $T_{trs}$  denote the triple point and temperature of  $\lambda$ -like transition.

The nature of this transition was studied by the X-ray crystallography. Table 6 lists crystallographic data of *n*-propylferrocene in vicinity of the "order-disorder" transition. The structure of *n*-propylferrocene at 200 K contains a propyl- group disordered between two positions (in ca. 2:1 ratio) obviously due to thermal motion (Fig. 9). The transition of the crystal II ( $T = 150$  K) to the crystal I ( $T = 200$  K) occurred with significant changes of the lattice parameters: basis vectors,  $a$ ,  $b$ ,  $c$ , angle,  $\beta$ , the volume,  $V$ , the number of molecule in the unit cell,  $Z$ , and the factors of the crystal structure quality,  $R_1$  and  $wR_2$ .

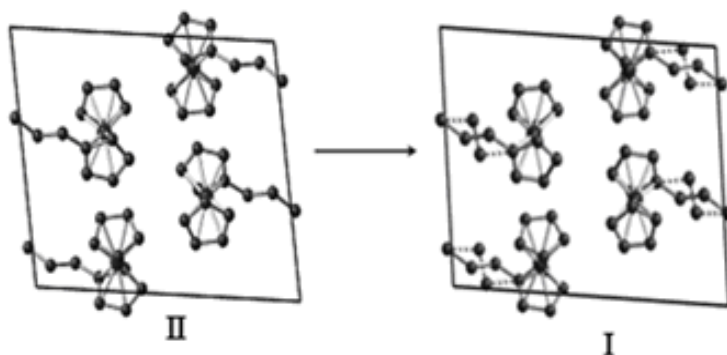


Fig. 9. The structure of the ferrocenyl-*n*-propane crystal at the temperatures 150 K (II) and 200 K (I)

$T/K$	symmetry, space group	$10^{-10}a / \text{m}$	$10^{-10}b / \text{m}$	$10^{-10}c / \text{m}$	$\beta / ^\circ$	$10^{-30}V / \text{m}^3$	$Z$	$R_1$	$wR_2$
150	monoclinic	13.738 (4)	7.541(1)	10.660(2)	99.76 (3)	1088.4(4)	4	0.061	0.112
200	$P2_1/c$	13.977 (4)	7.621(1)	10.521(2)	96.87 (3)	1112.6(4)		0.064	0.136

Table 6. Crystallographic data of *n*-propylferrocene crystal at the temperatures 150 K and 200 K

While both  $a$  and  $b$  parameters became larger during phase transition II to I, the  $c$  parameter and  $\beta$  angle slightly decreased (Table 6). Structure solution of the crystals I and II revealed that lower temperature modification of the *n*-propylferrocene molecule has only one orientation of the propyl group. Apparently, the transitions in reverse order occurred when cooling the crystals from (150 to 200) K. Thus, the solid-state anomaly of *n*-propylferrocene is caused by the onset of the internal rotation of propyl groups in the molecules and also by a small shift of the pentadienyl cycles around the axis passed through their centers. These variations led to some orientation disorder of the crystal phase II (Fig. 9).

The order-disorder conception is successfully used in exploring the plastic crystals. Adamantane and some of its derivatives form disordered plastic crystalline phases. The fusion of such substances occurs some times in two stages. First, an orientational disorder proceeds in the crystalline lattice because of high mobility of the molecules, and then the plastic crystals fuse owing to a translational molecular motion at higher temperature. In this

case, the magnitudes of enthalpy and entropy of the solid-to-solid transition are several times larger than those of the fusion. In accordance with the empirical Timmerman's criterion, the  $\Delta_{fus}S_m^0$  values for the plastic crystals are usually less than  $20 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ . There are two modifications of the plastic crystals with different molecular reorientations, isotropic and anisotropic. Some substances are known to form both of reorientations. Fig. 10 presents  $C_s - T$  curve of 1,3,5-trimethyladamantane (1,3,5-TMA) explored in this work (a) and the Raman spectrum of the substance (b). A spectroscopic investigation of 1,3,5-TMA was carried out together with that of 1,3-dimethyladamantane (1,3-DMA), which also formed plastic crystals. Both compounds have low values of the entropy of fusion,  $\Delta_{fus}S_m^0 = 8.1$  (6.2)  $\text{J}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$  and narrow temperature intervals of existence of the high temperature crystal I,  $\Delta T = 21$  (24.2) K, respectively. According to adiabatic calorimetry, these properties are typical for the plastic crystals. Distinct bands were observed in their spectra at the low temperatures. As it follows from joint spectra discussion, the number of bands for 1,3,5-TMA is about half of that for the 1,3-DMA crystal. This suggests that more symmetric molecules of the former compound compose the lattice of higher symmetry and/or with the less quantity of units in a primitive cell. In the vicinity of the phase transitions, the changes in spectra become more pronounced. At the transition points, all the bands disappear and transform into the wing of a broad Rayleigh scattering. This implies that all of them were caused by the librational modes in the low-temperature crystals of 1,3-DMA and 1,3,5-TMA and that the character of molecular motion is different in high-temperature solid phases. The absence of the preferential axes of molecular reorientation in the latter implies that there are isotropic plastic crystals. Analysis of Raman

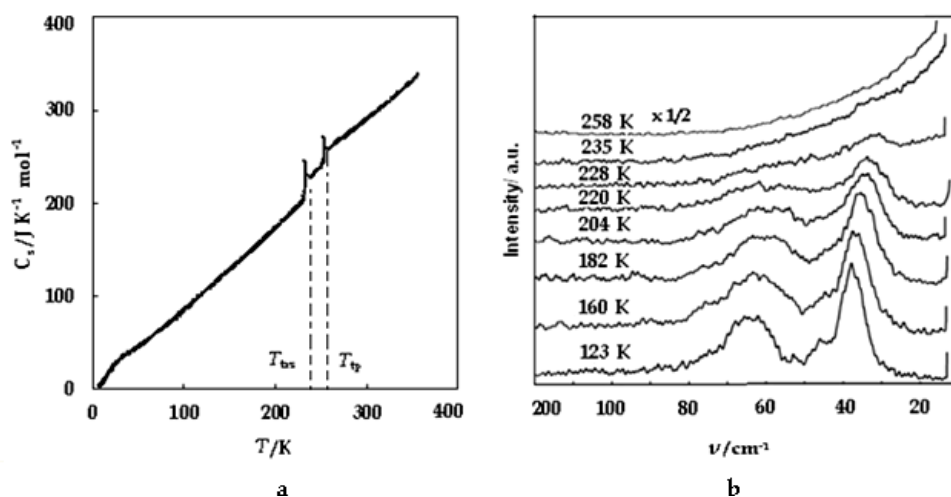


Fig. 10. (a) The heat capacity of 1,3,5-trimethyladamantane in dependence on temperature,  $T$ , where  $T_{tp}$ , and  $T_{trs}$  denote temperatures of the triple point and the solid-to-solid transition. (b) The temperature dependence of low-frequency Raman spectrum of 1,3,5-trimethyladamantane.



spectra is in agreement with the low values of transition entropies measured by adiabatic calorimetry for alkyl derivatives of adamantane. The *cis*- and *trans*- isomers of perfluorobicyclo(4,3,0)nonane are almost spherical "globular" molecules, which are able to form the plastic crystals due to unusually high molecular mobility. The thermodynamic properties of compounds are known to change when going from fluoroorganic compounds to their hydrogen containing analogous. With this in mind, the thermodynamic properties of the solid state transitions of perfluorocarbons are compared with those of appropriate hydrocarbons.

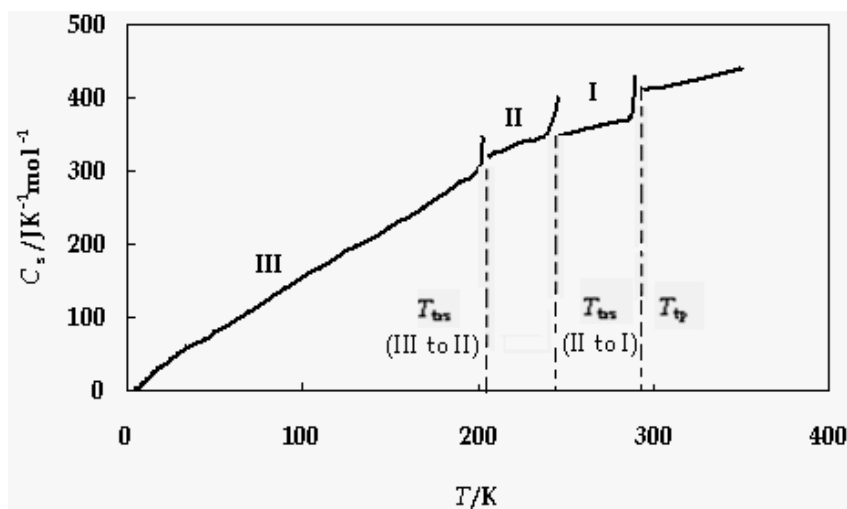


Fig. 11. Molar heat capacity,  $C_s$ , of *cis*-perfluorobicyclo(4,3,0)nonane as function of temperature,  $T$ , where  $T_{trs}$  (III to II) and  $T_{trs}$  (II to I) denote the temperatures of the solid-to-solid transitions and  $T_{tp}$  is the temperature of triple point

Fig. 11 presents a temperature dependence of the heat capacity of *cis*-perfluorobicyclo(4,3,0)nonane. Two solid states transitions can be interpreted as the order-disorder transformation caused by orientation disorder. Taking into account the high entropy values of  $\Delta_{trs}S_m^0 = 43.66 \pm 0.04 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$  (Table 5) for transition of the crystal III-to- crystal II, this solid phase conversion can be attributed to an anisotropic molecular reorientation about preferential common C-C axis. According to the empirical Timmerman's criterion, small values of the fusion entropy,  $\Delta_{fus}S_m^0 = 9.34 \pm 0.09 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$  of *cis*- $\text{C}_9\text{F}_{16}$  (Table 5) indicates the onset of the isotropic molecular reorientation and, thus the formation of the plastic crystals. As is seen from Table 5, the *cis*-bicyclo(4,3,0)nonane has thermodynamic properties of the solid-state transition, analogous to its perfluorated counterpart, and forms plastic crystals. The *trans*-perfluorobicyclo(4,3,0)nonane exhibits one solid-state transition and also forms the plastic crystals, but its hydrogen analogous, *trans*-bicyclo(4,3,0)nonane do not form the plastic crystals. These properties of the *trans*-isomers can be explained by their less spherical shapes as compared with the *cis*-isomers.

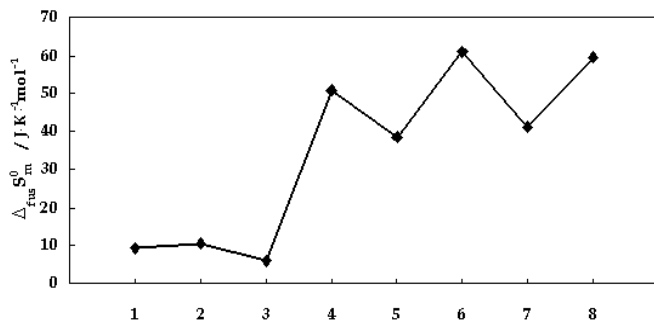


Fig. 12. Molar entropies of fusion  $\Delta_{fus} S_m^0$  for some bicyclic perfluorocarbons and appropriate hydrocarbons: *cis*- [1] and *trans*- [2] perfluorobicyclo(4,3,0)nonanes; *cis*- [3] and *trans*- [4] bicyclo (4,3,0) nonanes; *cis*- [5] and *trans*- [6] perfluorobicyclo(4,4,0)decanes; *cis*- [7] and *trans*- [8] bicyclo(4,4,0)decanes

Analysis of the entropies of fusion for *cis*- and *trans*-isomers of perfluorobicyclo (4,3,0)nonanes and -(4,4,0) decanes and appropriate values of hydrogen - containing analogous (Table 5, Fig. 12) permits to interpret the influence of the structure and chemical nature on molecular mobility and thermodynamic properties of the solids. Comparing the entropies of fusion shows that the mobility of the molecules increases in going from hydrocarbons to appropriate perfluorocarbons and from the *cis*-isomers to the *trans*-ones. The larger molecular mobility of the perfluorocarbons can be explained by more weak intermolecular interactions for these compounds compared with the hydrocarbons. The greater ability of the molecules of *cis*-isomer to reorient in the solid state seems to be due to the steric factors. The nature of the solid-to-solid transitions in *cis*- [7] and *trans*- [8] bicyclodecane and *cis*- [5] and *trans*- [6] perfluorobicyclodecane (Fig. 12) were discussed in the order-disorder concept in reference (Kolesov, 1995).

#### 4. Thermodynamic functions in the ideal gas states

The absolute entropy, changes of the enthalpy and Gibbs energy in three aggregate states are calculated on the basis of smoothed heat capacity values. The experimental data on the heat capacities of the substances under study in the temperature intervals from (6/8 to 372) K were fitted by polynomial (11). Extending the heat capacities to  $T \rightarrow 0$  were performed by Debye equation:

$$C_{s,m} = nD(\Theta/T) \quad (18)$$

where  $D$  is the Debye function,  $n = 3$ , and  $\Theta$  denote Debye characteristic temperature.

Testing the  $C_{s,m}$  values at  $T \rightarrow 0$  was performed by fitting the heat capacities in small temperature interval below (10 to 12) K by equation:

$$C_{s,m}/T = \alpha T^2 + \gamma, \quad (19)$$

where  $\alpha$  and  $\gamma$  are coefficients. If  $\gamma = 0$ , it can be accepted that  $C_{s,m} = 0$  and extrapolation of the heat capacity to  $T \rightarrow 0$  can be carried out by equation (18) or by Debye cube's law,

$C_{s,m} = AT^3$ . But, in the case  $\gamma \neq 0$ , appropriate extending of the heat capacity could be conducted by (18), provided the parameters  $n$  and  $\Theta$  are adjusted thus to allow one to obtain zero  $C_{s,m}$  values at  $T = 0$  K. The  $C_{p,m} - C_{s,m} = T(\partial V / \partial T)_p(\partial p / \partial T)_s$  difference evaluated at  $T = 298.15$  K was smaller than uncertainties of  $C_{s,m}$  for the substances under study and was not taken into account. The smoothed values of  $C_{p,m}$  and thermodynamic functions  $S_m^0(T)$ ,  $\{H_m^0(T) - H_m^0(0)\}$ , and  $\{G_m^0(T) - H_m^0(0)\}$  for the condensed states were calculated by numerical integrating the  $C_{p,m} = f(T)$  functions obtained by equations (18) and (11) and adding the enthalpies and entropies of the solid-to-solid transition and fusion. The errors of thermodynamic functions were estimated by the law of random errors accumulation using the uncertainties of the heat capacity measurements. The ideal gas absolute entropy,  $S_m^0(T)$ , the changes of the enthalpy and the free Gibbs energy at 298.15 K were calculated using the appropriate functions in the liquid state, enthalpies and entropy of vaporization and the entropy of the ideal gas compression,  $\Delta S = R \cdot \ln\{p(T) / (101.325)\text{kPa}\}$  calculated from the vapor pressure data.

#### 4.1 Theoretical calculations of the thermodynamic functions

The ideal gas absolute entropy and heat capacity were calculated by statistical thermodynamics, additive principle (Poling et al., 2001; Domalski et al., 1993; Sabbe et al., 2008), and empirical difference method of group equations (Cohen & Benson, 1993).

The statistical thermodynamic method was used with quantum mechanical [QM] calculation on the basis of the density functional theory [DFT]. The QM calculation was performed on the level B3LYP/6-31G(d,p) using the Gaussian 98 and 03 software packages (Frisch et al., 2003). As a result, the following constants can be calculated: the moments of inertia of the entire molecule, the moments of inertia for internal rotors, the normal vibrational frequencies, and the barrier to internal rotation. The potential functions of internal rotation were determined by scanning the torsion angles from (0 to 360)° at 10° increments and allowing all other structural parameters to be optimized at the same level with the subsequent frequency calculation. The calculated potential energies were fitted to the cosine-based Fourier series:

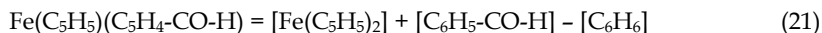
$$V(\varphi) = V_0 + 0.5 \sum_n V_n (1 - \cos n\varphi), \quad (20)$$

where  $V(\varphi)$  denotes potential energy function,  $\varphi$  is torsional angle.

The ideal gas entropies and heat capacities in dependence on the temperature were calculated by standard statistical thermodynamics formulae using the rigid-rotor harmonic oscillator [RRHO] approximation. To account for the internal rotation processes, the torsional frequencies were omitted in the calculation of thermodynamic function. A contribution of the internal rotation for each rotor was calculated by direct summation over the energy levels obtained by diagonalization of the one-dimensional Hamiltonian matrix associated with potential function from equation (20). The RRHO approximation, is known, results in overestimated entropy values for flexible molecules due to coupling the internal rotations. One-dimensional hindered rotor correction has been applied by (Vansteenkiste et al., 2003; Van Speybroeck et al., 2000) assuming decoupled internal rotations.

The method of group equation is suitable for calculation of some additive properties of a compound, namely  $S_m^0(T)$ ,  $C_{p,m}^0(T)$ ,  $\Delta_f H_m^0$  values on the basis of reliable appropriate

properties of the structurally similar compounds. For example, determination of  $S_m^{\circ}(T)$  value of ferrocenylmethanol using absolute entropies of the components of hypothetical reaction:



The calculation of the absolute entropy by additive methods requires taking into account corrections for the symmetry and the optical isomerism of the molecule. Otherwise, the principle of group additivity can be broken if these parameters alter when changing the structure of the molecules in the series of compounds. Therefore, an additive calculation of the absolute entropy,  $S_m^0$ , is conducted by using so-called intrinsic entropy,  $S_{m,\text{int}}^0$ , which allows to exclude an influence of the rotary components, depending on the symmetry and optical activity of the molecule:

$$S_{m,\text{int}}^0 = S_m^0 + R \cdot \ln(\sigma_{\text{tot}} / n) \quad (22)$$

where  $R \cdot \ln(\sigma_{\text{tot}} / n)$  is the correction for the symmetry and the chirality of the molecule,  $\sigma_{\text{tot}}$  and  $n$  denote a total number of the symmetry and the number of the optical isomers. The total symmetry of the molecule is calculated as  $\sigma_{\text{tot}} = \sigma_{\text{ex}} \cdot \sigma_{\text{in}}$ , where  $\sigma_{\text{ex}}$  is the number of symmetry of the external rotation of the molecule, as a whole, and  $\sigma_{\text{in}}$  denotes the number of the internal rotation equal to the product ( $P$ ) of the order of independent axes in the rotating group ( $l$ ) raised to the power of the number of these axes ( $k$ ),  $\sigma_{\text{in}} = P \cdot l^k$ .

Thus, an additive calculation of the absolute entropy by group equation method requires of a missing intrinsic entropy  $S_{m,\text{int}}^0$  determination and following recalculation it to the  $S_m^0$  value by formula (20).

Comparing between the values of absolute entropy determined on the basis of calorimetric measurements,  $S_m^0(g)(\text{expt})$ , group equation method, and statistical thermodynamic calculations,  $S_m^0(\text{stat})$ , are applied for prediction of the missing data, verification of their reliability, and mutual congruence in the series of the same type compounds, or homologous. Below, a critical analysis of the ideal gas entropies is performed for some series of functional organic compounds.

**Freons and chloroalkanes.** Table 7 lists the ideal gas absolute entropies determined on the basis of experimental data and those ones obtained by theoretical calculation for some freons [R] and chloroethanes [CIE]. The  $S_m^0(g)(\text{expt})$  and  $S_m^0(\text{stat})$  values for halogenethanes and freons, including  $S_m^0(g)$  values with marks (*a* and *b*), agree within errors limits: from (0.1 to 1.3) %. These compounds have rather compact and symmetric molecules and exist as an equilibrium mixture of *trans*- and *gauche*- conformers in three aggregate states and as single more stable conformers in the low-temperature crystal phase. There are disagreements between  $S_m^0(g)(\text{expt})$  and  $S_m^0(\text{stat})$  values in a group of three freons  $\text{CF}_2\text{ClCFCl}_2$  [R-113],  $\text{CFCl}_2\text{CHFCl}$  [R-122a], and  $\text{CF}_2\text{ClCHCl}_2$  [R-122]. A characteristic feature of these compounds is availability of residual entropies caused by orientational or conformational disorders. The mixtures of the *trans*- and two *gauche*- conformers of these freons were probably frozen at low helium temperatures. The entropy change caused by disorder of this type can be evaluated by the formula  $\Delta S = R \ln(N_2 / N_1)$ , where  $N_2$  and

$N_i$  are the number of states of disordered and ordered phases. Thus, the residual entropy of  $\text{CFCl}_2\text{CHFCl}$  and  $\text{CF}_2\text{ClCHCl}_2$  is  $S_m^0(0) = R \cdot \ln(3) = 9.1 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ . The value  $S_m^0(0) = 10.1 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$  of  $\text{CF}_2\text{ClCFCl}_2$ , obtained in (Higgins & Lielmers, 1965), is in good agreement with the  $R \cdot \ln(3)$  value. After taking into account residual entropies of these freons, the calorimetric and theoretical values of absolute entropies agree within errors limits from (0.1 to 1.3) %. At the same time, testing the low-temperature  $C_{p,m}$  values of  $\text{CFCl}_2\text{CHFCl}$  and  $\text{CF}_2\text{ClCHCl}_2$  by equation (19) showed an absence of the residual entropies for these freons. Their heat capacities in the temperature interval from (5 to 8) K obeyed Debye cub's law with experimental error of  $C_{p,m}$  value, 2 %. A disagreement of two methods evaluation of the residual entropy can not be explained on the basis of available physico-chemical data of these freons.

Compounds	R-112	R-113	R-114	R-122	R-122a	R-123	R-123a	R-243	R-253fa	1,1-DCIE	1,2-DCIE	1,1,1-TCIE
$S_m^0(g)$ ( <i>expt</i> ), $\text{J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$	382.3	378.9 382.0 <sup>a</sup>	363.8	363.1 372.2 <sup>b</sup>	366.2 375.4 <sup>b</sup>	352.4	365.6	394.7	365.3	304.4	307.8	321.0
$S_m^0(stat)$ $\text{J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$	385.7	386.9	364.2	372.5	378.6	352.6	368.1	399.1	367.1	305.2	306.0	320.1

<sup>a,b</sup> Absolute entropies were calculated as  $S_m^0(g)(expt) + S_m^0(0)$ , were  $S_m^0(0) = 10.1 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$  and  $S_m^0(0) = 9.1 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ .

Table 7. The absolute entropies at  $T = 298.15 \text{ K}$ , calculated by statistical thermodynamic,  $S_m^0(stat)$ , and determined from the calorimetric data on the heat capacity, entropy of vaporization, and the entropy of the ideal gas compression,  $S_m^0(g)(expt)$  for some freons and chloroethanes

Compounds	$\text{C}_9\text{F}_{16}$		$\text{C}_9\text{H}_{16}$		$\text{C}_{10}\text{F}_{18}$		$\text{C}_{10}\text{H}_{18}$		1,3-DMA	1,3,5-TMA	1-EA
	<i>cis</i> -	<i>trans</i> -	<i>cis</i> -	<i>trans</i> -	<i>cis</i> -	<i>cis</i> -	<i>trans</i> -				
$S_m^0(g)(expt)$ , $\text{J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$	605.3 $\pm 3.0$	602.8 $\pm 3.0$	374.20	365.39	630.7 $\pm 0.3$	378.86	374.43	392.9 $\pm 1.0$	415.2 $\pm 1.0$	400.9 $\pm 1.0$	
$S_m^0(stat)$ , $\text{J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$	610.83	606.41	377.32 <sup>a</sup>	365.30 <sup>a</sup>	633.33	378.81	373.81	391.4	413.4	400.0	

<sup>a</sup> Values were taken from (Frenkel et al., 1994).

Table 8. The absolute entropies at  $T = 298.15 \text{ K}$ , calculated by statistical thermodynamics,  $S_m^0(stat)$ , and determined from the calorimetric data on the heat capacity, entropy of vaporization, and the entropy of the ideal gas compression,  $S_m^0(g)(expt)$  for *cis*- and *trans*-perfluorobicyclo(4,3,0)nonanes [*cis*- and *trans*-  $\text{C}_9\text{F}_{16}$ ], *cis*- and *trans*- bicyclo(4,3,0)nonane, [*cis*- and *trans*-  $\text{C}_9\text{H}_{16}$ ], *cis*-perfluorobicyclo(4,4,0)decane, [*cis*- $\text{C}_{10}\text{F}_{18}$ ], *cis*- and *trans*-bicyclo(4,4,0)decanes [*cis*- and *trans*-  $\text{C}_{10}\text{H}_{18}$ ]; 1,3-dimethyladamahtane [1,3-DMA], 1,3,5-trimethyladamahtane [1,3,5-TMA] and 1-ethyladamahtane [1-EA]

**Cyclic hydrocarbons and perfluorocarbons.** The absolute entropies obtained by the third thermodynamic law from the experimental data and statistical thermodynamics for some cyclic and bicyclic hydrocarbons and perfluorocarbons are listed in Table 8. The values of the absolute entropies determined by independent methods agree within errors limits from (0.1 to 1) % that proves their mutual conformity and also the reliability of the all experimental data used for computing the value of  $S_m^0(g)(expt)$ , namely the heat capacities, enthalpies of vaporization, and saturation vapor pressures.

**Ferrocene derivatives [FD].** The experimental data on thermodynamics of the phase equilibriums for FD are very scarce. Therefore, critical analysis of the available data on thermodynamic properties and using them for science prognosis of failing properties are the urgent problems. A verification of reliability and mutual congruence of the properties of some derivatives of ferrocene were carried out on the basis of absolute entropies, which were computed by the third thermodynamic law from the experimental data and by statistical thermodynamics and group equation method. Comprising the  $S_m^0$  and  $C_{p,m}^0$  values determined by both theoretical methods showed that they agree within errors limits,  $\leq 1.2$  %. The values of absolute entropies in homologous series of alkyl- and acyl-ferrocenes obtained in this work and available in the literature are presented in Table 9.

Compound	$S_m^0(cond)$ <sup>a</sup>	$S_m^0(g)(expt)$ <sup>a</sup>	$S_m^0(g)(calc)$ <sup>b</sup>	$S_m^0(g)(recom)$ <sup>c</sup>
F (cr)	211.85 ± 1.1 <sup>d</sup>	361.0 ± 3.6	362.6 ± 7	362 ± 3.1
EF(liq)	(354.1 ± 1.1) <sup>e</sup>	(525.3 ± 2.9) <sup>e</sup>	453 ± 7 <sup>d</sup>	453 ± 7
<i>n</i> -PF (liq)	356.9 ± 0.9	494.3 ± 5.0	489 ± 7	492 ± 6
<i>i</i> -BF (liq)	377.6 ± 1.9	517.3 ± 5.2	513 ± 7	515 ± 6
<i>n</i> -BF (liq)	398.1 ± 1.2 <sup>e</sup>	(586.3 ± 3.4) <sup>e</sup>	527 ± 7	527 ± 7
FF (cr)	241.3 ± 0.7 <sup>e</sup>	(398.1 ± 2.8) <sup>e</sup>	435 ± 7 <sup>d</sup>	435 ± 7
AF (cr)	264.6 ± 0.8	478.3 ± 2.9	-	478 ± 3
POF (cr)	291.0 ± 1.2	502.7 ± 5.1	513 ± 7	507 ± 6
<i>i</i> -BTF (liq)	379.6 ± 1.2	-	541 ± 7	541 ± 7

<sup>a</sup>  $S_m^0(cond)$  and  $S_m^0(g)(expt)$  denote the entropies determined based on experimental data;

<sup>b</sup>  $S_m^0(g)(calc)$  are values calculated by DFT and group equation methods; <sup>c</sup>  $S_m^0(g)(recom)$  denotes recommended values; <sup>d</sup> (Emel'yanenko et al., 2010); <sup>e</sup> (Karyakin, et al., 2003).

Table 9. The absolute entropies ( $J \cdot K^{-1} \cdot mol^{-1}$ ) in condensed and ideal gas states of ferrocene, [F], and some derivatives of ferrocene: ethylferrocene [EF], *n*-propylferrocene [*n*-PF], *iso*-butylferrocene [*i*-BF], *n*-butylferrocene [*n*-BF], formylferrocene [FF], acetylferrocene [AF], propionylferrocene [POF], *iso*-butyrylferrocene [*i*-BTF] at  $T = 298.15$  K

The  $S_m^0(g)(expt)$  and  $S_m^0(g)(calc)$  values for ferrocene, *n*-propyl- and *iso*-butylferrocene agree to within their errors that proves the reliability of these values and all the thermodynamic data used for their calculation. The mutual consistency of the absolute entropies of *iso*-butyrylferrocene and *iso*-butylferrocene in the liquid state was revealed using the Benson additive method. The difference between the entropies of these compounds,  $\Delta S = 2 J \cdot K^{-1} \cdot mol^{-1}$ , is close to the entropy increment  $\Delta' S = 1.4 J \cdot K^{-1} \cdot mol^{-1}$ , which fit for replacing the  $CH_2$  by CO group in the passage from *iso*-butyl- to *iso*-butyrylferrocene. The difference between the  $S_m^0(g)(expt)$  and  $S_m^0(g)(calc)$  values for crystal propionylferrocene, which exceeds measurement errors, can be explained by uncertainties

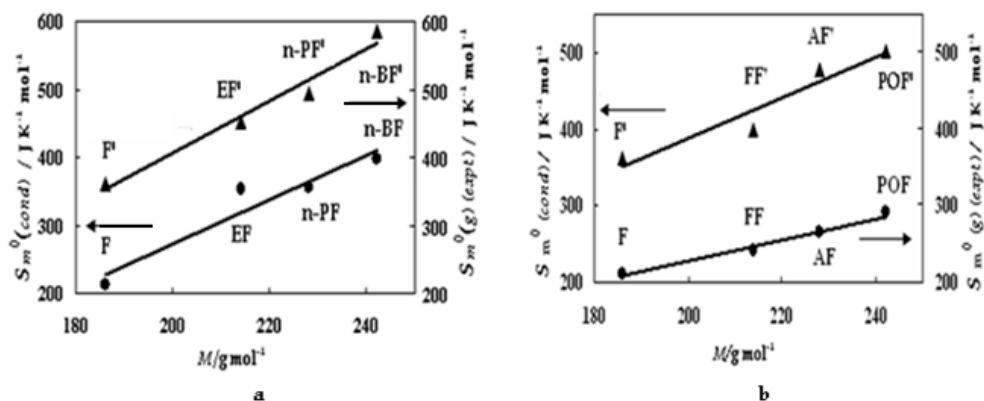


Fig. 13. Absolute entropies ( $\text{J K}^{-1} \text{mol}^{-1}$ ) as functions of molecular weights ( $M$ ) at  $T = 298.15 \text{ K}$  for (a) series of alkylferrocenes in condensed (F,  $n\text{-PF}$ , and  $n\text{-BF}$ ) and ideal gas states ( $F'$ ,  $EF'$ ,  $n\text{-PF}'$ , and  $n\text{-BF}'$ ) where straight lines correspond to equations:  $S_m^0(\text{cond}, AlF) = 3.250 \cdot M - 377.1$  and  $S_m^0(g, AlF) = 3.840 \cdot M - 361.8$ ; and (b) series of acylferrocenes in the condensed (F,  $FF$ ,  $AF$ , and  $POF$ ) and ideal gas states ( $F'$ ,  $FF'$ ,  $AF'$ , and  $POF'$ ) where the straight lines correspond to equations:  $S_m^0(\text{cond}, AcF) = 1.392 \cdot M - 50.68$  and  $S_m^0(g, AcF) = 2.664 \cdot M - 144.6$

of the entropy of vapor compression calculation over the liquid phase because there are no vapor pressure data over the crystals. The verification of reliability of the other  $S_m^0(g)(\text{expt})$  values was analyzed on the basis of the additivity principle for extensive properties in homologous series of compounds.

As is seen from Table 9, the discrepancies between  $S_m^0(g)(\text{expt})$  and  $S_m^0(g)$  values of  $EF$ ,  $n\text{-BF}$ , and  $FF$  are 5-10 times larger than the errors of these values. To find the reasons for these discrepancies, the linear correlations between the absolute entropies and molecular weights of the compounds were analyzed in homologous series of alkyl-,  $S_m^0(AlF)$ , and acyl-,  $S_m^0(AcF)$ , derivatives of ferrocene (Fig. 13).

The root-mean-square deviation [RMS] of the calculated  $S_m^0(\text{cond}, AlF)$  values from experimental ones was  $\sigma_s = 29 \text{ J K}^{-1} \text{mol}^{-1}$ . Besides, the entropy of ethylferrocene deviates by the value of 10.5 % from the linear dependence of the other homologues. These deviations are several times larger than the uncertainty of  $S_m^0(\text{cond})$  values and can be explained by errors in calorimetric heat capacities, obtained in (Karyakin, et al., 2003). After excluding of this value, the coefficient  $R^2$  of linear correlation increased from 0.9147 to 0.9992, and the  $\sigma_s$  value decreased to  $4 \text{ J K}^{-1} \text{mol}^{-1}$ . Thus, the value  $S_m^0(\text{cond})$  for  $EF$  is unreliable. The  $S_m^0(g)(\text{expt})$  entropies for  $n$ -propylferrocene and  $n$ -butylferrocene deviate from the smoothed straight line (Fig. 13) by only (3.9 and 3.2) %, nevertheless the RMS deviation of calculated  $S_m^0(g)(\text{expt})$  values from initial ones  $\sigma_s = 29 \text{ J K}^{-1} \text{mol}^{-1}$  is fairly large. After replacing the absolute entropy of  $n\text{-BF}$ , obtained from experimental data, on the  $S_m^0(g)(\text{calc})$  value, calculated by the empirical method of group equations, the  $\sigma_s$  deviation

decreased to  $6.2 \text{ J} \cdot \text{K}^{-1} \text{ mol}^{-1}$  and appropriate deviations of  $S_m^0(g)(\text{expt})$  values for *n*-PF and *n*-BF from the smoothed straight line did not exceed  $\sim 1 \%$ .

Correlations for the absolute entropies of acylferrocenes obtained for the condensed and ideal gas states (Table 9) are shown in Fig. 13, b. The RMS deviation  $\sigma_s = 6 \text{ J} \cdot \text{K}^{-1} \text{ mol}^{-1}$  of the experimental  $S_m^0(\text{cond}, \text{AcF})$  entropies from the smoothed straight line conforms to the errors of these values. Appropriate deviation for the ideal gas entropies  $S_m^0(\text{expt AcF})$  is substantially larger,  $\sigma_s = 23.4 \text{ J} \cdot \text{K}^{-1} \text{ mol}^{-1}$ . Besides, the deviation of the entropy of formylferrocene (Table 9) from smoothed straight line,  $27.5 \text{ J} \cdot \text{K}^{-1} \text{ mol}^{-1}$ , is  $\sim 10$  times larger than the experimental error, which can be explained by errors in vapor pressure and the entropies of sublimation for FF. After replacing this  $S_m^0(g)(\text{expt})$  value by the  $S_m^0(g)(\text{calc})$  entropy obtained by the method of statistical thermodynamics (Table 9), appropriate deviations of the ideal gas entropies from smoothed  $S_m^0(g, \text{AcF})$  line decreased to  $\leq 1.4 \%$  for the all compounds of the AcF – series. This deviation is close to the measurement errors, which indicate to unreliability of the  $S_m^0(g)(\text{expt})$  value for FF.

A critical analysis of the absolute entropies of the alkyl and acyl derivatives of ferrocene shows that the  $S_m^0(g)(\text{calc})$  and  $S_m^0(g)(\text{expt})$  values for *n*-propylferrocene, *iso*-butylferrocene, acetylferrocene, propionylferrocene, and *iso*-butyrylferrocene are reliable to within errors. The entropies of ethylferrocene in the condensed state and *n*-butylferrocene and formylferrocene in the ideal gas states are not reliable because of errors in EF heat capacity measurements and errors of the vapor pressures and enthalpy of vaporization and sublimation for *n*-BF and FF, respectively. On the basis of the critical analyses of the absolute entropies of the ferrocene derivatives, the recommended reliable  $S_m^0(g)(\text{recom})$  values have been presented in Table 9.

#### 4.2 Extrapolation of vapor pressure to entire range of liquid phase

A saturation vapor pressure of substances for entire liquid region was obtained by calculation methods on the basis of the experimental  $pT$  data of moderate "atmospheric range" from 2 to 101.6 kPa. Extending the  $pT$  parameters to the region of low pressures is performed by simultaneous processing the vapor pressure and differences between the heat capacities of the ideal gas and liquid. The vapor pressures above 100 kPa are calculated by the empirical bimodal equation obtained by processing the  $pT$  parameters and densities of liquids using the one-parameter corresponding states law in Filippov's version.

The extrapolation of vapor pressure to the low-temperature region down to the triple point is carried out by a system of two equations:

$$\begin{cases} \ln(p / \langle p \rangle) = A' + B' / T + C' \cdot \ln T + D' \cdot T \\ (\Delta C_{p,m} / 2) / R = (1/2) \cdot (C_{p,m}^0(g) - C_{p,m}^0(\text{liq})) \cdot R = C' / 2 + D' \cdot T \end{cases} \quad (23)$$

where  $\langle p \rangle$  is the vapor pressure at the average temperature of experimental range, and  $A'$ ,  $B'$ ,  $C'$ , and  $D'$  are coefficients of simultaneous processing of the  $pT$  data with  $\Delta C_{p,m}$  differences.

Mathematical processing by a system of equations (23) requires the use of heat capacities of the ideal gas and liquid in the neighbourhood of the triple point, where the linear temperature dependence of  $C_{p,m}$  is valid.



The method of extrapolation of the vapor pressure was tested with standard substances 1,1,1-trifluoro-2,2-dichloroethane and *n*-decane for which precision  $pT$  parameters and heat capacities of the ideal gas and liquid are available in a wide temperature range.

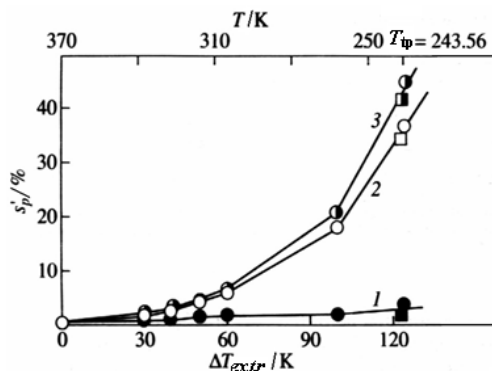


Fig. 14. Deviation  $s'_p = \{(p_{recom} - p_{extr}) / p_{recom}\} \times 100$  of extrapolated values  $p_{extr}$  of the vapor pressure of *n*-decane from experimental recommended values  $p_{recom}$  as functions of temperature range of extrapolation  $\Delta T_{extr}$  : 1 - calculation by a system of equations (23); 2 and 3 - calculation by equation (4) using full and shortened ( $\Delta T = 50$  K) interval of  $pT$  parameters

Fig. 14 presents comparison of the extrapolation capabilities of equations (4) and the system of equations (23) by processing the precision  $pT$  data of *n*-decane in moderate range of pressure and the  $\Delta C_{p,m}$  differences in the neighbourhood of triple point  $T_{tp} = 243.56$  K. The deviations of the extrapolated  $p$  values from the recommended experimental vapor pressures of *n*-decane (Boublik et al., 1984) and most reliable calculated values (Ruzicka & Majer, 1994) were evaluated using equation (4) (curves 2 and 3) and the system of equations (23) (curve 1). Curves 2 and 3 were obtained for the complete and shortened temperature ranges of the  $pT$  data  $\Delta T = (80$  and  $50)$  K, respectively. In so doing, the deviation of the calculated  $p$  values from experimental ones,  $s'_p$ , increase for equation (4) from (5 to 40) % in dependence on the temperature range of extrapolation,  $\Delta T_{extr}$ , from (50 to 120) K. The appropriate deviations for the system of equations (23) are much smaller and equal from (2 to 4) % for this range of extrapolation.

The results of approximation of  $pT$  data for 1,1,1-trifluoro-2,2-dichloroethane (Weber, 1992) by individual equation (4) and by system of equations (23) are compared in Table 10. The data were processed by the least squares method using the vapor pressure in the temperature range from 256 to 299 K and the heat capacities of the ideal gas (Frenkel et al., 1994) and liquid (Varushchenko et al., 2007) in the range from 150 to 240 K.

In so doing, the temperature range of extrapolation of vapor pressure from initial temperature 256 K to  $T_{tp} = 145.68$  K equals  $\Delta T_{extr} = 111$  K. Tentative errors of approximation coefficients were evaluated only for revealing the change of extrapolation prediction in going from equation (4) to that of (23). As is seen from Table 10, errors of

coefficients of equations (23) are approximately ten times lower than those of equation (4), which explains increasing the extrapolation capabilities of the system (23). The resultant value of the vapor pressure at triple point  $p_{tp}$  is about 25% lower than that of equation (4). Reducing the temperature range of initial  $pT$  data,  $\Delta T$ , from (25 to 50) K (accordingly, extending the range of extrapolation of vapor pressure to  $\sim 130$  K) leads to the variation of  $p_{tp}$  value by only 4-6%.

Equation	$A / A'$	$-(B / B')$	$-(C / C')$	$(D / D') \times 10^{-3}$	$p_{tp}, \text{Pa}$
(4)	98.82786 $\pm$ 11.09 (10%)	6061.251 $\pm$ 273.0 (4%)	14.58016 $\pm$ 1.97 (10%)	17.7902 $\pm$ 3.56 (20%)	0.155
(23)	78.44263 $\pm$ 0.496 (0.6%)	5677.915 $\pm$ 29.87 (0.5%)	10.76986 $\pm$ 0.093 (1%)	9.0292 $\pm$ 0.23 (2.5%)	0.114

Table 10. The coefficients of equations (4) and (23) and saturated vapor pressure of 1,1,1-trifluoro-2,2-dichloroethane at the triple point,  $T_{tp} = 145.68 \pm 0.02$  K (Weber, 1992)

Analysis of calculated data for *n*-decane and 1,1,1-trifluoro-2,2-dichloroethane revealed that the use of the system of equations (23) for extrapolation of the vapor pressure in a wide temperature range  $\Delta T_{extr} \leq (120 \text{ to } 130)$  K enabled one to obtain  $p(T_{tp})$  values with errors of  $\leq 10\%$ .

Extending the saturation vapor pressure to the critical region and computing the critical parameters of the substances were performed by the corresponding states law of Filippov's version (Filippov, 1988). A large class of normal (unassociated) liquids obey the one-parameter law of corresponding states [OLCS] which manifests an existence of the one-parameter family of  $\pi - \varphi - \tau$  surfaces of liquids. For the saturation curve, the OLCS equation has the form  $\pi = \pi(\tau, A') = 0$  or  $\varphi = \varphi(\tau, A')$ , where  $A'$  is the similarity criterion. The latter is assigned by ordinate  $\pi$  with abscissa  $\tau = 0.625$  on the reduced curve of saturated vapor pressure,  $A'_F = 100\pi$  at  $\tau = 0.625$ .

Few literature data are available for critical parameters of organic compounds. The law of corresponding states in Filippov's version enables one both to obtain the required critical parameters using precise  $pT$  data and density  $\rho$  of liquid and to calculate numerous thermophysical properties on the basis of known values of  $T_C$ ,  $V_C$ , and  $A'_F$ . Generalized equations for the calculation of the critical parameters and of the thermophysical properties were obtained by empirical method using the array of experimental data on the appropriate properties for studied compounds (Filippov, 1988). The critical parameters were calculated by algorithms (Filippov, 1988) and (Varushchenko et. al., 1987), in which the following pairs of input data were used:

$$\left. \begin{array}{l} T_{n.b.} \\ T_2 \\ T_3 \end{array} \right\} \begin{array}{l} p = 101.325 \\ p_2 \\ \rho \end{array} \quad \text{or} \quad \left. \begin{array}{l} T_{n.b.} \\ T_2 \\ T_3 \end{array} \right\} \begin{array}{l} p = 101.325 \\ d \ln(p) / d(1/T) \\ \rho \end{array} \quad (24)$$

The derivative  $d \ln(p) / d(1/T)$  was determined by differentiation of the equations  $\ln(p) = F(T)$  for investigated substances. A random sampling of 14 compounds, having reliable data on the vapor pressure, density, and critical parameters was used to demonstrate (Varushchenko et. al., 1987) that 1) the employed algorithm is capable to produce adequate results in computing  $T_c$ ,  $V_c$ , and  $A'_F$  and 2) the calculation algorithm originally developed for compounds with the similarity criterion  $1 \leq A'_F \leq 4$  is suitable for a wider class of compounds, including polyatomic molecules with the criterion  $0.5 \leq A'_F \leq 4$ . A prediction of the thermophysical properties by the OLCS of Filippov (Filippov, 1988) requires following basic quantities:  $T_c$ ,  $V_c$ , and  $A'_F$ . The errors of the latter calculation are within  $\pm 1 - 2\%$ ; the appropriate errors of  $P_c$ , are higher and amount to  $\pm 3 - 5\%$ . The saturation vapor pressure are extrapolated to the critical region using the empirical binodal equation with pseudocritical parameters  $T^* \geq T_c$  and  $P^* \geq P_c$ :

$$\log(P^* / p) = a \cdot \log(T^* / T) + [T^* / T - 1] \cdot [b + c(T^* / T)], \quad (25)$$

where  $a = 3.9726$ ,  $b = 0.3252$ , and  $c = 0.40529$ .

Two-parameter formula (25) weakly depends on similarity criterion  $A'_F$ ; therefore, it may be employed for curvilinear extrapolation of saturated vapor pressure using only two pairs of the  $pT$  data. The formula was derived by the similarity conversion, namely using superposition of curves  $-\log(\pi) / -\log(\tau)$  which have the same slope for different substances. As a result, the family of curves reduces to a single curve which has segments of vapor pressure in the supercritical region. The pair of the  $P^* - T^*$  parameters is of importance for remote point on the binodal curve, which enables one to calculate the vapor pressure of a substance in the region from the normal boiling temperature to the critical point. The error of the pressure calculation is  $\leq (3 \text{ to } 5) \%$  depending on the temperature range of  $pT$  data extending and their reliability.

Therefore, analysis of equations (4) and (23) demonstrated that equation (4) obtained by approximation of vapor pressure of the "atmospheric range" gives precision results when it used as an interpolation equation. It is further employed for extrapolation of the vapor pressure in the temperature range of  $\leq 50 \text{ K}$  with an error of  $(1 \text{ to } 2) \%$ . Extending the vapor pressure to the entire liquid phase from the critical to the triple points temperatures is performed by means of one-parameter law of corresponding states with errors from  $(3 \text{ to } 5) \%$  and, respectively, by simultaneous processing of the  $pT$  parameters and the differences between low-temperature heat capacities  $\Delta C_{p,m} = C_{p,m}^0(g) - C_{p,m}^0(liq)$  with uncertainties of  $\leq 10 \%$ , respectively. The data on extrapolation of the vapor pressure are suitable for many technological calculations.

## 5. Conclusion

The fundamental thermodynamic investigations of the phase equilibria of many functional organic compounds were carried out by experimental and calculation methods in Luginin's Thermochemistry Laboratory of the Moscow State University. Modified setups have been created for precise determination of the saturation vapor pressures by

comparative ebulliometry, the enthalpies of vaporization by evaporation calorimetry, and the low temperature heat capacity and phase transitions by vacuum adiabatic calorimetry.

The saturation vapor pressures were determined in moderate range of pressure  $2 \leq (p/\text{kPa}) \leq 101.6$  with accuracy of the temperature  $\pm 0.01$  K, and pressure,  $\pm 26$  Pa, which correspond to the modern precision levels. The temperature dependences of the saturation vapor pressure,  $\ln(p) = F(T)$ , and the enthalpies of vaporization  $\Delta H_{vap} = f(T)$  were obtained by mathematical processing of the  $pT$  parameters by equation (1), derived on the basis of Clapeyron equation using LSM with orthogonal functions. The latter allow one to calculate the errors of  $\Delta H_{vap}$  values, which are urgent problem because the indirect method is the main source for determination of the enthalpies of vaporization. An agreement of the  $\Delta H_{vap}$  values obtained by direct (calorimetric) and indirect (calculation) methods proves their reliability.

The precise saturated vapor pressure data are extended to entire region of the liquids under study. Extrapolation of the  $pT$  parameters down to the triple point temperature are carried out by simultaneous processing the vapor pressures and low-temperature differences  $\Delta C_{p,m} = C_{p,m}^0(g) - C_{p,m}^0(liq)$ , which are the second derivatives of the vapor pressure upon the temperature. Extending the  $pT$  parameters to the critical region and calculation of the critical quantities are performed by Filippov's one-parameter law of the corresponding states. The latter enables us both to calculate the critical parameters on the basis of more readily available  $pT$ -data and density of liquids and to predict numerous thermo-physical properties of the equilibrium liquid - vapor by means of the known critical quantities  $T_C$ ,  $V_C$ , and criterion of similarity  $A'_F$ .

The low temperature heat capacity in the temperature region (5 to 373) K, molecular motion in crystal and metastable phases, and solid state transitions and fusion were investigated by the adiabatic calorimetry. Uncertainties of the  $C_{p,m}$  measurements are on the average  $\sim 0.2\%$  which correspond up-to-date precision level.

An accurate calorimetric study of the solid states of the functional organic compounds revealed different polymorphic modifications of the molecules, order - disorder transitions involving orientational and conformational disorder, glass-like transitions, and plastic crystalline phases with anisotropic and isotropic reorientations of the molecules. For interpretation of these transformations, the X-ray crystallography, infrared and Raman molecular spectroscopy were got involved in investigation.

The main thermodynamic functions in three aggregate states: the absolute entropy by the 3<sup>d</sup> law of thermodynamics, the changes of the enthalpy and free Gibbs energy are derived on the basis of the heat capacity and vapor pressure measurements. A critical analysis and verification of the reliability of obtained data are very significant parts of the thermodynamic investigation. With this in mind, the experimental thermodynamic functions are compared with calculated ones by additive principles and by statistical thermodynamics coupled with quantum mechanical (QM) calculation on the basis of DFT method. The QM calculation are performed on the level B3LYP/6-31G(d,p) by Gaussian 98 and 03 software packages.

A qualitative analysis of thermodynamic properties in dependence on some parameters responsible for intermolecular interactions and short range order of the liquid phase has been carried out for verification of the mutual consistency of the properties in homologous series and the series of the same type of compounds. A quantitative verification of the

properties were performed by comparison of the absolute entropies, obtained by the third thermodynamic law,  $S_m^0(g)(expt)$ , and calculated by statistical thermodynamics,  $S_m^0(stat)$ . The agreement between these quantities confirms their reliability, as well as all experimental and calculation data used in computing process.

## 6. Acknowledgment

We are grateful to Professor O. Dorofeeva for providing Gaussian programs and assistance in quantum-chemical calculations of the ideal gas thermodynamic functions. Many thanks to Professor S. Verevkin for helping in determination of the vapor pressures of some derivatives of ferrocene and Dr. L. Pashchenko for taking part in determination of the vapor pressures of some freons. Special thanks are to Post-graduate student E. Tkachenko for providing illustrative materials of the chapter.

This work was financially supported by Russian Foundation for Basic Research under Grants No. 96-02-05445 and No. 05-02-17435.

## 7. Appendix

### 7.1 Mathematical processing of saturation vapor pressure

The vapor pressures in dependence on temperature are treated by a relationship:

$$\begin{aligned}
 -RT \ln(p) = \Delta H_{\langle T \rangle} - \alpha_1 \cdot T + \alpha_2 \cdot \{T - \langle T \rangle - T \cdot (\ln T - \ln \langle T \rangle)\} \\
 + \alpha_3 \cdot [T^2 / 2 - \langle T \rangle^2 / 2 - T \cdot \langle T \rangle \cdot (\ln T - \ln \langle T \rangle)]
 \end{aligned}
 \tag{A1}$$

where  $\alpha_1 = (\Delta H'_{\langle T \rangle} / \langle T \rangle) + R \ln(p_{\langle T \rangle})$ .

Coefficients of equation (A1) are calculated by LSM using orthogonal functions (Kornilov & Vidavski, 1969).

Two functions  $\varphi_k$  and  $\varphi_s$  are orthogonal if the following condition is satisfied:

$$[\varphi_k \varphi_s] = 0 \text{ at } k \neq s
 \tag{A2}$$

According to Gauss, the  $\sum$  sign was replaced by  $[ ]$ . If there are  $m$  functions  $\Phi_0, \Phi_1, \dots, \Phi_m$  they may be replaced by linear combination of the orthogonal functions  $\varphi_0, \varphi_1, \dots, \varphi_p$  using Shmidt method:

$$\varphi_0 = \Phi_0; \varphi_1 = \Phi_1 - \varepsilon_{10}\varphi_0; \varphi_2 = \Phi_2 - \varepsilon_{20}\varphi_0 - \varepsilon_{21}\varphi_1; \dots; \varphi_p = \Phi_p - \sum_{q=0}^{q=p-1} \varepsilon_{pq}\varphi_q \dots, \tag{A3}$$

where coefficients of orthogonality,  $\varepsilon_{pq}$ , are defined from equation (A3) and the orthogonal condition as:

$$\varepsilon_{pq} = [\Phi_p \varphi_q] / [\varphi_q^2] \quad (\text{A4})$$

A description of the technique of vapor pressures approximation by equation (A1) is given below.

The data on the saturation vapor pressure in dependence on the temperature have unequal accuracy; therefore treating those by the LSM were carried out using statistical weights  $w_i$  :

$$\{-RT \ln(p_i)\} \cdot w_i = \Phi(T_i) . \quad (\text{A5})$$

The value of  $w_i$  is introduced by inversely proportional to the dispersion:

$$w_i = 1 / \{2 \cdot s[-RT_i \ln(p_i)]\}^2 \quad (\text{A6})$$

In conformity with the law of an accumulation of the random errors, the dispersions were calculated from the formula:

$$s^2 \{-RT_i \ln(p_i)\} = \{-R \ln(p_i)\}^2 s_T^2 + \{-RT_i / p_i\}^2 s_p^2$$

For simplification of the calculation, equation (A1) is reduced to the linear form:

$$Y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3, \quad (\text{A7})$$

where  $Y = -RT \ln(p)$ ,  $a_0 = \Delta H' / \langle T \rangle$ ,  $a_1 = -\{\Delta H' / \langle T \rangle\} / \langle T \rangle + R \ln(p / \langle T \rangle)$ ,  $a_2 = \alpha_2$ ,

$$a_3 = -\alpha_3, \quad x_1 = T,$$

$$x_2 = \{T - \langle T \rangle - T[\ln(T) - \ln(\langle T \rangle)]\}, \quad x_3 = \{T^2 / 2 - \langle T \rangle^2 / 2 - T\langle T \rangle[\ln(T) - \ln(\langle T \rangle)]\}.$$

Owing to transition to the orthogonal functions, equation (A7) takes a form:

$$Y = b_0 \varphi_0 + b_1 \varphi_1 + b_2 \varphi_2 + b_3 \varphi_3, \quad (\text{A8})$$

where  $\varphi_0 = 1$ ,  $\varphi_1 = x_1 - \varepsilon_{10} \varphi_0$ ,  $\varphi_2 = x_2 - \varepsilon_{20} \varphi_0 - \varepsilon_{21} \varphi_1$ ,

$$\varphi_3 = x_3 - \varepsilon_{30} \varphi_0 - \varepsilon_{31} \varphi_1 - \varepsilon_{32} \varphi_2.$$

The coefficients of orthogonality,  $\varepsilon_{pq}$  (A4), expressed with the use of statistical weights, are:

$$\varepsilon_{10} = [x_1 \varphi_0 w] / [\varphi_0^2 w], \quad \varepsilon_{20} = [x_2 \varphi_0 w] / [\varphi_0^2 w], \quad \varepsilon_{30} = [x_3 \varphi_0 w] / [\varphi_0^2 w],$$

$$\varepsilon_{21} = [x_2 \varphi_1 w] / [\varphi_1^2 w],$$

$$\varepsilon_{31} = [x_3 \varphi_1 w] / [\varphi_1^2 w], \quad \varepsilon_{32} = [x_3 \varphi_2 w] / [\varphi_2^2 w]. \quad (\text{A9})$$

The coefficients  $b_p$  of equation (A8) were determined by LSM using a minimum of a sum of the deviations square of the experimental  $y_i = -\{RT \ln(p_i)\}w_i$  values from those ones calculated by (A8):

$$\sum_i [y_i - (b_0\phi_0 + b_1\phi_{1i} + b_2\phi_{2i} + b_3\phi_{3i})]^2 = \min \quad (\text{A10})$$

By differentiation of the expression (A10) upon  $b_p$  and followed by transformation of the obtained equations using a ratio (A2), a system of normal equations were obtained, every equation being contained only one unknown coefficient:

$$\begin{aligned} [y_i\phi_0w_i] - b_0[\phi_0^2w_i] &= 0 \\ [y_i\phi_1w_i] - b_1[\phi_1^2w_i] &= 0 \\ [y_i\phi_2w_i] - b_2[\phi_2^2w_i] &= 0 \\ [y_i\phi_3w_i] - b_3[\phi_3^2w_i] &= 0 \end{aligned} \quad (\text{A11})$$

Thus, the system of the normal equations is a diagonal matrix. The formulae intended for calculation of  $b_p$  coefficients were obtained on the basis of the normal system of equations (A11):

$$\begin{aligned} b_0 &= [y_i\phi_0w_i]/[\phi_0^2w_i], \quad b_1 = [y_i\phi_1w_i]/[\phi_1^2w_i], \quad b_2 = [y_i\phi_2w_i]/[\phi_2^2w_i], \\ b_3 &= [y_i\phi_3w_i]/[\phi_3^2w_i] \\ b_p &= [y_i\phi_pw_i]/[\phi_p^2w_i] \end{aligned} \quad (\text{A12})$$

The  $s^2(b_p)$  dispersions of the coefficients  $b_p$  were defined in accordance with the formula (A12), as:

$$s^2(b_p) = s_0^2(\partial b_p / \partial y)^2 = s_0^2 / [\phi_p^2w_i] \quad (\text{A13})$$

where  $s_0^2 = r^{-1} \sum (y_i - Y_i)^2$  is dispersion of individual measurement,  $r = n - l$  denotes a number of a freedom degree,  $n$  is a number of measurements,  $l$  is the number of terms of approximation equation,  $y_i$  are experimental values of the  $y(x)$  function,  $Y_i$  are the values of the function calculated from approximated equation.

The constants  $a_i$  of equation (A1) were calculated after the coefficients of  $b_p$  and  $\varepsilon_{pq}$  have been rounded off. The  $b_p$  and  $\varepsilon_{pq}$  coefficients were rounded up to the values of their uncertainties, but one spare digit being excepted. The errors of  $b_p$  and  $\varepsilon_{pq}$  coefficients were calculated by formulae:

$$\Delta b_p = t_{0.05} \{s_0^2 / [\varphi_p^2 w]\}^{1/2} \quad \text{and} \quad \Delta \varepsilon_{pq} = \Delta x_p / \{[\varphi_p^2 w]\}^{1/2},$$

where  $\Delta x_p$  is uncertainty of rounding of initial  $x_p$  value. Dispersion of  $s_0^2$  value and sums  $[\varphi_p^2 w]$ , which were used by computing the  $\Delta b_p$ ,  $\Delta \varepsilon_{pq}$ , and  $b_i$  constants (A12) were rounded off up to three significant figures. On the basis of formula (A7) and uncertainty of the temperature measurements of  $\pm 0.006$  K, the  $\Delta x_i$  errors were evaluated equal to:  $\Delta x_1 = 0.006$  K,  $\Delta x_2 = 0.024$  K and  $\Delta x_3 = 0.048$  K<sup>2</sup>.

The constants  $a_1$ ,  $a_2$ , and  $a_3$  are easily determined by means of coefficients of  $b_p$  and  $\varepsilon_{pq}$  as  $\varphi$ -functions are linear combination of the  $x_1$ ,  $x_2$ , and  $x_3$  variable values:

$$a_3 = b_3, \quad a_2 = b_2 - \varepsilon_{32} b_3, \quad a_1 = b_1 - \varepsilon_{21} b_2 - \varepsilon_{31} b_3 + \varepsilon_{32} \varepsilon_{21} b_3 \quad (\text{A14})$$

The constant  $a_0$  of equation (A7) are determined when replacing its terms by appropriate mean ones, namely  $\langle x \rangle_{iw}$  and  $\langle y \rangle_w$ :

$$a_0 = \langle y_w \rangle - a_1 \langle x_{1w} \rangle - a_2 \langle x_{2w} \rangle - a_3 \langle x_{3w} \rangle \quad (\text{A15})$$

A calculation of the  $a_p$  coefficients by the method described above is simpler than straight treating the  $pT$ -parameter by equation (A7) using LSM. For decreasing a number of tabulated coefficients, the final equation (A7) was transformed to the form:

$$\ln(p) = A + B/T + C \cdot \ln(T) + D \cdot T \quad (\text{A16})$$

Here  $A$ ,  $B$ ,  $C$ , and  $D$  are correlated constants, which were obtained by reducing the similar terms of equation (A7). A number of significant digits in the coefficients were chosen based on a condition: deviations of calculated vapor pressures from the experimental ones should not exceed the uncertainties of determination of the  $p$  values ( $13 \div 26$  Pa). Then, the accuracy of each coefficient of equation (A16) was designated depending on the accuracy of  $\ln(p) = f(T)$  function using derivatives  $\partial \ln(p) / \partial K$ , where  $K = A, B, C$  or  $D$  coefficients. If  $\partial \ln(p) / \partial K \approx \Delta \ln(p) / \Delta K = X$ , then  $\Delta K = \Delta \ln(p) / X$ . This formula was used for evaluation of  $\Delta K$  values on the basis maximum values of  $p = 101.6$  kPa and  $T = 530$  K, determined in modified ebulliometer.

$$\Delta A = \Delta \ln(p) = \frac{\Delta p}{p} = \frac{13 \cdot 10^{-3}}{101.6} = 0.00010$$

$$\Delta B = T \cdot \Delta \ln(p) = \frac{T \cdot \Delta p}{p} = \frac{530 \cdot 13 \cdot 10^{-3}}{101.6} = 0.07$$

$$\Delta C = \frac{\Delta \ln(p)}{\ln(T)} = \frac{\Delta p}{p \cdot \ln(T)} = \frac{13 \cdot 10^{-3}}{101.6 \cdot 6.3} = 0.00002$$

$$\Delta D = \frac{\Delta \ln(p)}{T} = \frac{\Delta p}{p \cdot T} = \frac{13 \cdot 10^{-3}}{101.6 \cdot 530} = 0.0000002$$



## 7.2 Calculation of the enthalpies of vaporization

Mathematical processing the saturated vapor pressure in dependence on temperature by LSM with orthogonal functions makes it possible to calculate both the enthalpy of vaporization and its dispersion  $s^2(\Delta_{vap}H)$ . The latter for the thermodynamic values  $F = \Delta H_T$  is designated based on the law of an accumulation of the random errors by the formula:

$$s^2(F) = s_0^2 \sum (\partial F / \partial y_i)^2 \quad (A17)$$

An equation for computing the temperature dependence of the enthalpy of vaporization was obtained by means of formula (A1) and Clapeyron equation (2) (part 2.1) taking account on interconnections between the  $\alpha_1, \alpha_2, \alpha_3$  constants of (A7) and (A15) and connections of these coefficients with  $b_p$  by (A14), respectively:

$$\begin{aligned} \Delta H'_T = \Delta H'_T + \alpha_2(T - \langle T \rangle) + \alpha_3(T^2 - \langle T \rangle^2) / 2 = \langle y_w \rangle - b_1 \langle x_{1w} \rangle + b_2(\varepsilon_{21} \langle x_{1w} \rangle + T - \langle T \rangle - \langle x_{2w} \rangle) + \\ + b_3 \{ (\varepsilon_{31} - \varepsilon_{32} \varepsilon_{21}) \langle x_{1w} \rangle + \varepsilon_{32} (\langle T \rangle + \langle x_{2w} \rangle) - \langle T \rangle^2 / 2 - \langle x_{3w} \rangle - \varepsilon_{32} T + T^2 / 2 \} \end{aligned} \quad (A18)$$

Computing the dispersion,  $s^2(\Delta H'_T)$ , caused by errors of the  $pT$ -data, was carried out by the formula, deduced based on (A17):

$$s^2(\Delta H'_T) = s_0^2 \sum_i \{ (\partial \Delta H'_T) / (\partial y_i) \}^2 = s_0^2 \left\{ \frac{1}{[w_i]} + \frac{\langle x_{1w} \rangle^2}{[\varphi_1^2 w_i]} + \frac{(T + \gamma)^2}{[\varphi_2^2 w_i]} + \frac{(\lambda - \varepsilon_{32} T + T^2 / 2)^2}{[\varphi_3^2 w_i]} \right\}, \quad (A19)$$

$$\text{where } \gamma = \varepsilon_{21} \langle T \rangle - \langle x_{2w} \rangle - \langle T \rangle,$$

$$\lambda = (\varepsilon_{31} - \varepsilon_{32} \varepsilon_{21}) \langle x_{1w} \rangle + \varepsilon_{32} (\langle T \rangle + \langle x_{2w} \rangle) - \langle T \rangle^2 / 2 - \langle x_{3w} \rangle$$

According to approximation accepted by deriving of the equation (A6), the enthalpy of vaporization,  $\Delta_{vap}H$ , equals to product of the  $(\Delta H'_T)$  and  $\Delta Z$  values:

$$\Delta_{vap}H = \{ \Delta H'_T + \alpha_2(T - \langle T \rangle) + \alpha_3(T - \langle T \rangle)^2 / 2 \} \cdot \Delta Z, \quad (A20)$$

where  $\Delta Z$  is the difference of the compressibility factors, taking into account vapor deviation from ideality and volume changes of the vapor and liquid. For decreasing a number of tabulated coefficients, the equation (A19) was transformed to the form:

$$\Delta_{vap}H = R(-B + CT + DT^2) \cdot \Delta Z \quad (A21)$$

The coefficients of equation (A20) are connected with correlations:

$$B = (\alpha_2 \langle T \rangle + \alpha_3 \langle T \rangle^2 / 2 - \Delta H'_m(T)) / R, \quad C = \alpha_2 / R, \quad \text{and} \quad D = \alpha_3 / 2R.$$

As an example, Table A1 lists the parameters of equation for calculating the uncertainties of the enthalpies of vaporization caused by errors of the vapor pressures for some freons.

Compounds	$t_{0.05}$	$a$	$b$	$-c$	$d \cdot 10^{-3}$	$-e$	$g$	$s\{\Delta H'_m(T)\},$ J·mol <sup>-1</sup>	$s\{\Delta H'_m\} / \Delta H'_m \cdot 10^2,$ %
CFCl <sub>2</sub> CFCl <sub>2</sub>	2.306	0.389	0.0104	339	0.66	56918	337.6	43.1	0.12
CF <sub>2</sub> ClCFCl <sub>2</sub>	2.775	114	21.64	308	-	-	-	130.6	0.45
CF <sub>2</sub> BrCF <sub>2</sub> Br	2.569	27.8	3.709	310	-	-	-	60.2	0.21

Table A1. Parameters of equation for calculation of the  $s\{\Delta H'_m(T)\}$  values by

$$s\{\Delta H'_m(T)\} / (J \cdot mol^{-1}) = \pm t_{0.05} \cdot [a + b \cdot \{(T/K + c)\}^2 + d \cdot \{e + g \cdot (T/K) - 0.5 \cdot (T/K)^2\}^2]^{1/2} \quad (A22)$$

for freons CFCl<sub>2</sub>CFCl<sub>2</sub>, CF<sub>2</sub>ClCFCl<sub>2</sub> and CF<sub>2</sub>BrCF<sub>2</sub>Br at  $T = 298.15$  K, where  $t_{0.05}$  denotes the Student criterion.

This equation was obtained by summing up dispersions of the orthogonal coefficients of equation (A19). Errors of the  $\Delta_{vap}H_m(T)$  values caused by uncertainties of  $pT$  data equal to those of the direct calorimetric methods (Table 2).

## 8. References

- Adachi K.; Suga H. & Seki S. (1968). Phase changes in crystalline and glassy-crystalline cyclohexanol. *Bulletin Chemical Society of Japan*, 41, No 5, (May 1968), p.p.1073-1087, ISSN 0009-2673
- Alexandrov, Yu.I. (1975). *Tochnaja Kriometrija Organicheskikh Veschestv* (Exact criometric of organic substances). Khimiya, Leningrad
- Boublik, T.; Fried, V. & Hala, E. (1984). *The vapour pressure of pure substances*, Elsevier, ISBN 0-444-41097-X, Amsterdam
- Chikos, J.S. & Acree Jr., W.E. (2003). Enthalpies of vaporization of organic and organometallic compounds. *Journal of Physical and Chemical Reference Data*, Vol. 32, No 2, p.p. 519-878, ISSN 0047-2689
- Cohen, N. & Benson, S.W. (1993). Estimation of Heat of Formation of Organic Compounds by Additivity Methods. *Chemical Reviews*. Vol. 93, (1993), p.p. 2419-2438
- Domalski, E.S. & Hearing, E.D. (1993). Estimation of the Thermodynamic Properties of C-H-N-O-S-Halogen Compounds at 298.15 K. *Journal of Physical and Chemical Reference Data*, Vol. 22, No 4, p.p. 805-1159, ISSN 0047-2689
- Emel'yanenko, V.N.; Verevkin, S.P.; Krol, O.V.; Varushchenko, R.M. & Chelovskaya, N.V. (2007). Vapour pressures and enthalpies of vaporization of a series of the ferrocene derivatives. *Journal of Chemical Thermodynamics.*, Vol. 39, No 4, (April 2007), p.p. 594-601, ISSN 0021-9614
- Emel'yanenko, V.N.; Krol, O.V.; Varushchenko, R.M.; Druzhinina, A.I. & Verevkin, S.P. (2010). The Thermodynamic Characteristics of Ferrocene Alkyl and Acyl Derivatives.

- Russian Journal of Physical Chemistry A*. Vol. 84, No 7, (July 2010), p.p. 1089-1097, ISSN 0036-0244
- Filippov, L.P. (1988). *Metodi rascheta i prognozirovaniya svoystv veschestv*. (Methods of calculation and prediction of properties of substances). Moscow State University, ISBN 5-211-00096-X, Moscow
- Frenkel, M.; Kabo, G.J.; Marsh, K.N.; Roganov, G.N. & Wilhoit, R.C. (1994). *Thermodynamics of Organic Compounds in the Gas State*, Thermodynamics Research Center, College Station, ISBN 1-883400-03-1 (Vol.1), ISBN 1-883400-05-8 (Vol.2), Texas, USA
- Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; et al., (2003). *GAUSSIAN 03*, Revision B.03, Gaussian, Inc, Pittsburgh, PA
- Higgins, E.R. & Lielmers, J. (1965). Thermodynamic Functions for Freon 113. *Journal of Chemical and Engineering Data*. Vol. 10, No 2, (April 1965), p.p. 178-179, ISSN 0021-9568
- Karyakin, N. V.; Kozlova, M. S.; Sheiman, M. S.; Kamelova, G.P. & Larina, V.N. (2003). Thermodynamic functions of derivatives of ferrocene. *Russian Journal of Physical Chemistry A*. Vol. 77, No 8, (August 2010), p.p. 1230-1236, ISSN 0036-0244
- Kolesov, V.P. (1995). The order - disorder transitions in some organic crystals. *Thermochimica Acta*, Vol. 266, (November 1995), p.p. 129-145, ISSN 0040-6031
- Kornilov, A.N. & Vidavski, L.M. (1969). Some questions of statistical processing of the thermodynamic data. III. Approximation of dependence  $\Delta G = \Delta H - T \cdot \Delta S$ . *Zh. Fiz. Khim.* Vol. 43, No 9, (September 1969), p.p. 2224- 2230, ISSN 0044-4537
- Kulikov, D.V; Verevkin, S.P. & Heintz, A. (2001). Enthalpies of Vaporization of a Series of Aliphatic Alcohols. Experimental Results and Values Predicted by the ERAS-Model. *International Journal Fluid Phase Equilibria*. Vol. 192, No 1-2, (December 2001), p.p. 187-207, ISSN 0378-3812
- Lawson, D.D.; Moacanin, J.; Scherer Jr., K.V; Terranova, T.F. & Ingham, J.D. (1978). Methods for the estimation of vapor pressures and oxygen solubility of fluorochemicals for possible application in artificial blood formulations. *Journal of Fluorine Chemistry*. Vol. 12, No 3, (September 1978) p.p. 221-236, ISSN 0022-1139
- Majer, V. & Svoboda, V. (1985). *Enthalpies of vaporization of organic compounds: A Critical Review and Data Compilation*. IUPAC. Blackwell Science, ISBN 0632015292, Oxford
- NIST Chemistry Webbook. Standard Reference Data. <http://webbook.nist.gov/chemistry>
- Poling, B.E.; Prausnitz, J.M. & O'Connell, J.P. (2001). *The Properties of Gases and Liquids*. 5-th Edition, McGraw-Hill, ISBN 0-07-011682-2 (Errata at <http://www.che.virginia.edu/PGL5> , (2010))
- Pvsev, F. & Malishev, V.M. (1994). Routine measurements of specific heat capacity and thermal conductivity of high- $T_c$  superconducting materials in the range 4 - 300 K using modular equipment. *Advances in Cryogenic Engineering*, Vol. 40, Proceedings of the 1993 International Cryogenic Materials Conference (ICMC), Albuquerque, New Mexico, p. 119-124, ISBN 0-306-44823-8
- Ries J.G. (1991). Fluorocarbon - Based in vivo Oxygen Transport and Delivery System. *Vox Sanguinis*. Vol. 61, No 4, (December 1991), p.p. 225-239, ISSN 1423-0410
- Ruzicka, K. & Majer, V. (1994). Simultaneous Treatment of Vapour Pressures and Related Thermal Data between the Triple and Normal Boiling Temperatures for n-Alkanes  $C_5 - C_{20}$ . *Journal of Physical and Chemical Reference Data*, Vol. 23, No 1, p.p. 1-39, ISSN 0047-2689

- Sabbe, M.K.; De Vleeschouwer, F.; Reyniers M.-F.; Waroquier M. & Marin G.B. (2008). First Principles Based Group Additive Values for the Gas Phase Standard Entropy and Heat Capacity of Hydrocarbons and Hydrocarbon Radicals. *Journal of Physical Chemistry, A*, Vol. 112, No 47, (November 2008), p.p. 12235-12251, ISSN 1089-5639
- Van Speybroeck, V.; Van Neck, D.; Waroquier, M.; Wauters, S.; Saeys, M. & Marin G.B. (2000). Ab Initio Study of Radical Addition Reactions: Addition of a Primary Ethylbenzene Radical to Ethene (I). *Journal of Physical Chemistry, A*, Vol. 104, No 46, (November 2008), p.p. 10939-10950, ISSN 1089-5639
- Vansteenkiste, P.; Van Speybroeck, V.; Marin G.B. & Waroquier, M. (2003). Ab Initio Calculation of Entropy and Heat Capacity of Gas-Phase *n*-Alkanes Using Internal Rotations. *Journal of Physical Chemistry, A*, Vol. 107, No 17, (April 2003), p.p. 3139-3145, ISSN 1089-5639
- Van Wijk, M.F.; & Smit W.M. (1960) Impurity determination by thermal analysis. I. The melting curve of gradually frozen sample. *Analytica Chimica Acta*. Vol. 23, No 6 (December 1960), p.p. 545-551, ISSN 0003-2670
- Varushchenko, R.M.; Halchenko, G.L. & Medvedev, V.A. (1977). Precision calorimetric device for measurement enthalpy of evaporation. *Zh. Fiz. Khim.* Vol. 51, No 4, (April 1977), p.p. 992-996, ISSN 0044-4537
- Varushchenko, R. M.; Druzhinina, A. I. & Pashchenko, L. L. (1987). Vapor pressure and critical parameters dichloroalkanes, *Zh. Fiz. Khim.* Vol. 61, No 9, (September 1987), p.p. 2327-2332, ISSN 0044-4537
- Varouchchenko, R.M. & Droujinina, A.I. (1995). Thermodynamics of vaporization of some perfluorotrialkylmines. *Journal of Chemical Thermodynamics.*, Vol. 27, No 4, (April 1995), p.p. 355-368, ISSN 0021-9614
- Varushchenko, R. M.; Druzhinina, A. I. & Sorokin, E. L. (1997a). Low-temperature heat capacity of 1-bromoperfluoro- octane. *Journal of Chemical Thermodynamics.*, Vol. 29, No 6, (June 1997), p.p. 623-637, ISSN 0021-9614
- Varushchenko, R. M.; Druzhinina, A. I. & Korshunova, M. V. (1997b). Low-temperature heat capacities and thermodynamic properties of 1,1-difluoro-1,2,2-trichloroethane and 1,2-difluoro-1,1,2-trichloroethane. *Journal of Chemical Thermodynamics.*, Vol. 29, No 10, (October 1997), p.p. 1059-1070, ISSN 0021-9614
- Varushchenko, R.M.; Druzhinina, A.I.; Kuramshina, G.M. & Dorofeeva, O.V. (2007). Thermodynamics of vaporization of some freons and halogenated ethanes and propanes. *International Journal Fluid Phase Equilibria*. Vol. 256, No 1-2, (August 2007), p.p. 112-122, ISSN 0378-3812
- Wadsö, I. (1966). A Heat of Vaporization Calorimeter for Work at 25 °C and for Small Amounts of Substances. *Acta Chemica Scandinavica* Vol. 20, No 2, (February 1966), p.p. 536-543, ISSN 0904-213X
- Weber, L.A. (1992). Ebulliometric Measurement of the Vapor Pressures of R123 and R141b. *International Journal Fluid Phase Equilibria*. Vol. 80, (November 1992), p.p. 141-148, ISSN 0378-3812
- Westrum, E.F. & McCullough, J.P. (1965). Thermodynamics of crystals, In: *Physics and Chemistry of the Organic Solid State*. Eds.: Fox, D., Labes, M.M. & Weissberger, A., Vol. 1, p.p. 5-155, Interscience, New-York – London

# Thermodynamics and Thermokinetics to Model Phase Transitions of Polymers over Extended Temperature and Pressure Ranges Under Various Hydrostatic Fluids

S  verine A.E. Boyer<sup>1</sup>, Jean-Pierre E. Grolier<sup>2</sup>,

Hirohisa Yoshida<sup>3</sup>, Jean-Marc Haudin<sup>4</sup> and Jean-Loup Chenot<sup>4</sup>

<sup>1</sup>*Institut P PRIME-P', ISAE-ENSMA, UPR CNRS 3346, Futuroscope Chasseneuil*

<sup>2</sup>*Universit   Blaise Pascal de Clermont-Ferrand, Laboratoire de Thermodynamique, UMR CNRS 6272, Aubi  re*

<sup>3</sup>*Tokyo Metropolitan University, Faculty of Urban Environmental Science, Tokyo*

<sup>4</sup>*MINES ParisTech, CEMEF, UMR CNRS 7635, Sophia Antipolis*

<sup>1,2,4</sup>*France*

<sup>3</sup>*Japan*

## 1. Introduction

A scientific understanding of the behaviour of polymers under extreme conditions of temperature and pressure becomes inevitably of the utmost importance when the objective is to produce materials with well-defined final in-use properties and to prevent the damage of materials during on-duty conditions. The proper properties as well as the observed damages are related to the phase transitions together with intimate pattern organization of the materials.

Thermodynamic and thermokinetic issues directly result from the thermodynamic independent variables as temperature, pressure and volume that can stay constant or be scanned as a function of time. Concomitantly, these variables can be coupled with a mechanical stress, the diffusion of a solvent, and/or a chemically reactive environment. A mechanical stress can be illustrated in a chemically inert environment by an elongation and/or a shear. Diffusion is typically described by the sorption of a solvent. A chemical environment is illustrated by the presence of a reactive environment as carbon dioxide or hydrogen for example.

Challenging aspects are polymer pattern multi scale organizations, from the nanometric to the macrometric scale, and their importance regarding industrial and technological problems, as described in the state of the art in Part 2. New horizons and opportunities are at hands through pertinent approaches, including advanced *ad hoc* experimental techniques with improved modelling and simulation. Four striking illustrations, from the interactions between a solvent and a polymer to the growth patterns, are illustrated in Part 3.

## 2. Multi-length scale pattern formation with *in-situ* advanced techniques

### 2.1 Structure formation in various materials

#### 2.1.1 Broad multi-length scale organization

The development of polymer-type patterns is richly illustrated in the case of biological materials and metals.

#### Pattern growth

Among the observed morphologies which extend from polymeric to metallic materials and to biologic species, similar pattern growth is observed. Patterns extend, with a multilevel branching, from the nanometric (**Fig. 1.a-b**) to the micrometric (**Fig. 1.c-d-e**) scale structures.

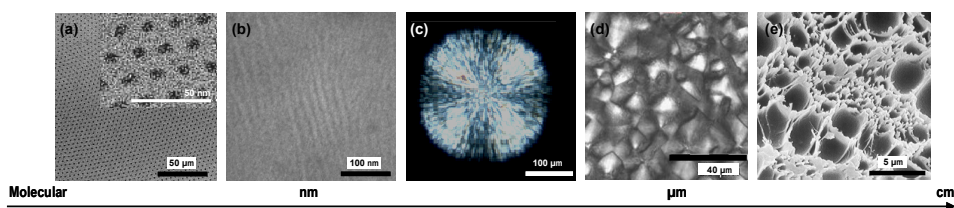


Fig. 1. Two-dimensional (2D) observations of various polymer patterns. **(a)** nanometric scale pattern of poly(ethylene-oxide) cylinders (PEO in black dots) in amphiphilic diblock copolymer  $\text{PEO}_m\text{-}b\text{-PMA}(\text{Az})_n$  (**a**, Iwamoto & Boyer, CREST-JSPS, Tokyo, Japan), **(b)** nanometric scale lamellae of an isotactic polypropylene (iPP, crystallization at  $0.1\text{ }^\circ\text{C}\cdot\text{min}^{-1}$ ,  $\text{RuCl}_3$  stained) with crystalline lamella thickness of 10 nm in order of magnitude, **(c)** micrometric scale of an iPP spherulite with lamellar crystals radiating from a nucleating point (iPP, crystallization at  $140\text{ }^\circ\text{C}$ ), **(d)** micrometric scale structure of a polyether block amide after injection moulding (**b-c-d**, Boyer, CARNOT-MINES-CEMEF, Sophia Antipolis, France), **(e)** micrometric scale cellular structure of a polystyrene damaged under carbon dioxide sorption at 317 K (**e**, Hilic & Boyer, Brite Euram POLYFOAM Project BE-4154, Clermont-Ferrand, France).

The polycrystalline features, formed by freezing an undercooled melt, are governed by dynamical processes of growth that depend on the material nature and on the thermodynamic environment. Beautiful illustrations are available in the literature. To cite a few, the rod-like eutectic structure is observed in a dual-phase pattern, namely for metallic with ceramic, and for polymeric (De Rosa et al., 2000; Park et al., 2003) systems like nanometric length scale of hexagonal structure of poly(ethylene-oxide) PEO cylinders in amphiphilic diblock copolymer  $\text{PEO}_m\text{-}b\text{-PMA}(\text{Az})_n$  with azobenzene part PMA(Az) (Tian et al., 2002). Dendritic patterns are embellished with images like snowflake ice dendrites from undercooled water (Kobayashi, 1993) and primary solidified phase in most metallic alloys (*e.g.*, steel, industrial alloys) (Trivedi & Laorchan, 1988a-b), and even dendrites in polymer blends (Ferreiro et al., 2002a) like PEO polymer dendrites formed under cooling PEO/polymethyl methacrylate PMMA blend (Gránásy et al., 2003; Okerberg et al., 2008). In the nanometric scale, immiscibility of polymer chains in block copolymers leads to microphase-separated structures with typical morphologies like hexagonally packed cylindrical structures, lamellae, spheres in centred cubic phases, double gyroid and double diamond networks (Park et al., 2003).

In polymer physics, the spherulitic crystallization (**Fig. 1.c**) represents a classic example of pattern formation. It is one of the most illustrated in the literature. Besides their importance in technical polymers, spherulitic patterns are also interesting from a biological point of view like semicrystalline amyloid spherulites associated with the Alzheimer and Kreutzfeld-Jacob diseases (Jin et al., 2003; Krebs et al., 2005). The spherulitic pattern depends on polymer chemistry (Ferreiro et al., 2002b). Stereo irregular atactic or low molecular weight compounds are considered as impurities, which are rejected by growing crystals. The openness of structure, from spherulite-like to dendrite-like, together with the coarseness of texture (a measure of the 'diameters' of crystalline fibres between which impurities become concentrated during crystal growth) was illustrated in the work of Keith & Padden (1964). These processes induce thermal and solute transport. Thus pattern formation is defined by the dynamics of the crystal/melt interface involving the interfacial energy. In the nanometric scale domain, spherulite is a cluster of locally periodic arrays of crystalline layers distributed as radial stacks of parallel crystalline lamellae separated by amorphous layers (**Fig. 1.b**). Molecular chains through the inter-lamellar amorphous layers act as tie molecules between crystalline layers, making a confined interphase crystalline lamellae/amorphous layer.

### **Cross fertilization between polymer crystallization and metal solidification**

Physical chemists and metallurgists alike are constantly confronted with materials properties related to (polymer) crystallization (*e.g.*, spherulite size distribution, lamellae spacing) or (metal) solidification (*e.g.*, grain size distribution, dendrite arm or eutectic spacing), respectively. In metal science, if accurate numerical modelling of dendritic growth remains a major challenge even with today's powerful computers, the growth kinetic theories, using accurate surface tension and/or kinetic anisotropies, are well advanced (Asta et al., 2009; Flemings, 1974). In polymer science, such approaches exist. But still insight into the physics/kinetics connection and morphologies is little known (Piorkowska et al., 2006). The most well-known growth kinetics theory is the one of Hoffman and coworkers (Hoffman, 1983) which is based on the concept of secondary nucleation; the nucleation and overall kinetics of crystallization have been also intensively studied (Avrami, 1939, 1940, 1941; Binsbergen, 1973; Haudin & Chenot, 2004).

### **2.1.2 Practical applications, importance of crystal organization**

The multi-length scale and semi-crystalline structure organizations are intimately linked with the chemical, physical, mechanical integrity and failure characteristics of the materials.

#### **Polymers with well-defined end-used properties**

Semi-crystalline polymers gain increasing importance in manufacturing (extended to recycling) industries where the control at the nano- to micro- up to macrometric hierarchical levels of the patterns constitutes a major engineering challenge (Lo et al., 2007). The domains extend from optics, electronics, magnetic storage, isolation to biosorption, medicine, packaging, membranes and even food industry (Rousset et al., 1998; Winter et al., 2002; Park et al., 2003; Nowacki et al., 2004; Scheichl et al., 2005; Sánchez et al., 2007; Wang et al., 2010).

#### **Control of polymer structure in processing conditions**

Industrial polymer activities, through processes like, for instance, extrusion coating (*i.e.*, the food industry with consumption products), injection moulding (*i.e.*, the industry with

engineering parts for automotive or medicine needs) (Devisme et al., 2007; Haudin et al., 2008), deal with polymer formulation and transformation. The viscous polymer melt partially crystallizes after undergoing a complex flow history or during flow, under temperature gradients and imposed pressure (Watanabe et al., 2003; Elmoumni & Winter, 2006) resulting into a non homogeneous final macrometric structure throughout the thickness of the processed part. The final morphologies are various sizes and shapes of more or less deformed spherulites resulting from several origins: *i*) isotropic spherulites by static crystallization (Ferreiro et al., 2002a; Nowacki et al., 2004), *ii*) highly anisotropic morphologies as oriented and row-nucleated structures (*i.e.*, shish-kebabs) by specific shear stress (Janeschitz-Kriegl, 2006; Ogino et al., 2006), *iii*) transcrystalline layer (as columnar pattern in metallurgy) by surface nucleation and/or temperature gradient, and *iv*) teardrop-shaped spherulites or “comets” (spherulites with a quasi-parabolic outline) by temperature gradients (Ratajski & Janeschitz-Kriegl, 1996; Pawlak et al., 2002).

Together with the deformation path (*e.g.*, tension, compression), the morphology strongly influences the behaviour of polymers. Some models have attempted to predict the properties of spherulites through a simulation of random distributions of flat ellipsoids (crystalline lamellae) embedded in an amorphous phase described by a finite extensible rubber network (Ahzi et al., 1991; Dahoun et al., 1991; Arruda & Boyce, 1993; Bedoui et al., 2006).

Moreover by considering the high-pressure technology, the use of specific fluids plays a non negligible role in pattern control. The thermodynamic phase diagrams of fluids implies the three coordinates (pressure-volume-temperature, *PVT*, variables) representation where the fluids can be in the solid, gaseous, liquid and even supercritical state. The so-called “signature of life” water ( $H_2O$ ) (Glasser, 2004) and the so-called “green solvent” in fact “clean safe” carbon dioxide ( $CO_2$ ) (Glasser, 2002) can be cited. The use of  $H_2O$  is encountered in injection moulding assisted with water.  $CO_2$  is known as a valuable agent in polymer processing thanks to its aptitude to solubilize, to plasticize (Boyer & Grolier, 2005), to reduce viscosity, to favour polymer blending or to polymerize (Varma-Nair et al., 2003; Nalawade et al., 2006). In polymer foaming, elevated temperatures and pressures are involved as well as the addition of chemicals, mostly penetrating agents that act as blowing agents (Tomasko et al., 2003; Lee et al., 2005).

### **Damage of polymer structure in on-duty conditions**

In the transport of fluids, in particular in the petroleum industry taken as an example, flexible hosepipes are used which engineering structures contain extruded thermoplastic or rubber sheaths together with reinforcing metallic armour layers. Transported fluids contain important amounts of dissolved species, which on operating temperature and pressure may influence the resistance of the engineering structures depending on the thermodynamic *T*, *P*-conditions and various phenomena as sorption/diffusion, chemical interactions (reactive fluids, *i.e.*, oxidation), mechanical (confinement) changes. The polymer damage occurs when rupture of the thermodynamic equilibrium (*i.e.*, after a sharp pressure drop) activates the blistering phenomenon, usually termed as ‘explosive decompression failure’ (XDF) process (Dewimille et al., 1993; Rambert et al., 2006; Boyer et al., 2007; Baudet et al., 2009). Damage is a direct result of specific interactions between semi-crystalline patterns and solvent with a preferential interaction (but not exclusive) in the amorphous phase (Klopffer & Flaconnèche, 2001).



## 2.2 Development of combined experimental procedures

The coupling of thermodynamic and kinetic effects (*i.e.*, confinement, shear flow, thermal gradient) with diffusion (*i.e.*, pressurizing sorption,) and chemical environment (*i.e.*, polar effect, oxidation), and the consideration of the nature of the polymers (*i.e.*, homopolymers, copolymers, etc.) require a broad range of indispensable *in-situ* investigations. They aim at providing well-documented thermodynamic properties and phase transitions profiles of polymers under various, coupled and extreme conditions.

### 2.2.1 Temperature control at atmospheric pressure

Usual developed devices are based on the control of temperature, while the main concerns are high cooling rate control and shearing rate.

The kinetic data of polymer crystallization are often determined in isothermal conditions or at moderate cooling rates. The expressions are frequently interpreted using simplified forms of Avrami's theory involving thus Avrami's exponent and a temperature function, which can be derived from Hoffman-Lauritzen's equation (Devisme et al., 2007). However, such an interpretation cannot be extrapolated to low crystallization temperatures encountered in polymer processing, *i.e.*, to high cooling rates (Magill, 1961, 1962, 2001; Haudin et al., 2008; Boyer et al., 2011b). In front of the necessity for obtaining crystallization data at high cooling rates, different technical solutions are proposed. Specific hot stages (Ding & Spruiell, 1996; Boyer & Haudin, 2010), quenching of thin polymer films (Brucato et al., 2002), and nanocalorimetry (Schick, 2009) are the main designs.

Similarly, to generate a controlled melt shearing, various shearing devices have been proposed, for instance, home-made sliding plate (Haudin et al., 2008) and rotating parallel plate devices (*e.g.*, Linkam temperature controlled stage, Haake Mars modular advanced rheometer system). The shear-induced crystallization can be performed according to a 'long' shearing protocol as compared to the 'short-term' shearing protocol proposed by the group of Janeschitz-Kriegl (Janeschitz-Kriegl et al., 2003, 2006; Baert et al., 2006).

### 2.2.2 Temperature-pressure-volume control

The design of devices based on the control of pressure requires breakthrough technologies. The major difficulty is to generate high pressure.

In polymer solidification, the effects of pressure can be studied through pressure-volume-temperature phase diagrams obtained during cooling at constant pressure. The effect of hydrostatic (or inert) pressure on phase transitions is to shift the equilibrium temperature to higher values, *e.g.*, the isotropic phase changes of complex compounds as illustrated in the works of Maeda et al. (2005) by high-pressure differential thermal analyzer and of Boyer et al. (2006a) by high-pressure scanning transitionometry, or the melting temperature in polymer crystallization as illustrated for polypropylene in the work of Fulchiron et al. (2001) by high-pressure dilatometry. However, classical dilatometers cannot be operated at high cooling rate without preventing the occurrence of a thermal gradient within the sample. This problem can be solved by modelling the dilatometry experiment (Fulchiron et al., 2001) or by using a miniaturized dilatometer (Van der Beek et al., 2005). Alternatively, other promising technological developments propose to couple the pressure and cooling rates as shown with an apparatus for solidification based on the confining fluid technique as described by Sorrentino et al. (2005). The coupling of pressure and shear is possible with the shear flow pressure-volume-temperature measurement system developed by Watanabe et

al. (2003). Presently, performing of *in-situ* observations of phase changes based on the optical properties of polymers (Magill, 1961, 2001) under pressure is the object of a research project developed by Boyer (Boyer et al., 2011a).

To estimate the solubility of penetrating agents in polymers, four main approaches are currently generating various techniques and methods, namely: gravimetric techniques, oscillating techniques, pressure decay methods, and flow methods. However, with many existing experimental devices, the gain in weight of the polymer is measured whereas the associated volume change is either estimated or sometimes neglected (Hilic et al., 2000; Nalawade et al., 2006; Li et al., 2008).

The determination of key thermo-mechanical parameters coupled with diffusion and chemical effects together with temperature and pressure control is not yet well established. Approaches addressing the prediction of the multifaceted thermo-difffuso-chemo-mechanical (TDCM) behaviour are being suggested. Constitutive equations are built within a thermomechanical framework, like the relation based on a rigorous thermodynamic approach (Boyer et al., 2007), and the proposed formalism based on as well rigorous mechanical approach (Rambert et al., 2006; Baudet et al., 2009).

### 3. Development and optimization of pertinent models

Modelling of polymer phase transitions with a specific thermodynamics- and thermokinetics-based approach assumes to consider the coupling between thermal, diffusion, chemical and mechanical phenomena and to develop advanced physically-based polymer laws taking into account the morphologies and associated growth. This implies a twofold decisive step, theoretical and experimental.

As regards specific industrial and technological problems, from polymer formulation to polymer damage, passing by polymer processing, the conceptualization involves largely different size scales with extensive and smart experimentation to suggest and justify suitable approximations for theoretical analyses.

#### 3.1 Thermodynamics as a means to understand and prevent macro-scale changes and damages resulting from molten or solid polymer/solvent interactions

Thermodynamics is a useful and powerful means to understand and prevent polymer macro-scale changes and damages resulting from molten or solid material/solvent interactions. Two engineering examples are illustrative: foaming processes with hydrochlorofluorocarbons (HCFCs) as blowing agents in extrusion processes with a concern on safeguarding the ozone layer and the global climate system, Montreal Protocol (Dixon, 2011), and transport of petroleum fluids with in-service pipelines made of structural semi-crystalline polymers which are then exposed to explosive fluctuating fluid pressure (Dewimille et al., 1993).

##### Solubility and concomitant swelling of solvent-saturated molten polymer

In the prediction of the relevant thermo-difffuso-chemo-mechanical behaviour of polymers, sorption is the central phenomenon. Sorption is by nature complex, since the effects of fluids solubility in polymers and of the concomitant swelling of these polymers cannot be separated. To experimentally extract reliable solubility data, the development of inventive equipments is required. In an original way, dynamic pendulum technology under pressure is used. The advanced development proposes to combine the features of the vibrating-wire viscometer

with a high pressure decay technique, the whole setup being operated under a fine control of the temperature. The limits and performances of this mechanical setup under extreme conditions, *i.e.*, pressure and environment of fluid, were theoretically assessed (Boyer et al., 2007). In the working equation of the vibrating-wire sensor (VW) (eq. (1)), unknowns are both the mass  $m_{sol}$  of solvent absorbed in the polymer and the associated change in volume  $\Delta V_{pol}$  of the polymer due to the sorption.

$$m_{sol} = \rho_g \Delta V_{pol} + \left[ (\omega_B^2 - \omega_0^2) \frac{4 L^2 R^2 \rho_s}{\pi g} + \rho_g (V_C + V_{pol}) \right] \quad (1)$$

The volume of the degassed polymer is represented by  $V_{pol}$  and  $\rho_g$  is the density of the solvent. The other parameters are the physical characteristics of the wire, namely,  $\omega_0$  and  $\omega_B$  which represent the natural (angular) frequencies of the wire in vacuum and under pressure, respectively. And  $L$ ,  $R$ ,  $\rho_s$  are, respectively, the length, the radius and the density of the wire.  $V_C$  is the volume of the polymer container.

The thermodynamics of solvent-polymer interactions can be theoretically expressed with a small number of adjustable parameters. The currently used models are the 'dual-mode' model (Vieth et al., 1976), the cubic equation of state (EOS) as Peng-Robinson (Zhong & Masuoka, 1998) or Soave-Redlich-Kwong (Orbey et al., 1998) EOSs, the lattice-fluid model of Sanchez-Lacombe equation of state (SL-EOS) (Lacombe & Sanchez, 1976; Sanchez & Lacombe, 1976) with the extended equation of Doghieri-Sarti (Doghieri & Sarti, 1996; Sarti & Doghieri, 1998), and the Statistical Associating Fluid Theory (SAFT) (Prigogine et al., 1957; Beret & Prausnitz, 1975; Behme et al., 1999).

From the state of the art, the thermodynamic SL-EOS was preferably selected to theoretically estimate the change in volume of the polymer *versus* pressures and temperatures found in eq. (1). In this model, phase equilibria of pure components or solutions are determined by equating chemical potentials of a component in coexisting phases. It is based on a well-defined statistical mechanical model, which extends the basic Flory-Huggins theory (Panayiotou & Sanchez, 1991). Only one binary adjustable interaction parameter  $k_{12}$  has to be calculated by fitting the sorption data eqs. (2-4). In the mixing rule appears the volume fraction of the solvent (index 1,  $\phi_1$ ) in the polymer (index 2,  $\phi_2$ ), ( $\rho_1^*$ ,  $p_1^*$ ,  $T_1^*$ ) and ( $\rho_2^*$ ,  $p_2^*$ ,  $T_2^*$ ) being the characteristic parameters of pure compounds.

$$p^* = \phi_1 p_1^* + \phi_2 p_2^* - \phi_1 \phi_2 \Delta p^* \quad (2)$$

$$T^* = \frac{p^*}{\frac{\phi_1 p_1^*}{T_1^*} + \frac{\phi_2 p_2^*}{T_2^*}} \quad (3)$$

The parameter  $\Delta p^*$  characterizes the interactions in the mixture. It is correlated with the binary adjustable parameter  $k_{12}$ .

$$\Delta p^* = k_{12} \sqrt{p_1^* p_2^*} \quad (4)$$

The mass fraction of solvent (the permeant),  $\omega_1$ , at the thermodynamical equilibrium is calculated with eq. (5).

$$\omega_1 = \frac{\phi_1}{\phi_1 + (1 - \phi_1) \frac{\rho_2^*}{\rho_1^*}} \quad (5)$$

Coupled with the equation of DeAngelis (DeAngelis et al., 1999), the change in volume  $\Delta V_{pol}$  of the polymer is accessible via **eq. (6)**:

$$\frac{\Delta V_{pol}}{V_0} = \frac{1}{\tilde{\rho} \rho^* (1 - \omega_1)} \frac{1}{\hat{v}_2^0} \quad (6)$$

$\rho^*$  and  $\tilde{\rho}$  are the mixture characteristic and reduced densities, respectively.  $\hat{v}_2^0$  is the specific volume of the pure polymer at fixed  $T$ ,  $P$  and composition. The correlation with the model is done in conjunction with the optimization of the parameter  $k_{12}$  that minimizes the Average of Absolute Deviations (AAD) between the experimental results and the results recalculated from the fit.

The critical comparison between the semi-experimental (or semi-theoretical) data of solubility and pure-experimental data available in the literature allows us to validate the consistency of the methodology of the calculations. The combination of coupled experimental and calculated data obtained from the vibrating-wire and theoretical analyses gives access to original solubility data that were not up to now available for high pressure in the literature. As an illustration in **Fig. 2.a-b** is given the solubility of carbon dioxide (CO<sub>2</sub>) and of 1,1,1,2-tetrafluoroethane (HFC-134a) in molten polystyrene (PS). HFC-134a is significantly more soluble in PS by a factor of two compared to CO<sub>2</sub>. The parameter  $k_{12}$  was estimated at 0.9232, 0.9342, 0.9140 and 0.9120 for CO<sub>2</sub> sorption respectively at 338, 362, 383 and 402 K. For HFC-134a sorption, it was estimated at 0.9897 and 0.9912 at 385 and 402 K, respectively. The maximum of the polymer volume change was in CO<sub>2</sub> of 13 % at 25 MPa and 338 K, 15 % at 25 MPa and 363 K, 14 % at 43 MPa and 383 K, 13 % at 44 MPa and 403 K, and in HFC-134a of 12 % at 16 MPa and 385K, 11 % at 20 MPa and 403 K. The thermodynamic behaviour of {PS-permeant} systems with temperature is comparable to a lower critical solution temperature (LCST) behaviour (Sanchez & Lacombe, 1976).

From these data, the aptitude of the thermodynamic SAFT EOS to predict the solubility of carbon dioxide and of 1,1,1,2-tetrafluoroethane (HFC-134a) in polystyrene (PS) is evaluated. The use of SAF theoretical model is rather delicate because the approach uses a reference fluid that incorporates both chain length (molecular size and shape) and molecular association. SAF Theory is then defined in terms of the residual Helmholtz energy  $a^{res}$  per mole. And  $a^{res}$  is represented by a sum of three intermolecular interactions, namely, segment-segment interactions, covalent chain-forming bonds among segments and site-site interactions such as hydrogen bond association. The SAFT equation satisfactorily applies for CO<sub>2</sub> dissolved in PS with a molecular mass in weight near about 1000 g.mol<sup>-1</sup>, while it is extended to HFC-134a dissolved in PS with a low molecular mass in weight.

### **Global cubic expansion coefficient of solvent saturated polymer as thermo-diffusio-chemo-mechanical parameter for preferential control of solid polymer/solvent interactions**

An essential additional information to solubility quantification, in direct relation with polymer damage by dissolved gases, is the expansion coefficient of the gas saturated polymer, *i.e.*, the mechanical cubic expansion coefficient of the polymer saturated in a solvent,  $\alpha_{pol-g-int}$ .

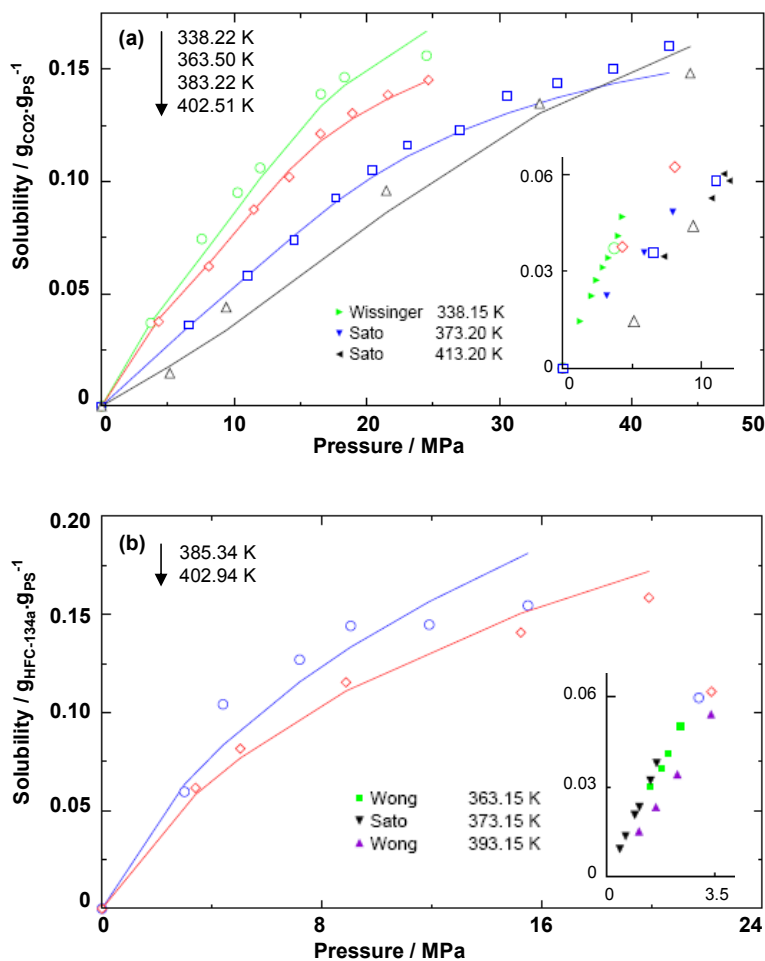


Fig. 2. Solubility of (a) CO<sub>2</sub> (critical pressure ( $P_c$ ) of 7.375 MPa, critical temperature ( $T_c$ ) of 304.13 K) and (b) HFC-134a ( $P_c$  of 4.056 MPa,  $T_c$  of 374.18 K) in PS with (a-insert) literature data from pressure decay measurement (Sato et al., 1996, pressure up to 20 MPa), from elongation measurement (Wissinger & Paulaitis, 1987, pressure up to 5 MPa), and (b-insert) literature data from volumetric measurement (Sato et al., 2000, pressure up to 3 MPa), from gravimetry (Wong et al., 1998, pressure up to 4 MPa). The correlation of CO<sub>2</sub> and HFC-134a solubility in PS with SAFT is illustrated with solid lines.

A precise experimental methodology and a mathematical development proposed by Boyer (Boyer et al., 2006b, 2007) use the thermodynamic approach of high-pressure-controlled scanning transitionometry (PCST) (Grolier et al., 2004; Bessières et al., 2005). The heat resulting from the polymer/solvent interactions is measured during pressurization/depressurization runs performed under isothermal scans. Several binary polymer/fluid systems with a more or less reactive pressurizing medium have been investigated with a view to illustrate the

importance of dissociating the purely hydrostatic effect from the fluid sorption over an extended pressure range.

Taking advantage of the differential mounting of the high pressure calorimetric detector and the proper use of the thermodynamic Maxwell's relation  $(\partial S / \partial P)_T = -(\partial V / \partial T)_P$ , a practical expression of the global cubic expansion coefficient  $\alpha_{pol-g-int}$  of the saturated polymer subjected to the compressed penetrating (permeant) solvent under isothermal conditions has been established as follows by eq. (7):

$$\alpha_{pol-g-int} = \frac{(Q_{diff, SS} - Q_{diff, pol}) + V_{SS, r} \alpha_{SS} T \Delta P}{V_{pol} T \Delta P} \quad (7)$$

$\alpha_{SS}$  is the cubic expansion coefficient of the stainless steel of which are made the cells.  $V_{pol}$  and  $V_{SS}$  are the volumes of the polymer sample placed in the measuring cell and of the stainless steel (reference) sample placed in the reference cell, respectively. The stainless steel sample is identical in volume to the initial polymer sample.  $Q_{diff, pol}$  is the differential heat between the measuring cell and the reference cell.  $Q_{diff, SS}$  is the measure of the thermodynamic asymmetry of the cells.  $\Delta P$  is the variation of gas-pressure during a scan at constant temperature  $T$ .

Three quite different pressure transmitting fluids, as regards their impact on a given polymer, have been selected: *i*) mercury (Hg), inert fluid, with well-established thermo-mechanical coefficients inducing exclusively hydrostatic effect, *ii*) a non-polar medium nitrogen ( $N_2$ ) qualified as "poor" solvent, and *iii*) "chemically active" carbon dioxide ( $CO_2$ ) (Glasser, 2002; Nalawade et al., 2006). While maintaining the temperature constant, the independent thermodynamic variables  $P$  or  $V$  can be scanned. Optimization and reliability of the results are verified by applying fast variations of pressure ( $P$  jumps), pressure scans ( $P$  scans) and volume scans ( $V$  scans) during pressurization and depressurization. Additionally, taking advantage of the differential arrangement of the calorimetric detector the comparative behaviour of two different polymer samples subjected to exactly the same supercritical conditions can be documented. As such, three main and original conclusions for quantifying the thermo-difffuso-chemo-mechanical behaviour of two polymers, a polyvinylidene fluoride (PVDF) and a medium density polyethylene (MDPE) with similar volume fraction of amorphous phase, can be drawn. This includes the reversibility of the solvent sorption/desorption phenomena, the role of the solvent (the permeant) state, *i.e.*, gaseous or supercritical state, the direct thermodynamic comparison of two polymers in real conditions of use.

The reversibility of the sorption/desorption phenomena is well observed when experiments are performed at the thermodynamic equilibrium, *i.e.*, at low rate volume scans. The preferential polymer/solvent interaction, when solvent is becoming a supercritical fluid, is emphasized with respect to the competition between plasticization and hydrostatic pressure effects. In the vicinity of the critical point of the solvent, a minimum of the  $\alpha_{pol-g-int}$  coefficient is observed. It corresponds to the domain of pressure where plasticization due to the solvent sorption is counterbalanced by the hydrostatic effect of the solvent. The significant influence of the 'active' supercritical  $CO_2$  is illustrated by more energetic interactions with PVDF than with MDPE at pressure below 30 MPa (Boyer et al., 2009). The hetero polymer/ $CO_2$  interactions appear stronger than the homo interactions between molecular chains. PVDF more easily dissolves  $CO_2$  than MDPE, the solubility being favoured by the presence of polar groups C-F

in the PVDF chain (Flaconnèche et al., 2001). This easiness for CO<sub>2</sub> to dissolve is observed at high pressure where the parameter  $\alpha_{pol-g-int}$  is smaller for highly condensed {PVDF-CO<sub>2</sub>} systems than for less condensed {MDPE-CO<sub>2</sub>} system (Boyer et al., 2007).

With the objective to scrutinize the complex interplay of the coupled diffusive, chemical and mechanical parameters under extreme conditions of  $P$  and  $T$ , thermodynamics plays a pivotal role. Precise experimental approaches are as crucial as numerical predictions for a complete understanding of polymer behaviour in interactions with a solvent.

### 3.2 Thermodynamics as a means to understand and control nanometric scale length patterns using preferential liquid-crystal polymer/solvent interactions

Thermodynamics is ideally suited to obtain specific nano-scale pattern formation, for instance 'selective decoration' of arrayed polymer structure through selected additives, by controlling simultaneously the phase diagrams of fluids and of semi-crystalline polymers.

The creation of hybrid metal-polymer composite materials, with a well-controlled structure organization at the nanometric scale, is of great practical interest (Grubbs, 2005; Hamley, 2009), notably for the new generation of microelectronic and optical devices. Inorganic nanoparticles possess unique size dependent properties, from electronic, optical to magnetic properties. Among them, noble gold nanoparticles (AuNPs) are prominent. Included into periodic structures, inorganic nanoparticles can potentially lead to new collective states stemming from precise positioning of the nanoparticles (Tapalin et al., 2009). When used as thin organic smart masks, block copolymers make ideal macromolecular templates. Especially, the unique microphase separated structure of asymmetric liquid-crystal (LC) diblock copolymer (BC), like PEO-*b*-PMA(Az), develops itself spontaneously by self assemblage to form PEO channels hexagonally packed (Tian et al., 2002; Watanabe et al., 2008). PEO<sub>*m*</sub>-*b*-PMA(Az)<sub>*n*</sub> amphiphilic diblock copolymer consists of hydrophilic poly(ethylene oxide) (PEO) entity and hydrophobic poly(methacrylate) (PMA) entity bearing azobenzene mesogens (Az) in the side chains, where  $m$  and  $n$  denote the degrees of polymerization of PEO and of photoisomerized molecules azobenzene moieties, respectively. By varying  $m$  and  $n$ , the size of the diameters of PEO cylinders is controlled from 5 to 10 nm while the distance between the cylinders is 10 to 30 nm. Four phase transitions during BC heating are ascribed to PEO crystal melting, PMA(Az) glass transition, liquid crystal transition from the smectic C (SmC) phase to the smectic A (SmA) phase and isotropic transition (Yoshida et al., 2004). In PEO<sub>114</sub>-*b*-PMA(Az)<sub>46</sub>, the temperatures of the transitions are about 311, 339, 368 and 388 K, respectively.

As such, for creating smart and noble polymer-metal hybrids possessing a structure in the nanometric domain, three original aspects are discussed. They include the initial thermodynamic polymer/pressure medium interaction, the modulation of the surface topology concomitantly with the swelling of the solvent-modified nano-phase-separated organization, the "decorative" particles distribution modulation. All the aspects have an eco-aware issue and they are characterized through a rigorous analysis of the specific interactions taking place in LC/solvent systems.

#### Polymer/pressurizing fluid interactions

The isobaric temperature-controlled scanning transitionometry (TCST) (Grolier et al., 2004; Bessières et al., 2005) is used to investigate the phase changes via the Clapeyron's equation while the pressure is transmitted by various fluids. The enthalpy, volume and entropy

changes are quantified *versus* the (high) pressure of either Hg, CO<sub>2</sub>, or N<sub>2</sub> (Yamada et al., 2007a-b). The hydrostatic effect of “more or less chemically active” solvent CO<sub>2</sub>, or N<sub>2</sub> is smaller than the hydrostatic effect of mercury. The adsorbed solvent induces smaller volume changes at the isotropic transition than the mercury pressure. This results from the low compressibility of solvent (gas) molecules compared to the free volume compressibility induced in BC. A particular behaviour is observed with “chemically active” CO<sub>2</sub> where the quadrupole-dipole interactions favour the CO<sub>2</sub> sorption into the PMA(Az) matrix during the isotropic liquid transition (Kamiya et al., 1998; Vogt et al., 2003). The hydrostatic effect by CO<sub>2</sub> overcomes above 40 MPa with a CO<sub>2</sub> desorption at higher pressures explained by the large change of molecular motions at the isotropic transition upon the disruption of  $\pi$ -bounds with azobenzene moieties.

### **Modulation of the surface topology and swelling of the CO<sub>2</sub>-modified nanometric-phase-separated organization**

Supercritical carbon dioxide (SCCO<sub>2</sub>) constitutes an excellent agent of microphase separation. From *ex-situ* Atomic Force Microscopy (AFM) and Transmission Electron Microscopy (TEM) analysis of the pattern organization, the fine control of the pressure together with the temperature at which the CO<sub>2</sub> treatment is achieved demonstrates the possibility to modulate the surface topology inversion between the copolymer phases concomitantly with the swelling of the nano-phase-separated organization. The observed phase contrast results from the coupled effect of the different elastic moduli of the two domains of the block-copolymer with chemo-diffuso phenomenology.

Remarkably, the preferential CO<sub>2</sub> affinity is associated with the thermodynamic state of CO<sub>2</sub>, from liquid (9 MPa, room temperature (r.t.)) to supercritical (9 MPa, 353 K) and then to gaseous (5 MPa, r.t.) state (Glasser, 2002). This is typically observed when annealing the copolymer for 2 hours to keep the dense periodic hexagonal honeycomb array (**Fig. 3.a-d**). Under gaseous CO<sub>2</sub>, the surface morphology of PEO cylinders is not significantly expanded (**Fig. 3.a-b**). However, liquid CO<sub>2</sub> induces a first drastic shift at the surface with the emergence of a new surface state of PEO cylinders. This surface state inversion is attributed to domain-selective surface disorganization. PMA(Az) in the glassy smectic C (SmC) phase cannot expand. PEO cylinders dissolve favourably within liquid CO<sub>2</sub>, with polar interactions, get molecular movement, expand preferentially perpendicularly to the surface substrate (**Fig. 3.c**). By increasing temperature, liquid CO<sub>2</sub> changes to supercritical CO<sub>2</sub>. The PMA(Az) domain is in the SmC phase and get potential molecular mobility. At this stage, the copolymer chains should be easily swelled. The easiness of SCCO<sub>2</sub> to dissolve within liquid PEO cylinders deals with a new drastic change of the surface topology where the absorbed SCCO<sub>2</sub> increases the diameter of the PEO nano-tubes (**Fig. 3.d**).

The preferential CO<sub>2</sub> affinities produce porous membranes with a selective sorption in hydrophilic semicrystalline ‘closed loop’, *i.e.*, PEO channels (Boyer et al., 2006a). More especially, under supercritical SCCO<sub>2</sub>, the PEO cylinders kept in the ordered hexagonal display exhibit the highest expansion in diameter. In the case of PEO<sub>114-b</sub>-PMA(Az)<sub>46</sub>, the exposure to SCCO<sub>2</sub> swells the PEO cylinders by 56 %, with arrays from 11.8 nm in diameter at r.t. to 18.4 nm in diameter at 353 K. The lattice of the PMA matrix, *i.e.*, periodic plane distance between PEO cylinders, slightly increases by 26 %, from 19.8 nm at r.t. to 24.9 nm at 353 K. This microphase separation is driven by disparity in free volumes between dissimilar segments of the polymer chain, as described from the entropic nature of the closed-loop miscibility gap (Lavery et al., 2006; Yamada et al., 2007a-b).



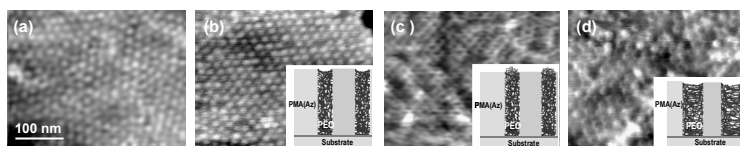


Fig. 3. Pattern control in the nanometric scale under multifaceted  $T$ ,  $P$  and  $\text{CO}_2$  constraints, 2 hrs annealed. AFM phase, tapping mode, illustrations on silicon substrate (a) neat  $\text{PEO}_{114}\text{-}b\text{-PMA}(\text{Az})_{46}$ , PEO 'softer' than  $\text{PMA}(\text{Az})$  appears brighter (whiter), (b)  $\text{GCO}_2$  saturation (5 MPa, r.t.), (c)  $\text{LCO}_2$  saturation (9 MPa, r.t.),  $\text{PMA}(\text{Az})$  surrounding PEO becomes 'softer', (d)  $\text{SCCO}_2$  saturation (9 MPa, 353 K), PEO becomes 'softer' while swelling. Inserts (b-c-d) are schematic representations of  $\text{CO}_2$ -induced changes of PEO cylinders. (BC film preparation before modification: 2 wt% toluene solution spin-coating, 2000 rpm, annealing at 423 K for 24 hrs in vacuum.)

### Modulation of the decorative particles distribution

To create nano-scale hybrid of metal-polymer composites, the favourable  $\text{SCCO}_2/\text{PEO}$  interactions are advantageously exploited, as illustrated in Fig. 4.a-b. They enable a tidy pattern of hydrophilic gold nano-particles (AuNPs). AuNPs are of about 3 nm in diameter and stabilized with thiol end-functional groups (Boal & Rotello, 2000). Preferentially, the metal NPs wet one of the two copolymer domains, the PEO channels, but de-wet the other, the  $\text{PMA}(\text{Az})$  matrix. This requires a high mobility contrast between the two copolymer domains, heightened by  $\text{CO}_2$  plasticization that enhances the free volume disparity between copolymer parts. Each  $\text{SCCO}_2$ -swollen PEO hydrophilic hexagonal honeycomb allows the metal NPs to cluster. A two-dimensional (2D) periodic arrangement of hydrophilic AuNPs is generated in the organic PEO in turn confined into smectic C phase of  $\text{PMA}(\text{Az})$  matrix which has potential molecular mobility. Additionally to the plasticizing action, the force of the trap is driving chemically. It is due to the hydrophilic compatibility of AuNPs in PEO cylinders by grafted polar groups (Watanabe et al., 2007).

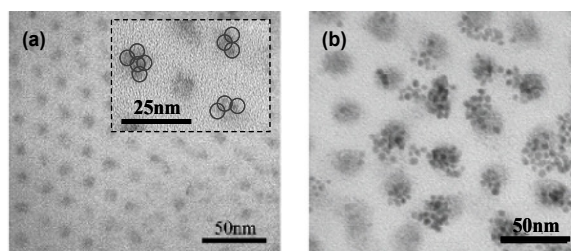


Fig. 4. Pattern control in the nanometric scale of  $\text{PEO}\text{-}b\text{-PMA}(\text{Az})$  under multifaceted  $T$ ,  $P$ ,  $\text{CO}_2$  constraints with AuNPs. TEM illustrations of BC on carbone coated copper grid (a)  $\text{PEO}_{114}\text{-}b\text{-PMA}(\text{Az})_{46}$ , (b)  $\text{PEO}_{454}\text{-}b\text{-PMA}(\text{Az})_{155}$  doped with AuNPs under  $\text{SCCO}_2$  (9 MPa, 353 K). Black spots are AuNPs wetted hexagonal PEO honeycomb, selectively. PEO is (a) 8.6, (b) 24.3 nm in diameter with a periodicity of (a) 17.1, (b) 36.6 nm. (Step 1, BC film preparation before modification: 2 wt% toluene solution solvent-casting, annealing at 423 K for 24 hrs in vacuum. Step 2, AuNPs deposition before modification: droplet of an ethanol solution of hydrophilic AuNPs (solvent in toluene of 1 %) on dried BC film, drying at r.t. for 2 hrs.)

The local affinities of AuNPs with PEO/SCCO<sub>2</sub> stabilize the thermodynamically unstable SCCO<sub>2</sub>-plasticized network and keep it stable with time, which cannot be observed without the insertion of gold nano-particles mainly because of diffusion effect of the solvent (Boyer et al., 2006a). The mean height of AuNPs layer is about 3 nm, which is 20 times smaller than PEO cylinders with a 60 nm in length. Thus PEO channels could be considered as nano-dots receptors, schematically as a “compact core-shell model” consisting of a spherical or isotropic AuNP “core” embedded into a PEO channel “shell”, consequently leading to isotropic two- and three-dimensional materials. Nicely, AuNPs clusters on PEO channel heads can be numerically expressed. The presence of, 4, 5 and 8 single Au nano-clusters for  $m = 114, 272$  and 454 is identified, respectively. It represents a linear function between the number of AuNPs on swollen PEO *versus* SCCO<sub>2</sub>-swollen diameter with half of ligands of AuNPs linked with PEO polymer chain.

From this understanding, a fine thermodynamic-mechanical control over extended  $T$  and  $P$  ranges would provide a precious way to produce artificial and reliable nanostructured materials. SCCO<sub>2</sub>-based technology guides a differential diffusion of hydrophilic AuNPs to cluster selectively along the hydrophilic PEO scaffold. As a result, a highly organized hybrid metal-polymer composite is produced. Such understanding would be the origin of a 2D nanocrystal growth.

### 3.3 Thermokinetics as a means to control macrometric length scale molecular organizations through molten to solid transitions under mechanical stress

A newly developed phenomenological model for pattern formation and growth kinetics of polymers uses thermodynamic parameters, as thermo-mechanical constraints and thermal gradient. It is a system of physically-based morphological laws-taking into account the kinetics of structure formation and similarities between polymer physics and metallurgy within the framework of Avrami’s assumptions.

Polymer crystallization is a coupled phenomenon. It results from the appearance (nucleation in a more or less sporadic manner) and the development (growth) of semi-crystalline entities (*e.g.*, spherulites) (Gadomski & Luczka, 2000; Panine et al., 2008). The entities grow in all available directions until they impinge on one another. The crystallization kinetics is described in an overall manner by the fraction  $\alpha(t)$  (surface fraction in two dimensions (2D) or volume fraction in three dimensions (3D)) transformed into morphological entities (disks in 2D or spheres in 3D) at each time  $t$ .

The introduction of an overall kinetics law for crystallization into models for polymer processing is usually based on the Avrami-Evans’s (AE) theory (Avrami, 1939, 1940, 1941; Evans, 1945). To treat non-isothermal crystallization, simplifying additional assumptions have often been used, leading to analytical expressions and allowing an easy determination of the physical parameters, *e.g.*, Ozawa (1971) and Nakamura et al. (1972) approaches. To avoid such assumptions, a trend is to consider the general AE equation, either in its initial form as introduced by Zheng & Kennedy (2004), or after mathematical transformations as presented by Haudin & Chenot (2004) and recalled here after.

#### General equations for quiescent crystallization

The macroscopic mechanism for the nucleation event proposed by Avrami remains the most widely used, partly because of its firm theoretical basis leading to analytical mathematical equations. In the molten state, there exist zones, the potential nuclei, from which the crystalline phase is likely to appear. They are uniformly distributed throughout the melt,

with an initial number per unit volume (or surface)  $N_0$ .  $N_0$  is implicitly considered as constant. The potential nuclei can only disappear during the transformation according to activation or absorption (“swallowing”) processes. An activated nucleus becomes a growing entity, without time lag. Conversely, a nucleus which has been absorbed cannot be activated any longer. In the case of a complex temperature history  $T(t)$ , the assumption of a constant number of nuclei  $N_0$  is no more valid, because  $N_0 = N_0(T) = N_0(T(t))$  may be different at each temperature. Consequently, additional potential nuclei can be created in the non-transformed volume during a cooling stage. All these processes are governed by a set of differential equations (Haudin & Chenot, 2004), differential equations seeming to be most suitable for a numerical simulation (Schneider et al., 1988).

**Avrami’s Equation**

Avrami’s theory (Avrami, 1939, 1940, 1941) expresses the transformed volume fraction  $\alpha(t)$  by the general differential equation **eq. (8)**:

$$\frac{d\alpha(t)}{dt} = (1 - \alpha(t)) \frac{d\tilde{\alpha}(t)}{dt} \tag{8}$$

$\tilde{\alpha}(t)$  is the “extended” transformed fraction, which, for spheres growing at a radial growth rate  $G(t)$ , is given by **eq. (9)**:

$$\tilde{\alpha}(t) = \frac{4\pi}{3} \int_0^t \left[ \int_{\tau}^t G(u) du \right]^3 \frac{d\tilde{N}_a(\tau)}{d\tau} d\tau \tag{9}$$

$d\tilde{N}_a(t)/dt$  is the “extended” nucleation rate,  $\frac{4\pi}{3} \left[ \int_{\tau}^t G(u) du \right]^3$  is the volume at time  $t$  of a sphere appearing at time  $\tau$ , and  $d\tilde{N}_a(\tau)$  are spheres created per unit volume between  $\tau$  and  $\tau + d\tau$ .

**Assumptions on Nucleation**

The number of potential nuclei decreases by activation or absorption, and increases by creation in the non-transformed volume during cooling. All these processes are governed by the following equations:

$$\frac{dN(t)}{dt} = -\frac{dN_a(t)}{dt} - \frac{dN_c(t)}{dt} + \frac{dN_g(t)}{dt} \tag{10a}$$

$$\frac{dN_a(t)}{dt} = q(t)N(t) \tag{10b}$$

$$\frac{dN_c(t)}{dt} = \frac{N(t)}{1 - \alpha(t)} \frac{d\alpha(t)}{dt} \tag{10c}$$

$$\frac{dN_g(t)}{dt} = (1 - \alpha(t)) \frac{dN_0(T)}{dT} \frac{dT}{dt} \tag{10d}$$

$N(t), N_a(t), N_c(t), N_g(t)$  are the number of potential, activated, absorbed and generated (by cooling) nuclei per unit volume (or surface) at time  $t$ , respectively.  $q(t)$  is the activation frequency of the nuclei at time  $t$ . The “extended” quantities  $\tilde{N}, \tilde{N}_a$  are related to the actual ones by:

$$N = (1 - \alpha)\tilde{N} \quad (11a)$$

$$\frac{dN_a}{dt} = qN = (1 - \alpha)\frac{d\tilde{N}_a}{dt} \quad (11b)$$

### The System of Differential Equations

The crystallization process equations are written into a non-linear system of six, **eqs. (12, 13a, 14-17)**, or seven, **eqs. (12, 13b, 14-18)**, differential equations in 2D or 3D conditions, respectively (Haudin & Chenot, 2004):

$$\frac{dN}{dt} = -N \left( q + \frac{1}{1 - \alpha} \frac{d\alpha}{dt} \right) + (1 - \alpha) \frac{dN_0(T)}{dT} \frac{dT}{dt} \quad (12)$$

$$\frac{d\alpha}{dt} = 2\pi(1 - \alpha)G(F\tilde{N}_a - P) \quad (13a)$$

$$\frac{d\alpha}{dt} = 4\pi(1 - \alpha)G(F^2\tilde{N}_a - 2FP + Q) \quad (13b)$$

$$\frac{dN_a}{dt} = qN \quad (14)$$

$$\frac{d\tilde{N}_a}{dt} = \frac{qN}{1 - \alpha} \quad (15)$$

$$\frac{dF}{dt} = G \quad (16)$$

$$\frac{dP}{dt} = F \frac{d\tilde{N}_a}{dt} = F \frac{qN}{1 - \alpha} \quad (17)$$

$$\frac{dQ}{dt} = F^2 \frac{d\tilde{N}_a}{dt} = F^2 \frac{qN}{1 - \alpha} \quad (18)$$

The initial conditions at time  $t = 0$  are:

$$N(0) = N_0$$

$$\alpha(0) = N_a(0) = \tilde{N}_a(0) = F(0) = P(0) = Q(0) = 0 \quad (19)$$

$F, P$  and  $Q$  are three auxiliary functions added to get a first-order ordinary differential system. The model needs three physical parameters, the initial density of potential nuclei  $N_0$ ,

the frequency of activation  $q$  of these nuclei and the growth rate  $G$ . In isothermal conditions, they are constant. In non-isothermal conditions, they are defined as temperature functions, e.g.:

$$N_0 = N_{00} \exp(-N_{01}(T - T_0)) \quad (20a)$$

$$q = q_0 \exp(-q_1(T - T_0)) \quad (20b)$$

$$G = G_0 \exp(-G_1(T - T_0)) \quad (20c)$$

### General equations for shear-induced crystallization

Crystallization can occur in the form of spherulites, shish-kebabs, or both. The transformed volume fraction is written as (Haudin et al., 2008):

$$\frac{d\alpha(t)}{dt} = \frac{d\beta(t)}{dt} + \frac{d\kappa(t)}{dt} \quad (21)$$

$\beta(t)$  and  $\kappa(t)$  are the thermo-dependent volume fractions transformed *versus* time into spherulites and into shish-kebabs, respectively.

### Spherulitic Morphology

Modification of eqs. (8) and (10a) gives:

$$\frac{d\beta(t)}{dt} = (1 - \alpha(t)) \frac{d\tilde{\beta}(t)}{dt} \quad (22)$$

$$\frac{dN(t)}{dt} = -\frac{dN_a(t)}{dt} - \frac{dN_c(t)}{dt} + \frac{dN_s(t)}{dt} + \frac{dN_\gamma(t)}{dt} \quad (23)$$

$\beta(t)$  and  $\tilde{\beta}(t)$  are the actual and extended volume fractions of spherulites, respectively.  $N_\gamma(t)$  is the number of nuclei per unit volume generated by shear. Two situations are possible, i.e., crystallization occurs after shear or crystallization occurs during shear.

If crystallization during shear remains negligible, the number of shear-generated nuclei is:

$$\frac{dN_\gamma}{dt} = a\dot{\gamma}(A - N_\gamma) \text{ if } a\dot{\gamma}(A - N_\gamma) \geq 0 \quad (24a)$$

$$\frac{dN_\gamma}{dt} = 0 \text{ if } a\dot{\gamma}(A - N_\gamma) \leq 0 \quad (24b)$$

$a$  and  $A_1$  are material parameters, eventually thermo-dependent. As a first approximation,  $A = A_1\dot{\gamma}$ , with  $\dot{\gamma}$  the shear rate.

If crystallization proceeds during shear, only the liquid fraction is exposed to shear and the shear rate  $\dot{\gamma}'$  is becoming:

$$\dot{\gamma}' = \dot{\gamma} / (1 - \alpha)^{1/3} \quad (25)$$

By defining  $\tilde{N}_\gamma$  as the extended number of nuclei per unit volume generated by shear in the total volume, then:

$$\frac{d\tilde{N}_\gamma}{dt} = a\dot{\gamma}'(A_1\dot{\gamma}' - \tilde{N}_\gamma) \quad (26)$$

The number  $N_\gamma$  of nuclei generated by shear in the liquid fraction is:

$$N_\gamma = (1 - \alpha)\tilde{N}_\gamma \quad (27)$$

Under shear, the activation frequency of the nuclei increases. If the total frequency is the sum of a static component,  $q_{st}$ , function of temperature, and of a dynamic one,  $q_{flow}$ , then:

$$q = q_{st} + q_{flow} \quad (28)$$

$q_{flow}$  is given by eq. (29) where as a first approximation  $q_2 = q_{02}\dot{\gamma}$  and  $q_3$  is constant.

$$q_{flow} = q_2(1 - \exp(-q_3\dot{\gamma})) \quad (29)$$

The system of differential equations (12, 13b, 14-18) is finally replaced by a system taking the influence of shear into account through the additional unknown  $N_\gamma$  and through the dynamic component of the activation frequency  $q_{flow}$ . Two cases are considered, *i.e.*, crystallization occurs after shear (37a) or crystallization occurs under (37b) shear.

$$\frac{dN}{dt} = -N \left( q + \frac{1}{1 - \alpha} \frac{d\alpha}{dt} \right) + (1 - \alpha) \frac{dN_0(T)}{dT} \frac{dT}{dt} + \frac{dN_\gamma}{dt} \quad (30)$$

$$\frac{d\beta}{dt} = 4\pi(1 - \alpha)G(F^2\tilde{N}_a - 2FP + Q) \quad (31)$$

$$\frac{dN_a}{dt} = qN \quad (32)$$

$$\frac{d\tilde{N}_a}{dt} = \frac{qN}{1 - \alpha} \quad (33)$$

$$\frac{dF}{dt} = G \quad (34)$$

$$\frac{dP}{dt} = F \frac{d\tilde{N}_a}{dt} = F \frac{qN}{1 - \alpha} \quad (35)$$

$$\frac{dQ}{dt} = F^2 \frac{d\tilde{N}_a}{dt} = F^2 \frac{qN}{1 - \alpha} \quad (36)$$

$$\frac{dN_\gamma}{dt} = a\dot{\gamma}(A_1\dot{\gamma} - N_\gamma) \quad (37a)$$

$$\frac{dN_\gamma}{dt} = a\dot{\gamma} \left( (1-\alpha)^{1/3} A_1\dot{\gamma} - \frac{N_\gamma}{(1-\alpha)^{1/3}} \right) - \frac{N_\gamma}{1-\alpha} \frac{d\alpha}{dt} \quad (37b)$$

The initial conditions at time  $t = 0$  are:

$$\begin{aligned} N(0) &= N_0 \\ \alpha(0) &= N_a(0) = \tilde{N}_a(0) = F(0) = P(0) = Q(0) = 0 \\ N_\gamma(0) &= 0 \end{aligned} \quad (38)$$

### Shish-Kebab Morphology

Firstly are introduced the notions of real and extended transformed volume fractions of shish-kebab,  $\kappa$  and  $\tilde{\kappa}$ , respectively. Both are related by **eq. (39)**:

$$\frac{d\kappa(t)}{dt} = (1-\alpha) \frac{d\tilde{\kappa}(t)}{dt} \quad (39)$$

$\alpha(t)$  is the total transformed volume fraction for both spherulitic and oriented phases. Shish-kebabs are modelled as cylinders with an infinite length. The growth rate  $H$  is deduced from the radius evolution of the cylinder. The general balance of the number of nuclei for the oriented structure is given as:

$$\frac{dM(t)}{dt} = -\frac{dM_a(t)}{dt} - \frac{dM_c(t)}{dt} + \frac{dM_\gamma(t)}{dt} \quad (40)$$

$M(t), M_a(t), M_c(t), M_\gamma(t)$  are the numbers of potential, activated, absorbed and generated (by shear) nuclei per unit volume, respectively. In the same way as for the spherulitic morphology, a set of differential equations can be defined where  $w$  is the activation frequency of the nuclei,  $b$  and  $B_1$  the material parameters:

$$\begin{aligned} \frac{dM}{dt} &= -M \left( w + \frac{1}{1-\alpha} \frac{d\alpha}{dt} \right) \\ &+ b\dot{\gamma} \left( (1-\alpha)^{1/3} B_1\dot{\gamma} - \frac{M}{(1-\alpha)^{1/3}} \right) - \frac{M}{1-\alpha} \frac{d\alpha}{dt} \end{aligned} \quad (41)$$

$$\frac{d\kappa}{dt} = 2\pi(1-\alpha)H(R\tilde{M}_a - S) \quad (42)$$

$$\frac{dM_a}{dt} = wM \quad (43)$$

$$\frac{d\tilde{M}_a}{dt} = \frac{wM}{1-\alpha} \quad (44)$$

$$\frac{dR}{dt} = H \quad (45)$$

$$\frac{dS}{dt} = R \frac{d\tilde{M}_a}{dt} = R \frac{wM}{1-\alpha} \quad (46)$$

$F, P, Q, R$  and  $S$  are five auxiliary functions giving a first-order ordinary differential system. The initial conditions at time  $t = 0$  are:

$$M(0) = M_0$$

$$\kappa(0) = M_a(0) = \tilde{M}_a(0) = R(0) = S(0) = 0 \quad (47)$$

### Inverse resolution method for a system of differential equations

The crystallization, and especially the nucleation stage, is by nature a statistical phenomenon with large discrepancies between the sets of experimental data. The analytical extraction of the relevant crystallization parameters must be then considered as a multi-criteria optimization problem. As such the Genetic Algorithm Inverse Method is considered. The Genetic Algorithm Inverse Method is a stochastic optimization method inspired from the Darwin theory of nature survival (Paszkowicz, 2009). In the present work, the Genetic Algorithm developed by Carroll (Carroll, "FORTRAN Genetic Algorithm Front-End Driver Code", site: <http://cuaerospace.com/ga>) is used (Smirnova et al., 2007; Haudin et al., 2008). The vector of solutions is represented by a parameter  $Z$ . In quiescent crystallization (eqs. 20a-c),  $Z = [N_{00}, N_{01}, q_0, q_1, G_0, G_1]$  with  $N_{00}, N_{01}, q_0, q_1, G_0, G_1$  the parameters of non-isothermal crystallization for a spherulitic morphology. In shear-induced crystallization,  $Z = [N_{00}, N_{01}, q_0, q_1, q_{02}, q_3, G_0, G_1, M_0, w, H, A_1, a, B_1, b]$  with  $(q_{02}, q_3, A_1, a)$  the parameters of shear-induced crystallization for a spherulitic morphology (eqs. 26,29) and  $(M_0, w, H, B_1, b)$  the parameters of shear-induced crystallization for an oriented, like shish-kebab, morphology (eqs. 41,43,45,47).

The optimization is applied to the experimental evolution of the overall kinetics coupled with one kinetic parameter at a lower scale, the number of entities (density of nucleation  $N_a(t)$ ). The system of differential equations is solved separately for each experimental set and gives the evolutions of  $\alpha(t)$  and of the nuclei density defining a corresponding data file. The optimization function  $Q_{total}$  is expressed as the sum of the mean square errors of the transformed volume fraction  $Q_\alpha$  and of the number of entities  $Q_{N_a}$ .

### Model-experiment-optimization confrontation

The structure development parameters are identifiable by using the optical properties of the crystallizing entities. The experimental investigations and their analysis are done thanks to crossed-polarized optical microscopy (POM) (Magill, 1962, 1962, 2001) coupled with optically transparent hot stages, a home-made sliding plate shearing device and a rotating parallel plate shearing device (e.g., Linkam). Data accessible directly are: *i*) the evolution of the transformed fraction  $\alpha(t)$ , and the number of activated nuclei  $N_a(t)$ , *ii*) the approximate values of the initial number of potential nuclei  $N_0(T)$ , activation frequency  $q(T)$ , and growth rate  $G(T)$  for isothermal conditions and their functions of temperature for non-isothermal



conditions (**eqs. 20a-c**). The exponential temperature evolution of the three key parameters  $N_0$ ,  $q$ ,  $G$  is possibly calculated from the values of the physical parameters obtained in three different ways: firstly, an approximate physical analysis with direct determination from the experiments (*APA*), secondly, the use of the Genetic Algorithm method for an optimization based on several experiments (at least 5) done with the same specimen, thirdly, an optimization based on several experiments (at least 8) involving different polymer samples for which an important dispersion of the number of nuclei is observed (Haudin et al., 2008, Boyer et al., 2009). These sets of optimized temperature functions made it possible to validate the mathematical model in the 2D version, as illustrated in **Fig. 5.a-b-inserts**. The selected polymer is a polypropylene that is considered as a 'model material' because of its aptitude to crystallize with well-defined spherulitic entities in quiescent conditions.

Shear-induced crystallization, with a spherulitic morphology, gives access to the function  $dN_\gamma/dt$  ( $N_\gamma$  is the number of nuclei per unit volume generated by shear (**eq. 23**)) versus time and to the shear dependence of the activation frequency for different relatively low shear rates (up to  $20\text{ s}^{-1}$ ). A set of seven optimized parameters are identifiable:  $N_{00}$ ,  $q_0$ ,  $G_0$  from quiescent isothermal crystallization, and  $(q_{02}, q_3, A_1, a)$  from isothermal shear-induced crystallization. The agreement between experiment and theory is better for higher shear rates associated with a shorter total time of crystallization. The mean square error does not exceed 12 %, the average mean square error for  $5\text{ s}^{-1}$  is equal to 6.7 %. The agreement between experiment and theory is less satisfactory for the number of spherulites, the mean square error reaches 25 %. Then, the new model is able to predict the overall crystallization kinetics under low shear with enough accuracy, when the entities are spherulitic.

Shear-induced crystallization, with both a spherulitic and an oriented morphology, is a different task. High shear rates (from  $75\text{ s}^{-1}$ ) enhance all the kinetics (nucleation, growth, overall kinetics) and lead to the formation of micron-size fibrillar (thread-like) structures immediately after shear, followed by the appearance of unoriented spherulitic structures at the later stages (**Fig. 6insert**). The determination of the parameters for this double crystallization becomes a complicated task for a twofold reason: the quantitative data for both oriented and spherulitic structures are not available at high shear rate, and the double crystallization kinetics model requires to additionally determine the four parameters  $(w, H, B_1, b)$ . So, optimization is based only on the evolution of the total transformed volume fraction (**eq. 21**). Parameters characterizing quiescent crystallization ( $N_{00}, q_0, G_0$ ) and shear-induced crystallization with the spherulitic morphology  $(q_{02}, q_3, A_1, a)$  are taken from the previous 'smooth' analysis, so that four parameters  $(w, H, B_1, b)$  characterizing the oriented structure have to be optimized.

**Fig. 6.** gathers the experimental and theoretical variations of the total transformed volume fraction for different shear rates. At the beginning, the experimental overall kinetics is faster than the calculated one most probably because the influence of shear rate on the activation frequency of the oriented structure is not taken into account. Since with higher shear rate thinner samples ( $\sim 30\text{ }\mu\text{m}$  at  $150\text{ s}^{-1}$ ) are used, and since numerically the growth of entities is considered as three dimensional, the condition of 3D experiment seems not perfectly respected and the experiments give a slower evolution at the end. The mean square errors between numerical and experimental evolutions of the total transformed volume fraction do not exceed 19%.

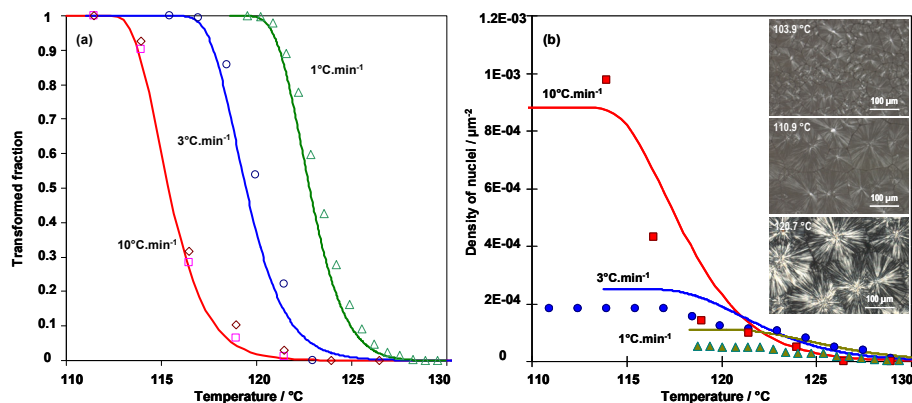


Fig. 5. Experimental (symbols / and numerically predicted (lines) of (a) the overall kinetics and (b) the number of activated nuclei *vs.* temperature at constant cooling-rate. The inserts illustrate the events at 10, 3 and 1 °C.min<sup>-1</sup>. Sample: iPP in 2D (5 μm-thick layer).

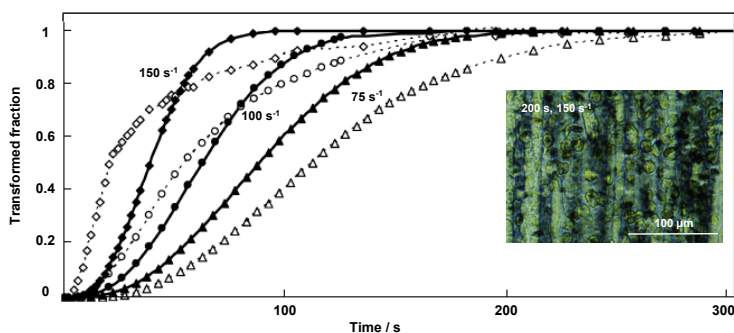


Fig. 6. Experimental (dashed-line curves) and numerically predicted (solid curves) total overall kinetics, *i.e.*, spherulitic and oriented structures, *vs.* time in constant shear,  $T = 132$  °C. The insert illustrates the event at 150 s<sup>-1</sup>. Sample: iPP in 2-3D (~30 μm-thick layer).

The present differential system, based on the nucleation and growth phenomena of polymer crystallization, is adopted to describe the crystalline morphology evolution *versus* thermo-mechanical constraints. It has been implemented into a 3D injection-moulding software. The implementation allows us to estimate its feasibility in complex forming conditions, *i.e.*, anisothermal flow-induced crystallization, and to test the sensitivity to the accuracy of the values of the parameters determined by the Genetic Algorithm Inverse Method.

#### 4. Conclusion

Fundamental understanding of the inherent links between multiscale polymer pattern and polymer behaviour/performance is firmly anchored on rigorous thermodynamics and

thermokinetics explicitly applied over extended temperature and pressure ranges, particularly under hydrostatic stress generated by pressure transmitting fluids of different physico-chemical nature.

Clearly, such an approach rests not only on the conjunction of pertinent coupled experimental techniques and of robust theoretical models, but also on the consistency and optimization of experimental and calculation procedures.

Illustration is made with selected examples like molten and solid polymers in interaction with various light molecular weight solvents, essentially gases. Data obtained allow evaluating specific thermal, chemical, mechanical behaviours coupled with sorption effect during solid to melt as well as crystallization transitions, creating smart and noble hybrid metal-polymer composites and re-visiting kinetic models taking into account similarities between polymer and metal transformations.

This work generates a solid platform for polymer science, addressing formulation, processing, long-term utilization of end-products with specific performances controlled via a clear conception of greatly different size scales, altogether with an environmental aware respect.

## 5. Acknowledgments

The principal author, Séverine A.E. Boyer, wishes to address her grateful acknowledgments for financial supports from Centre National de la Recherche Scientifique CNRS (France) ; Institut Français du Pétrole IFP (France) with Mrs. Marie-Hélène Klopffer and Mr. Joseph Martin ; Core Research for Evolutional Science and Technology - Japan Science and Technology Agency CREST-JST (Japan) with Prof. Tomokazu Iyoda (Tokyo Institute of Technology TIT, Japan) ; ARMINES-CARNOT-MINES ParisTech (France) ; Conseil Régional de Provence-Alpes-Côte d'Azur and Conseil Général des Alpes-Maritimes (France) for support in the development of «CRISTAPRESS» project.

Séverine A.E. Boyer wishes to express her acknowledgements to Intech for selectionning the current research that has been recognized as valuable and relevant to the given theme.

## 6. References

- Ahzi, S.; Parks, D.M.; Argon, A.S. (1991). Modeling of deformation textures evolution in semi-crystalline polymers. *Textures and Microstructures*, Vol.14-18, No1, (January 1991), pp. 1141-1146, ISSN 1687-5397(print) 1687-5400(web); doi: 10.1155/TSM.14-18.1141
- Arruda, E.M.; Boyce, M.C. (1993). A three-dimensional constitutive model for the large stretch behaviour of rubber elastic materials. *Journal of the Mechanics and Physics of Solids*, Vol.41, No2, (February 1993), pp. 389-412, ISSN 0022-5096; doi: 10.1016/0022.5096(93)900 13-6
- Asta, M.; Beckermann, C.; Karma, A.; Kurz, W.; Napolitano, R.; Plapp, M.; Purdy, G.; Rappaz, M.; Trivedi, R. (2009). Solidification microstructures and solid-state parallels: Recent developments, future directions. *Acta Materialia*, Vol.57, No4, (February 2009), pp. 941-971; ISSN 1359-6454; doi: 10.1016/j.actamat.2008.10.020

- Avrami, M. (1939). Kinetics of phase change. I General theory. *Journal of Chemical Physics*, Vol.7, No12, (December 1939), pp. 1103-1112, ISSN 0021-9606(print) 1089-7690(web); doi: 10.1063/1.1750380
- Avrami, M. (1940). Kinetics of phase change. II Transformation-time relations for random distribution of nuclei. *Journal of Chemical Physics*, Vol.8, No2, (February 1940), pp. 212-224, ISSN 0021-9606(print) 1089-7690(web); doi: 10.1063/1.1750631
- Avrami, M. (1941). Granulation, phase change, and microstructure. Kinetics of phase change. III. *Journal of Chemical Physics*, Vol.9, No2, (February 1941), pp. 177-184, ISSN 0021-9606(print) 1089-7690(web); doi: 10.1063/1.1750872
- Baert, J.; Van Puyvelde, P.; Langouche, F.; (2006). Flow-induced crystallization of PB-1: From the low shear rate region up to processing rates. *Macromolecules*, Vol.39, No26, (November 2006), pp. 9215-9222, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma062068q
- Baudet, C.; Grandidier, J.-C.; Cangémi, L. (2009). A two-phase model for the diffuse-mechanical behaviour of semicrystalline polymers in gaseous environment. *International Journal of Solids and Structures*, Vol.46, No6, (March 2009), pp. 1389-1401, ISSN 0020-7683; doi: 10.1016/j.ijsolstr.2008.11.010
- Bedoui, F.; Diani, J.; Régnier, G.; Seiler, W. (2006). Micromechanical modelling of isotropic elastic behaviour of semicrystalline polymers. *Acta Materialia*, Vol.54, No6, (April 2006), pp. 1513-1523, ISSN 1359-6454; doi: 10.1016/j.actamat.2005.11.028
- Behme, S.; Sadowski, G.; Arlt, W. (1999). Modeling of the separation of polydisperse polymer systems by compressed gases. *Fluid Phase Equilibria*, Vol.158-160, (June 1999), pp. 869-877, ISSN 0378-3812; doi: 10.1016/S0378-3812(99)00055-2
- Beret, S.; Prausnitz, J.M. (1975). Perturbed hard-chain theory: An equation of state for fluids containing small or large molecules. *AIChE Journal*, Vol.21, No6, (November 1975), pp. 1123-1132, ISSN 0001-1541(print) 1547-5905(web); doi: 10.1002/aic.690210612
- Bessières, D.; Lafitte, Th.; Daridon, J.-L.; Randzio, S.L. (2005). High pressure thermal expansion of gases: measurements and calibration. *Thermochimica Acta*, Vol.428, No1-2, (April 2005), pp. 25-30, ISSN 0040-6031; doi: 10.1016/j.tca.2004.09.020
- Binsbergen, F.L. (1973). Heterogeneous nucleation in the crystallization of polyolefins. III. Theory and mechanism. *Journal of Polymer Science: Polymer Physics Edition*, Vol.11, No1, (January 1973), pp. 117-135, ISSN 0098-1273(print) 1542-9385(web); doi: 10.1002/pol.1973.180110112
- Boal, A.K.; Rotello, V.M. (2000). *Intra-* and *intermonolayer* hydrogen bonding in amide-functionalized alkanethiol self-assembled monolayers on gold nanoparticles. *Langmuir*, Vol.16, No24, (November 2000), pp. 9527-9532, ISSN 0743-7463(print) 1520-5827(web); doi: 10.1021/la0012384
- Boyer, S.A.E.; Fournier, F.; Haudin, J.-M.; Gandin, Ch.-A. (2011a). Model experiments and structure development in high-pressure crystallization: the CRISTAPRESS project. *The Polymer Processing Society - 27<sup>th</sup> Annual Meeting* (Key Note), Marrakech (Maroc) May 10-14, 2011. (proceeding, 5 pages)
- Boyer, S.A.E.; Ganet, P.; Robinson, P.; Melis, J.-P.; Haudin, J.-M. (2011b). Crystallization of polypropylene at high cooling rates. Microscopic and calorimetric studies. *Journal of Applied Polymer Science - Invited conference*, 4<sup>th</sup> International Conference on Polymer Behavior, International Union of Pure and Applied Chemistry (IUPAC), Lodz, Poland, September 20-23, 2010-, (Accepted March 2011), ISSN 0021-8995(print) 1097-4628(web)

- Boyer, S.A.E.; Grolier, J.-P.E. (2005). Modification of the glass transitions of polymers by high-pressure gas solubility. *Pure and Applied Chemistry Science - Invited conference, The 11th International Symposium on Solubility Phenomena (11th ISSP)*, Aveiro, Portugal, July 25-29, 2004-, Vol.77, No3, (March 2005), pp. 593-603, ISSN 0033-4545(print) 1365-3075(web); doi: 10.1351/pac200577030593
- Boyer, S.A.E.; Grolier, J.-P.E.; Pison, L.; Iwamoto, C.; Yoshida, H.; Iyoda, T. (2006a). Isotropic transition behavior of an amphiphilic di-block copolymer under pressure. Carbon dioxide or mercury as pressure medium. *Journal of Thermal Analysis and Calorimetry*, Vol.85, No3, (September 2006), pp. 699-706, ISSN 1388-6150(print) 1572-8943(web); doi: 10.1007/s10973-006-7633-z
- Boyer, S.A.E.; Grolier, J.-P.E.; Yoshida, H.; Haudin, J.-M.; Chenot, J.-L. (2009). Phase transitions of polymers over T and P ranges under various hydraulic fluids: Polymer/supercritical gas systems and liquid to solid polymer transitions. *Journal of Molecular Liquids - Invited conference, Joint Conference of JMLG/EMLG Meeting 2007 and 30th Symposium on Solution Chemistry of Japan, Molecular Approaches to Complex Liquid Systems*, Fukuoka, Japan, November 21-25, 2007-, Vol.147, No1-2, (July 2009), pp. 24-29, ISSN 0167-7322; doi: 10.1016/j.molliq.2009.01.016
- Boyer, S.A.E.; Haudin, J.-M. (2010). Crystallization of polymers at constant and high cooling rates: A new hot stage microscopy set-up. *Polymer Testing - Invited conference, The 45th Japanese Conference on Calorimetry and Thermal Analysis*, Tokyo, Japan, September 28-30, 2009-, Vol.29, No4, (June 2010), pp. 445-452, ISSN 0142-9418; doi:10.1016/j.polymertesting.2010.02.003
- Boyer, S.A.E.; Klopffer, M.-H.; Martin, J.; Grolier, J.-P.E. (2007). Supercritical gas-polymer interactions with applications in the petroleum industry. Determination of thermophysical properties. *Journal of Applied Polymer Science*, Vol.103, No3, (February 2007), pp. 1706-1722, ISSN 0021-8995(print) 1097-4628(web); doi: 10.1002/app.25085
- Boyer, S.A.E.; Randzio, S.L.; Grolier, J.-P.E. (2006b). Thermal expansion of polymers submitted to supercritical CO<sub>2</sub> as a function of pressure. *Journal of Polymer Science Part B: Polymer Physics*, Vol.44, No1, (January 2006), pp. 185-194, ISSN 0887-6266(print) 1099-0488(web); doi: 10.1002/polb.20674
- Brucato, V.; Piccarolo, S.; La Carrubba, V. (2002). An experimental methodology to study polymer crystallization under processing conditions. The influence of high cooling rates. *Chemical Engineering Science*, Vol.57, No19, (October 2002), pp. 4129-4143, ISSN 0009-2509; doi: 10.1016/S0009-2509(02)00360-3
- Carroll, D.L. (2002). "FORTRAN Genetic Algorithm Front-End Driver Code", <http://cuaerospace.com/ga>.
- Dahoun, A.; Canova, G.R.; Molinari, A.; Philippe, M. J.; G'Sell, C. (1991). The modelling of large strain textures and stress-strain relations of polyethylene. *Textures and Microstructures*, Vol.14-18, No1, (January 1991), pp. 347-354, ISSN 1687-5397(print) 1687-5400(web); doi: 10.1155/TSM.14-18.347
- De Angelis, M.G.; Merkel, T.C.; Bondar, V.I.; Freeman, B.D.; Doghieri, F.; Sarti, G.C. (1999). Hydrocarbon and fluorocarbon solubility and dilation in poly(dimethylsiloxane): Comparison of experimental data with predictions of the Sanchez-Lacombe equation of state. *Journal of Polymer Science Part B: Polymer Physics*, Vol.37, No21, (November 1999), pp. 3011-3026, ISSN 0887-6266(print) 1099-0488(web); doi: 10.1002/(SICI)1099-0488(19991101)37:21<3011::AID-POLB11>3.0.CO;2-V

- De Rosa, C.; Park, C.; Thomas, E.L.; Lotz, B. (2000). Microdomain patterns from directional eutectic solidification and epitaxy. *Nature*, Vol.405, No6785, (May 2000), pp. 433-437, ISSN 0028-0836(print) 1476-4687(web); doi: 10.1038/35013018
- Devisme, S.; Haudin, J.-M.; Agassant, J.-F.; Rauline, D.; Chopinez, F. (2007). Numerical simulation of extrusion coating. Contribution to the understanding of adhesion mechanisms between grafted polypropylene and aluminium. *International Polymer Processing*, Vol.22, No1, pp. 90-104, ISSN 0930-777X
- Dewimille, B.; Martin, J.; Jarrin., J. (1993). Behavior of thermoplastic polymers during explosive decompressions in a petroleum environment. *Journal de Physique. IV*, Vol. 3(2), No7, (July 1993), pp. 1559-1564, ISSN 1155-4339; doi: 10.1051/jp4:19937243
- Ding, Z.; Spruiell, J.E. (1996). An experimental method for studying nonisothermal crystallization of polymers at very high cooling rates. *Journal of Polymer Science Part B: Polymer Physics*, Vol.34, No16, (September 1996), pp. 2783-2804, ISSN 0887-6266(print) 1099-0488(web); doi: 10.1002/(SICI)1099-0488(19961130) 34:16<2783::AID-POLB12>3.0.CO;2-6
- Dixon, R.K. (2011). Global environment facility investments in the phase-out of ozone-depleting substances. *Mitigation and Adaptation Strategies for Global Change* (online first™ January 2011), ISSN 1381-2386(print) 1573-1596(web); doi: 10.1007/s11027-011-9281-2
- Doghieri, F.; Sarti, G.C. (1996). Nonequilibrium lattice fluids: A predictive model for the solubility in glassy polymers. *Macromolecules*, Vol.29, No24, (November 1996), pp. 7885-7896, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma951366c
- Elmoumni, A.; Winter, H.H. (2006). Large strain requirements for shear-induced crystallization of isotactic polypropylene. *Rheologica Acta*, Vol.45, No6, (August 2006), pp. 793-801, ISSN 0035-4511(print) 1435-1528(web); doi: 10.1007/s00397-005-0082-y
- Evans, U.R., (1945). The laws of expanding circles and spheres in relation to the lateral growth of surface films and the grain-size of metals. *Transactions of the Faraday Society*, Vol.41, (1945), pp. 365-374, ISSN 0014-7672; doi: 10.1039/TF9454100365
- Ferreiro, V.; Douglas, J.F.; Warren, J.A.; Karim, A. (2002a). Nonequilibrium pattern formation in the crystallization of polymer blend films. *Physical review E, Statistical, nonlinear, and soft matter physics*, Vol.65, No4, (April 2002), pp. 042802 1-4, ISSN 1539-3755(print) 1550-2376(web); doi: 10.1103/PhysRevE.65.042802
- Ferreiro, V.; Douglas, J.F.; Warren, J.A.; Karim, A. (2002b). Growth pulsations in symmetric dendritic crystallization in thin polymer blend films. *Physical review E, Statistical, nonlinear, and soft matter physics*, Vol.65, No5, (May 2002), pp. 051606 1-16, ISSN 1539-3755(print) 1550-2376(web); doi: 10.1103/PhysRevE.65.051606
- Flaconnèche, B.; Martin, J.; Klopffer, M.-H. (2001). Permeability, diffusion, and solubility of gases in polyethylene, polyamide 11 and poly(vinylidene fluoride). *Oil & Gas Science and Technology – Rev. IFP*, Vol.56, No3, (May-June 2001), pp. 261-278, ISSN 1294-4475(print) 1953-8189(web); doi: 10.2516/ogst:2001023
- Flemings, M.C. (1974). Solidification processing. *Metallurgical and Materials Transactions B*, Vol.5, No10, (October 1974), pp. 2121-2134, ISSN 1073-5615(print) 1543-1916(web); doi: 10.1007/BF02643923
- Fulchiron, R.; Koscher, E.; Poutot, G.; Delaunay, D.; Régnier, G. (2001). Analysis of the pressure effect on the crystallization kinetics of polypropylene: Dilatometric measurements and thermal gradient modeling. *Journal of Macromolecular Science*,

- Part B: Physics*, Vol.40, No3&4, pp. 297-314, ISSN 0022-2348(print) 1525-609X(web); doi: 10.1081/MB-100106159
- Gadomski, A.; Łuczka, J. (2000). On the kinetics of polymer crystallization: A possible mechanism. *Journal of Molecular Liquids*, Vol.86, No1-3, (June 2000), pp. 237-247, ISSN 0167-7322; doi: 10.1016/S0167-7322(99)00145-2
- Glasser, L. (2002). Equations of state and phase diagrams. *Journal of Chemical Education*, Vol.79, No7, (July 2002), pp. 874-876, ISSN 0021-9584; doi: 10.1021/ed079p874
- Glasser, L. (2004). Water, water, everywhere: Phase diagrams of ordinary water substance. *Journal of Chemical Education*, Vol.81, No3, (March 2004), pp. 414-418, ISSN 0021-9584; doi: 10.1021/ed081p414
- Gránásy, L.; Pusztai, T.; Warren, J.A.; Douglas, J.F.; Börzsönyi, T.; Ferreiro, V. (2003). Growth of 'dizzy dendrites' in a random field of foreign particles. *Nature Materials*, Vol.2, No2, (February 2003), pp. 92-96, ISSN 1476-1122(print) 1476-4660(web); doi: 10.1038/nmat815
- Grolier, J.-P.E.; Dan, F.; Boyer, S.A.E.; Orlowska, M.; Randzio, S.L. (2004). The use of scanning transitiometry to investigate thermodynamic properties of polymeric systems over extended  $T$  and  $p$  ranges. *International Journal of Thermophysics*, Vol.25, No2, (March 2004), pp. 297-319, ISSN 0195-928X(print) 1572-9567(web); doi: 10.1023/B:IJOT.0000028469.17288.de
- Grubbs, R.B. (2005). Hybrid metal-polymer composites from functional block copolymers. *Journal of Polymer Science Part A: Polymer Chemistry*, Vol.43, No19, (October 2005), pp. 4323-4673, ISSN 0887-624X(print) 1099-0518(web); doi: 10.1002/pola.20946
- Hamley, I.W. (2009). Ordering in thin films of block copolymers: Fundamentals to potential applications. *Progress in Polymer Science*, Vol.34, No11, (November 2009), pp. 1161-1210, ISSN 0079-6700; doi: 10.1016/j.progpolymsci.2009.06.003
- Haudin, J.-M.; Chenot, J.-L. (2004). Numerical and physical modeling of polymer crystallization. Part I: Theoretical and numerical analysis. *International Polymer Processing*, Vol.19, No3, pp. 267-274, ISSN 0930-777X
- Haudin, J.-M.; Smirnova, J.; Silva, L.; Monasse, B.; Chenot, J.-L. (2008). Modeling of structure development during polymer processing. *Polymer Science, Series A. Polymer Physics*, Vol.50, No5, (May 2008), pp. 538-549, ISSN 0965-545X(print) 1555-6107(web); doi: 10.1134/S0965545X08050088
- Hilic, S.; Pádua, A.A.H.; Grolier, J.-P.E. (2000). Simultaneous measurement of the solubility of gases in polymers and of the associated volume change. *Review of Scientific Instruments*, Vol.71, No11, (November 2000), pp. 4236-4241, ISSN 0034-6748(print) 1089-7623(web); doi: 10.1063/1.1289675
- Hoffman, J.D. (1983). Regime III crystallization in melt-crystallized polymers: The variable cluster model of chain folding. *Polymer*, Vol.24, No1, (January 1983), pp. 3-26, ISSN 0032-3861; doi:10.1016/0032-3861(83)90074-5
- Janeschitz-Kriegl, H. (2006). Phases of flow-induced crystallization of i-PP: How remote pieces of the puzzle appear to fit. *Macromolecules*, Vol.39, No13, (June 2006), pp. 4448-4454, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma0601215
- Janeschitz-Kriegl, H.; Ratajski, E.; Stadlbauer, M. (2003). Flow as an effective promotor of nucleation in polymer melts: A quantitative evaluation. *Rheologica Acta*, Vol.42, No4, (July 2003), pp. 355-364, ISSN 0035-4511(print) 1435-1528(web); doi: 10.1007/s00397-002-0247-x
- Jin, L.-W.; Claborn, K.A.; Kurimoto, M.; Geday, M.A.; Maezawa, I.; Sohraby, F.; Estrada, M.; Kaminsky, W.; Kahr, B. (2003). Imaging linear birefringence and dichroism in

- cerebral amyloid pathologies. *Proceedings of the National Academy of Sciences of the United States of America*, Vol.100, No26, (December 2003), pp. 15294-15298, ISSN 0027-8424(print) 1091-6490(web); doi: 10.1073/pnas.2534647100
- Kamiya, Y.; Mizoguchi, K.; Terada, K.; Fujiwara, Y.; Wang, J.-S. (1998). CO<sub>2</sub> sorption and dilation of poly(methyl methacrylate). *Macromolecules*, Vol.31, No2, (January 1998), pp. 472-478, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma970456
- Keith, H.D.; Padden, Jr., F.J. (1964). Spherulitic crystallization from the melt. I. Fractionation and impurity segregation and their influence on crystalline morphology. *Journal of Applied Physics*, Vol.35, No4, (April 1964), pp. 1270-1285, ISSN 0021-8979(print) 1089-7550(web); doi:10.1063/1.1713606
- Klopffer, M.-H.; Flaconnèche, B. (2001). Transport properties of gases in polymers: Bibliographic review. *Oil & Gas Science and Technology – Rev. IFP*, Vol.56, No3, (May-June 2001), pp. 223-244, ISSN 1294-4475(print) 1953-8189(web); doi: 10.2516/ogst:2001021
- Kobayashi, R. (1993). Modeling and numerical simulations of dendritic crystal growth. *Physica D: Nonlinear Phenomena*, Vol.63, No3-4, (March 1993), pp. 410-423, ISSN 0167-2789; doi: 10.1016/0167-2789(93)90120-P
- Krebs, M.R.H.; Bromley, E.H.C.; Rogers, S.S.; Donald, A.M. (2005). The mechanism of amyloid spherulite formation by bovine insulin. *Biophysical Journal*, Vol.88, No3, (March 2005), pp. 2013-2021, ISSN 0006-3495(print) 1542-0086(web); doi: 10.1529/biophysj.104.051896
- Lacombe, R.H.; Sanchez, I.C. (1976). Statistical thermodynamics of fluid mixtures. *The Journal of Physical Chemistry C*, Vol.80, No23, (November 1976), pp. 2568-2580, ISSN 1932-7447(print) 1932-7455(web); doi: 10.1021/j100564a009
- Lavery, K.A.; Sievert, J.D.; Watkins, J.J.; Russell, T.P.; Ryu, D.Y.; Kim, J.K. (2006). Influence of carbon dioxide swelling on the closed-loop phase behavior of block copolymers. *Macromolecules*, Vol.39, No19, (September 2006), pp. 6580-6583, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma060329q
- Lee, L.J.; Zeng, C.; Cao, X.; Han, X.; Shen, J.; Xu, G. (2005). Polymer nanocomposite foams. *Composites Science and Technology*, Vol.65, No15-16, (December 2005), pp. 2344-2363, ISSN 0266-3538; doi: 10.1016/j.compscitech.2005.06.016
- Li, Y.G.; Park, C.B.; Li, H.B.; Wang, J. (2008). Measurement of the PVT property of PP/CO<sub>2</sub> solution. *Fluid Phase Equilibria*, Vol.270, No1-2, (August 2008), pp. 15-22, ISSN 0378-3812; doi 10.1016/j.fluid.2008.05.007
- Lo, C.-T.; Lee, B.; Pol, V.G.; Dietz Rago, N.L.; Seifert, S.; Winans, R.E.; Thiyagarajan, P. (2007). Effect of molecular properties of block copolymers and nanoparticles on the morphology of self-assembled bulk nanocomposites. *Macromolecules*, Vol.40, No23, (October 2007), pp. 8302-8310, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma070835v
- Maeda, Y.; Niori, T.; Yamamoto, J.; Yokoyama, H. (2005). Effect of pressure on phase behavior of a thermotropic cubic mesogen. *Thermochimica Acta*, Vol.428, No1-2, (April 2005), pp. 57-62, ISSN 0040-6031; doi: 10.1016/j.tca.2004.09.030
- Magill, J.H. (1961). Crystallization of isotactic polypropylene using a light depolarization technique. *Nature*, Vol.191, No4793, (September 1961), pp. 1092-1093, ISSN 0028-0836(print) 1476-4687(web); doi: 10.1038/1911092a0
- Magill, J.H. (1962). A new technique for following rapid rates of crystallization II Isotactic polypropylene. *Polymer*, Vol.3, (April 1962), pp. 35-42, ISSN 0032-3861; doi: 10.1016/0032-3861(62)90064-2



- Magill, J.H. (2001). Review Spherulites: A personal perspective. *Journal of Materials Science*, Vol.36, No13, (July 2001), pp. 3143-3164, ISSN 0022-2461(print) 1573-4803(web); doi: 10.1023/A:1017974016928
- Nakamura, K.; Watanabe, T.; Katayama, K.; Amano T. (1972). Some aspects of non-isothermal crystallization of polymers. I. Relationship between crystallization temperature, crystallinity, and cooling conditions. *Journal of Applied Polymer Science*, Vol.16, No5, (May1972), pp. 1077-1091, ISSN 0021-8995(print) 1097-4628(web); doi: 10.1002/app.1972.070160503
- Nalawade, S.P.; Picchioni, F.; Janssen, L.P.B.M. (2006). Supercritical carbon dioxide as a green solvent for processing polymer melts: Processing aspects and applications. *Progress in Polymer Science*, Vol.31, No1, (January 2006), pp. 19-43, ISSN 0079-6700; doi: 10.1016/j.progpolymsci.2005.08.002
- Nowacki, R.; Monasse, B.; Piorkowska, E.; Galeski, A.; Haudin, J.-M. (2004). Spherulite nucleation in isotactic polypropylene based nanocomposites with montmorillonite under shear. *Polymer*, Vol.45, No14, (June 2004), pp. 4877-4892, ISSN 0032-3861; doi: 10.1016/j.polymer.2004.04.058
- Ogino, Y.; Fukushima, H.; Takahashi, N.; Matsuba, G.; Nishida, K.; Kanaya, T. (2006). Crystallization of isotactic polypropylene under shear flow observed in a wide spatial scale. *Macromolecules*, Vol.39, No22, (October 2006), pp. 7617-7625, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma061254t
- Okerberg, B.C.; Marand, H.; Douglas, J.F. (2008). Dendritic crystallization in thin films of PEO/PMMA blends: A comparison to crystallization in small molecule liquids. *Polymer*, Vol.49, No2, (January 2008), pp. 579-587, ISSN 0032-3861; doi: 10.1016/j.polymer.2007.11.034
- Orbey, H.; Bokis, C.P.; Chen, C.-C. (1998). Equation of state modeling of phase equilibrium in the low-density polyethylene process: The Sanchez-Lacombe, statistical associating fluid theory, and polymer-Soave-Redlich-Kwong equations of state. *Industrial & Engineering Chemistry Research*, Vol.37, No11, (November 1998), pp. 4481-4491, ISSN 0888-5885(print) 1520-5045(web); doi: 10.1021/ie980220+
- Ozawa, T. (1971). Kinetics of non-isothermal crystallization. *Polymer*, Vol.12, No3, (March 1971), pp. 150-158, ISSN 0032-3861; doi: 10.1016/032-3861(71)90041-3
- Panayiotou, C.; Sanchez, I.C. (1991). Statistical thermodynamics of associated polymer solutions. *Macromolecules*, Vol.24, No23, (November 1991), pp. 6231-6237, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma00023a027
- Panine, P.; Di Cola, E.; Sztucki, M.; Narayanan, T. (2008). Early stages of polymer melt crystallization. *Polymer*, Vol.49, No3, (February 2008), pp. 676-680, ISSN 0032-3861; doi: 10.1016/j.polymer.2007.12.026
- Park, C.; Yoon, J.; Thomas, E.L. (2003). Enabling nanotechnology with self assembled block copolymer patterns. *Polymer*, Vol.44, No22, (October 2003), pp. 6725-6760, ISSN 0032-3861; doi: 10.1016/j.polymer.2003.08.011
- Paszkwicz, W. (2009). Genetic Algorithms, a nature-inspired tool: Survey of applications in materials science and related fields. *Materials and Manufacturing Processes*, Vol.24, No2, (February 2009), pp. 174-197, ISSN 1042-6914(print) 1532-2475(web); doi: 10.1080/10426910802612270
- Pawlak, A.; Chapel, J.-P.; Piorkowska, E. (2002). Morphology of iPP spherulites crystallized in a temperature gradient. *Journal of Applied Polymer Science*, Vol.86, No6, (November 2002), pp. 1318-1328, ISSN 0021-8995(print) 1097-4628(web); doi: 10.1002/app.11269

- Piorkowska, E.; Galeski, A.; Haudin, J.-M. (2006). Critical assessment of overall crystallization kinetics theories and predictions. *Progress in Polymer Science*, Vol.31, No6, (June 2006), pp. 549-575, ISSN 0079-6700; doi: 10.1016/j.progpolymsci.2006.05.001
- Prigogine, I.; Bellemans, A.; Mathot, V. (1957). The molecular theory of solutions. *North-Holland Pub. Co. (Amsterdam) 1957. ID 101-186-725 (Last edited on 2002/02/27 17:06:35 US/Mountain)*
- Rambert, G.; Jugla, G.; Grandidier, J.-C.; Cangémi, L. (2006). A modelling of the direct couplings between heat transfer, mass transport, chemical reactions and mechanical behaviour. Numerical implementation to explosive decompression. *Composites Part A: Applied Science and Manufacturing*, Vol.37, No4, (April 2006), pp. 571-584, ISSN 1359-835X; doi: 10.1016/j.compositesa.2005.05.021
- Ratajski, E.; Janeschitz-Kriegl, H. (1996). How to determine high growth speeds in polymer crystallization. *Colloid & Polymer Science*, Vol.274, No10, (October 1996), pp. 938-951, ISSN 0303-402X(print) 1435-1536(web); doi: 10.1007/BF00656623
- Rousset, Ph.A.; Rappaz, M.; Minner, E. (1998). Polymorphism and solidification kinetics of the binary system POS-SOS. *Journal of the American Oil Chemists' Society*, Vol.75, No7, (July 1998), pp. 857-864, ISSN 0003-021X(print) 1558-9331(web); doi: 10.1007/s11746-998-0237-y
- Sanchez, I.C.; Lacombe, R.H. (1976). An elementary molecular theory of classical fluids. Pure fluids. *The Journal of Physical Chemistry*, Vol.80, No21, (October 1976), pp. 2352-2362, ISSN 1932-7447; doi: 10.1021/j100562a008
- Sánchez, M.S.; Mathot, V.B.F.; Poel, G.V.; Gómez Ribelles, J.L. (2007). Effect of the cooling rate on the nucleation kinetics of poly(L-lactic acid) and its influence on morphology. *Macromolecules*, Vol.40, No22, (September 2007), pp. 7989-7997, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma0712706
- Sarti, G.C.; Doghieri, F. (1998). Predictions of the solubility of gases in glassy polymers based on the NELF model. *Chemical Engineering Science*, Vol.53, No19, (October 1998), pp. 3435-3447, ISSN 0009-2509; doi: 10.1016/S0009-2509(98)00143-2
- Sato, Y.; Yurugi, M.; Fujiwara, K.; Takishima, S.; Masuoka, H. (1996). Solubilities of carbon dioxide and nitrogen in polystyrene under high temperature and pressure. *Fluid Phase Equilibria*, Vol.125, No1-2, (October 1996), pp. 129-138, ISSN 0378-3812; doi: 10.1016/S0378-3812(96)03094-4
- Sato, Y.; Iketani, T.; Takishima, S.; Masuoka, H. (2000). Solubility of hydrofluorocarbon (HFC-134a, HFC-152a) and hydrochlorofluorocarbon (HCFC-142b) blowing agents in polystyrene. *Polymer Engineering & Science*, Vol.40, No6, (June 2000), pp. 1369-1375, ISSN 0032-3888(print) 1548-2634 (web); doi: 10.1002/pen.11266
- Schick, C. (2009). Differential scanning calorimetry (DSC) of semicrystalline polymers. *Analytical and Bioanalytical Chemistry*, Vol.395, No6, (November 2009), pp. 1589-1611, ISSN 1618-2642(print) 1618-2650(web); doi: 10.1007/s00216-009-3169-y
- Scheichl, R.; Klopffer, M.-H.; Benjelloun-Dabaghi, Z.; Flaconneche, B. (2005). Permeation of gases in polymers: Parameter identification and nonlinear regression analysis. *Journal of Membrane Science*, Vol.254, No1-2, (June 2005), pp. 275-293, ISSN 0376-7388; doi: 10.1016/j.memsci.2005.01.019
- Schneider, W.; Köppl, A.; Berger, J. (1988). Non-isothermal crystallization of polymers. System of rate equation. *International Polymer Processing*, Vol.2, No3-4, pp. 151-154, ISSN 0930-777X

- Smirnova, J.; Silva, L.; Monasse, B.; Haudin, J.-M.; Chenot, J.-L. (2007). Identification of crystallization kinetics parameters by genetic algorithm in non-isothermal conditions. *Engineering Computations*, Vol.24, No5, pp. 486-513, ISSN 0264-4401; doi: 10.1108/02644400710755889
- Sorrentino, A.; Pantani, R.; Picarella, D.; Titomanlio, G. (2005). A novel apparatus for solidification of polymer samples under simultaneous high pressures and high cooling rates. *Review of Scientific Instruments*, Vol.76, No8, (August 2005), pp. 083901-083905, ISSN 0034-6748(print) 1089-7623(web); doi: 10.1063/1.1986987
- Talapin, D.V.; Shevchenko, E.V.; Bodnarchuk, M.I.; Ye, X.; Chen, J.; Murray, C.B. (2009). Quasicrystalline order in self-assembled binary nanoparticle superlattices. *Nature*, Vol.461, No7266, (October 2009), pp. 964-967, ISSN 0028-0836(print) 1476-4687(web); doi: 10.1038/nature08439
- Tian, Y.; Watanabe, K.; Kong, X.; Abe, J.; Iyoda, T. (2002). Synthesis, nanostructures, and functionality of amphiphilic liquid crystalline block copolymers with azobenzene moieties. *Macromolecules*, Vol.35, No9, (April 2002), pp. 3739-3747, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma011859j
- Tomasko, D.L.; Li, H.; Liu, D.; Han, X.; Wingert, M.J.; Lee, L.J.; Koelling, K.W. (2003). A review of CO<sub>2</sub> applications in the processing of polymers. *Industrial & Engineering Chemistry Research*, Vol.42, No25, (December 2003), pp. 6431-6456, ISSN 0888-5885(print) 1520-5045(web); doi: 10.1021/ie030199z
- Trivedi, R.; Laorchan, V. (1988a). Crystallization from an amorphous matrix-I. Morphological studies. *Acta Metallurgica Materialia*, Vol.36, No8, (August 1988), pp. 1941-1950, ISSN 0001-6160; doi: 10.1016/0001-6160(88)90296-9
- Trivedi, R.; Laorchan, V. (1988b). Crystallization from an amorphous matrix-II. Growth kinetics. *Acta Metallurgica*, Vol.36, No8, (August 1988), pp. 1951-1959, ISSN 0001-6160; doi: 10.1016/0001-6160(88)90297-0
- Van der Beek, M.H.E.; Peters, G.W.M.; Meijer, H.E.H. (2005). The influence of cooling rate on the specific volume of isotactic poly(propylene) at elevated pressures. *Macromolecular Materials and Engineering*, Vol.290, No5, (May 2005), pp. 443-455, ISSN 1438-7492(print) 1439-2054(web); doi: 10.1002/mame.200500027
- Varma-Nair, M.; Handa, P.Y.; Mehta, A.K.; Agarwal, P. (2003). Effect of compressed CO<sub>2</sub> on crystallization and melting behavior of isotactic polypropylene. *Thermochimica Acta*, Vol.396, No1-2, (February 2003), pp. 57-65, ISSN 0040-6031; doi: 10.1016/S0040-6031(02)00516-6
- Vieth, W.R.; Howell, J.M.; Hsieh J.H. (1976). Dual sorption theory. *Journal of Membrane Science*, Vol.1, (1976), pp. 177-220, ISSN 0376-7388; doi: 10.1016/S0376-7388(00)82267-X
- Vogt, B.D.; Ramachandra Rao, V.S.; Gupta, R.R.; Lavery, K.A.; Francis, T.J.; Russell, T.P.; Watkins, J.J. (2003). Phase behavior of polystyrene-block-poly(*n*-alkyl methacrylate)s diluted with carbon dioxide. *Macromolecules*, Vol.36, No11, (April 2003), pp. 4029-4036, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma0300544
- Wang, L.; Li, Q.; Shen, C. (2010). The numerical simulation of the crystallization morphology evolution of semi-crystalline polymers in injection molding. *Polymer-Plastics Technology and Engineering*, Vol.49, No10, (August 2010), pp. 1036-1048, ISSN 0360-2559(print) 1525-6111(web); doi: 10.1080/03602559.2010.482088
- Watanabe, K.; Suzuki, T.; Masubuchi, Y.; Taniguchi, T.; Takimoto, J.-I.; Koyama, K. (2003). Crystallization kinetics of polypropylene under high pressure and steady shear

- flow. *Polymer*, Vol.44, No19, (September 2003), pp. 5843-5849, ISSN 0032-3861; doi: 10.1016/S0032-3861(03)00604-9
- Watanabe, S.; Fujiwara, R.; Hada, M.; Okazaki, Y.; Iyoda, T. (2007). Site-specific recognition of nanophase-separated surfaces of amphiphilic block copolymers by hydrophilic and hydrophobic gold nanoparticles. *Angewandte Chemie International Edition*, Vol.46, No7, (February 2007), pp. 1120-1123, ISSN 1433-7851(print) 1521-3773(web); doi: 10.1002/anie.200603516
- Watanabe, R.; Kamata, K.; Iyoda, T. (2008). Smart block copolymer masks with molecule-transport channels for total wet nanopatterning. *Journal of Materials Chemistry*, Vol.18, No45, (October 2008), pp. 5482-5491, ISSN 0959-9428(print+web) 1364-5501(web only); doi: 10.1039/B806378H
- Winter, H.H.; Gappert, G.; Ito, H. (2002). Rigid pore structure from highly swollen polymer gels. *Macromolecules*, Vol.35, No9, (March 2002), pp. 3325-3327, ISSN 0024-9297(print) 1520-5835(web); doi: 10.1021/ma0119225
- Wissinger, R.G.; Paulaitis, M.E. (1987). Swelling and sorption in polymer-CO<sub>2</sub> mixtures at elevated pressures. *Journal of Polymer Science Part B: Polymer Physics*, Vol.25, No12, (December 1987), pp. 2497-2510, ISSN 0887-6266(print) 1099-0488(web); doi: 10.1002/polb.1987.090251206
- Wong, B.; Zhang, Z.; Handa, Y.P. (1998). High-precision gravimetric technique for determining the solubility and diffusivity of gases in polymers. *Journal of Polymer Science Part B: Polymer Physics*, Vol.36, No12, (September 1998), pp. 2025-2032, ISSN 0887-6266(print) 1099-0488(web); doi: 10.1002/(SICI)1099-0488(19980915)36:12<2025::AID-POLB2>3.0.CO;2-W
- Yamada, T.; Boyer, S.A.E.; Iyoda, T.; Yoshida, H.; Grolier, J.-P.E. (2007a). Isotropic transition of amphiphilic side-chain type liquid crystalline di-block copolymers. Effects of nitrogen pressure. *Journal of Thermal Analysis and Calorimetry*, Vol.89, No1, (July 2007), pp. 9-12, ISSN 1388-6150(print) 1572-8943(web); doi: 10.1007/s10973-006-8451-z
- Yamada, T.; Boyer, S.A.E.; Iyoda, T.; Yoshida, H.; Grolier, J.-P.E. (2007b). Effects of CO<sub>2</sub> pressure on isotropic transition of amphiphilic side-chain type liquid crystalline di-block copolymers. *Journal of Thermal Analysis and Calorimetry*, Vol.89, No3, (September 2007), pp. 717-721, ISSN 1388-6150(print) 1572-8943(web); doi: 10.1007/s10973-006-7960-0
- Yoshida, H.; Watanabe, K.; Watanabe, R.; T. Iyoda (2004). Self assembled structure of amphiphilic di-block copolymer having azobenzene moieties. *Transactions of the Materials Research Society of Japan*. Vol.29, No3, (May 2004), pp. 861-864, ISSN 1382-3469
- Zheng, R.; Kennedy, P.K. (2004). A model for post-flow induced crystallization: General equations and predictions. *Journal of Rheology*, Vol.48, No4, (July 2004), pp. 823-842, ISSN 0148-6055; doi: 10.1122/1.1763944
- Zhong, C.; Masuoka, H. (1998). Modeling of gas solubilities in polymers with cubic equation of state. *Fluid phase equilibria*, Vol.144, No1-2, (February 1998), pp. 49-57, ISSN 0378-3812; doi: 10.1016/S0378-3812(97) 00243-4

# Thermodynamics and Reaction Rates

Miloslav Pekař  
*Brno University of Technology*  
*Czech Republic*

## 1. Introduction

Thermodynamics has established in chemistry principally as a science determining possibility and direction of chemical transformations and giving conditions for their final, equilibrium state. Thermodynamics is usually thought to tell nothing about rates of these processes, their velocity of approaching equilibrium. Rates of chemical reactions belong to the domain of chemical kinetics. However, as thermodynamics gives some restriction on the course of chemical reactions, similar restrictions on their rates are continuously looked for. Similarly, because thermodynamic potentials are often formulated as driving forces for various processes, a thermodynamic driving force for reactions rates is searched for.

Two such approaches will be discussed in this article. The first one are restrictions put by thermodynamics on values of rate constants in mass action rate equations. The second one is the use of the chemical potential as a general driving force for chemical reactions and also “directly” in rate equations. These two problems are in fact connected and are related to expressing reaction rate as a function of pertinent independent variables.

Relationships between chemical thermodynamics and kinetics traditionally emerge from the ways that both disciplines use to describe equilibrium state of chemical reactions (chemically reacting systems or mixtures in general). Equilibrium is the main domain of classical, equilibrium, thermodynamics that has elaborated elegant criteria (or, perhaps, definitions) of equilibria and has shown how they naturally lead to the well known equilibrium constant. On the other hand, kinetics describes the way to equilibrium, i.e. the nonequilibrium state of chemical reactions, but also gives a clear idea on reaction equilibrium. Combining these two views various results on compatibility between thermodynamics and kinetics, on thermodynamic restrictions to kinetics etc. were published. The main idea can be illustrated on the trivial example of decomposition reaction  $AB = A + B$  with rate (kinetic) equation  $r = \bar{k}c_{AB} - \bar{k}c_Ac_B$  where  $r$  is the reaction rate,  $\bar{k}, \bar{k}$  are the forward and reverse rate constants, and  $c_\alpha$  are the concentrations. In equilibrium, the reaction rate is zero, consequently  $\bar{k} / \bar{k} = (c_Ac_B / c_{AB})_{\text{eq}}$ . Because the right hand side corresponds to the thermodynamic equilibrium constant ( $K$ ) it is concluded that  $K = \bar{k} / \bar{k}$ . However, this is simplified approach not taking into account conceptual differences between the true thermodynamic equilibrium constant and the ratio of rate constants that is called here the kinetic equilibrium constant. This discrepancy is sometimes to be removed by restricting this approach to ideal systems of elementary reactions but even then some questions remain.

Chemical potential ( $\mu$ ) is introduced into chemical kinetics by similar straightforward way (Qian & Beard, 2005). If it is expressed by  $\mu_\alpha = \mu_\alpha^\circ + RT \ln c_\alpha$ , multiplied by stoichiometric coefficients, summed and compared with rate equation it is obtained for the given example that:

$$\Delta\mu \equiv -\mu_{AB} + \mu_A + \mu_B = RT \ln \frac{c_A c_B}{K c_{AB}} = RT \ln(\bar{r} / \bar{r}) \quad (1)$$

(note that the equivalence of thermodynamic and kinetic equilibrium constants is supposed again;  $\bar{r}, \bar{r}$  are the forward and reverse rates). Equation (1) used to be interpreted as determining the (stoichiometric) sum of chemical potentials ( $\Delta\mu$ ) to be some (thermodynamic) “driving force” for reaction rates. In fact, there is “no kinetics”, no kinetic variables in the final expression  $\Delta\mu = RT \ln(\bar{r} / \bar{r})$  and reaction rates are directly determined by chemical potentials what is questionable and calls for experimental verification.

## 2. Restrictions put by thermodynamics on values of rate constants

### 2.1 Basic thermodynamic restrictions on rate constants coming from equilibrium

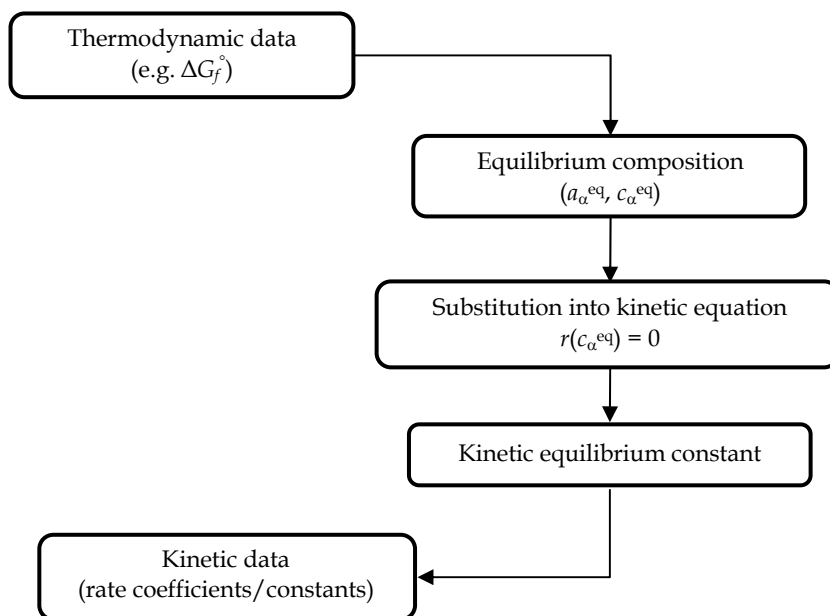
Perhaps the only one work which clearly distinguishes kinetic and thermodynamic equilibrium constant is the kinetic textbook by Eckert and coworkers (Eckert et al., 1986); the former is in it called the empirical equilibrium constant. This book stresses different approaches of thermodynamics and kinetics to equilibrium. In thermodynamics, equilibrium is defined as a state of minimum free energy (Gibbs energy) and its description is based on stoichiometric equation and thermodynamic equilibrium constant containing activities. Different stoichiometric equations of the same chemical equation can give different values of thermodynamic equilibrium constant, however, equilibrium composition is independent on selected stoichiometric equation. Kinetic description of equilibrium is based on zero overall reaction rate, on supposed reaction mechanism or network (reaction scheme) and corresponding kinetic (rate) equation. Kinetic equilibrium constant usually contains concentrations. According to that book, thermodynamic equilibrium data should be introduced into kinetic equations indirectly as shown in the Scheme 1.

Simple example reveals basic problems. Decomposition of carbon monoxide occurs (at the pressure  $p$ ) according to the following stoichiometric equation:



Standard state of gaseous components is selected as the ideal gas at 101 kPa and for solid component as the pure component at the actual pressure (due to negligible effects of pressure on behavior of solid components, the dependence of the standard state on pressure can be neglected here). Ideal behavior is supposed. Then  $a_\alpha = p_\alpha / p^\circ = p_{\text{rel}} n_\alpha / n_\Sigma$  for  $\alpha = \text{CO}$ ,  $\text{CO}_2$ , where  $p_{\text{rel}} = p / p^\circ$ , and  $a_{\text{C}} = 1$ ;  $a_\alpha$  is the activity,  $p_\alpha$  is the partial pressure,  $p^\circ$  the standard pressure,  $n_\alpha$  is the number of moles, and  $n_\Sigma$  the total number of moles. Thermodynamic equilibrium constant is then given by

$$K = \left( \frac{n_\Sigma n_{\text{CO}_2}}{p_{\text{rel}} n_{\text{CO}}^2} \right)_{\text{eq}} = \left[ \frac{(c_{\text{CO}} + c_{\text{CO}_2}) c_{\text{CO}_2}}{p_{\text{rel}} c_{\text{CO}}^2} \right]_{\text{eq}} \quad (2)$$



Scheme 1. Connecting thermodynamics and kinetics correctly (Eckert et al., 1986)

On contrary, the ratio of rate constants is given by

$$\left(\frac{\bar{k}}{\bar{k}}\right)_{\text{eq}} = \left(\frac{c_{\text{CO}_2} c_{\text{C}}}{c_{\text{CO}}^2}\right)_{\text{eq}} \quad (3)$$

It is clear that thermodynamic and kinetic equilibrium constants need not be equivalent even in ideal systems. For example, the former does not contain concentration of carbon and though this could be remedied by stating that carbon amount does not affect reaction rate and its concentration is included in the reverse rate constant, even then the kinetic equilibrium constant could depend on carbon amount in contrast to the thermodynamic equilibrium constant. Some discrepancies could not be remedied by restricting on elementary reactions only - in this example the presence of  $p_{\text{rel}}$  and of the total molar amount, generally, the presence of quantities transforming composition variables into standard state-related (activity-related) variables, and, of course, discrepancy in dimensionalities of the two equilibrium constants.

Let us use the same example to illustrate the procedure suggested by Eckert et al. (1986). At 1300 K and 202 kPa the molar standard Gibbs energies are (Novák et al., 1999):  $G_m^\circ(\text{CO}) = -395.3 \text{ kJ/mol}$ ,  $G_m^\circ(\text{CO}_2) = -712.7 \text{ kJ/mol}$ ,  $G_m^\circ(\text{C}) = -20.97 \text{ kJ/mol}$  and from them the value of thermodynamic equilibrium constant is calculated:  $K = 0.00515$ . Equilibrium molar balance gives  $(n_{\text{CO}_2})_{\text{eq}} = (n_{\text{C}})_{\text{eq}} = x$ ,  $(n_{\text{CO}})_{\text{eq}} = 1 - 2x$ ,  $n_\Sigma = 1 - x$ . Then from (2) follows  $x = 0.0107$  (Novák et al., 1999). Equilibrium composition is substituted into (3):

$$\left(\frac{\bar{k}}{\bar{k}}\right)_{\text{eq}} = \frac{0.0107 \times 0.0107}{0.09786^2} = 0.012 \quad (4)$$

and this is real and true result of thermodynamic restriction on values of rate constants valid at given temperature. More precisely, this is a restriction put on the ratio of rate constants, values of which are supposed to be independent on equilibrium, in other words, dependent on temperature (and perhaps on pressure) only and therefore this restriction is valid also out of equilibrium at given temperature. The numerical value of this restriction is dependent on temperature and should be recalculated at every temperature using the value of equilibrium constant at that temperature.

Thus, simple and safe way how to relate thermodynamics and kinetics, thermodynamic and kinetic equilibrium constants, and rate constants is that shown in Scheme 1. However, it gives no general equations and should be applied specifically for each specific reaction (reacting system) and reaction conditions (temperature, at least). There are also works that try to resolve relationship between the two types of equilibrium constant more generally and, in the same time, correctly and consistently. They were reviewed previously and only main results are presented here, in the next section. But before doing so, let us note that kinetic equilibrium constant can be used as a useful indicator of the distance of actual state of reacting mixture from equilibrium and to follow its approach to equilibrium. In the previous example, actual value of the fraction  $c_{\text{CO}_2} c_{\text{C}} / c_{\text{CO}}^2$  can be compared with the value of the ratio  $\bar{k} / \bar{k}$  and relative distance from equilibrium calculated, for more details and other examples see our previous work (Pekař & Koubek, 1997, 1999, 2000).

## 2.2 General thermodynamic restrictions on rate constants

As noted in the preceding section there are several works that do not rely on simple identification of thermodynamic and kinetic equilibrium constants. Hollingsworth (1952a, 1952b) generalized restriction on the ratio of forward and reverse reaction rates ( $f$ ) defined by

$$f(c_{\alpha}, T) = \bar{f}(c_{\alpha}, T) / \bar{f}(c_{\alpha}, T) \equiv \bar{r} / \bar{r} \quad (5)$$

Hollingsworth showed that sufficient condition for consistent kinetic and thermodynamic description of equilibrium is

$$F(Q_r, T) = \Phi(Q_r/K) \text{ and } \Phi(1) = 1 \quad (6)$$

where  $F$  is the function  $f$  with transformed variables,  $F(Q_r, T) = f(c_{\alpha}, T)$ , and  $Q_r$  is the well known reaction quotient. The first equality in (6) says that function  $F$  should be expressible as a function  $\Phi$  of  $Q_r/K$ . This is too general condition saying explicitly nothing about rate constants. Identifying kinetic equilibrium constant with thermodynamic one, condition (6) is specialized to

$$\Phi(Q_r/K) = (Q_r/K)^{-z} \quad (7)$$

where  $z$  is a positive constant. Equation (7) is a generalization of simple identity  $K = \bar{k} / \bar{k}$  from introduction. Hollingsworth also derived the necessary consistency condition:



$$f - 1 = (Q_r/K - 1) \Psi(c_w, T, u_j) \quad (8)$$

in the neighbourhood of  $Q_r/K = 1$  (i.e., of equilibrium);  $u_j$  stands for a set of non-thermodynamic variables. Example of practical application of Hollingsworth's approach in an ideal system is given by Boyd (Boyd, 1977).

Blum (Blum & Luus, 1964) considered a general mass action rate law formulated as follows:

$$r = \bar{k} \phi \prod_{\alpha=1}^m a_{\alpha}^{\omega_{\alpha}} - \bar{k} \phi \prod_{\alpha=1}^m a_{\alpha}^{\omega'_{\alpha}} \quad (9)$$

where  $\phi$  is some function of activities,  $a_w$  of reacting species,  $\omega_{\alpha}$  and  $\omega'_{\alpha}$  are coefficients which may differ from the stoichiometric coefficients ( $\nu_{\alpha}$ ), in fact, reaction orders. Supposing that both the equilibrium constant and the ratio of the rate constants are dependent only on temperature, they proved that

$$\bar{k} / \bar{k} = K^z \quad (10)$$

where

$$z = (\omega'_{\alpha} - \omega_{\alpha}) / \nu_{\alpha}; \quad \alpha = 1, \dots, n \quad (11)$$

General law (9) is rarely used in chemical kinetics, in reactions of ions it probably does not work (Laidler, 1965; Boudart, 1968). It can be transformed, particularly simply in ideal systems, to concentrations. Samohýl (personal communication) pointed out that criteria (11) may be problematic, especially for practically irreversible reactions. For example, reaction orders for reaction  $4 \text{NH}_3 + 6 \text{NO} = 5 \text{N}_2 + 6 \text{H}_2\text{O}$  were determined as follows:  $\omega_{\text{NH}_3} = 1$ ,  $\omega_{\text{NO}} = 0.5$ ,  $\omega_{\text{N}_2} = \omega_{\text{H}_2\text{O}} = 0$ . Orders for reversed direction are unknown, probably because of practically irreversible nature of the reaction. Natural selection could be, e.g.,  $\omega'_{\text{NO}} = 0$  (reaction is not inhibited by reactant), then  $z = 1/12$  and from this follows  $\omega'_{\text{NH}_3} = 2/3$  which seems to be improbable (rather strong inhibition by reactant).

### 2.3 Independence of reactions, Wegscheider conditions

Wegscheider conditions belong also among "thermodynamic restrictions" on rate constants and have been introduced more than one hundred years ago (Wegscheider, 1902). In fact, they are also based on equivalence between thermodynamic and kinetic equilibrium constants disputed in previous sections. Recently, matrix algebra approaches to find these conditions were described (Vlad & Ross, 2009). Essential part of them is to find (in)dependent chemical reactions. Problem of independent and dependent reactions is an interesting issue sometimes found also in studies on kinetics and thermodynamics of reacting mixtures. As a rule, a reaction scheme, i.e. a set of stoichiometric equations (whether elementary or nonelementary), is proposed, stoichiometric coefficients are arranged into stoichiometric matrix and linear (matrix) algebra is applied to find its rank which determines the number of linearly (stoichiometrically) independent reactions; all other reactions can be obtained as linear combinations of independent ones. This procedure can be viewed as an a posteriori analysis of the proposed reaction mechanism or network. Bowen has shown (Bowen, 1968) that using not only matrix but also vector algebra interesting results can be obtained on the basis of knowing only components of reacting

mixture, i.e. with no reaction scheme. This is a priori type of analysis and is used in continuum nonequilibrium (rational) thermodynamics. Because Bowen's results are important for this article they are briefly reviewed now for reader's convenience.

Let a reacting mixture be composed from  $n$  components (compounds) which are formed by  $z$  different atoms. Atomic composition of each component is described by numbers  $T_{\sigma\alpha}$  that indicate the number of atoms  $\sigma$  ( $= 1, 2, \dots, z$ ) in component  $\alpha$  ( $= 1, 2, \dots, n$ ). Atomic masses  $M_a^\sigma$  in combination with these numbers determine the molar masses  $M_\alpha$ :

$$M_\alpha = \sum_{\sigma=1}^z M_a^\sigma T_{\sigma\alpha} \quad (12)$$

Although compounds are destroyed or created in chemical reactions the atoms are preserved. If  $J^\alpha$  denotes the number of moles of the component  $\alpha$  formed or reacted per unit time in unit volume, i.e. the reaction rate for the component  $\alpha$  (component rate in short), then the persistence of atoms can be formulated in the form

$$\sum_{\alpha=1}^n T_{\sigma\alpha} J^\alpha = 0; \quad \sigma = 1, 2, \dots, z \quad (13)$$

This result expresses, in other words, the mass conservation.

Atomic numbers can be arranged in matrix  $\|T_{\sigma\alpha}\|$  of dimension  $z \times n$ . Chemical reactions are possible if its rank ( $h$ ) is smaller than the number of components ( $n$ ), otherwise the system (13) has only trivial solution, i.e. is valid only for zero component rates. If  $h < z$  then a new  $h \times n$  matrix  $\|S_{\sigma\alpha}\|$  with rank  $h$  can be constructed from the original matrix  $\|T_{\sigma\alpha}\|$  and used instead of it:

$$\sum_{\alpha=1}^n S_{\sigma\alpha} J^\alpha = 0; \quad \sigma = 1, 2, \dots, h \quad (14)$$

In this way only linearly independent relations from (13) are retained and from the chemical point of view it means that instead of (some) atoms with masses  $M_a^\sigma$  only some their linear combinations with masses  $M_e^\sigma$  should be considered as elementary building units of components:

$$M_\alpha = \sum_{\sigma=1}^h M_e^\sigma S_{\sigma\alpha} \quad (15)$$

Example. Mixture of  $\text{NO}_2$  and  $\text{N}_2\text{O}_4$  has the matrix  $\|T_{\sigma\alpha}\|$  of dimension  $2 \times 2$  and rank 1; the matrix  $\|S_{\sigma\alpha}\|$  is of dimension  $1 \times 2$  and can be selected as  $\begin{pmatrix} 1 & 2 \end{pmatrix}$  which means that the elementary building unit is  $\text{NO}_2$  and  $M_e^1 = M_a^1 + 2M_a^2 \equiv M_a^N + 2M_a^O$ .

Multiplying each of the  $z$  relations (13) by corresponding  $M_a^\sigma$  and summing the results for

all  $\sigma$  it follows that  $\sum_{\alpha=1}^n M_\alpha J^\alpha = 0$ . This fact can be much more effectively formulated in vector form because further important implications than follow. The last equality indicates

that component molar masses and rates should form two perpendicular vectors, i.e. vectors with vanishing scalar product. Let us introduce  $n$ -dimensional vector space, called the component space and denoted by  $U$ , with base vectors  $\mathbf{e}_\alpha$  and reciprocal base vectors  $\mathbf{e}^\alpha$  ( $\alpha = 1, 2, \dots, n$ ). Then the vector of molar masses  $\mathbf{M}$  and the vector of reaction rates  $\mathbf{J}$  are defined in this space as follows:

$$\mathbf{M} = \sum_{\alpha=1}^n M_\alpha \mathbf{e}^\alpha, \quad \mathbf{J} = \sum_{\alpha=1}^n J^\alpha \mathbf{e}_\alpha \quad (16)$$

To proceed further we use relations (14) and (15) because in contrast to relations (12) and (13) the matrix  $\|S_{\sigma\alpha}\|$  is of "full rank" (does not contain linearly dependent rows). The product of the two vectors can be then expressed in the following form:

$$\mathbf{M} \cdot \mathbf{J} = \left( \sum_{\alpha=1}^n \sum_{\sigma=1}^h M_e^\sigma S_{\sigma\alpha} \mathbf{e}^\alpha \right) \cdot \left( \sum_{\alpha=1}^n J^\alpha \mathbf{e}_\alpha \right) = \left( \sum_{\sigma=1}^h M_e^\sigma \sum_{\alpha=1}^n S_{\sigma\alpha} \mathbf{e}^\alpha \right) \cdot \left( \sum_{\alpha=1}^n J^\alpha \mathbf{e}_\alpha \right) = 0 \quad (17)$$

where the latter equality follows using (14). Because the matrix  $\|S_{\sigma\alpha}\|$  has rank  $h$ , the vectors

$$\mathbf{f}_\sigma = \sum_{\alpha=1}^n S_{\sigma\alpha} \mathbf{e}^\alpha; \quad \sigma = 1, 2, \dots, h \quad (18)$$

that appear in (17) are linearly independent and thus form a basis of a  $h$ -dimensional subspace  $W$  of the space  $U$  (remember that  $h < n$ ). This subspace unambiguously determines complementary orthogonal subspace  $V$  (of dimension  $n-h$ ), i.e.  $U = V \oplus W$ ,  $V \perp W$ . From (17) follows:

$$\mathbf{M} = \sum_{\sigma=1}^h M_e^\sigma \mathbf{f}_\sigma \quad (19)$$

which shows that  $\mathbf{M}$  can be expressed in the basis of the subspace  $W$  or  $\mathbf{M} \in W$ . From (14) and (16)<sub>2</sub> follows:

$$\mathbf{J} \cdot \mathbf{f}_\sigma = 0; \quad \sigma = 1, 2, \dots, h \quad (20)$$

which means that  $\mathbf{J}$  is perpendicular to all basis vectors of the subspace  $W$ , consequently,  $\mathbf{J}$  lies in the complementary orthogonal subspace  $V$ ,  $\mathbf{J} \in V$ .

Let us now select basis vectors in the subspace  $V$  and denote them  $\mathbf{d}^p$ ,  $p = 1, 2, \dots, n-h$ . Of course, these vectors lie also in the (original) space  $U$  and can be expressed using its basis vectors analogically to (16):

$$\mathbf{d}^p = \sum_{\alpha=1}^n P^{p\alpha} \mathbf{e}_\alpha \quad (21)$$

Because of orthogonality of subspaces  $V$  and  $W$ , their bases conform to equation

$$\mathbf{f}_\sigma \cdot \mathbf{d}^p = \sum_{\alpha=1}^n S_{\sigma\alpha} P^{p\alpha} = 0 \quad (22)$$

which can be alternatively written in matrix form as

$$\|P^{p\alpha}\| \times \|S_{\sigma\alpha}\|^T = \|0\| \quad (23)$$

Meaning of the matrix  $\|P^{p\alpha}\|$  can be deduced from two consequences. First, because the reaction vector  $\mathbf{J}$  lies in the subspace  $\mathbf{V}$ , it can be expressed also using its basis vectors,

$\mathbf{J} = \sum_{p=1}^{n-h} J_p \mathbf{d}^p$ . Substituting for  $\mathbf{J}$  from (16)<sub>2</sub> and for  $\mathbf{d}^p$  from (21), it follows:

$$J^\alpha = \sum_{p=1}^{n-h} J_p P^{p\alpha}; \quad \alpha = 1, 2, \dots, n \quad (24)$$

Second, because the vector of molar masses  $\mathbf{M}$  is in the subspace  $\mathbf{W}$ , it is perpendicular to all vectors  $\mathbf{d}^p$  and thus

$$0 = \mathbf{d}^p \cdot \mathbf{M} = \sum_{\alpha=1}^n P^{p\alpha} M_\alpha; \quad p = 1, 2, \dots, n-h \quad (25)$$

as follows after substitution from (19), (21), (22). Eq. (25) shows that matrix  $\|P^{p\alpha}\|$  enables to express component rates in  $n-h$  quantities  $J_p$  which are, in fact, rates of  $n-h$  independent reactions shown by (25) if instead of molar masses  $M_\alpha$  the corresponding chemical symbols are used. In other words  $\|P^{p\alpha}\|$  is the matrix of stoichiometric coefficients of component  $\alpha$  in (independent) reaction  $p$ .

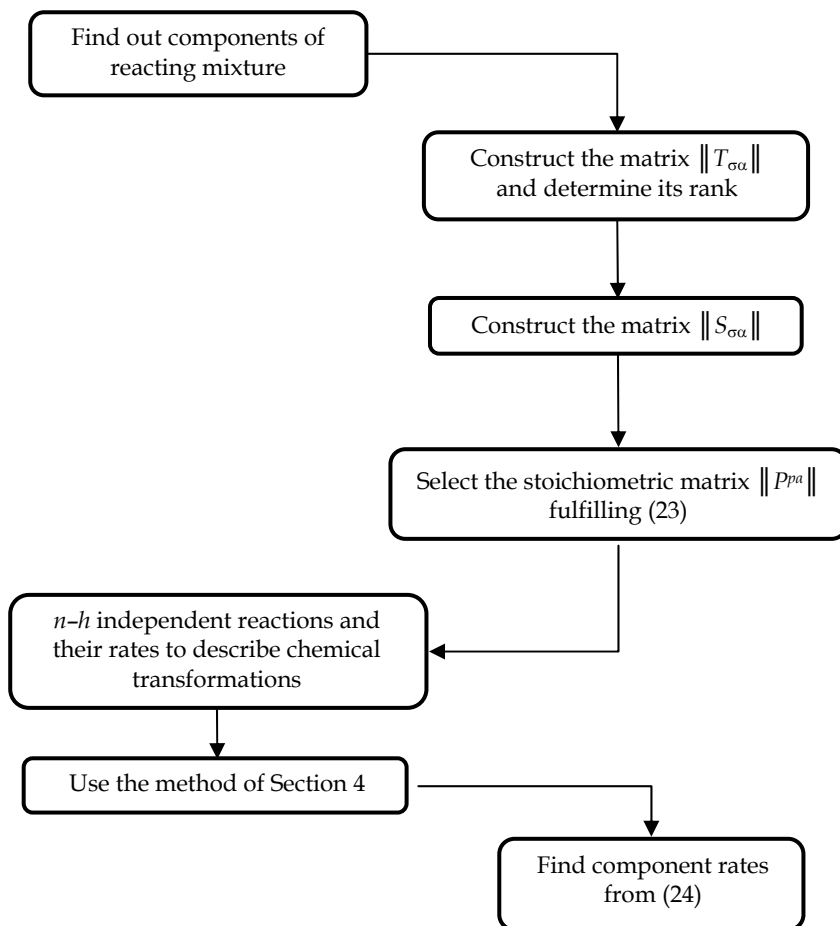
Vector algebra thus shows that chemical transformations fulfilling persistence of atoms (mass conservation) can be equivalently described either by component reaction rates or by rates of independent reactions. The number of the former is equal to the number of components ( $n$ ) whereas the number of the latter is lower ( $n-h$ ) which could decrease the dimensionality of the problem of description of reaction rates. In kinetic practice, however, changes in component concentrations (amounts) are measured, i.e. data on component rates and not on rates of individual reactions are collected. Reactions, in the form of reaction schemes, are suggested a posteriori on the basis of detected components, their concentrations changing in time and chemical insight. Then dependencies between reactions can be searched. Vector analysis offers rather different procedure outlined in Scheme 2. Dependencies are revealed at the beginning and then only independent reactions are included in the (kinetic) analysis. Vector analysis also shows how to transform (measured) component rates into (suggested, selected) rates of independent reactions. This transformation is made by standard procedure for interchange between vector bases or between vector coordinates in different bases. First, the contravariant metric tensor with components  $d^p = \mathbf{d}^r \cdot \mathbf{d}^p$  is constructed and then its inversion (covariant metric tensor) with components  $d_{rp}$  is found. From  $\mathbf{J} = \sum_{p=1}^{n-h} J_p \mathbf{d}^p$  it follows that  $\mathbf{J} \cdot \mathbf{d}_r = \sum_{p=1}^{n-h} J_p \mathbf{d}^p \cdot \mathbf{d}_r \equiv J_r$ . Using in

the latter equation the well known relationship between metric tensors and corresponding base vectors and the definition of base vectors (21) it finally follows:

$$J_p = \sum_{\alpha=1}^n \left( J^\alpha \sum_{r=1}^{n-h} P^{r\alpha} d_{rp} \right); \quad p = 1, 2, \dots, n-h \quad (26)$$

Of course, so far we have seen only relationships between reaction rates and no explicit equations for them like, e.g., the kinetic mass action law. Analysis based only on permanence of atoms cannot give such equations - they belong to the domain of chemical kinetics although they can also be devised by thermodynamics, see Section 4.

Simple example on Wegscheider conditions was presented by Vlad and Ross (Vlad & Ross, 2009) - isomerization taking place in two ways:



Scheme 2. Alternative procedure to find reaction rates

Vlad and Ross note that if the (thermodynamic) equilibrium constant is  $K = (c_B / c_A)_{\text{eq}}$  and if kinetic equations are expressed e.g.  $r_1 = \bar{k}_1 c_A - \bar{k}_1 c_B$  then the consistency between

thermodynamic and kinetic description of equilibrium is achieved only if the following (Wegscheider) condition holds:

$$\bar{k}_1 / \bar{k}_1 = \bar{k}_2 / \bar{k}_2 = K \quad (27)$$

It can be easily checked that in this mixture of one kind of atom and two components the rank of the matrix  $\|T_{\sigma\alpha}\|$  (dimension  $1 \times 2$ ) is 1 and there is only one independent reaction. The matrix  $\|S_{\sigma\alpha}\|$  can be selected as equal to the matrix  $\|T_{\sigma\alpha}\|$  and then the stoichiometric matrix can be selected as  $(-1 \ 1)$  which corresponds to the first reaction ( $A = B$ ) selected as the independent reaction. There is one base vector  $\mathbf{d}^1 = -\mathbf{e}_1 + \mathbf{e}_2$  giving one component contravariant tensor  $d^{11} = 2$  and corresponding component of covariant tensor  $d_{11} = 1/2$ . Consequently, the rate of the independent reaction is related to component reaction rates by:

$$J^A = J_1 P^{11} = -J_1, \quad J^B = J_1 P^{12} = J_1 \quad (28)$$

and  $J^A = -J^B$  which follows also from (14). Kinetics of transformations in a mixture of two isomers can be thus fully described by one reaction rate only – either from the two component rates can be measured and used for this purpose, the other component rate is then determined by it, can be calculated from it. At this stage of analysis there is no indication that two reactions should be considered and this should be viewed as some kind of “external” information coming perhaps from experiments. At the same time this analysis does not provide any explicit expression for reaction rate and its dependence on concentration – this is another type of external information coming usually from kinetics. Let us therefore suppose the two isomerization processes given above and their rates formulated in the form of kinetic mass action law:

$$r_1 = \bar{k}_1 c_A - \bar{k}_1 c_B, \quad r_2 = \bar{k}_2 c_A^2 - \bar{k}_2 c_A c_B \quad (29)$$

Then the only one independent reaction rate is in the form  $J_1 = r_1 + r_2$ . Note, that although the first reaction has been selected as the independent reaction, the rate of independent reaction is not equal to (its mass action rate)  $r_1$ . This interesting finding has probably no specific practical implication. However, individual traditional rates ( $r_i$ ) should not be independent. Let us suppose that  $r_2$  is dependent on  $r_1$ , i.e. can be expressed through it:  $r_2 = br_1$ ; then

$$\left(b\bar{k}_1 - \bar{k}_2 c_A\right) c_A - \left(b\bar{k}_1 - \bar{k}_2 c_A\right) c_B = 0 \quad (30)$$

should be valid for any concentrations. Sufficient conditions for this are  $b = \bar{k}_2 c_A / \bar{k}_1 = \bar{k}_2 c_A / \bar{k}_1$  and from them follows:

$$\bar{k}_1 \bar{k}_2 = \bar{k}_1 \bar{k}_2 \quad (31)$$

i.e. “kinetic part” of Wegscheider condition (27). Substituting derived expressions for  $b$  into  $br_1$  it can be easily checked that  $r_2$  really results. Although the derivation is rather straightforward and is not based on linear dependency with constant coefficients it points to assumption that Wegscheider conditions are not conditions for consistency of kinetics with thermodynamics but results of dependencies among reaction rates. Moreover, this derivation need not suppose equality of thermodynamic and kinetic equilibrium constant.

There is a thermodynamic method giving kinetic description in terms of independent reactions as noted in Scheme 2, see Section 4.

More complex reaction mixture and scheme was discussed by Ederer and Gilles (Ederer & Gilles, 2007). Their mixture was composed from six formal components (A, B, C, AB, BC, ABC) formed by three atoms (A, B, C). Three independent reactions are possible in this mixture while four reactions were considered by Ederer and Gilles (Ederer & Gilles, 2007)  $r_4 = b_1r_1 + b_2r_2 + b_3r_3$  with following mass action rate equations:

$$r_1 = \bar{k}_1c_Ac_B - \bar{k}_1c_{AB}, \quad r_2 = \bar{k}_2c_{AB}c_C - \bar{k}_2c_{ABC}, \quad r_3 = \bar{k}_3c_Bc_C - \bar{k}_3c_{BC}, \quad r_4 = \bar{k}_4c_Ac_{BC} - \bar{k}_4c_{ABC} \quad (32)$$

Let us suppose that the fourth reaction rate can be expressed through the other three rates:  $b_1r_1 + b_2r_2 + b_3r_3$ . By similar procedure as in the preceding example we arrive at conditions  $b_2 = \bar{k}_4 / \bar{k}_2$ ,  $b_3 = -\bar{k}_4c_A / \bar{k}_3$ , and  $b_1 = b_2\bar{k}_2c_C / \bar{k}_1 = -b_3\bar{k}_3c_C / \bar{k}_1c_A$  from which it follows that

$$\frac{\bar{k}_1\bar{k}_2\bar{k}_3\bar{k}_4}{\bar{k}_1\bar{k}_2\bar{k}_3\bar{k}_4} = 1 \quad (33)$$

i.e., Wegscheider condition derived in (Ederer & Gilles, 2007) from equilibrium considerations. Thus also here Wegscheider condition seems to be a result of mutual dependence of reaction rates and not a necessary consistency condition between thermodynamics and kinetics.

If reactions  $A + B = AB$ ,  $AB + C = ABC$ , and  $B + C = BC$  are selected as independent ones then (24) gives

$$J^A = -J_1, \quad J^B = -J_1 - J_3, \quad J^C = -J_2 - J_3, \quad J^{AB} = J_1 - J_2, \quad J^{BC} = J_3, \quad J^{ABC} = J_2 \quad (34)$$

Remember that, e.g.,  $J_1 \neq r_1$  but that the relationships between rates of independent reactions and mass action rates (32) follow from (34):

$$J_1 = r_1 + r_4, \quad J_2 = r_2 + r_4, \quad J_3 = r_3 - r_4 \quad (35)$$

Eq. (26) gives more complex expressions for independent rates, e.g.  $J_1 = -J^A/2 - J^B/4 + J^{AB}/4 - J^{BC}/4 + J^{ABC}/4$ , whereas from (24), i.e. (34), simply follow:  $J_1 = -J^A$ ,  $J_2 = J^{ABC}$ ,  $J_3 = J^{BC}$ . This is because the rates are considered as vector components - components  $J^\alpha$  of six dimensional space are transformed to components  $J_i$  in three dimensional subspace. Consequently, in practical applications (24) should be preferred in favor of (26) also to express  $J_i$  in terms of  $J^\alpha$ .

Message from the analysis of independence of reactions in this example is that it is sufficient to measure three component rates only ( $J^A$ ,  $J^{ABC}$ ,  $J^{BC}$ ); the remaining three component rates are determined by them. Although concentrations, i.e. component rates, are measured in kinetic experiments, results are finally expressed in reaction rates, rates of reactions occurring in suggested reaction scheme. Component rates are simply not sufficient in kinetic analysis and they are (perhaps always) translated into rates of reaction steps. However, from the three independent rates there cannot be unambiguously determined rates of four reactions in suggested reaction schemes as (35) demonstrates (three equations for four unknown  $r_i$ ). One equation more is needed and this is the above equation relating  $r_4$  to the remaining three rates. Equations containing  $r_i$  are too general and in practice are replaced by mass action expressions shown in (32) - eight parameters (rate constants) are thus

introduced in this example. They can be in principle determined from three equations (35) with the three measured independent reactions, four equations relating equilibrium composition (or thermodynamic equilibrium constant) and kinetic equilibrium constant and one Wegscheider condition (33), i.e. eight equations in total. Alternative thermodynamic method is described in Section 4.

Algebraically more rigorous is this analysis in the case of first order reactions as was illustrated on a mixture of three isomers and their triangular reaction scheme which is traditional example used to discuss consistency between thermodynamics and kinetics. Here, Wegscheider relations are consequences of linear dependence of traditional mass action reaction rates (Pekař, 2007).

#### 2.4 Note on standard states

Preceding sections demonstrated that one of the main problems to be solved when relating thermodynamics and kinetics is the transformation between activities and concentration variables. This is closely related to the selection of standard state (important and often overlooked aspect of relating thermodynamic and kinetic equilibrium constants) and to chemical potential. Standard states are therefore briefly reviewed in this section and chemical potential is subject of the following section.

Rates of chemical reactions are mostly expressed in terms of concentrations. Among standard states introduced and commonly used in thermodynamics there is only one based on concentration – the standard state of nonelectrolyte solute on concentration basis. Only this standard state can be directly used in kinetic equations. Standard state in gaseous phase or mixture is defined through (partial) pressure or fugacity. As shown above even in mixture of ideal gases it is impossible to simply use this standard state in concentration based kinetic equations. Although kinetic equations could be reformulated into partial pressures there still remains problem with the fact that standard pressure is fixed (at 1 atm or, nowadays, at  $10^5$  Pa) and its recalculation to actual pressure in reacting mixture may cause incompatibility of thermodynamic and kinetic equilibrium constants (see the factor  $p_{rel}$  in the example above in Section 2.1). This opens another problem – the very selection of standard state, particularly in relation to activity discussed in subsequent section. In principle, it can be selected arbitrarily, as dependent only on temperature or on temperature and pressure. Standard states strictly based on the (fixed) standard pressure are of the former type and only such will be considered in this article. All other states, including states dependent also on pressure, will be called the reference state; the same approach is used, e.g. by de Voe (de Voe, 2001).

The value of thermodynamic equilibrium constant and its dependence or independence on pressure is thus dependent on the selected standard (or reference) state. This is quite uncommon in chemical kinetics where the dependence of rate constants is not a matter of selection of standard states but result of experimental evidence or some theory of reaction rates. As a rule, rate constant is always function of temperature. Sometimes also the dependence on pressure is considered but this is usually the case of nonelementary reactions. Consequently, attempts to relate thermodynamic and kinetic equilibrium constants should select standard state consistently with functional dependence of rate constants. On the other hand, the method of Scheme 1 is self-consistent in this aspect because equilibrium composition is independent of the selection of standard state.



### 3. Chemical potential and activity revise

Chemical potential is used in discussions on thermodynamic implications on reaction rates, particularly in the form of (stoichiometric) difference between chemical potentials of reaction products and reactants and through its explicit relationship to concentrations (activities, in general). Before going into this type of analysis basic information is recapitulated.

Chemical potential is in classical, equilibrium thermodynamics defined as a partial derivative of Gibbs energy ( $G$ ):

$$\mu_{\alpha} = (\partial G / \partial n_{\alpha})_{T,p,n_{j \neq \alpha}} \quad (36)$$

Although another definitions through another thermodynamic quantities are possible (and equivalent with this one), the definition using the Gibbs energy is the most useful for chemical thermodynamics. Chemical potential expresses the effect of composition and this effect is also essential in chemical kinetics. To make the mathematical definition of the chemical potential applicable in practice its relationship to composition (concentration) should be stated explicitly. Practical chemical thermodynamics suggests that this is an easy task but we must be very careful and bear all (tacit) presumptions in mind to arrive at proper conclusions. Generally the explicit relationship between chemical composition and chemical potential is stated defining the activity of a component  $\alpha$ :

$$a_{\alpha} = \exp\left(\frac{\mu_{\alpha} - \mu_{\alpha}^{\circ}}{RT}\right) \quad (37)$$

which can be transformed to

$$\mu_{\alpha} = \mu_{\alpha}^{\circ} + RT \ln a_{\alpha} \quad (38)$$

but this still lacks direct interconnection/linkage to measurable concentrations. Just this is the main problem of applying chemical potential (and activities) in rate equations which systematically use molar concentrations. Even when reaction rates would be expressed using activities in place of concentrations the activities should be properly calculated from the measured concentrations, in other words, the concentrations should be correctly transformed to the activities. Activity is very easily related to measurable composition variable in the case of mixture of ideal gases. Providing that Gibbs energy is a function of temperature, pressure and molar amounts, following relation is well known from thermodynamics for the partial molar volume:  $\bar{V}_{\alpha} = (\partial \mu_{\alpha} / \partial p)_{T,n_{j \neq \alpha}}$ . In a mixture of ideal gases partial molar volumes are equal to the molar volume of the mixture,  $V_m$  (Silbey et al., 2005). Because  $V_m = RT/p$  we can write:

$$RT / p = (\partial \mu_{\alpha,g} / \partial p_{\alpha})(\partial p_{\alpha} / \partial p) = x_{\alpha} (\partial \mu_{\alpha,g} / \partial p_{\alpha}) \quad (39)$$

and

$$RT / p_{\alpha} = (\partial \mu_{\alpha,g} / \partial p_{\alpha}) \quad (40)$$

Integration from the standard state to some actual state then yields

$$\mu_{\alpha,g} = \mu_{\alpha,g}^{\circ} + RT \ln(p_{\alpha} / p^{\circ}) \quad (41)$$

Comparing with the definition of activity it follows

$$a_{\alpha} = p_{\alpha} / p^{\circ} \quad (\text{mixture of ideal gases}) \quad (42)$$

Application of this relationship was illustrated in the example given above. Note that (42) was not derived from the definition of activity but comparing the properties of chemical potential in the ideal gas mixture (41) with the definition of activity. Note also that the partial derivative in the original definition of chemical potential is in general a function of molar amounts (contents) of all components but eq. (42) states that the chemical potential of a component  $\alpha$  is a function only of the content of that component.

In a real gas mixture, non-idealities should be taken into account, usually by substituting fugacity ( $f_{\alpha}$ ) for the partial pressure:

$$\mu_{\alpha,g} = \mu_{\alpha,g}^{\circ} + RT \ln(f_{\alpha} / p^{\circ}) \quad (43)$$

The fugacity can be eliminated in favor of directly measurable quantities using the fugacity coefficient  $\phi_{\alpha}$

$$f_{\alpha} = \phi_{\alpha} p_{\alpha} \quad (44)$$

and its relationship to the partial molar volume and the total pressure (de Voe, 2001):

$$\mu_{\alpha,g} = \mu_{\alpha,g}^{\circ} + RT \ln(p_{\alpha} / p^{\circ}) + \int_0^{p_{\alpha}} (\bar{V}_{\alpha} - RT / p) dp \quad (45)$$

It should be stressed that in derivation of the expression for the fugacity coefficient it was assumed that the Gibbs energy is a function of (only) temperature, pressure, and molar amounts of all components. Comparing with the definition of activity we have

$$a_{\alpha} = f_{\alpha} / p^{\circ} \quad (\text{mixture of gases}) \quad (46)$$

If kinetic equations for mixture of real gases are written in partial pressures then thermodynamic and kinetic equilibrium constants are incompatible due to the presence of fugacity coefficient or the integral in eq. (45). Kinetic equations for mixture of real gases could be formulated in terms of fugacities instead of concentrations (or partial pressures) to achieve compatibility between thermodynamic and kinetic equilibrium constants but even then the same problem remains with the presence of the standard pressure in thermodynamic relations. Kinetic equations formulated in fugacities are really rare – some success in this way was demonstrated by Eckert and Boudart (Eckert & Boudart, 1963) while Mason (Mason, 1965) showed, using the same data, that fugacities need not remedy the whole situation.

Similar derivation for liquid state (solutions) has different basis. It stems from the equilibrium between liquid and gaseous phase in which the following identity holds:  $\mu_{\alpha,g} = \mu_{\alpha,l}$ . Introducing expression (41) or (43) and using either Raoult's or Henry's law for the

relationship between compositions of equilibrated liquid and gaseous phases final form of  $\mu_{\alpha,l}$  dependence on the composition of liquid is obtained. For example, with Raoult's law  $p_{\alpha} = x_{\alpha}p_{\alpha}^*$  and ideal gas phase we have this equation

$$\mu_{\alpha,l} = \mu_{\alpha,g}^{\circ} + RT \ln(x_{\alpha}p_{\alpha}^*/p^{\circ}) \equiv \mu_{\alpha}^{\text{ref}} + RT \ln x_{\alpha} \quad (47)$$

which has, in fact, inspired the definition of an ideal (liquid, solid, or gas) mixture as a mixture with the chemical potential defined, at a given  $T$  and  $p$ , as  $\mu_{\alpha} = \mu_{\alpha}^{\text{ref}} + RT \ln x_{\alpha}$  where  $\mu_{\alpha}^{\text{ref}}$  is a function of both  $T$  and  $p$ . This definition, as well as the identity in (47), can be simply related to the definition of activity only if the standard state is selected consistently with the reference state, i.e. if the former is a function of both  $T$  and  $p$ . If the standard state is selected as dependent on temperature, as it should be, than the pressure factor ( $\Gamma_{\alpha}$ ) should be introduced (see, e.g., de Voe, 2001)

$$\Gamma_{\alpha} = \exp\left(\frac{\mu_{\alpha}^{\text{ref}} - \mu_{\alpha}^{\circ}}{RT}\right) \quad (48)$$

Then the activity of a (non-electrolyte) component in real solution is written as  $a_{\alpha} = \Gamma_{\alpha}\gamma_{\alpha}x_{\alpha}$  where  $\gamma_{\alpha}$  is the activity coefficient introduced by the equation  $\mu_{\alpha} = \mu_{\alpha}^{\text{ref}} + RT \ln(\gamma_{\alpha}x_{\alpha})$ . Introducing activities in place of concentrations means in this case to know the pressure factor and to transform molar fractions into molar concentrations to be consistent with thermodynamics.

The main problems with using activities defined for liquid systems can be summarized as follows. Activity is based on molar fractions whereas kinetic uses concentrations. Although there are formulas for the conversion of these variables they do not allow direct substitution, they introduce other variables (e.g., solution density) and lead to rather complex expression of thermodynamic equilibrium constant in concentrations. Whereas concentrations of all species are independent (variables) this is not true for molar fractions - value of one from them is unambiguously determined by values of remaining ones. Chemical potential in liquid and activity based on it are introduced on the basis of (liquid-gas) equilibrium while kinetics essentially works with reactions out of equilibrium. Applicability of equilibrium-based formulated in fugacities are really rare in nonequilibrium states deserves further study. The problem with molar fractions can be resolved by the use of molar concentration based Henry's law giving for ideal-dilute solution  $\mu_{\alpha,l} = \mu_{\alpha,c}^{\text{ref}} + RT \ln c_{\alpha}/c^{\circ}$ , however, rate equations should be formulated with the standard concentration. Sometimes following relationship is used:  $\mu_{\alpha,l} = \mu_{\alpha,c}^{\text{ref}} + RT \ln c_{\alpha}/c_{\Sigma}$  (Ederer & Gilles, 2007) where  $c_{\Sigma}$  is the sum of all concentrations. In this case, the invertibility for  $c_{\alpha}$  is problematic because it is included in  $c_{\Sigma}$ ; reaction rates should be then formulated in  $c_{\alpha}/c_{\Sigma}$  instead of concentrations that is quite unusual. Of course, the value of activity is dependent on the selected standard state, anyway. All attempts to relate thermodynamic and kinetic equilibrium constants should pay great attention to the selection of standard state and its consequences to be really rigorous and correct.

It is clear from this basic overview that chemical potential, activity and their interrelation are in principle equilibrium quantities which, in kinetic applications, are to be used for

non-equilibrium situations. Let us now trace one relatively simple non-equilibrium approach to description of chemically reacting systems and its results regarding the chemical potential. Samohýl has developed rational thermodynamic approach for chemically reacting fluids with linear transport properties (henceforth called briefly linear fluids) and these fluids seem to include many (non-electrolyte) systems encountered in chemistry (Samohýl & Malijevský, 1976; Samohýl, 1982, 1987). This is a continuum mechanics based approach working with densities of quantities and specific quantities (considered locally, in other words, as fields but this is not crucial for the present text) therefore it primarily uses densities of components (more precisely, the density of component mass) instead of their molar concentrations or fractions that are common in chemistry. This density, in fact, is known in chemistry as a mass concentration with dimension of mass per (unit) volume and can be thus easily recalculated to concentration quantities more common in chemistry. Chemical potential of a reacting component  $\alpha$  is defined in this theory as follows:

$$g_\alpha = \partial(\rho\bar{f}) / \partial\rho_\alpha \quad (49)$$

Here  $\rho$  is the density of mixture, i.e. the sum of all component densities  $\rho_\alpha$ , and  $\bar{f}$  is the specific free energy of (reacting) mixture as a function of relevant independent variables (the value of this function is denoted by  $f$ ). Inspiration for this definition came from the entropic inequality (the “second law” of thermodynamics) as formulated in rational thermodynamics generally for mixtures and from the fact that this definition enabled to derive classical (equilibrium) thermodynamic relations in the special case that is covered by classical theory. The chemical potential  $g_\alpha$  thus has the dimensions of energy per mass. The product  $\rho\bar{f}$  essentially transforms the specific quantity to its density and the definition (49) can be viewed as a generalization of the classical definition (36) – partial derivative of mixture free energy (as a function) with respect to an independent variable expressing the amount of a component.

The specific free energy  $\bar{f}$  is function of various (mostly kinematic and thermal) variables but here it is sufficient to note that component densities are among them, of course.

In the case of linear fluids it can be proved that free energy is function of densities and temperature only,  $f = \bar{f}(\rho_1, \rho_2, \dots, \rho_n, T)$ . The same result is proved also for chemical potentials  $g_\alpha$  and also for reaction rates expressed as component mass created or destroyed by chemical reactions at a given place and time in unit volume,  $r_\alpha = \bar{r}_\alpha(\rho_1, \rho_2, \dots, \rho_n, T)$ . These rates can be easily transformed to molar basis much more common in chemistry using the molar mass  $M_\alpha$ :  $J^\alpha = r_\alpha/M_\alpha$ . Component densities are directly related to molar concentration by a similar equation:  $c_\alpha = \rho_\alpha/M_\alpha$ . In this way, the well known kinetic empirical law – the law of mass action – is derived theoretically in the form:  $J^\alpha = \bar{J}^\alpha(c_1, c_2, \dots, c_n, T)$ . Apparently, activities could be introduced into this function as independent variables controlling reaction rates by means of relations as  $a_\alpha = \Gamma_\alpha \gamma_\alpha c_\alpha / c^\circ$  but this is not rigorous because these relations are consequences of chemical potential and its explicit dependence on mixture composition and not definitions per se. Therefore, chemical potentials should be introduced as independent variables at first. This could be done providing that component densities can be expressed as functions of chemical

potential, i.e. providing that functions  $g_\alpha = \bar{g}_\alpha(\rho_1, \rho_2, \dots, \rho_n, T)$  are invertible (with respect to densities). This invertibility is not self-evident and the best way would be to prove it. Samohýl has proved (Samohýl, 1982, 1987) that if mixture of linear fluids fulfils Gibbs' stability conditions then the matrix with elements  $\partial \bar{g}_\alpha / \partial \rho_\gamma$  ( $\alpha, \gamma = 1, \dots, n$ ) is regular which ensures the invertibility. This stability is a standard requirement for reasonable behavior of many reacting systems of chemist's interest, consequently the invertibility can be considered to be guaranteed and we can transform the rate functions as follows:

$$J^\alpha = \bar{J}^\alpha(\rho_1, \rho_2, \dots, \rho_n, T) = \hat{J}^\alpha(g_1, g_2, \dots, g_n, T) = \check{J}^\alpha(\mu_1, \mu_2, \dots, \mu_n, T) \quad (50)$$

where the last transformation was made using the following transformation of (specific) chemical potential into the traditional chemical potential (which will be called the molar chemical potential henceforth):  $\mu_\alpha = g_\alpha M_\alpha$ . Using the definition of activity (37) another transformation, to activities, can be made providing that the standard state is a function of temperature only:

$$\check{J}^\alpha(\mu_1, \mu_2, \dots, \mu_n, T) = \tilde{J}^\alpha(a_1, a_2, \dots, a_n, T) \quad (51)$$

It should be stressed that chemical potential of component  $\alpha$  as defined by (49) is a function of densities of all components, i.e. of  $\rho_\gamma$ ,  $\gamma = 1, \dots, n$ , therefore also the molar chemical potential is following function of composition:  $\mu_\alpha = \bar{\mu}_\alpha(c_1, c_2, \dots, c_n, T)$ . Note that generally any rate of formation or destruction ( $J^\alpha$ ) is a function of densities, or chemical potentials, or activities, etc. of all components.

Although the functions (dependencies) given above were derived for specific case of linear fluids they are still too general. Yet simpler fluid model is the simple mixture of fluids which is defined as mixture of linear fluids constitutive (state) equations of which are independent on density gradients. Then it can be shown (Samohýl, 1982, 1987) that

$$\partial \bar{f}_\alpha / \partial \rho_\gamma = 0 \quad \text{for } \alpha \neq \gamma; \quad \alpha, \gamma = 1, \dots, n \quad (52)$$

and, consequently, also that  $g_\alpha = \bar{g}_\alpha(\rho_\alpha, T)$ , i.e. the chemical potential of any component is a function of density of this component only (and of temperature). Mixture of ideal gases is defined as a simple mixture with additional requirement that partial internal energy and enthalpy are dependent on temperature only. Then it can be proved (Samohýl, 1982, 1987) that chemical potential is given by

$$g_\alpha = g_\alpha^\circ(T) + R_\alpha T \ln(p_\alpha / p^\circ) \quad (53)$$

that is slightly more general than the common model of ideal gas for which  $R_\alpha = R/M_\alpha$ . Thus the expression (41) is proved also at nonequilibrium conditions and this is probably only one mixture model for which explicit expression for the dependence of chemical potential on composition out of equilibrium is derived. There is no indication for other cases while the function  $g_\alpha = \bar{g}_\alpha(\rho_\alpha, T)$  should be just of the logarithmic form like (47). Let us check conformity of the traditional ideal mixture model with the definition of simple mixture. For solute in an ideal-dilute solution following concentration-based expression is used:

$$\mu_{\alpha} = \mu_{\alpha}^{\text{ref}} + RT \ln(c_{\alpha} / c^{\circ}) \quad (54)$$

where  $\mu_{\alpha}^{\text{ref}}$  includes (among other) the gas standard state and concentration-based Henry's constant. Changing to specific quantities and densities we obtain:

$$g_{\alpha} = \mu_{\alpha}^{\text{ref}} / M_{\alpha} + (RT / M_{\alpha}) \ln(\rho_{\alpha} / M_{\alpha} c^{\circ}) \quad (55)$$

which looks like a function of  $\rho_{\alpha}$  and  $T$  only, i.e. the simple mixture function  $g_{\alpha} = \bar{g}_{\alpha}(\rho_{\alpha}, T)$ . However, the referential state is a function of pressure so this is not such function rigorously. Except ideal gases there is probably no proof of applicability of classical expressions for dependence of chemical potential on composition out of equilibrium and no proof of its logarithmic point. There are probably also no experimental data that could help in resolving this problem.

#### 4. Solution offered by rational thermodynamics

Rational thermodynamics offers certain solution to problems presented so far. It should be stressed that this is by no means totally general theory resolving all possible cases. But it clearly states assumptions and models, i. e. scope of its potential application.

The first assumption, besides standard balances and entropic inequality (see, e.g., Samohýl, 1982, 1987), or model is the mixture of linear fluids in which the functional form of reaction rates was proved:  $J^{\alpha} = \bar{J}^{\alpha}(c_1, c_2, \dots, c_n, T)$  (Samohýl & Malijevský, 1976; Samohýl, 1982, 1987). Only independent reaction rates are sufficient that can be easily obtained from component rates, cf. (26) from which further follows that they are function of the same variables. This function,  $J_i = \bar{J}_i(c_1, c_2, \dots, c_n, T)$ , is approximated by a polynomial of suitable degree (Samohýl & Malijevský, 1976; Samohýl, 1982, 1987). Equilibrium constant is defined for each independent reaction as follows:

$$-RT \ln K_p = \sum_{\alpha=1}^n \mu_{\alpha}^{\circ} P^{p_{\alpha}}; \quad p = 1, 2, \dots, n-h \quad (56)$$

Activity (37) is supposed to be equal to molar concentrations (divided by unit standard concentration), which is possible for ideal gases, at least (Samohýl, 1982, 1987). Combining

this definition of activity with the proved fact that in equilibrium  $\sum_{\alpha=1}^n (\mu_{\alpha})_{\text{eq}} P^{p_{\alpha}} = 0$  (Samohýl, 1982, 1987) it follows

$$K_p = \prod_{\alpha=1}^n [(c_{\alpha})_{\text{eq}}]^{p_{\alpha}} \quad (57)$$

Some equilibrium concentrations can be thus expressed using the others and (57) and substituted in the approximating polynomial that equals zero in equilibrium. Equilibrium polynomial should vanish for any concentrations what leads to vanishing of some of its coefficients. Because the coefficients are independent of equilibrium these results are valid

also out of it and the final simplified approximating polynomial, called thermodynamic polynomial, follows and represents rate equation of mass action type. More details on this method can be found elsewhere (Samohýl & Malijevský, 1976; Pekař, 2009, 2010). Here it is illustrated on two examples relevant for this article.

First example is the mixture of two isomers discussed in Section 2. 3. Rate of the only one independent reaction, selected as  $A = B$ , is approximated by a polynomial of the second degree:

$$J_1 = k_{00} + k_{10}c_A + k_{01}c_B + k_{20}c_A^2 + k_{02}c_B^2 + k_{11}c_Ac_B \quad (58)$$

The concentration of B is expressed from the equilibrium constant,  $(c_B)_{\text{eq}} = K(c_A)_{\text{eq}}$  and substituted into (58) with  $J_1 = 0$ . Following form of the polynomial in equilibrium is obtained:

$$0 = k_{00} + (k_{10} + K k_{01})(c_A)_{\text{eq}} + (k_{20} + K^2 k_{02} + K k_{11})(c_A)_{\text{eq}}^2 \quad (59)$$

Eq. (59) should be valid for any values of equilibrium concentrations, consequently

$$k_{00} = 0; \quad k_{10} = -K k_{01}; \quad k_{20} = -K^2 k_{02} - K k_{11} \quad (60)$$

Substituting (60) into (58) the final thermodynamic polynomial (of the second degree) results:

$$J_1 = k_{10}(-Kc_A + c_B) + k_{02}(-K^2c_A^2 + c_B^2) + k_{11}(-Kc_A^2 + c_Ac_B) \quad (61)$$

Note, that coefficients  $k_{ij}$  are functions of temperature only and can be interpreted as mass action rate constants (there is no condition on their sign, if some  $k_{ij}$  is negative then traditional rate constant is  $k_{ij}$  with opposite sign). Although only the reaction  $A = B$  has been selected as the independent reaction, its rate as given by (61) contains more than just traditional mass action term for this reaction. Remember that component rates are given by (28). Selecting  $k_{02} = 0$  two terms remain in (61) and they correspond to the traditional mass action terms just for the two reactions supposed in (R2). Although only one reaction has been selected to describe kinetics, eq. (61) shows that thermodynamic polynomial does not exclude other (dependent) reactions from kinetic effects and relationship very close to  $J_1 = r_1 + r_2$ , see also (29), naturally follows. No Wegscheider conditions are necessary because there are no reverse rate constants. On contrary, thermodynamic equilibrium constant is directly involved in rate equation; it should be stressed that because no reverse constant are considered this is not achieved by simple substitution of  $K$  for  $\bar{k}_j$  from (27). Eq. (61) also extends the scheme (R2) and includes also bimolecular isomerization path:  $2A = 2B$ .

This example illustrated how thermodynamics can be consistently connected to kinetics considering only independent reactions and results of nonequilibrium thermodynamics with no need of additional consistency conditions.

Example of simple combination reaction  $A + B = AB$  will illustrate the use of molar chemical potential in rate equations. In this mixture of three components composed from two atoms only one independent reaction is possible. Just the given reaction can be selected with equilibrium constant defined by (56):  $\ln K = (-\mu_A^\circ - \mu_B^\circ + \mu_{AB}^\circ)/(-RT)$  and equal to

$K = (c_{AB} / c_A c_B)_{\text{eq}}$ , cf. (57). The second degree thermodynamic polynomial results in this case in following rate equation:

$$J_1 = k_{110}(c_A c_B - K^{-1} c_{AB}) \quad (62)$$

that represents the function  $J_1 = \bar{J}_1(T, c_A, c_B, c_{AB})$ . Its transformation to the function  $J_1 = \check{J}_1(T, \mu_A, \mu_B, \mu_{AB})$  gives:

$$J_1 = k_{110} \exp\left(\frac{-\mu_A^\circ - \mu_B^\circ}{RT}\right) \left[ \exp\left(\frac{\mu_A + \mu_B}{RT}\right) - \exp\left(\frac{\mu_{AB}}{RT}\right) \right] \quad (63)$$

This is thermodynamically correct expression (for the supposed thermodynamic model) of the function  $\check{J}$  discussed in Section 3 and in contrast to (1). It is clear that proper “thermodynamic driving force” for reaction rate is not simple (stoichiometric) difference in molar chemical potentials of products and reactants. The expression in square brackets can be considered as this driving force. Equation (63) also lucidly shows that high molar chemical potential of reactants in combination with low molar chemical potential of products can naturally lead to high reaction rate as could be expected. On the other hand, this is achieved in other approaches, based on  $\sum v_i \mu_i$ , due to arbitrary selection of signs of stoichiometric coefficients. In contrast to this straightforward approach illustrated in introduction, also kinetic variable ( $k_{110}$ ) is still present in eq. (63), explaining why some “thermodynamically highly forced” reactions may not practically occur due to very low reaction rate. Equation (63) includes also explicit dependence of reaction rate on standard state selection (cf. the presence of standard chemical potentials). This is inevitable consequence of using thermodynamic variables in kinetic equations. Because also the molar chemical potential is dependent on standard state selection, it can be perhaps assumed that these dependences are cancelled in the final value of reaction rate.

Rational thermodynamics thus provides efficient connection to reaction kinetics. However, even this is not totally universal theory; on the other hand, presumptions are clearly stated. First, the procedure applies to linear fluids only. Second, as presented here it is restricted to mixtures of ideal gases. This restriction can be easily removed, if activities are used instead of concentrations, i.e. if functions  $\check{J}$  are used in place of functions  $\bar{J}$  – all equations remain unchanged except the symbol  $a_\alpha$  replacing the symbol  $c_\alpha$ . But then still remains the problem how to find explicit relationship between activities and concentrations valid at non equilibrium conditions. Nevertheless, this method seems to be the most carefully elaborated thermodynamic approach to chemical kinetics.

## 5. Conclusion

Two approaches relating thermodynamics and chemical kinetics were discussed in this article. The first one were restrictions put by thermodynamics on the values of rate constants in mass action rate equations. This can be also formulated as a problem of relation, or even equivalence, between the true thermodynamic equilibrium constant and the ratio of forward and reversed rate constants. The second discussed approach was the use of chemical potential as a general driving force for chemical reaction and “directly” in rate equations.



Both approaches are closely connected through the question of using activities, that are common in thermodynamics, in place of concentrations in kinetic equations and the problem of expressing activities as function of concentrations.

Thermodynamic equilibrium constant and the ratio of forward and reversed rate constants are conceptually different and cannot be identified. Restrictions following from the former on values of rate constants should be found indirectly as shown in Scheme 1.

Direct introduction of chemical potential into traditional mass action rate equations is incorrect due to incompatibility of concentrations and activities and is problematic even in ideal systems.

Rational thermodynamic treatment of chemically reacting mixtures of fluids with linear transport properties offers some solution to these problems whenever its clearly stated assumptions are met in real reacting systems of interest. No compatibility conditions, no Wegscheider relations (that have been shown to be results of dependence among reactions) are then necessary, thermodynamic equilibrium constants appear in rate equations, thermodynamics and kinetics are connected quite naturally. The role of ("thermodynamically") independent reactions in formulating rate equations and in kinetics in general is clarified.

Future research should focus attention on the applicability of dependences of chemical potential on concentrations known from equilibrium thermodynamics in nonequilibrium states, or on the related problem of consistent use of activities and corresponding standard states in rate equations.

Though practical chemical kinetics has been successfully surviving without special incorporation of thermodynamic requirements, except perhaps equilibrium results, tighter connection of kinetics with thermodynamics is desirable not only from the theoretical point of view but may be of practical importance considering increasing interest in analyzing of complex biochemical network or increasing computational capabilities for correct modeling of complex reaction systems. The latter when combined with proper thermodynamic requirements might contribute to more effective practical, industrial exploitation of chemical processes.

## 6. Acknowledgment

The author is with the Centre of Materials Research at the Faculty of Chemistry, Brno University of Technology; the Centre is supported by project No. CZ.1.05/2.1.00/01.0012 from ERDF. The author is indebted to Ivan Samohýl for many valuable discussions on rational thermodynamics.

## 7. References

- Blum, L.H. & Luus, R. (1964). Thermodynamic Consistency of Reaction Rate Expressions. *Chemical Engineering Science*, Vol.19, No.4, pp. 322-323, ISSN 0009-2509
- Boudart, M. (1968). *Kinetics of Chemical Processes*, Prentice-Hall, Englewood Cliffs, USA
- Bowen, R.M. (1968). On the Stoichiometry of Chemically Reacting Systems. *Archive for Rational Mechanics and Analysis*, Vol.29, No.2, pp. 114-124, ISSN 0003-9527
- Boyd, R.K. (1977). Macroscopic and Microscopic Restrictions on Chemical Kinetics. *Chemical Reviews*, Vol.77, No.1, pp. 93-119, ISSN 0009-2665

- De Voe, H. (2001). *Thermodynamics and Chemistry*, Prentice Hall, ISBN 0-02-328741-1, Upper Saddle River, USA
- Eckert, C.A. & Boudart, M. (1963). Use of Fugacities in Gas Kinetics. *Chemical Engineering Science*, Vol.18, No.2, 144-147, ISSN 0009-2509
- Eckert, E.; Horák, J.; Jiráček, F. & Marek, M. (1986). *Applied Chemical Kinetics*, SNTL, Prague, Czechoslovakia (in Czech)
- Ederer, M. & Gilles, E.D. (2007). Thermodynamically Feasible Kinetic Models of Reaction Networks. *Biophysical Journal*, Vol.92, No.6, pp. 1846-1857, ISSN 0006-3495
- Hollingsworth, C.A. (1952a). Equilibrium and the Rate Laws for Forward and Reverse Reactions. *Journal of Chemical Physics*, Vol.20, No.5, pp. 921-922, ISSN 0021-9606
- Hollingsworth, C.A. (1952b). Equilibrium and the Rate Laws. *Journal of Chemical Physics*, Vol.20, No.10, pp. 1649-1650, ISSN 0021-9606
- Laidler, K.J. (1965). *Chemical Kinetics*, McGraw-Hill, New York, USA
- Mason, D.M. (1965). Effect of Composition and Pressure on Gas Phase Reaction Rate Coefficient. *Chemical Engineering Science*, Vol.20, No.12, pp. 1143-1145, ISSN 0009-2509
- Novák, J.; Malijevský, A.; Voňka, P. & Matouš, J. (1999). *Physical Chemistry*, VŠCHT, ISBN 80-7080-360-6, Prague, Czech Republic (in Czech)
- Pekař, M. & Koubek, J. (1997). Rate-limiting Step. Does It Exist in the Non-Steady State? *Chemical Engineering Science*, Vol.52, No.14, pp. 2291-2297, ISSN 0009-2509
- Pekař, M. & Koubek, J. (1999). Concentration Forcing in the Kinetic Research in Heterogeneous Catalysis. *Applied Catalysis A*, Vol.177, No.1, pp. 69-77, ISSN 0926-860X
- Pekař, M. & Koubek, J. (2000). On the General Principles of Transient Behaviour of Heterogeneous Catalytic Reactions. *Applied Catalysis A*, Vol.199, No.2, pp. 221-226, ISSN 0926-860X
- Pekař, M. (2007). Detailed Balance in Reaction Kinetics – Consequence of Mass Conservation? *Reaction Kinetics and Catalysis Letters*, Vol. 90, No. 2, p. 323-329, ISSN 0133-1736
- Pekař, M. (2009). Thermodynamic Framework for Design of Reaction Rate Equations and Schemes. *Collection of the Czechoslovak Chemical Communications*, Vol.74, No.9, pp. 1375-1401, ISSN 0010-0765
- Pekař, M. (2010). Macroscopic Derivation of the Kinetic Mass-Action Law. *Reaction Kinetics, Mechanisms and Catalysis*, Vol.99, No. 1, pp. 29-35, ISSN 1878-5190
- Qian, H. & Beard, D.A. (2005). Thermodynamics of Stoichiometric Biochemical Networks in Living Systems Far From Equilibrium. *Biophysical Chemistry*, Vol.114, No.3, pp. 213-220, ISSN 0301-4622
- Samohýl, I. (1982). *Rational Thermodynamics of Chemically Reacting Mixtures*, Academia, Prague, Czechoslovakia (in Czech)
- Samohýl, I. (1987). *Thermodynamics of Irreversible Processes in Fluid Mixtures*, Teubner, Leipzig, Germany
- Samohýl, I. & Malijevský, A. (1976). Phenomenological Derivation of the Mass Action Law of homogeneous chemical kinetics. *Collection of the Czechoslovak Chemical Communications*, Vol.41, No.8, pp. 2131-2142, ISSN 0010-0765
- Silbey, R.J.; Alberty, R.A. & Bawendi M.G. (2005). *Physical Chemistry*, 4<sup>th</sup> edition, J.Wiley, ISBN 0-471-21504-X, Hoboken, USA
- Vlad, M.O. & Ross, J. (2009). Thermodynamically Based Constraints for Rate Coefficients of Large Biochemical Networks. *WIREs Systems Biology and Medicine*, Vol.1, No.3, pp. 348-358, ISSN 1939-5094
- Wegscheider, R. (1902). Über simultane Gleichgewichte und die Beziehungen zwischen Thermodynamik und Reaktionskinetik. *Zeitschrift für physikalische Chemie*, Vol. XXXIX, pp. 257-303

# The Thermodynamics *in* Planck's Law

Constantinos Ragazas<sup>1</sup>  
*The Lawrenceville School*  
 USA

## 1. Introduction

Quantum Physics has its historical beginnings with Planck's derivation of his formula for blackbody radiation, more than one hundred years ago. In his derivation, Planck used what latter became known as *energy quanta*. In spite of the best efforts at the time and for decades later, a more *continuous approach* to derive this formula had not been found. Along with Einstein's *Photon Hypothesis*, the *Quantization of Energy Hypothesis* thus became the foundations for much of the Physics that followed. This *physical view* has shaped our understanding of the Universe and has resulted in mathematical certainties that are counter-intuitive and contrary to our experience.

Physics provides *mathematical models* that seek to describe *what is* the Universe. We believe mathematical models of *what is* -- as with past metaphysical attempts -- are a never ending search getting us deeper and deeper into the 'rabbit's hole' [Frank 2010]. We show in this Chapter that a *quantum-view* of the Universe is not necessary. We argue that *a world without quanta* is not only possible, but desirable. We do not argue, however, with the mathematical formalism of Physics -- just the *physical view* attached to this.

We will present in this Chapter a mathematical derivation of *Planck's Law* that uses simple continuous processes, without needing *energy quanta* and *discrete statistics*. This *Law* is not true by Nature, but by Math. In our view, *Planck's Law* becomes a *Rosetta Stone* that enables us to translate known physics into simple and sensible formulations. To this end the quantity *eta* we introduce is fundamental. This is the time integral of energy that is used in our mathematical derivation of *Planck's Law*. In terms of this *prime physis* quantity *eta* (acronym for *energy-time-action*), we are able to define such physical quantities as energy, force, momentum, temperature and entropy. Planck's constant *h* (in units of energy-time) is such a quantity *eta*. Whereas currently *h* is thought as *action*, in our derivation of *Planck's Law* it is more naturally viewed as *accumulation of energy*. And while *h* is a constant, the quantity *eta* that appears in our formulation is a variable. Starting with *eta*, Basic Law can be mathematically derived and not be physically posited.

Is the Universe *continuous* or *discrete*? In my humble opinion this is a false dichotomy. It presents us with an impossible choice between two absolute views. And as it is always the case, making one side *absolute* leads to endless fabrications denying the opposite side. The Universe is neither *continuous* nor *discrete* because the Universe is both *continuous* and *discrete*. Our *view* of the Universe is *not* the Universe. The Universe simply *is*. In The Interaction of

---

<sup>1</sup> cragaza@lawrenceville.org

Measurement [Ragazas, 2010h] we argue with mathematical certainty that we cannot know through direct measurements what a physical quantity  $E(t)$  is as a function of time.

Since we are limited by our measurements of '*what is*', we should consider these as the beginning and end of our knowledge of '*what is*'. Everything else is just '*theory*'. There is nothing real about theory! As the ancient Greeks knew and as the very word '*theory*' implies. In Planck's Law is an Exact Mathematical Identity [Ragazas 2010f] we show *Planck's Law* is a mathematical truism that describes the *interaction of measurement*. We show that *Planck's Formula* can be *continuously* derived. But also we are able to explain *discrete* 'energy quanta'. In our view, *energy propagates continuously but interacts discretely*. Before there is *discrete manifestation* we argue there is *continuous accumulation* of energy. And this is based on the *interaction of measurement*.

Mathematics is a tool. It is a language of objective reasoning. But mathematical 'truths' are always 'conditional'. They depend on our presuppositions and our premises. They also depend, in my opinion, on the mental images we use to think. We phrase our explanations the same as we frame our experiments. In the single electron emission double-slit experiment, for example, it is assumed that the electron emitted at the source is the same electron detected at the screen. Our explanation of this experiment considers that these two electrons may be separate events. Not directly connected by some trajectory from *source* to *sensor*. [Ragazas 2010j]

We can have beautiful mathematics based on *any* view of the Universe we have. Consider the Ptolemy with their epicycles! But if the view leads to physical explanations which are counter-intuitive and defy common sense, or become too abstract and too removed from life and so not support life, than we must not confuse mathematical deductions with *physical realism*. Rather, we should change our view! And just as we can write bad literature using good English, we can also write bad physics using good math. In either case we do not fault the language for the story. We can't fault Math for the failings of Physics.

The failure of Modern Physics, in my humble opinion, is in not providing us with *physical explanations* that make sense; a *physical view* that is consistent with our experiences. A *view* that will not put us at odds with ourselves, with our understanding of our world and our lives. Math may not be adequate. Sense may be a better guide.

## 2. Mathematical results

We list below the main mathematical derivations that are the basis for the results in physics in this Chapter. The proofs can be found in the Appendix at the end. These mathematical results, of course, do not depend on Physics and are not limited to Physics. In *Stocks and Planck's Law* [Ragazas 2010l] we show how the same 'Planck-like' formula we derive here also describes a simple comparison model for stocks.

*Notation.*  $E(t)$  is a real-valued function of the real-variable  $t$

$\tau = \Delta t = t - s$  is an interval of  $t$

$\Delta E = E(t) - E(s)$  is the change of  $E$

$\eta = P = \int_s^t E(u) du$  is the accumulation of  $E$

$\bar{E} = E_{av} = \frac{1}{t-s} \int_s^t E(u) du$  is the average of  $E$

$$\mathcal{J} = \mathcal{J}_\eta = \left(\frac{1}{\kappa}\right) \frac{\eta}{\tau} \text{ where } \kappa \text{ is a scalar constant}$$

$D_x$  indicates differentiation with respect to  $x$   
 $r, \nu$  are constants, often a rate of growth or frequency

Characterization 1:  $E(t) = E_0 e^{rt}$  if and only if  $\Delta E = Pr$

Characterization 2:  $E(t) = E_0 e^{rt}$  if and only if  $\frac{Pr}{e^{r(t-s)} - 1} = E(s)$

Characterization 2a:  $E(t) = E_0 e^{rt}$  if and only if  $\frac{Pr}{e^{Pr/E_{av}} - 1} = E(s)$

Characterization 3:  $E(t) = E_0 e^{rt}$  if and only if  $\frac{\Delta E}{e^{\Delta E/E_{av}} - 1} = E(s)$

Characterization 4:  $E(t) = E_0 e^{rt}$  if and only if  $\frac{\Delta E}{E_{av}} = r\Delta t$

Theorem 1a:  $E(t) = E_0 e^{rt}$  if and only if  $\frac{Pr}{e^{Pr/E_{av}} - 1}$  is invariant with  $t$

Theorem 2: For any integrable function  $E(t)$ ,  $\lim_{t \rightarrow s} \frac{Pr}{e^{r\Delta t} - 1} = E(s)$

**2.1 'Planck-like' characterizations [Ragazas 2010a]**

Note that  $E_{av} = \kappa \mathcal{J}_\eta$ . We can re-write Characterization 2a above as,

$$E(t) = E_0 e^{\nu t} \text{ if and only if } E_0 = \frac{\eta \nu}{e^{\eta \nu / \kappa \mathcal{J}} - 1} \tag{1}$$

Planck's Law for blackbody radiation states that,  $E_0 = \frac{h\nu}{e^{h\nu/kT} - 1}$  (2)

where  $E_0$  is the intensity of radiation,  $\nu$  is the frequency of radiation and  $T$  is the (Kelvin) temperature of the blackbody, while  $h$  is Planck's constant and  $k$  is Boltzmann's constant. [Planck 1901, Eqn 11]. Clearly (1) and (2) have the exact same mathematical form, including the type of quantities that appear in each of these equations. We state the main results of this section as,

Result I: A 'Planck-like' characterization of simple exponential functions

$$E(t) = E_0 e^{\nu t} \text{ if and only if } E_0 = \frac{\eta \nu}{e^{\eta \nu / \kappa \mathcal{J}} - 1}$$

Using Theorem 2 above we can drop the condition that  $E(t) = E_0 e^{\nu t}$  and get,

Result II: A 'Planck-like' limit of any integrable function

$$\text{For any integrable function } E(t), E_0 = \lim_{t \rightarrow 0} \frac{\eta \nu}{e^{\eta \nu / \kappa \mathcal{J}} - 1}$$

We list below for reference some helpful variations of these mathematical results that will be used in this Chapter.

$$E_0 = \frac{\Delta E}{e^{\Delta E/E_{av}} - 1} = \frac{\eta V}{e^{\eta V/\kappa \mathcal{J}_\eta} - 1} \quad (\text{if } E(t) = E_0 e^{\nu t}) \quad (3)$$

$$E_0 \approx \frac{\Delta E}{e^{\Delta E/E_{av}} - 1} \approx \frac{\eta V}{e^{\eta V/\kappa \mathcal{J}_\eta} - 1} \quad (\text{if } E(t) \text{ is integrable}) \quad (4)$$

$$E_0 = \frac{\eta V}{e^{\eta V/\kappa \mathcal{J}_\eta} - 1} \text{ is exact if and only if } \frac{\eta V}{e^{\eta V/\kappa \mathcal{J}_\eta} - 1} \text{ is independent of } \eta \quad (5)$$

Note that in order to avoid using limit approximations in (4) above, by (3) we will assume an *exponential of energy* throughout this Chapter. This will allow us to explore the underlying ideas more freely and simply. Furthermore in **Section 10.0** of this Chapter, we will be able to justify such an exponential time-dependent local representation of energy [Ragazas 2010i]. Otherwise, all our results (with the exception of **Section 8.0**) can be thought as pertaining to a blackbody with perfect emission, absorption and transmission of energy.

### 3. Derivation of Planck's law without *energy quanta* [Ragazas 2010f]

*Planck's Formula* as originally derived describes what physically happens at the *source*. We consider instead what happens at the *sensor* making the measurement. Or, equivalently, what happens at the *site of interaction* where energy exchanges take place. We assume we have a blackbody medium, with perfect emission, absorption and transmission of energy. We consider that measurement involves an *interaction* between the *source* and the *sensor* that results in energy exchange. This interaction can be mathematically described as a functional relationship between  $E(s)$ , the energy locally at the *sensor* at time  $s$ ;  $\Delta E$ , the energy absorbed by the *sensor* making the measurement; and  $\bar{E}$ , the average energy at the *sensor* during measurement. Note that *Planck's Formula* (2) has the exact same mathematical form as the mathematical equivalence (3) and as the limit (4) above. By letting  $E(s)$  be an *exponential*, however, from (3) we get an *exact* formula, rather than the limit (4) if we assume that  $E(s)$  is only an integrable function. The argument below is one of several that can be made. The *Assumptions* we will use in this very simple and elegant derivation of *Planck's Formula* will themselves be justified in later **Sections 5.0, 6.0 and 10.0** of this Chapter.

*Mathematical Identity.* For any integrable function  $E(t)$ ,  $\eta = \int_s^{s+\eta/E_{av}} E(u) du$  (6)

*Proof:* (see Fig. 1)

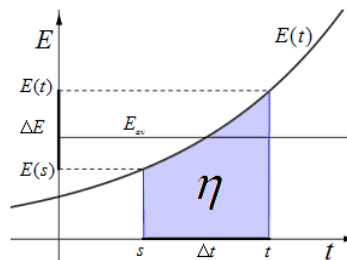


Fig. 1.

Assumptions: 1) Energy locally at the sensor at  $t = s$  can be represented by  $E(s) = E_0 e^{\nu s}$ , where  $E_0$  is the intensity of radiation and  $\nu$  is the frequency of radiation. 2) When measurement is made, the source and the sensor are in equilibrium. The average energy of the source is equal to the average energy at the sensor. Thus,  $\bar{E} = kT$ . 3) Planck's constant  $h$  is the minimal 'accumulation of energy' at the sensor that can be manifested or measured. Thus we have  $\eta = h$ .

Using the above Mathematical Identity (6) and Assumptions we have Planck's Formula,

$$h = \int_0^{\frac{h}{kT}} E_0 e^{\nu u} du = \frac{E_0}{\nu} \left[ e^{h\nu/kT} - 1 \right] \text{ and so, } E_0 = \frac{h\nu}{e^{h\nu/kT} - 1}$$

Planck's Formula is a mathematical truism that describes the interaction of energy. That is to say, it gives a mathematical relationship between the energy locally at the sensor, the energy absorbed by the sensor, and the average energy at the sensor during measurement. Note further that when an amount of energy  $\Delta E$  is absorbed by the sensor,  $E(t)$  resets to  $E_0$ .

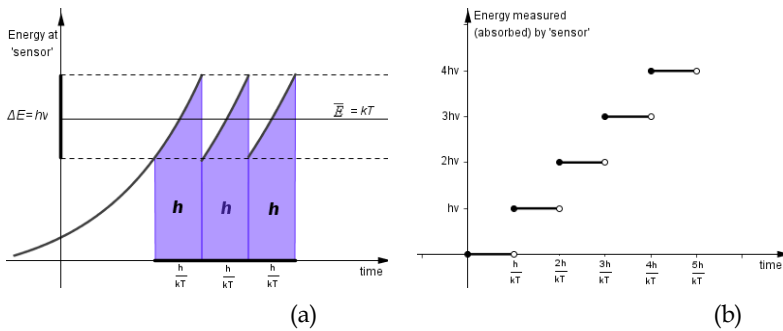


Fig. 2.

Note: Our derivation, showing that Planck's Law is a mathematical truism, can now clearly explain why the experimental blackbody spectrum is so indistinguishable from the theoretical curve. ([http://en.wikipedia.org/wiki/File:Firas\\_spectrum.jpg](http://en.wikipedia.org/wiki/File:Firas_spectrum.jpg))

Conclusions:

1. Planck's Formula is an exact mathematical truism that describes the interaction of energy.
2. Energy propagates continuously but interacts discretely. The absorption or measurement of energy is made in discrete 'equal size sips'(energy quanta).
3. Before manifestation of energy (when an amount  $\Delta E$  is absorbed or emitted) there is an accumulation of energy that occurs over a duration of time  $\Delta t$ .
4. The absorption of energy is proportional to frequency,  $\Delta E = h\nu$  (The Quantization of Energy Hypothesis).
5. There exists a time-dependent local representation of energy,  $E(t) = E_0 e^{\nu t}$ , where  $E_0$  is the intensity of radiation and  $\nu$  is the frequency of radiation. [Ragazas 2011a]
6. The energy measured  $\Delta E$  vs.  $\Delta t$  is linear with slope  $\nu kT$  for constant temperature  $T$ .
7. The time  $\Delta t$  required for an accumulation of energy  $h$  to occur at temperature  $T$  is given

$$\text{by } \Delta t = \frac{h}{kT}.$$

#### 4. Prime physis eta and the derivation of Basic Law [Ragazas 2010d]

In our derivation of *Planck's Formula* the quantity  $\eta$  played a prominent role. In this derivation  $\eta$  is the *time-integral of energy*. We consider this quantity  $\eta$  as *prime physis*, and define in terms of it other physical quantities. And thus mathematically derive Basic Law. Planck's constant  $h$  is such a quantity  $\eta$ , measured in units of *energy-time*. But whereas  $h$  is a constant,  $\eta$  is a variable in our formulation.

*Definitions: For fixed  $(\vec{x}_0, t_0)$  and along the x-axis for simplicity,*

*Prime physis:  $\eta = eta$  (energy-time-action)*

$$\text{Energy: } E = \frac{\partial \eta}{\partial t} \quad (7)$$

$$\text{Momentum: } p_x = \frac{\partial \eta}{\partial x} \quad (8)$$

$$\text{Force: } F_x = \frac{\partial^2 \eta}{\partial x \partial t} \quad (9)$$

Note that the quantity *eta* is undefined. But it can be thought as 'energy-time-action' in units of *energy-time*. *Eta* is both *action* as well as *accumulation of energy*. We make only the following assumption about  $\eta$ .

*Identity of Eta Principle: For the same physical process, the quantity  $\eta$  is one and the same.*

Note: This *Principle* is somewhat analogous to a physical system being described by the *wave function*. Hayrani Öz has also used originally and consequentially similar ideas in [Öz 2002, 2005, 2008, 2010].

##### 4.1 Mathematical derivation of Basic Law

Using the above definitions, and known mathematical theorems, we are able to derive the following Basic Law of Physics:

- Planck's Law,  $E_0 = \frac{h\nu}{e^{h\nu/kT} - 1}$ , is a mathematical truism (**Section 3.0**)
- The Quantization of Energy Hypothesis,  $\Delta E = nh\nu$  (**Section 3.0**)
- *Conservation of Energy and Momentum.* The gradient of  $\eta(\vec{x}, t)$  is  $\vec{\nabla} \eta = \left\langle \frac{\partial \eta}{\partial x}, \frac{\partial \eta}{\partial t} \right\rangle = \langle p_x, E \rangle$ . Since all gradient vector fields are *conservative*, we have the *Conservation of Energy and Momentum.*
- *Newton's Second law of Motion.* The second Law of motion states that  $F = ma$ . From definition (9) above we have,

$$F = \frac{\partial^2 \eta}{\partial x \partial t} = \frac{\partial^2 \eta}{\partial t \partial x} = \frac{\partial p_x}{\partial t} = \frac{\partial}{\partial t}(mv) = ma, \text{ since } p_x = mv.$$



- *Energy-momentum Equivalence.* From the definition of energy  $E = \frac{\partial \eta}{\partial t}$  and of momentum

$$p_x = \frac{\partial \eta}{\partial x} \text{ we have that, } \eta = \int_{t_0}^t E(u) du \text{ and } \eta = \int_{x_0}^x p_x(u) du .$$

Using the *Identity of Eta Principle*, the quantity  $\eta$  in these is one and the same.

Therefore,  $\int_{t_0}^t E(u) du = \int_{x_0}^x p_x(u) du$  . Differentiating with respect to  $t$ , we obtain,

$$E(t) = p_x(x) \cdot \frac{dx}{dt} \text{ or more simply, } E = p_x v \text{ (energy-momentum equivalence)}$$

- *Schroedinger Equation:* Once the extraneous constants are striped from Schroedinger's equation, this in essence can be written as  $\frac{\partial \psi}{\partial t} = H\psi$  , where  $\psi$  is the *wave function* ,  $H$  is the *energy operator*, and  $H\psi$  is the energy at any  $(\vec{x}, t)$  . The definition (7) of energy  $\frac{\partial \eta}{\partial t} = E$  given above is for a *fixed*  $(\vec{x}_0, t_0)$  . Comparing these we see that whereas our definition of energy is for *fixed*  $(\vec{x}_0, t_0)$  , Schroedinger equation is for *any*  $(\vec{x}, t)$  . But otherwise the two equations have the same form and so express the same underlying idea. Now (7) *defines* energy in terms of the more primary quantity  $\eta$  (which can be viewed as *accumulation of energy* or *action*) and so we can view Schroedinger Equation as in essence *defining* the energy of the system at any  $(\vec{x}, t)$  while the *wave function*  $\psi$  can be understood to express the *accumulation of energy* at any  $(\vec{x}, t)$  . This suggests that the *wave function*  $\psi$  is the same as the quantity  $\eta$  . We have the following interesting interpretation of the wave function.

- The wave function gives the distribution of the accumulation of energy of the system.
- *Uncertainty Principle:* Since  $\Delta E = \eta v$  , for  $\Delta t > \frac{1}{v}$  (a 'wavelength') we have

$$\Delta E \cdot \Delta t > \eta v \cdot \frac{1}{v} = \eta > h . \text{ Or equivalently, for } \frac{\Delta E}{E_{av}} > 1 , \text{ we again have } \Delta E \cdot \Delta t > \eta > h , \text{ since}$$

$h$  is the *minimal eta* that can be manifested. Note that since  $\frac{\Delta E}{E_{av}} = v \Delta t$  (*Characteristic 5*), we

have  $\frac{\Delta E}{E_{av}} > 1$  if and only if  $\Delta t > \frac{1}{v}$  . Since  $E_{av} = kT$  and entropy is defined as  $\Delta S = \frac{\Delta E}{T}$  ,

we have that

- $\Delta E \cdot \Delta t > h$  if and only if  $\Delta S > k$  .
- *Planck's Law and Boltzmann's Entropy Equation Equivalence:*

Starting with our *Planck's Law* formulation,  $E_0 = \frac{\Delta E}{e^{\Delta E/E_{av}} - 1}$  in (3) above and re-writing this

equivalently we have,  $e^{\Delta E/E_{av}} = 1 + \frac{\Delta E}{E_0} = \frac{E}{E_0}$  and so,  $\frac{\Delta E}{E_{av}} = \ln \left( \frac{E}{E_0} \right)$  . Using the definition of

thermodynamic entropy we get  $\Delta S_{\Theta} = \frac{\Delta E}{T} = k \cdot \frac{\Delta E}{E_{av}} = k \cdot \ln\left(\frac{E}{E_0}\right)$ . If  $\Omega(t)$  represents the number of microstates of the system at time  $t$ , then  $E(t) = A\Omega(t)$ , for some constant  $A$ . Thus, we get *Boltzmann's Entropy Equation*,  $S_{\Theta} = k \ln \Omega$ .

Conversely, starting with *Boltzmann's Entropy Equation*,  $\Delta S_{\Theta} = k \ln\left(\frac{\Omega}{\Omega_0}\right) = k \ln\left(\frac{E}{E_0}\right)$ .

Since  $\Delta S_{\Theta} = \frac{\Delta E}{T}$  we can rewrite this equivalently as  $\frac{\Delta E}{E_{av}} = \ln\left(\frac{E}{E_0}\right)$  and so

$e^{\Delta E/E_{av}} = 1 + \frac{\Delta E}{E_0} = \frac{E}{E_0}$ . From this we have, *Planck's Law*,  $E_0 = \frac{\Delta E}{e^{\Delta E/E_{av}} - 1}$  in (3) above.

- *Entropy-Time Relationship*:  $\Delta S = k\nu\Delta t$  where  $\nu$  is the *rate of evolution* of the system and  $\Delta t$  is the *time duration* of evolution, since  $E_{av} = \frac{\eta}{\Delta t}$  and  $\Delta E = \eta\nu$ .

- *The Fundamental Thermodynamic Relation*: It is a well known fact that the internal energy  $U$ , entropy  $S$ , temperature  $T$ , pressure  $P$  and volume  $V$  of a system are related by the equation  $dU = TdS - PdV$ . By using increments rather than differentials, and using the fact that work performed by the system is given by  $W = \int PdV$  this can be re-written as

$\Delta S = \frac{\Delta U}{T} + \frac{\Delta W}{T}$ . All the terms in this equation are various entropy quantities. The

fundamental thermodynamic relation can be interpreted thus as saying, “the total change of entropy of a system equals the sum of the change in the internal (unmanifested) plus the change in the external (manifested) entropy of the system”. Considering the *entropy-time relationship* above, this can be rephrased more intuitively as saying “the total lapsed time for a physical process equals the time for the 'accumulation of energy' plus the time for the 'manifestation of energy' for the process”. This relationship along with *The Second Law of Thermodynamics* establish a *duration of time* over which there is *accumulation of energy before manifestation of energy* – one of our main results in this Chapter and a premise to our explanation of the double-slit experiment. [Ragazas 2010g]

## 5. The temperature of radiation [Ragazas 2010g]

Consider the energy  $E(t)$  at a fixed point at time  $t$ . We define the *temperature of radiation* to

be given by  $\mathcal{T} = \mathcal{T}_{\eta} = \left(\frac{1}{\kappa}\right)\frac{\eta}{\tau}$  where  $\kappa$  is a scalar constant. Though in defining *temperature*

this way the *accumulation of energy*  $\eta$  can be any value, when considering a *temperature scale*  $\eta$  is fixed and used as a *standard for measurement*. To distinguish *temperature* and *temperature scale* we will use  $\mathcal{T}$  and  $\mathcal{T}_{\eta}$  respectively. We assume that *temperature* is characterized by the following property:

*Characterization of temperature*: For a fixed  $\eta$ , the temperature is inversely proportional to the duration of time for an accumulation of energy  $\eta$  to occur.

Thus if *temperature* is twice as high, the accumulation of energy will be twice as fast, and visa-versa. This *characterization of temperature* agrees well with our physical sense of temperature. It is also in agreement with *temperature* as being the average kinetic energy of the motion of molecules.

For fixed  $\eta$ , we can define  $\mathcal{T}_\eta = \left(\frac{1}{\kappa}\right)\frac{\eta}{\tau}$ , which will be unique up to an arbitrary scalar

constant  $\kappa$ . Conversely, for a given  $\mathcal{T}$  as characterized above, we will have  $\mathcal{T} = \rho \cdot \frac{1}{\tau}$ , where

$\rho$  is a proportionality constant. By setting  $\rho = \frac{\eta}{\kappa}$  we get  $\mathcal{T} = \mathcal{T}_\eta = \left(\frac{1}{\kappa}\right)\frac{\eta}{\tau}$ . We have the following *temperature-eta* correspondence:

*Temperature-eta Correspondence: Given  $\eta$ , we have  $\mathcal{T}_\eta = \left(\frac{1}{\kappa}\right)\frac{\eta}{\tau}$ , where  $\kappa$  is some arbitrary scalar*

*constant. Conversely, given  $\mathcal{T}$  we have  $\mathcal{T} = \mathcal{T}_\eta = \left(\frac{1}{\kappa}\right)\frac{\eta}{\tau}$ , for some fixed  $\eta$  and arbitrary scalar constant  $\kappa$ . Any temperature scale, therefore, will have some fixed  $\eta$  and arbitrary scalar constant  $\kappa$  associated with it.*

## 6. The meaning and existence of Planck's constant $h$ [Ragazas 2010c]

Planck's constant  $h$  is a fundamental universal constant of Physics. And although we can experimentally determine its value to great precision, the reason for its existence and what it really means is still a mystery. Quantum Mechanics has adapted it in its mathematical formalism. But QM does not explain the meaning of  $h$  or prove why it must exist. Why does the Universe need  $h$  and *energy quanta*? Why does the mathematical formalism of QM so accurately reflect physical phenomena and predict these with great precision? Ask any physicists and uniformly the answer is "that's how the Universe works". The units of  $h$  are in *energy-time* and the conventional interpretation of  $h$  is as a *quantum of action*. We interpret  $h$  as the *minimal accumulation of energy* that can be manifested. Certainly the units of  $h$  agree with such interpretation. Based on our results above we provide an explanation for the existence of Planck's constant -- what it means and how it comes about. We show that the existence of *Planck's constant* is not necessary for the Universe to exist but rather  $h$  exists by Mathematical necessity and inner consistency of our system of measurements.

Using *eta* we defined in **Section 5.0** above the *temperature of radiation* as being proportional to the ratio of *eta/time*. To obtain a *temperature scale*, however, we need to fix *eta* as a standard for measurement. We show below that the fixed *eta* that determines the Kelvin *temperature scale* is Planck's constant  $h$ .

In The Interaction of Measurement [Ragazas 2010h] we argue that direct measurement of a physical quantity  $E(t)$  involves a physical interaction between the *source* and the *sensor*. For measurement to occur an interval of time  $\Delta t$  must have lapsed and an incremental amount  $\Delta E$  of the quantity will be absorbed by the *sensor*. This happens when there is an *equilibrium* between the *source* and the *sensor*. At *equilibrium*, the 'average quantity  $E_{av}$  from the source' will equal to the 'average quantity  $E_{av}$  at the sensor'. *Nothing in our observable World can exist without time, when the entity 'is' in equilibrium with its environment and its 'presence' can be observed and measured.* Furthermore as we showed above in **Section 3.0** the *interaction of measurement* is described by *Planck's Formula*.

From the mathematical equivalence (5) above we see that  $\eta$  can be *any* value and  $\frac{\eta\nu}{e^{\eta\nu/\kappa\mathcal{T}_\eta} - 1}$  will be invariant and will continue to equal to  $E_0$ . We can in essence (Fig. 3) 'reduce' the formula  $E_0 = \frac{\eta\nu}{e^{\eta\nu/\kappa\mathcal{T}_\eta} - 1}$  by reducing the value of  $\eta$  and so the value of  $E_{av} = \kappa\mathcal{T}_\eta$  will correspondingly adjust, and visa versa. Thus we see that  $\eta$  and  $\mathcal{T}_\eta$  go *hand-in-hand* to maintain  $E_0 = \frac{\eta\nu}{e^{\eta\nu/\kappa\mathcal{T}_\eta} - 1}$  invariant. And though the mathematical equivalence (5) above allows these values to be anything, the calibrations of these quantities in Physics require their value to be specific. Thus, for  $\eta = h$  (Planck's constant) and  $\kappa = k$  (Boltzmann's constant), we get  $\mathcal{T}_\eta = T$  (Kelvin temperature) (see Fig. 3). Or, conversely, if we start with  $\mathcal{T}_\eta = T$  and set the arbitrary constant  $\kappa = k$ , then this will force  $\eta = h$ . Thus we see that *Planck's constant h, Boltzmann's constant k, and Kelvin temperature T* are so defined and calibrated to fit Planck's Formula. Simply stated, when  $\eta = h, \mathcal{T}_h = T$ .

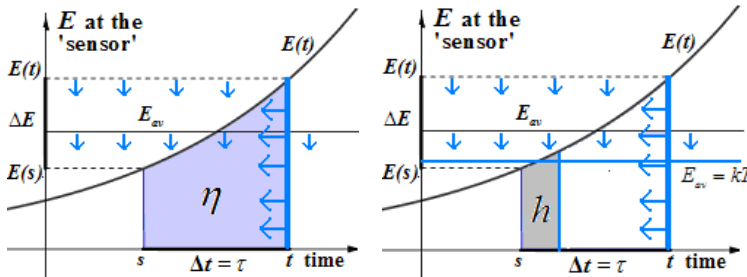


Fig. 3.  $E_0 = \frac{\Delta E}{\frac{\Delta E}{E_{av}} - 1} = \frac{\eta\nu}{e^{\eta\nu/\kappa\mathcal{T}_\eta} - 1}$ ,  $\Delta E = \eta\nu$ ,  $E_{av} = \kappa\mathcal{T}_\eta$ ,  $\mathcal{T}_\eta = \left(\frac{1}{\kappa}\right)\frac{\eta}{\tau}$ ,  $E(t) = E_0 e^{\nu t}$

*Conclusion: Physical theory provides a conceptual lens through which we 'see' the world. And based on this theoretical framework we get a measurement methodology. Planck's constant h is just that 'theoretical focal point' beyond which we cannot 'see' the world through our theoretical lens. Planck's constant h is the minimal eta that can be 'seen' in our measurements. Kelvin temperature scale requires the measurement standard eta to be h.*

*Planck's Formula is a mathematical identity that describes the interaction of measurement. It is invariant with time, accumulation of energy or amount of energy absorbed. Planck's constant exists because of the time-invariance of this mathematical identity. The calibration of Boltzmann's constant k and Kelvin temperature T, with kT being the average energy, determine the specific value of Planck's constant h.*

### 7. Entropy and the second law of thermodynamics [Ragazas 2010b]

The quantity  $\frac{\Delta E}{E_{av}}$  that appears in our *Planck's Law* formulation (3) is 'additive over time'. This

is so because under the assumption that *Planck's Formula* is exact we have that  $\frac{\Delta E}{E_{av}} = \nu\Delta t$ , by

*Characterization 4.* Interestingly, this quantity is essentially *thermodynamic entropy*, since  $E_{av} = kT$ , and so  $\Delta S = \frac{\Delta E}{T} = kv\Delta t$ . Thus entropy is *additive over time*. Since  $v$  can be thought as the *evolution rate* of the system (both positive or negative), entropy is a measure of the *amount of evolution* of the system over a duration of time  $\Delta t$ . Such connection between *entropy as amount of evolution* and *time* makes eminent intuitive sense, since *time* is generally thought in terms of *change*. But, of course, this is *physical time* and not some mathematical abstract parameter as in *spacetime continuum*.

Note that in the above, *entropy* can be both positive or negative depending on the *evolution rate*  $v$ . That the *duration of time*  $\Delta t$  is positive, we argue, is postulated by *The Second Law of Thermodynamics*. It is amazing that the most fundamental of all physical quantities *time* has no fundamental Basic Law pertaining to its nature. We argue *the Basic Law pertaining to time is The Second Law of Thermodynamics*. Thus, a more revealing rewording of this Law should state that *all physical processes take some positive duration of time to occur*. Nothing happens *instantaneously*. Physical time is really *duration*  $\Delta t$  (or  $dt$ ) and not *instantiation*  $t = s$ .

## 8. The photoelectric effect without photons [Ragazas 2010k]

Photoelectric emission has typically been characterized by the following experimental facts (*some of which can be disputed, as noted*):

1. For a given metal surface and frequency of incident radiation, the rate at which photoelectrons are emitted (the photoelectric current) is directly proportional to the intensity of the incident light.
2. The energy of the emitted photoelectron is independent of the intensity of the incident light but depends on the frequency of the incident light.
3. For a given metal, there exists a certain minimum frequency of incident radiation below which no photoelectrons are emitted. This frequency is called the threshold frequency. (*see below*)
4. The time lag between the incidence of radiation and the emission of photoelectrons is very small, less than  $10^{-9}$  second.

*Explanation of the Photoelectric Effect without the Photon Hypothesis:* Let  $v$  be the rate of radiation of an incident light on a metal surface and let  $\alpha$  be the rate of absorption of this radiation by the metal surface. The combined rate locally at the surface will then be  $v - \alpha$ . The radiation energy at a point on the surface can be represented by  $E(t) = E_0 e^{(v-\alpha)t}$ , where  $E_0$  is the intensity of radiation of the incident light. If we let  $\eta$  be the *accumulation of energy* locally at the surface over a time pulse  $\tau$ , then by *Characterization 1* we'll have that  $\Delta E = \eta(v - \alpha)$ . If we let Planck's constant  $h$  be the *accumulation of energy* for an electron, the number of electrons  $n_e$  over the pulse of time  $\tau$  will then be  $n_e = \frac{\eta}{h}$  and the energy of an electron  $\Delta E_e$  will be given by

$$\Delta E_e = \frac{\Delta E}{n_e} = h(v - \alpha) \quad (10)$$

Since  $\eta = \int_0^\tau E_0 e^{(v-\alpha)u} du = E_0 \left[ \frac{e^{(v-\alpha)\tau} - 1}{(v - \alpha)} \right]$ , we get the *photoelectric current*  $I$ ,

$$I = \frac{n_e}{\tau} = \frac{\eta}{h\tau} = E_0 \left[ \frac{e^{(\nu-\alpha)\tau} - 1}{h(\nu-\alpha)\tau} \right] \quad (11)$$

The absorption rate  $\alpha$  is a characteristic of the metal surface, while the pulse of time  $\tau$  is assumed to be constant for fixed experimental conditions. The quantity  $\left[ \frac{e^{(\nu-\alpha)\tau} - 1}{h(\nu-\alpha)\tau} \right]$  in equation (10) would then be *constant*.

Combining the above and using (10) and (11) we have *The Photoelectric Effect*:

1. For incident light of fixed frequency  $\nu$  and fixed metal surface, the photoelectric current  $I$  is proportional to the intensity  $E_0$  of the incident light. (by (11) above)
2. The energy  $\Delta E_e$  of a photoelectron depends only on the frequency  $\nu$  and not on the intensity  $E_0$  of the incident light. It is given by the equation  $\Delta E_e = h(\nu - \alpha)$  where  $h$  is Planck's constant and the absorption rate  $\alpha$  is a property of the metal surface. (by (10) above)
3. If  $\Delta E_e$  is taken to be the kinetic energy of a photoelectron, then for incident light with frequency  $\nu$  less than the 'threshold frequency'  $\alpha$  the kinetic energy of a photoelectron would be negative and so there will be no photoelectric current. (by (10) above) (*see Note below*)
4. The photoelectric current is almost instantaneous ( $< 10^{-9}$  sec.), since for a single photoelectron we have that  $\Delta t = \frac{h}{kT} < 10^{-9}$  sec. by *Conclusion 7 Section 3*.

*Note:* Many experiments since the classic 1916 experiments of Millikan have shown that there is photoelectric current even for frequencies below the threshold, contrary to the explanation by Einstein. In fact, the original experimental data of Millikan show an asymptotic behavior of the (photocurrent) vs (voltage) curves along the energy axis with no clear 'threshold frequency'. The photoelectric equations (10) and (11) we derived above agree with these experimental anomalies, however.

In an article Richard Keesing of York University, UK, states,

*I noticed that a reverse photo-current existed ... and try as I might I could not get rid of it. My first disquieting observation with the new tube was that the I/V curves had high energy tails on them and always approached the voltage axis asymptotically. I had been brought up to believe that the current would show a well defined cut off, however my curves just refused to do so. Several years later I was demonstrating in our first year lab here and found that the apparatus we had for measuring Planck's constant had similar problems. After considerable soul searching it suddenly occurred on me that there was something wrong with the theory of the photoelectric effect ... [Keesing 2001]*

In the same article, taking the original experimental data from the 1916 experiments by Millikan, Prof. Keesing plots the graphs in Fig. 4.

In what follows, we analyze the asymptotic behavior of equation (11) by using a function of the same form as (11).

$$f(x) = \frac{A(e^{b(x-c)} - 1)}{x - c} + d \quad (12)$$

*Note:* We use  $d$  since some graphs typically are shifted up a little for clarity.

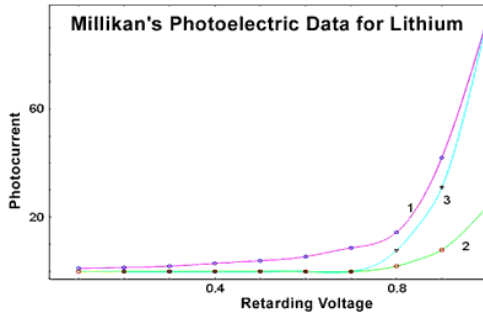


Fig. 4.

The graphs in Fig. 5 match the above experimental data to various graphs (in red) of equation (12)

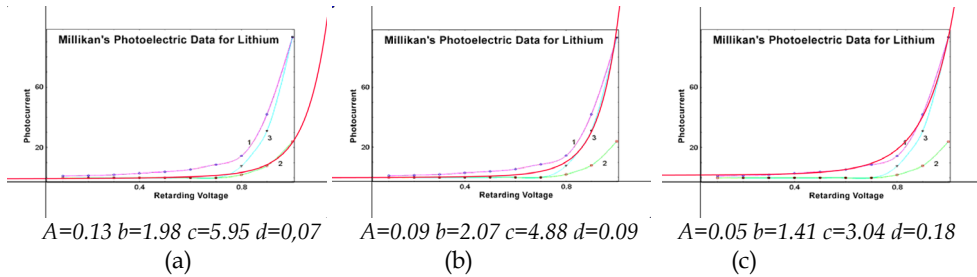


Fig. 5.

The above graphs (Fig. 5) seem to suggest that Eq. (11) agrees well with the experimental data showing the asymptotic behavior of the (photocurrent) v (energy) curves. But more systematic experimental work is needed.

**9. Meaning and derivation of the De Broglie equations [Ragazas 2011a]**

Consider  $\eta_0(x_0, t_0) \rightarrow \eta(x, t)$ . We can write  $\frac{\Delta\eta}{\eta_0} = \%$ -change of  $\eta =$  'cycle of change'. For

corresponding  $\Delta x$  and  $\Delta t$  we can write,  $\lambda = \frac{\Delta x}{\Delta\eta/\eta_0} =$  "distance per cycle of change" and

$$v = \frac{\Delta\eta/\eta_0}{\Delta t} = \text{"cycle of change per time"}. \text{ We can rewrite these as } \lambda = \frac{\eta_0}{\Delta\eta/\Delta x} \text{ and } v = \frac{\Delta\eta/\Delta t}{\eta_0}.$$

Taking limits and letting  $\eta_0 = h$  (Planck's constant being the minimal  $\eta$  that can be measured) we get the *de Broglie equations*:

$$\lambda = \frac{\eta_0}{\Delta\eta/\Delta x} \rightarrow \frac{h}{\partial\eta/\partial x} = \frac{h}{p} \text{ and } v = \frac{\Delta\eta/\Delta t}{\eta_0} \rightarrow \frac{\partial\eta/\partial t}{h} = \frac{E}{h}$$

Note: Since %-change in  $\eta$  can be both positive or negative,  $\lambda$  and  $\nu$  can be both positive or negative.

### 10. The 'exponential of energy' $E(t) = E_0 e^{\nu t}$ [Ragazas 2010i, 2011a]

From Section 9.0 above we have that  $\nu$  equals "%-change of  $\eta$  per unit of time". If we consider *continuous change*, we can express this as  $\eta = \eta_0 e^{\nu t}$ . Differentiating with respect to  $t$  we have,  $E(t) = \frac{\partial \eta}{\partial t} = \eta_0 \nu e^{\nu t}$  and  $E_0 = \eta_0 \nu$ . Thus,  $E(t) = E_0 e^{\nu t}$

### 11. Proposition: "If the speed of light is constant, then light is a wave" [Ragazas 2011b]

*Proof:* We have that  $\lambda = \frac{h}{p}$ ,  $\nu = \frac{E}{h}$  and  $\lambda \nu = c$ . Since  $p = \frac{\partial \eta}{\partial x}$  and  $E = \frac{\partial \eta}{\partial t}$ , we have that

$\lambda \nu = \frac{\partial \eta / \partial t}{\partial \eta / \partial x}$ . Differentiating, we get

$$D_t(\lambda \nu) = \frac{\frac{\partial^2 \eta}{\partial t^2} \cdot \frac{\partial \eta}{\partial x} - \frac{\partial \eta}{\partial t} \cdot \frac{\partial^2 \eta}{\partial t \partial x}}{\left(\frac{\partial \eta}{\partial x}\right)^2} \quad \text{and} \quad D_x(\lambda \nu) = \frac{\frac{\partial^2 \eta}{\partial x \partial t} \cdot \frac{\partial \eta}{\partial x} - \frac{\partial \eta}{\partial t} \cdot \frac{\partial^2 \eta}{\partial x^2}}{\left(\frac{\partial \eta}{\partial x}\right)^2}$$

Since  $\lambda \nu = c$ , we have that  $D_t(\lambda \nu) = 0$  and  $D_x(\lambda \nu) = 0$ . Therefore,

$$\frac{\partial^2 \eta}{\partial t^2} \cdot \frac{\partial \eta}{\partial x} - \frac{\partial \eta}{\partial t} \cdot \frac{\partial^2 \eta}{\partial t \partial x} = 0 \quad \text{and} \quad \frac{\partial^2 \eta}{\partial x \partial t} \cdot \frac{\partial \eta}{\partial x} - \frac{\partial \eta}{\partial t} \cdot \frac{\partial^2 \eta}{\partial x^2} = 0$$

Using  $\lambda \nu = \frac{\partial \eta / \partial t}{\partial \eta / \partial x}$  and  $\lambda \nu = c$ , these can be written as,

$$\frac{\partial^2 \eta}{\partial t^2} = c \cdot \frac{\partial^2 \eta}{\partial t \partial x} \quad \text{and} \quad \frac{\partial^2 \eta}{\partial x \partial t} = c \cdot \frac{\partial^2 \eta}{\partial x^2}$$

Since 'mixed partials are equal', these equations combine to give us,

$$\frac{\partial^2 \eta}{\partial t^2} = c^2 \cdot \frac{\partial^2 \eta}{\partial x^2}, \quad \text{the wave equation in one dimension}$$

Thus, for the speed of light to be constant the 'propagation of light'  $\eta$  must be a solution to the wave equation. *q.e.d*

### 12. The double-slit experiment [Ragazas 2011a]

The 'double-slit experiment' (where a beam of light passes through two narrow parallel slits and projects onto a screen an interference pattern) was originally used by Thomas Young in 1803, and latter by others, to demonstrate the wave nature of light. This experiment later



came in direct conflict, however, with Einstein's *Photon Hypothesis* explanation of the Photoelectric Effect which establishes the particle nature of light. Reconciling these logically antithetical views has been a major challenge for physicists. The double-slit experiment embodies this quintessential mystery of Quantum Mechanics.

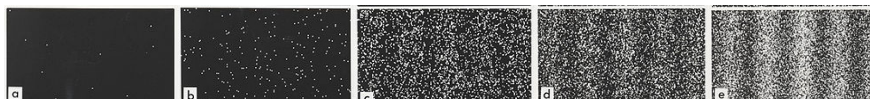


Fig. 6.

There are many variations and strained explanations of this simple experiment and new methods to prove or disprove its implications to Physics. But the 1989 Tonomura 'single electron emissions' experiment provides the clearest expression of this wave-particle enigma. In this experiment single emissions of electrons go through a simulated double-slit barrier and are recorded at a detection screen as 'points of light' that over time randomly fill in an interference pattern. The picture frames in Fig. 6 illustrate these experimental results. We will use these results in explaining the *double-slit experiment*.

### 12.1 Plausible explanation of the double-slit experiment

The basic logical components of this *double-slit experiment* are the 'emission of an electron at the source' and the subsequent 'detection of an electron at the screen'. It is commonly assumed that these two events are directly connected. The electron emitted at the source is assumed to be the same electron as the electron detected at the screen. We take the view that this may not be so. Though the two events (emission and detection) are related, they may not be directly connected. That is to say, there may not be a 'trajectory' that directly connects the electron emitted with the electron detected. And though many explanations in Quantum Mechanics do not seek to trace out a trajectory, nonetheless in these interpretations the detected electron is tacitly assumed to be the same as the emitted electron. This we believe is the source of the dilemma. We further adapt the view that while energy propagates continuously as a wave, the measurement and manifestation of energy is made in discrete units (*equal size sips*). This view is supported by all our results presented in this Chapter. And just as we would never characterize the nature of a vast ocean as consisting of discrete 'bucketfuls of water' because that's how we draw the water from the ocean, similarly we should not conclude that energy consists of discrete energy quanta simply because that's how energy is absorbed in our measurements of it.

The 'light burst' at the detection screen in the Tonomura *double-slit experiment* may not signify the arrival of "the" electron emitted from the source and going through one or the other of the two slits as a particle strikes the screen as a 'point of light'. The 'firing of an electron' at the source and the 'detection of an electron' at the screen are two separate events. What we have at the detection screen is a separate event of a light burst at some point on the screen, having absorbed enough energy to cause it to 'pop' (like popcorn at seemingly random manner once a seed has absorbed enough heat energy). The parts of the detection screen that over time are illuminated more by energy will of course show more 'popping'. The emission of an electron at the source is a separate event from the detection of a light burst at the screen. Though these events are connected they are not directly connected. There is no trajectory that connects these two electrons as being one and the same. The electron 'emitted' is not the same electron 'detected'.

What is emitted as an electron is a burst of energy which propagates continuously as a wave and going through both slits illuminates the detection screen in the typical interference pattern. This interference pattern is clearly visible when a large beam of energy illuminates the detection screen all at once. If we systematically lower the intensity of such electron beam the intensity of the illuminated interference pattern also correspondingly fades. For small bursts of energy, the interference pattern illuminated on the screen may be undetectable as a whole. However, when at a point on the screen *local equilibrium* occurs, we get a 'light burst' that in effect discharges the screen of an amount of energy equal to the energy burst that illuminated the screen. These points of discharge will be more likely to occur at those areas on the screen where the illumination is greatest. Over time we would get these dots of light filling the screen in the interference pattern.

We have a 'reciprocal relation' between 'energy' and 'time'. Thus, 'lowering energy intensity' while 'increasing time duration' is equivalent to 'increasing energy intensity' and 'lowering time duration'. But the resulting phenomenon is the same: the interference pattern we observe. This explanation of the *double-slit* experiment is logically consistent with the 'probability distribution' interpretation of Quantum Mechanics. The view we have of energy propagating continuously as a wave while manifesting locally in discrete units (*equal size sips*) when *local equilibrium* occurs, helps resolve the *wave-particle dilemma*.

### 12.2 Explanation summary

The argument presented above rests on the following ideas. These are consistent with all our results presented in this Chapter.

1. The 'electron emitted' is not be the same as the 'electron detected'.
2. Energy 'propagates continuously' but 'interacts discretely' when equilibrium occurs
3. We have 'accumulation of energy' before 'manifestation of energy'.

Our thinking and reasoning are also guided by the following attitude of *physical realism*:

- a. Changing our detection devices while keeping the experimental setup the same can reveal something 'more' of the examined phenomenon but not something 'contradictory'.
- b. If changing our detection devices reveals something 'contradictory', this is due to the detection device design and not to a change in the physics of the phenomenon examined.

Thus, using *physical realism* we argue that if we keep the experimental apparatus constant but only replace our 'detection devices' and as a consequence we detect something contradictory, the physics of the double slit experiment does not change. The experimental behavior has not changed, just the display of this behavior by our detection device has changed. The 'source' of the beam has not changed. The effect of the double slit barrier on that beam has not changed. So if our detector is now telling us that we are detecting 'particles' whereas before using other detector devices we were detecting 'waves', *physical realism* should tell us that this is entirely due to the change in our methods of detection. For the same input, our instruments may be so designed to produce different outputs.

### 13. Conclusion

In this Chapter we have sought to present a thumbnail sketch of *a world without quanta*. We started at the very foundations of Modern Physics with a simple and continuous mathematical derivation of *Planck's Law*. We demonstrated that *Planck's Law is an exact mathematical identity that describes the interaction of energy*. This fact alone explains why *Planck's Law* fits so exceptionally well the experimental data.

Using our derivation of *Planck's Law* as a *Rosetta Stone* (linking Mechanics, Quantum Mechanics and Thermodynamics) we considered the *quantity eta* that naturally appears in our derivation as *prime physis*. Planck's constant  $h$  is such a quantity. Energy can be defined as the time-rate of *eta* while momentum as the space-rate of *eta*. Other physical quantities can likewise be defined in terms of *eta*. Laws of Physics can and must be mathematically derived and not physically posited as Universal Laws chiseled into cosmic dust by the hand of God.

We postulated the *Identity of Eta Principle*, derived the Conservation of Energy and Momentum, derived Newton's Second Law of Motion, established the intimate connection between entropy and time, interpreted Schoedinger's equation and suggested that the *wave-function*  $\psi$  is in fact *prime physis*  $\eta$ . We showed that The Second Law of Thermodynamics pertains to *time* (and not entropy, which can be both positive and negative) and should be reworded to state that '*all physical processes take some positive duration of time to occur*'. We also showed the unexpected mathematical equivalence between *Planck's Law* and Boltzmann's *Entropy Equation* and proved that '*if the speed of light is a constant, then light is a wave*'.

#### 14. Appendix: Mathematical derivations

The proofs to many of the derivations below are too simple and are omitted for brevity. But the propositions are listed for purposes of reference and completeness of exposition.

*Notation.* We will consistently use the following notation throughout this APPENDIX:

$E(t)$  is a real-valued function of the real-variable  $t$

$\Delta t = t - s$  is an 'interval of  $t$ '

$\Delta E = E(t) - E(s)$  is the 'change of  $E$ '

$P = \int_s^t E(u) du$  is the 'accumulation of  $E$ '

$\bar{E} = E_{av} = \frac{1}{t-s} \int_s^t E(u) du$  is the 'average of  $E$ '

$D_x$  indicates 'differentiation with respect to  $x$ '

$r$  is a constant, often an 'exponential rate of growth'

##### 14.1 Part I: Exponential functions

We will use the following characterization of exponential functions without proof:

*Basic Characterization:*  $E(t) = E_0 e^{rt}$  if and only if  $D_t E = rE$

*Characterization 1:*  $E(t) = E_0 e^{rt}$  if and only if  $\Delta E = Pr$

*Proof:* Assume that  $E(t) = E_0 e^{rt}$ . We have that  $\Delta E = E(t) - E(s) = E_0 e^{rt} - E_0 e^{rs}$ ,

while  $P = \int_s^t E_0 e^{ru} du = \frac{1}{r} [E_0 e^{rt} - E_0 e^{rs}] = \frac{\Delta E}{r}$ . Therefore  $\Delta E = Pr$ .

Assume next that  $\Delta E = Pr$ . Differentiating with respect to  $t$ ,  $D_t E = rD_t P = rE$ .

Therefore by the *Basic Characterization*,  $E(t) = E_0 e^{rt}$ . *q.e.d*

*Theorem 1:*  $E(t) = E_0 e^{rt}$  if and only if  $\frac{Pr}{e^{r\Delta t} - 1}$  is invariant with respect to  $t$

*Proof:* Assume that  $E(t) = E_0 e^{rt}$ . Then we have, for fixed  $s$ ,

$$P = \int_s^t E_0 e^{ru} du = \frac{E_0}{r} [e^{rt} - e^{rs}] = \frac{E_0 e^{rs}}{r} [e^{r(t-s)} - 1] = \frac{E(s)}{r} (e^{r(t-s)} - 1)$$

and from this we get that  $\frac{Pr}{e^{r\Delta t} - 1} = E(s) = \text{constant}$ . Assume next that  $\frac{Pr}{e^{r\Delta t} - 1} = C$  is constant with respect to  $t$ , for fixed  $s$ .

Therefore,  $D_t \left[ \frac{Pr}{e^{r\Delta t} - 1} \right] = \frac{rE(t) \cdot [e^{r\Delta t} - 1] - rP \cdot [re^{r\Delta t}]}{(e^{r\Delta t} - 1)^2} = 0$  and so,  $E(t) = \left( \frac{Pr}{e^{r\Delta t} - 1} \right) e^{r\Delta t} = C \cdot e^{r\Delta t}$

where  $C$  is constant. Letting  $t = s$  we get  $E(s) = C$ . We can rewrite this as  $E(t) = E(s)e^{r(t-s)} = E_0 e^{rt}$ . *q.e.d*

From the above, we have

*Characterization 2:*  $E(t) = E_0 e^{rt}$  if and only if  $\frac{Pr}{e^{r(t-s)} - 1} = E(s)$

Clearly by definition of  $E_{av}$ ,  $r\Delta t = \frac{Pr}{E_{av}}$ . We can write  $\frac{Pr}{e^{r\Delta t} - 1}$  equivalently as  $\frac{Pr}{e^{Pr/E_{av}} - 1}$  in the above. *Theorem 1* above can therefore be restated as,

*Theorem 1a:*  $E(t) = E_0 e^{rt}$  if and only if  $\frac{Pr}{e^{Pr/E_{av}} - 1}$  is invariant with  $t$

The above *Characterization 2* can then be restated as

*Characterization 2a:*  $E(t) = E_0 e^{rt}$  if and only if  $\frac{Pr}{e^{Pr/E_{av}} - 1} = E(s)$ .

But if  $\frac{Pr}{e^{Pr/E_{av}} - 1} = E(s)$ , then by *Characterization 2a*,  $E(t) = E_0 e^{rt}$ . Then, by *Characterization 1*,

we must have that  $\Delta E = Pr$ . And so we can write equivalently  $\frac{\Delta E}{e^{\Delta E/E_{av}} - 1} = E(s)$ . We have the following equivalence,

*Characterization 3:*  $E(t) = E_0 e^{rt}$  if and only if  $\frac{\Delta E}{e^{\Delta E/E_{av}} - 1} = E(s)$

As we've seen above, it is always true that  $\frac{Pr}{E_{av}} = r\Delta t$ . But for exponential functions  $E(t)$  we also have that  $\Delta E = Pr$ . So, for exponential functions we have the following.

*Characterization 4:*  $E(t) = E_0 e^{rt}$  if and only if  $\frac{\Delta E}{E_{av}} = r\Delta t$

## 14.2 Part II: Integrable functions

We next consider that  $E(t)$  is any function. In this case, we have the following.

Theorem 2: a) For any differentiable function  $E(t)$ ,  $\lim_{t \rightarrow s} \frac{\Delta E}{e^{\Delta E/E_{av}} - 1} = E(s)$

b) For any integrable function  $E(t)$ ,  $\lim_{t \rightarrow s} \frac{Pr}{e^{r\Delta t} - 1} = E(s)$

Proof: Since  $\frac{\Delta E}{e^{\Delta E/E_{av}} - 1} \rightarrow \frac{0}{0}$  and  $\frac{Pr}{e^{r\Delta t} - 1} \rightarrow \frac{0}{0}$  as  $t \rightarrow s$ , we apply L'Hopital's Rule.

$$\begin{aligned} \lim_{t \rightarrow s} \frac{\Delta E}{e^{\Delta E/\bar{E}} - 1} &= \lim_{t \rightarrow s} \frac{D_t E(t)}{e^{\Delta E/\bar{E}} \cdot \left[ \frac{D_t E(t) \cdot \bar{E} - D_t \bar{E} \cdot \Delta E}{\bar{E}^2} \right]} \\ &= \lim_{t \rightarrow s} \frac{\bar{E}^2 \cdot D_t E(t)}{e^{\Delta E/\bar{E}} \cdot [D_t E(t) \cdot \bar{E} - D_t \bar{E} \cdot \Delta E]} = E(s) \end{aligned}$$

since  $\Delta E \rightarrow 0$  and  $\bar{E} \rightarrow E(s)$  as  $t \rightarrow s$ .

Likewise, we have  $\lim_{t \rightarrow s} \frac{Pr}{e^{r\Delta t} - 1} = \lim_{t \rightarrow s} \frac{E(s)r}{e^{r\Delta t} \cdot r} = E(s)$ . *q.e.d.*

Corollary A:  $\frac{\Delta E}{e^{\Delta E/\bar{E}} - 1}$  is invariant with  $t$  if and only if  $E(s) = \frac{\Delta E}{e^{\Delta E/\bar{E}} - 1}$

Proof: Using Theorem 2 we have  $\lim_{t \rightarrow s} \frac{\Delta E}{e^{\Delta E/E_{av}} - 1} = E(s)$ . Since  $\frac{\Delta E}{e^{\Delta E/E_{av}} - 1}$  is constant with respect to  $t$ , we have  $E(s) = \frac{\Delta E}{e^{\Delta E/E_{av}} - 1}$ . Conversely, if  $E(s) = \frac{\Delta E}{e^{\Delta E/E_{av}} - 1}$ , then by Characterization 3,  $E(s) = E_0 e^{rs}$ . Since  $E(s)$  is a constant,  $\frac{\Delta E}{e^{\Delta E/E_{av}} - 1}$  is invariant with respect to  $t$ . *q.e.d.*

Since it is always true by definitions that  $r\Delta t = \frac{Pr}{E_{av}}$ , Theorem 2 can also be written as,

Theorem 2a: For any integrable function  $E(t)$ ,  $\lim_{t \rightarrow s} \frac{Pr}{e^{Pr/E_{av}} - 1} = E(s)$

As a direct consequence of the above, we have the following interesting and important result:

Corollary B:  $E(s) = \frac{\Delta E}{e^{\Delta E/E_{av}} - 1}$  and  $E(s) = \frac{Pr}{e^{Pr/E_{av}} - 1}$  are independent of  $\Delta t$ ,  $\Delta E$ .

### 14.3 Part III: Independent proof of Characterization 3

In the following we provide a direct and independent proof of Characterization 3.

We first prove the following,

Lemma: For any  $E$ ,  $D_t \bar{E}(t) = \frac{E(t) - \bar{E}}{t - s}$  and  $D_s \bar{E}(s) = \frac{\bar{E} - E(s)}{t - s}$

Proof: We let  $\Delta t = t - s$  and  $\bar{E} = \frac{1}{t - s} \int_s^t E(u) du$ .

Differentiating with respect to  $t$  we have  $(t - s) \cdot D_t \bar{E}(t) + \bar{E} = E(t)$ .

Rewriting, we have  $D_t \bar{E}(t) = \frac{E(t) - \bar{E}}{t - s}$ . Differentiating with respect to  $s$  we have

$$(t - s) \cdot D_s \bar{E}(s) - \bar{E} = -E(s). \text{ Rewriting, we have } D_s \bar{E}(s) = \frac{\bar{E} - E(s)}{t - s}. \text{ q.e.d.}$$

Characterization 3:  $E(t) = E_0 e^{rt}$  if and only if  $\frac{\Delta E}{e^{\Delta E/E_{av}} - 1} = E(s)$

Proof: Assume that  $E(t) = E_0 e^{rt}$ . From,

$$P = \int_s^t E_0 e^{ru} du = \frac{E_0}{r} [e^{rt} - e^{rs}] = \frac{E_0 e^{rs}}{r} [e^{r\Delta t} - 1] = \frac{E(s)}{r} [e^{r\Delta t} - 1]$$

we get,  $E(s) = \frac{Pr}{e^{r\Delta t} - 1}$ . This can be rewritten as,  $E(s) = \frac{Pr}{e^{Pr/E_{av}} - 1}$ . Since  $\Delta E = Pr$ , this can

further be written as  $E(s) = \frac{\Delta E}{e^{\Delta E/E_{av}} - 1}$ .

Conversely, consider next a function  $E(s)$  satisfying

$$E(s) = \frac{\Delta E}{e^\xi - 1}, \text{ where } \begin{cases} \Delta E = E(t) - E(s) \\ \Delta t = t - s \\ \xi = \frac{\Delta E}{\bar{E}} \\ \bar{E} = \frac{1}{\Delta t} \int_s^t E(u) du \end{cases} \text{ and } t \text{ can be any real value.}$$

From the above, we have that  $e^\xi = \frac{\Delta E}{E(s)} + 1 = \frac{E(t) - E(s) + E(s)}{E(s)} = \frac{E(t)}{E(s)}$ .

Differentiating with respect to  $s$ , we get  $e^\xi \cdot D_s \xi = \frac{-E(t) \cdot D_s E(s)}{E(s)^2} = -e^\xi \cdot \frac{D_s E(s)}{E(s)}$

$$\text{and so, } D_s \xi = -\frac{D_s E(s)}{E(s)} \tag{A1}$$

From the above Lemma we have

$$D_s \bar{E}(s) = \frac{\bar{E} - E(s)}{t - s} \tag{A2}$$

Differentiating  $\xi = \frac{\Delta E}{\bar{E}}$  with respect to  $s$  we get,

$$D_s \xi = \frac{-D_s E(s) \cdot \bar{E} - \Delta E \cdot D_s \bar{E}(s)}{\bar{E}^2} \tag{A3}$$

and combining (A1), (A2), and (A3) we have

$$-\frac{D_s E(s)}{E(s)} = \frac{-D_s E(s) \cdot \bar{E} - \frac{\Delta E}{\Delta t} (\bar{E} - E(s))}{\bar{E}^2} = -\frac{D_s E(s)}{\bar{E}} - \frac{\Delta E}{\Delta t} \cdot \frac{(\bar{E} - E(s))}{\bar{E}^2}$$

We can rewrite the above as follows,

$$\frac{D_s E(s)}{E(s)} - \frac{D_s E(s)}{\bar{E}} = D_s E(s) \left( \frac{\bar{E} - E(s)}{E(s) \cdot \bar{E}} \right) = \frac{\Delta E}{\Delta t} \cdot \frac{(\bar{E} - E(s))}{\bar{E}^2}$$

and so, 
$$\frac{D_s E(s)}{E(s)} = \frac{\Delta E}{\Delta t} \cdot \frac{1}{\bar{E}}.$$

Using (A1), this can be written as

$$-D_s \xi = \frac{\xi}{\Delta t}, \text{ or as } \xi = -D_s \xi \cdot \Delta t. \quad (\text{A4})$$

Differentiating (A4) above with respect to  $s$ , we get  $D_s \xi = -D_s^2 \xi \cdot \Delta t + D_s \xi$ . Therefore,  $D_s^2 \xi = 0$ . Working backward, this gives  $D_s \xi = -r = \text{constant}$ .

From (A1), we then have that  $\frac{D_s E(s)}{E(s)} = r$  and therefore  $E(s) = E_0 e^{rs}$ . *q.e.d.*

## 15. Acknowledgement

I am indebted to Segun Chanillo, Prof. of Mathematics, Rutgers University for his encouragement, when all others thought my efforts were futile. Also, I am deeply grateful to Hayrani Oz, Prof. of Aerospace Engineering, Ohio State University, who discovered my posts on the web and was the first to recognize the significance of my results in Physics. Special thanks also to Miguel Bayona of The Lawrenceville School for his friendship and help with the graphics in this chapter. And Alexander Morisse who is my best and severest critic of the Physics in these results.

## 16. References

- Frank, Adam (2010), *Who Wrote the Book of Physics?* Discover Magazine (April 2010)
- Keesing, Richard (2001). *Einstein, Millikan and the Photoelectric Effect*, Open University Physics Society Newsletter, Winter 2001/2002 Vol 1 Issue 4  
<http://www.oufusion.org.uk/pdf/FusionNewsWinter01.pdf>
- Öz, H., Algebraic Evolutionary Energy Method for Dynamics and Control, in: *Computational Nonlinear Aeroelasticity for Multidisciplinary Analysis and Design*, AFRL, VA-WP-TR-2002-XXXX, 2002, pp. 96-162.
- Öz, H., Evolutionary Energy Method (EEM): An Aerothermoservoelastoc Application.: *Variational and Extremum Principles in Macroscopic Systems*, Elsevier, 2005, pp. 641-670.
- Öz, H., *The Law Of Evolutionary Enerxaction and Evolutionary Enerxaction Dynamics*, Seminar presented at Cambridge University, England, March 27, 2008,  
<http://talks.cam.ac.uk/show/archive/12743>
- Öz, Hayrani; John K. Ramsey, *Time modes and nonlinear systems*, Journal of Sound and Vibration, 329 (2010) 2565-2602, doi:10.1016/j.jsv.2009.12.021
- Planck, Max (1901) *On the Energy Distribution in the Blackbody Spectrum*, Ann. Phys. 4, 553, 1901
- Ragazas, C. (2010) *A Planck-like Characterization of Exponential Functions*, knol

- <http://knol.google.com/k/constantinos-ragazas/a-planck-like-characterization-of/ql47o1qdr604/7#>
- Ragazas, C. (2010) *Entropy and 'The Arrow of Time'*, knol <http://knol.google.com/k/constantinos-ragazas/entropy-and-the-arrow-of-time/ql47o1qdr604/17#>
- Ragazas, C. (2010) *"Let there be h": An Existence Argument for Planck's Constant*, knol <http://knol.google.com/k/constantinos-ragazas/let-there-be-h-an-existence-argument/ql47o1qdr604/12#>
- Ragazas, C. (2010) *Prime 'physis' and the Mathematical Derivation of Basic Law*, knol <http://knol.google.com/k/constantinos-ragazas/prime-physis-and-the-mathematical/ql47o1qdr604/10#>
- Ragazas, C. (2010) *"The meaning of  $\psi$ ": An Interpretation of Schroedinger's Equations*, knol <http://knol.google.com/k/constantinos-ragazas/the-meaning-of-psi-an-interpretation-of/ql47o1qdr604/14#>
- Ragazas, C. (2010) *Planck's Law is an Exact Mathematical Identity*, knol <http://knol.google.com/k/constantinos-ragazas/planck-s-law-is-an-exact-mathematical/ql47o1qdr604/3#>
- Ragazas, C. (2010) *The Temperature of Radiation*, knol <http://knol.google.com/k/constantinos-ragazas/the-temperature-of-radiation/ql47o1qdr604/6#>
- Ragazas, C. (2010) *The Interaction of Measurement*, knol <http://knol.google.com/k/constantinos-ragazas/the-interaction-of-measurement/ql47o1qdr604/11#>
- Ragazas, C. (2010) *A Time-dependent Local Representation of Energy*, knol <http://knol.google.com/k/constantinos-ragazas/a-time-dependent-local-representation/ql47o1qdr604/9#>
- Ragazas, C. (2010) *A Plausible Explanation of the Double-slit Experiment in Physics*, knol <http://knol.google.com/k/constantinos-ragazas/a-plausible-explanation-of-the-double/ql47o1qdr604/4#>
- Ragazas, C. (2010) *The Photoelectric Effect Without Photons*, knol <http://knol.google.com/k/constantinos-ragazas/the-photoelectric-effect-without-photons/ql47o1qdr604/8#>
- Ragazas, C. (2010) *Stocks and Planck's Law*, knol <http://knol.google.com/k/constantinos-ragazas/stocks-and-planck-s-law/ql47o1qdr604/2#>
- Ragazas, C. (2011) *What is The Matter With de Broglie Waves?* knol <http://knol.google.com/k/constantinos-ragazas/what-is-the-matter-with-de-broglie-waves/ql47o1qdr604/18#>
- Ragazas, C. (2011) *"If the Speed of Light is a Constant, Then Light is a Wave"*, knol <http://knol.google.com/k/constantinos-ragazas/if-the-speed-of-light-is-a-constant/ql47o1qdr604/19#>
- Tonomura (1989) <http://www.hitachi.com/rd/research/em/doubleslit.html>
- Wikipedia, (n.d.) [http://en.wikipedia.org/wiki/File:Firas\\_spectrum.jpg](http://en.wikipedia.org/wiki/File:Firas_spectrum.jpg)



# Statistical Thermodynamics

Anatol Malijeuský

*Department of Physical Chemistry, Institute of Chemical Technology, Prague  
Czech Republic*

## 1. Introduction

This chapter deals with the statistical thermodynamics (statistical mechanics) a modern alternative of the classical (phenomenological) thermodynamics. Its aim is to determine thermodynamic properties of matter from forces acting among molecules. Roots of the discipline are in kinetic theory of gases and are connected with the names Maxwell and Boltzmann. Father of the statistical thermodynamics is Gibbs who introduced its concepts such as the statistical ensemble and others, that have been used up to present.

Nothing can express an importance of the statistical thermodynamics better than the words of Richard Feynman Feynman et al. (2006), the Nobel Prize winner in physics: *If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generations of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis (or the atomic fact, or whatever you wish to call it) that **All things are made of atoms – little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another.***

In that one sentence, you will see, there is an enormous amount of information about the world, if just a little imagination and thinking are applied.

The chapter is organized as follows. Next section contains axioms of the phenomenological thermodynamics. Basic concepts and axioms of the statistical thermodynamics and relations between the partition function and thermodynamic quantities are in Section 3. Section 4 deals with the ideal gas and Section 5 with the ideal crystal. Intermolecular forces are discussed in Section 6. Section 7 is devoted to the virial expansion and Section 8 to the theories of dense gases and liquids. The final section comments axioms of phenomenological thermodynamics in the light of the statistical thermodynamics.

## 2. Principles of phenomenological thermodynamics

The phenomenological thermodynamics or simply thermodynamics is a discipline that deals with the thermodynamic system, a macroscopic part of the world. The thermodynamic state of system is given by a limited number of thermodynamic variables. In the simplest case of one-component, one-phase system it is for example volume of the system, amount of substance (e.g. in moles) and temperature. Thermodynamics studies changes of thermodynamic quantities such as pressure, internal energy, entropy, *e.t.c.* with thermodynamic variables.

The phenomenological thermodynamics is based on six axioms (or postulates if you wish to call them), four of them are called the laws of thermodynamics:

- **Axiom of existence of the thermodynamic equilibrium**

For thermodynamic system at unchained external conditions there exists a state of the thermodynamic equilibrium in which its macroscopic parameters remain constant in time. The thermodynamic system at unchained external conditions always reaches the state of the thermodynamic equilibrium.

- **Axiom of additivity**

Energy of the thermodynamic system is a sum of energies of its macroscopic parts. This axiom allows to define extensive and intensive thermodynamic quantities.

- **The zeroth law of thermodynamics**

When two systems are in the thermal equilibrium, *i.e.* no heat flows from one system to the other during their thermal contact, then both systems have the same temperature as an intensive thermodynamic parameter. If system A has the same temperature as system B and system B has the same temperature as system C, then system A also has the same temperature as system C (temperature is transitive).

- **The first law of thermodynamics**

There is a function of state called internal energy  $U$ . For its total differential  $dU$  we write

$$dU = \delta W + \delta Q, \quad (1)$$

where the symbols  $\delta Q$  and  $\delta W$  are not total differentials but represent infinitesimal values of heat  $Q$  and work  $W$  supplied to the system.

- **The second law of thermodynamics**

There is a function of state called entropy  $S$ . For its total differential  $dS$  we write

$$dS = \frac{\delta Q}{T}, \quad [\text{reversible process}], \quad (2)$$

$$dS > \frac{\delta Q}{T}, \quad [\text{irreversible process}]. \quad (3)$$

- **The third law of thermodynamics**

At temperature of 0 K, entropy of a pure substance in its most stable crystalline form is zero

$$\lim_{T \rightarrow 0} S = 0. \quad (4)$$

This postulate supplements the second law of thermodynamics by defining a natural referential value of entropy. The third law of thermodynamics implies that temperature of 0 K cannot be attained by any process with a finite number of steps.

Phenomenological thermodynamics using its axioms radically reduces an amount of experimental effort necessary for a determination of the values of thermodynamic quantities. For example enthalpy or entropy of a pure fluid need not be measured at each temperature and pressure but they can be calculated from an equation of state and a temperature dependence of the isobaric heat capacity of ideal gas. However, empirical constants in an equation of state and in the heat capacity must be obtained experimentally.

### 3. Principles of statistical thermodynamics

#### 3.1 Basic concepts

The statistical thermodynamics considers thermodynamic system as an assembly of a very large number (of the order of  $10^{23}$ ) of mutually interacting particles (usually molecules). It uses the following concepts:

- **Microscopic state of system**

The microscopic state of thermodynamic system is given by positions and velocities of all particles in the language of the Newton mechanics, or by the quantum states of the system in the language of quantum mechanics. There is a huge number of microscopic states that correspond to a given thermodynamic (macroscopic) state of the system.

- **Statistical ensemble**

Statistical ensemble is a collection of all systems that are in the same thermodynamic state but in the different microscopic states.

- **Microcanonical ensemble** or **NVE ensemble** is a collection of all systems at a given number of particles  $N$ , volume  $V$  and energy  $E$ .

- **Canonical ensemble** or **NVT ensemble** is a collection of all systems at a given number of particles  $N$ , volume  $V$  and temperature  $T$ .

There is a number of ensembles, *e.g.* the grandcanonical ( $\mu VT$ ) or isothermal isobaric (**NPT**) that will not be considered in this work.

- **Time average of thermodynamic quantity**

The time average  $\bar{X}_\tau$  of a thermodynamic quantity  $X$  is given by

$$\bar{X}_\tau = \frac{1}{\tau} \int_0^\tau X(t) dt, \quad (5)$$

where  $X(t)$  is a value of  $X$  at time  $t$  and,  $\tau$  is a time interval of a measurement.

- **Ensemble average of thermodynamic quantity**

The ensemble average  $\bar{X}_s$  of a thermodynamic quantity  $X$  is given by

$$\bar{X}_s = \sum_i P_i X_i, \quad (6)$$

where  $X_i$  is a value in the quantum state  $i$ , and  $P_i$  is the probability of the quantum state.

#### 3.2 Axioms of the statistical thermodynamics

The statistical thermodynamics is based on two axioms:

##### Axiom on equivalence of average values

It is postulated that the time average of thermodynamic quantity  $X$  is equivalent to its ensemble average

$$\bar{X}_\tau = \bar{X}_s. \quad (7)$$

##### Axiom on probability

Probability  $P_i$  of a quantum state  $i$  is only a function of energy of the quantum state,  $E_i$ ,

$$P_i = f(E_i). \quad (8)$$

### 3.3 Probability in the microcanonical and canonical ensemble

From Eq.(8) relations between the probability and energy can be derived:

#### Probability in the microcanonical ensemble

All the microscopic states in the microcanonical ensemble have the same energy. Therefore,

$$P_i = \frac{1}{W} \quad \text{for } i = 1, 2, \dots, W, \quad (9)$$

where  $W$  is a number of microscopical states (the statistical weight) of the microcanonical ensemble.

#### Probability in the canonical ensemble

In the canonical ensemble it holds

$$P_i = \frac{\exp(-\beta E_i)}{Q}, \quad (10)$$

where  $\beta = \frac{1}{k_B T}$ ,  $k_B$  is the Boltzmann constant,  $T$  temperature and  $Q$  is the *partition function*

$$Q = \sum_i \exp(-\beta E_i), \quad (11)$$

where the sum is over the microscopic states of the canonical ensemble.

### 3.4 The partition function and thermodynamic quantities

If the partition function is known thermodynamic quantities may be determined. The following relations between the partition function in the canonical ensemble and thermodynamic quantities can be derived

$$A = -k_B T \ln Q \quad (12)$$

$$U = k_B T^2 \left( \frac{\partial \ln Q}{\partial T} \right)_V \quad (13)$$

$$S = k_B \ln Q + k_B T \left( \frac{\partial \ln Q}{\partial T} \right)_V. \quad (14)$$

$$C_V = \left( \frac{\partial U}{\partial T} \right)_V = k_B T^2 \frac{\partial^2 \ln Q}{\partial T^2} + 2k_B T \left( \frac{\partial \ln Q}{\partial T} \right)_V, \quad (15)$$

$$p = - \left( \frac{\partial A}{\partial V} \right)_T = k_B T \left( \frac{\partial \ln Q}{\partial V} \right)_T, \quad (16)$$

$$H = U + pV = k_B T^2 \left( \frac{\partial \ln Q}{\partial T} \right)_V + V k_B T \left( \frac{\partial \ln Q}{\partial V} \right)_T, \quad (17)$$

$$G = A + pV = -k_B T \ln Q + V k_B T \left( \frac{\partial \ln Q}{\partial V} \right)_T, \quad (18)$$

$$C_p = \left( \frac{\partial H}{\partial T} \right)_V = C_V + V k_B \frac{\partial^2 \ln Q}{\partial V \partial T}. \quad (19)$$

$A$  is Helmholtz free energy,  $U$  internal energy,  $S$  entropy,  $C_V$  isochoric heat capacity,  $p$  pressure,  $H$  enthalpy,  $G$  Gibbs free energy and  $C_p$  isobaric heat capacity.

Unfortunately, the partition function is known only for the simplest cases such as the ideal gas (Section 4) or the ideal crystal (Section 5). In all the other cases, real gases and liquids considered here, it can be determined only approximatively.

### 3.5 Probability and entropy

A relation between entropy  $S$  and probabilities  $P_i$  of quantum states of a system can be proved in the canonical ensemble

$$S = -k_B \sum_i P_i \ln P_i. \quad (20)$$

For the microcanonical ensemble a similar relation holds

$$S = k_B \ln W, \quad (21)$$

where  $W$  is a number of accessible states. This equation (with log instead of ln) is written in the grave of Ludwig Boltzmann in Central Cemetery in Vienna, Austria.

## 4. Ideal gas

The ideal gas is in statistical thermodynamics modelled by a assembly of particles that do not mutually interact. Then the energy of  $i$ -th quantum state of system,  $E_i$ , is a sum of energies of individual particles

$$E_i = \sum_{j=1}^N \epsilon_{i,j}. \quad (22)$$

In this way a problem of a determination of the partition function of system is dramatically simplified. For one-component system of  $N$  molecules it holds

$$Q = \frac{q^N}{N!}, \quad (23)$$

where

$$q = \sum_j \exp(-\beta \epsilon_j) \quad (24)$$

is the partition function of molecule.

The partition function of molecule can be further simplified. The energy of molecule can be approximated by a sum of the translational  $\epsilon_{\text{trans}}$ , the rotational  $\epsilon_{\text{rot}}$ , the vibrational  $\epsilon_{\text{vib}}$ , and the electronic  $\epsilon_{\text{el}}$  contributions (subscript  $j$  in  $\epsilon_j$  is omitted for simplicity of notation)

$$\epsilon = \epsilon_0 + \epsilon_{\text{trans}} + \epsilon_{\text{rot}} + \epsilon_{\text{vib}} + \epsilon_{\text{el}}, \quad (25)$$

where  $\epsilon_0$  is the zero point energy. The partition function of system then becomes a product

$$Q = \frac{\exp(-N\beta\epsilon_0)}{N!} q_{\text{trans}} q_{\text{rot}} q_{\text{vib}} q_{\text{el}}. \quad (26)$$

Consequently all thermodynamic quantities of the ideal gas become sums of the corresponding contributions. For example the Helmholtz free energy is

$$\begin{aligned} A &= -k_B T \ln Q \\ &= k_B T \ln N! + U_0 - Nk_B T \ln q_{\text{tr}} - Nk_B T \ln q_{\text{rot}} - Nk_B T \ln q_{\text{vib}} - Nk_B T \ln q_{\text{el}} \\ &= k_B T \ln N! + U_0 + A_{\text{tr}} + A_{\text{rot}} + A_{\text{vib}} + A_{\text{el}}, \end{aligned} \quad (27)$$

where  $U_0 = N\epsilon_0$  and  $A_{\text{tr}}, A_{\text{rot}}, A_{\text{vib}}, A_{\text{el}}$  are the translational, rotational, vibrational, electronic contributions to the Helmholtz free energy, respectively.

### 4.1 Translational contributions

Translational motions of a molecule are modelled by a particle in a box. For its energy a solution of the Schrödinger equation gives

$$\epsilon_{\text{tr}} = \frac{h^2}{8m} \left( \frac{n_x^2}{a^2} + \frac{n_y^2}{b^2} + \frac{n_z^2}{c^2} \right), \quad (28)$$

where  $h$  is the Planck constant,  $m$  mass of molecule, and  $abc = V$  where  $V$  is volume of system. Quantities  $n_x, n_y, n_z$  are the quantum numbers of translation. The partition function of translation is

$$q_{\text{tr}} = \left( \frac{2\pi mk_B T}{h^2} \right)^{3/2} V. \quad (29)$$

Translational contribution to the Helmholtz energy is

$$A_{\text{tr}} = -RT \ln q_{\text{tr}} = -RT \ln (\lambda^{-3} V), \quad (30)$$

where  $R = Nk_B$  is the gas constant and  $\lambda = h/\sqrt{2\pi mk_B T}$  is the Broglie wavelength. The remaining thermodynamic functions are as follows

$$S_{\text{tr}} = - \left( \frac{\partial A_{\text{tr}}}{\partial T} \right)_V = R \ln (\lambda^{-3} V) + \frac{3}{2} R, \quad (31)$$

$$p_{\text{tr}} = - \left( \frac{\partial A_{\text{tr}}}{\partial V} \right)_T = \frac{RT}{V}, \quad (32)$$

$$U_{\text{tr}} = A_{\text{tr}} + TS_{\text{tr}} = \frac{3}{2} RT, \quad (33)$$

$$H_{\text{tr}} = U_{\text{tr}} + p_{\text{tr}} V = \frac{5}{2} RT, \quad (34)$$

$$G_{\text{tr}} = A_{\text{tr}} + p_{\text{tr}} V = -RT \ln (\lambda^{-3} V) + RT, \quad (35)$$

$$C_{V,\text{tr}} = \left( \frac{\partial U_{\text{tr}}}{\partial T} \right)_V = \frac{3}{2} R, \quad (36)$$

$$C_{p,\text{tr}} = \left( \frac{\partial H_{\text{tr}}}{\partial T} \right)_p = \frac{5}{2} R. \quad (37)$$

### 4.2 Rotational contributions

Rotations of molecule are modelled by the rigid rotator. For linear molecules there are two independent axes of rotation, for non-linear molecules there are three.

#### 4.2.1 Linear molecules

For the partition function of rotation it holds

$$q_{\text{rot}} = \frac{8\pi I k_B T}{\sigma h^2}, \quad (38)$$

where  $\sigma$  is the symmetry number of molecule and  $I$  its moment of inertia

$$I = \sum_1^n m_i r_i^2,$$

with  $n$  a number of atoms in molecule,  $m_i$  their atomic masses and  $r_i$  their distances from the center of mass. Contributions to the thermodynamic quantities are

$$A_{\text{rot}} = -RT \ln q_{\text{rot}} = -RT \ln \left( \frac{8\pi^2 I k_B T}{\sigma h^2} \right), \quad (39)$$

$$S_{\text{rot}} = R \ln \left( \frac{8\pi^2 I k_B T}{\sigma h^2} \right) + R, \quad (40)$$

$$p_{\text{rot}} = 0, \quad (41)$$

$$U_{\text{rot}} = RT, \quad (42)$$

$$H_{\text{rot}} = U_{\text{rot}}, \quad (43)$$

$$G_{\text{rot}} = F_{\text{rot}}, \quad (44)$$

$$C_{V,\text{rot}} = R, \quad (45)$$

$$C_{p,\text{rot}} = C_{V,\text{rot}}. \quad (46)$$

#### 4.2.2 Non-linear molecules

The partition function of rotation of a non-linear molecule is

$$q_{\text{rot}} = \frac{1}{\sigma} \left( \frac{8\pi^2 k_B T}{h^2} \right)^{3/2} (\pi I_A I_B I_C)^{1/2}, \quad (47)$$

where  $I_A$ ,  $I_B$  and  $I_C$  the principal moments of inertia. Contributions to the thermodynamic quantities are

$$A_{\text{rot}} = -RT \ln q_{\text{rot}} = -RT \ln \left[ \frac{1}{\sigma} \left( \frac{8\pi^2 k_B T}{h^2} \right)^{3/2} (\pi I_A I_B I_C)^{1/2} \right], \quad (48)$$

$$S_{\text{rot}} = R \ln \left[ \frac{1}{\sigma} \left( \frac{8\pi^2 k_B T}{h^2} \right)^{3/2} (\pi I_A I_B I_C)^{1/2} \right] + \frac{3}{2} R, \quad (49)$$

$$p_{\text{rot}} = 0, \quad (50)$$

$$U_{\text{rot}} = \frac{3}{2} RT, \quad (51)$$

$$H_{\text{rot}} = U_{\text{rot}}, \quad (52)$$

$$G_{\text{rot}} = A_{\text{rot}}, \quad (53)$$

$$C_{V,\text{rot}} = \frac{3}{2} R, \quad (54)$$

$$C_{p,\text{rot}} = C_{V,\text{rot}}. \quad (55)$$

### 4.3 Vibrational contributions

Vibrations of atoms in molecule around their equilibrium states may be at not very high temperatures approximated by harmonic oscillators.

#### 4.3.1 Diatomic molecules

In a diatomic molecule there is only one vibrational motion. Its partition function is

$$q_{\text{vib}} = [1 - \exp(h\nu_0/k_B T)]^{-1}, \quad (56)$$

where  $\nu_0$  is the fundamental harmonic frequency. Vibrational contributions to thermodynamic quantities are

$$A_{\text{vib}} = -RT \ln q_{\text{vib}} = RT \ln (1 - e^{-x}) , \quad (57)$$

$$S_{\text{vib}} = R \frac{x e^{-x}}{1 - e^{-x}} - R \ln (1 - e^{-x}) , \quad (58)$$

$$p_{\text{vib}} = 0 , \quad (59)$$

$$U_{\text{vib}} = RT \frac{x e^{-x}}{1 - e^{-x}} , \quad (60)$$

$$H_{\text{vib}} = U_{\text{vib}} , \quad (61)$$

$$G_{\text{vib}} = A_{\text{vib}} , \quad (62)$$

$$C_{V,\text{vib}} = R \frac{x^2 e^{-x}}{(1 - e^{-x})^2} , \quad (63)$$

$$C_{p,\text{vib}} = C_{V,\text{vib}} , \quad (64)$$

where  $x = \frac{h\nu_0}{k_B T}$ .

#### 4.3.2 Polyatomic molecules

In  $n$ -atomic molecule there is  $f$  fundamental harmonic frequencies  $\nu_i$  where

$$f = \begin{cases} 3n - 5 & \text{linear molecule} \\ 3n - 6 & \text{non-linear molecule} \end{cases}$$

The partition function of vibration is

$$q_{\text{vib}} = \prod_{i=1}^f \frac{1}{1 - \exp(-h\nu_i/k_B T)} . \quad (65)$$

For the thermodynamic functions of vibration we get

$$A_{\text{vib}} = -RT \ln q_{\text{vib}} = RT \sum_{i=1}^f \ln (1 - e^{-x_i}) , \quad (66)$$

$$S_{\text{vib}} = R \sum_{i=1}^f \frac{x_i e^{-x_i}}{1 - e^{-x_i}} - R \sum_{i=1}^f \ln (1 - e^{-x_i}) , \quad (67)$$

$$p_{\text{vib}} = 0 , \quad (68)$$

$$U_{\text{vib}} = RT \sum_{i=1}^f \frac{x_i e^{-x_i}}{1 - e^{-x_i}} , \quad (69)$$

$$H_{\text{vib}} = U_{\text{vib}} , \quad (70)$$

$$G_{\text{vib}} = A_{\text{vib}} , \quad (71)$$

$$C_{V,\text{vib}} = R \sum_{i=1}^f \frac{x_i^2 e^{-x_i}}{(1 - e^{-x_i})^2} , \quad (72)$$

$$C_{p,\text{vib}} = C_{V,\text{vib}} , \quad (73)$$



where  $x_i = \frac{h\nu_i}{k_B T}$ .

#### 4.4 Electronic contributions

The electronic partition function reads

$$q_{\text{el}} = \sum_{\ell=0}^{\infty} g_{\text{el},\ell} e^{-\varepsilon_{\text{el},\ell}/k_B T}, \quad (74)$$

where  $\varepsilon_{\text{el},\ell}$  the energy level  $\ell$ , and  $g_{\text{el},\ell}$  is its degeneracy. In most cases the electronic contributions to the thermodynamic functions are negligible at not very high temperatures. Therefore they are not written here.

#### 4.5 Ideal gas mixture

Let us consider two-component mixture of  $N_1$  non-interacting molecules of component 1 and  $N_2$  non-interacting molecules of component 2 (extension to the case of a multi-component mixture is straightforward). The partition function of mixture is

$$Q = \frac{q_1^{N_1} q_2^{N_2}}{N_1! N_2!} \quad (75)$$

where  $q_1$  and  $q_2$  are the partition functions of molecules 1 and 2, respectively. Let us denote  $X_{m,i}$  the molar thermodynamic quantity of pure component  $i$ ,  $i = 1, 2$  and  $x_i = \frac{N_i}{N_1 + N_2}$  its mole fraction. Then

$$A = RT (x_1 \ln x_1 + x_2 \ln x_2) + x_1 A_{m,1} + x_2 A_{m,2} \quad (76)$$

$$S = -R (x_1 \ln x_1 + x_2 \ln x_2) + x_1 S_{m,1} + x_2 S_{m,2}, \quad (77)$$

$$G = RT (x_1 \ln x_1 + x_2 \ln x_2) + x_1 G_{m,1} + x_2 G_{m,2}, \quad (78)$$

$$p = x_1 p_{m,1} + x_2 p_{m,2} = x_1 \frac{RT}{V_m} + x_2 \frac{RT}{V_m}, \quad (79)$$

$$U = x_1 U_{m,1} + x_2 U_{m,2}, \quad (80)$$

$$H = x_1 H_{m,1} + x_2 H_{m,2}, \quad (81)$$

$$C_V = x_1 C_{V,m,1} + x_2 C_{V,m,2}, \quad (82)$$

$$C_p = x_1 C_{p,m,1} + x_2 C_{p,m,2}. \quad (83)$$

### 5. Ideal crystal

We will call the ideal crystal an assembly of molecules displayed in a regular lattice without any impurities or lattice deformations. Distances among lattice centers will not depend on temperature and pressure. For simplicity we will consider one-atomic molecules. The partition function of crystal is

$$Q = e^{-U_0/k_B T} Q_{\text{vib}}, \quad (84)$$

where  $U_0$  is the lattice energy.

We will discuss here two models of the ideal crystal: the Einstein approximation and the Debye approximation.

### 5.1 Einstein model

An older and simpler Einstein model is based on the following postulates

1. Vibrations of molecules are independent:

$$Q_{\text{vib}} = q_{\text{vib}}^N, \quad (85)$$

where  $q_{\text{vib}}$  is the vibrational partition function of molecule.

2. Vibrations are isotropic:

$$q_{\text{vib}} = q_x q_y q_z = q_x^3. \quad (86)$$

3. Vibrations are harmonical

$$q_x = \sum_{v=0}^{\infty} e^{-\epsilon_v/k_B T}, \quad (87)$$

where

$$\epsilon_v = h\nu \left( v + \frac{1}{2} \right)$$

is the energy in quantum state  $v$  and  $\nu$  is the fundamental vibrational frequency.

Combining these equations one obtains

$$Q = e^{-U_0/k_B T} \left( \frac{e^{-\Theta_E/(2T)}}{1 - e^{-\Theta_E/T}} \right)^{3N}, \quad (88)$$

where

$$\Theta_E = \frac{h\nu}{k_B}$$

is the *Einstein characteristic temperature*.

For the isochoric heat capacity it follows

$$C_V = 3Nk_B \left( \frac{\Theta_E}{T} \right)^2 \frac{e^{-\Theta_E/T}}{(1 - e^{-\Theta_E/T})^2}. \quad (89)$$

### 5.2 Debye model

Debye considers crystal as a huge molecule (*i.e.* he replaces the postulates of independence and isotropy in the Einstein model) of an ideal gas; the postulate of harmonicity of vibrations remains. From these assumptions it can be derived for the partition function

$$\ln Q = -\frac{U_0}{k_B T} - \frac{9}{8} N \frac{\Theta_D}{T} - 9N \left( \frac{T}{\Theta_D} \right)^3 \int_0^{\Theta_D/T} x^2 \ln(1 - e^{-x}) dx, \quad (90)$$

where

$$\Theta_D = \frac{h\nu_{\text{max}}}{k_B}$$

is the *Debye characteristic temperature* with  $\nu_{\text{max}}$  being the highest frequency of crystal.

For the isochoric heat capacity it follows

$$C_V = 3R \left( 4D(u) - \frac{3u}{e^u - 1} \right). \quad (91)$$

where  $u = \Theta_D/T$  and

$$D(u) = \frac{3}{u^3} \int_0^u \frac{x^3}{e^x - 1} dx.$$

It can be proved that at low temperatures the heat capacity becomes a cubic function of temperature

$$C_V = 36R \left( \frac{T}{\Theta_D} \right)^3 \int_0^\infty \frac{x^3}{e^x - 1} dx = a T^3,$$

while the Einstein model incorrectly gives

$$C_V = 3R \left( \frac{\Theta_E}{T} \right)^2 e^{-\Theta_E/T}.$$

Both models give a correct high-temperature limit (the Dulong-Petit law)

$$C_V = 3R.$$

The same is true for the zero temperature limit

$$\lim_{T \rightarrow 0} C_V = 0.$$

### 5.3 Beyond the Debye model

Both the Einstein and the Debye models assume harmonicity of lattice vibrations. This is not true at high temperatures near the melting point. The harmonic vibrations are not assumed in the lattice theories (the cell theory, the hole theory, ...) that used to be popular in forties and fifties of the last century for liquids. It was shown later that they are poor theories of liquids but very good theories for solids.

Thermodynamic functions cannot be obtained analytically in the lattice theories.

## 6. Intermolecular forces

Up to now forces acting among molecules have been ignored. In the ideal gas (Section 4) molecules are assumed to exert no forces upon each other. In the ideal crystal (Section 5) molecules are imprisoned in the lattice, and the intermolecular forces are counted indirectly in the lattice energy and in the Einstein or Debye temperature. For real gases and liquids the intermolecular force must be included explicitly.

### 6.1 The configurational integral and the molecular interaction energy

The partition function of the real gas or liquid may be written in a form

$$Q = \frac{1}{N!} \exp(-N\beta\epsilon_0) q_{\text{int}}^N \left( \frac{2\pi mk_B T}{h^2} \right)^{\frac{3}{2}N} Z. \quad (92)$$

where  $q_{\text{int}} = q_{\text{rot}} q_{\text{vib}} q_{\text{el}}$  is the partition function of the internal motions in molecule. Quantity  $Z$  is the configurational integral

$$Z = \int_{(V)} \int_{(V)} \cdots \int_{(V)} \exp[-\beta u_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)] d\vec{r}_1 d\vec{r}_2 \dots d\vec{r}_N, \quad (93)$$

where symbol

$$\int_{(V)} \cdots d\vec{r}_i = \int_0^L \int_0^L \int_0^L \cdots dx_i dy_i dz_i \quad \text{and} \quad L^3 = V.$$

The quantity  $u_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$  is the potential energy of an assembly of  $N$  molecules. Here and in Eq.(93) one-atomic molecules are assumed for simplicity. More generally, the potential energy is a function not only positions of centers of molecules  $\vec{r}_i$  but also of their orientations  $\vec{\omega}_i$ . However, we will use the above simplified notation.

The interaction potential energy  $u_N$  of system may be written as an expansion in two-body, three-body, *e.t.c* contributions

$$u_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) = \sum_{i<j} u_2(\vec{r}_i, \vec{r}_j) + \sum_{i<j<k} u_3(\vec{r}_i, \vec{r}_j, \vec{r}_k) + \cdots \quad (94)$$

Most often only the first term is considered. This approximation is called *the rule of pairwise additivity*

$$u_N = \sum_{i<j} u_2(\vec{r}_i, \vec{r}_j), \quad (95)$$

where  $u_2$  is the pair intermolecular potential. The three-body potential  $u_3$  is used rarely at very accurate calculations, and  $u_4$  and higher order contributions are omitted as a rule.

## 6.2 The pair intermolecular potential

The pair potential depends of a distance between centers of two molecules  $r$  and on their mutual orientation  $\vec{\omega}$ . For simplicity we will omit the angular dependence of the pair potential (it is true for the spherically symmetric molecules) in further text, and write

$$u_2(\vec{r}_i, \vec{r}_j) = u_2(r_{ij}, \vec{\omega}_{ij}) = u(r)$$

where subscripts 2 and  $ij$  are omitted, too.

The following model pair potentials are most often used.

### 6.2.1 Hard spheres

It is after the ideal gas the simplest model. It ignores attractive interaction between molecules, and approximates strong repulsive interactions at low intermolecular distances by an infinite barrier

$$u(r) = \begin{cases} \infty & r < \sigma \\ 0 & r > \sigma \end{cases} \quad (96)$$

where  $\sigma$  is a diameter of molecule.

### 6.2.2 Square well potential

Molecules behave like hard spheres surrounded by an area of attraction

$$u(r) = \begin{cases} \infty & r < \sigma \\ -\epsilon & \sigma < r < \lambda\sigma \\ 0 & r > \lambda\sigma \end{cases} \quad (97)$$

Here  $\sigma$  is a hard-sphere diameter,  $\epsilon$  a depth of the attractive well, and the attraction region ranges from  $\sigma$  to  $\lambda\sigma$ .

### 6.2.3 Lennard-Jones potential

This well known pair intermolecular potential realistically describes a dependence of pair potential energy on distance

$$u(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]. \quad (98)$$

$\epsilon$  is a depth of potential at minimum, and  $2^{1/6}\sigma$  is its position.

More generally, the Lennard-Jones n-m potential is

$$u(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^n - \left( \frac{\sigma}{r} \right)^m \right]. \quad (99)$$

### 6.2.4 Pair potentials of non-spherical molecules

There are analogues of hard spheres for non-spherical particles: hard diatomics or dumbbells made of two fused hard spheres, hard triatomics, hard multiatomics, hard spherocylinders, hard ellipsoids, and so on.

Examples of soft pair potentials are Lennard-Jones multiatomics, molecules whose atoms interact according to the Lennard-Jones potential (98).

Another example is the Stockmayer potential, the Lennard-Jones potential with an indebted dipole moment

$$u(r, \theta_1, \theta_2, \phi) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] - \frac{\mu^2}{r^3} [2 \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 \cos \phi], \quad (100)$$

where  $\mu$  is the dipole moment.

### 6.2.5 Pair potentials of real molecules

The above model pair potentials, especially the Lennard-Jones potential and its extensions, may be used to calculate properties of the real substances. In this case their parameters, for example  $\epsilon$  and  $\sigma$ , are fitted to the experimental data such as the second virial coefficients, rare-gas transport properties and molecular properties.

More sophisticated approach involving a realistic dependence on the interparticle separation with a number of adjustable parameters was used by Aziz, see Aziz (1984) and references therein.

For simple molecules, there is a fully theoretical approach without any adjustable parameters utilizing the first principle quantum mechanics calculations, see for example Slaviček et al. (2003) and references therein.

## 6.3 The three-body potential

The three-body intermolecular interactions are caused by polarizabilities of molecules. The simplest and the most often used is the Axilrod-Teller-Muto term

$$u(r, s, t) = \frac{\nu}{rst} (3 \cos \theta_1 \cos \theta_2 \cos \theta_3 + 1), \quad (101)$$

where  $\nu$  is a strength parameter. It is a first term (DDD, dipole-dipole-dipole) in the multipole expansion. Analytical formulae and corresponding strength parameters are known for higher order terms (DDQ, dipole-dipole-quadrupole, DQQ, dipole-quadrupole-quadrupole,...) as well.

More accurate three-body potentials can be obtained using quantum chemical *ab initio* calculations Malijevský et al. (2007).

## 7. The virial equation of state

The virial equation of state in the statistical thermodynamics is an expansion of the compressibility factor  $z = \frac{pV}{RT}$  in powers of density  $\rho = \frac{N}{V}$

$$z = 1 + B_2\rho + B_3\rho^2 + \dots, \quad (102)$$

where  $B_2$  is the second virial coefficient,  $B_3$  the third, *et.c.* The virial coefficients of pure gases are functions of temperature only. For mixtures they are functions of temperature and composition.

The first term in equation (102) gives the equation of state of ideal gas, the first two terms or three give corrections to non-ideality. Higher virial coefficients are not available experimentally. However, they can be determined from knowledge of intermolecular forces. The relations among the intermolecular forces and the virial coefficients are exact, the pair and the three-body of potentials are subjects of uncertainties, however.

### 7.1 Second virial coefficient

For the second virial coefficient of spherically symmetric molecules we find

$$B = -2\pi \int_0^\infty f(r) r^2 dr = -2\pi \int_0^\infty (e^{-\beta u(r)} - 1) r^2 dr, \quad (103)$$

where

$$f(r) = \exp[-\beta u(r)] - 1$$

is the *Mayer function*. For linear molecules we have

$$B = -\frac{1}{4} \int_0^\infty \int_0^\pi \int_0^\pi \int_0^{2\pi} [e^{-\beta u(r, \theta_1, \theta_2, \phi)} - 1] r^2 \sin \theta_1 \sin \theta_2 dr d\theta_1 d\theta_2 d\phi. \quad (104)$$

For general non-spherical molecules we obtain

$$B = -\frac{2\pi}{\int_{\vec{\omega}_1} \int_{\vec{\omega}_2} d\vec{\omega}_1 d\vec{\omega}_2} \int_0^\infty \int_{\omega_1} \int_{\omega_2} [e^{-\beta u(r, \vec{\omega}_1, \vec{\omega}_2)} - 1] r^2 dr d\vec{\omega}_1 d\vec{\omega}_2. \quad (105)$$

### 7.2 Third virial coefficient

The third virial coefficient may be written for spherically symmetric molecules as

$$C = C_{\text{add}} + C_{\text{nadd}}, \quad (106)$$

where

$$C_{\text{add}} = -\frac{8}{3}\pi^2 \int_0^\infty \int_0^r \int_{|r-s|}^{r+s} (e^{-\beta u(r)} - 1) (e^{-\beta u(s)} - 1) (e^{-\beta u(t)} - 1) r s t dr ds dt, \quad (107)$$

and

$$C_{\text{nadd}} = \frac{8}{3}\pi^2 \int_0^\infty \int_0^r \int_{|r-s|}^{r+s} e^{-\beta u(r)} e^{-\beta u(s)} e^{-\beta u(t)} \{ \exp[-\beta u_3(r, s, t)] - 1 \} r s t dr ds dt, \quad (108)$$

where  $u_3(r, s, t)$  is the three-body potential. Analogous equations hold for non-spherical molecules.

### 7.3 Higher virial coefficients

Expressions for higher virial coefficients become more and more complicated due to an increasing dimensionality of the corresponding integrals and their number. For example, the ninth virial coefficient consists of 194 066 integrals with the Mayer integrands, and their dimensionalities are up to 21 Malijevský & Kolafa (2008) in a simplest case of spherically symmetric molecules. For hard spheres the virial coefficients are known up to ten, which is at the edge of a present computer technology Labík et al. (2005).

### 7.4 Virial coefficients of mixtures

For binary mixture of components 1 and 2 the second virial coefficient reads

$$B_2 = x_1^2 B_2(11) + 2x_1 x_2 B_2(12) + x_2^2 B_2(22), \quad (109)$$

where  $x_i$  are the mole fractions,  $B_2(ii)$  the second virial coefficients of pure components and  $B_2(12)$  the crossed virial coefficient representing an influence of the interaction between molecule 1 and molecule 2.

The third virial coefficient reads

$$B_3 = x_1^3 B_3(111) + 3x_1^2 x_2 B_3(112) + 3x_1 x_2^2 B_3(122) + x_2^3 B_3(222). \quad (110)$$

Extensions of these equations on multicomponent mixtures and higher virial coefficients is straightforward.

## 8. Dense gas and liquid

Determination of thermodynamic properties from intermolecular interactions is much more difficult for dense fluids (for gases at high densities and for liquids) than for rare gases and solids. This fact can be explained using a definition of the Helmholtz free energy

$$A = U - TS. \quad (111)$$

Free energy has a minimum in equilibrium at constant temperature and volume. At high temperatures and low densities the term  $TS$  dominates because not only temperature but also entropy is high. A minimum in  $A$  corresponds to a maximum in  $S$  and system, thus, is in the gas phase. Ideal gas properties may be calculated from a behavior of individual molecules only. At somewhat higher densities thermodynamic quantities can be expanded from their ideal-gas values using the virial expansion.

At low temperatures the energy term in equation (111) dominates because not only temperature but also entropy is small. For solids we may start from a concept of the ideal crystal.

No such simple molecular model as the ideal gas or the ideal crystal is known for liquid and dense gas. Theoretical studies of liquid properties are difficult and uncompleted up to now.

### 8.1 Internal structure of fluid

There is no internal structure of molecules in the ideal gas. There is a long-range order in the crystal. The fluid is between of the two extremal cases: it has a local order at short intermolecular distances (as crystal) and a long-range disorder (as gas).

The fundamental quantity describing the internal structure of fluid is the pair distribution function  $g(r)$

$$g(r) = \frac{\rho(r)}{\rho}, \quad (112)$$

where  $\rho(r)$  is local density at distance  $r$  from the center of a given molecule, and  $\rho$  is the average or macroscopic density of system. Here and in the next pages of this section we assume spherically symmetric interactions and the rule of the pair additivity of the intermolecular potential energy.

The pair distribution function may be written in terms of the intermolecular interaction energy  $u_N$

$$g(r) = V^2 \frac{\int_{(V)} \cdots \int_{(V)} e^{-\beta u_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)} d\vec{r}_3 \dots d\vec{r}_N}{\int_{(V)} \cdots \int_{(V)} e^{-\beta u_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)} d\vec{r}_1 \dots d\vec{r}_N}. \quad (113)$$

It is related to the thermodynamic quantities using the pressure equation

$$z \equiv \frac{pV}{RT} = 1 - \frac{2}{3} \pi \rho \beta \int_0^\infty \frac{du(r)}{dr} g(r) r^3 dr, \quad (114)$$

the energy equation

$$\frac{U}{RT} = \frac{U^0}{RT} + 2\pi\rho\beta \int_0^\infty u(r)g(r)r^2 dr, \quad (115)$$

where  $U^0$  internal energy if the ideal gas, and the compressibility equation

$$\beta \left( \frac{\partial p}{\partial \rho} \right)_\beta = \left\{ 1 + 4\pi\rho \int_0^\infty [g(r) - 1] r^2 dr \right\}^{-1}. \quad (116)$$

Present mainstream theories of liquids can be divided into two large groups: perturbation theories and integral equation theories Hansen & McDonald (2006), Martynov (1992).

## 8.2 Perturbation theories

A starting point of the perturbation theories is a separation of the intermolecular potential into two parts: a harsh, short-range repulsion and a smoothly varying long-range attraction

$$u(r) = u^0(r) + u^p(r). \quad (117)$$

The term  $u^0(r)$  is called *the reference potential* and the term  $u^p(r)$  *the perturbation potential*. In the simplest case of the first order expansion of the Helmholtz free energy in the perturbation potential it holds

$$\frac{A}{RT} = \frac{A^0}{RT} + 2\pi\rho\beta \int_0^\infty u^p(r)g^0(r)r^2 dr, \quad (118)$$

where  $A^0$  is the Helmholtz free energy of a reference system.

In the perturbation theories knowledge of the pair distribution function and the Helmholtz free energy of the reference system is supposed. On one hand the reference system should be simple (the ideal gas is too simple and brings nothing new; a typical reference system is a fluid of hard spheres), and the perturbation potential should be small on the other hand. As a result of a battle between a simplicity of the reference potential (one must know its structural and thermodynamic properties) and an accuracy of a truncated expansion, a number of methods have been developed.



### 8.3 Integral equation theories

Among the integral equation theories the most popular are those based on the Ornstein-Zernike equation

$$h(r_{12}) = c(r_{12}) + \rho \int_{(V)} h(r_{13})c(r_{32}) d\vec{r}_3. \quad (119)$$

where  $h(r) = g(r) - 1$  is the total correlation function and  $c(r)$  the direct correlation function. This equation must be closed using a relation between the total and the direct correlation functions called *the closure* to the Ornstein-Zernike equation. From the diagrammatic analysis it follows

$$h(r) = \exp[-\beta u(r) + \gamma(r) + B(r)] - 1, \quad (120)$$

where

$$\gamma(r) = h(r) - c(r)$$

is the indirect (chain) correlation function and  $B(r)$  is the bridge function, a sum of elementary diagrams. Equation (120) does not yet provide a closure. It must be completed by an approximation for the bridge function. The mostly used closures are in listed in Malijevský & Kolafa (2008). The simplest of them are the hypernetted chain approximation

$$B(r) = 0 \quad (121)$$

and the Percus-Yevick approximation

$$B(r) = \gamma(r) - \ln[\gamma(r) + 1]. \quad (122)$$

Let us compare the perturbation and the integral equation theories. The first ones are simpler but they need an extra input - the structural and thermodynamic properties of a reference system. The accuracy of the second ones depends on a chosen closure. Their examples shown here, the hypernetted chain and the Percus-Yevick, are too simple to be accurate.

### 8.4 Computer simulations

Besides the above theoretical approaches there is another route to the thermodynamic quantities called the computer experiments or pseudoexperiments or simply simulations. For a given pair intermolecular potential they provide values of thermodynamic functions in the dependence on the state variables. In this sense they have characteristics of real experiments. Similarly to them they do not give an explanation of the bulk behavior of matter but they serve as tests of approximative theories. The thermodynamic values are free of approximations, or more precisely, their approximations such as a finite number of molecules in the basic box or a finite number of generated configurations can be systematically improved Kolafa et al. (2002). The computer simulations are divided into two groups: the Monte Carlo simulations and the molecular dynamic simulations. The Monte Carlo simulations generate the ensemble averages of structural and thermodynamic functions while the molecular dynamics simulations generate their time averages. The methods are described in detail in the monograph of Allen and Tildesley Allen & Tildesley (1987).

## 9. Interpretation of thermodynamic laws

In Section 2 the axioms of the classical or phenomenological thermodynamics have been listed. The statistical thermodynamics not only determines the thermodynamic quantities from knowledge of the intermolecular forces but also allows an interpretation of the phenomenological axioms.

### 9.1 Axiom on existence of the thermodynamic equilibrium

This axiom can be explained as follows. There is a very, very large number of microscopic states that correspond to a given macroscopic state. At unchained macroscopic parameters such as volume and temperature of a closed system there is much more equilibrium states than the states out of equilibrium. Consequently, a spontaneous transfer from non-equilibrium to equilibrium has a very, very high probability. However, a spontaneous transfer from an equilibrium state to a non-equilibrium state is not excluded.

Imagine a glass of whisky on rocks. This two-phase system at a room temperature transfers spontaneously to the one-phase system - a solution of water, ethanol and other components. It is not excluded but it is highly improbable that a glass with a dissolved ice will return to the initial state.

### 9.2 Axiom of additivity

This axiom postulates that the internal energy of the macroscopic system is a sum of its two macroscopic parts

$$U = U_1 + U_2. \quad (123)$$

However, if we consider a macroscopic system as an assembly of molecules, equation (123) does not take into account intermolecular interactions among molecules of subsystem 1 and molecules of subsystem 2. Correctly the equation should be

$$U = U_1 + U_2 + U_{12}. \quad (124)$$

Due to the fact that intermolecular interactions vanish at distances of the order of a few molecule diameters, the term  $U_{12}$  is negligible in comparison with  $U$ .

### 9.3 The zeroth law of thermodynamics and the negative absolute temperatures

The statistical thermodynamics introduces temperature formally as parameter  $\beta = \frac{1}{k_B T}$  in the expression (11) for the partition function

$$Q = \sum_i \exp(-\beta E_i).$$

As energies of molecular systems are positive and unbounded, temperature must be positive otherwise the equation diverges. For systems with bounded energies

$$E_{\min} \leq E_i \leq E_{\max}$$

both negative and positive temperatures are allowed. Such systems are in lasers, for example.

### 9.4 The second law of thermodynamics

From equation (3) it follows that entropy of the adiabatically isolated system either grows for spontaneous processes or remains constant in equilibrium

$$dS \geq 0. \quad (125)$$

Entropy in the statistical thermodynamics is connected with probability via equation (20)

$$S = -k_B \sum_i P_i \ln P_i.$$

Thus, entropy may spontaneously decrease but with a low probability.

### 9.5 Statistical thermodynamics and the arrow of time

Direction of time from past to future is supported by three arguments

- **Cosmological time**

The cosmological time goes according the standard model of Universe from the Big Bang to future.

- **Psychological time**

We as human beings remember (as a rule) what was yesterday but we do not "remember" what will be tomorrow.

- **Thermodynamic time**

Time goes in the direction of the growth of entropy in the direction given by equation (125).

The statistical thermodynamics allows due to its probabilistic nature a change of a direction of time "**from coffin to the cradle**" but again with a very, very low probability.

### 9.6 The third law of thermodynamics

Within the statistical thermodynamics the third law may be easily derived from equation (20) relating entropy and the probabilities. The state of the ideal crystal at  $T = 0$  K is one. Its probability  $P_0 = 1$ . By substituting to the equation we get  $S = 0$ .

## 10. Acknowledgement

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic under the project No. 604 613 7316.

#### List of symbols

A	Helmholtz free energy
B	second virial coefficient
$B_i$	i-th virial coefficient
$B(r)$	bridge function
$C_V$	isochoric heat capacity
$C_p$	isobaric capacity
$c(r)$	direct correlation function
E	energy
$\epsilon$	energy of molecule
G	Gibbs free energy
$g(r)$	pair distribution function
H	enthalpy
h	Planck constant
$h(r)$	total correlation function
$k_B$	Boltzmann constant
N	number of molecules, Avogadro number
P	probability
p	pressure
Q	heat
Q	partition function
q	partition function of molecule
R	(universal) gas constant (8.314 in SI units)
S	entropy
T	temperature

$\tau$	time
U	internal energy
$u_N$	potential energy of N particles
u	pair potential
W	work
W	number of accessible states
X	measurable thermodynamic quantity
x	mole fraction

## 11. References

- Feynman, R. P.; Leghton, R. B. & Sands, M. (2006). *The Feynman lectures on physics*, Vol. 1, Pearson, ISBN 0-8053-9046-4, San Francisco.
- Lucas K. (1991). *Applied Statistical Thermodynamics*, Springer-Verlag, ISBN 0-387-52007-4, New York, Berlin, Heidelberg.
- Aziz, R.A. (1984). Interatomic Potentials for Rare-Gases: Pure and Mixed Interactions, In: *Inert Gases* M. L. Klein (Ed.), 5 - 86, Springer-Verlag, ISBN 3-540-13128-0, Berlin, Heidelberg.
- Slaviček, P.; Kalus, R.; Paška, P.; Odvárková, I.; Hobza, P. & Malijevský, A. (2003). State-of-the-art correlated ab initio potential energy curves for heavy rare gas dimers: Ar<sub>2</sub>, Kr<sub>2</sub>, and Xe<sub>2</sub>. *Journal of Chemical Physics*, 119, 4, 2102-2119, ISSN:0021-9606.
- Malijevský, Alexandr; Karlický, F.; Kalus, R. & Malijevský, A. (2007). Third Virial Coefficients of Argon from First Principles. *Journal of Physical Chemistry C*, 111, 43, 15565 - 15568, ISSN:1932-7447.
- Malijevský, A. & Kolafa, J. (2008). Introduction to the thermodynamics of Hard Spheres and Related Systems In: *Theory and Simulation of Hard-Sphere Fluids and Related Systems* A. Mulero (Ed.) 27 - 36, Springer-Verlag, ISBN 978-3-540-78766-2, Berlin, Heidelberg.
- Labík, S.; Kolafa, J. & Malijevský, A. (2005). Virial coefficients of hard spheres and hard disks up to the ninth. *Physical Review E*, 71, 2-1, 021105/1-021105/8, ISSN:1539-3755.
- Allen, M. P. & Tildesley, D. J. (1987). *Computer Simulation of Liquids*, Clarendon Press, ISBN: 0-19-855645-4, Oxford.
- Hansen, J.-P. & McDonald, I. R. (2006). *Theory of Simple Fluids*, Elsevier, ISBN: 978-0-12-370535-8, Amsterdam.
- Martynov, G. A. (1992). *Fundamental Theory of Liquids*, Adam Hilger, ISBN: 0-7503-0069-8, Bristol, Philadelphia and New York.
- Malijevský, A. & Kolafa, J. (2008). Structure of Hard Spheres and Related Systems In: *Theory and Simulation of Hard-Sphere Fluids and Related Systems* A. Mulero (Ed.) 1 - 26, Springer-Verlag, ISBN 978-3-540-78766-2, Berlin, Heidelberg.
- Kolafa, J.; Labík & Malijevský, A. (2002). The bridge function of hard spheres by direct inversion of computer simulation data. *Molecular Physics*, 100, 16, 2629 - 2640, ISSN: 0026-8976.
- Baus, M. & Tejero, C., F. (2008). *Equilibrium Statistical Physics: Phases of Matter and Phase Transitions*, Springer, ISBN: 978-3-540-74631-7, Heidelberg.
- Ben-Naim, A. (2010). *Statistical Thermodynamics Based on Information: A Farewell to Entropy*, World-Scientific, ISBN: 978-981-270-707-9, Philadelphia, New York.
- Plischke, M. & Bergesen, B. (2006). *Equilibrium Statistical Physics*, University of British Columbia, Canada, ISBN: 978-981-256-048-3.



temperatures is favored for this type of adsorption, it increases with the increase of temperatures. For example, materials that contain silica aluminates or calcium oxide such as silica sand, kaolinite, bauxite, limestone, and aluminum oxide, were used as sorbents to capture heavy metals at high temperatures. The adsorption efficiency of the sorbents are influenced by operating temperature [2-7]. Usually, the removal of the chemisorbed species from the surface may be possible only under extreme conditions of temperature or high vacuum, or by some suitable chemical treatment of the surface. In deed, the chemisorption process depends on the surface area [8]. It too increases with an increase of surface area because the adsorbed molecules are linked to the surface by valence bonds. Normally, the chemi-adsorbed material forms a layer over the surface, which is only one chemisorbed molecule thick, *i.e.* they will usually occupy certain *adsorption sites* on the surface, and the molecules are not considered free to move from one surface site to another [9]. When the surface is covered by the monomolecular layer (monolayer adsorption), the capacity of the adsorbent is essentially exhausted. In addition, this type of adsorption is irreversible [10], wherein the chemical nature of the adsorbent(s) may be altered by the surface dissociation or reaction in which the original species cannot be recovered *via* desorption process [11]. In general, the adsorption isotherms indicated two distinct types of adsorption—reversible (composed of both physisorption and weak chemisorption) and irreversible (strongly chemisorbed) [10-11].

On the other hand *Physisorption* is a physical adsorption involving intermolecular forces (Van der Waals forces), which do not involve a significant change in the electronic orbital patterns of the species [12]. The energy of interaction between the adsorbate and adsorbent has the same order of magnitude as, but is usually greater than the energy of condensation of the adsorptive. Therefore, no activation energy is needed. In this case, low temperature is favourable for the adsorption. Therefore, the *physisorption* decreases with increase temperatures [13]. In physical adsorption, equilibrium is established between the adsorbate and the fluid phase resulting multilayer adsorption. Physical adsorption is relatively non specific due to the operation of weak forces of attraction between molecules. The adsorbed molecule is not affixed to a particular site on the solid surface, but is free to move about over the surface. Physical adsorption is generally is reversible in nature; *i.e.*, with a decrease in concentration the material is desorbed to the same extent that it was originally adsorbed [14]. In this case, the adsorbed species are chemically identical with those in the fluid phase, so that the chemical nature of the fluid is not altered by adsorption and subsequent desorption; as result, it is not specific in nature. In addition, the adsorbed material may condense and form several superimposed layers on the surface of the adsorbent [15].

Some times, both physisorption and chemisorption may occur on the surface at the same time, a layer of molecules may be physically adsorbed on a top of an underlying chemisorbed layer [16].

In summary, based on the different reversibility and specific of physical and chemical adsorption processes, thermal desorption of the adsorbed sorbent could provide important information for the study of adsorption mechanism.

## 2. Factors affecting adsorption

In general, the adsorption reaction is known to proceed through the following three steps [16]:

1. Transfer of adsorbate from bulk solution to adsorbent surface, which is usually mentioned as diffusion.
2. Migration of adsorbate ( $\text{Fe}^{3+}$  ion, where its ionic radius = 0.064 nm) into pores.
3. Interaction of  $\text{Fe}^{3+}$  ion with available sites on the interior surface of pores.

From the previous studies, it was shown that the rate-determining step for the adsorption of  $\text{Fe}^{3+}$  ion is step (3).

Normally, the driving force for the adsorption process is the concentration difference between the adsorbate in the solution at any time and the adsorbate in the solution at equilibrium ( $C-C_e$ ) [17]. but, there are some important factors affecting adsorption, such as the factors affecting the adsorption of  $\text{Fe}^{3+}$  ions in the aqueous solution:

### 2.1 Surface area of adsorbent

Larger surface area imply a greater adsorption capacity, for example, carbon and activated carbon [18].

### 2.2 Particle size of adsorbent

Smaller particle sizes reduce internal diffusion and mass transfer limitation to penetrate of the adsorbate inside the adsorbent (i.e., equilibrium is more easily achieved and nearly full adsorption capability can be attained). Figure 2 represents the removal efficiency  $\text{Fe}^{3+}$  ions by natural zeolite through three different particle sizes (45, 125 and 250  $\mu\text{m}$ ). It can be observed that the maximum adsorption efficiency is achieved with particle size 45  $\mu\text{m}$ . This is due to the most of the internal surface of such particles might be utilized for the adsorption. The smaller particle size gives higher adsorption rates, in which the  $\text{Fe}^{3+}$  ion has short path to transfer inside zeolite pores structure of the small particle size [19].

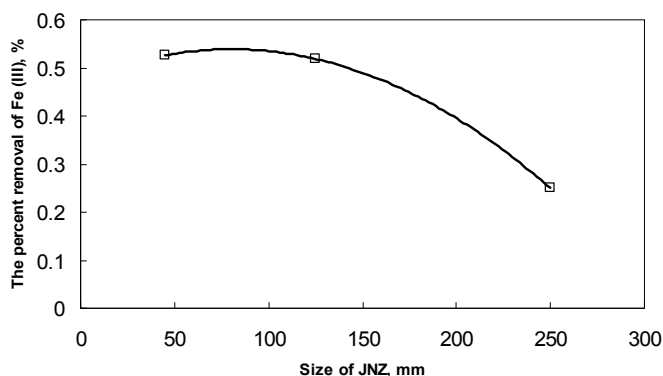


Fig. 2. Percent removal of  $\text{Fe}^{3+}$  ions (1000 ppm) vs. natural zeolite particle size: 1 g adsorbent/ 50 ml  $\text{Fe}^{3+}$  ion solution, initial pH of 1%  $\text{HNO}_3$ , and 300 rpm.

### 2.3 Contact time or residence time

The longer residence time means the more complete the adsorption will be. Therefore, the required contact time for sorption to be completed is important to give insight into a sorption process. This also provides an information on the minimum time required for considerable adsorption to take place, and also the possible diffusion control mechanism

between the adsorbate, for example  $\text{Fe}^{3+}$  ions, as it moves from the bulk solution towards the adsorbent surface, for example natural zeolite [19].

For example, the effect of contact time on sorption of  $\text{Fe}^{3+}$  ions is shown in Figure 3. At the initial stage, the rate of removal of  $\text{Fe}^{3+}$  ions using natural quartz (NQ) and natural bentonite (NB) is higher with uncontrolled rate. The initial faster rate may be due to the availability of the uncovered surface area of the adsorbent such as NQ and NB initially [20]. This is because the adsorption kinetics depends on: (i) the surface area of the adsorbent, (ii) the nature and concentration of the surface groups (active sites), which are responsible for interaction with the  $\text{Fe}^{3+}$  ions. Therefore, the adsorption mechanism on both adsorbent has uncontrolled rate during the first 10 minutes, where the maximum adsorption is achieved. Afterward, at the later stages, there is no influence for increasing the contact time. This is due to the decreased or lesser number of active sites. Similar results have been shown in our results using zeolite and olive cake as well as other reported in literatures for the removal of dyes, organic acids and metal ions by various adsorbents [19, 21].

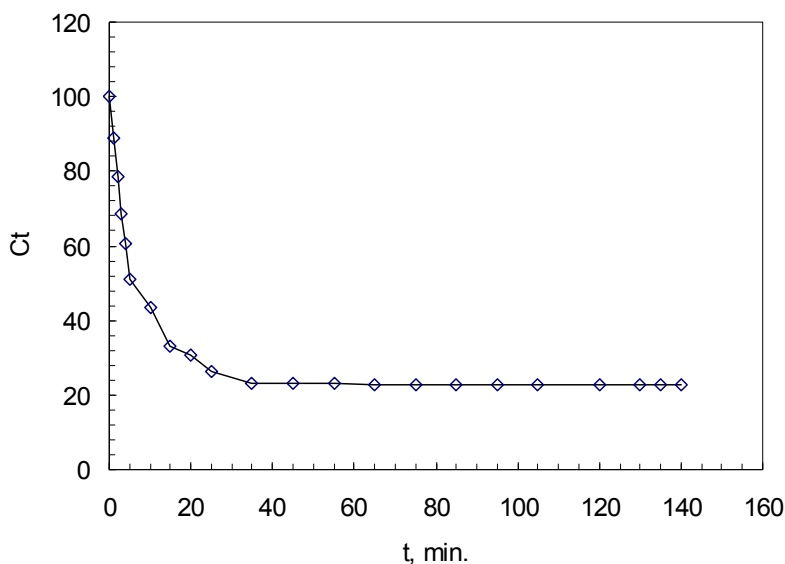


Fig. 3. Adsorption of  $\text{Fe}^{3+}$  ions onto olive cake. Variation of the  $\text{Fe}^{3+}$  ions concentration with time. (Initial concentration of  $\text{Fe}^{3+}$  ions: 100 ppm, Agitation speed: 100 rpm, pH: 4.5, temperature 28 °C).

#### 2.4 Solubility of adsorbent/ heavy metals in wastewater/ water

The slightly soluble metal ions in water will be more easily removed from water (i.e., adsorbed) than substances with high solubility. Also, non-polar substances will be more easily removed than polar substances since the latter have a greater affinity for adsorption.

#### 2.5 Affinity of the solute for the adsorbent

If the surface of adsorbent is slightly polar, the non-polar substances will be more easily picked up by the adsorbent than polar ones (the opposite is correct).



## 2.6 Size of the molecule with respect to size of the pores

Large molecules may be too large to enter small pores. This may reduce adsorption independently of other causes.

## 2.7 Degree of ionization of the adsorbate molecule

More highly ionized molecules are adsorbed to a smaller degree than neutral molecules.

## 2.8 pH

The degree of ionization of a species is affected by the pH (e.g., a weak acid or a weak basis). This, in turn, affects adsorption. For example, the precipitation of  $\text{Fe}^{3+}$  ions occurred at pH greater than 4.5 (see Figure 4). The decrease in the  $\text{Fe}^{3+}$  ions removal capacity at pH > 4.5 may have been caused by the complexing  $\text{Fe}^{3+}$  ions with hydroxide. Therefore, the removal efficiency increases with increasing initial pH. For example, at low pH, the concentration of proton is high. Therefore, the positively charged of the  $\text{Fe}^{3+}$  ions and the protons compete for binding on the adsorbent sites in Zeolite, Bentonite, Quartz, olive cake, Tripoli in which, this process decrease the uptake of iron ions. The concentration of proton in the solution decrease as pH gradually increases in the ranges from 2 to 4.5. In this case, little protons have the chance to compete with  $\text{Fe}^{3+}$  ions on the adsorption sites of the olive cake. Thus, higher pH in the acidic media is facilitated the greater uptake of  $\text{Fe}^{3+}$  ions. Above pH 4.5, the removal efficiency decreases as pH increases, this is inferred to be attributable to the hydrolysis [19-22].

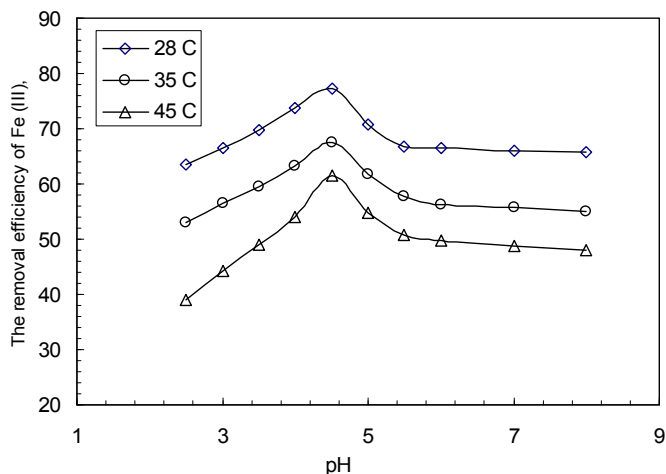


Fig. 4. Effect of initial pH on the removal efficiency, %, of  $\text{Fe}^{3+}$  ions at different temperatures. (Initial concentration of  $\text{Fe}^{3+}$  ions: 100 ppm, Agitation speed: 100 rpm, Mass of olive cake: 1 g, Dose: 5 g/l, Contact time: 24 hr).

## 2.9 Effect of initial concentration

At high-level concentrations, the available sites of adsorption become fewer. This behaviour is connected with the competitive diffusion process of the  $\text{Fe}^{3+}$  ions through the micro-

channel and pores in NB [20]. This competitive will lock the inlet of channel on the surface and prevents the metal ions to pass deeply inside the NB, *i.e.* the adsorption occurs on the surface only. These results indicate that energetically less favorable sites by increasing metal concentration in aqueous solution. This results are found matching with our recently studies using natural zeolite [19] and olive cake [21], in addition to other reported by Rao *et al.* [23] and Karthikeyan *et al.* [24]. The removal efficiency of  $\text{Fe}^{3+}$  ions on NQ and NB as well as zeolite at different initial concentrations (50, 100, 200, 300 and 400 ppm) is shown in the Figures 5-6. It is evident from the figure that the percentage removal of  $\text{Fe}^{3+}$  ions on NQ is slightly depended on the initial concentration. While the removal efficiency of  $\text{Fe}^{3+}$  ions using NB decreases as the initial concentration of  $\text{Fe}^{3+}$  ions increases. For example, the percentage removal is calculated 98 % using the initial concentration of 50 ppm, while it is found 28 % using high-level of 400 ppm [20].

On the other hand, it is clear from Figure 6 that the removal efficiency of  $\text{Fe}^{3+}$  ions using NQ is less affected by the initial concentration. For instance, the percentage removal using 50 ppm of the initial concentration is found 35 %, while is found 34.9 % using high-level concentrations (400 ppm). This means that the high concentration of  $\text{Fe}^{3+}$  ions will create and activate of some new activation sites on the adsorbent surface [20, 25].

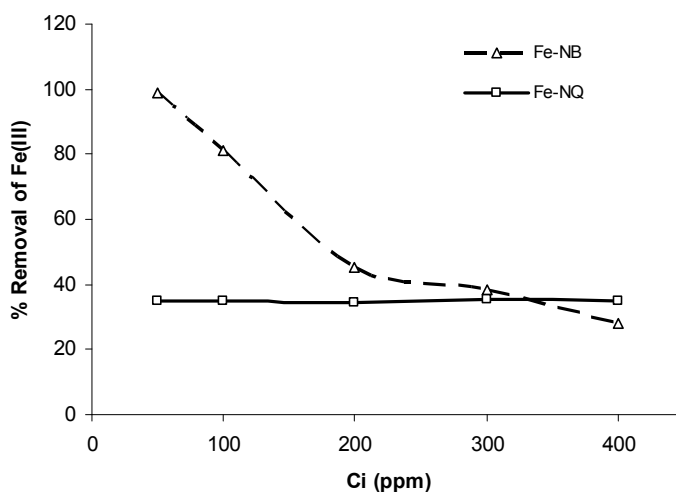


Fig. 5. The effect of initial concentration namely 50, 100, 200, 300 and 400 ppm of  $\text{Fe}^{3+}$  ions at constant contact time (2.5 hours), adsorbent dosage 4 g/L of natural NQ and NB, Temperature (30 °C) and agitation speed (300 rpm).

### 2.10 Dosage effect

The removal efficiency is generally increased as the concentration dose increases over these temperature values. This can be explained by the fact that more mass available, more the contact surface offered to the adsorption. The effect of the Jordanian Natural Zeolite (JNZ) dosage on the removal of  $\text{Fe}^{3+}$  ions is shown in Figure 6 [19]. The adsorbent dosage is varied from 10 to 40 g/l. The initial  $\text{Fe}^{3+}$  ions concentration, stirrer speed, initial pH and temperature are 1000 ppm, 300 rpm, 1%  $\text{HNO}_3$ , and 30 °C, respectively. This figure shows that the maximum removal of 69.15 % is observed with the dosage of 40 g/l. We observed

that the removal efficiency of adsorbents generally improved with increasing amount of JNZ. This is expected because the higher dose of adsorbent in the solution, the greater availability of exchangeable sites for the ions, *i.e.* more active sites are available for binding of  $\text{Fe}^{3+}$  ions. Moreover, our recent studies using olive cake, natural quartz and natural bentonite and tripoli [19-22] are qualitatively in a good agreement with each other and with those found in the literatures [26].

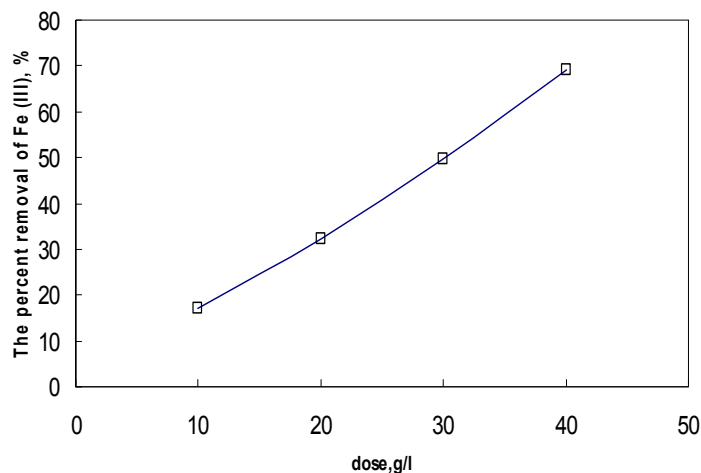


Fig. 6. The adsorbent dose of JNZ vs. Percent removal of  $\text{Fe}^{3+}$  ions: 1 g adsorbent/ 50 ml  $\text{Fe}^{3+}$  ions solution, 30 °C, initial pH of 1%  $\text{HNO}_3$ , 300 rpm, and constant initial concentration (1000 ppm).

### 3. Adsorption operation

Adsorption from solution is usually conducted using either the column or the batch operation. It should be possible to characterize the solution - adsorbent system by both technique operations and arrive at the same result. This is due to the physical and/or chemical forces applicable in each case must be identical. Furthermore, the results obtained from the batch experiment should be somewhat more reliable. Among the most serious objections of the column experiments are: (1) the inherent difficulties associated to maintain a constant flow rate; (2) the difficulty of ensuring a constant temperature throughout the column; (3) the appreciable probability of presence the channels within the packed column; and (4) the relatively large expenditure both in time and manpower required for a column experiment.

#### 3.1 Batch operation

In a batch operation, fixed amount of adsorbent is mixed all at once with specific volume of adsorbate (with the range of initial concentration). Afterwards, the system kept in agitation for a convenient period of time. Separation of the resultant solution is accomplished by filtering, centrifuging, or decanting. The optimum pH, contact time, agitation speed and optimum temperature are fixed and used in this technique. For instance, the contact time study, the experiment are carried out at constant initial concentration, agitation speed, pH, and temperature. During the adsorption progress, the mixture container must be covered by

alumina foil to avoid the evaporation. The samples are withdrawn at different time intervals, for example, every 5 minutes or every 15 minutes.

The uptake of heavy metal ions was calculated from the mass balance, which was stated as the amount of solute adsorbed onto the solid. It equal the amount of solute removed from the solution. Mathematically can be expressed in equation 1 [27]:

$$q_e = \frac{(C_i - C_e)}{S} \quad (1)$$

$q_e$ : Heavy metal ions concentration adsorbed on adsorbent at equilibrium (mg of metal ion/g of adsorbent).

$C_i$ : Initial concentration of metal ions in the solution (mg/l).

$C_e$ : Equilibrium concentration or final concentration of metal ions in the solution (mg/l).  $S$ : Dosage (slurry) concentration and it is expressed by equation 2:

$$S = \frac{m}{v} \quad (2)$$

Where  $v$  is the initial volume of metal ions solution used (L) and  $m$  is the mass of adsorbent. The percent adsorption (%) was also calculated using equation 3:

$$\% \text{ adsorption} = \frac{C_i - C_e}{C_i} \times 100\% \quad (3)$$

### 3.2 Column operation

In a column operation, the solution of adsorbate such as heavy metals (with the range of initial concentration) is allowed to percolate through a column containing adsorbent (ion exchange resin, silica, carbon, etc.) usually held in a vertical position. For instance, column studies were carried out in a column made of Pyrex glass of 1.5 cm internal diameter and 15 cm length. The column was filled with 1 g of dried PCA by tapping so that the maximum amount of adsorbent was packed without gaps. The influent solution was allowed to pass through the bed at constant flow rate of 2 mL/min, in down flow manner with the help of a fine metering valve. The effluent solution was collected at different time intervals.

The breakthrough adsorption capacity of adsorbate (heavy metal ions) was obtained in column at different cycles using the equation 4 [28].

$$q_e = [(C_i - C_e)/m] \times bv \quad (4)$$

Where  $C_i$  and  $C_e$  denote the initial and equilibrium (at breakthrough) of heavy metal ions concentration (mg/L) respectively.  $bv$  was the breakthrough volume of the heavy metal ions solution in liters, and  $m$  was the mass of the adsorbent used (g). After the column was exhausted, the column was drained off the remaining aqueous solution by pumping air. The adsorption percent is given by equation 5.

$$\% \text{ Desorption} = (C_e/C_i) \times 100 \quad (5)$$

## 4. Thermodynamic and adsorption isotherms

Adsorption isotherms or known as equilibrium data are the fundamental requirements for the design of adsorption systems. The equilibrium is achieved when the capacity of the

adsorbent materials is reached, and the rate of adsorption equals the rate of desorption. The theoretical adsorption capacity of an adsorbent can be calculated with an adsorption isotherm. There are basically two well established types of adsorption isotherm the Langmuir and the Freundlich adsorption isotherms. The significance of adsorption isotherms is that they show how the adsorbate molecules (metal ion in aqueous solution) are distributed between the solution and the adsorbent solids at equilibrium concentration on the loading capacity at different temperatures. That mean, the amount of sorbed solute versus the amount of solute in solution at equilibrium.

#### 4.1 Langmuir adsorption isotherm

Langmuir is the simplest type of theoretical isotherms. Langmuir adsorption isotherm describes quantitatively the formation of a monolayer of adsorbate on the outer surface of the adsorbent, and after that no further adsorption takes place. Thereby, the Langmuir represents the equilibrium distribution of metal ions between the solid and liquid phases [29]. The Langmuir adsorption is based on the view that every adsorption site is identical and energetically equivalent (thermodynamically, each site can hold one adsorbate molecule). The Langmuir isotherm assume that the ability of molecule to bind and adsorbed is independent of whether or not neighboring sites are occupied. This mean, there will be no interactions between adjacent molecules on the surface and immobile adsorption. Also mean, trans-migration of the adsorbate in the plane of the surface is precluded. In this case, the Langmuir isotherms is valid for the dynamic equilibrium adsorption – desorption processes on completely homogeneous surfaces with negligible interaction between adsorbed molecules that exhibit the form:

$$q_e = (Q \times b \times C_e) / (1 + b \times C_e) \quad (6)$$

$C_e$  = The equilibrium concentration in solution

$q_e$  = the amount adsorbed for unit mass of adsorbent

$Q$  and  $b$  are related to standard monolayer adsorption capacity and the Langmuir constant, respectively.

$$q_{\max} = Q \times b \quad (7)$$

$q_{\max}$  = is the constant related to overall solute adsorptivity (l/g).

Equation 6 could be re-written as:

$$C_e / q_e = 1 / (q_{\max} \times b) + (1 / q_{\max}) \times C_e \quad (8)$$

In summary, the Langmuir model represent one of the the first theoretical treatments of non-linear sorption and suggests that uptake occurs on a homogenous surface by monolayer sorption without interaction between adsorbed molecules. The Langmuir isotherm assumes that adsorption sites on the adsorbent surfaces are occupied by the adsorbate in the solution. Therefore the Langmuir constant ( $b$ ) represents the degree of adsorption affinity the adsorbate. The maximum adsorption capacity ( $Q$ ) associated with complete monolayer cover is typically expressed in (mg/g). High value of  $b$  indicates for much stronger affinity of metal ion adsorption.

The shape of the isotherm (assuming the (x) axis represents the concentration of adsorbing material in the contacting liquid) is a gradual positive curve that flattens to a constant value.

A plot of  $C_e/q_e$  versus  $C_e$  gives a straight line of slope  $1/q_{\max}$  and intercept  $1/(q_{\max} \times b)$ , for example, as shown in Figure 7.

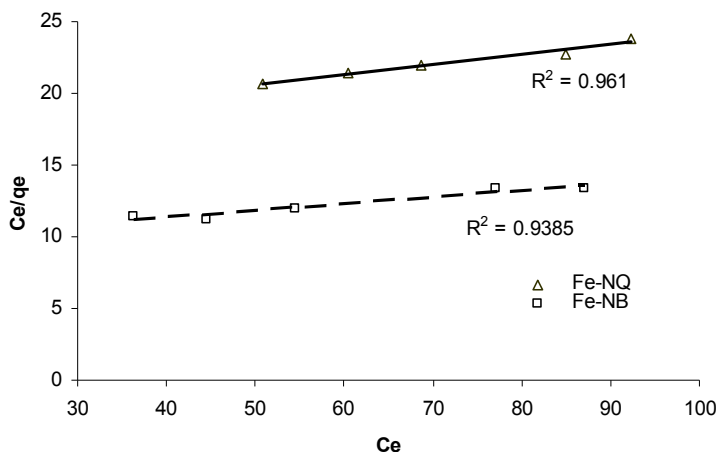


Fig. 7. The linearized Langmuir adsorption isotherms for  $\text{Fe}^{3+}$  ions adsorption by natural quartz (NQ) and bentonite (NB) at constant temperature 30 °C. (initial concentration: 400 ppm, 300 rpm and contact time: 2.5 hours).

The effect of isotherm shape is discussed from the direction of the predicting whether and adsorption system is "favorable" or "unfavorable". Hall et al (1966) proposed a dimensionless separation factor or equilibrium parameter,  $R_L$ , as an essential feature of the Langmuir Isotherm to predict if an adsorption system is "favourable" or "unfavourable", which is defined as [30]:

$$R_L = 1/(1+bC_i) \quad (9)$$

$C_i$  = reference fluid-phase concentration of adsorbate (mg/l) (initial  $\text{Fe}^{3+}$  ions concentration)  
 $b$  = Langmuir constant ( $\text{ml mg}^{-1}$ )

Value of  $R_L$  indicates the shape of the isotherm accordingly as shown in Table 1 below. For a single adsorption system,  $C_i$  is usually the highest fluid-phase concentration encountered.

Value of $R_L$	Type of Isotherm
$0 < r < 1$	Favorable
$r > 1$	Unfavorable
$r = 1$	Linear
$R = 0$	Irreversible

Table 1. Type of isotherm according to value of  $R_L$

#### 4.2 Freundlich adsorption isotherms

Freundlich isotherm is commonly used to describe the adsorption characteristics for the heterogeneous surface [31]. It represents an initial surface adsorption followed by a condensation effect resulting from strong adsorbate-adsorbate interaction. Freundlich

isotherm curves in the opposite way of Langmuir isotherm and is exponential in form. The heat of adsorption, in many instances, decreases in magnitude with increasing extent of adsorption. This decline in heat is logarithmic implying that the adsorption sites are distributed exponentially with respect to adsorption energy. This isotherm does not indicate an adsorption limit when coverage is sufficient to fill a monolayer ( $\theta = 1$ ). The equation that describes such isotherm is the Freundlich isotherm, given as [31]:

$$q_e = K_f (C_e)^{1/n} \quad n > 1 \quad (10)$$

$K_f$  = Freundlich constant related to maximum adsorption capacity (mg/g). It is a temperature-dependent constant.

$n$  = Freundlich constant related to surface heterogeneity (dimensionless). It gives an indication of how favorable the adsorption processes.

With  $n = 1$ , the equation reduces to the linear form:  $q_e = k \times C_e$

The plotting  $q_e$  versus  $C_e$  yield a non-regression line, which permits the determination of  $(1/n)$  and  $K_f$  values of  $(1/n)$  ranges from 0 to 1, where the closer value to zero means the more heterogeneous the adsorption surface. On linearization, these values can be obtained by plotting  $(\ln q_e)$  versus  $(\ln C_e)$  as presented in equation 11. From the plot, the values  $K_f$  and  $n$  can be obtained.

$$\ln q_e = \ln K_f + (1/n) \ln C_e \quad (11)$$

where, the slope =  $(1/n)$ , and the intercept =  $\ln K_f$

### 4.3 Dubinin–Kaganer–Radushkevich (DKR)

The DKR isotherm is reported to be more general than the Langmuir and Freundlich isotherms. It helps to determine the apparent energy of adsorption. The characteristic porosity of adsorbent toward the adsorbate and does not assume a homogenous surface or constant sorption potential [32].

The Dubinin–Kaganer–Radushkevich (DKR) model has the linear form

$$\ln q_e = \ln X_m - \beta \varepsilon^2 \quad (12)$$

where  $X_m$  is the maximum sorption capacity,  $\beta$  is the activity coefficient related to mean sorption energy, and  $\varepsilon$  is the Polanyi potential, which is equal to

$$\varepsilon = RT \ln \left( 1 + \frac{1}{C_e} \right) \quad (13)$$

where  $R$  is the gas constant (kJ/kmol- K) .

The slope of the plot of  $\ln q_e$  versus  $\varepsilon^2$  gives  $\beta$  ( $\text{mol}^2/\text{J}^2$ ) and the intercept yields the sorption capacity,  $X_m$  (mg/g) as shown in Fig. 6. The values of  $\beta$  and  $X_m$ , as a function of temperature are listed in table 1 with their corresponding value of the correlation coefficient,  $R^2$ . It can be observed that the values of  $\beta$  increase as temperature increases while the values of  $X_m$  decrease with increasing temperature.

The values of the adsorption energy,  $E$ , was obtained from the relationship [33]

$$E = (-2\beta)^{-1/2}$$

#### 4.4 Thermodynamics parameters for the adsorption

In order to fully understand the nature of adsorption the thermodynamic parameters such as free energy change ( $\Delta G^\circ$ ) and enthalpy change ( $\Delta H^\circ$ ) and entropy change ( $\Delta S^\circ$ ) could be calculated. It was possible to estimate these thermodynamic parameters for the adsorption reaction by considering the equilibrium constants under the several experimental conditions. They can be calculated using the following equations [34]:

$$\Delta G = -R \ln K_d(T) \quad (14)$$

$$\ln K_d = \Delta S/R - \Delta H/RT \quad (15)$$

$$\Delta G = \Delta H - T\Delta S \quad (16)$$

The  $K_d$  value is the adsorption coefficient obtained from Langmuir equation. It is equal to the ratio of the amount adsorbed ( $x/m$  in mg/g) to the adsorptive concentration ( $y/a$  in mg/dm<sup>3</sup>)

$$K_d = (x/m) \cdot (y/a) \quad (17)$$

These parameters are obtained from experiments at various temperatures using the previous equations. The values of  $\Delta H^\circ$  and  $\Delta S^\circ$  are determined from the slope and intercept of the linear plot of ( $\ln K_d$ ) vs. ( $1/T$ ).

In general these parameters indicate that the adsorption process is spontaneous or not and exothermic or endothermic. The standard enthalpy change ( $\Delta H^\circ$ ) for the adsorption process is: (i) positive value indicates that the process is endothermic in nature. (ii) negative value indicates that the process is exothermic in nature and a given amount of heat is evolved during the binding metal ion on the surface of adsorbent. This could be obtained from the plot of percent of adsorption (%  $C_{ads}$ ) vs. Temperature (T). The percent of adsorption increases with increase in temperature, this indicates for the endothermic processes and the opposite is correct [35]. The positive value of ( $\Delta S^\circ$ ) indicates an increase in the degree of freedom (or disorder) of the adsorbed species.

#### 5. Motivation for the removal and sorption Fe<sup>3+</sup> ions

In practice from recent studies, the natural zeolite [19], activated carbon [36], olive cake [21], quartz and bentonite [20] and jojoba seeds [37] are used as an adsorbent for the adsorption of mainly trivalent iron ions in aqueous solution. The motivation for the removal and sorption of Fe<sup>3+</sup> ions is that iron ions cause serious problems in the aqueous streams especially at high levels of concentration [38 - 39]. Usually, the iron ions dissolve from rocks and soils toward the ground water at low levels, but it can occur at high levels either through a certain geological formation or through the contamination by wastes effluent of the industrial processes such as pipeline corrosion, engine parts, metal finishing and galvanized pipe manufacturing [40 - 41].

The presence of iron at the high-levels in the aqueous streams makes the water unusable for several considerations: Firstly, aesthetic consideration such as discoloration, the metallic taste even at low concentration (1.8 mg/l), odor, and turbidity, staining of laundry and plumbing fixtures. Secondly, the health consideration where the high level of iron ions precipitates as an insoluble Fe<sup>3+</sup>-hydroxide under aerobic conditions at neutral or alkaline pH [42]. This can generate toxic derivatives within the body by using drinking



water and can contribute to disease development in several ways. For instance, an excessive amounts of iron ions in specific tissues and cells (iron-loading) promote development of infection, neoplasia, cardiomyopathy, arthropathy, and various endocrine and possibly neurodegenerative disorders. Finally, the industrial consideration such as blocking the pipes and increasing of corrosion. In addition to that, iron oxides promote the growth of micro-organism in water which inhibit many industrial processes in our country [43].

In response to the human body health, its environmental problems and the limitation water sources especially in Jordan [44], the high-levels of  $\text{Fe}^{3+}$  ions must be removed from the aqueous stream to the recommended limit 5.0 and 0.3 ppm for both inland surface and drinking water, respectively. These values are in agreement with the Jordanian standard parameters of water quality[45]. For tracing  $\text{Fe}^{3+}$  ions into recommended limit, many chemical and physical processes were used such as supercritical fluid extraction, bioremediation, oxidation with oxidizing agent [46]. These techniques were found not effective due to either extremely expensive or too inefficient to reduce such high levels of ions from the large volumes of water [47 - 48]. Therefore, the effective process must be low cost-effective technique and simple to operate [49 - 52]. It found that the adsorption process using natural adsorbents realize these prerequisites. In addition to that, the natural adsorbents are environmental friend, existent in a large quantities and has good adsorption properties. The binding of Iron(III) ion with the surface of the natural adsorbent could change their forms of existence in the environment. In general they may react with particular species, change oxidation states and precipitate [53]. In spite of the abundant reported researches in the adsorption for the removal of the dissolved heavy metals from the aqueous streams, however the iron(III) ions still has limited reported studies. Therefore our studies are concentrated in this field. From our previous work, the natural zeolite [19], quartz and Bentonite [20], olive cake [21], in addition to the chitin [24], activated carbon [54 - 55] and alumina [56] have been all utilized for this aspect at low levels. The adsorption isotherm models (Langmuir and Freundlich) are used in order to correlate the experimental results.

### 5.1 Sorption $\text{Fe}^{3+}$ ions using natural quartz (NQ) and bentonite (NB))

It is known from the chemistry view that surface of NB and NQ is ending with  $(\text{Si-O})^-$  negatively charged. These negative entities might bind metal ions *via* the coordination aspects especially at lower pH values as known in the literatures.  $\text{Fe}^{3+}$  ions are precipitated in the basic medium. Therefore, the 1 %  $\text{HNO}_3$  stock solution is used to soluble  $\text{Fe}^{3+}$  ions and then achieving the maximum adsorption percentages [20].

The binding of metal ions might be influenced on the surface of NB more than NQ. This is due to the expected of following ideas: (i) NQ have pure silica entity with homogeneous negatively charged, therefore the binding will be homogeneous. (ii) The natural bentonite has silica surface including an inner-layer of alumina and iron oxide, which cause a heterogeneous negatively charged. Therefore, the binding  $\text{Fe}^{3+}$  ions on the surface on NB might be complicated.

The adsorption thermodynamics modelsof  $\text{Fe}^{3+}$  ions on NQ and NB at 30 °C are examined [20]. The calculated results of the Langmuir and Freundlich isotherm constants are given in Table 2. The high values of  $R^2$  (>95%) indicates that the adsorption of  $\text{Fe}^{3+}$  ions onto both NQ and NB was well described by Freundlich isotherms. It can also be seen that the  $q_{\text{max}}$  and

the adsorption intensity values of NB are higher than that of NQ. The calculated  $b$  values indicate the interaction forces between NB surface and  $\text{Fe}^{3+}$  ions are stronger than in case of using NQ, this is in agreement with the higher ionic potential of  $\text{Fe}^{3+}$ . This means that the NB is more powerful adsorbent than NQ. Furthermore, based on this information, we found that the adsorption using NQ and NB is much higher as compared to carbon used by other authors [57].

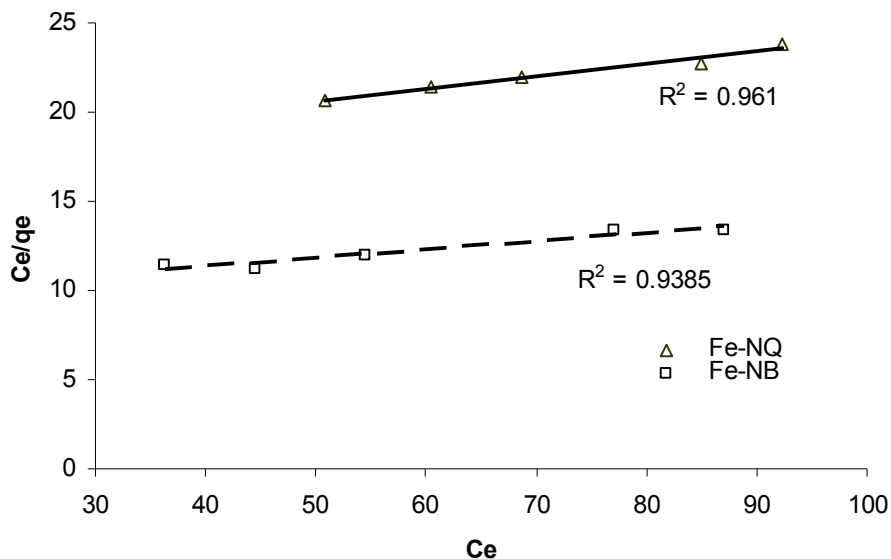


Fig. 8. The linearized Langmuir adsorption isotherms for  $\text{Fe}^{3+}$  ions adsorption by natural quartz (NQ) and bentonite (NB) at constant temperature 30 °C. (Initial concentration: 400 ppm, 300 rpm and contact time: 2.5 hours).

Langmuir Constants Adsorbent	$q_{\max}$	$b$ (L/mol)	$\Delta G/1000$ (kJ/mol)
NQ	14.4	226.3	-13.4
NB	20.96	283.8	-13.9

NQ = Natural Quartz

NB = Natural Bentonite

Table 2. Langmuir constants for adsorption of  $\text{Fe}^{3+}$  ions on NQ and NB

The obtained experimental data also has well described by Freundlich isotherm model into both NQ and NB. The negative value of  $\Delta G^\circ$  (- 13.4 and - 13.9 KJ/mol, respectively) confirms the feasibility of the process and the spontaneous nature of adsorption with a high preference for metal ions to adsorb onto NB more easily than NQ in pseudo second order rate reaction.

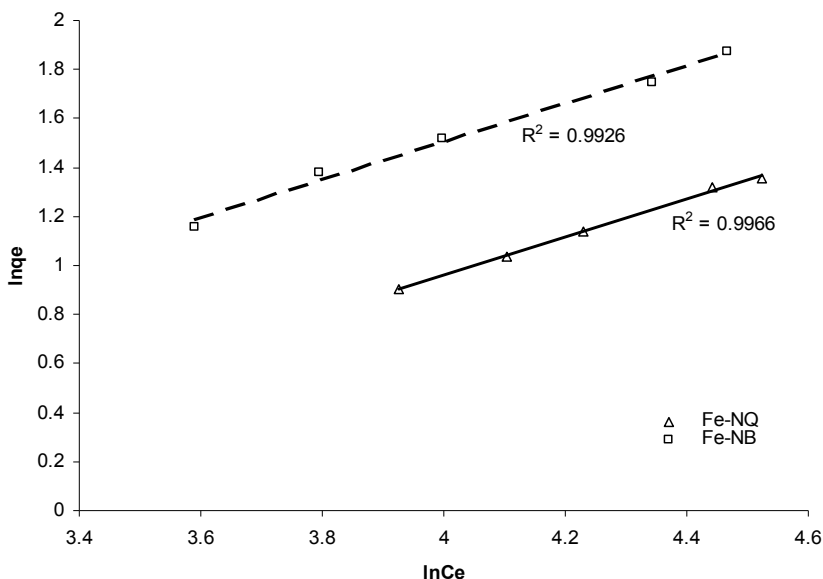


Fig. 9. The linearized Freundlich adsorption isotherms for  $\text{Fe}^{3+}$  ions adsorption by natural Quartz (NQ) and bentonite (NB) at constant temperature 30 °C. (initial concentration: 400 ppm, normal 1%  $\text{HNO}_3$  aqueous solution, 300 rpm, contact time: 2.5 hours).

## 5.2 Sorption isotherms of $\text{Fe}^{3+}$ ions processes using Jordanian natural zeolite (JNZ)

Isotherm studies are conducted at 30 °C by varying the initial concentration of  $\text{Fe}^{3+}$  ions [19]. Representative initial concentration (1000, 800, 600, 400 ppm) of  $\text{Fe}^{3+}$  ions are mixed with slurry concentrations (dose) of 20g/L for 150 min., which is the equilibrium time for the zeolite and  $\text{Fe}^{3+}$  ions reaction mixture. The equilibrium results are obtained at the 1%  $\text{HNO}_3$  model solution of  $\text{Fe}^{3+}$  ions. The Langmuir isotherm model is applied to the experimental data as presented in Figure 10. Our experimental results give correlation regression coefficient,  $R^2$ , equals to 0.998, which are a measure of goodness-of-fit and the general empirical formula of the Langmuir model by

$$\frac{c_e}{q_e} = 0.136c_e + 8.72$$

Our results are in a good qualitatively agreements with those found from adsorption of  $\text{Fe}^{3+}$  on the palm fruit bunch and maize cob [58].

Figure 11 represents the fitting data into the Freundlich model. We observe that the empirical formula of this model is found as  $\ln q_e = 0.1058 \ln C_e + 1.2098$  with  $R^2$  value equals to 0.954. It can be seen that the Langmuir model has a better fitting model than Freundlich as the former have higher correlation regression coefficient than the latter.

The value of standard Gibbs free energy change calculated at 30 °C is found to be -16.98 kJ/mol. The negative sign for  $\Delta G^0$  indicates to the spontaneous nature of  $\text{Fe}^{3+}$  ions adsorption on the JNZ.

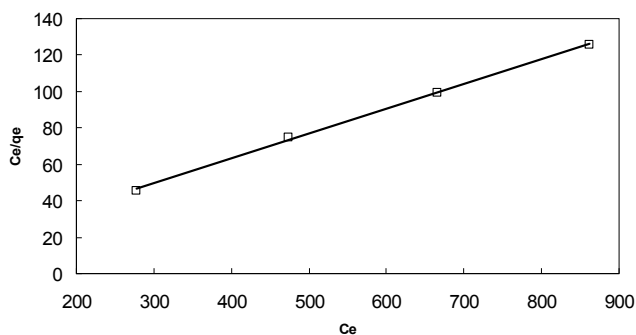


Fig. 10. Langmuir isotherm model plot of  $q_e/C_e$  of  $\text{Fe}^{3+}$  ions on JNZ vs.  $C_e$  (ppm) of  $\text{Fe}^{3+}$  ions.

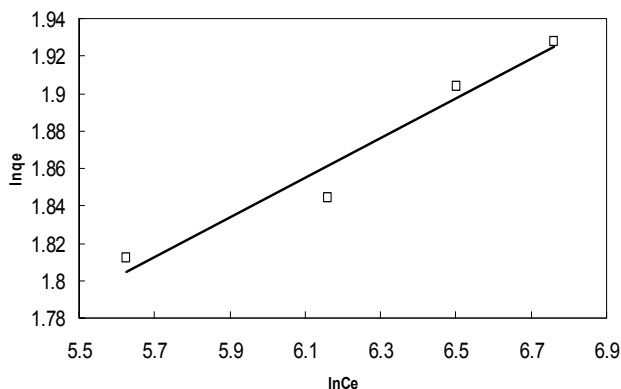


Fig. 11. Freundlich isotherm plot of  $\ln q_e$  vs.  $\ln C_e$ , where  $C_e$  is the equilibrium concentration of  $\text{Fe}^{3+}$  ions concentration.

### 5.3 Sorption isotherms of $\text{Fe}^{3+}$ ions processes using olive cake

To conduct the isotherm were studied at the initial pH of solution which was adjusted at the optimum value (pH = 4.5) and the mass of olive cake which was taken as 0.3, 0.5, 0.75 and 1.0 g at different temperatures of 28, 35 and 45 °C. Three adsorption isotherms models were used: Langmuir, Freundlich and Dubinin-Kaganer-Radushkevich (DKR) [21]. Figure 12 shows the experimental data that were fitted by the linear form of Langmuir model,  $(C_e/q_e)$  versus  $C_e$ , at temperatures of 28, 35 and 45 °C. The values of  $q_{\max}$  and  $b$  were evaluated

from the slope and intercept respectively for the three isothermal lines. These values of  $q_{\max}$  and  $b$  are listed in Table 1 with their uncertainty and their regression coefficients,  $R^2$ . Table 3 shows that the values of  $q_{\max}$  and  $b$  are decreased when the solution temperature increased from 28 to 45 °C. The decreasing in the values of  $q_{\max}$  and  $b$  with increasing temperature indicates that the  $\text{Fe}^{3+}$  ions are favorably adsorbed by olive cake at lower temperatures, which shows that the adsorption process is exothermic. In order to justify the validity of olive cake as an adsorbent for  $\text{Fe}^{3+}$  ions adsorption, its adsorption potential must be compared with other adsorbents like eggshells [59] and chitin [24] used for this purpose. It may be observed that the maximum sorption of  $\text{Fe}^{3+}$  ions on olive cake is approximately greater 10 times than those on the chitin and eggshells.

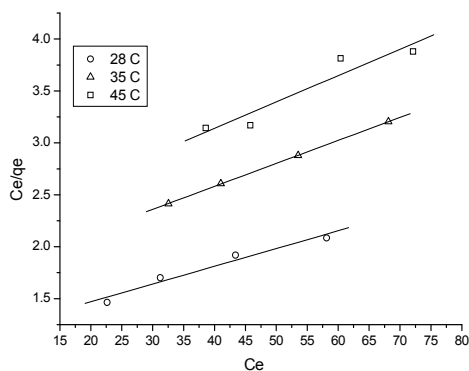


Fig. 12. Langmuir isotherm model plot of  $q_e/C_e$  of  $\text{Fe}^{3+}$  ions on olive cake vs.  $C_e$  (ppm) of  $\text{Fe}^{3+}$  ions.

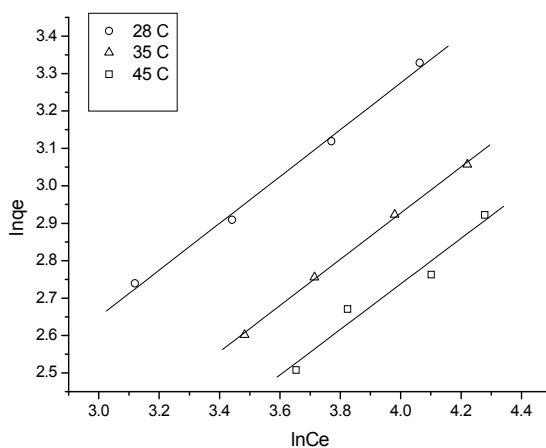


Fig. 13. The linearized Freundlich adsorption isotherms for  $\text{Fe}^{3+}$  ions adsorption by olive cake at different temperatures. (Initial concentration of  $\text{Fe}^{3+}$  ions: 100 ppm, Agitation speed: 100 rpm, pH: 4.5, Contact time: 24 hr).

The Freundlich constants  $K_f$  and  $n$ , which respectively indicating the adsorption capacity and the adsorption intensity, are calculated from the intercept and slope of plot  $\ln q_e$  versus  $\ln C_e$  respectively, as shown in Figure 13. These values of  $K_f$  and  $n$  are also listed in Table 3 with their regression coefficients. It can be observed that the values of  $K_f$  are decreased with increasing the temperature of solution from 28 to 45 °C. The decreasing in these values with temperature confirms also that the adsorption process is exothermic. It can be also seen that the values of  $1/n$  decreases as the temperature increases. Our experimental data of values  $K_f$  and  $1/n$  are considered qualitatively consistence with those that found in adsorption of  $\text{Fe}^{3+}$  ions on eggshells [59] and chitin [24].

The negative value of  $\Delta G^0$  (Table 4) confirms the feasibility of the process and the spontaneous nature of sorption. The values of  $\Delta H^0$  and  $\Delta S^0$  are found to be -10.83 kJ/mol and 19.9 J/mol-K, respectively (see Table 4). The negative values of  $\Delta H^0$  indicate and exothermic sorption reaction process. The positive value of  $\Delta S^0$  shows the increasing randomness at the solid/liquid interface during the sorption of  $\text{Fe}^{3+}$  ions onto olive cake.

T (°C)	Langmuir			Freundlich			DKR			
	$b$	$q_{\max}$	$R^2$	$1/n$	$k_f$	$R^2$	$\beta$	$X_m$	$E$	$R^2$
28	0.0152 ± 0.00011	58.479 ± 3.44	0.91	0.626	2.164 ± 0.98	0.99	-0.00006	28.321 ± 1.04	91.287	0.91
35	0.0130 ± 0.00013	45.249 ± 1.17	0.99	0.618	1.578 ± 0.15	0.99	-0.00009	23.539 ± 2.12	74.535	0.97
45	0.012 ± 0.00014	39.370 ± 2.07	0.96	0.609	1.354 ± 0.12	0.96	-0.0001	20.816 ± 1.01	70.711	0.95

Table 3. Langmuir, Freundlich and DKR constants for adsorption of  $\text{Fe}^{3+}$  ions on olive cake

T(°C)	$b(\text{L/mol})$	$-\Delta G^0$ (kJ/mol)	$-\Delta H^0$ (kJ/mol)	$\Delta S^0$ (J/mol-K)
28	847.2174	16.8718		
35	727.8947	16.8755	10.83	19.9
45	667.7246	17.1953		

Table 4. Thermodynamics parameters for the adsorption of  $\text{Fe}^{3+}$  ions on olive cake.

#### 5.4 Sorption isotherms of $\text{Fe}^{3+}$ ions processes using chitosan and cross-linked chitosan beads

A batch adsorption system was applied to study the adsorption of  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  ions from aqueous solution by chitosan and cross-linked chitosan beads [60 - 61]. The adsorption capacities and rates of  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  ions onto chitosan and cross-linked chitosan beads were evaluated as shown in Figure 14. Experiments were carried out as function of pH, agitation period, agitation rate and concentration of  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  ions. Langmuir and Freundlich

adsorption models were applied to describe the isotherms and isotherm constants. Equilibrium data agreed very well with the Langmuir model.

The calculated results of the Langmuir and Freundlich isotherm constants are given in Table 5. It is found that the adsorptions of  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  ions on the chitosan and cross-linked chitosan beads correlated well ( $R > 0.99$ ) with the Langmuir equation as compared to the Freundlich equation under the concentration range studied.

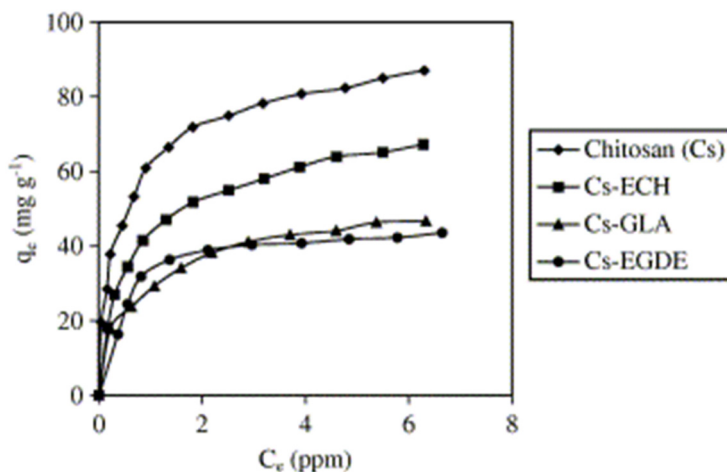


Fig. 14. Adsorption isotherms of  $\text{Fe}^{3+}$  ions on chitosan and cross-linked chitosan beads. chitosan, chitosan-GLA and chitosan-ECH bead

Iron	Beads	Langmuir			Freundlich		
		$Q$ ( $\text{mg g}^{-1}$ )	$b$ ( $\text{ml mg}^{-1}$ )	$R$	$K_F$ ( $\text{mg g}^{-1}$ )	$n$	$R$
Fe(II)	Chitosan	64.10	2197	0.9986	42.74	4.77	0.7730
	Chitosan-ECH	57.47	1891	0.9998	33.25	3.19	0.9597
	Chitosan-GLA	45.25	1023	0.9995	21.84	2.71	0.9744
	Chitosan-EGDE	38.61	762	0.9985	17.15	2.79	0.9985
Fe(III)	Chitosan	90.09	2413	0.9989	55.27	3.32	0.9824
	Chitosan-ECH	72.46	1550	0.9987	39.35	2.98	0.9788
	Chitosan-GLA	51.55	1405	0.9989	28.63	3.47	0.9881
	Chitosan-EGDE	46.30	2076	0.9991	28.36	3.61	0.8982

Table 5. Langmuir and Freundlich isotherm constants and correlation coefficients

Table 6 lists the calculated results. Based on the effect of separation factor on isotherm shape, the  $R_L$  values are in the range of  $0 < R_L < 1$ , which indicates that the adsorptions of  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  ions on chitosan and cross-linked chitosan beads are favourable. Thus, chitosan and cross-linked chitosan beads are favourable adsorbents. As mentioned earlier, chitosan and cross-linked chitosan beads are microporous biopolymers, therefore pores are large enough to let  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  ions through. The mechanism of ion adsorption on porous adsorbents may involve three steps [60 - 61]: (i) diffusion of the ions to the external surface of adsorbent; (ii) diffusion of ions into the pores of adsorbent; (iii) adsorption of the ions on the internal

surface of adsorbent. The first step of adsorption may be affected by metal ion concentration and agitation period. The last step is relatively a rapid process.

Iron	Initial concentration (ppm)	$R_L$ value			
		Chitosan	Chitosan-ECH	Chitosan-GLA	Chitosan-EGDE
Fe(II)	3	0.1317	0.1498	0.2457	0.3044
	6	0.0705	0.0810	0.1401	0.1795
	9	0.0481	0.0555	0.0980	0.1273
Fe(III)	3	0.1214	0.1769	0.1917	0.1383
	6	0.0646	0.0971	0.1060	0.0743
	9	0.0440	0.0669	0.0732	0.0508

Table 6.  $R_L$  values based on the Langmuir equation

### 5.5 Adsorption of $Fe^{3+}$ ions on activated carbons obtained from bagasse, pericarp of rubber fruit and coconut shell

The adsorptions of  $Fe^{3+}$  ions from aqueous solution at room temperature on activated carbons obtained from bagasse, pericarp of rubber fruit and coconut shell have been studied [62]. The adsorption behavior of  $Fe^{3+}$  ions on these activated carbons could be interpreted by Langmuir adsorption isotherm as monolayer coverage. The maximum amounts of  $Fe^{3+}$  ions adsorbed per gram of these activated carbons were 0.66 mmol/g, 0.41 mmol/g and 0.18 mmol/g, respectively. The mechanism by which the adsorption of  $Fe^{3+}$  ions onto the activated carbon can be performed after being reduced to  $Fe^{2+}$  ions [63].

Figures 15 to 17 show the Langmuir plots that have the greatest values of iron adsorption on three types of activated carbons. The maximum adsorption at monolayer coverage on bagasse, pericarp of rubber fruit and coconut shell are in the range 0.25 - 0.66 mmol/g, 0.11 - 0.41 mmol/g and 0.12 - 0.19 mmol/g, respectively. The experimental result shows that the amount of iron ion adsorbed on activated carbons decreased with increasing adsorption temperature. This suggested that the adsorption mechanism was physical adsorption, in contrast to chemical adsorption in which the amount of adsorbate adsorbed on an adsorbent increases with increasing adsorption temperature [63].

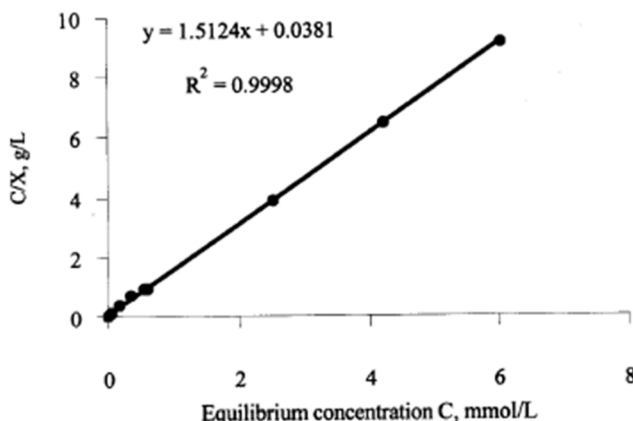


Fig. 15. Langmuir plot for the adsorption of  $Fe^{3+}$  ions on activated carbon obtained from bagasse.



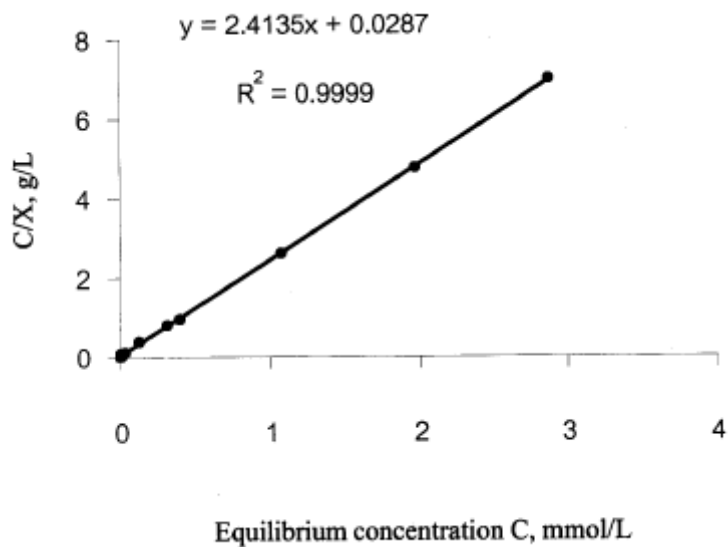


Fig. 16. Langmuir plot for the adsorption of  $\text{Fe}^{3+}$  ions on activated carbon obtained from pericarp of rubber fruit.

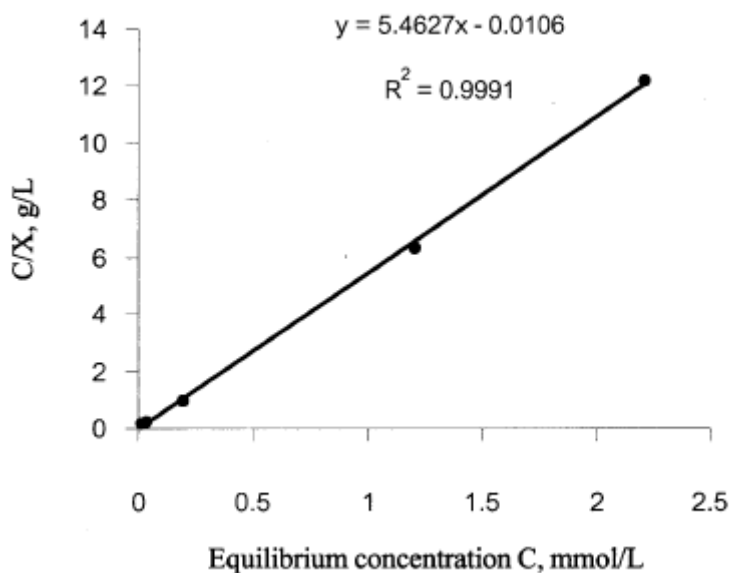


Fig. 17. Langmuir plot for the adsorption of  $\text{Fe}^{3+}$  ions on activated carbon obtained from coconut shell.

Study of the temperature dependence on these adsorptions has revealed them to be exothermic processes with the heats of adsorption of about -8.9 kJ/mol, -9.7 kJ/mol and -5.7 kJ/mol for bagasse, pericarp of rubber fruit and coconut shell, respectively. The value of Langmuir isotherm constant for the maximum adsorption at monolayer coverage ( $X_{max}$ ) and the heats of adsorption ( $\Delta H_{ads}$ ) of  $Fe^{3+}$  ions on three types of activated carbons was summarized in Table 7.

Raw Materials	$X_{max}$	$\Delta H_{ads}$ (KJ/mol)
Bagasse	0.66	- 8.9
Pericarp of Rubber Fruit	0.41	- 9.7
Coconut Shell	0.19	- 5.7

Table 7. The maximum adsorption of iron ion at monolayer coverage ( $X_{max}$ ) and the heats of adsorption ( $\Delta H_{ads}$ ) for iron ion on three types of activated carbons and activation temperature at 600 °C

### 5.6 Adsorption of $Fe^{3+}$ ions on unmodified raphia palm (*Raphia Hookeri*) fruit endocarp

The adsorption of aqueous  $Fe^{3+}$  ions onto the surface of *Raphia* palm fruit endocarp (nut) was studied in a batch system [64 – 67]. The influence of initial  $Fe^{3+}$  ions concentration, temperature and particle size was investigated and the results showed that particle size and temperature affected the sorption rate and that the adsorption was fast with a maximum percentage adsorption of 98.7% in 20 min as initial metal ion concentration was increased. There is a general decrease in sorption efficiency as the particle size is increased. The increased sorption with smaller particle size means that there is higher external surface area available for adsorption with smaller particle at a constant total mass.

Four isotherms; Langmuir, Freundlich, Dubinin-Radushkevich (D-R) and Temkin were used to model the equilibrium sorption experimental data. The sorption process was found to follow chemisorption mechanism. From Dubinin-Radushkevich (D-R) isotherm, the apparent energy of adsorption was 353.55 KJ/mol. The apparent energy shows if the sorption process follows physisorption, chemisorption or ion exchange. It has been reported:

1. Physisorption processes have adsorption energies <40 KJ/mol
2. Chemisorption processes have adsorption energies > 40 KJ/mol
3. Chemical ion exchange have adsorption energies between 8.0 and 16 KJ/mol
4. Adsorption is physical in nature have adsorption energies <8.0 KJ/mol

From the result obtained, the sorption of  $Fe^{3+}$  ions onto *Raphia* palm fruit endocarp (nut) was chemisorption process.

In order to describe the thermodynamic behaviour of the sorption of  $Fe^{3+}$  ion onto *Raphia* palm fruit endocarp (nut) from aqueous solution, thermodynamic parameters including  $\Delta G^\circ$ ,  $\Delta H^\circ$ ,  $\Delta S^\circ$ , were evaluated. The value of  $\Delta H^\circ$  is negative indicating exothermic process. The standard Gibbs free energy indicates that the the sorption process is spontaneous in nature and also feasible. The decreasing in  $\Delta G^\circ$  values with increasing temperature shows a decrease in feasibility of sorption at higher temperature.

<b>Isotherm Constants</b>	<b>Fe(III)</b>
Langmuir	- 16.3500
$q_{max}$	- 0.0800
b	0.9321
$R^2$	33.1000
$\Delta q(\%)$	
Freundlich	
n	0.6900
$K_f$	3.9800
$R^2$	0.9222
$\Delta q(\%)$	8.8100
Dubinin-Radushkevich	
$\beta$	$4.0 \cdot 10^{-6}$
$q_D$	104.4800
E	353.5500
$R^2$	0.9461
$\Delta q(\%)$	24.5200

Table 8. Isotherm constant for adsorption of  $Fe^{3+}$  ions onto *Raphia* palm fruit endocarp (nut) from aqueous solution

<b>Constants (KJ/mol/K)</b>	<b>Fe(III)</b>
$\Delta H$	- 2560.4600
$\Delta S$	19.1900
$E_A$	2094.8000
$S^n$	0.0900
$R^2$ (Vant Hoff)	0.4633
$R^2$ (Sticking Probability)	0.4747

Table 9. The activation energy of Fe(III)

The activation energy of any reaction process depicts the energy barrier which the reactants must overcome before any reaction could take place. High activation energy to react hence decrease in reaction rate.

## 6. Conclusion

High and low concentration level of  $Fe^{3+}$  ions can be adsorbed on different types of natural adsorbents. This process can be used to remove  $Fe^{3+}$  ions from aqueous solutions. The thermodynamic isotherms indicate the behavior picture of  $Fe^{3+}$  ions onto the surface of natural adsorbent as homogenous or heterogeneous monolayer coverage. It depends on the chemical nature of adsorbent surfaces. Mostly, the thermodynamic parameters show the spontaneous and exothermic adsorption processes of  $Fe^{3+}$  ions onto the surfaces of natural adsorbents, indicating of easier handling.

## 7. References

- [1] Sawyer, C. N., and McCarty, P. L. (1978). *Chemistry for environmental engineering*, 3rd Ed., McGraw-Hill, Singapore, 85-90
- [2] Lee, S. H. D., and Johnson, I. J. (1980). "Removal of gaseous alkali metal compounds from hot flue gas by particulate sorbents." *J. Engrg. Power*, 102, 397-402.
- [3] Uberoi, M., and Shadman, F. (1990). "Sorbents for removal of lead compounds from hot flue gases." *AICHE J.*, 36(2), 307-309.
- [4] Ho, T. C., Chen, C., Hopper, J. R., and Oberacker, D. A. (1992). "Metal capture during fluidized bed incineration of wastes contaminated with lead chloride." *Combustion Sci. and Technol.*, 85, 101-116.
- [5] Ho, T. C., Chen, J. M., Shukla, S., and Hopper, J. R. (1990). "Metal capture during fluidized bed incineration of solid wastes." *AICHE Symp. Ser.*, 276(86), 51-60.
- [6] Ho, T. C., Chu, H. W., and Hopper, J. R. (1993). "Metal volatilization and separation during incineration." *Waste Mgmt.*, 13, 455-466.
- [7] Ho, T. C., Tan, T., Chen, C., and Hopper, J. R. (1991). "Characteristics of metal capture during fluidized bed incineration." *Aiche Symp. Ser.*, 281(87), 118-126. a) M.C. Macias-Pbrez, C. Salinas-Martinez de Lecea, M.J. Muiloz- Guillena and A. Linares-Solano, *Low Temeratijre SO<sub>2</sub> C A P m by Calcium Based Sorbents: Caaracterization of the Active Calcium*. b) MuRoz-Guillena, M.J., Linares-Solano, A. and Salinas-Martinez, de Lecea, C. *Appl. Surf. Sci.*. 1995, 89, 197. Chau-Hwa Yang and James G. Goodwin, Jr., *Journal of Catalysis* 78: 1, 1982, Pages 182-187A. Guerrero-Ruiz, *Reaction Kinetics and Catalysis Letters*, 49:1, 53-60,  
<http://www.webref.org/chemistry/p/physisorption.htm>, International Union of Pure and Applied Chemistry,  
<http://wikichemistry.com/konfuciy.asp?tda=dt&t=13145&fs=physisorption+-+charactersitics>
- [8] P. Somasundaran, Somil C. Mehta, X. Yu, and S. Krishnakumar, *handbook of Surface and Colloid Chemistry, Third Edition, Colloid Systems and Interfaces Stability of Dispersions through Polymer and Surfactant Adsorption*, chapter 6.
- [9] Deguo Kong 2009 , Master Thesis, Department of Land and Water Resources Engineering , Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden.
- [10] M. G. Lee, J.K. Cheon, S.K. Kam, *J. Ind. Chem. Eng.* 9(2) (2003) 174-180. a) M. Al-Anber and Z. Al-Anber, *Utilization of natural zeolite as ion-exchange and sorbent material in the removal of iron. Desalination*, 255, (2008) 70 - 81. b) M. Al-anber, *Removal of Iron(III) from Model Solution Using Jordanian Natural Zeolite: Magnetic Study*, *Asian J. Chem. Vol. 19, No. 5 (2007)*, 3493-3501.
- [11] A. Guerrero-Ruiz, *Reaction Kinetics and Catalysis Letters*, 49:1, 53-60.
- [12] <http://www.webref.org/chemistry/p/physisorption.htm>, International Union of Pure and Applied Chemistry.
- [13] <http://wikichemistry.com/konfuciy.asp?tda=dt&t=13145&fs=physisorption+-+charactersitics>
- [14] P. Somasundaran, Somil C. Mehta, X. Yu, and S. Krishnakumar, *handbook of Surface and Colloid Chemistry, Third Edition, Colloid Systems and Interfaces Stability of Dispersions through Polymer and Surfactant Adsorption*, chapter 6.
- [15] Deguo Kong 2009 , Master Thesis, Department of Land and Water Resources Engineering , Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden.

- a) R.C. Weast, CRC Handbook of Chemistry and Physics, CRC Press, Boca Raton, FL, 1979, p. F-214.]= b) Y. Onganer, C. Temur, Adsorption dynamics of Fe(III) from aqueous solution onto activated carbon, *J. Colloid Interf. Sci.* 205 (1998) 241-244.
- [16] Adil Denizli, Ridvan Say and Yakup Arica Separation and Purification Technology, 21 (2000) 181-190.
- [17] M. G. Lee, J.K. Cheon, S.K. Kam, *J. Ind. Chem. Eng.* 9(2) (2003) 174-180.
- [18] a) M. Al-Anber and Z. Al-Anber, Utilization of natural zeolite as ion-exchange and sorbent material in the removal of iron. *Desalination*, 255, (2008) 70 – 81. b) M. Al-anber, Removal of Iron(III) from Model Solution Using Jordanian Natural Zeolite: Magnetic Study, *Asian J. Chem. Vol.* 19, No. 5 (2007), 3493-3501.
- [19] Mohammad Al-Anber, Removal of High-level  $Fe^{3+}$  from Aqueous Solution using Jordanian Inorganic Materials: Bentonite and Quartz, *Desalination* 250 (2010) 885-891.
- [20] Z. Al-Anber and M. Al-Anber, Thermodynamics and Kinetic Studies of Iron(III) Adsorption by Olive Cake in a Batch System, *J. Mex. Chem. Soc.* 2008, 52(2), 108-115.
- [21] Tayel El-Hasan, Zaid A. Al-Anber, Mohammad Al-Anber\*, Mufeed Batarseh, Farah Al-Nasr, Anf Ziadat, Yoshigi Kato and Anwar Jiries. Removal of  $Zn^{2+}$ ,  $Cu^{2+}$  and  $Ni^{2+}$  Ions from Aqueous Solution via Tripoli: Simple Component with Single Phase Model, *Current World Environment*, 3(1) (2008) 01-14
- [22] L.N. Rao, K.C.K. Krishnaiah, A. Ashutosh, *Indian J. Chem. Technol.* 1 (1994) 13.
- [23] G. Karthikeyan, N.M. Andal and K Anbalagan, Adsorption studies of iron(III) on chitin, *J. Chem. Sci.* 117 (6) (2005) 663–672.
- [24] a) N. Khalid, S.Ahmad, S.N. Kiani, J.Ahmed, Removal of mercury from aqueous solutions by adsorption to rice husks, *Sep. Sci. Technol.* 34 (16) (1999) 3139–3153. b) N. Khalid, S. Ahmed, S.N. Kiani, J. Ahmed, Removal of Lead from Aqueous Solutions Using Rice Husk, *Sep. Sci. Technol.* 33 (1998) 2349–2362.
- [25] G. Karthikeyan, N.M. Andal and K Anbalagan, Adsorption studies of iron(III) on chitin, *J. Chem. Sci.* 117 (6) (2005) 663–672.
- [26] *Journal of Research of the National Bureau of Standards – A. Physics and Chemistry*, 66A:6 (1962) 503-515
- [27] N. Kannan and T. Veemaraj, Removal of Lead(II) Ions by Adsorption onto Bamboo Dust and Commercial Activated Carbons -A Comparative Study, *E-Journal of Chemistry*, 2009, 6(2), 247-256.
- [28] I. Langmuir, adsorption of gases on plain surfaces of glass mica platinum, *J. Am. Chem. Soc.* 40 (1918) 136-403. b) J.M. Coulson, J.F. Richardson with J.R. Backhurst, J.H. Harker, (1991) *Chemical Engineering*, Vol 2, 4<sup>th</sup> Edt. "practical Technology and separation processes" Pergamon Press, Headington Hill Hall, Oxford. c) (Domenico, P.A. and Schwartz, F.W., "Physical and Chemical Hydrogeology", 1st Ed., John Wiley and Sons, New York, 1990). d) Reddi, L.N. and Inyang, H.I., "Geo-Environmental Engineering Principles and Applications", Marcel Decker Inc., New York, 2001. e) Nitzsche, O. and Vereecken, H., "Modelling Sorption and Exchange Processes in Column Experiments and Large Scale Field Studies". *Mine Water and the Environment*, 21, 15-23, 2002. f) F. Banat, S. Al-Asheh, D. Al-Rousan, A comparative study of copper and zinc ion adsorption on to activated date-pits. *Adsorption Science Technology* 2002, 20 (4) 319-335.

- [29] Vermeulan et al., 1966 T.H. Vermeulan, K.R. Hall, L.C. Eggleton and A. Acrivos *Ind. Eng. Chem. Fundam.* 5 (1966), pp. 212–223.
- [30] a) Metcalf and Eddy. (2003). *Wastewater Engineering, Treatment, Disposal and Reuse*, 3rd Ed. McGraw-Hill: New York. b) Zeldowitsch, 1934 J. Zeldowitsch, Über den Mechanismus der katalytischen Oxydation von CO an MnO<sub>2</sub>, *Acta Physicochim. URSS* 1 (1934), pp. 364–449. c) A. EDWIN VASU, *E-Journal of Chemistry*, 5:1, (2008) 1-9.
- [31] Hutson N D, Yang R T, Adsorption. 1997, 3, 189. Krishna B S, Murty D S R, Jai Prakash B S, *J. Colloid Interf Sci.* 2000, 229, 230.
- [32] a) Arivoli S, Venkatraman B R, Rajachandrasekar T and Hema M, Adsorption of ferrous ion from aqueous solution by low cost activated carbon obtained from natural plant material, *Res J Chem Environ.* 2007, 17, 70-78. b) Arivoli S, Kalpana K, Sudha R and Rajachandrasekar T, Comparative study on the adsorption kinetics and thermodynamics of metal ions onto acid activated low cost carbon, *E J Chem*, 2007, 4, 238-254. c) Renmin Gong, Yingzhi Sun, Jian Chen, Huijun Liu, Chao yang, Effect of chemical modification on dye adsorption capacity of peanut hull, *Dyes and Pigments*, 2005, 67, 179.
- [33] Jj S Arivoli, P.Martin Deva Prasath and M Thenkuzhali, *EJEAFCh*, 6 (9), (2007) 2323-2340
- [34] O. Sirichote, W. Innajitara, L. Chuenchom, D. Chunchit and K. Naweean. Songklanakarin *J. Sci. Technol.*, 2002, 24(2) : 235-242.
- M. A. AL-Anber, Z. AL-Anber, I. AL-Momani, Thermodynamics and Kinetic Studies of Iron(III) Adsorption by Jojoba seeds in a Batch System, in preparation (2011).
- [35] J. H. Duffus, "Heavy metals"- A Meaningless Term? IUPAC Technical Report, *Pure Appl. Chem.* 74:5 (2002) 793-807.
- [36] E. D. Weinberg, Iron Loading and Disease Surveillance. *Emerging Infectious Diseases*, 5 (3) 1999.
- [37] B. Das, P. Hazarika, G. Saikia, H. Kalita, D.C. Goswami, H.B. Das, S.N. Dube, R.K. Dutta, Removal of iron from groundwater by ash: A systematic study of a traditional method, *Journal of Hazardous Materials* 141 (2007) 834–841.
- [38] a) Gedge G (1992) Corrosion of cast Iron in potable water service. In: *Proceedings of the Institute of Materials Conference*. London, UK. b) Rice, O. Corrosion Control with Calgon. *Journal AWWA*, 39(6), (1947) 552. c) Hidmi, L.; Gladwell, D. & Edwards, M. Water Quality and Lead, Copper, and Iron Corrosion in Boulder Water. Report to the City of Boulder, CO (1994).
- [39] J.D. Zuane, *Handbook of Drinking Water Quality*. Van Nostrand Reinhold, New York, 1990.
- a) G. Chen, *Electrochemical technologies in wastewater treatment. Sep. Purif. Technol.* 38 (1), (2004) 11–41. b) M. Wessling-Resnick, *Biochemistry of iron uptake. Crit. Rev. Biochem. Mol. Biol.* 34, (1999) 285–314. c) G. J. Kontoghiorghes and E.D. Weinberg Iron: mammalian defense systems, mechanisms of disease, and chelation therapy approaches. *Blood Rev.* 9 (1995) 33-45. d) E.D. Weinberg and G.A. Weinberg. The role of iron in infection. *Current Opinion in Infectious Diseases*, 8 (1995) 164-9. e) E.D. Weinberg, The role of iron in cancer. *Eur. J. Cancer. Prev.* 5 (1996) 19-36. f) J.R. Connor and J.L. Beard, Dietary iron supplements in the elderly: to use or not to use? *Nutrition Today*, 32 (1997) 102-9. j) T-P. Tuomainen, K. Punnonen, K.

- Nyyssonen, J.T. Salonen. Association between body iron stores and the risk of acute myocardial infarction in men. *Circulation*, 97 (1998) 1461-6. h) J.M. McCord, Effects of positive iron status at a cellular level. *Nutr. Rev.*, 54 (1996) 85-8. i) D. Weinberg, Patho-ecological implications of microbial acquisition of host iron. *Reviews in Medical Microbiology*, 9 (1998) 171-8.
- [40] P. Sarin, V.L. Snoeyink, J. Bebee, K.K. Jim, M. A. Beckett, W.M. Kriven, J. A. Clement, Iron release from corroded iron pipes in drinking water distribution systems, *Water Research*, 38,5, (2004)1259-1269.
- [41] <http://www.unu.edu/unupress/unupbooks/80858e/80858E02.htm>
- [42] F.R. Spellman, *Handbook for Waterworks Operator Certification*, Vol. 2, Technomic Publishing Company Inc., Lancaster, USA, 2001, pp 6-11, 81-83.
- W.C. Andersen, T.J. Bruno, Application of gas-liquid entraining rotor to supercritical fluid extraction: removal of iron (III) from water, *Anal. Chim. Acta* 485 (2003) 1-8.
- [43] a) D. Ellis, C. Bouchard, G. Lantagne, Removal of iron and manganese from groundwater by oxidation and microfiltration, *Desalination* 130 (2000) 255-264. b) Hikmet Katircioglu, Belma Aslm, Ali Rehber Tu'rker, Tahir Atic, Yavuz Beyatl, Removal of cadmium(II) ion from aqueous system by dry biomass, immobilized live and heat-inactivated *Oscillatoria* sp. H1 isolated from freshwater (Mogan Lake) *Bioresource Technology* 99 (2008) 4185-4191
- [44] P. Berbenni, A. Pollice, R. Canziani, L. Stabile, F. Nobili, Removal of iron and manganese from hydrocarbon-contaminated groundwaters, *Bioresour. Technol.* 74 (2000) 109-114.
- [45] a) M. Kalin, W.N. Wheeler, G. Meinrath, The removal of uranium from mining waste water using algal/microbial biomass. *J. Environ. Radioact.* 78 (2005) 151-177. b) F. Veglio, F. Beolchini, Removal of metals by biosorption: a review. *Hydrometall.* 44 (1997) 301-316.
- [46] M. Uchida, S. Ito, N. Kawasaki, T. Nakamura, S. Tanada, Competitive adsorption of chloroform and iron ion onto activated carbon fiber, *J. Colloid Interf. Sci.* 220 (1999) 406-409.
- [47] M. Pakula, S. Biniak, A. Swiatkowski, Chemical and electrochemical studies of interactions between iron (III) ions and activated carbon surface, *Langmuir* 14 (1998) 3082-3089.
- [48] C. Huang, W.P. Cheng, Thermodynamic parameters of iron-cyanide adsorption onto -Al<sub>2</sub>O<sub>3</sub>, *J. Colloid Interf. Sci.* 188 (1997) 270-274.
- [49] a) E. A. Sigworth and S. B. Smith, Adsorption of inorganic compound by activated carbon, *J. Am. Water Works Assoc.*, 64, 386 (1972). b) E. Jackwerth, J. Lohmar and G. Wittler, *Fresenius' Z. Anal. Chem.*, 266, 1(1973). c) A. Edwin Vasu, Adsorption of Ni(II), Cu(II) and Fe(III) from Aqueous Solutions Using Activated Carbon, *E-Journal of Chemistry*, 5 (1), 2008, 1-9. d) A. Ucer, A. Uyanik, S.F. Aygun, Adsorption of Cu<sup>2+</sup>, Cd<sup>2+</sup>, Zn<sup>2+</sup>, Mn<sup>2+</sup> and Fe<sup>3+</sup> ions by tannic acid immobilised activated carbon, *Separation and Purification Technology* 47 (2006) 113-118.
- [50] Hideko Koshima, Adsorption of Iron(III) on Activated Carbon from Hydrochloric Acid Solution, *Analytical Sciences* (1), 1985, 195.
- [51] a) N. Cvjetićanin, D. Cvjetićanin, D. Golobočanin, and M. Pravica, Adsorption of colloidal trivalent iron on alumina, *J. Radioanalytical and Nuclear Chem.*, 54(1-2), 1979, 149-158.

- [52] A. Edwin Vasu, *E-Journal of Chemistry*, 5:1, (2008) 1-9.
- [53] M.M. Nassar, K.T. Ewida, E.E. Ebrahiem, Y.H. Magdy and M.H. Mheaedi, Adsorption of iron and manganese ions using low-cost materials as adsorbents, *Adsorp. Sci. Technol.*, 22(1) (2004) 25-37.
- [54] Nacéra Yeddou, Aicha Bensmaili, Equilibrium and kinetic modeling of iron adsorption by eggshells in a batch system: effect of temperature, *Desalination* 206 (2007) 127-134.
- [55] W. S. Wan Ngah, S. Ab Ghani and A. Kamari, *Bioresource Technology* Volume 96, Issue 4, March 2005, Pages 443-450.
- [56] Peniche-Covas, C., Alvarez, L.W., Arguelles-Monal, W., 1992. The adsorption of mercuric ions by chitosan. *J. Appl. Polym. Sci.* 46, 1147-1150.
- [57] O. Sirichote, W. Innajitara, L. Chuenchom, D. Chunchit and K. Naweean. Songklanakarin *J. Sci. Technol.*, 2002, 24(2) : 235-242.
- [58] Uchida, M., Shinohara, O., Ito, S., Kawasaki, N., Nakamura, T. and Tanada, S. 2000. Reduction of iron(III) ion by activated carbon fiber. *J. Colloid Interface Sci.* 224: 347-350.
- [59] C.Y. Abasi, A.A. Abia and J.C. Igwe, *Environmental Research Journal*, 5(3), 2011, Page No.: 104-113.
- [60] Y. S. Ho, C. T. Huang, H. W. Huang, equilibrium sorption isotherm for metal ions on tree fern. *Process biochem.* 37 (2002) 1421-1430.
- [61] B. A. Shah, A. V. Shah, R. Singh, Sorption isotherms and kinetics of chromium uptake from wastewater using natural sorbent materials. *Intern. J. Environ. Sci. Technology.* 6, (2009) 77-90.
- [62] a) M. Horsfall, A.I. Spiff, A.A. Abia, studies the influence of mercaptoacetic acid (MAA) modification of cassava (*Manihot esculenta* cranz) waste biomass on the adsorption of  $\text{Cu}^{2+}$  and  $\text{Cd}^{2+}$  from aqueous solution. *Bull Korean Chem. Soc.* 25 (2004) 969-976. b) P. Loderio, J.L. Barriada, R. Herrero, M.E. Sastre-Vicente, The marine macroalga *Crustoserira baccata* as biosorbent for  $\text{Cd}(\text{II})$  and  $\text{Pb}(\text{II})$  removal: kinetic and equilibrium studies. *Environ. Pollution*, 142 (2006) 264-273.



# Thermodynamics as a Tool for the Optimization of Drug Binding

Ruth Matesanz, Benet Pera and J. Fernando Díaz  
*Centro de Investigaciones Biológicas (C.S.I.C.)*  
Spain

## 1. Introduction

A non-covalent interaction is a kind of chemical bond, typically between macromolecules, that involves dispersed variations of electromagnetic interactions (Alberts *et al.* 1994; Connors & Mecozi 2010). Non-covalent interactions are individually weak as compared with covalent bonds, but their net strength is higher than the sum of that of the individual interactions. There are few drugs that bind irreversibly to their targets, in pharmacology, most drugs establish non-covalent interactions with their target molecules (usually proteins).

From a chemical point of view, the affinity constant ( $K_a$ ) is a very useful measurement for the study of binding reactions as it provides much information about the mechanism. In many cases some chemical or physical properties of ligand or target change with the interaction between them, these changes might help to measure binding constants. It is important to establish the stoichiometry of the complex to be sure that the constants are accurately calculated. From the affinity constants measured it is possible to calculate the standard thermodynamic quantities for the binding reaction: free-energy ( $\Delta G$ ), enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ).

Our group has already demonstrated that, in some cases, binding affinity measurements are very helpful for the optimization of ligand binding as it can be determined the contribution of every single chemical modification of the ligand to the binding affinity (Buey *et al.* 2004; Matesanz *et al.* 2008)

One of the objectives of drug development is the search of new or modified compounds with improved properties such as better potency, higher selectivity, better pharmacokinetics or superior drug resistance profiles. An important goal in this objective is the optimization of drugs binding affinity towards their targets, as binding affinity is directly related to potency (Ruben *et al.* 2006). Moreover, it has been shown that extremely high affinity drugs reflect as well changes in other properties like selectivity (Ohtaka *et al.* 2004; Ohtaka & Freire 2005) or resistance overcoming ability (Matesanz *et al.* 2008).

Examples of the importance of ligand affinity in drug optimization can be observed in the development of HIV-1 protease inhibitors and statins (cholesterol lowering drugs) over the years as remarked in (Freire 2008).

In this chapter we will study the nature of non-covalent interactions and the concept of binding constant for these interactions. Examples of methodologies to measure binding constants of small ligands to macromolecules will be introduced and we will emphasize the

need to determine the stoichiometry of the studied system to calculate accurately the constants. Once the thermodynamic concepts were introduced, we will show the use of these kind of studies for the optimization of drug binding to its target. We will detail the role of single chemical modifications in the molecule of study to modulate its binding affinity, and the way to quantify these changes. We will finally further discuss how the selection of the best substituents can result in the optimization of binding.

## 2. Non-covalent interactions

Non-covalent interactions are chemical bonds that do not involve sharing of electron pairs between orbitals of different atoms, there are no orbital overlapping in these interactions which have an electrostatic nature and are not highly directional. Covalent bonds are generally shorter than 2Å while the non-covalent ones are within the range of several angstroms. Another difference between these two types of bonds is the energy released in its formation, non-covalent interactions are weaker, with energies below 40 kJ/mol whereas covalent bonds energies range 80-800 kJ/mol.

These weak interactions have important roles in the binding of macromolecules with each other and with other molecules in the cell, in the maintenance of the three dimensional structure of large macromolecules such as proteins or nucleic acids (e.g. DNA double helix) and they are the forces found in the majority of the drug-proteins interactions in pharmacology.

### 2.1 Types of non-covalent interactions

There are four commonly mentioned fundamental non-covalent interaction types including ionic interactions, hydrogen bonds, hydrophobic interactions and van der Waals forces (dispersion attractions, dipole-dipole and dipole-induced dipole interactions). All these weak interactions must work together to have significant effects. Their combined bond effect is greater than the sum of the individual ones. The free energy of multiple bonds between two molecules is different than the sum of the enthalpies of each bond due to entropic effects.

#### 2.1.1 Ionic interactions

Ionic bonds result from the electrostatic attraction between two ionized groups of opposite charge such as carboxyl (-COO-) and amino (-NH<sub>3</sub><sup>+</sup>). These ionic interactions are directly proportional to the product of the interacting charges and inversely proportional to the dielectric constant of the medium and the distance separating the charges. This relationship is defined by Coulomb's law:

$$E = \frac{kq_1q_2}{Dr} \quad (1)$$

where E is the energy, q<sub>1</sub> and q<sub>2</sub> are the charges of two atoms, r is the distance between them, D is the dielectric constant, and k is a proportionality constant. A charged group on a molecule can attract an oppositely charged group from another molecule. By contract, an attractive interaction has a negative energy. The dielectric constant is important for the medium. In water, these bonds are very weak as the dielectric constant is much higher (D=80) than in vacuum (D=1). As an example, the electrostatic interaction between two atoms bearing single opposite charges separated by 3 Å in water has an energy of 5.9 kJ/mol (k=1389 kJ/mol).

### 2.1.2 Van der Waals forces

Van der Waals forces are short range attractive forces between chemical groups in contact. The forces are caused by slight charge displacements. The distribution of electronic charge around an atom changes with time. At any moment, the charge distribution is not perfectly symmetric. This transient asymmetry in the electronic charge around an atom induces a complementary asymmetry in the electron distribution around its neighboring atoms. These induced dipole effects give rise to the so called van der Waals interactions, also known as dispersion forces. The attraction between two atoms increases as they come closer to each other, until they are separated by the so called van der Waals contact distance. At a shorter distance, very strong repulsive forces become dominant because the outer electron clouds overlap. The van der Waals radius of an atom is defined where the net force between two atoms is zero. The van der Waals potential is then best described as a balance between attraction and repulsion.

Van der Waals forces are non-directional. Energies associated with them are quite small; typical interactions contribute from 2 to 4 kJ/mol per atom pair. However, when the surfaces of two large molecules come together, a large number of atoms are in van der Waals contact, and the net effect, summed over many atom pairs, can be substantial.

### 2.1.3 Hydrogen bonds

A hydrogen bond is an interaction between a proton donor group (a hydrogen atom covalently bound to an electronegative atom -e.g. F, O, N, S-) and a proton acceptor atom (another electronegative atom). It is a very important interaction responsible for the structure and properties of water, as well as the structure and properties of biological macromolecules (e.g. hydrogen bonds are responsible of specific base-pair formation in the DNA double helix).

Hydrogen bonds are fundamentally electrostatic interactions. The relatively electronegative atom to which the hydrogen atom is covalently bonded pulls electron density away from the hydrogen atom so that it develops a partial positive charge ( $\delta^+$ ). Thus, it can interact with an atom having a partial negative charge ( $\delta^-$ ) through an electrostatic interaction. However, this interaction is more than just an ionic or dipole-dipole interaction between the donor and the acceptor groups. Here, the distance between the hydrogen and acceptor atoms is less than the sum of their respective van der Waals radii.

Hydrogen bonds are directional toward the electronegative atom. The strongest hydrogen bonds have a tendency to be approximately straight, such that the proton donor group, the hydrogen atom, and the acceptor atom lie along a straight line, with significant weakening of the interaction if they are not colinear. They are somewhat longer than are covalent bonds. Hydrogen bonds are constantly being made and remade. Their half-life is about 10 seconds. These bonds have only 5% or so of the strength of covalent bonds. They have energies of 5-15 kJ/mol compared with approximately 420 kJ/mol for a carbon-hydrogen covalent bond. However, when many hydrogen bonds can form between two molecules (or parts of the same molecule), the resulting union can be sufficiently strong as to be quite stable. Examples of multiple hydrogen bonds are widely found in biological systems, they hold secondary structures of polypeptides, help in binding of enzymes to their substrate or antibodies to their antigen, help also transcription factors bind to each other or to DNA.

### 2.1.4 Hydrophobic interactions

Hydrophobic interactions result when non-polar molecules are in a polar solvent (e.g. water). The non-polar molecules group together to exclude water so that they minimize the

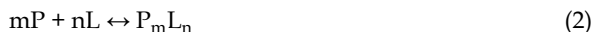
surface area in contact with the polar solvent. Unlike the non-covalent interactions mentioned above, which are pairwise interactions between atoms or parts of molecules, the nature of the hydrophobic interaction is very different. It involves a considerable number of (water) molecules, and does not arise from a direct force between the non-polar molecules.

Nonpolar molecules are not good acceptors of hydrogen bonds. When a non-polar molecule is placed in water, the hydrogen bonding network of water is disrupted. Water molecules must reorganize around the solute and make a kind of cage, similar to the structure of water in ice, in order to gain back the broken hydrogen bonds. This reorganization results in a considerable loss in the configurational entropy of water and therefore, in an increase in the free energy. If there are more than one such non-polar molecules, the configuration in which they are clustered together is preferred because now the hydrogen bonding network of water is disrupted in just one (albeit bigger) pocket, rather than in several small pockets. Therefore, the entropy of water is larger when the non-polar molecules are clustered together, leading to a decrease in the free energy.

Hydrophobic interactions have strengths comparable in energy to hydrogen bonds.

### 3. Binding constants

Most drugs have a non-covalent binding to their targets, thus these interactions are of great importance for our studies. Measurements of equilibrium constants, their dependence with temperature, the determination of stoichiometry, provide main information on the mechanism of the chemical process involved. The basic process can be taken out of the association of ligand (or ligands) to its target. The binding reaction can be written as follows:



Regardless of mechanism, every reversible reaction reaches an equilibrium distribution of reactants and products. At some point the rates of the opposing reactions (association and dissociation in our case) become equal and there would no longer be any change in the concentration of the molecules implied.

$$v_{\text{ass}} = k_{\text{ass}} [P]^m [L]^n \quad (3)$$

$$v_{\text{diss}} = k_{\text{diss}} [P_m L_n] \quad (4)$$

Under these conditions ( $v_{\text{ass}} = v_{\text{diss}}$ ):

$$\frac{k_{\text{ass}}}{k_{\text{diss}}} = \frac{[P_m L_n]}{[P]^m [L]^n} = K_a \quad (5)$$

that will be the equilibrium association constant assuming that activities are equal to concentrations.

In this section we will discuss the cases for one single site in the target, multiple sites with same affinities and multiple sites with different affinities.

#### 3.1 One-site binding

In the simplest case, where there is only one site per target molecule,  $n$  and  $m$  are 1. It is possible to define the fraction of occupied binding sites ( $\theta$ ) as:

$$v = \frac{L_{\text{bound}}}{P_{\text{total}}} = \frac{[PL]}{[PL] + [P]} \quad (6)$$

Determining  $v$  is often easy in spectrophotometric manipulation as will be discussed later. Given  $v$ , equation 5 can then be solved for  $[PL]$  and the answer substituted into equation 6 to obtain the quantitative 1:1 stoichiometric model:

$$v = \frac{K_a[L]}{1 + K_a[L]} \quad (7)$$

This equation is the 1:1 binding isotherm also known as the Langmuir isotherm or the "direct" plot. Its functional form is a rectangular hyperbola whose midpoint will yield  $K_a$ . Chemical interpretation of 1:1 binding is that the target P has a single "binding site", as has the ligand L; and when the complex PL forms, no further sites are available for the binding of any additional ligand. To test the 1:1 stoichiometry equation 7 may be rearrange into a linear plotting form. Since  $v$  is the bound fraction, then  $1-v$  is the free one:  $(1-v) = 1/(1+K_a[L])$ . Thus  $v/(1-v) = K_a[L]$ , and:

$$\log \frac{v}{1-v} = \log [L] + \log K_a \quad (8)$$

This log-log plot should be linear with a slope of one if the stoichiometry is really 1:1. This is called a Hill plot. Equation 7 can be also rearranged to three different non-logarithmic linear plotting forms. Taking simply the reciprocal of the equation yields the double-reciprocal plot (used by plotting  $1/v$  against  $1/[L]$ ):

$$\frac{1}{v} = \frac{1}{K_a[L]} + 1 \quad (9)$$

In spectroscopic studies this plot is commonly known as the Benesi-Hildebrand plot (Benesi & Hildebrand 1949).

Another plot is that of  $[L]/v$  against  $[L]$  which is expected to be linear:

$$\frac{[L]}{v} = [L] + \frac{1}{K_a} \quad (10)$$

And the third plotting of  $v/[L]$  against  $v$ , sometimes called Scatchard plot (Scatchard 1949):

$$\frac{v}{[L]} = vK_a + K_a \quad (11)$$

Linearity in all of these plots is a necessary condition if the 1:1 model is valid; and from the parameters of equations  $K_a$  can be evaluated. Usually  $v$  is not measured directly but rather some experimental quantity related to it, so that the interpretation of the plots depends on the particular experimental methodology.

### 3.2 Multi-site binding

Most biological systems tend to have more than one binding site, that is the case of many systems of small molecules binding to proteins. In these cases we may consider that  $n$  ligands may bind to a single target molecule. The average number of ligand molecules bound per target molecule ( $b$ ) is defined as:

$$b = \frac{L_{\text{total}} - [L]}{P_{\text{total}}} \quad (12)$$

Assuming that all  $n$  binding sites in the target molecule are identical and independent, it is possible to establish:

$$b = \frac{nk[L]}{1 + k[L]} \quad (13)$$

where  $k$  is the constant for binding to a single site. According to this equation this system follows the hyperbolic function characteristic for the one-site binding model. To define the model  $n$  and  $k$  can be evaluated from a Scatchard plot. The affinity constant  $k$  is an average over all binding sites, it is in fact constant if all sites are truly identical and independent. A stepwise binding constant ( $K_{st}$ ) can be defined which would vary statistically depending on the number of target sites previously occupied. It means that for a target with  $n$  sites will be much easier for the first ligand added to find a binding site than it will be for each successive ligand added. The first ligand would have  $n$  sites to choose while the  $n$ th one would have just one site to bind. The stepwise binding constant can be defined as:

$$K_{st} = \frac{\text{number of free target sites}}{\text{number of bound sites}} \quad k = \frac{n - b + 1}{b} k \quad (14)$$

It is interesting to notice that a deviation from linearity in the Scatchard plot (and to a lesser extent in the Benesi-Hildebrand) gives information on the nature of binding sites. A curved plot denotes that the binding sites are not identical and independent.

### 3.3 Allosteric interactions

Another common situation in biological systems is the cooperative effect, in that case several identical but dependent binding sites are found in the target molecule. It is important to define the effect of the binding of successive ligands to the target to describe the system. A useful model for that issue is the Hill plot (Hill 1910). In this case the number of ligands bound per target molecule will be (take into account that the situation in this system for equation 2 is  $m=1$  and  $n \neq 1$ ):

$$b = \frac{n[PL_n]}{[PL_n] + [P]} \quad (15)$$

if equation 5 is solved for  $[PL_n]$  and substitute into equation 15, then:

$$b = \frac{nK_a[L]^n}{K_a[L]^n + 1} \quad (16)$$

This expression can be rewritten as:

$$\frac{b}{n - b} = K_a[L]^n \quad (17)$$

Note that the fraction of sites bound,  $v$  (see equation 6), is the number of sites occupied,  $b$ , divided by the number of sites available,  $n$ . Then equation 17 becomes:

$$\frac{v}{1 - v} = K_a[L]^n \quad (18)$$

Equation 18 is known as the Hill equation. From the Hill equation we arrive at the Hill plot by taking logarithms at both sides:

$$\log \frac{v}{1 - v} = n_H \log [L] + \log K_a \quad (19)$$

Plotting  $\log(v/(1-v))$  against  $\log[L]$  will yield a straight line with slope  $n_H$  (called the Hill coefficient). The Hill coefficient is a qualitative measure of the degree of cooperativity and it is experimentally less than the actual number of binding sites in the target molecule. When  $n_H > 1$ , the system is said to be positively cooperative, while if  $n_H < 1$ , it is said to be anti-cooperative. Positively cooperative binding means that once the first ligand is bound to its target molecule the affinity for the next ligand increases, on the other hand the affinity for subsequent ligand binding decreases in negatively cooperative (anti-cooperative) systems. In the case of  $n_H = 1$  a non-cooperative binding occurs, here ligand affinity is independent of whether another ligand is already bound or not.

Since equation 19 assumes that  $n_H = n$ , it does not describe exactly the real situation. When a Hill plot is constructed over a wide range of ligand concentrations, the continuity of the plot is broken at the extremes concentrations. In fact, the slope at either end is approximately one. This phenomenon can be easily explained: when ligand concentration is either very low or very high, cooperativity does not exist. For low concentrations it is more probable for individual ligands to find a target molecule "empty" rather than to occupy successive sites on a pre-bound molecule, thus single-binding is happening in this situation. At the other extreme, for high concentrations, every binding-site in the target molecule but one will be filled, thus we find again single-binding situation. The larger the number of sites in a single target molecule is, the wider range of concentrations the Hill plot will show cooperativity.

#### 4. Determination of binding constants

As discussed above the binding constant provides important and interesting information about the system studied. We will present a few of the multiple experimental possibilities to measure this constant (further information could be found in the literature (Johnson *et al.* 1960; Connors 1987; Hirose 2001; Connors&Mecozzi 2010; Pollard 2010)). It is essential to keep in mind some crucial details to be sure to calculate the constants properly: it is important to control the temperature, to be sure that the system has reached the equilibrium and to use the correct equilibrium model. One common mistake that should be avoided is to confuse the total and free concentrations in the equilibrium expression.

Different techniques are commonly used to study the binding of ligands to their targets. These techniques can be classified as calorimetry, spectroscopy and hydrodynamic methods. Hydrodynamic techniques are typically separation methodologies such as different chromatographies, ultracentrifugation or equilibrium dialysis with which free ligand, free target and complex are physically separated from each other at equilibrium, thus concentrations of each can be measured. Spectroscopic methodologies include optical spectroscopy (e.g. absorbance, fluorescence), nuclear magnetic resonance or surface plasmon resonance. Calorimetry includes isothermal titration and differential scanning. Calorimetry and spectroscopy methods allow accurate determination of thermodynamics and kinetics of the binding, as well as can give information about the structure of binding sites.

Once the bound (or free) ligand concentration is measured, the binding proportion can be calculated. Other thermodynamic parameters can be calculated by varying ligand or target concentrations or the temperature of the system.

##### 4.1 Determination of stoichiometry. Continuous variation method.

Since correct reaction stoichiometry is crucial for correct binding constant determination we will study how it can be evaluated. There are different methods of calculating the

stoichiometry: continuous variation method, slope ratio method, mole ratio method, being the first one, the continuous variation method the most popular. In order to determine the stoichiometry by this method the concentration of the produced complex (or any property proportional to it) is plotted versus the mole fraction ligand ( $[L]_{\text{total}}/([P]_{\text{total}}+[L]_{\text{total}})$ ) over a number of titration steps where the sum of  $[P]_{\text{total}}$  and  $[L]_{\text{total}}$  is kept constant ( $\alpha$  changing  $[L]_{\text{total}}$  from 0 to  $\alpha$ ). The maxima of this plot (known as Job's plot, (Job 1928; Ingham 1975)) indicates the stoichiometry of the binding reaction: 1:1 is indicated by a maximum at 0.5 since this value corresponds to  $n/(n+m)$ . For the understanding of the theoretical background of the method, it is important to remember equations 2 and 5; notice that:

$$[P]_{\text{total}} = [P] + m[P_mL_n] \quad (20)$$

$$[L]_{\text{total}} = [L] + n[P_mL_n] \quad (21)$$

$$\alpha = [L]_{\text{total}} + [P]_{\text{total}} \quad (22)$$

$$x = \frac{[L]_{\text{total}}}{[P]_{\text{total}} + [L]_{\text{total}}} \quad (23)$$

$$y = [P_mL_n] \quad (24)$$

Substitution of  $[P]_{\text{total}}$  and  $[L]_{\text{total}}$  by the functions of  $\alpha$  and  $x$  from equation 23 and 24 yields:

$$[P]_{\text{total}} = \alpha - \alpha x \quad (25)$$

$$[L]_{\text{total}} = \alpha x \quad (26)$$

from equations 2, 5, 20, 21, 24, 25, 26:

$$y = K_a(\alpha - my - \alpha x)^m (\alpha x - ny)^n \quad (27)$$

Equation 27 is differentiated, and the  $dy/dx$  substituted by zero to obtain the  $x$ -coordinate at the maximum:

$$x = \frac{n}{n+m} \quad (28)$$

This equation shows the correlation between stoichiometry and the  $x$ -coordinate at the maximum in Job's plot. That's why a maximum at  $x = 0.5$  means a 1:1 stoichiometry ( $n = m = 1$ ). In the case of 1:2 the maximum would be at  $x = 1/3$ .

## 4.2 Calorimetry

Isothermal titration calorimetry (ITC) is a useful tool for the characterization of thermodynamics and kinetics of ligands binding to macromolecules. With this method the rate of heat flow induced by the change in the composition of the target solution by titration of a ligand (or vice versa) is measured. This heat is proportional to the total amount of binding. Since the technique measures heat directly, it allows simultaneous determination of the stoichiometry ( $n$ ), the binding constant ( $K_a$ ) and the enthalpy ( $\Delta H^0$ ) of binding. The free energy ( $\Delta G^0$ ) and the entropy ( $\Delta S^0$ ) are easily calculated from  $\Delta H^0$  and  $K_a$ . Note that the binding constant is related to the free energy by:

$$\Delta G^0 = -RT \ln K_a \quad (29)$$



where  $R$  is the gas constant and  $T$  the absolute temperature. The free energy can be dissected into enthalpic and entropic components by:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 \quad (30)$$

On the other hand, the heat capacity ( $\Delta C_p$  -p subscript indicates that the system is at constant pressure-) of a reaction predicts the change of  $\Delta H^0$  and  $\Delta S^0$  with temperature and can be expressed as:

$$\Delta C_p = \frac{\Delta H^0_{T_2} - \Delta H^0_{T_1}}{T_2 - T_1} \quad (31)$$

or

$$\Delta C_p = \frac{\Delta S^0_{T_2} - \Delta S^0_{T_1}}{\ln \frac{T_2}{T_1}} \quad (32)$$

In an ITC experiment a constant temperature is set, a precise amount of ligand is added to a known target molecule concentration and the heat difference is measured between reference and sample cells. To eliminate heats of mixing effects, the ligand and target as well as the reference cell contain identical buffer composition. Subsequent injections of ligand are done until no further heat of binding is observed (all sites are then bound with ligand molecules). The remaining heat generated now comes from dilution of ligand into the target solution. Data should be corrected for the heat of dilution. The heat of binding calculated for every injection is plotted versus the molar ratio of ligand to protein.  $K_a$  is related to the curve shape and binding capacity ( $n$ ) determined from the ratio of ligand to target at the equivalence point of the curve. Data must be fitted to a binding model. The type of binding must be known from other experimental techniques. Here, we will study the simplest model with a single site. Equations 6 and 7 can be rearranged to find the following relation between  $v$  and  $K_a$ :

$$K_a = \frac{v}{(1-v)[L]} \quad (33)$$

Total ligand concentration is known and can be represented as (remember that we are assuming  $m=n=1$ ):

$$[L]_{\text{total}} = [L] + v[P]_{\text{total}} \quad (34)$$

Combining equations 33 and 34 gives:

$$v^2 - \left( \frac{[L]_{\text{total}}}{[P]_{\text{total}}} + \frac{1}{K_a[P]_{\text{total}}} + 1 \right) v + \frac{[L]_{\text{total}}}{[P]_{\text{total}}} = 0 \quad (35)$$

Solving for  $v$ :

$$v = \frac{1}{2} \left[ \left( \frac{[L]_{\text{total}}}{[P]_{\text{total}}} + \frac{1}{K_a[P]_{\text{total}}} + 1 \right) - \sqrt{\left( \frac{[L]_{\text{total}}}{[P]_{\text{total}}} + \frac{1}{K_a[P]_{\text{total}}} + 1 \right)^2 - \frac{4[L]_{\text{total}}}{[P]_{\text{total}}}} \right] \quad (36)$$

The total heat content ( $Q$ ) in the sample cell at volume ( $V$ ) can be defined as:

$$Q = [PL]\Delta H^0V = v[P]_{\text{total}}\Delta H^0V \quad (37)$$

where  $\Delta H^0$  is the heat of binding of the ligand to its target. Substituting equation 36 into 37 yields:

$$Q = \frac{[P]_{\text{total}} \Delta H^0 V}{2} \left[ \left( \frac{[L]_{\text{total}}}{[P]_{\text{total}}} + \frac{1}{K_a [P]_{\text{total}}} + 1 \right) - \sqrt{\left( \frac{[L]_{\text{total}}}{[P]_{\text{total}}} + \frac{1}{K_a [P]_{\text{total}}} + 1 \right)^2 - \frac{4 [L]_{\text{total}}}{[P]_{\text{total}}}} \right] \quad (38)$$

Therefore  $Q$  is a function of  $K_a$  and  $\Delta H^0$  (and  $n$ , but here we considered it as 1 for simplicity) since  $[P]_{\text{total}}$ ,  $[L]_{\text{total}}$  and  $V$  are known for each experiment.

### 4.3 Optical spectroscopy

The goal to be able to determine binding affinity is to measure the equilibrium concentration of the species implied over a range of concentrations of one of the reactants (P or L). Measuring one of them should be sufficient as total concentrations are known and therefore the others can be calculated by difference from total concentrations and measured equilibrium concentration of one of the species. Plotting the concentration of the complex (PL) against the free concentration of the varying reactant, the binding constant could be calculated.

#### 4.3.1 Absorbance

As an example a 1:1 stoichiometry model will be shown, wherein the Lambert-Beer law is obeyed by all the reactants implied. To use this technique we should ensure that the complex (PL) has a significantly different absorption spectrum than the target molecule (P) and a wavelength at which both molar extinction coefficients are different should be selected. At these conditions the absorbance of the target molecule in the absence of ligand will be:

$$\text{Abs}_0 = \epsilon_P l [P]_{\text{total}} \quad (39)$$

If ligand is added to a fixed total target concentration, the absorbance of the mix can be written as:

$$\text{Abs}_{\text{mix}} = \epsilon_P l [P] + \epsilon_L l [L] + \epsilon_{PL} l [PL] \quad (40)$$

Since  $[P]_{\text{total}} = [P] + [PL]$  and  $[L]_{\text{total}} = [L] + [PL]$ , equation 40 can be rewritten as:

$$\text{Abs}_{\text{mix}} = \epsilon_P l [P]_{\text{total}} + \epsilon_L l [L]_{\text{total}} + \Delta \epsilon l [PL] \quad (41)$$

where  $\Delta \epsilon = \epsilon_{PL} - \epsilon_P - \epsilon_L$ . If the blank solution against which samples are measured contains  $[L]_{\text{total}}$ , then the observed absorbance would be:

$$\text{Abs}_{\text{obs}} = \epsilon_P l [P]_{\text{total}} + \Delta \epsilon l [PL] \quad (42)$$

Subtracting equation 39 from 42 and incorporating  $K_a$  (equation 5):

$$\Delta \text{Abs} = K_a \Delta \epsilon l [P] [L] \quad (43)$$

$[P]_{\text{total}}$  can be written as  $[P]_{\text{total}} = [P](1 + K_a [L])$  which included in equation 43 yields:

$$\frac{\Delta \text{Abs}}{l} = \frac{[P]_{\text{total}} K_a \Delta \epsilon [L]}{1 + K_a [L]} \quad (44)$$

which is the direct plot expressed in terms of spectrophotometric observation. Note that the dependence of  $\Delta\text{Abs}/l$  on  $[\text{L}]$  is the same as the one shown in equation 7.

The free ligand concentration is actually unknown. The known concentrations are  $[\text{P}]_{\text{total}}$  to which a known  $[\text{L}]_{\text{total}}$  is added. In a similar way as shown above for  $[\text{P}]_{\text{total}}$ ,  $[\text{L}]_{\text{total}}$  can be written as:

$$[\text{L}]_{\text{total}} = [\text{L}] \frac{[\text{P}]_{\text{total}} K_a [\text{L}]}{1 + K_a [\text{L}]} \quad (45)$$

From equations 44 and 45 a complete description of the system is obtained. If  $[\text{L}]_{\text{total}} \gg [\text{P}]_{\text{total}}$  we will have that  $[\text{L}]_{\text{total}} \approx [\text{L}]$  from equation 45, equation 44 can be then analysed with this approximation. With this first rough estimate of  $K_a$ , equation 45 can be solved for the  $[\text{L}]$  value for each  $[\text{L}]_{\text{total}}$ . These values can be used in equation 44 to obtain an improved estimation of  $K_a$ , and this process should be repeated until the solution for  $K_a$  reaches a constant value. Equation 44 can be solved graphically using any of the plots presented in section 3.1.

#### 4.3.2 Fluorescence

Fluorescence spectroscopy is a widely used tool in biochemistry due to its ease, sensitivity to local environmental changes and ability to describe target-ligand interactions qualitatively and quantitatively in equilibrium conditions. In this technique the fluorophore molecule senses changes in its local environment. To analyse ligand-target interactions it is possible to take advantage of the nature of ligands, exceptionally we can find molecules which are essentially non or weakly fluorescent in solution but show intense fluorescence upon binding to their targets (that is the case, for example, of colchicines and some of its analogues). Fluorescence moieties such as fluorescein can be also attached to naturally non-fluorescent ligands to make use of these methods. The fluorescent dye may influence the binding, so an essential control with any tagged molecule is a competition experiment with the untagged molecule. Finally, in a few favourable cases the intrinsic tryptophan fluorescence of a protein changes when a ligand binds, usually decreasing (fluorescence quenching). Again, increasing concentrations of ligand to a fixed concentration of target (or vice versa) are incubated at controlled temperature and fluorescence changes measured until saturation is reached. Binding constant can be determined by fitting data according to equation 11 (Scatchard plot). From fluorescence data ( $F$ ),  $\nu$  can be calculated from the relationship:

$$\nu = \frac{F_{\text{max}} - F}{F_{\text{max}}} \quad (46)$$

If free ligand has an appreciable fluorescence as compared to ligand bound to its target, then the fluorescence enhancement factor ( $Q$ ) should be determined.  $Q$  is defined as (Mas & Colman 1985):

$$Q = \frac{F_{\text{bound}}}{F_{\text{free}}} - 1 \quad (47)$$

To determine it, a reverse titration should be done. The enhancement factor can be obtained from the intercept of linear plot of  $1/((F/F_0)-1)$  against  $1/P$ , where  $F$  and  $F_0$  are the observed fluorescence in the presence and absence of target, respectively. Once it is known, the concentration of complex can be determined from a fluorescence titration experiment using:

$$[PL] = [L]_{\text{total}} \frac{(F/F_0) - 1}{Q - 1} \quad (48)$$

Thus the binding constant can be determined from the Scatchard plot as described above.

### 4.3.3 Fluorescence anisotropy

Fluorescence anisotropy measures the rotational diffusion of a molecule. The effective size of a ligand bound to its target usually increases enormously, thus restricting its motion considerably. Changes in anisotropy are proportional to the fraction of ligand bound to its target. Using suitable polarizers at both sides of the sample cuvette, this property can be measured. In a titration experiment similar to the ones described above, the fraction of ligand bound ( $X_L = [PL]/[L]_{\text{total}}$ ) is determined from:

$$X_L = \frac{r - r_0}{r_{\text{max}} - r_0} \quad (49)$$

where  $r$  is the anisotropy of ligand in the presence of the target molecule,  $r_0$  is the anisotropy of ligand in the absence of target and  $r_{\text{max}}$  is the anisotropy of ligand fully bound to its target (note that equation 49 can be used only in the case where ligand fluorescence intensity does not change, otherwise appropriate corrections should be done, see (Lakowicz 1999)).  $[P]$  can be calculated from:

$$[P] = [P]_{\text{total}} - X_L [L]_{\text{total}} \quad (50)$$

The binding constant can be determined from the hyperbola:

$$X_L = \frac{K_a [P]}{1 + K_a [P]} \quad (51)$$

### 4.4 Competition methods

The characterization of a ligand binding let us determine the binding constant of any other ligand competing for the same binding site. Measurements of ligand (L), target (P), reference ligand (R) and both complexes (PR and PL) concentrations in the equilibrium permit the calculation of the binding constant ( $K_L$ ) from equation 53 (see below) as the binding constant of the reference ligand ( $K_R$ ) is already known.



$$K_L = K_R \frac{[PL][R]}{[L][PR]} \quad (53)$$

In the case that the reference ligand has been characterized due to the change of a ligand physical property (i.e. fluorescence, absorbance, anisotropy) upon binding, would permit us also following the displacement of this reference ligand from its site by competition with a ligand „blind“ to this signal (Diaz & Buey 2007). In this kind of experiment equimolar concentrations of the reference ligand and the target molecule are incubated, increasing concentrations of the problem ligand added and the appropriate signal measured. It is possible then to determine the concentration of ligand at which half the reference ligand is bound to its site ( $EC_{50}$ ). Thus  $K_L$  is calculated from:

$$K_L = \frac{1 + [R]K_R}{EC_{50}} \quad (54)$$

## 5. Drug optimization

Microtubule stabilizing agents (MSA) comprise a class of drugs that bind to microtubules and stabilize them against disassembly. During the last years, several of these compounds have been approved as anticancer agents or submitted to clinical trials. That is the case of taxanes (paclitaxel, docetaxel) or epothilones (ixabepilone) as well as discodermolide (reviewed in (Zhao *et al.* 2009)). Nevertheless, anticancer chemotherapy has still unsatisfactory clinical results, being one of the major reasons for it the development of drug resistance in treated patients (Kavallaris 2010). Thus one interesting issue in this field is drug optimization with the aim of improving the potential for their use in clinics: minimizing side-effects, overcoming resistances or enhancing their potency.

Our group has studied the influence of different chemical modifications on taxane and epothilone scaffolds in their binding affinities and the consequently modifications in ligand properties like cytotoxicity. The results from these studies firmly suggest thermodynamic parameters as key clues for drug optimization.

### 5.1 Epothilones

Epothilones are one of the most promising natural products discovered with paclitaxel-like activity. Their advantages come from the fact that they can be produced in large amounts by fermentation (epothilones are secondary metabolites from the myxobacterium *Sorangium celulosum*), their higher solubility in water, their simplicity in molecular architecture which makes possible their total synthesis and production of many analogs, and their effectiveness against multi-drug resistant cells due to they are worse substrates for P-glycoprotein.

The structure affinity-relationship of a group of chemically modified epothilones was studied. Epothilones derivatives with several modifications in positions C12 and C13 and the side chain in C15 were used in this work.

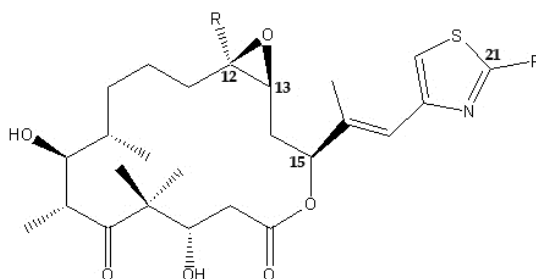


Fig. 1. Epothilone atom numbering.

Epothilone binding affinities to microtubules were measured by displacement of Flutax-2, a fluorescent taxoid probe (fluorescein tagged paclitaxel). Both epothilones A and B binding constants were determined by direct sedimentation which further validates Flutax-2 displacement method.

All compounds studied are related by a series of single group modifications. The measurement of the binding affinity of such a series can be a good approximation of the incremental binding energy provided by each group. Binding free energies are easily calculated from binding constants applying equation 29. The incremental free energies ( $\Delta G^0$ ) change associated with the modification of ligand L into ligand S is defined as:

$$\Delta\Delta G^0(L \rightarrow S) = \Delta G^0(L) - \Delta G^0(S) \quad (55)$$

These incremental binding energies were calculated for a collection of 20 different epothilones as reported in (Buey *et al.* 2004).

Site	Modification	Compounds	$\Delta\Delta G$
C15	<i>S</i> → <i>R</i>	4 → 17	~ 27
		7 → 18	~ 27
		14 → 16	17.8 ± 0.3
	Thiazole → Pyridine	5 → 7	-2.9 ± 0.2
		6 → 8	-2.1 ± 0.3
		14 → 4	-0.2 ± 0.4
		16 → 17	~ 9.4
C21	Methyl → Thiomethyl	2 → 3	-2.8 ± 0.8
		5 → 10	-5.9 ± 0.6
		6 → 11	-3.6 ± 0.3
		8 → 12	2.6 ± 0.3
	Methyl → Hydroxymethyl	8 → 9	1.4 ± 0.3
5-Thiomethyl-pyridine → 6-Thiomethyl-pyridine	12 → 13	4.1 ± 0.5	
C12	<i>S</i> → <i>R</i>	4 → 7	-2.1 ± 0.3
		14 → 5	0.6 ± 0.3
		17 → 18	~ -2
		19 → 11	9.0 ± 0.6
		20 → 8	1.9 ± 0.4
	Epoxide → Cyclopropyl	1 → 14	-4.7 ± 0.4
		3 → 19	-5.4 ± 0.8
	Cyclopropyl → Cyclobutyl	5 → 15	4.1 ± 0.2
	<i>S</i> H → Methyl	1 → 2	-8.1 ± 0.6
		4 → 20	-1.8 ± 0.5
	<i>R</i> H → Methyl	5 → 6	0.4 ± 0.3
		7 → 8	1.2 ± 0.2
		10 → 11	2.7 ± 0.7

Table 1. Incremental binding energies of epothilone analogs to microtubules. ( $\Delta\Delta G$  in kJ/mol at 35°C). Data from (Buey *et al.* 2004).

The data in table 1 show that the incremental binding free energy changes of single modifications give a good estimation of the binding energy provided by each group. Moreover, the effect of the modifications is accumulative, resulting the epothilone derivative with the most favourable modifications (a thiomethyl group at C21 of the thiazole side chain, a methyl group at C12 in the *S* configuration, a pyridine side chain with C15 in the *S* configuration and a cyclopropyl moiety between C12 and C13) the one with the highest affinity of all the compounds studied ( $K_a$  2.1±0.4 × 10<sup>10</sup> M<sup>-1</sup> at 35°C).

The study of these compounds showed also a correlation between their citotoxic potencial and their affinities to microtubules. The plot of log IC<sub>50</sub> in human ovarian carcinoma cells versus log K<sub>a</sub> shows a good correlation (figure 2), suggesting binding affinity as an important parameter affecting citotoxicity.

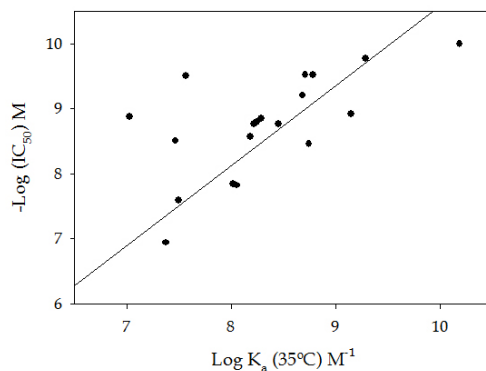


Fig. 2. Dependence of the  $IC_{50}$  of epothilone analogs against 1A9 cells on their  $K_a$  to microtubules. Data from (Buey *et al.* 2004).

## 5.2 Taxanes

Paclitaxel and docetaxel are widely used in the clinics for the treatment of several carcinoma and Kaposi's sarcoma. Nevertheless, their effectiveness is limited due to the development of resistance, being its main cause the overexpression and drug efflux activity of transmembrane proteins like P-glycoprotein (Shabbits *et al.* 2001).

We have studied the thermodynamics of binding of a set of nearly 50 taxanes to crosslinked stabilized microtubules with the aim to quantify the contributions of single modifications at four different locations of the taxane scaffold (C2, C13, C7 and C10).

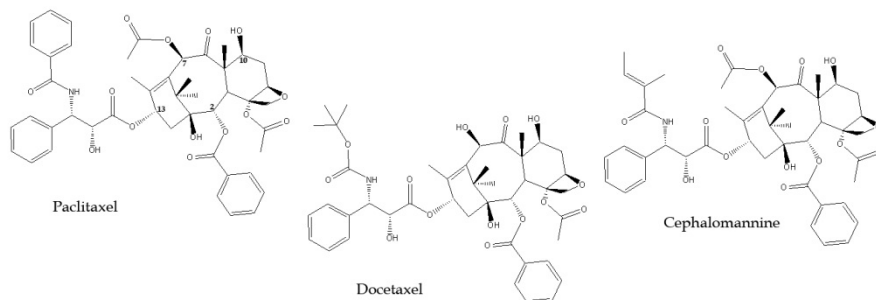


Fig. 3. Taxanes head compounds. Atom numbering

Once confirmed that all the compounds were paclitaxel-like MSA, their affinities were measured using the same competition method mentioned above (section 5.1. displacement of Flutax-2). Seven of the compounds completely displaced Flutax-2 at equimolar concentrations indicating that they have very high affinities and so they are in the limit of the range to be accurately calculated by this method (Diaz&Buey 2007). The affinities of these compounds were then measured using a direct competition experiment with epothilone-B, a higher-affinity ligand ( $K_a$   $75.0 \times 10^7$  at  $35^\circ\text{C}$  compared with  $3.0 \times 10^7$  for Flutax-2). With all the binding constants determined at a given temperature, it is possible to determine the changes in binding free energy caused by every single modification as discussed above for epothilones (table 2).

Site	Modification	Compounds	$\Delta\Delta G$	Average
C2	benzoyl → benzylether	T → 25	13.2	+13.0 ± 0.2
		21 → 24	12.8	
	benzoyl → benzylsulphur	T → 27	13.6	+15.9 ± 2.3
		21 → 26	18.1	
	benzoyl → benzylamine	T → 38	18.6	+20.1 ± 1.5
		21 → 39	21.6	
	benzoyl → thiobenzoyl	T → 23	19.6	+15.9 ± 3.8
		21 → 22	12.1	
	benzoyl → benzamide	21 → 42	19.2	
	benzamide → 3-methoxy-benzamide	42 → 43	-3.4	
	benzamide → 3-Cl-benzamide	42 → 44	5.3	
	benzoyl → 3 methyl- 2 butenoyl	1 → 2	6.2	
	benzoyl → 3 methyl- 3 butenoyl	1 → 3	4.9	
	benzoyl → 2(E)-butenoyl	1 → 9	7.3	
	benzoyl → 3 methyl- butanoyl	1 → 10	6.3	
	benzoyl → 2-debenzoyl-1,2-carbonate	C → 16	5.8	
	benzoyl → 3-azido-benzoyl	1 → 4	-8	-11.2 ± 1.3
		T → 12	-13.9	
		C → 14	-12.2	
		18 → 20	-10.6	
	benzoyl → 3-methoxy-benzoyl	1 → 5	-6.2	-7.2 ± 0.6
		T → 11	-8.3	
		C → 13	-8.1	
		18 → 19	-6.3	
	benzoyl → 3-Cl-benzoyl	1 → 6	-3.1	
	benzoyl → 3-Br-benzoyl	1 → 34	-2.3	
	benzoyl → 3-I-benzoyl	1 → 30	-3.3	
	benzoyl → 3-ciano-benzoyl	1 → 7	0.6	
	benzoyl → 3-methyl-benzoyl	1 → 8	0	
	benzoyl → 3-hydroxymethyl-benzoyl	1 → 36	7.2	
benzoyl → 3-hydroxy-benzoyl	18 → 37	9.2		
3-Cl-benzoyl → 2,4-di-Cl-benzoyl	6 → 29	4.8		
benzoyl → 2,4-di-F-benzoyl	1 → 28	2.7		
3-methoxy-benzoyl → 2,5-di-methoxy-benzoyl	5 → 35	4.6		
benzoyl → 2-thienoyl	1 → 31	4.1		
benzoyl → 3-thienoyl	1 → 32	1.8		
benzoyl → 6-carboxy-pyran-2-one	1 → 41	8.1		
C13	paclitaxel → cephalomannine	T → C	1.9	+2.0 ± 0.2
		11 → 13	1.9	
		12 → 14	1.6	
		15 → 17	2.4	
	paclitaxel → docetaxel	23 → 22	-1.7	-3.2 ± 0.9
		25 → 24	-6.2	



		27 → 26	-1.3	
		38 → 39	-2.8	
		T → 21	-4.2	
	cephalomannine → docetaxel	C → 21	-3.8	-5.6 ± 1.1
		17 → D	-7.7	
		20 → 40	-5.2	
C10	acetyl → hydroxyl	T → 15	-1.3	-1.7 ± 0.8
		C → 17	-0.7	
		21 → D	-3.2	
	propionyl → hydroxyl	18 → 17	0.9	
	acetyl → propionyl	C → 18	-1.6	-0.5 ± 0.4
	13 → 19	0.2		
	14 → 20	0		
C7	propionyl → hydroxyl	17 → 1	-1.6	

Table 2. Incremental binding energies of taxane analogs to microtubules. ( $\Delta\Delta G$  in kJ/mol at 35°C). Data from (Matesanz *et al.* 2008).

In this way, it is possible to select the most favourable substituents at the positions studied and design optimized taxanes. According to the data obtained, the optimal taxane should have the docetaxel side chain at C13, a 3-N<sub>3</sub>-benzoyl at C2, a propionyl at C10, and a hydroxyl at C7. From compound 1 with a binding energy of -39.4 kJ/mol, the modifications selected would increase the binding affinity in -5.6 kJ/mol from the change of the cephalomannine side chain at C13 to the docetaxel one, -11.2 kJ/mol from the introduction of 3-N<sub>3</sub>-benzoyl instead of benzoyl at C2, -1.6 kJ/mol from the substitution of a propionyl at C7 with a hydroxyl, and -0.9 kJ/mol from the change of a hydroxyl at C10 to a propionyl. Thus, this optimal taxane would have a predicted  $\Delta G$  at 35°C of -58.7 kJ/mol. This molecule was synthesized (compound 40) and its binding affinity measured using the epothilone-B displacement method and the value obtained is in good correspondence with the predicted one:  $K_a = 6.28 \pm 0.15 \times 10^9 \text{ M}^{-1}$ ;  $\Delta G = -57.7 \pm 0.1 \text{ kJ/mol}$  (Matesanz *et al.* 2008). This value means a 500-fold increment over the paclitaxel affinity.

It is also possible to check the influence of the modifications on the cytotoxic activity determining the IC<sub>50</sub> of each compound in the human ovarian carcinoma cells A2780 and their MDR counterparts (A2780AD). The plots of log IC<sub>50</sub> versus log K<sub>a</sub> (figure 4) indicate that, as in the case of epothilones, both magnitudes are related, and the binding affinity acts as a good predictor of cytotoxicity. In this type of MDR cells the high-affinity drugs are circa 100-fold more cytotoxic than the clinically used taxanes (paclitaxel and docetaxel) and exhibit very low resistance indexes.

The plot of log resistance index against log K<sub>a</sub> shows a bell-shaped curve (figure 5). Resistance index present a maximum for taxanes with similar affinities for microtubules and P-glycoprotein, then rapidly decreases when the affinity for microtubules either increases or decreases. To find an explanation for this behaviour we should note that the intracellular free concentration of the high-affinity compounds will be low. To be pumped out by P-glycoprotein ligands must first bind it, so ligand outflow will decrease with lower free ligand concentrations (discussed in (Matesanz *et al.* 2008)). In the case of the low-affinity drugs, the concentrations needed to exert their cytotoxicity are so high that the pump gets saturated and cannot effectively reduced the intracellular free ligand concentration.

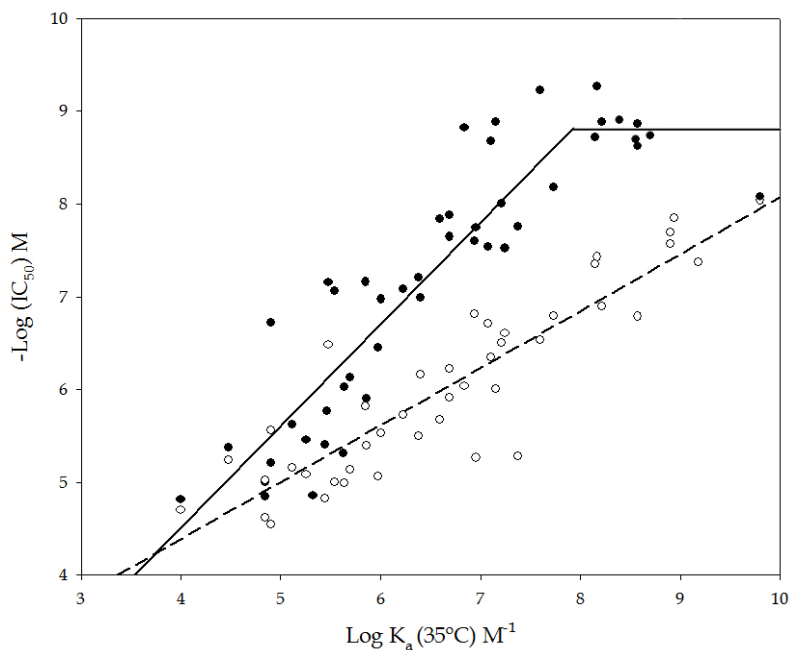


Fig. 4. Dependence of the  $IC_{50}$  of taxane analogs against A2780 non-resistant cells (black circles, solid line) and A2780AD resistant cells (white circles, dashed line) on their  $K_a$  to microtubules. Data from (Matesanz *et al.* 2008).

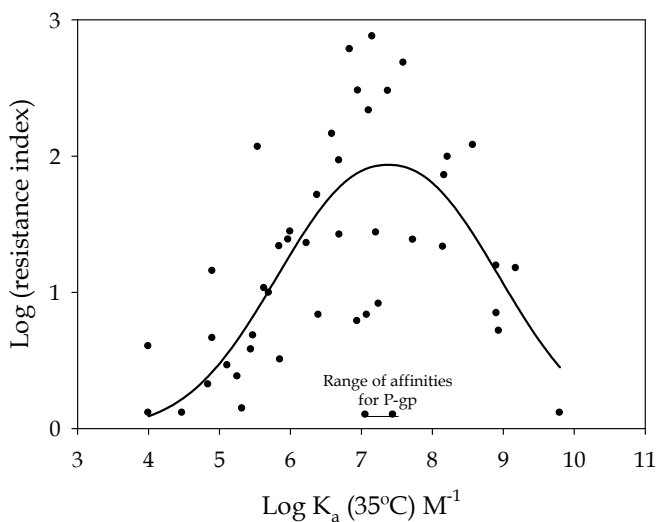


Fig. 5. Dependence of the resistance index of the A2780AD MDR cells on the  $K_a$  of the taxanes to microtubules. Data from (Yang *et al.* 2007; Matesanz *et al.* 2008).

## 6. Conclusion

We found a correlation between binding affinities of paclitaxel-like MSA to microtubules and their cytotoxicities in tumoral cells both MDR and non-resistant. The results with taxanes further validate the binding affinity approach as a tool to be used in drug optimization as it was previously discussed for the case of epothilones. Moreover, from the thermodynamic data we could design novel high-affinity taxanes with the ability to overcome resistance in P-glycoprotein overexpressing cells. Anyway, there is a limit concentration below which MSA are not able to kill cells (discussed in (Matesanz *et al.* 2008)), the highest-affinity compounds studied have no dramatically better cytotoxicities than paclitaxel or docetaxel have. Thus, the goal is not to find the drug with the highest cytotoxicity possible but rather to find one able to overcome resistances. The study of taxanes indicates that increased drug affinity could be an improvement in this direction. The extreme example of that come from the covalent binding of cyclostreptin (Buey *et al.* 2007) (that might be considered as infinite affinity) having a resistance index close to one. However, in the case of chemically diverse paclitaxel-like MSA, the inhibition of cell proliferation correlates better with enthalpy change than with binding constants (Buey *et al.* 2005) suggesting that favourable enthalpic contributions to the binding are important to improve drug activity as it has been shown for statins and HIV protease inhibitors (Freire 2008).

## 7. References

- Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts & J. D. Watson, Eds. (1994). *Molecular Biology of the Cell*. New York, Garland Science.
- Benesi, H. A. & J. H. Hildebrand (1949). "A Spectrophotometric Investigation of the Interaction of Iodine with Aromatic Hydrocarbons." *Journal of the American Chemical Society* 71(8): 2703-2707.
- Buey, R. M., I. Barasoain, E. Jackson, A. Meyer, P. Giannakakou, I. Paterson, S. Mooberry, J. M. Andreu & J. F. Diaz (2005). "Microtubule interactions with chemically diverse stabilizing agents: thermodynamics of binding to the paclitaxel site predicts cytotoxicity." *Chem Biol* 12(12): 1269-1279.
- Buey, R. M., E. Calvo, I. Barasoain, O. Pineda, M. C. Edler, R. Matesanz, G. Cerezo, C. D. Vanderwal, B. W. Day, E. J. Sorensen, J. A. Lopez, J. M. Andreu, E. Hamel & J. F. Diaz (2007). "Cyclostreptin binds covalently to microtubule pores and luminal taxoid binding sites." *Nat Chem Biol* 3(2): 117-125.
- Buey, R. M., J. F. Diaz, J. M. Andreu, A. O'Brate, P. Giannakakou, K. C. Nicolaou, P. K. Sasmal, A. Ritzen & K. Namoto (2004). "Interaction of epothilone analogs with the paclitaxel binding site: relationship between binding affinity, microtubule stabilization, and cytotoxicity." *Chem Biol* 11(2): 225-236.
- Connors, K. A., Ed. (1987). *Binding Constants: The Measurement of Molecular Complex Stability*. New York, Wiley-Interscience.
- Connors, K. A. & S. Mecozzi, Eds. (2010). *Thermodynamics of Pharmaceutical Systems. An Introduction to Theory and Applications*. New York, Wiley-Interscience.
- Diaz, J. F. & R. M. Buey (2007). "Characterizing ligand-microtubule binding by competition methods." *Methods Mol Med* 137: 245-260.

- Freire, E. (2008). "Do enthalpy and entropy distinguish first in class from best in class?" *Drug Discovery Today* 13(19-20): 869-874.
- Hill, A. V. (1910). "The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves." *The Journal of Physiology* 40(Suppl): iv-vii.
- Hirose, K. (2001). "A Practical Guide for the Determination of Binding Constants." *Journal of Inclusion Phenomena and Macrocyclic Chemistry* 39(3): 193-209.
- Ingham, K. C. (1975). "On the application of Job's method of continuous variation to the stoichiometry of protein-ligand complexes." *Analytical Biochemistry* 68(2): 660-663.
- Job, P. (1928). "Formation and stability of inorganic complexes in solution." *Annali di Chimica* 9: 113-203.
- Johnson, I. S., H. F. Wright, G. H. Svoboda&J. Vlantis (1960). "Antitumor principles derived from Vinca rosea Linn. I. Vincalukoblastine and leurosine." *Cancer Res* 20: 1016-1022.
- Kavallaris, M. (2010). "Microtubules and resistance to tubulin-binding agents." *Nat Rev Cancer* 10(3): 194-204.
- Lakowicz, J. R. (1999). *Principles of fluorescence spectroscopy*. New York, Kluwer Academic/ Plenum Publishers.
- Mas, M. T.&R. F. Colman (1985). "Spectroscopic studies of the interactions of coenzymes and coenzyme fragments with pig heart oxidized triphosphopyridine nucleotide specific isocitrate dehydrogenase." *Biochemistry* 24(7): 1634-1646.
- Matesanz, R., I. Barasoain, C. G. Yang, L. Wang, X. Li, C. de Ines, C. Coderch, F. Gago, J. J. Barbero, J. M. Andreu, W. S. Fang&J. F. Diaz (2008). "Optimization of taxane binding to microtubules: binding affinity dissection and incremental construction of a high-affinity analog of paclitaxel." *Chem Biol* 15(6): 573-585.
- Ohtaka, H.&E. Freire (2005). "Adaptive inhibitors of the HIV-1 protease." *Progress in Biophysics and Molecular Biology* 88(2): 193-208.
- Ohtaka, H., S. Muzammil, A. Schön, A. Velazquez-Campoy, S. Vega&E. Freire (2004). "Thermodynamic rules for the design of high affinity HIV-1 protease inhibitors with adaptability to mutations and high selectivity towards unwanted targets." *The International Journal of Biochemistry & Cell Biology* 36(9): 1787-1799.
- Pollard, T. D. (2010). "A Guide to Simple and Informative Binding Assays." *Mol. Biol. Cell* 21(23): 4061-4067.
- Ruben, A. J., Y. Kiso&E. Freire (2006). "Overcoming Roadblocks in Lead Optimization: A Thermodynamic Perspective." *Chemical Biology & Drug Design* 67(1): 2-4.
- Scatchard, G. (1949). "The attractions of proteins for small molecules and ions." *Annals of the New York Academy of Sciences* 51(4): 660-672.
- Shabbits, J. A., R. Krishna&L. D. Mayer (2001). "Molecular and pharmacological strategies to overcome multidrug resistance." *Expert Rev Anticancer Ther* 1(4): 585-594.
- Yang, C. G., I. Barasoain, X. Li, R. Matesanz, R. Liu, F. J. Sharom, D. L. Yin, J. F. Diaz&W. S. Fang (2007). "Overcoming Tumor Drug Resistance with High-Affinity Taxanes: A SAR Study of C2-Modified 7-Acyl-10-Deacetyl Cephalomannines." *ChemMedChem* 2(5): 691-701.
- Zhao, Y., W.-S. Fang&K. Pors (2009). "Microtubule stabilising agents for cancer chemotherapy." *Expert Opinion on Therapeutic Patents* 19(5): 607-622.

# On the Chlorination Thermodynamics

Brocchi E. A. and Navarro R. C. S.

*Pontifical Catholic University of Rio de Janeiro  
Brazil*

## 1. Introduction

Chlorination roasting has proven to be a very important industrial route and can be applied for different purposes. Firstly, the chlorination of some important minerals is a possible industrial process for producing and refining metals of considerable technological importance, such as titanium and zirconium. Also, the same principle is mentioned as a possible way of recovering rare earth from concentrates (Zang et al., 2004) and metals, of considerable economic value, from different industrial wastes, such as, tailings (Cechi et al., 2009), spent catalysts (Gabalah & Djona, 1995), slags (Brocchi & Moura, 2008) and fly ash (Murase et al., 1998). The chlorination processes are also presented as environmentally acceptable (Neff, 1995, Mackay, 1992).

In general terms the chlorination can be described as reaction between a starting material (mineral concentrate or industrial waste) with chlorine in order to produce some volatile chlorides, which can then be separated by, for example, selective condensation. The most desired chloride is purified and then used as a precursor in the production of either the pure metal (by reacting the chloride with magnesium) or its oxide (by oxidation of the chloride).

The chlorination reaction has been studied on respect of many metal oxides (Micco et. al., 2011; Gaviria & Bohe, 2010; Esquivel et al., 2003; Oheda et al., 2002) as this type of compound is the most common in the mentioned starting materials. Although some basic thermodynamic data is enclosed in these works, most of them are related to kinetics aspects of the gas - solid reactions. However, it is clear that the understanding of the equilibrium conditions, as predicted by classical thermodynamics, of a particular oxide reaction with chlorine can give strong support for both the control and optimization of the process. In this context, the impact of industrial operational variables over the chlorination efficiency, such as the reaction temperature and the reactors atmosphere composition, can be theoretically appreciated and then quantitatively predicted. On that sense, some important works have been totally devoted to the thermodynamics of the chlorination and became classical references on the subject (Kellog, 1950; Patel & Jere, 1960; Pilgrim & Ingraham, 1967; Sano & Belton, 1980).

Originally, the approach applied for the study of chemical equilibrium studies was based exclusively on  $\Delta G_r^0 \times T$  and predominance diagrams. Nowadays, however, advances in computational thermodynamics enabled the development of softwares that can perform more complex calculations. This approach, together with the one accomplished by simpler techniques, converge to a better understanding of the intimate nature of the equilibrium states for the reaction system of interest. Therefore, it is understood that the time has come

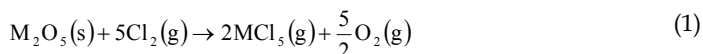
for a review on chlorination thermodynamics which can combine its basic aspects with a now available new kind of approach.

The present chapter will first focus on the thermodynamic basis necessary for understanding the nature of the equilibrium states achievable through chlorination reactions of metallic oxides. Possible ways of graphically representing the equilibrium conditions are discussed and compared. Moreover, the chlorination of  $V_2O_5$ , both in the absence as with the presence of graphite will be considered. The need of such reducing agent is clearly explained and discussed. Finally, the equilibrium conditions are appreciated through the construction of graphics with different levels of complexity, beginning with the well known  $\Delta G_f^\circ \times T$  diagrams, and ending with gas phase speciation diagrams, rigorously calculated through the minimization of the total Gibbs energy of the system.

## 2. Chemical reaction equilibria

The equilibrium state achieved by a system where a group of chemical reactions take place simultaneously can be entirely modeled and predicted by applying the principles of classical thermodynamics.

Supposing that we want to react some solid transition metal oxide, say  $M_2O_5$ , with gaseous  $Cl_2$ . Lets consider for simplicity that the reaction can result in the formation of only one gaseous chlorinated specie, say  $MCl_5$ . The transformation is represented by the following equation:



In this system there are only two phases, the pure solid oxide  $M_2O_5$  and a gas phase, whose composition is characterized by definite proportions of  $Cl_2$ ,  $O_2$ , and  $MCl_5$ . If temperature, total pressure, and the total molar amounts of O, Cl, and M are fixed, the chemical equilibrium is calculated by finding the global minimum of the total Gibbs energy of the system (Robert, 1993).

$$G = n_{M_2O_5}^s g_{M_2O_5}^s + G^g \quad (2)$$

Where  $g_{M_2O_5}^s$  represents the molar Gibbs energy of pure solid  $M_2O_5$  at reaction's temperature and total pressure,  $n_{M_2O_5}^s$  the number of moles of  $M_2O_5$  and  $G^g$  the molar Gibbs energy of the gaseous phase, which can be computed through the knowledge of the *chemical potential* of all molecular species present ( $\mu_{Cl_2}^g, \mu_{O_2}^g, \mu_{MCl_5}^g$ ):

$$G^g = n_{Cl_2}^g \mu_{Cl_2}^g + n_{O_2}^g \mu_{O_2}^g + n_{MCl_5}^g \mu_{MCl_5}^g$$

$$\mu_{Cl_2}^g = \left( \frac{\partial G^g}{\partial n_{Cl_2}^g} \right)_{T, P, n_{O_2}, n_{MCl_5}} \quad (3)$$

The minimization of function (2) requires that for the restrictions imposed to the system, the first order differential of  $G$  must be equal to zero. By fixing the reaction temperature ( $T$ ) and pressure ( $P$ ) and total amount of each one of the elements, this condition can be written according to equation (4) (Robert, 1993).

$$dG_{T,P,n_{O_2},n_{Cl_2},n_{M_2O_5}} = 0$$

$$dG = g_{M_2O_5}^s dn_{M_2O_5} + \mu_{Cl_2}^g dn_{Cl_2} + \mu_{O_2}^g dn_{O_2} + \mu_{MCl_5}^g dn_{MCl_5} \quad (4)$$

$$g_{M_2O_5}^s dn_{M_2O_5} + \mu_{Cl_2}^g dn_{Cl_2} + \mu_{O_2}^g dn_{O_2} + \mu_{MCl_5}^g dn_{MCl_5} = 0$$

The development of the chlorination reaction can be followed through introduction of a reaction coordinate called *degree of reaction* ( $\varepsilon$ ), whose first differential is computed by the ratio of its molar content variation of each specie participating in the reaction and the stoichiometric coefficient (Eq. 1).

$$d\varepsilon = \frac{dn_{Cl_2}}{(-5)} = \frac{dn_{M_2O_5}}{(-1)} = \frac{dn_{O_2}}{(+5/2)} = \frac{dn_{MCl_5}}{(+2)} \quad (5)$$

The numbers inside the parenthesis in the denominators of the fractions contained in equation (5) are the stoichiometric coefficient of each specie multiplied by “-1” if it is represented as a reactant, or “+1” if it is a product. The equilibrium condition (Eq. 4) can now be rewritten in the following mathematical form:

$$-g_{M_2O_5}^s d\varepsilon - 5\mu_{Cl_2}^g d\varepsilon + \frac{5}{2}\mu_{O_2}^g d\varepsilon + 2\mu_{MCl_5}^g d\varepsilon = 0$$

$$\left( -g_{M_2O_5}^s - 5\mu_{Cl_2}^g + \frac{5}{2}\mu_{O_2}^g + 2\mu_{MCl_5}^g \right) d\varepsilon = 0 \quad (6)$$

At the desired equilibrium state the condition defined by Eq. (6) must be valid for all possible values of the differential  $d\varepsilon$ . This can only be accomplished if the term inside the parenthesis is equal to zero. This last condition is the simplest mathematical representation for the chemical equilibrium associated with reaction (1).

$$-g_{M_2O_5}^s - 5\mu_{Cl_2}^g + \frac{5}{2}\mu_{O_2}^g + 2\mu_{MCl_5}^g = 0 \quad (7)$$

The chemical potentials can be computed through knowledge of the molar Gibbs energy of each pure specie in the gas phase, and its chemical activity. For the chloride  $MCl_5$ , for example, the following function can be used (Robert, 1993):

$$\mu_{MCl_5}^g = g_{MCl_5}^g + RT \ln a_{MCl_5}^g \quad (8)$$

Where  $a_{MCl_5}^g$  represents the chemical activity of the component  $MCl_5$  in the gas phase. By introducing equations analogous to Eq. (8) for all components of the gas phase, Eq. (7) can be rewritten according to Eq. (9). There, the activity of  $M_2O_5$  is not present in the term located at the left hand side because, as this oxide is assumed to be pure, its activity must be equal to one (Robert, 1993).

$$\ln \left( \frac{a_{MCl_5}^g a_{O_2}^{5/2}}{a_{Cl_2}^5} \right) = - \frac{\left( 2g_{MCl_5}^g + \frac{5}{2}g_{O_2}^g - 5g_{Cl_2}^g - g_{M_2O_5}^s \right)}{RT} = - \frac{\Delta G_r}{RT} \quad (9)$$

The numerator of the right side of Eq. (9) represents the molar Gibbs energy of reaction (1). It involves only the molar Gibbs energies of the species participating in the reaction as pure substances, at  $T$  and  $P$  established in the reactor. The molar Gibbs energy of a pure component is only a function of  $T$  and  $P$  (Eq. 10), so the same must be valid for the reactions Gibbs energy (Robert, 1993).

$$dg(T, P) = -sdT + vdP \quad (10)$$

Where  $s$  and  $v$  denote respectively the molar entropy and molar volume of the material, which for a pure substance are themselves only a function of  $T$  and  $P$ .

It is a common practice in treating reactions involving gaseous species to calculate the Gibbs energy of reaction not at the total pressure prevailing inside the reactor, but to fix it at 1 atm. This is in fact a reference pressure, and can assume any suitable value we desire. The molar Gibbs energy of reaction is in this case referred to as the *standard molar Gibbs energy of reaction*. According to this definition, the standard Gibbs energy of reaction must depend only on the reactor's temperature.

By assuming that the total pressure inside the reactor ( $P$ ) is low enough for neglecting the effect of the interactions among the species present in the gas phase, Eq. (9) can be rewritten in the following form:

$$\frac{P_{\text{MCl}_5}^2 P_{\text{O}_2}^{5/2}}{P_{\text{Cl}_2}^5} = \exp \left( - \frac{\left( 2g_{\text{MCl}_5}^g + \frac{5}{2}g_{\text{O}_2}^g - 5g_{\text{Cl}_2}^g - g_{\text{M}_2\text{O}_5}^s \right)}{RT} \right) \quad (11)$$

The activities were calculated as the ratio of the partial pressure of each component and the reference pressure chosen ( $P = 1$  atm). This proposal is based on the thermodynamic description of an ideal gas (Robert, 1993). For  $\text{MCl}_5$ , for example, the chemical activity is calculated as follows:

$$a_{\text{MCl}_5} = \frac{P_{\text{MCl}_5}}{1} = P_{\text{MCl}_5} = x_{\text{MCl}_5}^g P \quad (12)$$

Where  $x_{\text{MCl}_5}^g$  stands for the mol fraction of  $\text{MCl}_5$  in the gas phase. Similar relations hold for the other species present in the reactor atmosphere. The activity is then expressed as the product of the mol fraction of the specie and the total pressure exerted by the gaseous solution. The right hand side of Eq. (11) defines the equilibrium constant ( $K$ ) of the reaction in question. This quantity can be calculated as follows:

$$K = \exp \left( - \frac{\Delta G_r^o}{RT} \right) \quad (13)$$

The symbol "o" is used to denote that The molar reaction Gibbs energy ( $\Delta G_r^o$ ) is calculated at a reference pressure of 1 atm.

At this point, three possible situations arise. If the standard molar Gibbs energy of the reaction is negative, then  $K > 1$ . If it is positive,  $K < 1$  and if it is equal to zero  $K = 1$ . The first



situation defines a process where in the achieved equilibrium state, the atmosphere tends to be richer in the desired products. The second situation characterizes a reaction where the reactants are present in higher concentration in equilibrium. Finally, the third possibility defines the situation where products and reactants are present in amounts of the same order of magnitude.

## 2.1 Thermodynamic driving force and $\Delta G_r^\circ$ vs. T diagrams

Equation (6) can be used to formulate a mathematical definition of the thermodynamic driving force for a chlorination reaction. If the reaction proceeds in the desired direction, then  $d\varepsilon$  must be positive. Based on the fact that by fixing  $T$ ,  $P$ ,  $n(\text{O})$ ,  $n(\text{Cl})$ , and  $n(\text{M})$  the total Gibbs energy of the system is minimum at the equilibrium, the reaction will develop in the direction of the final equilibrium state, if and only if, the value of  $G$  reduces, or in other words, the following inequality must then be valid:

$$-g_{\text{M}_2\text{O}_5}^s - 5\mu_{\text{Cl}_2}^g + \frac{5}{2}\mu_{\text{O}_2}^g + 2\mu_{\text{MCl}_5}^g < 0 \quad (14)$$

The left hand side of inequality (14) defines the thermodynamic driving force of the reaction ( $\Delta\mu_r$ ).

$$\Delta\mu_r = -g_{\text{M}_2\text{O}_5}^s - 5\mu_{\text{Cl}_2}^g + \frac{5}{2}\mu_{\text{O}_2}^g + 2\mu_{\text{MCl}_5}^g \quad (15)$$

If  $\Delta\mu_r$  is negative, classical thermodynamics says that the process will develop in the direction of obtaining the desired products. However, a positive value is indicative that the reaction will develop in the opposite direction. In this case, the formed products react to regenerate the reactants. By using the mathematical expression for the chemical potentials (Eq. 8), it is possible to rewrite the driving force in a more familiar way:

$$\Delta\mu_r = \Delta G_r^\circ + RT \ln \left( \frac{P_{\text{MCl}_5}^2 P_{\text{O}_2}^{5/2}}{P_{\text{Cl}_2}^5} \right) = \Delta G_r^\circ + RT \ln Q \quad (16)$$

According to Eq. (16), the ratio involving the partial pressure of the components defines the so called *reaction coefficient* ( $Q$ ). This parameter can be specified in a given experiment by injecting a gas with the desired proportion of  $\text{O}_2$  and  $\text{Cl}_2$ . The partial pressure of  $\text{MCl}_5$ , on the other hand, would then be near zero, as after the formation of each species, the fluxing gas removes it from the atmosphere in the neighborhood of the sample.

At a fixed temperature and depending on the value of  $Q$  and the standard molar Gibbs energy of the reaction considered, the driving force can be positive, negative or zero. In the last case the reaction ceases and the equilibrium condition is achieved. It is important to note, however, that by only evaluating the reactions Gibbs energy one is not in condition to predict the reaction path followed, then even for positive values of  $\Delta G_r^\circ$ , it is possible to find a value  $Q$  that makes the driving force negative. This is a usual situation faced in industry, where the desired equilibrium is forced by continuously injecting reactants, or removing products. In all cases, however, for computing reaction driving forces it is vital to know the temperature dependence of the reaction Gibbs energy.

### 2.1.1 Thermodynamic basis for the construction of $\Delta G_r^\circ \times T$ diagrams

To construct the  $\Delta G_r^\circ \times T$  diagram of a particular reaction we must be able to compute its standard Gibbs energy in the whole temperature range spanned by the diagram.

$$\begin{aligned}\Delta G_r^\circ &= \Delta H_r^\circ - T\Delta S_r^\circ \\ \Delta H_r^\circ &= \Delta H_{298} + \int_{298.15 \text{ K}}^T \Delta C_P^\circ dT \\ \Delta S_r^\circ &= \Delta S_{298} + \int_{298.15 \text{ K}}^T \frac{\Delta C_P^\circ}{T} dT\end{aligned}\quad (17)$$

$$\Delta C_P^\circ = \frac{d\Delta H_r^\circ}{dT} = 2C_{P,\text{MCl}_5}^{\circ,g} + \frac{5}{2}C_{P,\text{O}_2}^g - 5C_{P,\text{Cl}_2}^g - C_{P,\text{M}_2\text{O}_5}^s$$

$$\Delta H_{298} = 2H_{298,\text{MCl}_5}^\circ + \frac{5}{2}H_{298,\text{O}_2}^g - 5H_{298,\text{Cl}_2}^g - H_{298,\text{M}_2\text{O}_5}^s$$

$$\Delta S_{298} = 2S_{298,\text{MCl}_5}^\circ + \frac{5}{2}S_{298,\text{O}_2}^g - 5S_{298,\text{Cl}_2}^g - S_{298,\text{M}_2\text{O}_5}^s$$

For accomplishing this task one needs a mathematical model for the molar standard heat capacity at constant pressure, valid for each participating substance for  $T$  varying between 298.15 K and the final desired temperature, its molar enthalpy of formation ( $H_{298}^\circ$ ) and its molar entropy of formation ( $S_{298}^\circ$ ) at 298.15 K

For the most gas – solid reactions both the molar standard enthalpy ( $\Delta H_r^\circ$ ) and entropy of reaction ( $\Delta S_r^\circ$ ) do not depend strongly on temperature, as far no phase transformation among the reactants and or products are present in the considered temperature range. So, the observed behavior is usually described by a line (Fig. 1), whose angular coefficient gives us a measurement of  $\Delta S_r^\circ$  and  $\Delta H_r^\circ$  is defined by the linear coefficient.

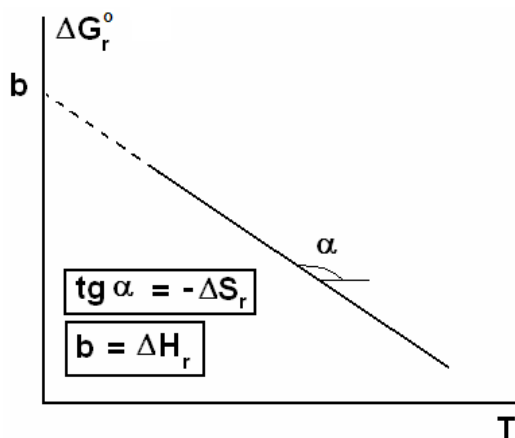


Fig. 1. Hypothetical  $\Delta G_r^\circ \times T$  diagram

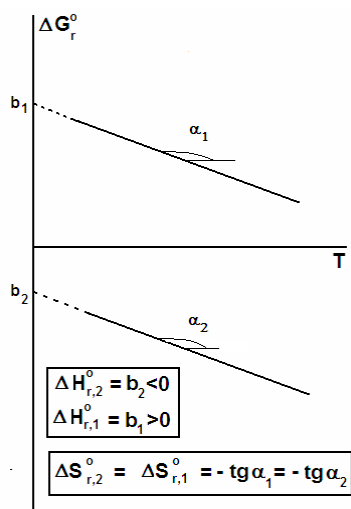


Fig. 2. Endothermic and exothermic reactions

Further, for a reaction defined by Eq. (1) the number of moles of gaseous products is higher than the number of moles of gaseous reactants, which, based on the ideal gas model, is indicative that the chlorination leads to a state of greater disorder, or greater entropy. In this particular case then, the straight line must have negative linear coefficient ( $-\Delta S_r^0 < 0$ ), as depicted in the graph of Figure (1).

The same can not be said about the molar reaction enthalpy. In principle the chlorination reaction can lead to an evolution of heat (exothermic process, then  $\Delta H_r^0 < 0$ ) or absorption of heat (endothermic process, then  $\Delta H_r^0 > 0$ ). In the first case the linear coefficient is positive, but in the later it is negative. Hypothetical cases are presented in Fig. (2) for the chlorination of two oxides, which react according to equations identical to Eq. (1). The same molar reaction entropy is observed, but for one oxide the molar enthalpy is positive, and for the other it is negative.

Finally, it is worthwhile to mention that for some reactions the angular coefficient of the straight line can change at a particular temperature value. This can happen due to a phase transformation associated with either a reactant or a product. In the case of the reaction (1), only the oxide  $M_2O_5$  can experience some phase transformation (melting, sublimation, or ebullition), all of them associated with an increase in the molar enthalpy of the phase. According to classical thermodynamics, the molar entropy of the compound must also increase (Robert, 1993).

$$\Delta S_t = \frac{\Delta H_t}{T_t} \quad (18)$$

Where  $\Delta S_t$ ,  $\Delta H_t$  and  $T_t$  represent respectively, the molar entropy, molar enthalpy and temperature of the phase transformation in question. So, to include the effect for melting of  $M_2O_5$  at a temperature  $T_t$ , the molar reaction enthalpy and entropy must be modified as follows.

$$\Delta H_r^\circ = \int_{298.15}^{T_i} \Delta C_p^\circ dT - \Delta H_{t,M_2O_5} + \int_{T_i}^T \Delta C_p^\circ dT \quad (19)$$

$$\Delta S_r^\circ = \int_{298.15}^{T_i} \frac{\Delta C_p^\circ}{T} dT - \frac{\Delta H_{t,M_2O_5}}{T_t} + \int_{T_i}^T \frac{\Delta C_p^\circ}{T} dT$$

It should be observed that the molar entropy and enthalpy associated with the phase transition experienced by the oxide  $M_2O_5$  were multiplied by its stoichiometric number “-1”, which explains the minus sign present in both relations of Eq. (19).

An analogous procedure can be applied if other phase transition phenomena take place. One must only be aware that the mathematical description for the molar reaction heat capacity at constant pressure ( $\Delta C_p^\circ$ ) must be modified by substituting the heat capacity of solid  $M_2O_5$  for a model associated with the most stable phase in each particular temperature range. If, for example, in the temperature range of interest  $M_2O_5$  melts at  $T_t$ , for  $T > T_t$  the molar heat capacity of solid  $M_2O_5$  must be substituted for the model associated with the liquid state (Eq. 20).

$$\Delta C_p^\circ = 2C_{P,MCl_5}^{\circ,g} + \frac{5}{2}C_{P,O_2}^{\circ,g} - 5C_{P,Cl_2}^{\circ,g} - C_{P,M_2O_5}^s \quad (T < T_t)$$

$$\Delta C_p^\circ = 2C_{P,MCl_5}^{\circ,g} + \frac{5}{2}C_{P,O_2}^{\circ,g} - 5C_{P,Cl_2}^{\circ,g} - C_{P,M_2O_5}^l \quad (T > T_t) \quad (20)$$

The effect of a phase transition over the geometric nature of the  $\Delta G_r^\circ \times T$  curve can be directly seen. The melting of  $M_2O_5$  makes its molar enthalpy and entropy higher. According to Eq. (19), such effects would make the molar reaction enthalpy and entropy lower. So the curve should experience a decrease in its first order derivative at the melting temperature (Figure 3).

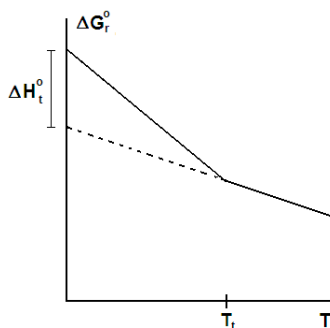


Fig. 3. Effect of  $M_2O_5$  melting over the  $\Delta G_r^\circ \times T$  diagram

Based on the definition of the reaction Gibbs energy (Eq. 17), similar transitions involving a product would produce an opposite effect. The reaction Gibbs energy would in these cases dislocate to more negative values. In all cases, though, the magnitude of the deviation is proportional to the magnitude of the molar enthalpy associated with the particular transition observed. The effect increases in the following order: melting, ebullition and sublimation.

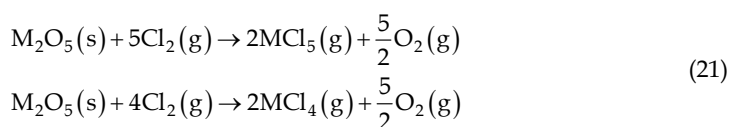
## 2.2 Multiple reactions

In many situations the reaction of a metallic oxide with  $\text{Cl}_2$  leads to the formation of a family of chlorinated species. In these cases, multiple reactions take place. In the present section three methods will be described for treating this sort of situation, the first of them is of qualitative nature, the second semi-qualitative, and the third a rigorous one, that reproduces the equilibrium conditions quantitatively.

The first method consists in calculating  $\Delta G_r^0 \times T$  diagrams for each reaction in the temperature range of interest. The reaction with the lower molar Gibbs energy must have a greater thermodynamic driving force. The second method involves the solution of the equilibrium equations independently for each reaction, and plotting on the same space the concentration of the desired chlorinated species. Finally, the third method involves the calculation of the thermodynamic equilibrium by minimizing the total Gibbs energy of the system. The concentrations of all species in the phase ensemble are then simultaneously computed.

### 2.2.1 Methods based on $\Delta G_r^0 \times T$ diagrams

It will be supposed that the oxide  $\text{M}_2\text{O}_5$  can generate two gaseous chlorinated species,  $\text{MCl}_4$  and  $\text{MCl}_5$ :



The first reaction is associated with a reduction of the number of moles of gaseous species ( $\Delta n_g = -0.5$ ), but in the second the same quantity is positive ( $\Delta n_g = 0.5$ ). If the gas phase is described as an ideal solution, the first reaction should be associated with a lower molar entropy than the second. The greater the number of mole of gaseous products, the greater the gas phase volume produced, and so the greater the entropy generated. By plotting the molar Gibbs energy of each reaction as a function of temperature, the curves should cross each other at a specific temperature ( $T_c$ ). For temperatures greater than  $T_c$  the formation of  $\text{MCl}_4$  becomes thermodynamically more favorable (see Figure 4).

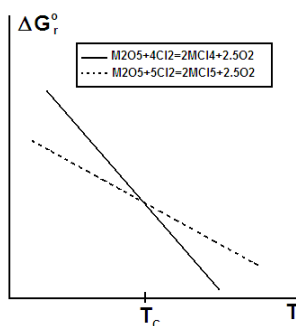


Fig. 4. Hypothetical  $\Delta G_r^0 \times T$  curves with intercept.

An interesting situation occurs, if one of the chlorides can be produced in the condensed state (liquid or solid). Let's suppose that the chloride  $\text{MCl}_5$  is liquid at lower temperatures.

The ebullition of  $MCl_5$ , which occur at a definite temperature ( $T_t$ ), dislocates the curve to lower values for temperatures higher than  $T_t$ . Such an effect would make the production of  $MCl_5$  in the gaseous state thermodynamically more favorable even for temperatures greater than  $T_c$  (Figure 5). Such fact the importance of considering phase transitions when comparing  $\Delta G_r^\circ \times T$  curves for different reactions.

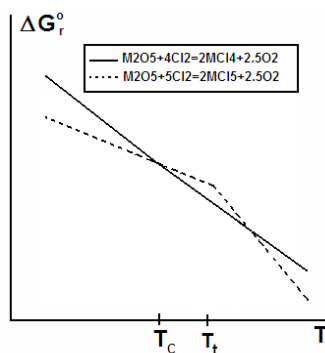


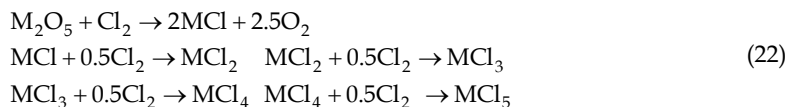
Fig. 5. Effect of  $MCl_5$  boiling temperature

Although simple, the method based on the comparison of  $\Delta G_r^\circ \times T$  diagrams is of limited application. The problem is that for discussing the thermodynamic viability of a reaction one must actually compute the thermodynamic driving force (Eq. 15 and 16), and by doing so, one must fix values for the concentration of  $Cl_2$  and  $O_2$  in the reactor's atmosphere, which, in the end, define the value of the reaction coefficient.

If the  $\Delta G_r^\circ \times T$  curves of two reactions lie close to one another (difference lower than 10 KJ/mol), it is impossible to tell, without a rigorous calculation, which chlorinated specie should have the highest concentration in the gaseous state, as the computed driving forces will lie very close from each other. In these situations, other methods that can address the direct effect of the reactor's atmosphere composition should be applied.

Apart from its simplicity, the  $\Delta G_r^\circ \times T$  diagrams have another interesting application in relation to the proposal of reactions mechanisms. From the point of view of the kinetics, the process of forming higher chlorinated species by the "collision" of one molecule of the oxide  $M_2O_5$  and a group of molecules of  $Cl_2$ , and vice versa, shall have a lower probability than the one defined by the first formation of a lower chlorinated specie, say  $MCl_2$ , and the further reaction of it with one or two  $Cl_2$  molecules (Eq. 22).

Let's consider that M can form the following chlorides:  $MCl$ ,  $MCl_2$ ,  $MCl_3$ ,  $MCl_4$ , and  $MCl_5$ . The synthesis of  $MCl_5$  can now be thought as the result of the coupled reactions represented by Eq. (22).



By plotting the  $\Delta G_r^\circ \times T$  diagrams of all reactions presented in Eq. (22) it is possible to evaluate if the thermodynamic stability of the chlorides follows the trend indicated by the

proposed reaction path. If so, the curves should lay one above the other. The standard reaction Gibbs energy would then grow in the following order:  $\text{MCl}$ ,  $\text{MCl}_2$ ,  $\text{MCl}_3$ ,  $\text{MCl}_4$  and  $\text{MCl}_5$  (Figure 6).

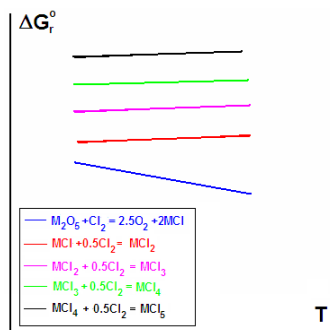


Fig. 6. Hypothetic  $\Delta G_r^\circ \times T$  curves for successive chlorination reactions

Another possibility is that the curve for the formation of one of the higher chlorinated species is associated with lower Gibbs energy values in comparison with the curve of a lower chlorinated compound. A possible example thereof is depicted on Figure (7), where the  $\Delta G_r^\circ \times T$  curve for the production of  $\text{MCl}_3$  lies below the curve associated with the formation of  $\text{MCl}_2$ .

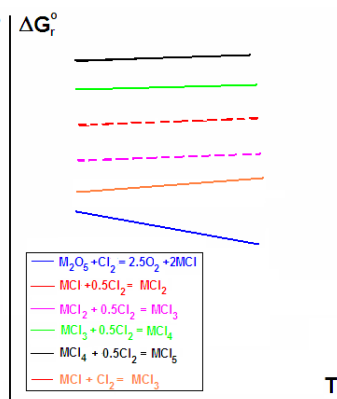


Fig. 7. Successive chlorination reactions - direct formation of  $\text{MCl}_3$  from  $\text{MCl}$

The formation of the species  $\text{MCl}_2$  would be thermodynamically less favorable, and  $\text{MCl}_3$  is preferentially produced directly from  $\text{MCl}$  ( $\text{MCl} + \text{Cl}_2 = \text{MCl}_3$ ). In this case, however, for the diagram to remain thermodynamically consistent, the curves associated with the formation of  $\text{MCl}_2$  from  $\text{MCl}$  and  $\text{MCl}_3$  from  $\text{MCl}$  (broken lines) should be substituted for the curve associated with the direct formation of  $\text{MCl}_3$  from  $\text{MCl}$  for the entire temperature range.

The same effect could originate due to the occurrence of a phase transition. Let's suppose that in the temperature range considered  $\text{MCl}_3$  sublimates at  $T_s$ . Because of this

phenomenon the curve for the formation of  $MCl_2$  crosses the curve for the formation of the last chloride at  $T_c$ , so that for  $T > T_c$  its formation is associated with a higher thermodynamic driving force (Figure 8). So, for  $T > T_c$   $MCl_3$  is formed directly from  $MCl$ , resulting in the same modification in the reaction mechanism as mentioned above.

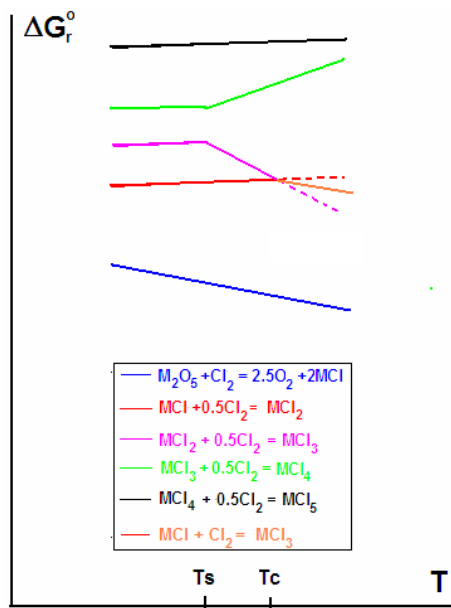


Fig. 8. Direct formation of  $MCl_3$  from  $MCl$  stimulated by  $MCl_3$  sublimation

For temperatures higher than  $T_c$ , the diagram of Figure (8) loses its thermodynamic consistency, as, according to what was mentioned in the last paragraph, the formation of  $MCl_2$  from  $MCl$  is impossible in this temperature range. The error can be corrected if, for  $T > T_c$ , the curves associated with the formation of  $MCl_2$  and  $MCl_3$  (broken lines) are substituted for the curve associated with the formation of  $MCl_3$  directly from  $MCl$ .

A direct consequence of that peculiar thermodynamic fact, as described in Figures (7) and (8), is that under these conditions, a predominance diagram would contain a straight line showing the equilibrium between  $MCl$  and  $MCl_3$ , and the field corresponding to  $MCl_2$  would not appear.

### 2.2.2 Method of Kang and Zuo

Kang & Zuo (1989) introduced a simple method for comparing the thermodynamic tendencies of formation of compounds obtained by gas - solid reactions, in that each equilibrium equation is solved independently, and the concentration of the desired species plotted as a function of the gas phase concentration and or temperature. The method will be illustrated for the reactions defined by Eq. (21). The concentrations of  $MCl_4$  and  $MCl_5$  in the gaseous phase can be computed as a function of temperature, partial pressure of  $Cl_2$ , and partial pressure of  $O_2$ .



$$\begin{aligned}
 P_{\text{MCl}_5} &= \sqrt{\frac{P_{\text{Cl}_2}^5}{P_{\text{O}_2}^{5/2}} \exp\left(-\frac{\left(2g_{\text{MCl}_5}^g + \frac{5}{2}g_{\text{O}_2}^g - 5g_{\text{Cl}_2}^g - g_{\text{M}_2\text{O}_5}^s\right)}{RT}\right)} \\
 P_{\text{MCl}_4} &= \sqrt{\frac{P_{\text{Cl}_2}^4}{P_{\text{O}_2}^{5/2}} \exp\left(-\frac{\left(2g_{\text{MCl}_4}^g + \frac{5}{2}g_{\text{O}_2}^g - 4g_{\text{Cl}_2}^g - g_{\text{M}_2\text{O}_5}^s\right)}{RT}\right)}
 \end{aligned} \tag{23}$$

Next, two intensive properties must be chosen, whose values are fixed, for example, the partial pressure of  $\text{Cl}_2$  and the temperature. The partial pressure of each chlorinated species becomes in this case a function of only the partial pressure of  $\text{O}_2$ .

$$\begin{aligned}
 P_{\text{MCl}_5} &= f_{\text{MCl}_5}(T, P_{\text{Cl}_2}) P_{\text{O}_2}^{5/2} \\
 P_{\text{MCl}_4} &= f_{\text{MCl}_4}(T, P_{\text{Cl}_2}) P_{\text{O}_2}^{5/2} \\
 f_{\text{MCl}_5}(T, P_{\text{Cl}_2}) &= P_{\text{Cl}_2}^{-5/2} \exp\left(-\frac{\left(2g_{\text{MCl}_5}^g + \frac{5}{2}g_{\text{O}_2}^g - 5g_{\text{Cl}_2}^g - g_{\text{M}_2\text{O}_5}^s\right)}{2RT}\right) \\
 f_{\text{MCl}_4}(T, P_{\text{Cl}_2}) &= P_{\text{Cl}_2}^{-2} \exp\left(-\frac{\left(2g_{\text{MCl}_4}^g + \frac{5}{2}g_{\text{O}_2}^g - 4g_{\text{Cl}_2}^g - g_{\text{M}_2\text{O}_5}^s\right)}{2RT}\right)
 \end{aligned} \tag{24}$$

By fixing  $T$  and  $P(\text{Cl}_2)$  the application of the natural logarithm to both sides of Eq. (24) results in a linear behavior.

$$\begin{aligned}
 \ln P_{\text{MCl}_5} &= \ln f_{\text{MCl}_5} + 2.5 \ln P_{\text{O}_2} \\
 \ln P_{\text{MCl}_4} &= \ln f_{\text{MCl}_4} + 2.5 \ln P_{\text{O}_2}
 \end{aligned} \tag{25}$$

The lines associated with the formation of  $\text{MCl}_4$  and  $\text{MCl}_5$  would have the same angular coefficient, but different linear coefficients. If the partial pressure of  $\text{Cl}_2$  is equal to one (pure  $\text{Cl}_2$  is injected into the reactor), the differences in the standard reaction Gibbs energy controls the values of the linear coefficients observed. If the lowest Gibbs energy values are associated with the formation of  $\text{MCl}_5$ , its line would have the greatest linear coefficient (Figure 9).

An interesting situation occurs if the curves obtained for the chlorinated species of interest cross each other (Figure 10). This fact would indicate that for some critical value of  $P(\text{O}_2)$  there would be a different preference for the system to generate each one of the chlorides. One of them prevails for higher partial pressure values and the other for values of  $P(\text{O}_2)$  lower than the critical one. Such a behavior could be exemplified if the chlorination of  $M$  also generates the gaseous oxychloride  $\text{MOCl}_3$  ( $\text{M}_2\text{O}_5 + 2\text{Cl}_2 = 2\text{MOCl}_3 + 1.5\text{O}_2$ ).

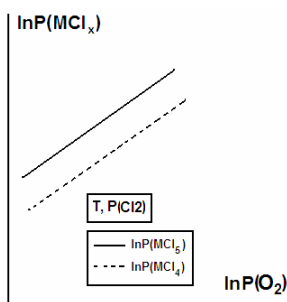


Fig. 9. Concentrations of  $MCl_4$  and  $MCl_5$ , as a function of  $P(O_2)$

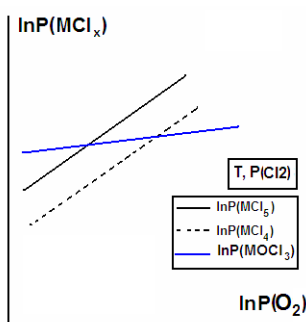


Fig. 10. Concentrations of  $MOCl_3$ ,  $MCl_4$  and  $MCl_5$  as a function of  $P(O_2)$

$$\ln P_{MOCl_3} = \ln f_{MOCl_3} + 1.5 \ln P_{O_2} \quad (26)$$

The linear coefficient of the line associated with the  $MOCl_3$  formation is higher for the initial value of  $P(O_2)$  than the same factor computed for  $MCl_4$  and  $MCl_5$ . As the angular coefficient is lower for  $MOCl_3$ , The graphic of Figure (10) depicts a possible result.

According to Figure (10), three distinct situations can be identified. For the initial values of  $P(O_2)$ , the partial pressure of  $MOCl_3$  is higher than the partial pressure of the other chlorinated compounds.

By varying  $P(O_2)$ , a critical value is approached after which  $P(MCl_5)$  assumes the highest value, being followed by  $P(MOCl_3)$  and then  $P(MCl_4)$ . A second critical value of  $P(O_2)$  can be identified in the graphic above. For  $P(O_2)$  values higher than this, the atmosphere should be more concentrated in  $MCl_5$  and less concentrated in  $MOCl_3$ ,  $MCl_4$  assuming a concentration value in between.

### 2.2.3 Minimization of the total gibbs energy

The most general way of describing equilibrium is to fix a number of thermodynamic variables (physical parameters that can be controlled in laboratory), and to chose an appropriate thermodynamic potential, whose maxima or minima describe the possible equilibrium states available to the system.

By fixing  $T$ ,  $P$ , and total amounts of the components  $M$ ,  $O$ , and  $Cl$  ( $n(O)$ ,  $n(M)$ , and  $n(Cl)$ ), the global minimum of the total Gibbs energy describes the equilibrium state of interest,

which is characterized by a proper phase ensemble, their amounts and compositions. This method is equivalent to solve all chemical equilibrium equations at the same time, so, that the compositions of the chlorinated species in each one of the phases present are calculated simultaneously.

For treating the equilibrium associated with the chlorination processes, two type of diagrams are important: *predominance diagrams*, and *phase speciation diagrams*. The first sort of diagram describes the equilibrium phase ensemble as a function of temperature, and or partial pressure of  $\text{Cl}_2$  or  $\text{O}_2$ . The second type describes how the composition of individual phases varies with temperature and or concentration of  $\text{Cl}_2$  or  $\text{O}_2$ .

The first step is to change the initial constraint vector  $(T, P, n(\text{O}), n(\text{M}), n(\text{Cl}))$ , by modifying the definition of the components. Instead considering as components the elements O, M, and Cl, we can describe the global composition of the system by specifying amounts of M,  $\text{Cl}_2$  and  $\text{O}_2$   $(T, P, n(\text{O}_2), n(\text{M}), n(\text{Cl}_2))$ .

According to the phase-rule (Eq. 27) applied to a system with three components (M,  $\text{Cl}_2$  and  $\text{O}_2$ ), by specifying five degrees of freedom (intensive variables or restriction equations) the equilibrium calculation problem has a unique solution:

$$\begin{aligned} L &= C + 2 - F \\ F &= 0 \quad C = 3 \\ L &= 5 \end{aligned} \tag{27}$$

Where  $F$  denotes the number of phases present (as we do not know the nature of the phase ensemble,  $F = 0$  at the beginning),  $C$  is the number of components, and  $L$  defines the number of degrees of freedom (equations and or intensive variables) to be specified. So, with  $L = 5$ , the constraint vector must have five coordinates  $(T, P, n(\text{O}_2), n(\text{M}), n(\text{Cl}_2))$ .

In reality, the chlorination system is described as an open system, where a gas flux of definite composition is established. The constraint vector defined so far is consistent with the definition of a closed system, which by definition does not allow matter to cross its boundaries. The calculation can become closer to the physical reality of the process if we specify the chemical activities of  $\text{Cl}_2$  and  $\text{O}_2$  in the gas phase, instead of fixing their global molar amounts. Such a restriction would be analogous as fixing the inlet gas composition. Further, if the gas is considered to behave ideally, the chemical activities can be replaced by the respective values of the partial pressure of the gaseous components. So, the final constraint vector should be defined as follows:  $T, P, n(\text{M}), P(\text{Cl}_2), P(\text{O}_2)$ .

The two types of computation mentioned in the first paragraph can now be discussed. For generating a speciation diagram, only one of the parameters  $T, P(\text{Cl}_2)$ , or  $P(\text{O}_2)$  is varied in a definite range. The composition of some phase of interest, for example the gas, can then be plotted as a function of the thermodynamic coordinate chosen. On the other hand, by systematically varying two of the parameter defined in the group  $T, P(\text{Cl}_2)$ , or  $P(\text{O}_2)$ , a predominance diagram can be constructed (Figure 11). The diagram is usually drawn in space  $P(\text{Cl}_2) \times P(\text{O}_2)$  and is composed by cells, which describe the stability limits of individual phases. A line describes the equilibrium condition involving two phases, and a point the equilibrium involving three phases.

Let's take a closer look in the nature of a predominance diagram applied to the case studied so far. In this situation, one must consider the gas phase, the solid metal M, and possible oxides, MO,  $\text{MO}_2$ , and  $\text{M}_2\text{O}_5$ , obtained through oxidation of the element M at different oxygen potentials. The equilibrium involving two oxides defines a unique value of the partial pressure of  $\text{O}_2$ , which is independent of the  $\text{Cl}_2$  concentration.

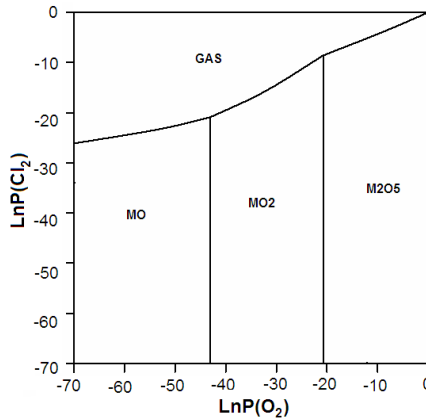


Fig. 11. Hypothetical predominance diagram chlorides mixed in the gas phase

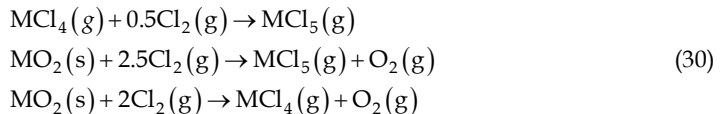
For the equilibrium between MO and MO<sub>2</sub>, for example, Eq. (28) enables the determination of the  $P(O_2)$  value, which is fixed by choosing  $T$  and is independent of the Cl<sub>2</sub> partial pressure. As a consequence, such equilibrium states are defined by a vertical line.

$$P_{O_2} = \exp \left( - \frac{ \left( g_{MO_2}^s - g_{MO}^s - 0.5g_{O_2}^g \right) }{ RT } \right) \quad (28)$$

The equilibrium when the phase ensemble is defined by the gas and one of the metal oxides, say MO<sub>2</sub>, is also defined by a line, whose inclination is determined by fixing  $T$ ,  $P$ ,  $n(M)$  and  $P(O_2)$ . This time the concentration of Cl<sub>2</sub>, MCl<sub>4</sub> and MCl<sub>5</sub> are computed by solving the group of non-linear equations presented below (Eq. 29). The first equation defines the restriction that the molar quantity of M is constant (mass conservative restriction). The second equation represents the conservation of the total mass of the gas phase (the summation of all mol fractions must be equal to one).

$$\begin{aligned} n_M &= n_{MO_2} + n^g \left( x_{MCl_4}^g + x_{MCl_5}^g \right) \\ 1 &= x_{MCl_4}^g + x_{MCl_5}^g + x_{O_2}^g + x_{Cl_2}^g \\ \mu_{MCl_5}^g - \mu_{MCl_4}^g - \frac{\mu_{Cl_2}^g}{2} &= 0 \\ 2\mu_{MCl_5}^g + 2.5\mu_{O_2}^g - g_{M_2O_5}^s - 5\mu_{Cl_2}^g &= 0 \\ 2\mu_{MCl_4}^g + 2.5\mu_{O_2}^g - g_{M_2O_5}^s - 4\mu_{Cl_2}^g &= 0 \end{aligned} \quad (29)$$

The other three relations define, respectively, the equilibrium conditions for the following group of reactions:



So, we have five equations and five unknowns ( $n^g$ ,  $n_{\text{MO}_2}$ ,  $x_{\text{MCl}_4}^g$ ,  $x_{\text{MCl}_5}^g$ ,  $x_{\text{Cl}_2}^g$ ), indicating that the equilibrium calculation admits a unique solution.

Finally by walking along a vertical line associated with the coexistence of two metallic oxides, for example MO and  $\text{MO}_2$ , a condition is achieved where the gaseous chlorides are formed. The equilibrium between the two oxides and the gas phase is defined by a point. In other words by fixing  $T$  and  $P$ , all equilibrium properties are uniquely defined. The equation associated with the coexistence of MO and  $\text{MO}_2$  (Eq. 28) is added and the partial pressure of  $\text{O}_2$  is allowed to vary, resulting in six variables and six equations (Eq. 31).

Equations (30) and (31) were presented here only with a didactic purpose. In praxis, the majority of the thermodynamic software (*Thermocalc*, for example) are designed to minimize the total Gibbs energy of the system. The algorithm varies systematically the composition of the equilibrium phase ensemble until the global minimum is achieved. By doing so the same algorithm can be implemented for dealing with all possible equilibrium conditions, eliminating at the end the difficulty of proposing a group of linear independent chemical equations, which for a system with a great number of components can become a complicated task.

$$\begin{aligned}
 n_{\text{M}} &= n_{\text{MO}_2} + n^g (x_{\text{MCl}_4}^g + x_{\text{MCl}_5}^g) \\
 1 &= x_{\text{MCl}_4}^g + x_{\text{MCl}_5}^g + x_{\text{O}_2}^g + x_{\text{Cl}_2}^g \\
 \mu_{\text{MCl}_5}^g - \mu_{\text{MCl}_4}^g - 0.5\mu_{\text{Cl}_2}^g &= 0 \\
 2\mu_{\text{MCl}_5}^g + 2.5\mu_{\text{O}_2}^g - g_{\text{M}_2\text{O}_5}^s - 5\mu_{\text{Cl}_2}^g &= 0 \\
 2\mu_{\text{MCl}_4}^g + 2.5\mu_{\text{O}_2}^g - g_{\text{M}_2\text{O}_5}^s - 4\mu_{\text{Cl}_2}^g &= 0 \\
 g_{\text{MO}_2}^s - g_{\text{MO}}^s - 0.5\mu_{\text{O}_2}^g &= 0
 \end{aligned} \tag{31}$$

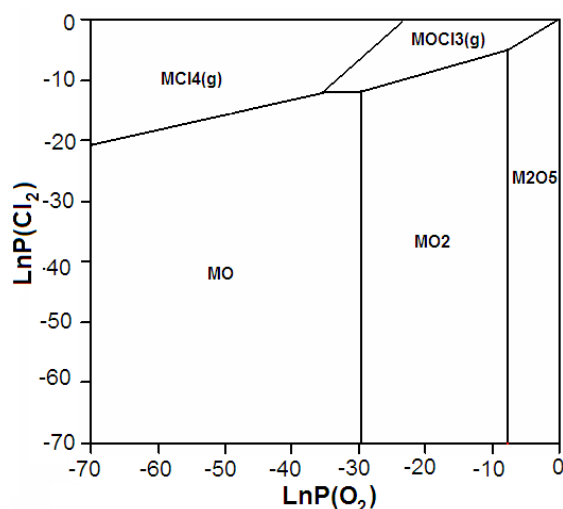


Fig. 12. Hypothetical predominance diagram: pure gaseous chlorides

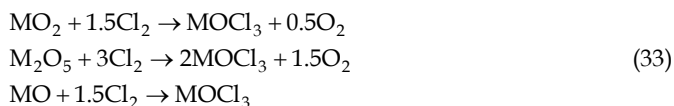
A simplified version of the predominance diagram of Figure (11) can be achieved through considering each possible gaseous chloride as a pure substance. In this case, the field representing the gas phase will be divided into sub-regions, each one representative of the stability of each gaseous chlorinated compound. By considering, that, besides  $\text{MCl}_5$  and  $\text{MCl}_4$ , gaseous  $\text{MOCl}_3$  can also be formed, a diagram similar to the one presented on Figure (12) would represent possible stability limits found in equilibrium.

The diagram of Figure (12) is associated with a temperature value where gaseous  $\text{MCl}_5$  can not be present in equilibrium for any suitable value of  $P(\text{Cl}_2)$  and  $P(\text{O}_2)$  chosen. It is interesting to note, that in this sort of diagram, there is a direct relation between the inclination of a line representative of the equilibrium between a gaseous chloride or oxychloride and an oxide, with the stoichiometric coefficients of the chemical reaction behind the transformation.

According to Eq. (32), the inclination of the line associated with the equilibrium between  $\text{MOCl}_3$  and  $\text{MO}_2$  should be lower than the one associated with the equilibrium between  $\text{MOCl}_3$  and  $\text{M}_2\text{O}_5$ . On the other hand, in the case of the equilibrium between  $\text{MO}$  and  $\text{MOCl}_3$ , the line is horizontal (does not depend on  $P(\text{O}_2)$ ), as the same number of oxygen atoms is present in the reactant and products, so  $\text{O}_2$  does not participate in the reaction.

$$\begin{aligned}\ln P_{\text{Cl}_2} &= \frac{1}{3} \ln P_{\text{O}_2} - \frac{2}{3} \ln K_{\text{MO}_2}(T) \\ \ln P_{\text{Cl}_2} &= \frac{1}{2} \ln P_{\text{O}_2} - \frac{1}{3} \ln K_{\text{M}_2\text{O}_5}(T) \\ \ln P_{\text{Cl}_2} &= -\frac{2}{3} \ln K_{\text{MO}}(T)\end{aligned}\quad (32)$$

Where,  $K_{\text{MO}_2}$ ,  $K_{\text{M}_2\text{O}_5}$  and  $K_{\text{MO}}$  represent respectively the equilibrium constants for the formation of  $\text{MOCl}_3$  from  $\text{MO}_2$ ,  $\text{M}_2\text{O}_5$  and  $\text{MO}$  (Eq. 33).



The diagrams of Figures (11) and (12) depict a behavior, where no condensed chlorinated phases are present. For many oxides, however, there is a tendency of formation of solid or liquid chlorides and or oxychlorides, which must appear in the predominance diagram as fields between the pure oxides and the gas phase regions. Such a behavior can be observed in the equilibrium states accessible to the system V - O - Cl.

### 3. The system V - O - Cl

Vanadium is a transition metal that can form a variety of oxides. At ambient temperature and oxygen potential, the form  $\text{V}_2\text{O}_5$  is the most stable. It is a solid stoichiometric oxide, where vanadium occupies the +5 oxidation state. By lowering the partial pressure of  $\text{O}_2$ , the valence of vanadium varies considerably, making it is possible to produce a family of stoichiometric oxides:  $\text{V}_2\text{O}_4$ ,  $\text{V}_3\text{O}_5$ ,  $\text{V}_4\text{O}_7$ ,  $\text{VO}$ ,  $\text{VO}_2$  and  $\text{V}_2\text{O}_3$ . Recently, it has been discovered that vanadium can also form a variety of non-stoichiometric oxygenated compounds (Brewer & Ebinghaus, 1988), however, to simplify the treatment of the present chapter, these

phases will not be included in the data-base used for the following computations. Additionally, it was considered that the concentration of the oxides in gas phase is low enough to be neglected. Further, on what touches the computations that follows, the software *Thermocalc* was used in all cases, and it will always be assumed that equilibrium is achieved, or in other words, kinetic effects can be neglected.

The relative stability of the possible vanadium oxides can be assessed through construction of a predominance diagram in the space  $T - P(\text{O}_2)$  (see Figure 13). As thermodynamic constraints we have  $n(\text{V})$  (number of moles of vanadium metal - it will be supposed that  $n(\text{V}) = 1$ ),  $T$ ,  $P$  and  $P(\text{O}_2)$ . The reaction temperature will be varied in the range between 1073 K and 1500 K and the partial pressure of  $\text{O}_2$  in the range between  $8.2 \cdot 10^{-40}$  atm and 1 atm.

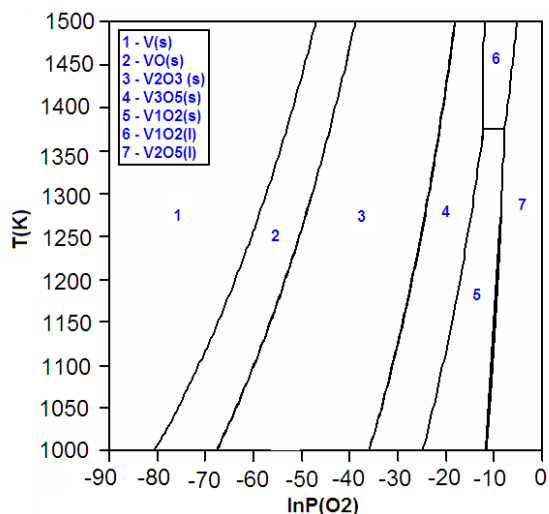


Fig. 13. Predominance diagram for the system V - O

The total pressure was fixed at 1 atm. It can be seen that for the temperature range considered and a partial pressure of  $\text{O}_2$  in the neighborhood of 1 atm,  $\text{V}_2\text{O}_5$  is formed in the liquid state. Through lowering the oxygen potential, crystalline vanadium oxides precipitate,  $\text{VO}_2$  being formed first, followed by  $\text{V}_2\text{O}_3$ ,  $\text{VO}$ , and finally  $\text{V}$ . The horizontal line between fields "5" and "6" indicates the melting of  $\text{V}_1\text{O}_2$ , which according to classical thermodynamics must occur at a fixed temperature. Next it will be considered the species formed by vanadium, chlorine and oxygen.

### 3.1 Vanadium oxides and chlorides

The already identified species formed between vanadium, chlorine and oxygen are:  $\text{VCl}$ ,  $\text{VCl}_2$ ,  $\text{VCl}_3$ ,  $\text{VCl}_4$ ,  $\text{VOCl}$ ,  $\text{VOCl}_2$ ,  $\text{VOCl}_3$ ,  $\text{VO}_2\text{Cl}$ .

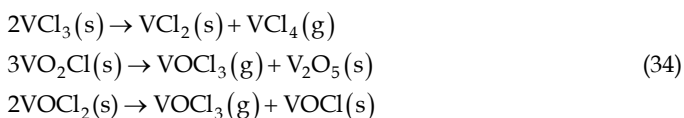
On Table (1) it was included information regarding the physical states at ambient conditions and some references related to phase equilibrium studies conducted on samples of specific vanadium chlorinated compounds.

Only a few studies were published in literature in relation to the thermodynamics of vanadium chlorinated phases. On Table (1) some references are given for earlier

investigations associated with measurements of the vapor pressure for the sublimation of  $VCl_2$  and  $VCl_3$ , and the boiling of  $VOCl_3$  and  $VCl_4$ . There are also evidences for the occurrence of specific thermal decomposition reactions (Eq. 34), such as those of  $VCl_3$ ,  $VOCl_2$  and  $VO_2Cl$  (Oppermann, 1967).

Chloride	Physical state	Equilibrium data	Reference
VCl	-	-	-
$VCl_2$	Solid	Sublimation(	McCarley & Roddy (1964)
$VCl_3$	Solid	Sublimation/ Thermal decomposition	McCarley & Roddy (1964)
$VCl_4$	Liquid	Ebultion	Oppermann (1962a)
$VOCl_3$	Liquid	Ebultion(	Oppermann (1967)
$VO_2Cl$	Solid	Thermal decomposition(	Oppermann (1967)
$VOCl_2$	Solid	Thermal decomposition(	Oppermann (1967)
VOCl	Solid	Synthesis and characterization(	Schäffer at al. (1961)

Table 1. Physical nature and phase equilibrium data for vanadium chlorinated compounds



Chromatographic measurements conducted recently confirmed the possible formation of VCl,  $VCl_2$ ,  $VCl_3$ , and  $VCl_4$  in the gas phase (Hildenbrand et al., 1988). In this study the molar Gibbs energy models for the mentioned chlorides were revised, and new functions proposed. In the case vanadium oxychlorides, models for the molar Gibbs energies of gaseous VOCl,  $VOCl_3$ , and  $VOCl_2$  have already been published (Hackert et al., 1996).

For gaseous  $VO_2Cl$ , on the other hand, no thermodynamic model exists, indicating the low tendency of this oxychloride to be stabilized in the gaseous state.

### 3.1.1 The V – O<sub>2</sub> – Cl<sub>2</sub> stability diagram

The relative stability of the possible chlorinated compounds of vanadium can be assessed through construction of predominance diagrams by fixing the temperature and systematic varying the values of  $P(Cl_2)$  and  $P(O_2)$ .

For the temperature range usually found in chlorination praxis, three temperatures were considered, 1073 K, 1273 K and 1573 K. The partial pressure of  $Cl_2$  and  $O_2$  were varied in the range between  $3.98 \cdot 10^{-31}$  atm and 1 atm. All chlorinated species are considered to be formed at the standard state (pure at 1 atm). The predominance diagrams can be observed on Figures (14), (15) and (16).

The stability field of  $VCl_2(l)$  grows in relation to those associated to  $VCl_4$  and  $VOCl_3$ . At 1573 K the  $VCl_2(l)$  area is the greatest among the chlorides and the  $VCl_3(g)$  field appears. So, as temperature achieves higher values the concentration of  $VCl_3$  in the gas phase should increase in comparison with the other chlorinated species, including  $VCl_2$ . This behavior agrees with the one observed during the computation of the gas phase speciation and will be better discussed on topic (3.1.3.2).



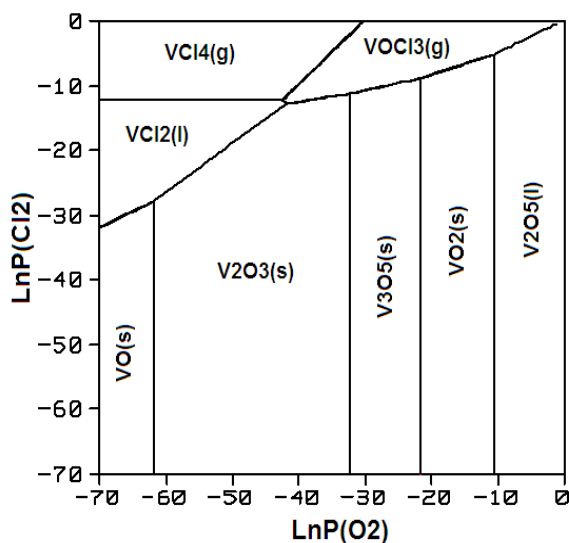


Fig. 14. Predominance diagram for the system V - O - Cl at 1073 K

Finally, by starting in a state inside a field representing the formation of  $\text{VCl}_4$  or  $\text{VCl}_3$  and by making  $P(\text{O}_2)$  progressively higher, a value is reached, after which  $\text{VOCl}_3(\text{g})$  appears. So, the mol fraction of  $\text{VCl}_4$  and  $\text{VCl}_3$  in gas should reduce when  $P(\text{O}_2)$  achieves higher values. This is again consistent with the speciation computations developed on topic (3.1.3.2).

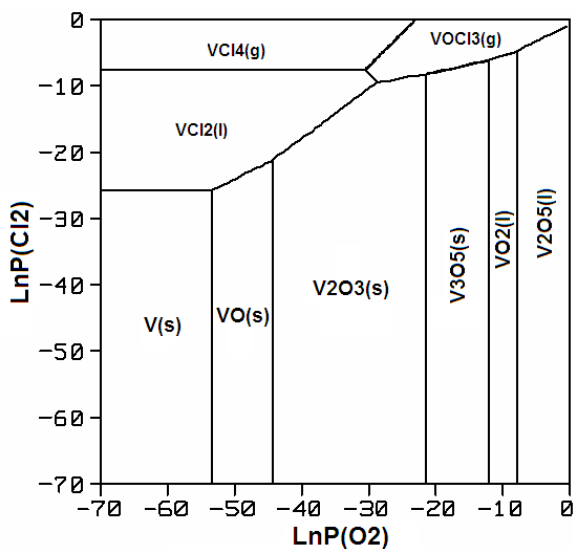


Fig. 15. Predominance diagram for the system V - O - Cl at 1273 K

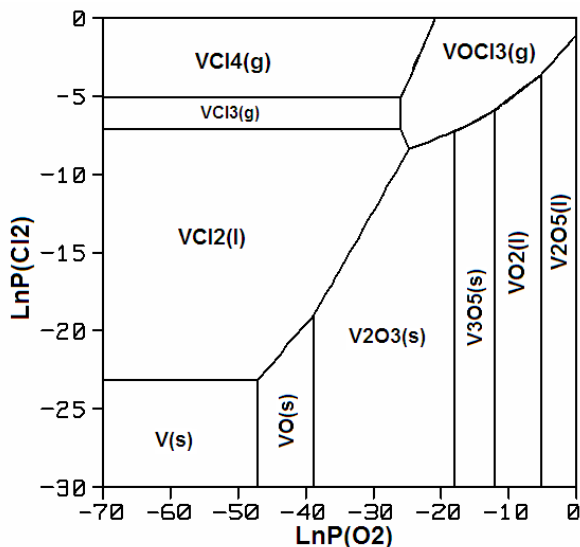
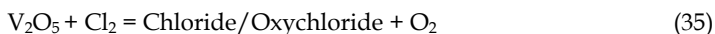


Fig. 16. Predominance diagram for the system V - O - Cl at at 1573 K

### 3.1.2 $\text{V}_2\text{O}_5$ direct chlorination and the effect of the reducing agent

The direct chlorination of  $\text{V}_2\text{O}_5$  is a process, which consists in the reaction of a  $\text{V}_2\text{O}_5$  sample with gaseous  $\text{Cl}_2$ .



In praxis, temperature lies usually between 1173 K and 1473 K. The chlorination equilibrium could then be dislocated in the direction of the formation of chlorides and oxychlorides if one removes  $\text{O}_2$  and or adds  $\text{Cl}_2$  to the reactors atmosphere. So, for low  $P(\text{O}_2)$  ( $< 10^{-20}$  atm) and high  $P(\text{Cl}_2)$  (between 0.1 and 1 atm) values, according to the predominance diagrams of Figures (14) and (15),  $\text{VCl}_4$  should be the most stable vanadium chloride, which is produced according to Eq. (36).



K	T (K)
$1.76257 \cdot 10^{-13}$	1173
$5.82991 \cdot 10^{-11}$	1273
$1.0397 \cdot 10^{-08}$	1473

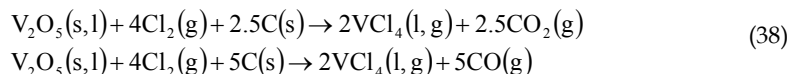
Table 2. Equilibrium constant for the reaction represented by Eq. (37)

The equilibrium constant for reaction represented by Eq. (36) is associated with very low values between 1173 K and 1473 K (see Table 2). So, it can be concluded that the formation of  $\text{VCl}_4$  has a very low thermodynamic driving force in the temperature range considered. One possibility to overcome this problem is to add to the reaction system some carbon bearing compound (Allain et al., 1997, Gonzalvez et al., 2002a; González et al., 2002b; Jena et

al., 2005). The compound decomposes producing graphite, which reacts with oxygen dislocating the chlorination equilibrium in the desired direction. A simpler route, however, would be to admit carbon as graphite together with the oxide sample into the reactor. If graphite is present in excess, the  $O_2$  concentration in the reactor's atmosphere is maintained at very low values, which are achievable through the formation of carbon oxides (Eq. 37)



So, for the production of  $VCl_4$  in the presence of graphite, the reaction of C with  $O_2$  can lead to the evolution of gaseous CO or  $CO_2$  (Eq. 38).



The effect of the presence of graphite over the  $\Delta G_r^\circ \times T$  curves for the formation of  $VCl_4$  can be seen in the diagram of Figure (17). As a matter of comparison, the plot for the formation of the same species in the absence of graphite is also shown, together with the curves for the reactions associated with the formation of CO and  $CO_2$  for one mole of  $O_2$  (Eq. 37).

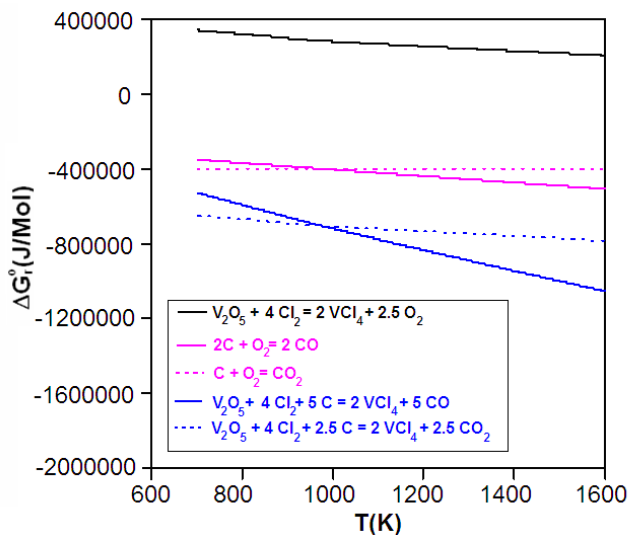
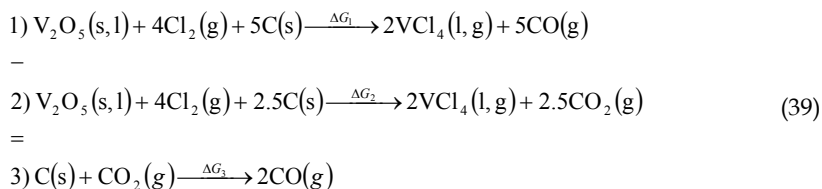


Fig. 17.  $\Delta G_r^\circ$  vs. T for the formation of  $VCl_4$

It can be readily seen that graphite strongly reduces the standard molar Gibbs energy of reaction, promoting in this way considerably the thermodynamic driving force associated with the chlorination process. The presence of graphite has also an impact over the standard molar reaction enthalpy. The direct action of  $Cl_2$  is associated with an endothermic reaction (positive linear coefficient), but by adding graphite the processes become considerably exothermic (negative linear coefficient).

The curves associated with the  $VCl_4$  formation in the presence of the reducing agent cross each other at 973 K, the same temperature where the curves corresponding to the formation of CO and  $CO_2$  have the same Gibbs energy value. This point is defined by the temperature, where the Gibbs energy of the Boudouard reaction ( $C + CO_2 = 2CO$ ) is equal to zero.

The equivalence of this point and the intersection associated with the curves for the formation of  $VCl_4$  can be perfectly understood, as the Boudouard reaction can be obtained through a simple linear combination, according to Eq. (39). So, the molar Gibbs energy associated with the Boudouard reaction is equal to the difference between the molar Gibbs energy of the  $VCl_4$  formation with the evolution of CO and the same quantity for the reaction associated with the  $CO_2$  production. When the curves for the formation of  $VCl_4$  crosses each other, the difference between their molar Gibbs energies is zero, and according to Eq. (39) the same must happen with the molar Gibbs energy of the Boudouard reaction.



$$\Delta G_3 = \Delta G_1 - \Delta G_2$$

$$\lim_{T \rightarrow 973K} (\Delta G_3) = \lim_{T \rightarrow 973K} (\Delta G_1 - \Delta G_2) = \Delta G - \Delta G = 0$$

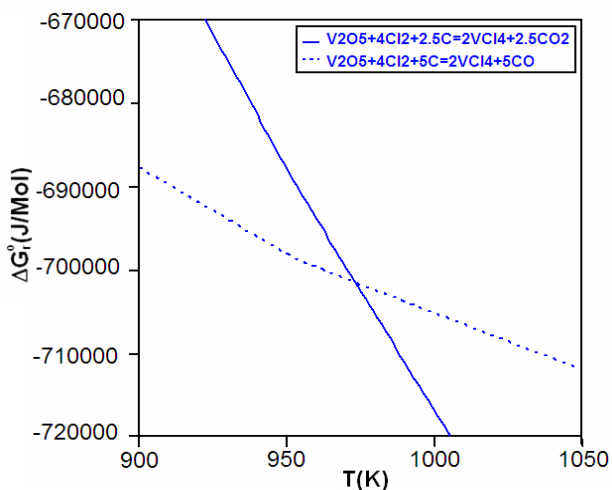


Fig. 18.  $\Delta G_f^\circ$  vs. T the formation of  $VCl_4$  – melting of  $V_2O_5$

The inflexion point present on the curves of Figure (17) is associated with the melting of  $V_2O_5$ . This inflexion is better evidenced on the graphic of Figure (18). As  $V_2O_5$  is a reactant, according to the concepts developed on topic (2.2.1), the curve should experience a

reduction of its inclination at the melting temperature of the oxide. However, the presence of the inflexion point is much more evident for the reactions with the lowest variation of number of moles of gaseous reactants, as is the case for the direct action of  $\text{Cl}_2$ , which leads to the evolution of  $\text{CO}_2$  ( $\Delta n_g = 0.5$ ).

The quantity  $\Delta n_g$  controls the molar entropy of the reaction. By lowering the magnitude  $\Delta n_g$  the value of the reaction entropy reduces, and the effect of melting of  $\text{V}_2\text{O}_5$  over the standard molar reaction Gibbs energy becomes more evident.

Based on the predominance diagrams of topic (3.1.1),  $\text{VOCl}_3$  should be formed for  $P(\text{Cl}_2)$  close to 1atm as  $P(\text{O}_2)$  gets higher. The presence of graphite has the same effect over the molar Gibbs energy of formation of  $\text{VOCl}_3$ , promoting in this way the thermodynamic driving force for the reaction. Its curve is compared with the one for the formation of  $\text{VCl}_4$  on Figure (19). The inflexion around 954 K is again associated with the melting of  $\text{V}_2\text{O}_5$ . As the reaction associated with the formation of  $\text{VCl}_4$ , the formation of  $\text{VOCl}_3$  has a negative molar reaction enthalpy. So, if the gas phase is considered ideal, for the production of both chlorinated compounds the system should transfer heat to its neighborhood (exothermic reaction).

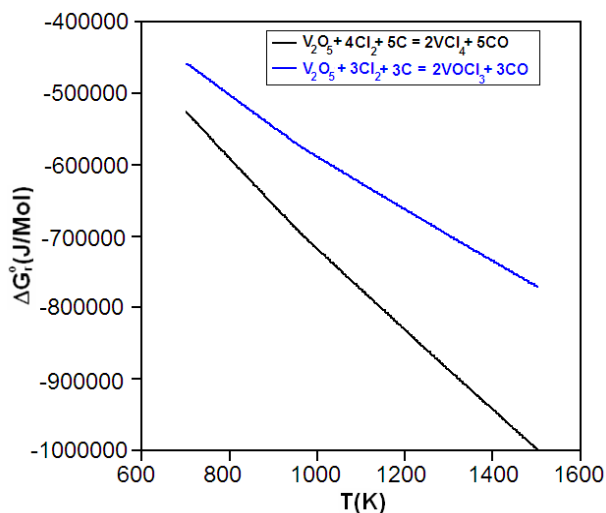


Fig. 19.  $\Delta G_r^0$  vs. T for the formation of  $\text{VOCl}_3$  and  $\text{VCl}_4$

On what touches the molar reaction entropy, the graphic of Figure (19) indicates, that the reaction associated with the formation of  $\text{VCl}_4$  should generate more entropy (more negative angular coefficient for the entire temperature range). This can be explained by the fact, that in the case of  $\text{VCl}_4$  the variation of the number of mole of gaseous reactants and products ( $\Delta n_g = 3$ ) is higher than the value for the formation of  $\text{VOCl}_3$  ( $\Delta n_g = 2$ ). This illustrates how important the magnitude of  $\Delta n_g$  is for the molar entropy of a gas - solid reaction.

Finally, it should be pointed out that the standard molar Gibbs energy has the same order of magnitude for both chlorinated species considered. So, only by appreciating the  $\Delta G_r^0 \times T$  curves of these chlorides it is impossible to tell case which species should be found in the gas with the highest concentration. This problem will be covered on topic (3.2).

### 3.1.2.1 Successive chlorination steps

As discussed on topic (2.2.1), the standard free energy vs. temperature diagram is a valuable tool for suggesting possible reactions paths. Let's consider first the formation of  $VCl_4$ . Such a process could be thought as the result of three stages. In the first one, a lower chlorinated compound ( $VCl$ ) is formed. The precursor then reacts with  $Cl_2$  resulting in higher chlorinated species (Eq. 40).

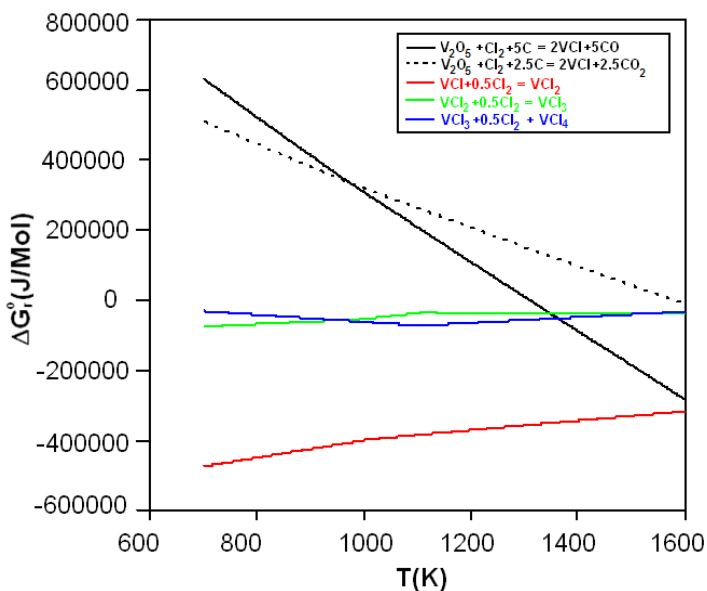
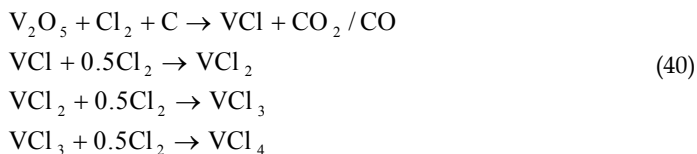


Fig. 20.  $\Delta G_r^0 \times T$  for reaction paths of Eq. (38)

The  $\Delta G_r^0 \times T$  plots associated with reactions paths represented by mechanisms of Eq. (40) were included on Figure (20). Two inflexion points are evidenced in the diagram of Figure (20). The first one around 1000 K is associated with  $VCl_2$  melting. The second one, around 1100 K, is associated with the sublimation of  $VCl_3$ . It can be deduced that only for temperatures greater than 1600 K the path described by Eq. (40) would be possible. For lower temperatures, the molar Gibbs energy of the first step is higher than the one associated with the second.

Another mechanism can be thought for the production of  $VCl_4$ . This time,  $VCl_2$  is formed first, which then reacts to give  $VCl_3$  and finally  $VCl_4$  (Eq. 41). The characteristic  $\Delta G_r^0 \times T$  curves for the reactions defined in Eq. (41) are presented on Figures (21) and (22).

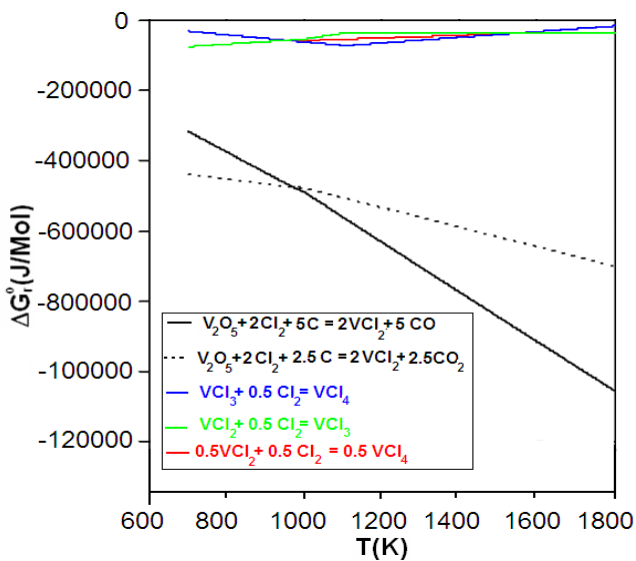


Fig. 21.  $\Delta G_r^0 \times T$  for reaction paths of Eq. (41)

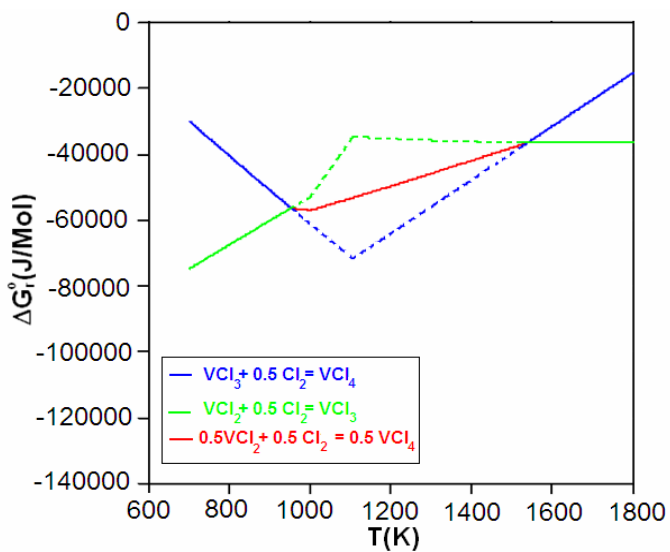


Fig. 22.  $\Delta G_r^0 \times T$  for reaction paths of Eq. (41)

The inflexion points have the same meaning as described for diagram of Figure (20). It can be seen that the first step has a much higher thermodynamic tendency as the other. Also, for temperatures lower than 953 K the second step leads to the formation of  $\text{VOCl}_3$ , which then reacts to give  $\text{VOCl}_4$ . However, for temperatures higher than 953 K and lower than 1539 K, the step associated with the formation of  $\text{VOCl}_4$  is the one with the lowest standard Gibbs energy. So, in this temperature range,  $\text{VOCl}_4$  should be formed directly from  $\text{VOCl}_2$ , as suggested by Eq. (42). In order to achieve thermodynamic consistency in the mentioned temperature interval, the curves associated with the formation of  $\text{VOCl}_3$  and  $\text{VOCl}_4$  according to Eq. (41) should be substituted for the curve associated with reaction defined by Eq. (42), which was represented with red color in the plots presented on Figures (21) and (22).

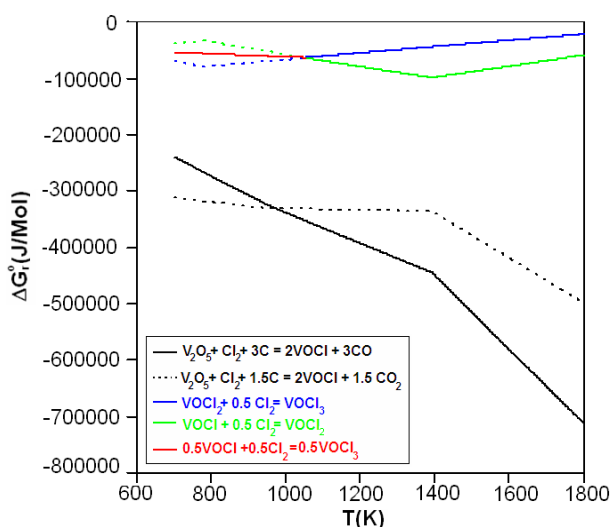
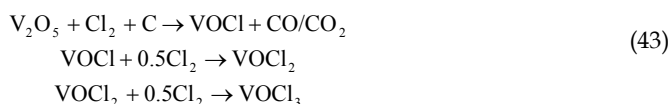


Fig. 23.  $\Delta G_r^0 \times T$  for reaction paths of Eq. (43)

For temperatures higher than 1539 K, however, the mechanism is again described by Eq. (41),  $\text{VOCl}_3$  being formed first, which then reacts leading to  $\text{VOCl}_4$ . It is also interesting to recognize that the sublimation of  $\text{VOCl}_3$  is responsible for the inversion of the behavior for temperatures higher than approximately 1400 K, where the second reaction step is again the one with the second lowest Gibbs energy of reaction.

On what touches the synthesis of  $\text{VOCl}_3$ , a reaction path can be proposed (Eq. 43), in that  $\text{VOCl}$  is formed first, which then reacts to give  $\text{VOCl}_2$ , which by itself then reacts to form  $\text{VOCl}_3$ . The  $\Delta G_r^0 \times T$  diagrams associated with these reactions are presented on Figure (23).

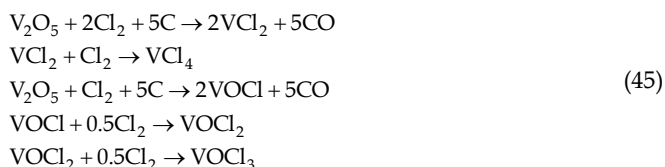




The inflexion point around 800 K is associated with the sublimation of  $\text{VOCl}_2$ , and around 1400 K with the sublimation of  $\text{VOCl}$ . According to the  $\Delta G_r^0 \times T$  curves presented on Figure (23), it can be deduced that the reaction steps will follow the proposed order only for temperatures higher than 1053 K. At lower temperatures  $\text{VOCl}_2$  should be formed directly from  $\text{VOCl}$  (Eq. 44). It is interesting to note that the sublimation of  $\text{VOCl}_2$  is the phenomenon responsible for the described inversion of behavior. Again, to attain thermodynamic consistency for temperatures higher than 1053 K, the curves associated with the formation of  $\text{VOCl}_2$  and  $\text{VOCl}_3$  according to Eq. (43) must be substituted for the curve associated with reaction represented by Eq. (44), which was drawn with red color in the diagram plotted on Figure (23). It should be mentioned indeed, that the reaction equations compared must be written with the same stoichiometric coefficient for  $\text{Cl}_2$ , or equivalently, the Gibbs energy of reaction (44) must be multiplied by 1/2.



Finally, some remarks may be constructed about the possible reaction order values in relation to  $\text{Cl}_2$ . According to the discussion developed so far, for the temperature range between 1100 K and 1400 K, Eq. (45) describes the most probable reactions paths for the formation of  $\text{VCl}_4$  and  $\text{VOCl}_3$ . As a result, depending on the nature of the slowest step, the reaction order in respect with  $\text{Cl}_2$  can be equal to one, two or  $1/2$ .

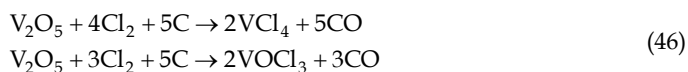


### 3.1.3 Relative stability of $\text{VCl}_4$ and $\text{VOCl}_3$

As is evident from the discussion developed on topic (3.1.2), the chlorinated compounds  $\text{VCl}_4$  and  $\text{VOCl}_3$  are the most stable species in the gas phase as the atmosphere becomes concentrated in  $\text{Cl}_2$ . The relative stability of these two chlorinated compounds will be first accessed on topic (3.1.3.1) by applying the method introduced by Kang & Zuo (1989) and secondly on topic (3.1.3.2) through computing some speciation diagrams for the gas phase.

#### 3.1.3.1 Method of Kang and Zuo

As shown in this topic (2.2.2) the concentrations of  $\text{VCl}_4$  and  $\text{VOCl}_3$  can be directly computed by considering that each chlorinated compound is generated independently. It will be assumed that the inlet gas is composed of pure  $\text{Cl}_2$  ( $P(\text{Cl}_2) = 1 \text{ atm}$ ). Further, two temperature values were investigated, 1073 K and 1373 K. At these temperatures, the presence of graphite makes the atmosphere richer in CO, so that for the computations the following reactions will be considered:



The concentrations of  $\text{VOCl}_3$  and  $\text{VCl}_4$  can then be expressed as a function of  $P(\text{CO})$  and temperature according to Eq. (47).

$$\begin{aligned}
 P_{\text{VCl}_4} &= \sqrt{P_{\text{CO}}^5 K_1} \rightarrow \ln P_{\text{VCl}_4} = \frac{\ln K_1}{2} + \frac{5}{2} \ln P_{\text{CO}} \\
 P_{\text{VOCl}_3} &= \sqrt{P_{\text{CO}}^3 K_2} \rightarrow \ln P_{\text{VOCl}_3} = \frac{\ln K_2}{2} + \frac{3}{2} \ln P_{\text{CO}}
 \end{aligned}
 \tag{47}$$

Where  $K_1$  and  $K_2$  represent, respectively, the equilibrium constants for the reactions associated with the formation of  $\text{VCl}_4$  and  $\text{VOCl}_3$  (Eq. 46). By applying Eq. (47) the partial pressure of  $\text{VCl}_4$  and  $\text{VOCl}_3$  were computed as a function of  $P(\text{CO})$ . The results were plotted on graphic contained in Figure (24). The significant magnitude of the partial pressure values computed for  $\text{VOCl}_3$  and  $\text{VCl}_4$  is a consequence of the huge negative standard Gibbs energy of reaction associated with the formation of these species in the temperature range considered (see Figure 19).

According to Figure (24),  $\text{VCl}_4$  is the chloride with the highest partial pressure for both specified temperatures. Also, for both temperatures,  $P(\text{VOCl}_3)$  becomes higher than  $P(\text{VCl}_4)$  only for significant values of  $P(\text{CO})$ . At 1073 K, for example, the partial pressures of the species have equal values only for  $P(\text{CO})$  equal to  $2.98 \cdot 10^3 \text{ atm}$ , and at 1373 K the same happens for  $P(\text{CO})$  equal to  $8.91 \cdot 10^3 \text{ atm}$ .

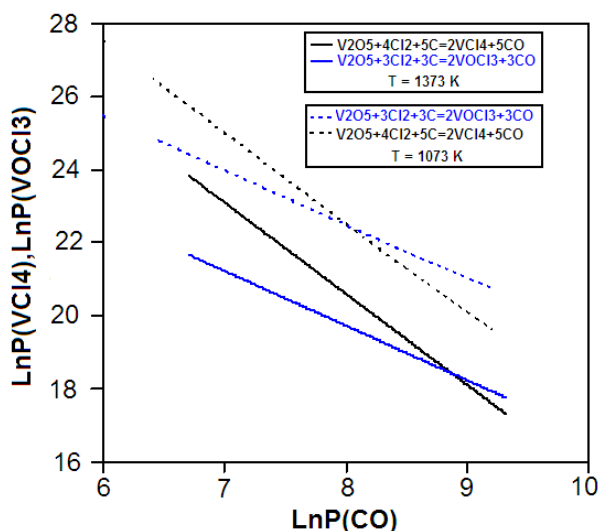


Fig. 24.  $\ln(P(\text{VCl}_4))$  and  $\ln(P(\text{VOCl}_3))$  as a function of  $\ln(P(\text{CO}))$

As a result, it is expected that the formation of gaseous  $\text{VCl}_4$  should have a much greater tendency of occurrence in the temperature range studied. These results will be confirmed through construction of speciation diagrams for the gas phase, a task that will be accomplished on topic (3.1.3.2).

### 3.1.3.2 Gas phase speciation

The construction of speciation diagrams for the gas phase enables the elaboration of a complementary picture of the chlorination process in question. The word “speciation” means the concentration of all species in gas. This brings another level of complexity to the

quantitative description of equilibrium, as the species build a solution, and as so, their concentrations must be determined at the same time. This sort of information can only arise if one solves the system of equilibrium equations associated to all possible chemical reactions involving the species that form the gas. For the present system (V - O - Cl - C) this task becomes very tricky, as the number of possible species present is pretty significant (ex. CO, CO<sub>2</sub>, O<sub>2</sub>, VCl<sub>2</sub>, VCl<sub>3</sub>, VCl, VCl<sub>4</sub>, VOCl<sub>3</sub>, VOCl), and so the number of possible chemical reactions connecting them. So, we must think in another route for simultaneously computing the concentration of the gaseous species produced by our chlorination process. The only possible way consists in minimizing the total Gibbs energy of the system (see topic 2.2.3).

The equilibrium state is defined by fixing  $T$ ,  $P$ ,  $n(\text{V}_2\text{O}_5)$ ,  $P(\text{O}_2)$  and  $P(\text{Cl}_2)$ . The number of moles of V<sub>2</sub>O<sub>5</sub> is fixed at one. If graphite is present in excess, the partial pressure of O<sub>2</sub> is controlled by according to the Boudouard equilibrium (Eq. 39), so that, its presence forces  $P(\text{O}_2)$  to attain very low values (typically lower than 10<sup>-20</sup>atm). The total pressure is fixed at 1atm and  $T$  varies in the range between 1073 K and 1473 K.

An excess of graphite is desirable, so that the chlorination reactions can achieve a considerable driving force at the desired conditions. Computationally speaking, this can be done in two ways. One possibility is to define an amount of carbon much greater than the number of moles of V<sub>2</sub>O<sub>5</sub>. Other possibility, which has been made accessible through modern computational thermodynamic software, consists in defining the phase "solid graphite" as fixed with a definite amount. The equilibrium compositions (intensive variables) are not a function of the amount of phases present (size of the system), depending only of temperature and total pressure. So we are free to choose any suitable value we desire, such for example zero ( $n_{\text{graphite}} = 0$ ). This last alternative was implemented in the computations conducted in the present topic.

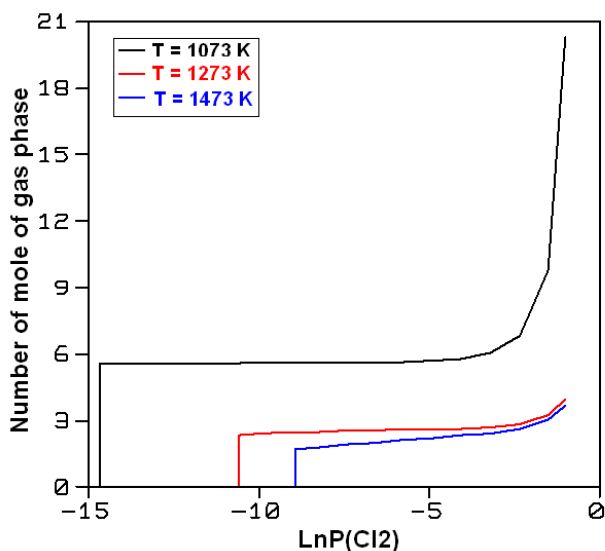


Fig. 25. Number of moles of gas as a function of  $P(\text{Cl}_2)$

On Figure (25), the number of moles of gas produced was plotted as a function of  $P(\text{Cl}_2)$  for  $T$  equal to 1073 K, 1273 K, and 1473 K. The partial pressure of  $\text{O}_2$  was fixed at  $1.93 \cdot 10^{-22}$  atm, and the partial pressure of  $\text{Cl}_2$  is varied between  $3.6 \cdot 10^{-7}$  atm and 0.61 atm.

Each curve is defined by three stages. First, for very low values of  $P(\text{Cl}_2)$ , no gas is formed. At this conditions  $\text{VCl}_2(\text{l})$  is present in equilibrium with graphite. The equilibrium ensemble does not experience any modification until a critical  $P(\text{Cl}_2)$  value is reached, at which a discontinuity can be evidenced. The gas phase appears in equilibrium and for any  $P(\text{Cl}_2)$  higher than the critical one, the number of moles  $\text{VCl}_2(\text{l})$  becomes equal to zero. This condition defines the second stage, where for higher  $P(\text{Cl}_2)$  values the gas composition changes accordingly, through forming of chlorides and oxychlorides. Finally, a  $P(\text{Cl}_2)$  value is reached, where all capacity of the system for forming chlorinated compounds is exhausted, and the effect of adding more  $\text{Cl}_2$  is only the dilution of the chlorinated species formed. As a consequence, the number of mole of gas phase experiences a significant elevation. At 1073 K, for example, Figure (26) describes the effect of  $P(\text{Cl}_2)$  over the gas phase composition during the second and third stages. We see that the mol fraction of  $\text{VCl}_4$  raises (second stage) and after achieving a maximum value starts to decrease (third stage). The concentration variations during the second stage can be ascribed to the occurrence of reactions represented by Eq. (48), which have at 1073 K equilibrium constants much higher than unity (Table 3). The reduction of the mol fraction of  $\text{VOCl}_3$  can be understood as a dilution effect, which is motivated by the elevation of the mol fractions of  $\text{VCl}_4$  and  $\text{Cl}_2$ .



K	Chemical reaction
$1.95 \cdot 10^5$	$\text{VCl}_2 + \text{Cl}_2 \rightarrow \text{VCl}_4$
$2.12 \cdot 10^3$	$\text{VCl}_3 + 0.5\text{Cl}_2 \rightarrow \text{VCl}_4$

Table 3. Equilibrium constants at 1073 K for the reactions represented by Eq. (48)

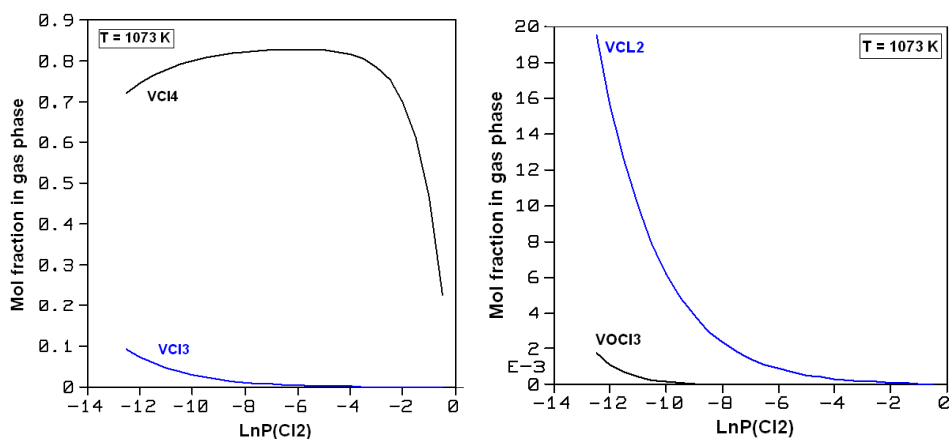


Fig. 26. Concentration of vanadium chlorides and oxychlorides as a function of  $P(\text{Cl}_2)$  at 1073 K

Besides  $P(\text{Cl}_2)$ , temperature should also have an effect over the composition of the gas phase. This was studied as follows. Six temperature values were chosen in the range between 1073 K and 1473 K. Next, for each temperature the critical  $P(\text{Cl}_2)$  value (the one associated with the formation of the first gas molecules) is identified. The composition of the most stable gaseous species is then computed and is presented in Table (4). During the calculations the partial pressure of  $\text{O}_2$  was fixed at  $1.93 \cdot 10^{-22} \text{atm}$ .

T(K)	X(CO)	X(CO <sub>2</sub> )	X(VOCl <sub>3</sub> )	X(VCl <sub>2</sub> )	X(VCl <sub>3</sub> )	X(VCl <sub>4</sub> )
1073	0.16	$3.64 \cdot 10^{-3}$	$1.95 \cdot 10^{-2}$	$1.74 \cdot 10^{-3}$	$9.27 \cdot 10^{-2}$	0.72
1100	0.12	$1.23 \cdot 10^{-3}$	$1.02 \cdot 10^{-2}$	$2.61 \cdot 10^{-3}$	0.12	0.75
1200	$4.21 \cdot 10^{-2}$	$3.36 \cdot 10^{-5}$	$1.08 \cdot 10^{-3}$	$9.87 \cdot 10^{-3}$	0.26	0.69
1300	$1.76 \cdot 10^{-2}$	$1.60 \cdot 10^{-6}$	$1.45 \cdot 10^{-4}$	$2.98 \cdot 10^{-2}$	0.43	0.52
1373	$1.00 \cdot 10^{-2}$	$2.29 \cdot 10^{-7}$	$3.69 \cdot 10^{-5}$	$5.97 \cdot 10^{-2}$	0.56	0.37
1400	$8.26 \cdot 10^{-3}$	$1.17 \cdot 10^{-7}$	$2.27 \cdot 10^{-5}$	$7.6 \cdot 10^{-2}$	0.59	0.32
1473	$5.07 \cdot 10^{-3}$	$2.18 \cdot 10^{-8}$	$6.22 \cdot 10^{-6}$	0.14	0.66	0.19

Table 4. Composition of the "first" gas formed as a function of temperature

As expected, the mol fraction of CO is greater than the mol fraction of CO<sub>2</sub> for the entire temperature range studied. Also, the chloride VCl<sub>4</sub> has the highest concentration at 1073 K, a phase which occupies a large area of the predominance diagram at this temperature (Figure 14). As temperature attains higher values, the mol fraction of VCl<sub>4</sub> and VOCl<sub>3</sub> become progressive lower and the atmosphere more concentrated in VCl<sub>2</sub> and VCl<sub>3</sub>. So, at 1473 K the situation is significant different from the equilibrium state observed at 1073 K. Such behavior is again consistent with the information contained on the predominance diagrams (Figures 14, 15 and 16) where can be seen that the stability fields of VCl<sub>4</sub>(g) and VOCl<sub>3</sub>(g) shrink while the area representing the phase VCl<sub>3</sub>(g) grows. At 1573 K it occupies a visible amount of the diagrams space (Figure 16).

It is worthwhile to mention that a more detailed look on the results seems to incorporate apparent inconsistencies. i) The minimum partial pressure of Cl<sub>2</sub> for the formation of pure VCl<sub>4</sub>(g) at 1073 K (Figure 14) is higher than the critical pressure for the formation of the first gaseous species at this temperature (Figure 25). ii) Measurable amounts of VCl<sub>3</sub> (greater or equal to 0.1) were detected for temperatures higher than 1100 K (Table 3) but no VCl<sub>3</sub>(g) field was observed in the predominance diagram computed at 1273 K (Figure 15). iii) Also, no field associated with the formation of VCl<sub>2</sub>(g) could be detected even at 1573 K (Figure 16) but the speciation computation predicts its presence in measurable amounts at the last temperature ( $x(\text{VCl}_2) = 0.14$ ) (Table 3). All these thermodynamic values differences are a consequence of the fact that the pure molar Gibbs energy of each component is higher than its chemical potential in the ideal gas solution, the former model being used for the predominance diagrams construction while the later is applied to the speciation calculations. Therefore, the driving force for the formation of the gaseous compounds is reduced accordingly to Eq. (49).

$$\mu_{\text{VCl}_3}^g - g_{\text{VCl}_3}^g = RT \ln x_{\text{VCl}_2}^g < 0 \quad (49)$$

Another possible type of computation is to study the effect of  $P(\text{O}_2)$  over the composition of the gas phase. This variable is restricted by the fact that the amount of graphite phase is fixed. So there is a maximum value of  $P(\text{O}_2)$  at each temperature for which the

thermodynamic modeling remains consistent and the computation can be performed. By fixing the temperature at 1373K, the upper limit for  $P(O_2)$  was equal to to  $1.56 \cdot 10^{-18}$ atm and the value of  $P(Cl_2)$  associated with the appearance of the first gaseous molecules is identified as  $2.05 \cdot 10^{-4}$ atm. The composition of the gas is then computed by fixing  $P(Cl_2)$  at  $2.05 \cdot 10^{-4}$ atm. Three different  $P(O_2)$  levels were studied,  $1.3 \cdot 10^{-24}$ ,  $5.4 \cdot 10^{-22}$  and  $1.56 \cdot 10^{-18}$ atm. The results are presented on Table (5).

$P(O_2)$ (atm)	X(CO)	X(CO <sub>2</sub> )	X(VOCl <sub>3</sub> )	X(VCl <sub>2</sub> )	X(VCl <sub>3</sub> )	X(VCl <sub>4</sub> )
$1.30 \cdot 10^{-24}$	$8.22 \cdot 10^{-4}$	$1.54 \cdot 10^{-9}$	$3.05 \cdot 10^{-6}$	0.0597	0.56	0.38
$5.24 \cdot 10^{-22}$	$1.65 \cdot 10^{-2}$	$6.22 \cdot 10^{-7}$	$6.03 \cdot 10^{-5}$	0.0588	0.55	0.37
$1.56 \cdot 10^{-18}$	0.902	$1.85 \cdot 10^{-3}$	$3.21 \cdot 10^{-4}$	$5.72 \cdot 10^{-3}$	0.054	0.036

Table 5. Gas phase speciation as function of  $P(O_2)$

The mol fractions of CO and CO<sub>2</sub> gets higher for higher values of  $P(O_2)$ . This is consistent with the dislocation of the equilibrium represented by Eq. (50) in the direction of the formation of the two carbon oxides.



Also, the Boudouard equilibrium demands that at the chosen temperature (1373 K) the atmosphere is more concentrated in CO. This was indeed observed for each equilibrium state investigated. It is interesting to observe that for  $P(O_2)$  varying between  $5.24 \cdot 10^{-22}$ atm and  $1.56 \cdot 10^{-18}$ atm the mol fraction of CO and CO<sub>2</sub> experience a much higher variation in comparison with the one observed for lower  $P(O_2)$  values.

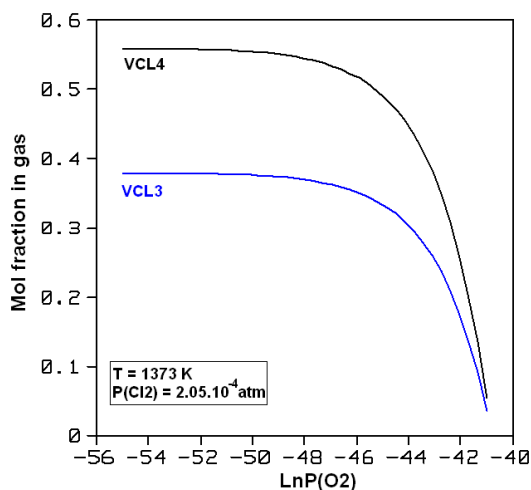


Fig. 27. Mol fractions of VCl<sub>3</sub> and VCl<sub>4</sub> as a function of  $P(O_2)$

In the case of the vanadium chlorides and oxychlorides an interesting trend is evidenced. The concentration of VOCl<sub>3</sub> grows and of VCl<sub>2</sub>, VCl<sub>3</sub> and VCl<sub>4</sub> reduce appreciably for the same O<sub>2</sub>

partial pressure range. The concentration variations associated with the vanadium chlorinated compounds is analogous to the variations observed in the concentrations of CO and CO<sub>2</sub>.

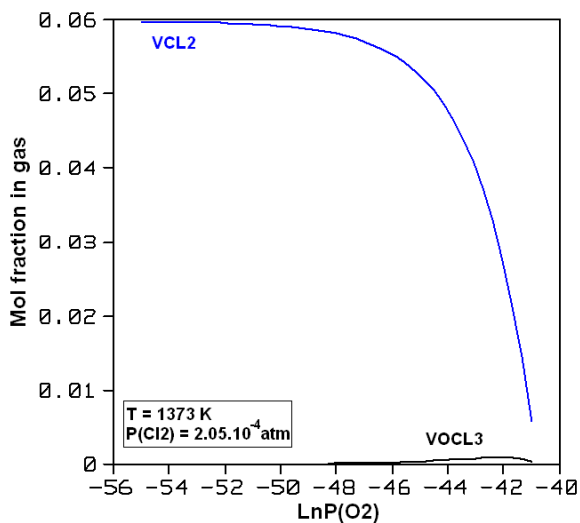


Fig. 28. Mol fractions of  $\text{VOCl}_3$  and  $\text{VCl}_2$  as a function  $P(\text{O}_2)$

For  $P(\text{O}_2)$  lower than  $5.24 \cdot 10^{-22} \text{ atm}$  the variations are much less significant. To get a better picture of the trend observed for the chlorides and oxychlorides, their concentrations were plotted as a function of  $P(\text{O}_2)$ , which was varied in the range spanned by the data of Table (5) (Figures 27, 28 and 29)

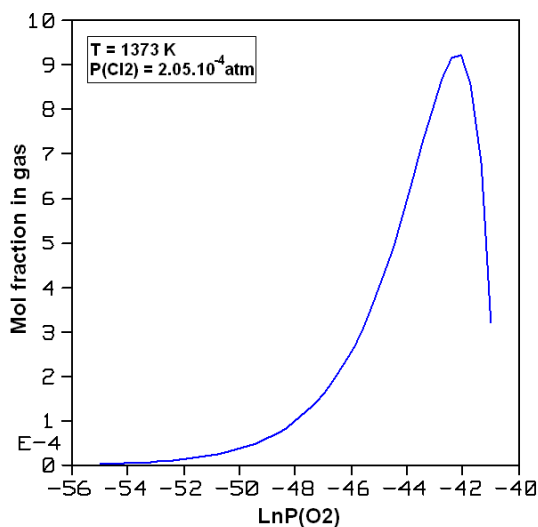
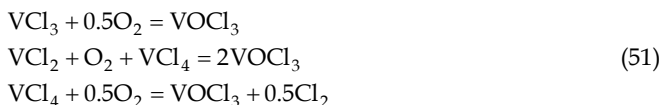


Fig. 29. Mol fraction of  $\text{VOCl}_3$  as a function of  $P(\text{O}_2)$

The variations depicted on Figures (27), (28) and (29) are consistent with the occurrence of reactions represented by Eq. (51). As  $P(\text{O}_2)$  achieves higher values, it reacts with  $\text{VCl}_3$ ,  $\text{VCl}_2$  and or  $\text{VCl}_4$  resulting in  $\text{VOCl}_3$ . Such phenomena could explain the significant reduction of  $\text{VCl}_3$ ,  $\text{VCl}_4$  and  $\text{VCl}_2$  concentrations, and the concomitant elevation of the  $\text{VOCl}_3$  mol fraction.



The participation of  $\text{VCl}_4$  in the second reaction is supported by the fact that its equilibrium concentration lowering is more sensible to  $\text{Ln}P(\text{O}_2)$  than observed for  $\text{VCl}_3$  (Figure 27). The consumption of  $\text{VCl}_4$  by the second reaction is also consistent with the maximum observed in the curve obtained for  $\text{VOCl}_3$  concentration (Figure 29). As less  $\text{VCl}_4$  is available, less  $\text{VOCl}_3$  can be produced.

K	Chemical reaction
8.01.10 <sup>6</sup>	$\text{VCl}_3 + 0.5\text{O}_2 = \text{VOCl}_3$
1.89.10 <sup>13</sup>	$\text{VCl}_2 + \text{O}_2 + \text{VCl}_4 = 2\text{VOCl}_3$
1.01.10 <sup>5</sup>	$\text{VCl}_4 + 0.5\text{O}_2 = \text{VOCl}_3 + 0.5\text{Cl}_2$

Table 6. Equilibrium constants at 1373 K for reactions represented by Eq. (51)

The occurrence of reactions represented by Eq. (51) is supported by classical thermodynamics, as the equilibrium constant ( $K$ ) computed at 1373 K for all chemical reactions above assume values appreciably greater than unity (see Table 6).

### 3.1.3.3 V2O5 chlorination enthalpy

For the implementation of an industry process based on chemical reaction is fundamental to know the amount of heat generated or absorbed from that. Exothermic processes (heat is released) reach higher temperatures, and frequently demand engineering solutions for protecting the oven structure against the tremendous heat generated by the chemical phenomena. In this context, endothermic processes (heat is absorbed) are easier controlled, but the energy necessary to stimulate the reaction must be continuously supplied, making the energy investment larger.

The variation of the total enthalpy of the system for the chlorination process in question was calculated as a function of  $P(\text{Cl}_2)$  (Figure 30). The partial pressure of  $\text{O}_2$  was forced to be equal to  $1.93 \cdot 10^{-22}$  atm and four temperature levels were studied, 1000 K, 1100 K, 1300 K and 1700 K. It can be seen that the total enthalpy for the process conducted at 1000 K reduces with the advent of the chlorination reactions, indicating that the chlorination process is exothermic. However, the molar enthalpy magnitude is progressively lower up to a certain temperature where it is zero. Above that, the molar reaction enthalpy becomes positive, and Figure (30) illustrates its value for 1700 K.

This is perfectly consistent with the results presented on Table (3). As temperature gets higher, the mol fractions of  $\text{VCl}_4$  and  $\text{VOCl}_3$  reduce and that for  $\text{VCl}_3$  and  $\text{VCl}_2$  experience a significant elevation. For some temperature between 1300 K and 1373 K the mol fractions of  $\text{VCl}_4$  and  $\text{VCl}_3$  assume equal values. This point is related to the condition where the chlorination enthalpy is zero. For higher temperatures, where  $x(\text{VCl}_4) < x(\text{VCl}_3)$  the process



becomes progressively more endothermic. It is interesting to see that the explained behavior is consistent with the fact that the global formation reactions of  $\text{VCl}_3$  and  $\text{VCl}_2$  are associated with positive molar reaction enthalpies and that of  $\text{VOCl}_3$  and  $\text{VCl}_4$  with negative molar reaction enthalpies (Figure 31).

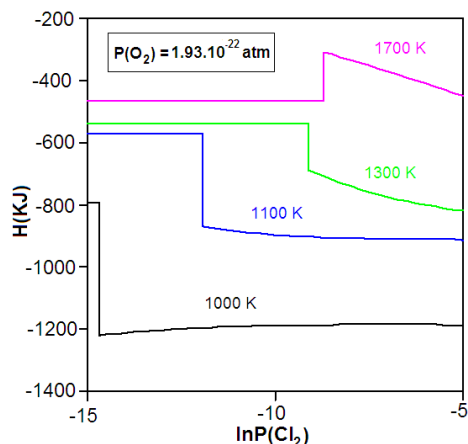


Fig. 30. Total as a function of  $P(\text{Cl}_2)$

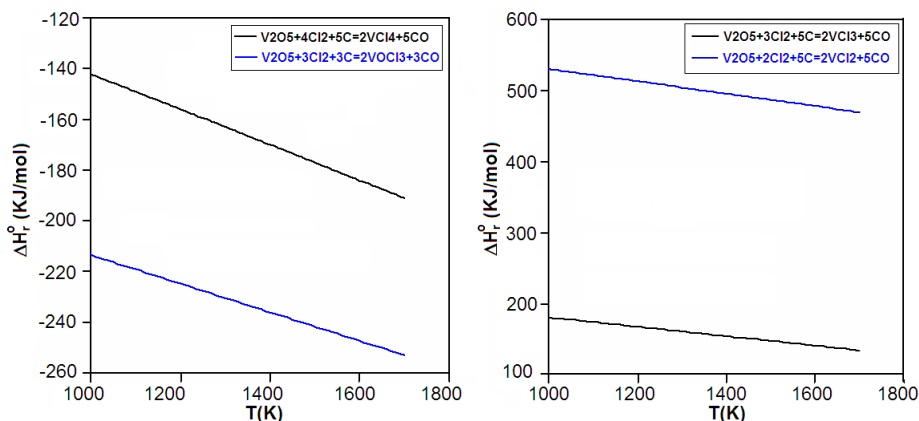


Fig. 31. Molar reaction enthalpy for the formation of gaseous  $\text{VCl}_3$ ,  $\text{VCl}_2$ ,  $\text{VCl}_4$  and  $\text{VOCl}_3$  as a function of temperature

#### 4. Final remarks

In this chapter three different approaches to the chlorination equilibrium study of an oxide were presented. The first two are based on the construction of  $\Delta G_r^0 \times T$  diagrams (topic 2.2.1) and on the calculations, first introduced by Kang & Zuo (1989) (topic 2.2.2), respectively. Both of them take into consideration that each chlorinated compound is produced independently. The third one has its fundamental based on the total Gibbs energy

minimization of the reaction system and the gas phase equilibrium composition is calculated considering that the formed species are produced simultaneously (topic 2.2.3).

The method based on the construction of  $\Delta G_r^\circ \times T$  was applied on topic (3.1.2) for studying the thermodynamic viability of the reaction between gaseous  $\text{Cl}_2$  and  $\text{V}_2\text{O}_5$ . The discussion evidenced that the chlorination is thermodynamically feasible only in the presence of a reducing agent (graphite in the case of the present work) and was initially focused on the production of  $\text{VCl}_4$  and  $\text{VOCl}_3$ . The same approach was employed for studying the possible mechanisms associated with the formation of  $\text{VCl}_4$  and  $\text{VOCl}_3$ . According to the results (topic 3.1.2.1), the synthesis of these two compounds is subdivided in different stages, which can vary in nature, depending on the temperature range considered. In global terms though, both  $\text{VOCl}_3$  and  $\text{VCl}_4$  have molar reaction Gibbs energies of the same magnitude order. So, it was not possible to clearly identify, which one of them should be produced in greater quantities. The problem of the relative stability between  $\text{VOCl}_3$  and  $\text{VCl}_4$  was then addressed by the implementation of the method of Kang & Zuo (1989) (topic 3.1.3.1). The results indicated that  $\text{VCl}_4$  should have a higher concentration in comparison with  $\text{VOCl}_3$  in the temperature range between 1073 K and 1373 K.

It can be said that both, the method based on the  $\Delta G_r^\circ \times T$  diagrams construction as well as the Kang & Zuo method (1989), incorporate some simplifications and are very easy to implement. However, they lead to only a superficial knowledge of the true nature of the equilibrium state achievable. Thanks to the development of computational thermodynamic software, of which *Thermocalc* is a good example, more complex computations can be realized. For example, by allowing the chlorides and oxychlorides to build a gaseous solution, the minimization of the total Gibbs energy of the system results in the direct computation of the mol fraction of each chlorinated species present in the gas phase (topic 3.1.3.2). This method can be seen as an improvement of the idea put forward by Kang & Zuo (1989), in that all equilibrium equations are solved simultaneously, with the further advantage that one does not need to formulate a group of independent reactions that cover all possible chemical interactions among the components, a task that can become very complex for metals, as in case of vanadium, which can produce a family of chlorides and oxychlorides.

The conclusion that graphite strongly promotes the thermodynamic driving force necessary to chlorination and that  $\text{VCl}_4$  should be formed preferentially in relation to  $\text{VOCl}_3$  are perfectly consistent with the results based on the total Gibbs energy minimization. However, by the application of this last method it was possible to go a little further, through investigation of the effect of  $P(\text{Cl}_2)$  over the chlorination enthalpy and by studying the effect of temperature,  $\text{Cl}_2$  and  $\text{O}_2$  partial pressures over the concentrations of vanadium chlorides and oxychlorides in the gas phase.

The predictions associated with the effect of temperature over the gas phase speciation (Table 4) indicate that the mol fractions of  $\text{VCl}_2$  and  $\text{VCl}_3$  grow significantly in the range between 1073 K and 1473 K and, as a result, the concentrations of  $\text{VOCl}_3$ ,  $\text{VCl}_4$ , CO and  $\text{CO}_2$  exhibit a significant reduction (topic 3.1.3.2). This finding agrees with the tendency depicted by the predominance diagrams constructed for the system V - O - Cl, where the  $\text{VCl}_4(\text{g})$  and  $\text{VOCl}_3(\text{g})$  fields shrink and that of  $\text{VCl}_3(\text{g})$  grows (Figures 14, 15 and 16). Also, the calculated mol fraction of CO is at all temperatures much higher than the mol fraction of  $\text{CO}_2$ , a fact that is consistent with the establishment of the Bourdouard equilibrium for temperatures higher than 973 K, where the concentrations of the two mentioned carbon oxides have the same magnitude.

The fact that the speciation computation indicates appreciable amounts of  $\text{VCl}_3$  for temperatures higher than 1100 K and of  $\text{VCl}_2$  for temperatures higher than 1473 K,

apparently contradicting the information contained in the predominance diagrams of Figures (15) and (16), is a mere consequence of the fact that on topic (3.1.3.2) the gaseous chlorides build an ideal gas solution. The chemical potentials of  $VCl_3$  and also of  $VCl_2$  are lower than their pure molar Gibbs energies. The species become more stable in the gaseous solution in relation to the pure state, and their mol fractions assume higher values for the same temperature imposed. The same idea explains why  $VCl_4$  is formed in significant amounts at 1073 K for a  $P(Cl_2)$  value lower than the one observed in the predominance diagram of Figure (14) for the equilibrium between  $VCl_2(l)$  and  $VCl_4(g)$ .

The effect of adding more  $Cl_2$  after all vanadium has been converted to gaseous chlorinated compounds is also consistent with the expectations. At 1073 K the results indicate that the mol fraction of  $VCl_4$  grows while all other relevant chlorinated species reduces (topic 3.1.3.2). This can be explained by the reaction of  $VCl_2$  and  $VCl_3$  with  $Cl_2$  resulting in  $VCl_4$ , which have a significant negative driving force at 1073 K (Table 3).

The study of the impact of varying  $P(O_2)$  over the gas phase composition at 1373 K indicated that the mol fractions of CO and  $CO_2$  experience significant elevation as  $P(O_2)$  becomes higher, a fact that is also observed in the case of  $VOCl_3$  (Table 4). The concentration of all other chlorinated compounds reduces for the same studied range of  $P(O_2)$ . The influence of the oxygen chemical activity over the gas phase speciation can be explained by a group of proposed reactions between  $VCl_4$ ,  $VCl_3$  and  $VCl_2$  with  $O_2$  resulting in  $VOCl_3$  (topic 3.1.3.2). All these reactions have equilibrium constants much higher than one, indicating an expressive thermodynamic driving force at 1373 K (Table 6).

The conclusions about the exothermic nature of the chlorination process in the temperature range between 1000 K and 1300 K and the observation that it becomes progressively more endothermic as 1700 K is approached (topic 3.1.3.3), are perfectly consistent with the fact that the atmosphere becomes progressively diluted in  $VCl_4$  and  $VOCl_3$ , whose formations are associated with negative molar enthalpies and becomes richer in  $VCl_2$  and  $VCl_3$ , whose molar enthalpy of formation are considerably positive (Figure 31).

Finally, we can conclude that the study of the equilibrium states achievable through the reaction between a transition metal oxide and gaseous  $Cl_2$ , can be now approached through the implementation of methods of different complexity levels. The most general one, in which the total Gibbs energy of the reaction system is minimized, enables the construction of a more detailed picture of the equilibrium state. However, as it is evident from the comparisons explained above, the most general method must be consistent with the tendencies predicted by simpler calculations.

## 5. References

- Allain E., Djona M., Gaballah I. Kinetics of Chlorination and Carbochlorination of Pure Tantalum and Niobium Pentoxides. Metallurgical and materials transactions B, v. 28, p. 223 - 232, 1997.
- Brewer L., Ebinghaus, B. B. The thermodynamics of the solid oxides of vanadium. Thermochimica Acta, v. 129, p. 49 - 55, 1988.
- Brocchi, E. A.; Moura, F. J. Chlorination methods applied to recover refractory metals from tin slags. Minerals Engineering, v. 21, n. 2, p. 150-156, 2008.
- Cecchi, E. et al. A feasibility study of carbochlorination of chrysotile tailings. International Journal of Mineral Processing, v. 93, n. 3 - 4, p. 278-283, 2009.
- Esquivel, M. R., Bohé, A. E., Pasquevich, D. M. Carbochlorination of samarium sesquioxide. Thermoquimica Acta, v. 403, p. 207 - 218, 2003.

- Gaballah, I., Djona, M. Recovery of Co, Ni, Mo, and V from unroasted spent hydrorefining catalysts by selective chlorination. *Metallurgical and Materials Transactions B*, v. 26, n. 1, p. 41-50, 1995.
- Gaviria, J. P., Bohe, A. E. Carbochlorination of yttrium oxide. *Thermochimica Acta*, v. 509, n. 1-2, p. 100-110, 2010.
- Gonzalez J. et al.  $\beta$ - $\text{Ta}_2\text{O}_5$  Carbochlorination with different types of carbon. *Canadian Metallurgical Quarterly*, v. 41, n.1, p. 29 - 40, 2002.
- Hackert A., Plies V. and Gruehn R. Nachweis und thermochemische Charakterisierung des Gasphasenmolekuls  $\text{VOCl}$ . *Zeitschrift für anorganische und allgemeine Chemie*, v. 622, p. 1651-1657, 1996.
- Hildenbrand, D. L. et al. Thermochemistry of the Gaseous Vanadium Chlorides  $\text{VCl}$ ,  $\text{VCl}_2$ ,  $\text{VCl}_3$ , and  $\text{VCl}_4$ . *Journal of Physical Chemistry A*, v. 112, p. 9978 - 9982, 2008.
- Jena, P. K.; Brocchi, E. A.; Gonzalez, J. Kinetics of low-temperature chlorination of vanadium pentoxide by carbon tetrachloride vapor. *Metallurgical and Materials Transactions B*, v. 36, n. 2, p. 195-199, 2005.
- Kang, S. X.; Zuo, Y. Z. Chloridizing roasting of complex material containing low tin and high iron at high temperature. *Kunming Metall. Res. Inst. Report*, v. 89, n. 3, 1989.
- Kellog, H. H. Thermodynamic relationships in chlorine metallurgy. *Journal of metals*, v. 188, p. 862 - 872, 1950.
- Mackay, D. Is chlorine the evil element? *Environmental Science & Engineering*, p. 49 - 52, 1992.
- McCarley R. E., Roddy J. W. The Vapor Pressures of Vanadium(II) Chloride, Vanadium(III) Chloride, Vanadium(II) Bromide, and Vanadium(III) Bromide by Knudsen Effusion. *Inorganic Chemistry*, v. 3. n. 1, p. 60 - 63, 1964.
- Micco, G.; Bohe, A. E.; Sohn, H. Y. Intrinsic Kinetics of Chlorination of  $\text{WO}_3$  Particles With  $\text{Cl}_2$  Gas Between 973 K and 1223 K (700°C and 950 °C). *Metallurgical and Materials Transactions B*, v. 42, n. 2, p. 316-323, 2011.
- Murase, K. et al. Recovery of vanadium, nickel and magnesium from a fly ash of bitumen-in-water emulsion by chlorination and chemical transport. *Journal of Alloys and Compounds*, v. 264, n. 1 - 2, p. 151-156, 1998.
- Neff, D. V. Environmentally acceptable chlorination processes. *Aluminium Cast House Technology: Theory & Practice, Australasian, Asian, Pacific Conference*, p. 211-225, 1995.
- Oheda, M. W., Rivarola, J. B., Quiroga, O. D. Study of the chlorination of molybdenum trioxide mixed with carbon black. *Minerals Engineering*, v. 15, p. 585 - 591, 2002.
- Oppermann, H. Preparation and properties of  $\text{VCl}_4$ ,  $\text{VCl}_3$ , and  $\text{VOCl}_3$ . *Zeitschrift für Chemie*, v. 2, p. 376 - 377, 1962.
- Oppermann H. Gleichgewichte mit  $\text{VOCl}_3$ ,  $\text{VO}_2\text{Cl}$ , und  $\text{VOCl}_2$ . *Zeitschrift für anorganische und allgemeine Chemie*, v. 331, n. 3 - 4, p. 113 - 224, 1967.
- Patel, C. C., Jere, G. V. Some Thermodynamical considerations in the chlorination of Ilmetite. *Transactions of the metallurgical society of AIME*, v. 218, p. 219 - 225, 1960.
- Pilgrim, R. F., Ingraham, T. R. Thermodynamics of chlorination of iron, cobalt, nickel and copper sulphides. *Canadian Metallurgical Quarterly*, v. 6, n. 4, p. 333 - 346, 1967.
- Robert, T. D. Thermodynamics in materials science, 1993.
- Sano N., Belton, G. The thermodynamics of chlorination of Vanadium Pentoxide. *Transactions of the Japan Institute of Metals*, v.21, n. 9, p. 597 - 600, 1980.
- Schäffer H., Wartenpfehl H. Über die Herstellung von  $\text{VOCl}$ . *Journal of the less common metals*, v. 3, p. 29 - 33, 1961.
- Zhang, L. et al. Rare earth extraction from bastnaesite concentrate by stepwise carbochlorination - chemical vapor transport-oxidation. *Metallurgical and Materials Transactions B*, v. 35, n. 2, p. 217-221, 2004.

# Thermodynamics of Reactions Among $\text{Al}_2\text{O}_3$ , $\text{CaO}$ , $\text{SiO}_2$ and $\text{Fe}_2\text{O}_3$ During Roasting Processes

Zhongping Zhu, Tao Jiang, Guanghui Li, Yufeng Guo and Yongbin Yang  
*School of Minerals Processing & Bioengineering,  
Central South University, Changsha, Hunan 410083,  
China*

## 1. Introduction

The thermodynamic of the chemical reactions among  $\text{Al}_2\text{O}_3$ ,  $\text{CaO}$ ,  $\text{SiO}_2$  and  $\text{Fe}_2\text{O}_3$  in the roasting processes was investigated in this chapter. The chemical reactions are classified into  $\text{SiO}_2$ - $\text{Al}_2\text{O}_3$  system,  $\text{Fe}_2\text{O}_3$ - $\text{Al}_2\text{O}_3$  system,  $\text{SiO}_2$ - $\text{Fe}_2\text{O}_3$  system,  $\text{CaO}$ - $\text{Al}_2\text{O}_3$  system,  $\text{SiO}_2$ - $\text{CaO}$  system,  $\text{SiO}_2$ -calcium aluminates system,  $\text{CaO}$ - $\text{Fe}_2\text{O}_3$  system,  $\text{Al}_2\text{O}_3$ -calcium ferrites system and  $\text{Al}_2\text{O}_3$ - $\text{CaO}$ - $\text{SiO}_2$ - $\text{Fe}_2\text{O}_3$  system. When the roasting temperature is over 1100K,  $3\text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$  is preferentially formed in  $\text{SiO}_2$ - $\text{Al}_2\text{O}_3$  system;  $\text{FeO} \cdot \text{Al}_2\text{O}_3$  can be formed in  $\text{Fe}_2\text{O}_3$ - $\text{Al}_2\text{O}_3$  system; ferric oxide and  $\text{SiO}_2$  could not generate iron silicate;  $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$  is preferentially formed in  $\text{CaO}$ - $\text{Al}_2\text{O}_3$  system when one mole  $\text{Al}_2\text{O}_3$  reacts with  $\text{CaO}$ ;  $2\text{CaO} \cdot \text{SiO}_2$  is preferentially formed in  $\text{SiO}_2$ - $\text{CaO}$  system; except for  $\text{CaO} \cdot 2\text{Al}_2\text{O}_3$  and  $\text{CaO} \cdot \text{Al}_2\text{O}_3$ , the other calcium aluminates can transform into calcium silicate by reacting with  $\text{SiO}_2$  in  $\text{SiO}_2$ -calcium aluminates system;  $2\text{CaO} \cdot \text{Fe}_2\text{O}_3$  is preferentially formed in  $\text{CaO}$ - $\text{Fe}_2\text{O}_3$  system; alumina is unable to form  $3\text{CaO} \cdot \text{Al}_2\text{O}_3$  with calcium ferrites ( $2\text{CaO} \cdot \text{Fe}_2\text{O}_3$  and  $\text{CaO} \cdot \text{Fe}_2\text{O}_3$ ), but able to form  $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$  with  $2\text{CaO} \cdot \text{Fe}_2\text{O}_3$ ; when  $\text{CaO}$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{SiO}_2$  coexist, they are more likely to form ternary compound  $2\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{SiO}_2$  and  $4\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{Fe}_2\text{O}_3$ .

## 2. Binary compounds

### 2.1 $\text{Fe}_2\text{O}_3$ - $\text{Al}_2\text{O}_3$ - $\text{CaCO}_3$ system

$\text{Fe}_2\text{O}_3$  and  $\text{Al}_2\text{O}_3$  can all react with limestone during roasting to generate corresponding aluminates and ferrites. In  $\text{Fe}_2\text{O}_3$ - $\text{Al}_2\text{O}_3$ - $\text{CaO}$  system, the reaction  $\text{Fe}_2\text{O}_3$  and  $\text{Al}_2\text{O}_3$  with  $\text{CaCO}_3$  coexist, and the reactions equations are as followed:

Reactions	A, J/mol	B, J/K.mol	Temperature, K
$\text{CaCO}_3 + \text{Al}_2\text{O}_3 = \text{CaO} \cdot \text{Al}_2\text{O}_3 + \text{CO}_2$	161088.3	-244.1	298~1200
$\text{CaCO}_3 + \text{Fe}_2\text{O}_3 = \text{CaO} \cdot \text{Fe}_2\text{O}_3 + \text{CO}_2$	151677.8	-220.9	298~1200

Table 1. The  $\Delta G_T^\theta$  of  $\text{Fe}_2\text{O}_3$ - $\text{Al}_2\text{O}_3$ - $\text{CaCO}_3$  system ( $\Delta G_T^\theta = A + BT$ , J/mol;  $P_{\text{CO}_2} = 30\text{Pa}$ , i.e., the partial pressure of  $\text{CO}_2$  in the air)

The relationships between Gibbs free energy ( $\Delta G_T^\theta$ ) and temperature (T) are as shown in figure 1.

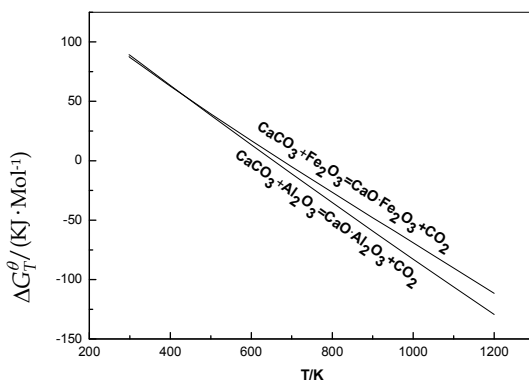


Fig. 1. Relationships between  $\Delta G_T^\theta$  and temperature in  $\text{Fe}_2\text{O}_3\text{-Al}_2\text{O}_3\text{-CaCO}_3$  system

Figure 1 shows that, the Gibbs free energy of reactions on  $\text{Fe}_2\text{O}_3$  and  $\text{Al}_2\text{O}_3$  with  $\text{CaCO}_3$  decreased with the rise of temperature in normal roasting process (due to decomposition of  $\text{CaCO}_3$  over 1200K, so the curve has no drawing above 1200K), and the reactions all can automatically react to generate the corresponding calcium aluminate and calcium ferrite. The  $\Delta G_T^\theta$  of reaction with  $\text{Al}_2\text{O}_3$  is more negative than the  $\Delta G_T^\theta$  of reaction with  $\text{Fe}_2\text{O}_3$  at the same temperature.  $\text{CaCO}_3$  has actually decomposed at 1473~1673K industrial roasting temperature, therefore, only  $\text{CaO}$  is taken into account on the following analysis.

## 2.2 $\text{SiO}_2\text{-Al}_2\text{O}_3$ system

$\text{SiO}_2$  mainly comes from the ore and coke ash in the roasting process.  $\text{SiO}_2$  reacts with  $\text{Al}_2\text{O}_3$  to form aluminium silicates. The aluminium silicates mainly include  $\text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$  ( $\text{AS}_2$ ),  $\text{Al}_2\text{O}_3 \cdot \text{SiO}_2$  ( $\text{AS}$ , andalusite),  $\text{AS}$  (kyanite),  $\text{AS}$  (fibrolite),  $3\text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$  ( $\text{A}_3\text{S}_2$ ). Thermodynamic calculation indicates that,  $\text{AS}_2$  can not be formed from the reaction of  $\text{Al}_2\text{O}_3$  and  $\text{SiO}_2$  under the roasting condition. The others equations are shown in table 2.

Reactions	A, J/mol	B, J/K.mol	Temperature, K
$\text{Al}_2\text{O}_3 + \text{SiO}_2 = \text{Al}_2\text{O}_3 \cdot \text{SiO}_2$ (kyanite)	-8469.3	9.0	298~1696
$\text{Al}_2\text{O}_3 + \text{SiO}_2 = \text{Al}_2\text{O}_3 \cdot \text{SiO}_2$ (fibrolite)	-4463.8	-0.9	298~1696
$\text{Al}_2\text{O}_3 + \text{SiO}_2 = \text{Al}_2\text{O}_3 \cdot \text{SiO}_2$ (andalusite)	-6786.1	0.6	298~1696
$\frac{3}{2}\text{Al}_2\text{O}_3 + \text{SiO}_2 = (\frac{1}{2})3\text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$	12764.7	-16.7	298~1696

Table 2. The  $\Delta G_T^\theta$  of  $\text{Al}_2\text{O}_3\text{-SiO}_2$  system ( $\Delta G_T^\theta = A + BT$ , J/mol)

The relationships of  $\Delta G_T^\theta$  and temperature in  $\text{Al}_2\text{O}_3\text{-SiO}_2$  system is shown in figure 2.

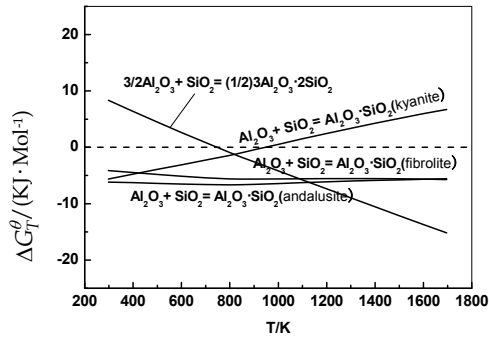


Fig. 2. Relationships of  $\Delta G_T^\theta$  and temperature in  $\text{Al}_2\text{O}_3$ - $\text{SiO}_2$  system

Figure 2 shows that, the  $\Delta G_T^\theta$  of kyanite is greater than zero at 1000~1700K, so the reaction cannot happen; the  $\Delta G_T^\theta$  of andalusite and fibrolite alter little with temperature changes; the  $\Delta G_T^\theta$  of  $\text{A}_3\text{S}_2$  decreases with the rise of temperature. The thermodynamic order of forming aluminium silicates is  $\text{A}_3\text{S}_2$ , AS(andalusite), AS(fibrolite) at 1100~1700K.

### 2.3 $\text{Fe}_2\text{O}_3$ - $\text{Al}_2\text{O}_3$ system

$\text{Al}_2\text{O}_3$  does not directly react with  $\text{Fe}_2\text{O}_3$ , but  $\text{Al}_2\text{O}_3$  may react with wustite ( $\text{FeO}$ ) produced during roasting process to form  $\text{FeO} \cdot \text{Al}_2\text{O}_3$ . No pure ferrous oxide ( $\text{FeO}$ ) exists in the actual process. The ratio of oxygen atoms to iron atoms is more than one in wustite, which is generally expressed as  $\text{Fe}_x\text{O}$  ( $x=0.83\sim 0.95$ ), whose crystal structure is absence type crystallogoly. For convenience,  $\text{FeO}$  is expressed as wustite in this thesis.  $\text{Al}_2\text{O}_3$  may react with wustite( $\text{FeO}$ ) to form  $\text{FeO} \cdot \text{Al}_2\text{O}_3$  in the roasting process. The relationship of  $\Delta G_T^\theta$  and temperature is shown in figure 2, and the chemical reaction of the equation is as followed:

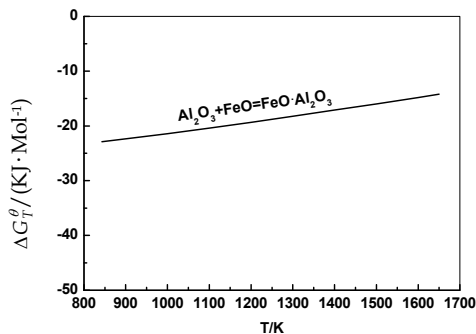
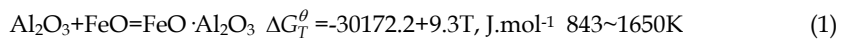


Fig. 3. Relationship of  $\Delta G_T^\theta$  and temperature in  $\text{Fe}_2\text{O}_3$ - $\text{Al}_2\text{O}_3$  system

Figure 3 shows that, the  $\Delta G_T^\theta$  is negative at 843~1650K, reaction can happen and generate  $\text{FeO} \cdot \text{Al}_2\text{O}_3$ ; the  $\Delta G_T^\theta$  rises with the temperature, the higher temperature is, the lower thermodynamic reaction trends.

#### 2.4 $\text{SiO}_2\text{-Fe}_2\text{O}_3$ system

$\text{SiO}_2$  also does not directly react with  $\text{Fe}_2\text{O}_3$ , but  $\text{Al}_2\text{O}_3$  may react with wustite ( $\text{FeO}$ ) to form  $\text{FeO} \cdot \text{SiO}_2$  (FS) and  $2\text{FeO} \cdot \text{SiO}_2$  ( $\text{F}_2\text{S}$ ). The relationships of  $\Delta G_T^\theta$  and temperature is shown in figure 4, and the chemical reactions of the equations are shown in table 3.

Reactions	A, J/mol	B, J/K.mol	Temperature, K
$\text{FeO} + \text{SiO}_2 = \text{FeO} \cdot \text{SiO}_2$	26524.6	18.8	847~1413
$2\text{FeO} + \text{SiO}_2 = 2\text{FeO} \cdot \text{SiO}_2$	-13457.3	30.3	847~1493

Table 3. The  $\Delta G_T^\theta$  of  $\text{SiO}_2\text{-Al}_2\text{O}_3$  system ( $\Delta G_T^\theta = A + BT$ , J/mol)

Figure 4 shows that, the  $\Delta G_T^\theta$  of  $\text{SiO}_2\text{-Al}_2\text{O}_3$  system are above zero at 847~1500K, so all of the reactions can not happen to form ferrous silicates (FS and  $\text{F}_2\text{S}$ ).

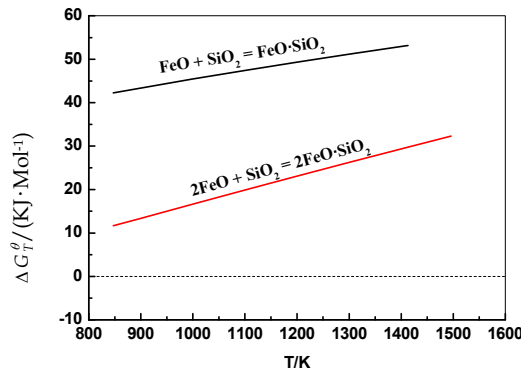


Fig. 4. Relationships of  $\Delta G_T^\theta$  and temperature in  $\text{SiO}_2\text{-Fe}_2\text{O}_3$  system

#### 2.5 $\text{CaO-Al}_2\text{O}_3$ system

$\text{Al}_2\text{O}_3$  can react with  $\text{CaO}$  to form calcium aluminates such as  $3\text{CaO} \cdot \text{Al}_2\text{O}_3$  ( $\text{C}_3\text{A}$ ),  $12\text{CaO} \cdot 7\text{Al}_2\text{O}_3$  ( $\text{C}_{12}\text{A}_7$ ),  $\text{CaO} \cdot \text{Al}_2\text{O}_3$  ( $\text{CA}$ ) and  $\text{CaO} \cdot 2\text{Al}_2\text{O}_3$  ( $\text{CA}_2$ ). As regard as the calcium aluminates only  $\text{C}_{12}\text{A}_7$  can be totally soluble in soda solution,  $\text{C}_3\text{A}$  and  $\text{CA}$  dissolve with a slow speed, and the other calcium aluminates such as  $\text{CA}_2$  are completely insoluble. Equations that  $\text{Al}_2\text{O}_3$  reacted with  $\text{CaO}$  to form  $\text{C}_3\text{A}$ ,  $\text{C}_{12}\text{A}_7$ ,  $\text{CA}$  and  $\text{CA}_2$  are presented in table 4.

Figure 5 shows that, the  $\Delta G_T^\theta$  of reactions of  $\text{Al}_2\text{O}_3$  with  $\text{CaO}$  decreases with the rise of temperature; all reactions automatically proceed to generate the corresponding calcium aluminates at normal roasting temperature (1473~1673K, same as follows); At the same



roasting temperature, the thermodynamic order that one mole Al<sub>2</sub>O<sub>3</sub> reacts with CaO to generate calcium aluminates such as C<sub>12</sub>A<sub>7</sub>, C<sub>3</sub>A, CA, CA<sub>2</sub>.

Reactions	A, J/mol	B, J/K.mol	Temperature, K
3CaO+ Al <sub>2</sub> O <sub>3</sub> =3CaO·Al <sub>2</sub> O <sub>3</sub>	-9.9	-28.4	298~1808
$\frac{12}{7}$ CaO+Al <sub>2</sub> O <sub>3</sub> =( $\frac{1}{7}$ )12CaO·7Al <sub>2</sub> O <sub>3</sub>	318.3	-44.5	298~1800
CaO+ Al <sub>2</sub> O <sub>3</sub> =CaO·Al <sub>2</sub> O <sub>3</sub>	-15871.5	-18.1	298~1878
$\frac{1}{2}$ CaO+Al <sub>2</sub> O <sub>3</sub> =( $\frac{1}{2}$ )CaO·2Al <sub>2</sub> O <sub>3</sub>	-6667.2	-13.8	298~2023

Table 4. The  $\Delta G_T^\theta$  of Al<sub>2</sub>O<sub>3</sub>-CaO system ( $\Delta G_T^\theta = A + BT$ , J/mol)

The relationships between  $\Delta G_T^\theta$  and temperature (T) are shown in figure 5.

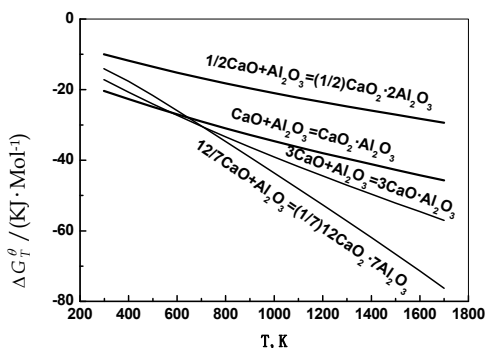


Fig. 5. Relationships between  $\Delta G_T^\theta$  and temperature in Al<sub>2</sub>O<sub>3</sub>-CaO system

Reactions	A, J/mol	B, J/K.mol	Temperature, K
$(\frac{4}{3})3CaO \cdot Al_2O_3 + Al_2O_3 = (\frac{1}{3})12CaO \cdot 7Al_2O_3$	13939.7	-65.8	298~1800
$(\frac{1}{2})3CaO \cdot Al_2O_3 + Al_2O_3 = (\frac{3}{2})CaO \cdot Al_2O_3$	-18843.8	-13.0	298~1878
$(\frac{1}{5})3CaO \cdot Al_2O_3 + Al_2O_3 = (\frac{3}{5})CaO \cdot 2Al_2O_3$	-6011.2	-10.9	298~2023
$(\frac{1}{5})12CaO \cdot 7Al_2O_3 + Al_2O_3 = (\frac{12}{5})CaO \cdot Al_2O_3$	-38544.8	18.8	298~1878
$(\frac{1}{17})12CaO \cdot 7Al_2O_3 + Al_2O_3 = (\frac{12}{17})CaO \cdot 2Al_2O_3$	-9541.1	-1.2	298~2023
CaO · Al <sub>2</sub> O <sub>3</sub> + Al <sub>2</sub> O <sub>3</sub> =CaO · 2Al <sub>2</sub> O <sub>3</sub>	2543.8	-9.5	298~2023

Table 5. The  $\Delta G_T^\theta$  of Al<sub>2</sub>O<sub>3</sub>-calcium aluminates system ( $\Delta G_T^\theta = A + BT$ , J/mol)

When CaO is insufficient, redundant  $\text{Al}_2\text{O}_3$  may promote the newly generated high calcium-to-aluminum ratio (CaO to  $\text{Al}_2\text{O}_3$  mole ratio) calcium aluminates to transform into lower calcium-to-aluminum ratio calcium aluminates. The reactions of the equations are presented in table 5:

The relationships between  $\Delta G_T^\theta$  of reactions of  $\text{Al}_2\text{O}_3$ -calcium aluminates system and temperature (T) are shown in figure 6.

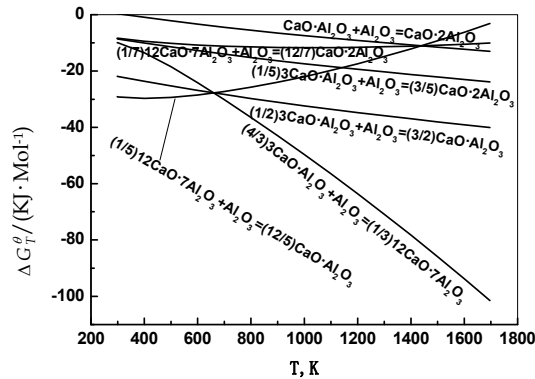


Fig. 6. Relationships between  $\Delta G_T^\theta$  of reactions  $\text{Al}_2\text{O}_3$ -calcium aluminates system and temperature

Figure 6 shows that, Gibbs free energy of the reaction of  $\text{Al}_2\text{O}_3$ -calcium aluminates system are negative at 400~1700K, and all the reactions automatically proceed to generate the corresponding low calcium-to-aluminum ratio calcium aluminates; Except for the reaction of  $\text{Al}_2\text{O}_3$ - $\text{C}_{12}\text{A}_7$ , the  $\Delta G_T^\theta$  of the rest reactions decreases with the rise of temperature and becomes more negative. Comparing figure 4 with figure 5, it can be found that  $\text{Al}_2\text{O}_3$  reacts with CaO easily to generate  $\text{C}_{12}\text{A}_7$ .

## 2.6 $\text{SiO}_2$ - CaO system

$\text{SiO}_2$  can react with CaO to form  $\text{CaO}\cdot\text{SiO}_2$  (CS),  $3\text{CaO}\cdot 2\text{SiO}_2$  ( $\text{C}_3\text{S}_2$ ),  $2\text{CaO}\cdot\text{SiO}_2$  ( $\text{C}_2\text{S}$ ) and  $3\text{CaO}\cdot\text{SiO}_2$  ( $\text{C}_3\text{S}$ ) in roasting process. The reactions are shown in table 6, and the relationships between  $\Delta G^\theta$  of the reactions of  $\text{SiO}_2$  with CaO and temperature are shown in figure 7.

Reactions	A, J/mol	B, J/K.mol	Temperature, K
$\text{CaO} + \text{SiO}_2 = \text{CaO}\cdot\text{SiO}_2$ (pseud-wollastonite)	-83453.0	-3.4	298~1817
$\text{CaO} + \text{SiO}_2 = \text{CaO}\cdot\text{SiO}_2$ (wollastonite)	-89822.9	-0.3	298~1817
$\frac{3}{2}\text{CaO} + \text{SiO}_2 = (\frac{1}{2})3\text{CaO}\cdot 2\text{SiO}_2$	-108146.6	-3.1	298~1700
$3\text{CaO} + \text{SiO}_2 = 3\text{CaO}\cdot\text{SiO}_2$	-111011.9	-11.3	298~1800
$2\text{CaO} + \text{SiO}_2 = 2\text{CaO}\cdot\text{SiO}_2$ ( $\beta$ )	-125875.1	-6.7	298~2403
$2\text{CaO} + \text{SiO}_2 = 2\text{CaO}\cdot\text{SiO}_2$ ( $\gamma$ )	-137890.1	3.7	298~1100

Table 6. The  $\Delta G_T^\theta$  of  $\text{SiO}_2$ -CaO system ( $\Delta G_T^\theta = A + BT$ , J/mol)

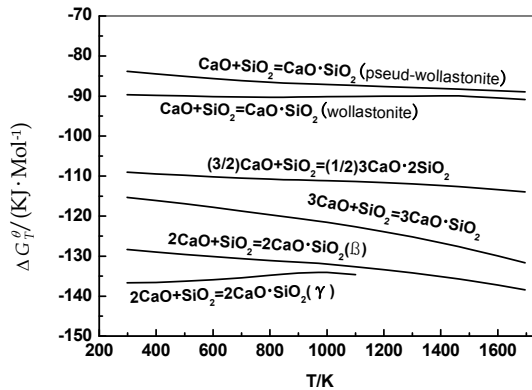


Fig. 7. Relationships between  $\Delta G_T^\theta$  and temperature

Figure 7 shows that,  $\text{SiO}_2$  reacts with  $\text{CaO}$  to form  $\gamma\text{-C}_2\text{S}$  when temperature below 1100K, but  $\beta\text{-C}_2\text{S}$  comes into being when the temperature above 1100K. At normal roasting temperature, the thermodynamic order of forming calcium silicate is  $\text{C}_2\text{S}$ ,  $\text{C}_3\text{S}$ ,  $\text{C}_3\text{S}_2$ ,  $\text{CS}$ .

Figure 5 ~ figure 7 show that,  $\text{CaO}$  reacts with  $\text{SiO}_2$  and  $\text{Al}_2\text{O}_3$  firstly to form  $\text{C}_2\text{S}$ , and then  $\text{C}_{12}\text{A}_7$ . Therefore, it is less likely to form aluminium silicates in roasting process.

### 2.7 $\text{SiO}_2$ - calcium aluminates system

In the  $\text{CaO-Al}_2\text{O}_3$  system, if there exists some  $\text{SiO}_2$ , the newly formed calcium aluminates are likely to react with  $\text{SiO}_2$  to transform to calcium silicates and  $\text{Al}_2\text{O}_3$  because  $\text{SiO}_2$  is more acidity than that of  $\text{Al}_2\text{O}_3$ . The reaction equations are presented in table 7, the relationships between  $\Delta G_T^\theta$  and temperature are shown in figure 8.

Figure 8 shows that, the  $\Delta G_T^\theta$  of all the reactions increases with the temperature increases; the reaction  $(3\text{CA}_2 + \text{SiO}_2 = \text{C}_3\text{S} + 6\text{Al}_2\text{O}_3)$  can not happen when the roasting temperature is above 900K, i.e., the lowest calcium-to-aluminum ratio calcium aluminates cannot transform to the highest calcium-to-silicon ratio (CaO to  $\text{SiO}_2$  molecular ratio) calcium silicate; when the temperature is above 1500K, the  $\Delta G_T^\theta$  of reaction  $(3\text{CA} + \text{SiO}_2 = \text{C}_3\text{S} + 3\text{Al}_2\text{O}_3)$  is also more than zero; but the other calcium aluminates all can react with  $\text{SiO}_2$  to generate calcium silicates at 800~1700K. The thermodynamic sequence of calcium aluminates reaction with  $\text{SiO}_2$  is firstly  $\text{C}_3\text{A}$ , and then  $\text{C}_{12}\text{A}_7$ ,  $\text{CA}$ ,  $\text{CA}_2$ .

Reactions	A, J/mol	B, J/K.mol	Temperature, K
$(3)\text{CaO} \cdot 2\text{Al}_2\text{O}_3 + \text{SiO}_2 = 3\text{CaO} \cdot \text{SiO}_2 + 6\text{Al}_2\text{O}_3$	-69807.8	70.8	298~1800
$(3)\text{CaO} \cdot \text{Al}_2\text{O}_3 + \text{SiO}_2 = 3\text{CaO} \cdot \text{SiO}_2 + 3\text{Al}_2\text{O}_3$	-62678.8	42.6	298~1800
$(\frac{1}{4})12\text{CaO} \cdot 7\text{Al}_2\text{O}_3 + \text{SiO}_2 = 3\text{CaO} \cdot \text{SiO}_2 + \frac{7}{4}\text{Al}_2\text{O}_3$	-111820.6	66.7	298~1800
$(2)\text{CaO} \cdot 2\text{Al}_2\text{O}_3 + \text{SiO}_2 = 2\text{CaO} \cdot \text{SiO}_2 + 4\text{Al}_2\text{O}_3$	-98418.8	48.1	298~1710
$(\frac{3}{2})\text{CaO} \cdot 2\text{Al}_2\text{O}_3 + \text{SiO}_2 = (\frac{1}{2})3\text{CaO} \cdot 2\text{SiO}_2 + 3\text{Al}_2\text{O}_3$	-87585.9	38.0	298~1700
$\text{CaO} \cdot 2\text{Al}_2\text{O}_3 + \text{SiO}_2 = \text{CaO} \cdot \text{SiO}_2 + 2\text{Al}_2\text{O}_3$	-76146.6	27.1	298~1817
$\text{CaO} \cdot \text{Al}_2\text{O}_3 + \text{SiO}_2 = \text{CaO} \cdot \text{SiO}_2 + \text{Al}_2\text{O}_3$	-73770.2	17.7	298~1817
$(\frac{3}{2})\text{CaO} \cdot \text{Al}_2\text{O}_3 + \text{SiO}_2 = (\frac{1}{2})3\text{CaO} \cdot 2\text{SiO}_2 + \frac{3}{2}\text{Al}_2\text{O}_3$	-84021.4	23.8	298~1700
$(2)\text{CaO} \cdot \text{Al}_2\text{O}_3 + \text{SiO}_2 = 2\text{CaO} \cdot \text{SiO}_2 + 2\text{Al}_2\text{O}_3$	-93666.1	29.2	298~1710
$(\frac{1}{12})12\text{CaO} \cdot 7\text{Al}_2\text{O}_3 + \text{SiO}_2 = \text{CaO} \cdot \text{SiO}_2 + \frac{7}{12}\text{Al}_2\text{O}_3$	-90150.8	25.7	298~1800
$(\frac{1}{8})12\text{CaO} \cdot 7\text{Al}_2\text{O}_3 + \text{SiO}_2 = (\frac{1}{2})3\text{CaO} \cdot 2\text{SiO}_2 + \frac{7}{8}\text{Al}_2\text{O}_3$	-108592.3	35.9	298~1700
$(\frac{1}{6})12\text{CaO} \cdot 7\text{Al}_2\text{O}_3 + \text{SiO}_2 = 2\text{CaO} \cdot \text{SiO}_2 + \frac{7}{6}\text{Al}_2\text{O}_3$	-126427.4	45.3	298~1710
$(\frac{1}{3})3\text{CaO} \cdot \text{Al}_2\text{O}_3 + \text{SiO}_2 = \text{CaO} \cdot \text{SiO}_2 + \frac{1}{3}\text{Al}_2\text{O}_3$	-86654.2	9.4	298~1808
$3\text{CaO} \cdot \text{Al}_2\text{O}_3 + \text{SiO}_2 = 3\text{CaO} \cdot \text{SiO}_2 + \text{Al}_2\text{O}_3$	-100774.6	16.9	298~1808
$(\frac{1}{2})3\text{CaO} \cdot \text{Al}_2\text{O}_3 + \text{SiO}_2 = (\frac{1}{2})3\text{CaO} \cdot 2\text{SiO}_2 + \frac{1}{2}\text{Al}_2\text{O}_3$	-103069.3	11.0	298~1700
$(\frac{2}{3})3\text{CaO} \cdot \text{Al}_2\text{O}_3 + \text{SiO}_2 = 2\text{CaO} \cdot \text{SiO}_2 + \frac{2}{3}\text{Al}_2\text{O}_3$	-119063.3	12.1	298~1710

Table 7. The  $\Delta G_T^\theta$  of the reactions  $\text{SiO}_2$  with calcium aluminates ( $\Delta G_T^\theta = A + BT$ , J/mol)

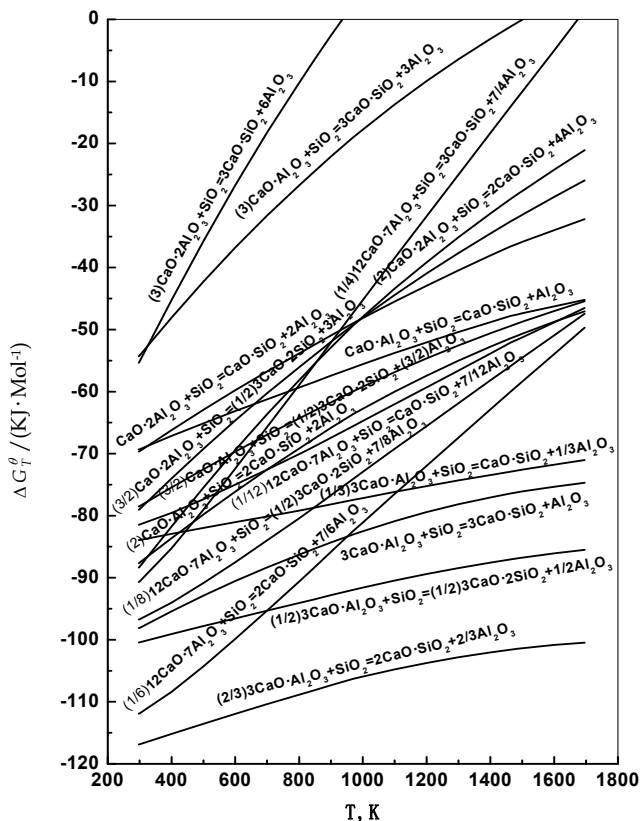


Fig. 8. Relationships between  $\Delta G_T^0$  and temperature in  $\text{SiO}_2$ -calcium aluminates system

### 2.8 CaO- $\text{Fe}_2\text{O}_3$ system

$\text{Fe}_2\text{O}_3$  can react with  $\text{CaO}$  to form  $\text{CaO} \cdot \text{Fe}_2\text{O}_3(\text{CF})$  and  $2\text{CaO} \cdot \text{Fe}_2\text{O}_3(\text{C}_2\text{F})$ . When  $\text{Fe}_2\text{O}_3$  is used up, the newly formed  $\text{C}_2\text{F}$  can react with  $\text{Fe}_2\text{O}_3$  to form  $\text{CF}$ . The reaction equations are shown in table 8, and the relationships between  $\Delta G^0$  and temperature are shown in figure 9.

Figure 9 shows that,  $\text{Fe}_2\text{O}_3$  reacts with  $\text{CaO}$  much easily to form  $\text{C}_2\text{F}$ ;  $\text{CF}$  is not from the reaction of  $\text{C}_2\text{F}$  and  $\text{Fe}_2\text{O}_3$ , but from the directly reaction of  $\text{Fe}_2\text{O}_3$  with  $\text{CaO}$ . When  $\text{Fe}_2\text{O}_3$  is excess,  $\text{C}_2\text{F}$  can react with  $\text{Fe}_2\text{O}_3$  to form  $\text{CF}$ .

Reactions	A, J/mol	B, J/K.mol	Temperature, K
$\text{CaO} + \text{Fe}_2\text{O}_3 = \text{CaO} \cdot \text{Fe}_2\text{O}_3$	-19179.9	-11.1	298~1489
$2\text{CaO} + \text{Fe}_2\text{O}_3 = 2\text{CaO} \cdot \text{Fe}_2\text{O}_3$	-40866.7	-9.3	298~1723
$2\text{CaO} \cdot \text{Fe}_2\text{O}_3 + \text{Fe}_2\text{O}_3 = (2)\text{CaO} \cdot \text{Fe}_2\text{O}_3$	2340.8	-12.6	298~1489

Table 8. The  $\Delta G_T^0$  of  $\text{Fe}_2\text{O}_3$ - $\text{CaO}$  system ( $\Delta G_T^0 = A + BT$ , J/mol)

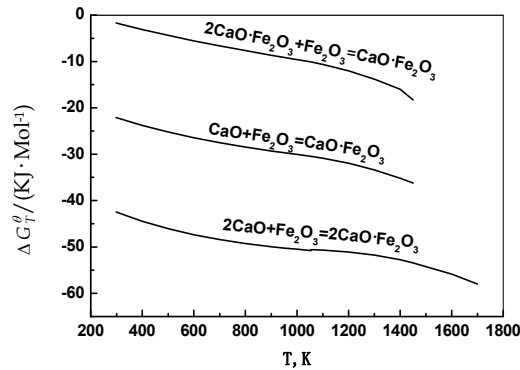


Fig. 9. Relationships between  $\Delta G_T^0$  and temperature in  $\text{Fe}_2\text{O}_3$ -CaO system

### 2.9 $\text{Al}_2\text{O}_3$ - calcium ferrites system

Figure 1 shows that, the  $\Delta G_T^0$  of the reaction of  $\text{Al}_2\text{O}_3$  with  $\text{CaCO}_3$  is more negative than that of  $\text{Fe}_2\text{O}_3$  with  $\text{CaCO}_3$ , therefore, the reaction of  $\text{Fe}_2\text{O}_3$  with  $\text{CaCO}_3$  occurs after the reaction of  $\text{Al}_2\text{O}_3$  with  $\text{CaCO}_3$  under the conditions of excess  $\text{CaCO}_3$ . The new generated calcium ferrites are likely to transform into calcium aluminates when  $\text{CaCO}_3$  is insufficient, the reactions are as followed:

Reactions	A, J/mol	B, J/K.mol	Temperature, K
$(3)\text{CaO}\cdot\text{Fe}_2\text{O}_3 + \text{Al}_2\text{O}_3 = 3\text{CaO}\cdot\text{Al}_2\text{O}_3 + 3\text{Fe}_2\text{O}_3$	47922.7	4.5	298~1489
$(\frac{3}{2})2\text{CaO}\cdot\text{Fe}_2\text{O}_3 + \text{Al}_2\text{O}_3 = 3\text{CaO}\cdot\text{Al}_2\text{O}_3 + \frac{3}{2}\text{Fe}_2\text{O}_3$	49.6	$-1.2\times 10^{-2}$	298~1723
$(\frac{12}{7})\text{CaO}\cdot\text{Fe}_2\text{O}_3 + \text{Al}_2\text{O}_3 = (\frac{1}{7})12\text{CaO}\cdot 7\text{Al}_2\text{O}_3 + \frac{12}{7}\text{Fe}_2\text{O}_3$	32685.1	-24.5	298~1489
$(\frac{6}{7})2\text{CaO}\cdot\text{Fe}_2\text{O}_3 + \text{Al}_2\text{O}_3 = (\frac{1}{7})12\text{CaO}\cdot 7\text{Al}_2\text{O}_3 + \frac{6}{7}\text{Fe}_2\text{O}_3$	34514.4	-35.0	298~1723
$\text{CaO}\cdot\text{Fe}_2\text{O}_3 + \text{Al}_2\text{O}_3 = \text{CaO}\cdot\text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3$	3626.6	-7.5	298~1489
$(\frac{1}{2})\text{CaO}\cdot\text{Fe}_2\text{O}_3 + \text{Al}_2\text{O}_3 = (\frac{1}{2})\text{CaO}\cdot 2\text{Al}_2\text{O}_3 + \frac{1}{2}\text{Fe}_2\text{O}_3$	3215.1	-8.8	298~1489
$(\frac{1}{4})2\text{CaO}\cdot\text{Fe}_2\text{O}_3 + \text{Al}_2\text{O}_3 = (\frac{1}{2})\text{CaO}\cdot 2\text{Al}_2\text{O}_3 + \frac{1}{4}\text{Fe}_2\text{O}_3$	3168.6	-11.0	298~1723
$(\frac{1}{2})2\text{CaO}\cdot\text{Fe}_2\text{O}_3 + \text{Al}_2\text{O}_3 = \text{CaO}\cdot\text{Al}_2\text{O}_3 + \frac{1}{2}\text{Fe}_2\text{O}_3$	4009.5	-12.8	298~1723

Table 9. The  $\Delta G_T^0$  of the reaction  $\text{Al}_2\text{O}_3$  with calcium ferrites ( $\Delta G_T^0 = A + BT$ , J/mol)

The relationships between  $\Delta G_T^0$  and temperature (T) are shown in figure 10. Figure 10 shows that,  $\text{Al}_2\text{O}_3$  cannot replace the  $\text{Fe}_2\text{O}_3$  in calcium ferrites to generate  $\text{C}_3\text{A}$ , and also cannot replace the  $\text{Fe}_2\text{O}_3$  in  $\text{CaO}\cdot\text{Fe}_2\text{O}_3(\text{CF})$  to generate  $\text{C}_{12}\text{A}_7$ , but it can replace the  $\text{Fe}_2\text{O}_3$  in  $2\text{CaO}\cdot\text{Fe}_2\text{O}_3(\text{C}_2\text{F})$  to generate  $\text{C}_{12}\text{A}_7$  when the temperature is above 1000K, the higher temperature is, the more negative Gibbs free energy is;  $\text{Al}_2\text{O}_3$  can react with CF and  $\text{C}_2\text{F}$  to

form CA or CA<sub>2</sub>, the higher temperature, more negative  $\Delta G_T^\theta$ . Because Fe<sub>2</sub>O<sub>3</sub> reacts with CaO more easily to generate C<sub>2</sub>F (Fig.9), therefore, C<sub>12</sub>A<sub>7</sub> is the reaction product at normal roasting temperature(1073~1673K) under the conditions that CaO is sufficient in batching and the ternary compounds are not considered.

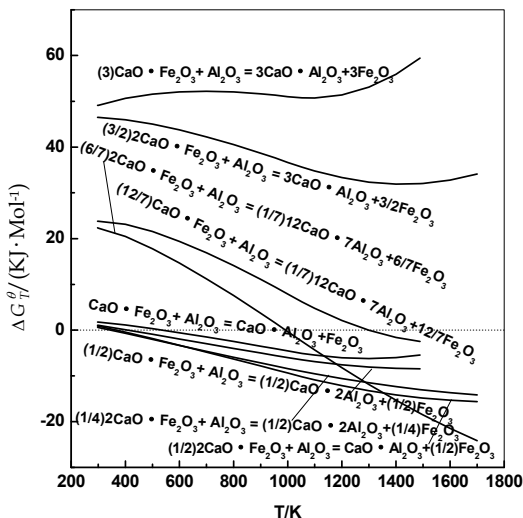


Fig. 10. Relationship between  $\Delta G_T^\theta$  and temperature in Al<sub>2</sub>O<sub>3</sub>- calcium ferrites system

### 3. Ternary compounds in Al<sub>2</sub>O<sub>3</sub>-CaO-SiO<sub>2</sub>-Fe<sub>2</sub>O<sub>3</sub> system

The ternary compounds formed by CaO, Al<sub>2</sub>O<sub>3</sub> and SiO<sub>2</sub> in roasting process are mainly 2CaO·Al<sub>2</sub>O<sub>3</sub>·SiO<sub>2</sub>(C<sub>2</sub>AS), CaO·Al<sub>2</sub>O<sub>3</sub>·2SiO<sub>2</sub>(C<sub>1</sub>AS<sub>2</sub>), CaO·Al<sub>2</sub>O<sub>3</sub>·SiO<sub>2</sub>(C<sub>1</sub>AS) and 3CaO·Al<sub>2</sub>O<sub>3</sub>·3SiO<sub>2</sub>(C<sub>3</sub>AS<sub>3</sub>). In addition, ternary compound 4CaO·Al<sub>2</sub>O<sub>3</sub>·Fe<sub>2</sub>O<sub>3</sub>(C<sub>4</sub>AF) is formed from CaO, Al<sub>2</sub>O<sub>3</sub> and Fe<sub>2</sub>O<sub>3</sub>. The equations are shown in table 10:

Reactions	A, J/mol	B, J/K.mol	Temperature, K
CaO·SiO <sub>2</sub> +CaO·Al <sub>2</sub> O <sub>3</sub> =2CaO·Al <sub>2</sub> O <sub>3</sub> ·SiO <sub>2</sub>	-30809.41	0.60	298~1600
$\frac{1}{2}$ Al <sub>2</sub> O <sub>3</sub> + $\frac{1}{2}$ CaO + SiO <sub>2</sub> = ( $\frac{1}{2}$ )CaO·Al <sub>2</sub> O <sub>3</sub> ·2SiO <sub>2</sub>	-47997.55	-7.34	298~1826
Al <sub>2</sub> O <sub>3</sub> + 2CaO + SiO <sub>2</sub> =2CaO·Al <sub>2</sub> O <sub>3</sub> ·SiO <sub>2</sub>	-50305.83	-9.33	298~1600
Al <sub>2</sub> O <sub>3</sub> + CaO + SiO <sub>2</sub> =CaO·Al <sub>2</sub> O <sub>3</sub> ·SiO <sub>2</sub>	-72975.54	-9.49	298~1700
$\frac{1}{3}$ Al <sub>2</sub> O <sub>3</sub> + CaO + SiO <sub>2</sub> = ( $\frac{1}{3}$ )3CaO·Al <sub>2</sub> O <sub>3</sub> ·3SiO <sub>2</sub>	-112354.51	20.86	298~1700
4CaO + Al <sub>2</sub> O <sub>3</sub> + Fe <sub>2</sub> O <sub>3</sub> =4CaO·Al <sub>2</sub> O <sub>3</sub> ·Fe <sub>2</sub> O <sub>3</sub>	-66826.92	-62.5	298~2000
Al <sub>2</sub> O <sub>3</sub> + 2CaO + SiO <sub>2</sub> =2CaO·Al <sub>2</sub> O <sub>3</sub> ·SiO <sub>2</sub> (cacoclasite)	-136733.59	-17.59	298~1863

Table 10. The  $\Delta G_T^\theta$  of forming ternary compounds ( $\Delta G_T^\theta = A + BT$ , J/mol)

The relationships between  $\Delta G_T^\theta$  and temperature (T) are shown in figure 11. Figure 11 shows that, except for  $C_3AS_3$ (Hessonite), all the  $\Delta G_T^\theta$  of the reactions get more negative with the temperature increasing; the thermodynamic order of generating ternary compounds at sintering temperature of 1473K is:  $C_2AS$ (cacoclasite),  $C_4AF$ ,  $CAS$ ,  $C_3AS_3$ ,  $C_2AS$ ,  $CAS_2$ .

$C_2AS$  may also be formed by the reaction of CA and CS, the curve is presented in figure 11.

Figure 11 shows that, the  $\Delta G_T^\theta$  of reaction ( $Al_2O_3+CaO+SiO_2$ ) is lower than that of reaction of CA and CS to generate  $C_2AS$ . So  $C_2AS$  does not form from the binary compounds CA and CS, but from the direct combination among  $Al_2O_3$ ,  $CaO$ ,  $SiO_2$ . Qiusheng Zhou thinks that,  $C_4AF$  is not formed by mutual reaction of calcium ferrites and sodium aluminates, but from the direct reaction of  $CaO$ ,  $Al_2O_3$  and  $Fe_2O_3$ . Thermodynamic analysis of figure 1~figure11 shows that, reactions of  $Al_2O_3$ ,  $Fe_2O_3$ ,  $SiO_2$  and  $CaO$  are much easier to form  $C_2AS$  and  $C_4AF$ , as shown in figure 12.

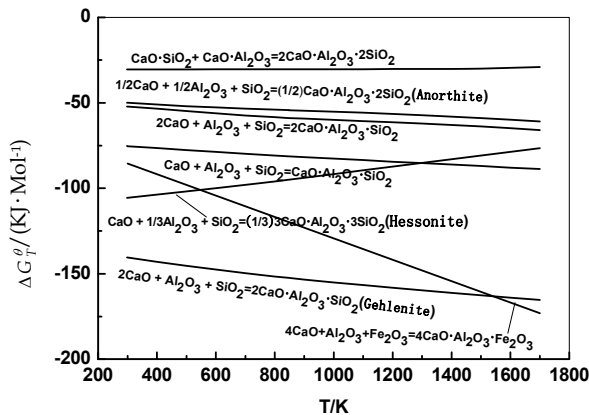


Fig. 11. Relationships between  $\Delta G_T^\theta$  of ternary compounds and temperature

Figure 12 shows that, in thermodynamics,  $C_2AS$  and  $C_4AF$  are firstly formed when  $Al_2O_3$ ,  $Fe_2O_3$ ,  $SiO_2$  and  $CaO$  coexist, and then calcium silicates, calcium aluminates and calcium ferrites are generated.

#### 4. Summary

1) When  $Al_2O_3$  and  $Fe_2O_3$  simultaneously react with  $CaO$ , calcium silicates are firstly formed, and then calcium ferrites. In thermodynamics, when one mole  $Al_2O_3$  reacts with  $CaO$ , the sequence of generating calcium aluminates are  $12CaO \cdot 7Al_2O_3$ ,  $3CaO \cdot Al_2O_3$ ,  $CaO \cdot Al_2O_3$ ,  $CaO \cdot 2Al_2O_3$ . When  $CaO$  is insufficient, redundant  $Al_2O_3$  may promote the newly generated high calcium-to-aluminum ratio calcium aluminates to transform to lower calcium-to-aluminum ratio calcium aluminates.  $Fe_2O_3$  reacts with  $CaO$  easily to form  $2CaO \cdot Fe_2O_3$ , and  $CaO \cdot Fe_2O_3$  is not from the reaction of  $2CaO \cdot Fe_2O_3$  and  $Fe_2O_3$  but from the directly combination of  $Fe_2O_3$  with  $CaO$ .  $Al_2O_3$  cannot replace the  $Fe_2O_3$  in calcium ferrites to generate  $3CaO \cdot Al_2O_3$ , and also cannot replace the  $Fe_2O_3$  in  $CaO \cdot Fe_2O_3$  to generate  $12CaO \cdot 7Al_2O_3$ , but can replace the  $Fe_2O_3$  in  $2CaO \cdot Fe_2O_3$  to generate  $12CaO \cdot 7Al_2O_3$  when the temperature is above 1000K;  $Al_2O_3$  can react with calcium ferrites to form  $CaO \cdot Al_2O_3$  or  $CaO \cdot 2Al_2O_3$ .



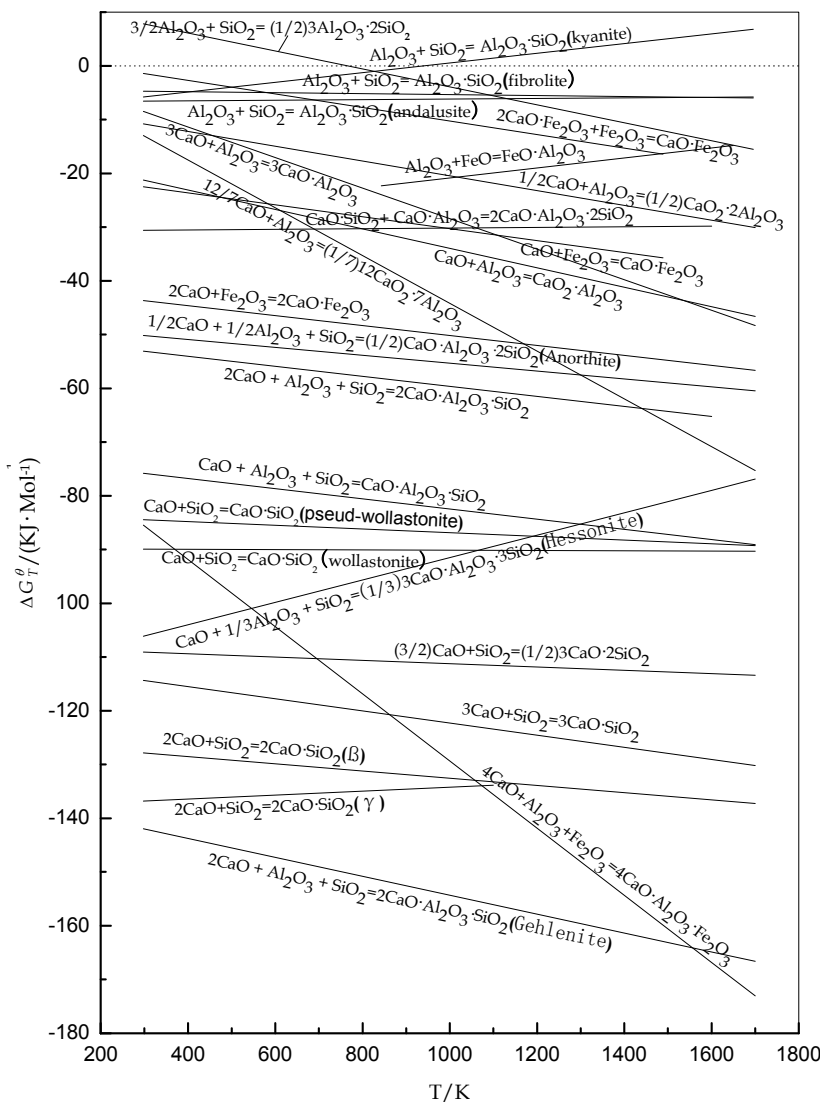


Fig. 12. Relationships between  $\Delta G_T^0$  and temperature in  $\text{Al}_2\text{O}_3$ - $\text{CaO}$ - $\text{SiO}_2$ - $\text{Fe}_2\text{O}_3$  system

2) One mole  $\text{SiO}_2$  reacts with  $\text{Al}_2\text{O}_3$  much easily to generate  $3\text{Al}_2\text{O}_3 \cdot 2\text{SiO}_2$ ,  $\text{Fe}_2\text{O}_3$  can not react with  $\text{SiO}_2$  in the roasting process in the air.  $\text{Al}_2\text{O}_3$  can not directly react with  $\text{Fe}_2\text{O}_3$ , but can react with wustite ( $\text{FeO}$ ) to form  $\text{FeO} \cdot \text{Al}_2\text{O}_3$ .

3) In thermodynamics, the sequence of one mole  $\text{SiO}_2$  reacts with  $\text{CaO}$  to form calcium silicates is  $2\text{CaO} \cdot \text{SiO}_2$ ,  $3\text{CaO} \cdot \text{SiO}_2$ ,  $3\text{CaO} \cdot 2\text{SiO}_2$  and  $\text{CaO} \cdot \text{SiO}_2$ . Calcium aluminates can react with  $\text{SiO}_2$  to transform to calcium silicates and  $\text{Al}_2\text{O}_3$ .  $\text{CaO} \cdot 2\text{Al}_2\text{O}_3$  can not transform to  $3\text{CaO} \cdot \text{SiO}_2$  when the roasting temperature is above 900K; when the temperature is above

1500K,  $3\text{CaO} \cdot \text{Al}_2\text{O}_3$  can not transform to  $3\text{CaO} \cdot \text{SiO}_2$ ; but the other calcium aluminates all can all react with  $\text{SiO}_2$  to generate calcium silicates at 800~1700K.

4) Reactions among  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{SiO}_2$  and  $\text{CaO}$  easily form  $2\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{SiO}_2$  and  $4\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{Fe}_2\text{O}_3$ .  $2\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{SiO}_2$  does not form from the reaction of  $\text{CaO} \cdot \text{Al}_2\text{O}_3$  and  $\text{CaO} \cdot \text{SiO}_2$ , but from the direct reaction among  $\text{Al}_2\text{O}_3$ ,  $\text{CaO}$ ,  $\text{SiO}_2$ . And  $4\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{Fe}_2\text{O}_3$  is also not formed via mutual reaction of calcium ferrites and sodium aluminates, but from the direct reaction of  $\text{CaO}$ ,  $\text{Al}_2\text{O}_3$  and  $\text{Fe}_2\text{O}_3$ . In thermodynamics, when  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{SiO}_2$  and  $\text{CaO}$  coexist,  $2\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{SiO}_2$  and  $4\text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{Fe}_2\text{O}_3$  are firstly formed, and then calcium silicates, calcium aluminates and calcium ferrites.

## 5. Symbols used

Thermodynamic temperature: T, K

Thermal unit: J

Amount of substance: mole

Standard Gibbs free energy:  $\Delta G_r^\theta$ , J

## 6. References

- Li, B.; Xu, Y. & Choi, J. (1996). Applying Machine Learning Techniques, *Proceedings of ASME 2010 4th International Conference on Energy Sustainability*, pp.14-17, ISBN 842-6508-23-3, Phoenix, Arizona, USA, May 17-22, 2010
- Rayi H. S. ; Kundu N.(1986). Thermal analysis studies on the initial stages of iron oxide reduction, *Thermochimi, Acta.* 101:107~118,1986
- Coats A.W. ; Redfern J.P.(1964). Kinetic parameters from thermogravimetric data, *Nature*, 201:68,1964
- LIU Gui-hua, LI Xiao-bin, PENG Zhi-hong, ZHOU Qiu-sheng(2003). Behavior of calcium silicate in leaching process. *Trans Nonferrous Met Soc China*, January 213–216,2003
- Paul S. ; Mukherjee S.(1992). Nonisothermal and isothermal reduction kinetics of iron ore agglomerates, *Ironmaking and steelmaking*, March 190~193, 1992
- ZHU Zhongping, JIANG Tao, LI Guanghui, HUANG Zhucheng(2009). Thermodynamics of reaction of alumina during sintering process of high-iron gibbsite-type bauxite, *The Chinese Journal of Nonferrous Metals*, Dec 2243~2250, 2009
- ZHOU Qiusheng, QI Tiangui, PENG Zhihong, LIU Guihua, LI Xiaobin(2007). Thermodynamics of reaction behavior of ferric oxide during sinter-preparing process, *The Chinese Journal of Nonferrous Metals*, Jun 974~978, 2007
- Barin I., Knacke O.(1997). *Thermochemical properties of inorganic substances*, Berlin:Supplement, 1997
- Barin I., Knacke O.(1973). *Thermochemical properties of inorganic substances*, Berlin: Springer, 1973

# Thermodynamic Perturbation Theory of Simple Liquids

Jean-Louis Bretonnet

*Laboratoire de Physique des Milieux Denses, Université Paul Verlaine de Metz  
France*

## 1. Introduction

This chapter is an introduction to the thermodynamics of systems, based on the correlation function formalism, which has been established to determine the thermodynamic properties of simple liquids. The article begins with a preamble describing few general aspects of the liquid state, among others the connection between the phase diagram and the pair potential  $u(r)$ , on one hand, and between the structure and the pair correlation function  $g(r)$ , on the other hand. The pair correlation function is of major importance in the theory of liquids at equilibrium, because it is required for performing the calculation of the thermodynamic properties of systems modeled by a given pair potential. Then, the article is devoted to the expressions useful for calculating the thermodynamic properties of liquids, in relation with the most relevant features of the potential, and provides a presentation of the perturbation theory developed in the four last decades. The thermodynamic perturbation theory is founded on a judicious separation of the pair potential into two parts. Specifically, one of the greatest successes of the microscopic theory has been the recognition of the quite distinct roles played by the repulsive and attractive parts of the pair potential in predicting many properties of liquids. Much attention has been paid to the hard-sphere potential, which has proved very efficient as natural reference system because it describes fairly well the local order in liquids. As an example, the Yukawa attractive potential is also mentioned.

## 2. An elementary survey

### 2.1 The liquid state

The ability of the liquids to form a free surface differs from that of the gases, which occupy the entire volume available and have diffusion coefficients ( $\sim 0,5 \text{ cm}^2\text{s}^{-1}$ ) of several orders of magnitude higher than those of liquids ( $\sim 10^{-5} \text{ cm}^2\text{s}^{-1}$ ) or solids ( $\sim 10^{-9} \text{ cm}^2\text{s}^{-1}$ ). Moreover, if the dynamic viscosity of liquids (between  $10^{-5} \text{ Pa}\cdot\text{s}$  and  $1 \text{ Pa}\cdot\text{s}$ ) is so lower compared to that of solids, it is explained in terms of competition between *configurational* and *kinetic* processes. Indeed, in a solid, the displacements of atoms occur only after the breaking of the bonds that keep them in a stable configuration. At the opposite, in a gas, molecular transport is a purely kinetic process perfectly described in terms of exchanges of energy and momentum. In a liquid, the continuous rearrangement of particles and the molecular transport combine together in appropriate proportion, meaning that the liquid is an intermediate state between the gaseous and solid states.

The characterization of the three states of matter can be done in an advantageous manner by comparing the kinetic energy and potential energy as it is done in figure (1). The nature and intensity of forces acting between particles are such that the particles tend to attract each other at great distances, while they repel at the short distances. The particles are in equilibrium when the attraction and repulsion forces balance each other. In gases, the kinetic energy of particles, whose the distribution is given by the Maxwell velocity distribution, is located in the region of unbound states. The particles move freely on trajectories suddenly modified by binary collisions; thus the movement of particles in the gases is essentially an *individual movement*. In solids, the energy distribution is confined within the potential well. It follows that the particles are in tight bound states and describe harmonic motions around their equilibrium positions; therefore the movement of particles in the solids is essentially a *collective movement*. When the temperature increases, the energy distribution moves towards high energies and the particles are subjected to anharmonic movements that intensify progressively. In liquids, the energy distribution is almost entirely located in the region of bound states, and the movements of the particles are strongly anharmonic. On approaching the critical point, the energy distribution shifts towards the region of unbound states. This results in important fluctuations in concentrations, accompanied by the destruction and formation of aggregates of particles. Therefore, the movement of particles in liquids is thus the result of a combination of individual and collective movements.

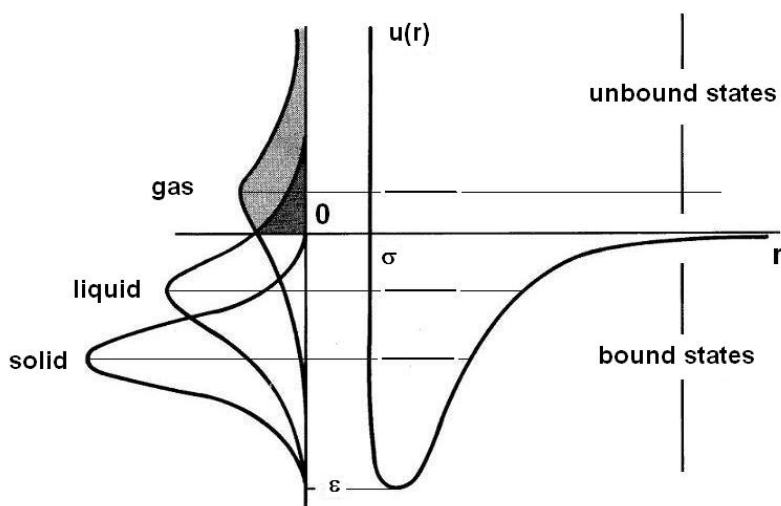


Fig. 1. Comparison of kinetic and potential energies in solids, liquids and gases.

When a crystalline solid melts, the long-range order of the crystal is destroyed, but a residual local order persists on distances greater than several molecular diameters. This local order into liquid state is described in terms of the *pair correlation function*,  $g(r) = \frac{\rho(r)}{\rho_\infty}$ , which is defined as the ratio of the mean molecular density  $\rho(r)$ , at a distance  $r$  from an arbitrary molecule, to the bulk density  $\rho_\infty$ . If  $g(r)$  is equal to unity everywhere, the fluid is completely disordered, like in diluted gases. The deviation of  $g(r)$  from unity is a measure of the local order in the arrangement of near-neighbors. The representative curve of  $g(r)$  for a liquid is formed of maxima and minima rapidly damped around unity, where the first maximum corresponds

to the position of the nearest neighbors around an origin atom. It should be noted that the pair correlation function  $g(r)$  is accessible by a simple Fourier transform of the experimental *structure factor*  $S(q)$  (intensity of scattered radiation).

The pair correlation function is of crucial importance in the theory of liquids at equilibrium, because it depends strongly on the *pair potential*  $u(r)$  between the molecules. In fact, one of the goals of the theory of liquids at equilibrium is to predict the thermodynamic properties using the pair correlation function  $g(r)$  and the pair potential  $u(r)$  acting in the liquids. There are a large number of potential models (hard sphere, square well, Yukawa, Gaussian, Lennard-Jones...) more or less adapted to each type of liquids. These interaction potentials have considerable theoretical interest in statistical physics, because they allow the calculation of the properties of the liquids they are supposed to represent. But many approximations for calculating the pair correlation function  $g(r)$  exist too.

Note that there is a great advantage in comparing the results of the theory with those issued from the numerical simulation with the aim to test the models developed in the theory. Beside, the comparison of the theoretical results to the experimental results allows us to test the potential when the theory itself is validated. Nevertheless, comparison of simulation results with experimental results is the most efficient way to test the potential, because the simulation provides the exact solution without using a theoretical model. It is a matter of fact that simulation is generally identified to a numerical experience. Even if they are time consuming, the simulation computations currently available with thousands of interacting particles gives a role increasingly important to the simulation methods.

In the theory of simple fluids, one of the major achievements has been the recognition of the quite distinct roles played by the repulsive and attractive parts of the pair potential in determining the microscopic properties of simple fluids. In recent years, much attention has been paid in developing analytically solvable models capable to represent the thermodynamic and structural properties of real fluids. The hard-sphere (HS) model - with its diameter  $\sigma$  - is the natural *reference system* for describing the general characteristics of liquids, i.e. the local atomic order due to the excluded volume effects and the *solidification* process of liquids into a solid ordered structure. In contrast, the HS model is not able to predict the *condensation* of a gas into a liquid, which is only made possible by the existence of dispersion forces represented by an attractive long-ranged part in the potential.

Another reference model that has proved very useful to stabilize the local structure in liquids is the hard-core potential with an attractive Yukawa tail (HCY), by varying the hard-sphere diameter  $\sigma$  and screening length  $\lambda$ . It is an advantage of this model for modeling real systems with widely different features (1), like rare gases with a screening length  $\lambda \sim 2$  or colloidal suspensions and protein solutions with a screening length  $\lambda \sim 8$ . An additional reason that does the HCY model appealing is that analytical solutions are available. After the search of the original solution with the mean-spherical approximation (2), valuable simplifications have been progressively brought giving simple analytical expressions for the thermodynamic properties and the pair correlation function. For this purpose, the expression for the free energy has been used under an expanded form in powers of the inverse temperature, as derived by Henderson *et al.* (3).

At this stage, it is perhaps salutary to claim that no attempt will be made, in this article, to discuss neither the respective advantages of the pair potentials nor the ability of various approximations to predict the structure, which are necessary to determine the thermodynamic properties of liquids. In other terms, nothing will be said on the theoretical aspect of correlation functions, except a brief summary of the experimental determination of the pair correlation function. In contrast, it will be useful to state some of the concepts

of statistical thermodynamics providing a link between the microscopic description of liquids and classical thermodynamic functions. Then, it will be given an account of the thermodynamic perturbation theory with the analytical expressions required for calculating the thermodynamic properties. Finally, the HCY model, which is founded on the perturbation theory, will be presented in greater detail for investigating the thermodynamics of liquids. Thus, a review of the thermodynamic perturbation theory will be set up, with a special effort towards the pedagogical aspect. We hope that this paper will help readers to develop their inductive and synthetic capacities, and to enhance their scientific ability in the field of thermodynamic of liquids. It goes without saying that the intention of the present paper is just to initiate the readers to that matter, which is developed in many standard textbooks (4).

## 2.2 Phase stability limits versus pair potential

One success of the numerical simulation was to establish a relationship between the shape of the pair potential and the phase stability limits, thus clarifying the circumstances of the liquid-solid and liquid-vapor phase transitions. It has been shown, in particular, that the hard-sphere (HS) potential is able to correctly describe the atomic structure of liquids and predict the liquid-solid phase transition (5). By contrast, the HS potential is unable to describe the liquid-vapor phase transition, which is essentially due to the presence of attractive forces of dispersion. More specifically, the simulation results have shown that the liquid-solid phase transition depends on the steric hindrance of the atoms and that the coexistence curve of liquid-solid phases is governed by the details of the repulsive part of potential. In fact, this was already contained in the phenomenological theories of melting, like the Lindemann theory that predicts the melting of a solid when the mean displacement of atoms from their equilibrium positions on the network exceeds the atomic diameter of 10%. In other words, a substance melts when its volume exceeds the volume at 0 K of 30%.

In restricting the discussion to simple centrosymmetric interactions from the outset, it is necessary to consider a realistic pair potential adequate for testing the phase stability limits. The most natural prototype potential is the Lennard-Jones (LJ) potential given by

$$u_{LJ}(r) = 4\varepsilon_{LJ} \left[ \left( \frac{\sigma_{LJ}}{r} \right)^m - \left( \frac{\sigma_{LJ}}{r} \right)^n \right], \quad (1)$$

where the parameters  $m$  and  $n$  are usually taken to be equal to 12 and 6, respectively. Such a functional form gives a reasonable representation of the interactions operating in real fluids, where the well depth  $\varepsilon_{LJ}$  and the collision diameter  $\sigma_{LJ}$  are independent of density and temperature. Figure (2a) displays the general shape of the Lennard-Jones potential ( $m - n$ ) corresponding to equation (1). Each substance has its own values of  $\varepsilon_{LJ}$  and  $\sigma_{LJ}$  so that, in reduced form, the LJ potentials have not only the same shape for all simple fluids, but superimpose each other rigorously. This is the condition for substances to conform to the *law of corresponding states*.

Figure (2b) represents the diagram  $p(T)$  of a pure substance. We can see how the slope of the coexistence curve of solid-liquid phases varies with the repulsive part of potential: the higher the value of  $m$ , the steeper the repulsive part of the potential (Fig. 2a) and, consequently, the more the coexistence curve of solid-liquid phases is tilted (Fig. 2b).

We can also remark that the LJ potential predicts the liquid-vapor coexistence curve, which begins at the triple point T and ends at the critical point C. A detailed analysis shows that the length of the branch TC is proportional to the depth  $\varepsilon$  of the potential well. As an example, for rare gases, it is verified that  $(T_C - T_T)k_B \simeq 0,55 \varepsilon$ . It follows immediately from this condition that the liquid-vapor coexistence curve disappears when the potential well is absent ( $\varepsilon = 0$ ).

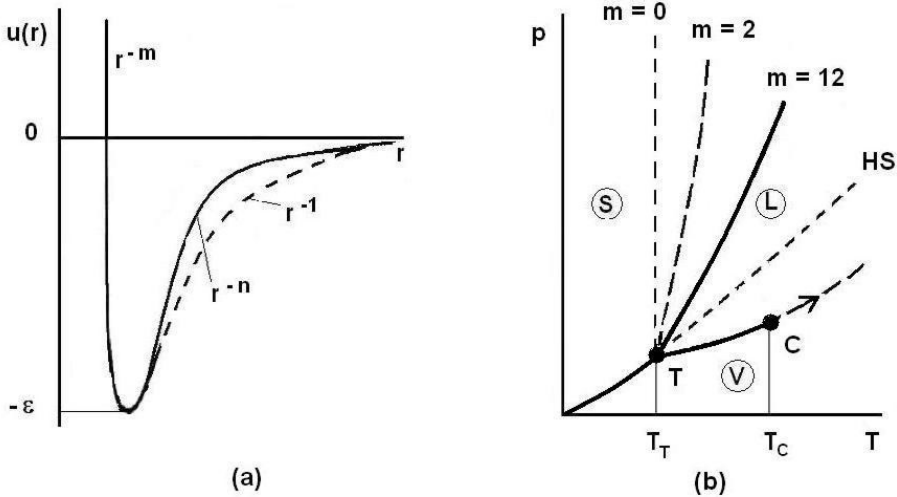


Fig. 2. Schematic representations of the Lennard-Jones potential ( $m - n$ ) and the diagram  $p(T)$ , as a function of the values of the parameters  $m$  and  $n$ .

The value of the slope of the branch TC also depends on the attractive part of the potential as shown by the Clausius-Clapeyron equation:

$$\frac{dp}{dT} = \frac{L_{vap}}{T_{vap}(V_{vap} - V_{liq})}, \tag{2}$$

where  $L_{vap}$  is the latent heat of vaporization at the corresponding temperature  $T_{vap}$  and  $(V_{vap} - V_{liq})$  is the difference of specific volumes between vapor and liquid. To evaluate the slope  $\frac{dp}{dT}$  of the branch TC at ambient pressure, we can estimate the ratio  $\frac{L_{vap}}{T_{vap}}$  with Trouton's rule ( $\frac{L_{vap}}{T_{vap}} \simeq 85 \text{ J.K}^{-1}.\text{mol}^{-1}$ ), and the difference in volume  $(V_{vap} - V_{liq})$  in terms of width of the potential well. Indeed, in noting that the quantity  $(V_{vap} - V_{liq})$  is an increasing function of the width of potential well, which itself increases when  $n$  decreases, we see that, for a given well depth  $\epsilon$ , the slope of the liquid-vapor coexistence curve decreases as  $n$  decreases.

For liquid metals, it should be mentioned that the repulsive part of the potential is softer than for liquid rare gases. Moreover, even if  $\epsilon$  is slightly lower for metals than for rare gases, the quantity  $\frac{(T_C - T_T)k_B}{\epsilon}$  is much higher (between 2 and 4), which explains the elongation of the TC curve compared to that of rare gases. It is worth also to indicate that some *flat-bottomed* potentials (6) are likely to give a good description of the physical properties of substances that have a low value of the ratio  $\frac{T_T}{T_C}$ . Such a potential is obviously not suitable for liquid rare gases, whose ratio  $\frac{T_T}{T_C} \simeq 0,56$ , or for organic and inorganic liquids, for which  $0,25 < \frac{T_T}{T_C} < 0,45$ . In return, it might be useful as empirical potential for metals with low melting point such as mercury, gallium, indium, tin, etc., the ratio of which being  $\frac{T_T}{T_C} < 0,1$ .

### 3. The structure of liquids

#### 3.1 Scattered radiation in liquids

The pair correlation function  $g(r)$  can be deduced from the experimental measurement of the structure factor  $S(q)$  by X-ray, neutron or electron diffractions. In condensed matter, the scatterers are essentially individual atoms, and diffraction experiments can only measure the structure of monatomic liquids such as rare gases and metals. By contrast, they provide no information on the structure of molecular liquids, unless they are composed of spherical molecules or monatomic ions, like in some molten salts.

Furthermore, each type of radiation-matter interaction has its own peculiarities. While the electrons are diffracted by all the charges in the atoms (electrons and nuclei), neutrons are diffracted by nuclei and X-rays are diffracted by the electrons localized on stable electron shells. The electron diffraction is practically used for fluids of low density, whereas the beams of neutrons and X-rays are used to study the structure of liquids, with their advantages and disadvantages. For example, the radius of the nuclei being 10,000 times smaller than that of atoms, it is not surprising that the structure factors obtained with neutrons are not completely identical to those obtained with X-rays.

To achieve an experience of X-ray diffraction, we must irradiate the liquid sample with a monochromatic beam of X-rays having a wavelength in the range of the interatomic distance ( $\lambda \sim 0,1$  nm). At this radiation corresponds a photon energy ( $h\nu = \frac{hc}{\lambda} \sim 10^4$  eV), much larger than the mean energy of atoms that is of the order of few  $k_B T$ , namely about  $10^{-1}$  eV. The large difference of the masses and energies between a photon and an atom makes that the photon-atom collision is elastic (constant energy) and that the liquid is transparent to the radiation. Naturally, the dimensions of the sample must be sufficiently large compared to the wavelength  $\lambda$  of the radiation, so that there are no side effects due to the walls of the enclosure - but not too much though for avoiding excessive absorption of the radiation. This would be particularly troublesome if the X-rays had to pass across metallic elements with large atomic numbers.

The incident radiation is characterized by its wavelength  $\lambda$  and intensity  $I_0$ , and the diffraction patterns depend on the structural properties of the liquids and on the diffusion properties of atoms. In neutron scattering, the atoms are characterized by the scattering cross section  $\sigma = 4\pi b^2$ , where  $b$  is a parameter approximately equal to the radius of the core ( $\sim 10^{-14}$  m). Note that the parameter  $b$  does not depend on the direction of observation but may vary slightly, even for a pure element, with the isotope. By contrast, for X-ray diffraction, the property corresponding to  $b$  is the atomic scattering factor  $A(q)$ , which depends on the direction of observation and electron density in the isolated atom. The structure factor  $S(q)$  obtained by X-ray diffraction has, in general, better accuracy at intermediate values of  $q$ . At the ends of the scale of  $q$ , it is less precise than the structure factor obtained by neutron diffraction, because the atomic scattering factor  $A(q)$  is very small for high values of  $q$  and very poorly known for low values of  $q$ .

#### 3.2 Structure factor and pair correlation function

When a photon of wave vector  $\mathbf{k} = \frac{2\pi}{\lambda} \mathbf{u}$  interacts with an atom, it is deflected by an angle  $\theta$  and the wave vector of the scattered photon is  $\mathbf{k}' = \frac{2\pi}{\lambda} \mathbf{u}'$ , where  $\mathbf{u}$  and  $\mathbf{u}'$  are unit vectors. If the scattering is elastic it results that  $|\mathbf{k}'| = |\mathbf{k}|$ , because  $E \propto k^2 = cte$ , and that the scattering vector (or transfer vector)  $\mathbf{q}$  is defined by the Bragg law:

$$\mathbf{q} = \mathbf{k}' - \mathbf{k}, \quad \text{and} \quad |\mathbf{q}| = 2 |\mathbf{k}| \sin \frac{\theta}{2} = \frac{4\pi}{\lambda} \sin \frac{\theta}{2}. \quad (3)$$



Now, if we consider an assembly of  $N$  identical atoms forming the liquid sample, the intensity scattered by the atoms in the direction  $\theta$  (or  $\mathbf{q}$ , according to Bragg's law) is given by:

$$I(q) = A_N A_N^* = A_0 A_0^* \sum_{j=1}^N \sum_{l=1}^N \exp \left[ i\mathbf{q} \left( \mathbf{r}_j - \mathbf{r}_l \right) \right].$$

In a crystalline solid, the arrangement of atoms is known once and for all, and the representation of the scattered intensity  $I$  is given by spots forming the Laue or Debye-Scherrer patterns. But in a liquid, the atoms are in continuous motion, and the diffraction experiment gives only the mean value of successive configurations during the experiment. Given the absence of translational symmetry in liquids, this mean value provides no information on long-range order. By contrast, it is a good measure of short-range order around each atom chosen as origin. Thus, in a liquid, the scattered intensity must be expressed as a function of  $q$  by the statistical average:

$$I(q) = I_0 \left\langle \sum_{l \neq j=1}^N \exp \left[ i\mathbf{q} \left( \mathbf{r}_j - \mathbf{r}_l \right) \right] \right\rangle + I_0 \left\langle \sum_{j=1}^N \sum_{l \neq j}^N \exp \left[ i\mathbf{q} \left( \mathbf{r}_j - \mathbf{r}_l \right) \right] \right\rangle. \quad (4)$$

The first mean value, for  $l = j$ , is worth  $N$  because it represents the sum of  $N$  terms, each of them being equal to unity. To evaluate the second mean value, one should be able to calculate the sum of exponentials by considering all pairs of atoms ( $j, l$ ) in all configurations counted during the experiment, then carry out the average of all configurations. However, this calculation can be achieved only by numerical simulation of a system made of a few particles. In a real system, the method adopted is to determine the mean contribution brought in by each pair of atoms ( $j, l$ ), using the probability of finding the atoms  $j$  and  $l$  in the positions  $\mathbf{r}'$  and  $\mathbf{r}$ , respectively. To this end, we rewrite the double sum using the Dirac delta function in order to calculate the statistical average in terms of the *density of probability*  $P_N(\mathbf{r}^N, \mathbf{p}^N)$  of the *canonical ensemble*<sup>1</sup>. Therefore, the statistical average can be written by using the *distribution*

<sup>1</sup> It seems useful to remember that the *probability density* function in the canonical ensemble is:

$$P_N(\mathbf{r}^N, \mathbf{p}^N) = \frac{1}{N! h^{3N} Q_N(V, T)} \exp \left[ -\beta H_N(\mathbf{r}^N, \mathbf{p}^N) \right],$$

where  $H_N(\mathbf{r}^N, \mathbf{p}^N) = \sum \frac{p^2}{2m} + U(\mathbf{r}^N)$  is the *Hamiltonian* of the system,  $\beta = \frac{1}{k_B T}$  and  $Q_N(V, T)$  the *partition function* defined as:

$$Q_N(V, T) = \frac{Z_N(V, T)}{N! \Lambda^{3N}},$$

with the *thermal wavelength*  $\Lambda$ , which is a measure of the thermodynamic uncertainty in the localization of a particle of mass  $m$ , and the *configuration integral*  $Z_N(V, T)$ , which is expressed in terms of the total *potential energy*  $U(\mathbf{r}^N)$ . They read:

$$\Lambda = \sqrt{\frac{h^2}{2\pi m k_B T}},$$

$$\text{and } Z_N(V, T) = \int \int_N \exp \left[ -\beta U(\mathbf{r}^N) \right] d\mathbf{r}^N.$$

Besides, the partition function  $Q_N(V, T)$  allows us to determine the free energy  $F$  according to the relation:

$$F = E - TS = -k_B T \ln Q_N(V, T).$$

The reader is advised to consult statistical-physics textbooks for further details.

function<sup>2</sup>  $\rho_N^{(2)}(\mathbf{r}, \mathbf{r}')$  in the form:

$$\left\langle \sum_{j=1}^N \sum_{l \neq j}^N \exp [i\mathbf{q}(\mathbf{r}_j - \mathbf{r}_l)] \right\rangle = \int \int_6 d\mathbf{r} d\mathbf{r}' \exp [i\mathbf{q}(\mathbf{r}' - \mathbf{r})] \rho_N^{(2)}(\mathbf{r}, \mathbf{r}').$$

If the liquid is assumed to be homogeneous and isotropic, and that all atoms have the same properties, one can make the changes of variables  $\mathbf{R} = \mathbf{r}$  and  $\mathbf{X} = \mathbf{r}' - \mathbf{r}$ , and explicit the pair correlation function  $g(|\mathbf{r}' - \mathbf{r}|) = \frac{\rho_N^{(2)}(\mathbf{r}, \mathbf{r}')}{\rho^2}$  in the statistical average as<sup>3</sup>:

$$\left\langle \sum_{j=1}^N \sum_{l \neq j}^N \exp [i\mathbf{q}(\mathbf{r}_j - \mathbf{r}_l)] \right\rangle = 4\pi\rho^2 V \int_0^\infty \frac{\sin(qr)}{qr} g(r) r^2 dr. \quad (5)$$

One sees that the previous integral diverges because the integrand increases with  $r$ . The problem comes from the fact that the scattered intensity, for  $q = 0$ , has no physical meaning and can not be measured. To overcome this difficulty, one rewrites the scattered intensity  $I(q)$  defined by equation (4) in the equivalent form (cf. footnote 3):

$$I(q) = NI_0 + NI_0\rho \int_V \exp(i\mathbf{q}\mathbf{r}) [g(r) - 1] d\mathbf{r} + NI_0\rho \int_V \exp(i\mathbf{q}\mathbf{r}) d\mathbf{r}. \quad (6)$$

To large distances,  $g(r)$  tends to unity, so that  $[g(r) - 1]$  tends towards zero, making the first integral convergent. As for the second integral, it corresponds to the Dirac delta function<sup>4</sup>,

<sup>2</sup> It should be stressed that the distribution function  $\rho_N^{(2)}(\mathbf{r}^2)$  is expressed as:

$$\rho_N^{(2)}(\mathbf{r}, \mathbf{r}') = \rho^2 g(|\mathbf{r}' - \mathbf{r}|) = \frac{N!}{(N-2)!Z_N} \int \int_{3(N-2)} \exp[-\beta U(\mathbf{r}^N)] d\mathbf{r}_3 \dots d\mathbf{r}_N.$$

<sup>3</sup> To evaluate an integral of the form:

$$I = \int_V d\mathbf{r} \exp(i\mathbf{q}\mathbf{r}) g(r),$$

one must use the spherical coordinates by placing the vector  $\mathbf{q}$  along the  $z$  axis, where  $\theta = (\mathbf{q}, \mathbf{r})$ . Thus, the integral reads:

$$I = \int_0^{2\pi} d\varphi \int_0^\pi \int_0^\infty \exp(iqr \cos \theta) g(r) r^2 \sin \theta d\theta dr,$$

with  $\mu = \cos \theta$  and  $d\mu = -\sin \theta d\theta$ . It follows that:

$$I = -2\pi \int_0^\infty \left[ \int_{+1}^{-1} \exp(iqr\mu) d\mu \right] g(r) r^2 dr = 4\pi \int_0^\infty \frac{\sin(qr)}{qr} g(r) r^2 dr.$$

<sup>4</sup> The generalization of the Fourier transform of the Dirac delta function to three dimensions is:

$$\delta(\mathbf{r}) = \frac{1}{(2\pi)^3} \int \int \int_{-\infty}^{+\infty} \delta(\mathbf{q}) \exp(-i\mathbf{q}\mathbf{r}) d\mathbf{q} = \frac{1}{(2\pi)^3},$$

and the inverse transform is:

$$\delta(\mathbf{q}) = \int \int \int_{-\infty}^{+\infty} \delta(\mathbf{r}) \exp(i\mathbf{q}\mathbf{r}) d\mathbf{r} = \frac{1}{(2\pi)^3} \int \int \int_{-\infty}^{+\infty} \exp(i\mathbf{q}\mathbf{r}) d\mathbf{r}.$$

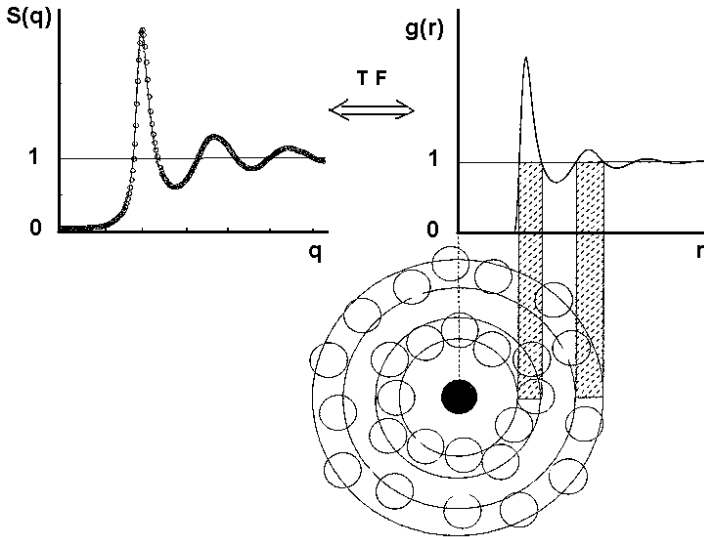


Fig. 3. Structure factor  $S(q)$  and pair correlation function  $g(r)$  of simple liquids.

which is zero for all values of  $q$ , except in  $q = 0$  for which it is infinite. In using the delta function, the expression of the scattered intensity  $I(q)$  becomes:

$$I(q) = NI_0 + NI_0\rho \int_V \exp(i\mathbf{q}\mathbf{r}) [g(r) - 1] d\mathbf{r} + NI_0\rho(2\pi)^3\delta(\mathbf{q}).$$

From the experimental point of view, it is necessary to exclude the measurement of the scattered intensity in the direction of the incident beam ( $q = 0$ ). Therefore, in practice, the structure factor  $S(q)$  is defined by the following normalized function:

$$S(q) = \frac{I(q) - (2\pi)^3 NI_0\rho\delta(\mathbf{q})}{NI_0} = 1 + 4\pi\rho \int_0^\infty \frac{\sin(qr)}{qr} [g(r) - 1] r^2 dr. \quad (7)$$

Consequently, the pair correlation function  $g(r)$  can be extracted from the experimental results of the structure factor  $S(q)$  by performing the numerical Fourier transformation:

$$\rho [g(r) - 1] = TF [S(q) - 1].$$

The pair correlation function  $g(r)$  is a dimensionless quantity, whose the graphic representation is given in figure (3). The gap around unity measures the probability of finding a particle at distance  $r$  from a particle taken in an arbitrary origin. The main peak of  $g(r)$  corresponds to the position of first neighbors, and the successive peaks to the next close neighbors. The pair correlation function  $g(r)$  clearly shows the existence of a short-range order that is fading rapidly beyond four or five interatomic distances. In passing, it should be mentioned that the structure factor at  $q = 0$  is related to the isothermal compressibility by the exact relation  $S(0) = \rho k_B T \chi_T$ .

## 4. Thermodynamic functions of liquids

### 4.1 Internal energy

To express the internal energy of a liquid in terms of the pair correlation function, one must first use the following relation from statistical mechanics :

$$E = k_B T^2 \frac{\partial}{\partial T} \ln Q_N(V, T),$$

where the partition function  $Q_N(V, T)$  depends on the configuration integral  $Z_N(V, T)$  and on the thermal wavelength  $\Lambda$ , in accordance with the equations given in footnote (1). The derivative of  $\ln Q_N(V, T)$  with respect to  $T$  can be written:

$$\frac{\partial}{\partial T} \ln Q_N(V, T) = \frac{\partial}{\partial T} \ln Z_N(V, T) - 3N \frac{\partial}{\partial T} \ln \Lambda,$$

with:

$$\frac{\partial}{\partial T} \ln Z_N(V, T) = \frac{1}{Z_N(V, T)} \int \int \left[ \frac{1}{k_B T^2} U(\mathbf{r}^N) \right] \exp \left[ -\beta U(\mathbf{r}^N) \right] d\mathbf{r}^N$$

and

$$\frac{\partial}{\partial T} \ln \Lambda = \frac{1}{\Lambda} \left( -\frac{1}{2T^{3/2}} \sqrt{\frac{h^2}{2\pi m k_B}} \right) = -\frac{1}{2T}.$$

Then, the calculation is continued by admitting that the total potential energy  $U(\mathbf{r}^N)$  is written as a sum of pair potentials, in the form  $U(\mathbf{r}^N) = \sum_i \sum_{j>i} u(r_{ij})$ . The internal energy reads:

$$E = \frac{3}{2} N k_B T + \frac{1}{Z_N(V, T)} \int \int \left[ \sum_i \sum_{j>i} u(r_{ij}) \right] \exp \left[ -\beta U(\mathbf{r}^N) \right] d\mathbf{r}^N. \quad (8)$$

The first term on the RHS corresponds to the kinetic energy of the system; it is the ideal gas contribution. The second term represents the potential energy. Given the assumption of additivity of pair potentials, we can assume that it is composed of  $N(N-1)/2$  identical terms, permitting us to write:

$$\sum_i \sum_{j>i} \frac{1}{Z_N(V, T)} \int \int u(r_{ij}) \exp \left[ -\beta U(\mathbf{r}^N) \right] d\mathbf{r}^N = \frac{N(N-1)}{2} \langle u(r_{ij}) \rangle,$$

where the mean value is expressed in terms of the pair correlation function as:

$$\langle u(r_{12}) \rangle = \rho^2 \frac{(N-2)!}{N!} \int \int_6 u(r_{12}) \left[ g_N^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \right] d\mathbf{r}_1 d\mathbf{r}_2.$$

For a homogeneous and isotropic fluid, one can perform the change of variables  $\mathbf{R} = \mathbf{r}_1$  and  $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ , where  $\mathbf{R}$  and  $\mathbf{r}$  describe the system volume, and write the expression of internal energy in the integral form:

$$E = \frac{3}{2} N k_B T + 2\pi\rho N \int_0^\infty u(r) g(r) r^2 dr. \quad (9)$$

Therefore, the calculation of internal energy of a liquid requires knowledge of the pair potential  $u(r)$  and the pair correlation function  $g(r)$ . For the latter, the choice is to employ either the experimental values or values derived from the microscopic theory of liquids. Note that the integrand in equation (9) is the product of the pair potential by the pair correlation function, weighted by  $r^2$ . It should be also noted that the calculation of  $E$  can be made taking into account the three-body potential  $u_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$  and the three-body correlation function  $g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ . In this case, the correlation function at three bodies must be determined only by the theory of liquids (7), since it is not accessible by experiment.

#### 4.2 Pressure

The expression of the pressure is obtained in the same way that the internal energy, in considering the equation:

$$p = k_B T \frac{\partial}{\partial V} \ln Q_N(V, T) = k_B T \frac{\partial}{\partial V} \ln Z_N(V, T).$$

The derivation of the configuration integral with respect to volume requires using the reduced variable  $\mathbf{X} = \frac{\mathbf{r}}{V^{1/3}}$  that allows us to find the dependence of the potential energy  $U(\mathbf{r}^N)$  versus volume. Indeed, if the volume element is  $d\mathbf{r} = V d\mathbf{X}$ , the scalar variable  $dr = V^{1/3} dX$  leads to the derivative:

$$\frac{dr}{dV} = \frac{1}{3} V^{-2/3} X = \frac{1}{3V} r. \quad (10)$$

In view of this, the configuration integral and its derivative with respect to  $V$  are written in the following forms with reduced variables:

$$Z_N(V, T) = V^N \int_{3N} \exp[-\beta U(\mathbf{r}^N)] d\mathbf{X}_1 \dots d\mathbf{X}_N,$$

$$\frac{\partial}{\partial V} \ln Z_N(V, T) = \frac{N}{V} + \frac{V^N}{Z_N(V, T)} \int_{3N} \left[ -\beta \frac{\partial U(\mathbf{r}^N)}{\partial V} \right] \exp[-\beta U(\mathbf{r}^N)] d\mathbf{X}_1 \dots d\mathbf{X}_N.$$

Assuming that the potential energy is decomposed into a sum of pair potentials, and with the help of equation (10), the derivation of the potential energy versus volume is performed as:

$$\frac{\partial U(\mathbf{r}^N)}{\partial V} = \frac{1}{3V} \sum_i \sum_{j>i} r_{ij} \frac{\partial u(r_{ij})}{\partial r_{ij}},$$

so that the expression of the pressure becomes:

$$p = k_B T \frac{N}{V} - \frac{1}{3V} \frac{1}{Z_N(V, T)} \sum_i \sum_{j>i} \int_{3N} \left[ r_{ij} \frac{\partial u(r_{ij})}{\partial r_{ij}} \right] \exp[-\beta U(\mathbf{r}^N)] d\mathbf{r}_1 \dots d\mathbf{r}_N. \quad (11)$$

Like for the calculation of internal energy, the additivity assumption of pair potentials permits us to write the sum of integrals of the previous equation as:

$$\sum_i \sum_{j>i} \frac{1}{Z_N(V, T)} \int \int \left[ r_{ij} \frac{\partial u(r_{ij})}{\partial r_{ij}} \right] \exp[-\beta U(\mathbf{r}^N)] d\mathbf{r}^N = \frac{N(N-1)}{2} \left\langle r_{ij} \frac{\partial u(r_{ij})}{\partial r_{ij}} \right\rangle,$$

where the mean value is expressed with the pair correlation function by:

$$\left\langle r_{12} \frac{\partial u(r_{12})}{\partial r_{12}} \right\rangle = \rho^2 \frac{(N-2)!}{N!} \int \int_6 r_{12} \frac{\partial u(r_{12})}{\partial r_{12}} \left[ g_N^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \right] d\mathbf{r}_1 d\mathbf{r}_2.$$

For a homogeneous and isotropic fluid, one can perform the change of variables  $\mathbf{R} = \mathbf{r}_1$  and  $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ , and simplify the expression of pressure as:

$$p = k_B T \frac{N}{V} - \frac{2\pi}{3} \rho^2 \int_0^\infty r^3 \frac{\partial u(r)}{\partial r} g(r) dr. \quad (12)$$

The previous equation provides the pressure of a liquid as a function of the pair potential and the pair correlation function. It is the so-called *pressure equation of state* of liquids. It should be stressed that this equation of state is not unique, as we will see in presenting the hard-sphere reference system (§ 4. 4). As the internal energy, the pressure can be written with an additional term containing the three-body potential  $u_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$  and the three-body correlation function  $g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ .

### 4.3 Chemical potential and entropy

We are now able to calculate the internal energy (Eq. 9) and pressure (Eq. 12) for any system, of which the potential energy is made of a sum of pair potentials  $u(r)$  and the pair correlation function  $g(r)$  is known. Beside this, all other thermodynamic properties can be easily derived. Traditionally, it is appropriate to derive the chemical potential  $\mu$  as a function of  $g(r)$  by integrating the partition function with respect to a parameter  $\lambda$  to be defined (8). Firstly, the formal expression of the chemical potential is defined by the energy required to introduce a new particle in the system:

$$\mu = F(V, T, N) - F(V, T, N - 1) = \left( \frac{\partial F}{\partial N} \right)_{V, T}.$$

From footnote (1), the free energy  $F$  is written:

$$F(V, T, N) = -k_B T \ln Q_N(V, T) = -k_B T \left[ \ln Z_N(V, T) - \ln N! - N \ln \Lambda^3 \right],$$

so that the chemical potential can be simplified as:

$$\mu = k_B T \left[ -\ln \frac{Z_N(V, T)}{Z_{N-1}(V, T)} + \ln N + \ln \Lambda^3 \right]. \quad (13)$$

Secondly, the procedure requires to write the potential energy as a function of the *coupling parameter*  $\lambda$ , under the following form, in order to assess the argument of the logarithm in the above relation:

$$U(\mathbf{r}^N, \lambda) = \lambda \sum_{j=2}^N u(r_{1j}) + \sum_i^N \sum_{j>i \geq 2}^N u(r_{ij}). \quad (14)$$

Varying from 0 to 1, the coupling parameter  $\lambda$  measures the degree of coupling of the particle to which it is assigned (1 in this case) with the rest of the system. In the previous relation,  $\lambda = 1$

means that particle 1 is completely coupled with the other particles, while  $\lambda = 0$  indicates a zero coupling, that is to say the absence of the particle 1 in the system. This allows the writing of the important relations:

$$U(\mathbf{r}^N, 1) = \sum_{j=2}^N u(r_{1j}) + \sum_i \sum_{j>i \geq 2}^N u(r_{ij}) = \sum_i \sum_{j>i \geq 1}^N u(r_{ij}) = U(\mathbf{r}^N),$$

and

$$U(\mathbf{r}^N, 0) = \sum_i \sum_{j>i \geq 2}^N u(r_{ij}) = U(\mathbf{r}^{N-1}).$$

Under these conditions, the configuration integrals for a total coupling ( $\lambda = 1$ ) and a zero coupling ( $\lambda = 0$ ) are respectively:

$$Z_N(V, T, \lambda = 1) = \int \int_{3N} \exp[-\beta U(\mathbf{r}^N)] d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N = Z_N(V, T), \quad (15)$$

$$Z_N(V, T, \lambda = 0) = \int_V d\mathbf{r}_1 \int \int_{3(N-1)} \exp[-\beta U(\mathbf{r}^{N-1})] d\mathbf{r}_2 \dots d\mathbf{r}_N = V Z_{N-1}(V, T). \quad (16)$$

These expressions are then used to calculate the logarithm of the ratio of configuration integrals in equation (13):

$$\ln \frac{Z_N(V, T)}{Z_{N-1}(V, T)} = \ln \frac{Z_N(V, T, \lambda = 1)}{Z_N(V, T, \lambda = 0)} + \ln V \quad (17)$$

$$= \ln V + \int_0^1 \frac{\partial \ln Z_N}{\partial \lambda} d\lambda. \quad (18)$$

But with the configuration integral  $Z_N(V, T, \lambda)$ , in which potential energy is given by equation (14), we can easily evaluate the partial derivatives  $\frac{\partial Z_N}{\partial \lambda}$  and  $\frac{\partial \ln Z_N}{\partial \lambda}$ . In particular, with the result of the footnote (1), we can write  $\frac{\partial \ln Z_N}{\partial \lambda}$  as a function of the pair correlation function as:

$$\frac{\partial \ln Z_N(V, T, \lambda)}{\partial \lambda} = -\beta \rho^2 \frac{(N-1)(N-2)!}{N!} \int \int_6 u(r_{12}) \left\{ g_N^{(2)}(\mathbf{r}_1, \mathbf{r}_2, \lambda) \right\} d\mathbf{r}_1 d\mathbf{r}_2.$$

In addition, if the fluid is homogeneous and isotropic, the above relation simplifies under the following form:

$$\frac{\partial \ln Z_N(V, T, \lambda)}{\partial \lambda} = -\frac{\beta \rho^2}{N} V \int_0^\infty u(r) g(r, \lambda) 4\pi r^2 dr,$$

that remains only to be substituted in equation (18) for obtaining the logarithm of the ratio of configuration integrals. And by putting the last expression in equation (13), one ultimately arrives to the expression of the chemical potential:

$$\mu = k_B T \ln \rho \Lambda^3 + 4\pi \rho \int_0^1 \int_0^\infty u(r) g(r, \lambda) r^2 dr d\lambda. \quad (19)$$

Thus, like the internal energy (Eq. 9) and pressure (Eq. 12), the chemical potential (Eq. 19) is calculated using the pair potential and pair correlation function.

Finally, one writes the entropy  $S$  in terms of the pair potential and pair correlation function, owing to the expressions of the internal energy (Eq. 9), pressure (Eq. 12) and chemical potential (Eq. 19) (cf. footnote 1):

$$S = \frac{E - F}{T} = \frac{E}{T} - \frac{\mu N}{T} + \frac{pV}{T}. \quad (20)$$

It should be noted that the entropy can also be estimated only with the pair correlation function  $g(r)$ , without recourse to the pair potential  $u(r)$ . The reader interested by this issue should refer to the original articles (9).

#### 4.4 Application to the hard-sphere potential

In this subsection we determine the equation of state of the hard-sphere system, of which the pair potential being:

$$u(r) = \begin{cases} \infty & \text{if } r < \sigma \\ 0 & \text{if } r > \sigma, \end{cases}$$

where  $\sigma$  is the hard-sphere diameter. The Boltzmann factor associated with this potential has a significant feature that enable us to express the thermodynamic properties under particularly simple forms. Indeed, the representation of the Boltzmann factor

$$\exp[-\beta u(r)] = \begin{cases} 0 & \text{if } r < \sigma \\ 1 & \text{if } r > \sigma, \end{cases}$$

is a step function (Fig. 4) whose derivative with respect to  $r$  is the Dirac delta function, i. e.:

$$\frac{\partial}{\partial r} \exp[-\beta u(r)] = -\beta \frac{\partial u}{\partial r} \exp[-\beta u(r)] = \delta(r - \sigma).$$

In substituting  $\frac{\partial u}{\partial r}$ , taken from the previous relation, in equation (12) we find the expression of the pressure:

$$p = k_B T \frac{N}{V} - \frac{2\pi}{3} \rho^2 \int_0^\infty r^3 \left\{ -\frac{1}{\beta} \frac{\delta(r - \sigma)}{\exp[-\beta u(r)]} \right\} g(r) dr,$$

or:

$$p = k_B T \frac{N}{V} + \frac{2\pi}{3} k_B T \rho^2 \sigma^3 g(\sigma) \exp[\beta u(\sigma)]. \quad (21)$$

It is important to recall that, for moderately dense gases, the pressure is usually expressed under the form of the virial expansion

$$\frac{p}{\rho k_B T} = 1 + \rho B_2(T) + \rho^2 B_3(T) + \rho^3 B_4(T) + \dots = \frac{p^{GP}}{\rho k_B T} + \frac{p^{ex}}{\rho k_B T}.$$



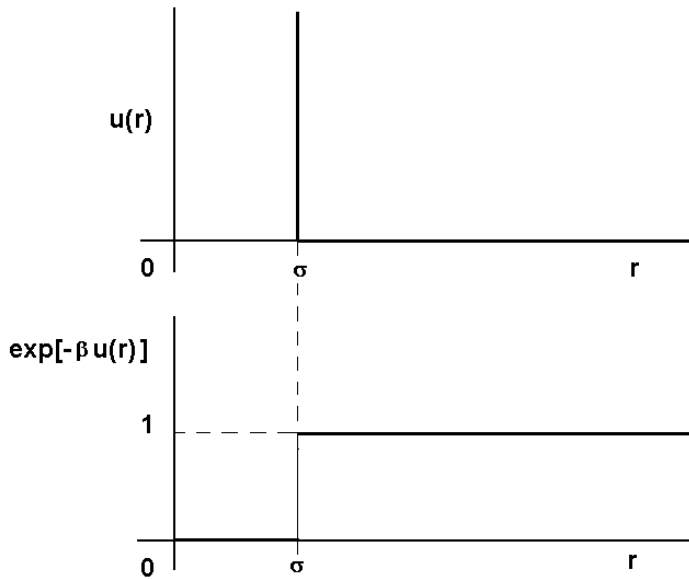


Fig. 4. Representation of the hard-sphere potential and its Boltzmann factor.

The first term of the last equality represents the contribution of the ideal gas, and the excess pressure  $p^{ex}$  comes from the interactions between particles. They are written:

$$\frac{p^{GP}}{\rho k_B T} = 1,$$

and

$$\frac{p^{ex}}{\rho k_B T} = 4\eta + \eta^2 B'_3(T) + \eta^3 B'_4(T) + \dots$$

where  $\eta$  is the *packing fraction* defined by the ratio of the volume actually occupied by the  $N$  spherical particles on the total volume  $V$  of the system, that is to say:

$$\eta = \frac{1}{V} \frac{4\pi}{3} \left(\frac{\sigma}{2}\right)^3 N = \frac{\pi}{6} \rho \sigma^3. \tag{22}$$

Note that the first 6 coefficients of the excess pressure  $p^{ex}$  have been calculated analytically and by molecular dynamics (10), with great accuracy. In addition, Carnahan and Starling (11) have shown that the excess pressure of the hard-sphere fluid can be very well predicted by rounding the numerical values of the 6 coefficients towards the nearest integer values, according to the expansion:

$$\frac{p^{ex}}{\rho k_B T} \simeq 4\eta + 10\eta^2 + 18\eta^3 + 28\eta^4 + 40\eta^5 + 54\eta^6 \dots \simeq \sum_{k=1}^{\infty} (k^2 + 3k)\eta^k. \tag{23}$$

In combining the first and second derivatives of the geometric series  $\sum_{k=1}^{\infty} \eta^k$ , it is found that equation (23) can be transformed into a rational fraction<sup>5</sup> enabling the deduction of the excess pressure in the form:

$$\frac{p^{ex}}{\rho k_B T} \simeq \sum_{k=1}^{\infty} (k^2 + 3k)\eta^k = \frac{4\eta - 2\eta^2}{(1 - \eta)^3}.$$

Consequently, the equation of state of the hard-sphere fluid is written with excellent precision as:

$$\frac{p}{\rho k_B T} = \frac{1 + \eta + \eta^2 - \eta^3}{(1 - \eta)^3}. \quad (24)$$

It is also possible to calculate the internal energy of the hard-sphere fluid by substituting  $u(r)$  in equation (9). Given that  $u(r)$  is zero when  $r > \sigma$  and  $g(r)$  is zero when  $r < \sigma$ , it follows that the integral is always zero, and that the internal energy of the hard-sphere fluid is equal to that of the ideal gas  $E = \frac{3}{2} N k_B T$ .

As for the free energy  $F$ , it is determined by integrating the pressure over volume with the equation:

$$p = - \left( \frac{\partial F}{\partial V} \right)_T = - \left( \frac{\partial F^{GP}}{\partial V} \right)_T - \left( \frac{\partial F^{ex}}{\partial V} \right)_T,$$

where  $F^{GP}$  is the free energy of ideal gas (cf. footnote 1, with  $Z_N(V, T) = V^N$ ):

$$F^{GP} = N k_B T \left( \ln \rho \Lambda^3 - 1 \right),$$

and  $F^{ex}$  the excess free energy, calculated by integrating equation (23) as follows:

$$F^{ex} = - \int p^{ex} dV = - \int \frac{N k_B T}{V} \left( 4\eta + 10\eta^2 + 18\eta^3 + 28\eta^4 + 40\eta^5 + 54\eta^6 \dots \right) \frac{dV}{d\eta} d\eta.$$

<sup>5</sup> To obtain the rational fraction, one must decompose the sum as:

$$\sum_{k=1}^{\infty} (k^2 + 3k)\eta^k = \sum_{k=1}^{\infty} (k^2 - k)\eta^k + \sum_{k=1}^{\infty} 4k\eta^k,$$

and combine the geometric series  $\sum_{k=1}^{\infty} \eta^k$  with its first and second derivatives:

$$\sum_{k=1}^{\infty} \eta^k = \eta + \eta^2 + \eta^3 + \dots = \frac{\eta}{1 - \eta},$$

$$\sum_{k=1}^{\infty} k\eta^{k-1} = \frac{1}{(1 - \eta)^2},$$

$$\text{and} \quad \sum_{k=1}^{\infty} k(k-1)\eta^{k-2} = \frac{2}{(1 - \eta)^3},$$

to see appear the relation:

$$\sum_{k=1}^{\infty} (k^2 + 3k)\eta^k = \frac{2\eta^2}{(1 - \eta)^3} + \frac{4\eta}{(1 - \eta)^2}.$$

But, with equation (22) that gives  $\frac{dV}{d\eta} = -\frac{V}{\eta}$ ,  $F^{ex}$  is then reduced to the series expansion:

$$\frac{F^{ex}}{Nk_B T} = 4\eta + 5\eta^2 + 6\eta^3 + 7\eta^4 + 8\eta^5 + 9\eta^6 \dots = \sum_{k=1}^{\infty} (k+3)\eta^k.$$

Like the pressure, this expansion is written as a rational function by combining the geometric series  $\sum_{k=1}^{\infty} \eta^k$  with its first derivative<sup>6</sup>. The expression of the excess free energy is:

$$\frac{F^{ex}}{Nk_B T} = \sum_{k=1}^{\infty} (k+3)\eta^k = \frac{4\eta - 3\eta^2}{(1-\eta)^2},$$

and the free energy of the hard-sphere fluid reduces to the following form:

$$\frac{F}{Nk_B T} = \frac{F^{GP}}{Nk_B T} + \frac{F^{ex}}{Nk_B T} = \ln \rho \Lambda^3 - 1 + \frac{4\eta - 3\eta^2}{(1-\eta)^2}. \quad (25)$$

Now, the entropy is obtained using the same method of calculation, by deriving the free energy with respect to temperature:

$$S = - \left( \frac{\partial F}{\partial T} \right)_V = - \left( \frac{\partial F^{GP}}{\partial T} \right)_V - \left( \frac{\partial F^{ex}}{\partial T} \right)_V,$$

where  $S^{GP}$  is the entropy of the ideal gas given by the Sackur-Tetrode equation:

$$S^{GP} = -Nk_B \left( \ln \rho \Lambda^3 - \frac{5}{2} \right),$$

and where the excess entropy  $S^{ex}$  arises from the relation:

$$S^{ex} = - \left( \frac{\partial F^{ex}}{\partial T} \right)_V = -Nk_B \frac{4\eta - 3\eta^2}{(1-\eta)^2},$$

hence the expression of the entropy of the hard-sphere fluid:

$$\frac{S}{Nk_B} = -\ln \rho \Lambda^3 + \frac{5}{2} - \frac{4\eta - 3\eta^2}{(1-\eta)^2}. \quad (26)$$

Finally, combining equations (25) and (24), with the help of equation (20), one reaches the chemical potential of the hard-sphere fluid that reads:

$$\frac{\mu}{k_B T} = \frac{F}{Nk_B T} + \frac{p}{\rho k_B T} = \ln \rho \Lambda^3 - 1 + \frac{1 + 5\eta - 6\eta^2 + 2\eta^3}{(1-\eta)^3}. \quad (27)$$

<sup>6</sup> Indeed, the identity:

$$\sum_{k=1}^{\infty} (k+3)\eta^k = \sum_{k=1}^{\infty} k\eta^k + \sum_{k=1}^{\infty} 3\eta^k,$$

is yet written:

$$\sum_{k=1}^{\infty} (k+3)\eta^k = \frac{\eta}{(1-\eta)^2} + \frac{3\eta}{(1-\eta)},$$

Since they result from equation (23), the expressions of thermodynamic properties ( $p$ ,  $F$ ,  $S$  and  $\mu$ ) of the hard-sphere fluid make up a homogeneous group of relations related to the Carnahan and Starling equation of state. But other expressions of thermodynamic properties can also be determined using the pressure equation of state (Eq. 12) and the compressibility equation of state, which will not be discussed here. Unlike the Carnahan and Starling equation of state, these two equations of state require knowledge of the pair correlation function of hard spheres,  $g_{HS}(r)$ . The latter is not available in analytical form. The interested reader will find the Fortran program aimed at doing its calculation, in the book by McQuarrie (12), page 600. It should be mentioned that the thermodynamic properties ( $p$ ,  $F$ ,  $S$  and  $\mu$ ), obtained with the equations of state of pressure and compressibility, have analytical forms similar to those from the Carnahan and Starling equation of state, and they provide results whose differences are indistinguishable to low densities.

### 5. Thermodynamic perturbation theory

All theoretical and experimental studies have shown that the structure factor  $S(q)$  of simple liquids resembles that of the hard-sphere fluid. For proof, just look at the experimental structure factor of liquid sodium (13) at 373 K, in comparison with the structure factor of hard-sphere fluid (14) for a value of the packing fraction  $\eta$  of 0.45. We can see that the agreement is not bad, although there is a slight shift of the oscillations and ratios of peak heights significantly different. Besides, numerical calculations showed that the structure factor obtained with the Lennard-Jones potential describes the structure of simple fluids (15) and looks like the structure factor of hard-sphere fluid whose diameter is chosen correctly.

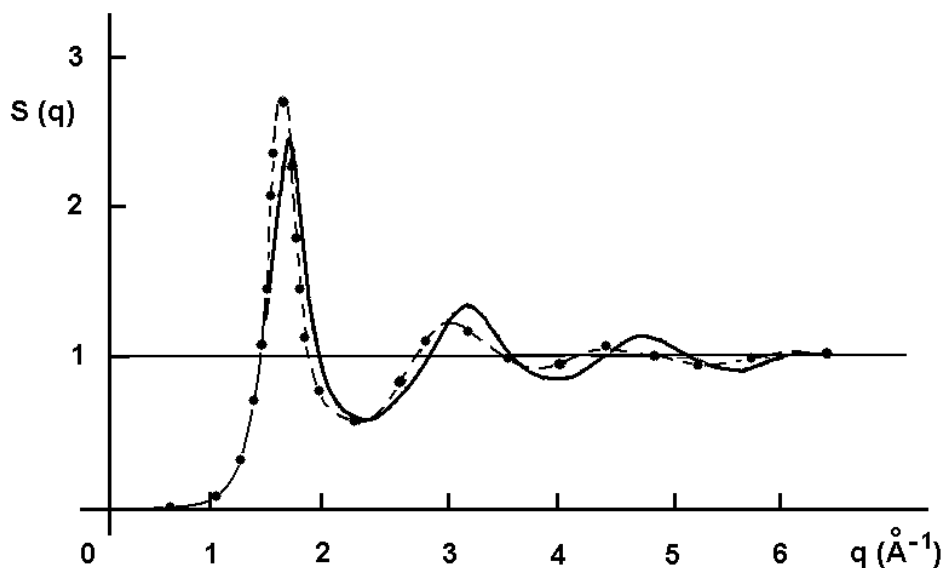


Fig. 5. Experimental structure factor of liquid sodium at 373 K (points), and hard-sphere structure factor (solid curve), with  $\eta = 0,45$ .

Such a qualitative success emphasizes the role played by the repulsive part of the pair potential to describe the structure factor of liquids, while the long-ranged attractive contribution has a minor role. It can be said for simplicity that the repulsive contribution of the potential determines the structure of liquids (stacking of atoms and steric effects) and the attractive contribution is responsible for their cohesion.

It is important to remember that the thermodynamic properties of the hard-sphere fluid (Eqs. 24, 25, 26, 27) and the structure factor  $S_{HS}(q)$  can be calculated with great accuracy. That suggests replacing the repulsive part of potential in real systems by the hard-sphere potential that becomes the *reference system*, and predict the structural and thermodynamic properties of real systems with those of the hard-sphere fluid, after making the necessary adaptations. To perform these adaptations, the attractive contribution of potential should be treated as a perturbation to the reference system.

The rest of this subsection is devoted to a summary of thermodynamic perturbation methods<sup>7</sup>. It should be noted, from the outset, that the calculation of thermodynamic properties with the thermodynamic perturbation methods requires knowledge of the pair correlation function  $g_{HS}(r)$  of the hard-sphere system and not that of the real system.

### 5.1 Zwanzig method

In perturbation theory proposed by Zwanzig (16), it is assumed that the total potential energy  $U(\mathbf{r}^N)$  of the system can be divided into two parts. The first part,  $U_0(\mathbf{r}^N)$ , is the energy of the unperturbed system considered as reference system and the second part,  $U_1(\mathbf{r}^N)$ , is the energy of the perturbation which is much smaller than  $U_0(\mathbf{r}^N)$ . More precisely, it is posed that the potential energy depends on the coupling parameter  $\lambda$  by the relation:

$$U(\mathbf{r}^N) = U_0(\mathbf{r}^N) + \lambda U_1(\mathbf{r}^N)$$

in order to vary continuously the potential energy from  $U_0(\mathbf{r}^N)$  to  $U(\mathbf{r}^N)$ , by changing  $\lambda$  from 0 to 1, and that the free energy  $F$  of the system is expanded in Taylor series as:

$$F = F_0 + \lambda \left( \frac{\partial F}{\partial \lambda} \right) + \frac{\lambda^2}{2} \left( \frac{\partial^2 F}{\partial \lambda^2} \right) + \dots \quad (28)$$

By replacing the potential energy  $U(\mathbf{r}^N)$  in the expression of the configuration integral (cf. footnote 1), one gets:

$$Z_N(V, T) = \int \int_{3N} \exp[-\beta U_0(\mathbf{r}^N)] d\mathbf{r}^N \times \frac{\int \int_{3N} \{ \exp[-\beta \lambda U_1(\mathbf{r}^N)] \} \exp[-\beta U_0(\mathbf{r}^N)] d\mathbf{r}^N}{\int \int_{3N} \exp[-\beta U_0(\mathbf{r}^N)] d\mathbf{r}^N}.$$

The first integral represents the configuration integral  $Z_N^{(0)}(V, T)$  of the reference system, and the remaining term can be regarded as the average value of the quantity  $\exp[-\beta \lambda U_1(\mathbf{r}^N)]$ , so that the previous relation can be put under the general form:

$$Z_N(V, T) = Z_N^{(0)}(V, T) \left\langle \exp[-\beta \lambda U_1(\mathbf{r}^N)] \right\rangle_0, \quad (29)$$

where  $\langle \dots \rangle_0$  refers to the statistical average in the canonical ensemble of the reference system. After the substitution of the configuration integral (Eq. 29) in the expression of the free energy

<sup>7</sup> The interested reader will find all useful adjuncts in the books either by J. P. Hansen and I. R. McDonald or by D. A. McQuarrie.

(cf. footnote 1), this one reads:

$$-\beta F = \ln \frac{Z_N^{(0)}(V, T)}{N! \Lambda^{3N}} + \ln \left\langle \exp \left[ -\beta \lambda U_1(\mathbf{r}^N) \right] \right\rangle_0. \quad (30)$$

The first term on the RHS stands for the free energy of the reference system, denoted  $(-\beta F_0)$ , and the second term represents the free energy of the perturbation:

$$-\beta F_1 = \ln \left\langle \exp \left[ -\beta \lambda U_1(\mathbf{r}^N) \right] \right\rangle_0. \quad (31)$$

Since the perturbation  $U_1(\mathbf{r}^N)$  is small,  $\exp(-\beta \lambda U_1)$  can be expanded in series, so that the statistical average  $\langle \exp[-\beta \lambda U_1(\mathbf{r}^N)] \rangle_0$ , calculated on the reference system, is expressed as:

$$\left\langle \exp \left[ -\beta \lambda U_1(\mathbf{r}^N) \right] \right\rangle_0 = 1 - \beta \lambda \langle U_1 \rangle_0 + \frac{1}{2!} \beta^2 \lambda^2 \langle U_1^2 \rangle_0 - \frac{1}{3!} \beta^3 \lambda^3 \langle U_1^3 \rangle_0 + \dots \quad (32)$$

Incidentally, we may note that the coefficients of  $\beta$  in the preceding expansion represent statistical moments in the strict sense. Given the shape of equation (32), it is still possible to write equation (31) by expanding  $\ln \langle \exp[-\beta \lambda U_1(\mathbf{r}^N)] \rangle_0$  in Taylor series. After simplifications, equation (31) reduces to:

$$\begin{aligned} \ln \left\langle \exp \left[ -\beta \lambda U_1(\mathbf{r}^N) \right] \right\rangle_0 &= -\beta \lambda \langle U_1 \rangle_0 + \frac{1}{2!} \beta^2 \lambda^2 \left[ \langle U_1^2 \rangle_0 - \langle U_1 \rangle_0^2 \right] \\ &\quad - \beta^3 \lambda^3 \left[ \frac{1}{3!} \langle U_1^3 \rangle_0 - \frac{1}{2} \langle U_1 \rangle_0 \langle U_1^2 \rangle_0 + \frac{1}{3} \langle U_1 \rangle_0^3 \right] + \beta^4 \lambda^4 [\dots] - \dots \end{aligned}$$

Now if we set:

$$c_1 = \langle U_1 \rangle_0, \quad (33)$$

$$c_2 = \frac{1}{2!} \left[ \langle U_1^2 \rangle_0 - \langle U_1 \rangle_0^2 \right], \quad (34)$$

$$c_3 = \frac{1}{3!} \left[ \langle U_1^3 \rangle_0 - 3 \langle U_1 \rangle_0 \langle U_1^2 \rangle_0 + 2 \langle U_1 \rangle_0^3 \right], \quad \text{etc.} \quad (35)$$

we find that:

$$\ln \left\langle \exp \left[ -\beta \lambda U_1(\mathbf{r}^N) \right] \right\rangle_0 = -\lambda \beta c_1 + \lambda^2 \beta^2 c_2 - \lambda^3 \beta^3 c_3 + \dots$$

The contribution of the perturbation (Eq. 31) to the free energy is then written in the compact form:

$$-\beta F_1 = \ln \left\langle \exp \left[ -\beta \lambda U_1(\mathbf{r}^N) \right] \right\rangle_0 = -\lambda \beta \sum_{n=1}^{\infty} c_n (-\lambda \beta)^{n-1}, \quad (36)$$

and the expression of the free energy  $F$  of the real system is found by substituting equation (36) into equation (30), as follows:

$$F = F_0 + F_1 = F_0 + \lambda c_1 - \lambda^2 \beta c_2 + \lambda^3 \beta^2 c_3 + \dots, \quad (37)$$

where the free energy of the real system is obtained by putting  $\lambda = 1$ . This expression of the free energy of liquids in power series expansion of  $\beta$  corresponds to the *high temperature approximation*.

### 5.2 Van der Waals equation

As a first application of thermodynamic perturbation method, we search the phenomenological van der Waals equation of state. In view of this, consider equation (37) at zero order in  $\beta$ . The simplest assumption to determine  $c_1$  is to admit that the total potential energy may be decomposed into a sum of pair potentials in the form:

$$U(\mathbf{r}^N) = U_0(\mathbf{r}^N) + U_1(\mathbf{r}^N) = \sum_i \sum_{j>i} u_0(r_{ij}) + \sum_i \sum_{j>i} u_1(r_{ij}).$$

Therefore, the free energy of the perturbation to zero order in  $\beta$  is given by equation (33), that is to say:

$$c_1 = \langle U_1 \rangle_0 = \frac{\int \int_{3N} \left[ \sum_i \sum_{j>i} u_1(r_{ij}) \right] \exp[-\beta U_0(\mathbf{r}^N)] d\mathbf{r}^N}{\int \int_{3N} \exp[-\beta U_0(\mathbf{r}^N)] d\mathbf{r}^N}.$$

To simplify the above relation, we proceed as for calculating the internal energy of liquids (Eq. 8) by revealing the pair correlation function of the reference system in the numerator. If we assume that the sum of pair potentials is composed of equivalent terms equal to  $\sum_i \sum_{j>i} u_1(r_{ij}) = \frac{N(N-1)}{2} u_1(r_{12})$ , the expression of  $c_1$  is simplified as:

$$c_1 = \frac{1}{Z_0(V, T)} \int \int \frac{N(N-1)}{2} u_1(r_{12}) \left\{ \int \int_{3(N-2)} \exp[-\beta U_0(\mathbf{r}^N)] d\mathbf{r}_3 \dots d\mathbf{r}_N \right\} d\mathbf{r}_1 d\mathbf{r}_2.$$

The integral in between the braces is then expressed as a function of the pair correlation function (cf. footnote 2), and  $c_1$  reduces to:

$$c_1 = \frac{\rho^2}{2} \int d\mathbf{r} \int u_1(r) g_0(r) d\mathbf{x}. \quad (38)$$

Yet, to find the equation of van der Waals we have to choose the hard-sphere system of diameter  $\sigma$ , as reference system, and suppose that the perturbation is a long-range potential, weakly attractive, the form of which is not useful to specify (Fig. 6a). Since one was unaware of the existence of the pair correlation function when the model was developed by van der Waals, it is reasonable to estimate  $g_0(r)$  by a function equal to zero within the particle, and to one at the outside. According to van der Waals, suppose further that the available volume per particle<sup>8</sup> is  $b = \frac{2}{3} \pi \sigma^3$  and the unoccupied volume is  $(V - Nb)$ .

With these simplifications in mind, the configuration integral and free energy of the reference system are respectively (cf. footnote 1):

$$Z_0(V, T) = \int \int_{3N} \exp[-\beta U_0(\mathbf{r}^N)] d\mathbf{r}^N = (V - Nb)^N,$$

and

$$F_0 = -k_B T \ln \left[ \frac{(V - Nb)^N}{N! \Lambda^{3N}} \right] = -Nk_B T \left[ \ln \frac{(V - Nb)}{N} - 3 \ln \Lambda + 1 \right].$$

<sup>8</sup> The parameter  $b$  introduced by van der Waals is the covolume. Its expression comes from the fact that if two particles are in contact, half of the excluded volume  $\frac{4}{3} \pi \sigma^3$  must be assigned to each particle (Fig. 6b).

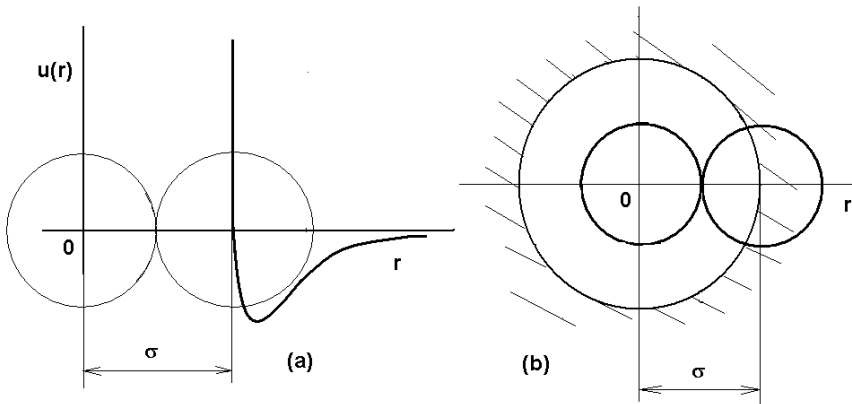


Fig. 6. Schematic representation of the pair potential by a hard-sphere potential plus a perturbation. (b) Definition of the covolume by the quantity  $b = \frac{1}{2} \left( \frac{4}{3} \pi \sigma^3 \right)$ .

As for the coefficient  $c_1$  (Eq. 38), it is simplified as:

$$c_1 = 2\pi\rho^2V \int_{\sigma}^{\infty} u_1(r)r^2 dr = -a\rho N, \quad (39)$$

with  $a = -2\pi \int_{\sigma}^{\infty} u_1(r)r^2 dr.$

Therefore, the expression of free energy (Eq. 37) corresponding to the model of van der Waals is:

$$F = F_0 + c_1 = -Nk_B T \left[ \ln \frac{(V - Nb)}{N} - 3 \ln \Lambda + 1 \right] - a\rho N,$$

and the van der Waals equation of state reduces to:

$$p = - \left( \frac{\partial F}{\partial V} \right)_T = \frac{Nk_B T}{V - Nb} - a \frac{N^2}{V^2}.$$

With  $b = a = 0$  in the previous equation, it is obvious that one recovers the equation of state of ideal gas. In return, if one wishes to improve the quality of the van der Waals equation of state, one may use the expression of the free energy (Eq. 25) and pair correlation function  $g_{HS}(r)$  of the hard-sphere system to calculate the value of the parameter  $a$  with equation (38). Another way to improve performance is to calculate the term  $c_2$ . Precisely what will be done in the next subsection.

### 5.3 Method of Barker and Henderson

To evaluate the mean values of the perturbation  $U_1(\mathbf{r}^N)$  in equations (34) and (35), Barker and Henderson (17) suggested to discretize the domain of interatomic distances into sufficiently small intervals  $(r_1, r_2), (r_2, r_3), \dots, (r_i, r_{i+1}), \dots$ , and assimilate the perturbing elemental potential in each interval by a constant. Assuming that the perturbing potential in the interval  $(r_i, r_{i+1})$  is  $u_1(r_i)$  and that the number of atoms subjected to this potential is  $N_i$ , the total perturbation can be written as the sum of elemental potentials:



$$U_1(\mathbf{r}^N) = \sum_i N_i u_1(r_i),$$

before substituting it in the configuration integral. The advantage of this method is to calculate the coefficients  $c_n$  and free energy (Eq. 37), not with the mean values of the perturbation, but with the fluctuation number of particles. Thus, each perturbing potential  $u_1(r_i)$  is constant in the interval which it belongs, so that we can write  $\langle U_1 \rangle_0^2 = \sum_i \sum_j \langle N_i \rangle_0 \langle N_j \rangle_0 u_1(r_i) u_1(r_j)$ . In view of this, the coefficient  $c_2$  defined by equation (34) is:

$$c_2 = \frac{1}{2!} [\langle U_1^2 \rangle_0 - \langle U_1 \rangle_0^2] = \frac{1}{2!} \sum_i \sum_j [\langle N_i N_j \rangle_0 - \langle N_i \rangle_0 \langle N_j \rangle_0] u_1(r_i) u_1(r_j).$$

With the *local compressibility* approximation (LC), where  $\rho$  and  $g_0$  depend on  $p$ , the expression of  $c_2$  obtained by Barker and Henderson according to the method described above is written:

$$c_2(LC) = \frac{\pi N \rho}{\beta} \left( \frac{\partial \rho}{\partial p} \right)_0 \frac{\partial}{\partial \rho} \left[ \int \rho u_1^2(r) g_0(r) r^2 dr \right]. \quad (40)$$

Incidentally, note the *macroscopic compressibility* approximation (MC), where only  $\rho$  is assumed to be dependent on  $p$ , has also been tested on a system made of the hard-sphere reference system and the square-well potential as perturbation. At low densities, the results of both approximations are comparable. But at intermediate densities, the results obtained with the LC approximation are in better agreement with the simulation results than the MC approximation. Note also that the coefficient  $c_3$  has been calculated by Mansoori and Canfield (18) with the macroscopic compressibility approximation.

At this stage of the presentation of the thermodynamic perturbation theory, we are in position to calculate the first terms of the development of the free energy  $F$  (Eq. 37), using the hard-sphere system as reference system. But there is not yet a criterion for choosing the diameter  $d$  of hard spheres. This point is important because all potentials have a repulsive part that must be replaced by a hard-sphere potential of diameter properly chosen. Decisive progress has been made to solve this problem in three separate ways followed, respectively, by Barker and Henderson (19), Mansoori and Canfield (20) and Week, Chandler and Andersen (21).

**Prescription of Barker and Henderson.** To choose the best reference system, that is to say, the optimal diameter of hard spheres, Barker and Henderson (19) proposed to replace the potential separation  $u(r) = u_0(r) + \lambda u_1(r)$ , where  $u_0(r)$  is the reference potential,  $u_1(r)$  the perturbation potential and  $\lambda$  the coupling parameter, by a more complicated separation associated with a potential  $v(r)$  whose the Boltzmann factor is:

$$\begin{aligned} \exp[-\beta v(r)] = & \left[ 1 - \Xi \left( d + \frac{r-d}{\alpha} - \sigma \right) \right] \exp \left[ -\beta u \left( d + \frac{r-d}{\alpha} \right) \right] \\ & + \Xi \left( d + \frac{r-d}{\alpha} - \sigma \right) + \Xi(r - \sigma) \{ \exp[-\beta \lambda u(r)] - 1 \}, \end{aligned} \quad (41)$$

where  $\Xi(x)$  is the Heaviside function, which is zero when  $x < 0$  and is worth one when  $x > 0$ . Note that here  $\sigma$  is the value of  $r$  at which the real potential  $u(r)$  vanishes and  $d$  is the hard-sphere diameter of the reference potential, to be determined. Moreover, the parameters  $\lambda$  and  $\alpha$  are coupling parameters that are 0 or 1. If one looks at equation (41) at the same time as figure(7a), it is seen that the function  $v(r)$  reduces to the real potential  $u(r)$  when  $\alpha = \lambda = 1$ ,

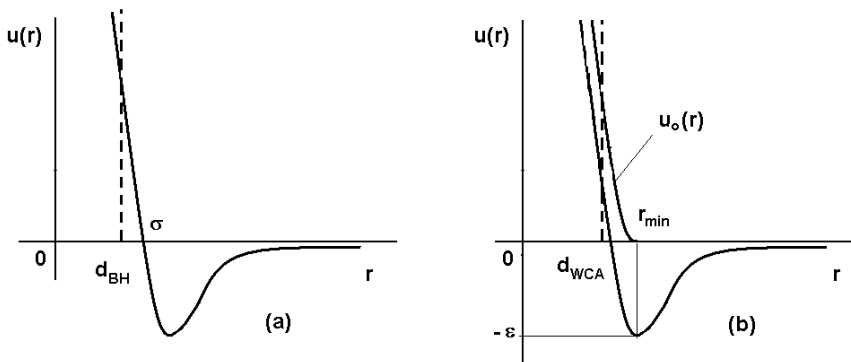


Fig. 7. Separation of the potential  $u(r)$  according to (a) the method of Barker and Henderson and (b) the method of Weeks, Chandler and Andersen.

and it behaves approximately as the hard-sphere potential of diameter  $d$  when  $\alpha \sim \lambda \sim 0$ . The substitution of equation (41) in the configuration integral (Eq. 29), followed by the related calculations not reproduced here, enable us to express the free energy  $F$  of the real system as a series expansion in powers of  $\alpha$  and  $\lambda$ , which makes the generalization of equation (28), namely:

$$F = F_{HS} + \lambda \left( \frac{\partial F}{\partial \lambda} \right) + \alpha \left( \frac{\partial F}{\partial \alpha} \right) + \frac{\lambda^2}{2} \left( \frac{\partial^2 F}{\partial \lambda^2} \right) + \frac{\alpha^2}{2} \left( \frac{\partial^2 F}{\partial \alpha^2} \right) + \dots \quad (42)$$

By comparing equations (37) and (42), we see that the first derivative  $\left( \frac{\partial F}{\partial \lambda} \right)$  coincides with  $c_1$  and the second derivative  $\frac{1}{2} \left( \frac{\partial^2 F}{\partial \lambda^2} \right)$  with  $(-\beta c_2)$ . Concerning the derivatives of  $F$  with respect to  $\alpha$ , they are complicated functions of the pair potential and the pair correlation function of the hard-sphere system. The first derivative  $\left( \frac{\partial F}{\partial \alpha} \right)$ , whose the explicit form given without proof, reads:

$$\left( \frac{\partial F}{\partial \alpha} \right) = -2\pi N\rho k_B T d^2 g^{HS}(d) \left[ d - \int_0^\sigma \{1 - \exp[-\beta u(r)]\} dr \right].$$

Since the Barker and Henderson prescription is based on the proposal to cancel the term  $\left( \frac{\partial F}{\partial \alpha} \right)$ , the criterion for choosing the hard-sphere diameter  $d$  is reduced to the following equation:

$$d = \int_0^\sigma \{1 - \exp[-\beta u(r)]\} dr. \quad (43)$$

In applying this criterion to the Lennard-Jones potential, it is seen that  $d$  depends on temperature but not on the density. Also, the calculations show that the terms of the expansion of  $F$  in  $\alpha^2$  and  $\alpha\lambda$  are negligible compared to the term in  $\lambda^2$ .

Therefore, using equation (38) to evaluate  $c_1$  and equation (40) to evaluate  $c_2$ , the expression of the free energy  $F$  of the real system (Eq. 42) is:

$$F = F_{HS} + 2\pi\rho N \int_d^\infty u_1(r)g_{HS}(\eta;r)r^2 dr - \pi N\rho \left(\frac{\partial\rho}{\partial p}\right)_{HS} \frac{\partial}{\partial\rho} \left[ \int_d^\infty \rho u_1^2(r)g_{HS}(\eta;r)r^2 dr \right], \quad (44)$$

where the first term on the RHS represents the free energy of the hard-sphere system (Eq. 25), and the partial derivative  $\left(\frac{\partial\rho}{\partial p}\right)_{HS}$  can be deduced from the Carnahan and Starling equation of state (Eq. 24). Recall that the pair correlation function of hard-sphere system,  $g_{HS}(\eta;r)$ , must be only determined numerically. It depends on the density  $\rho$  and diameter  $d$  via the packing fraction  $\eta (= \frac{\pi}{6}\rho d^3)$ . Since  $g_{HS}(\eta;r) = 0$  when  $r < d$ , either 0 or  $d$  can be used as lower limit of integration in equation (44).

#### 5.4 Prescription of Mansoori and Canfield.

An important consequence of the high temperature approximation to first order in  $\beta$  is to mark out the free energy of the real system by an upper limit that can not be exceeded, because the sum of the terms beyond  $c_1$  is always negative. The easiest way to proof this, is to consider the expression of the free energy (Eq. 30) and to write the perturbation  $U_1(\mathbf{r}^N)$  around its mean value  $\langle U_1 \rangle_0$  as:

$$U_1 = \langle U_1 \rangle_0 + \Delta U_1.$$

After replacing  $U_1$  in equation (30), we obtain:

$$-\beta F = -\beta F_0 - \beta \langle U_1 \rangle_0 + \ln \langle \exp[-\beta \Delta U_1]_0 \rangle. \quad (45)$$

However, considering the series expansion of an exponential, the above relation is transformed into the so-called *Gibbs-Bogoliubov inequality*<sup>9</sup>:

$$F \leq F_0 + \langle U_1 \rangle_0. \quad (46)$$

A thorough study of this inequality shows that it is always valid, and it is unnecessary to consider values of  $n$  greater than zero. Equation (46), at the base of the variational method, allows us to find the value of the parameter  $d$  that makes the free energy  $F$  *minimum*. If the reference system is that of hard spheres, the value of the free energy obtained with this value of  $d$  (or  $\eta = \frac{\pi}{6}\rho d^3$ ) is considered as the best estimate of the free energy of the real system. Its expression is:

$$F \leq F_{HS} + \frac{\rho N}{2} \int u_1(r)g_{HS}(\eta;r)d\mathbf{r}, \quad (47)$$

<sup>9</sup> The Mac-Laurin series of the exponential naturally leads to the inequality:

$$\exp(-\beta\Delta U_1) \geq \sum_{k=0}^{2n+1} \frac{(-\beta\Delta U_1)^k}{k!} \quad (n = 0, 1, 2, \dots).$$

For  $n = 0$ , the last term of equation (45) behaves as:

$$\ln \left\langle \sum_{k=0}^1 \frac{(-\beta\Delta U_1)^k}{k!} \right\rangle_0 = \ln \langle 1 - \beta\Delta U_1 \rangle_0 \simeq -\beta \langle \Delta U_1 \rangle_0 = 0,$$

since the mean value of the deviation,  $\langle \Delta U_1 \rangle_0$ , is zero.

where  $F_0 = F_{HS}$  is given by equation (25) and  $\langle U_1 \rangle_0 = c_1$  by equation (38). Since  $g_{HS}(\eta; r) = 0$  when  $r < d$  and  $u_1(r) = u(r) - u_{HS}(r) = u(r)$  when  $r \geq d$ , equation (47) can be written as:

$$F \leq F_{HS} + 2\pi\rho N \int_0^\infty u(r)g_{HS}(\eta; r)r^2 dr. \quad (48)$$

Practically, we vary the value of  $d$  (or  $\eta$ ) until the result of the integral is minimum. And the value of  $F$  thus obtained is the best estimate of the free energy of the real system, in the sense of the variational method (20).

### 5.5 Prescription of Weeks, Chandler and Andersen.

At the same time that the perturbation theory was developing, Weeks, Chandler and Andersen (21) formulated another prescription for finding the hard-sphere diameter  $d$ . Its originality lies in the idea that a particle in the liquid is less sensitive to the sign of the potential  $u(r)$  than to the sign of the strength, that is to say, to the derivative of the potential  $\left(-\frac{\partial u}{\partial r}\right)_{\rho, T}$ . That is why the authors proposed to separate the real potential into a purely repulsive contribution and a purely attractive perturbation. This separation, shown in figure (7b), is defined by the relation  $u(r) = u_0(r) + u_1(r)$  with:

$$u_0(r) = \begin{cases} u(r) + \varepsilon & \text{if } r < r_{\min} \\ 0 & \text{if } r \geq r_{\min} \end{cases}$$

$$u_1(r) = \begin{cases} -\varepsilon & \text{if } r < r_{\min} \\ u(r) & \text{if } r \geq r_{\min}, \end{cases}$$

where  $\varepsilon$  is the depth of the potential well, i. e. the value of  $u(r_{\min}) = -\varepsilon$ .

To follow the same sketch that for the Barker and Henderson prescription, define the potential  $v(r)$  by the Boltzmann factor:

$$\exp[-\beta v(r)] = \exp[-\beta u_{HS}(d)] + \alpha \{ \exp[-\beta(u_0(r) + \lambda u_1(r))] - \exp[-\beta u_{HS}(d)] \}. \quad (49)$$

However, from the simultaneous observation of equation (49) and figure (7b) one remarks that the potential  $v(r)$  reduces to the real potential  $u(r)$  when  $\alpha = \lambda = 1$ , and behaves like the hard-sphere potential of diameter  $d$  when  $\alpha = 0$ .

The substitution of equation (49) in the configuration integral, and subsequent calculations, enables us to express the free energy  $F$  of the system as a series expansion in powers of  $\alpha$  and  $\lambda$  identical to that of equation (42). The first term of this expansion,  $F_{HS}$ , is the free energy of the hard-sphere system. As for the first derivative of  $F$  with respect to  $\alpha$ , it is written in this case:

$$\left(\frac{\partial F}{\partial \alpha}\right) = -\frac{2\pi N\rho}{\beta} \int_0^\infty \{ \exp[-\beta u_0(r)] - \exp[-\beta u_{HS}(d)] \} y_{HS}(\eta; r)r^2 dr, \quad (50)$$

where  $y_{HS}(\eta; r)$  is the *cavity function* of great importance in the microscopic theory of liquids<sup>10</sup>. It is defined by means of the pair correlation function  $g_{HS}(\eta; r)$  and the Boltzmann factor  $\exp[-\beta u_{HS}(d)]$  of the hard-sphere potential as:

$$y_{HS}(\eta; r) = \exp[\beta u_{HS}(d)] g_{HS}(\eta; r). \quad (51)$$

<sup>10</sup> The cavity function  $y_{HS}(\eta; r)$  does not exist in analytical form. It must be calculated at all reduced densities ( $\rho d^3$ ) by solving an integro-differential equation.

It should be stressed that the cavity function depends only weakly on the potential, this is why Weeks, Chander and Anderson (WCA) suggested to use the same cavity function  $y_{HS}(\eta; r)$  for all potentials, and to express the pair correlation functions of each potential as a function of  $y_{HS}(\eta; r)$ . As a result, the pair correlation function  $g_0(r)$  related to the repulsive contribution  $u_0(r)$  of the potential can be approximately written as:

$$g_0(r) \simeq \exp[-\beta u_0(r)] y_{HS}(\eta; r). \tag{52}$$

The suggestion of WCA for choosing the hard-sphere diameter  $d$  is to cancel  $\left(\frac{\partial F}{\partial \alpha}\right)$ , which amounts to solving the nonlinear equation given by equation (50), i. e.:

$$\int_0^\infty \{\exp[-\beta u_0(r)] - \exp[-\beta u_{HS}(d)]\} y_{HS}(\eta; r) r^2 dr = 0. \tag{53}$$

It is interesting to note that equation (53) has a precise physical meaning that appears by writing it in terms of the pair correlation functions  $g_{HS}(\eta; r)$  and  $g_0(r)$  drawn, respectively, from equations (51) and (52). It reads:

$$\int_0^\infty [g_0(r) - 1] r^2 dr = \int_0^\infty [g_{HS}(\eta; r) - 1] r^2 dr.$$

According to equation (7), the previous expression is equivalent to the equality  $S_0(0) = S_{HS}(0)$ , meaning the equality between the isothermal compressibility of the repulsive potential  $u_0(r)$  and that of the hard-sphere potential  $u_{HS}(d)$ .

Ultimately, the expression of free energy (Eq. 28) of the real system is the sum of  $F_{HS}$ , calculated with the value of  $d$  issued from equation (53), and the term  $\left(\frac{\partial F}{\partial \lambda}\right) = c_1$ , calculated with equation (38) in using the pair correlation function  $g_0(r) \simeq \exp[-\beta u_0(r)] y_{HS}(\eta; r)$ , i. e.:

$$F = F_{HS} + 2\pi\rho N \int_0^\infty u_1(r) \{\exp[-\beta u_0(r)] y_{HS}(\eta; r)\} r^2 dr. \tag{54}$$

Note that the value of  $d$  obtained with the WCA prescription (*blip function* method) is significantly larger than that obtained with the Barker and Henderson (BH) prescription. By the fact that  $y_{HS}(\eta; r)$  depends on density, the value of  $d$  calculated with equation (53) depends on temperature and density, while that calculated with equation (43) depends only on temperature.

In order to compare the relative merits of equations (44) and (54), simply note that the WCA prescription shows that not only the terms of the expansion of  $F$  in  $\alpha^2$  and in  $\alpha\lambda$  are negligible, but also the term in  $\lambda^2$ . This means that the WCA treatment is a theory of first order in  $\lambda$ , while the BH one is a theory of second order in  $\lambda$ . In addition, the BH treatment is coherent as  $F_{HS}$ ,  $c_1$  and  $c_2$  are calculated with the same reference system, whereas the WCA treatment is not coherent because it uses the hard-sphere system to calculate  $F_{HS}$  and the repulsive potential  $u_0(r)$  to calculate  $c_1$ , via the pair correlation function  $g_0(r)$ . In contrast, the WCA treatment does not require the calculation of  $c_2$  and improves the convergence of calculations. As an additional advantage, the WCA treatment can be used to predict the structure factor of simple liquids (22).

### 5.6 Application to the Yukawa attractive potential

As an application of the perturbation theory, consider the hard-core attractive Yukawa potential (HCY). This potential consists of the hard-sphere potential  $u_{HS}(d)$  and the attractive perturbation  $u_1(r)$ . Its traditional analytical form is:

$$u(r) = \begin{cases} \infty & r < d \\ -(\varepsilon d/r) \exp[-\lambda(r/d - 1)] & r \geq d, \end{cases} \quad (55)$$

where  $d$  is the diameter of hard spheres,  $\varepsilon$  the potential value at  $r = d$  and  $\lambda$  a parameter that measures the rate of exponential decay. The shape of this potential is portrayed in figure (8a)

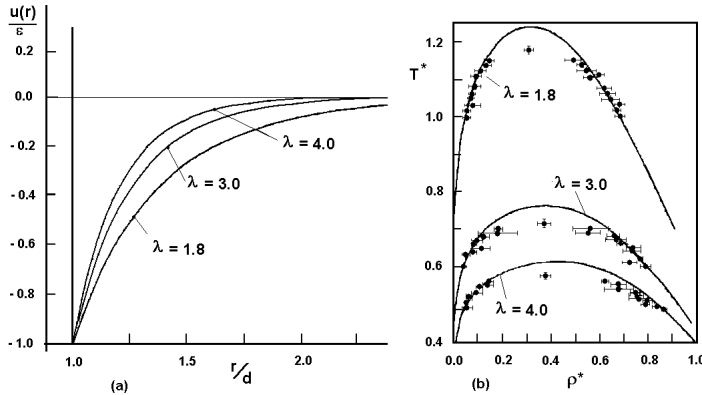


Fig. 8. (a) Representation of the HCY potential for three values of  $\lambda$  (1.8, 3.0, 4.0). (b) Phase diagram of the HCY potential for the same three values of  $\lambda$  (1.8, 3.0, 4.0).

for three different values of  $\lambda$ . For guidance, note that a value of  $\lambda \sim 1.8$  predicts the thermodynamic properties of liquid rare gases quite so well as the Lennard-Jones potential. By contrast, a value of  $\lambda \sim 8$  enables us to obtain the thermodynamic properties of colloidal suspensions or globular proteins. This wide possible range of  $\lambda$  values explains why the HCY potential has been used in many applications and has been the subject of numerous theoretical studies after its analytical solution was obtained by Waisman (2). Without going into details of the resolution of the problem that involves the microscopic theory of liquids, indicate that the solution reduces to determining a fundamental parameter  $\Gamma$  as the root of the quartic equation (23):

$$\Gamma(1 + \lambda\Gamma)(1 + \psi\Gamma)^2 + \beta\varepsilon w = 0, \quad (56)$$

where  $\varepsilon$  and  $\lambda$  are the two parameters of the HCY potential, and  $w$  and  $\psi$  two additional parameters explicitly depending on  $\lambda$  and  $\eta$  by the following relations:

$$w = \frac{6\eta}{\phi_0^2},$$

$$\psi = \lambda^2(1 - \eta)^2 \frac{[1 - \exp(-\lambda)]}{L(\lambda) \exp(-\lambda) + S(\lambda)}$$

$$- 12\eta(1 - \eta) \frac{[1 - \lambda/2 - (1 + \lambda/2) \exp(-\lambda)]}{L(\lambda) \exp(-\lambda) + S(\lambda)},$$

with:

$$\begin{aligned}\phi_0 &= \frac{L(\lambda) \exp(-\lambda) + S(\lambda)}{\lambda^3(1-\eta)^2}, \\ L(\lambda) &= 12\eta[1 + 2\eta + (1 + \eta/2)\lambda], \\ S(\lambda) &= (1-\eta)^2\lambda^3 + 6\eta(1-\eta)\lambda^2 + 18\eta^2\lambda - 12\eta(1+2\eta).\end{aligned}$$

By expressing  $\Gamma$  as a function of  $\beta$ , Henderson *et al.* (3) managed to write the free energy  $F$  of the system according to the series expansion in powers of  $\beta$ :

$$F = F_{HS} - \frac{Nk_B T}{2} \sum_{n=1}^{\infty} \frac{v_n}{n} (\beta\varepsilon)^n, \quad (57)$$

where  $F_{HS}$  is the free energy of the hard-sphere system calculated, for example, with equation (25), and where the first five terms  $v_n$  have expressions derived by the authors<sup>11</sup>.

At this stage, we see that the free energy of the HCY potential can be approximated for all values of the triplet  $(\eta, \varepsilon, \lambda)$ , using the analytical expressions of the five coefficients  $v_n$  given in footnote (11). From equation (57), we can also deduce the internal energy (Eq. 9), excess pressure (Eq. 12), chemical potential (Eq. 19) and entropy (Eq. 20) of the HCY system.

The HCY equation of state, which is based on the perturbation theory and expressed in terms of the relevant features of the potential, is a very handy tool for investigating the thermodynamics of systems governed by an effective hard-sphere interaction plus an attractive tail. Alternatively, the HCY system could be used vicariously to approximate any available interatomic potential for real fluids.

The reduced phase diagrams ( $T^* = \frac{k_B T}{\varepsilon}$  versus  $\rho^* = \rho\sigma^3 = \frac{6\eta}{\pi}$ ) predicted by this equation of state, for values of  $\lambda = 1.8, 3$  and  $4$ , are displayed in figure (8b), together with simulation data. A rapid glance at figure (8b) indicates that the critical temperature  $T_C^*$  predicted by the perturbation theory is noticeably greater than that predicted by the simulation. The reason is that the HCY equation of state involves a truncated series. But, also, we can speculate that the

<sup>11</sup> The first terms of the series expansion are:

$$\begin{aligned}v_0 &= 0, \\ v_1 &= \frac{2\alpha_0}{\phi_0}, \\ v_2 &= \frac{2w(1 - \alpha_1 + \alpha_0\psi)}{\lambda\phi_0}, \\ v_3 &= \frac{2w^2(1 - \alpha_1 + \alpha_0\psi)(1 + 3\lambda\psi)}{\lambda^3\phi_0}, \\ v_4 &= \frac{4w^3(1 - \alpha_1 + \alpha_0\psi)(1 + 4\lambda\psi + 6\lambda^2\psi^2)}{\lambda^5\phi_0}, \\ v_5 &= \frac{10w^4(1 - \alpha_1 + \alpha_0\psi)(1 + 5\lambda\psi + 11\lambda^2\psi^2 + 11\lambda^3\psi^3)}{\lambda^7\phi_0},\end{aligned}$$

with:

$$\alpha_0 = \frac{L(\lambda)}{\lambda^2(1-\eta)^2} \quad ; \quad \alpha_1 = \frac{12\eta(1+\lambda/2)}{\lambda^2(1-\eta)} \quad \text{and} \quad (1 - \alpha_1 + \alpha_0\psi)\phi_0 = 1.$$

correlation functions, inherent in the calculations of the pressure and the chemical potential, do not represent correctly the growing correlation lengths in approaching the critical point. Looking at the evolution of the binodal lines as a function of the rate of decay  $\lambda$  of the Yukawa potential, we remark that higher the critical temperature is, lower  $\lambda$  is. At the same time, the domain below the binodal line shrinks. In contrast, when  $\lambda$  increases (i. e., the attraction range of the potential becomes shorter), the critical temperature decreases, and the liquid and gaseous phases become indistinguishable. For the hard-sphere potential, as a border case ( $\lambda \rightarrow \infty$ ), there is no longer gas-liquid phase transition. On the other hand, one can see that the phase diagrams obtained with the perturbation theory agree with simulation data more favorably for the vapor branch than for the liquid branch. This is not surprising in the extent that the perturbation theory works better for low densities. Lastly, it should be mentioned that the structure and thermodynamic properties of the HCY potential have been extensively studied for the two last decades, as much by computer simulations and integral equation theory as by means of perturbation theory. Note that many other studies of the HCY potential with these various methods are available in literature (24).

## 6. Concluding remarks

In the foregoing, we started with a brief discussion of the liquid state, compared with the gaseous and solid states. Among other things, we mentioned that one of the most successes of the microscopic theory of liquids has been to emphasize the importance of the *pair potential*  $u(r)$  to describe a wide variety of physical properties of liquids. Then, we learned a rudimentary knowledge of the structure of liquids. The latter is essentially described by the static *structure factor*  $S(q)$ , which can be measured directly by elastic scattering of neutrons or X-rays. But  $S(q)$ , or more precisely the *pair correlation function*  $g(r)$ , can also be determined by numerical simulations (known as virtual experiments) and with models of the microscopic theory of liquids once the nature of the pair potential  $u(r)$  is known. The comparison between simulation results and those of analytical models essentially allows us to test the models, whereas the comparison of simulation results with experimental results is the ultimate test to judge the efficiency of the pair potential. Subsequently, we described in minute detail the calculations of the thermodynamic properties in terms of the pair potential and pair correlation function. In particular, with the pressure  $p$  and the chemical potential  $\mu$  to apply equilibrium conditions, it has been shown that we are in a position to determine theoretical estimates for the liquid-vapor coexistence curve. Finally, we got down to basics of the thermodynamic perturbation theory, before presenting the liquid-vapor coexistence curve for the HCY fluid.

In a 1990 time warp, it became apparent that the usefulness of thermodynamic perturbation theory was on the decline compared to the more powerful simulation methods. This was primarily due to the rapid increase in the power of computers. At the same time, the integral equation theory enjoyed renewed popularity with the extensively employed concept of thermodynamic consistency (25), but this aspect goes beyond the scope of this short review. Incidentally, we recall that the thermodynamic consistency consists to adjusting the isothermal compressibility obtained by two different routes. Nevertheless, the thermodynamic perturbation theory remains undoubtedly the most tractable approach to predict the thermodynamic properties of liquids. Let us just quote few articles containing investigations of the HCY fluid with the integral equation theory (26).

As one would expect, the thermodynamic perturbation theory described previously is also important for mixtures. The thermodynamic properties established for studying pure liquids can be applied to binary mixtures, with only few modifications. Specifically, analytic



expressions for the hard-sphere free energy  $F_{HS}(d_1, d_2)$  and subsequent thermodynamic quantities are readily generalized for a mixture of hard spheres of different diameters  $d_1$  and  $d_2$ . Therefore, if one estimates that a real binary mixture behaves like binary hard-sphere fluid, the variational method consists of minimizing  $F_{HS}(d_1, d_2)$  versus the two hard-sphere diameters, and by taking the resulting minimum upper bound as an approximation to the free energy. The variational method is not limited neither to hard-sphere reference system nor to a specific fluid. It should be stressed that systems like pure liquid metals composed of ions embedded in an electron gas can not be treated as a binary mixture, in the strict sense of the word, but must be suitably reduced to a one-component system of pseudoions. In contrast, a liquid metal made up of two species of pseudoions forms a binary mixture, for which the approach originally developed for simple liquids can be applied. What is it that determines whether or not two metals will mix to form an alloy is a crucial issue to be answered by thermodynamic perturbation theory (27).

I would like to express my acknowledgement to Jean-Marc Bomont for its stimulating discussions.

## 7. References

- [1] M. H. J. Hagen and D. Frenkel, *J. Chem. Phys.* 101, 4093 (1994).
- [2] E. Waisman, *Mol. Phys.* 25, 45 (1973).
- [3] D. Henderson, G. Stell and E. Waisman, *J. Chem. Phys.* 62, 4247 (1975). D. Henderson, L. Blum and J. P. Noworyta, *J. Chem. Phys.* 102, 4973 (1995).
- [4] See, for instance, the book of J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, Acad. Press., London (2006).
- [5] W. G. Hoover and F. H. Ree, *J. Chem. Phys.* 47, 4873 (1967); 49, 3609 (1968).
- [6] E. Chacon, M. Reinaldo-Falagan, E. Velasco and P. Tarazona, *Phys. Rev. Lett.* 87, 166101 (2001). D. Li and S. A. Rice, *J. Phys. Chem. B.* 108, 19640 (2004). J. M. Bomont and J. L. Bretonnet, *J. Chem. Phys.* 124, 054504 (2006).
- [7] N. Jakse and J. L. Bretonnet, *J. Phys.: Condens. Matter* 15, S3455 (2003).
- [8] J. G. Kirkwood, *J. Chem. Phys.* 3, 300 (1935). J. G. Kirkwood and E. Monroe, *J. Chem. Phys.* 9, 514 (1941). See also D. A. McQuarrie, *Statistical Mechanics*, Harper and Row, New-York (1976), p. 263.
- [9] R. E. Nettleton and M. S. Green, *J. Chem. Phys.* 29, 1365 (1958). H. J. Raveché, *J. Chem. Phys.* 55, 2242 (1971).
- [10] B. J. Alder and T. E. Wainwright, *Phys. Rev.* 127, 359 (1962); F. H. Ree and W. G. Hoover, *J. Chem. Phys.* 40, 939 (1964).
- [11] N. F. Carnahan and K. E. Starling, *J. Chem. Phys.* 51, 635 (1969).
- [12] D. A. McQuarrie, *Statistical Mechanics*, Harper and Row, New York (1976).
- [13] A. J. Greenfield, J. Wellendorf and N. Wiser, *Phys. Rev. A* 4, 1607 (1971).
- [14] N. W. Ashcroft and J. Lekner, *Phys. Rev.* 156, 83 (1966).
- [15] L. Verlet, *Phys. Rev.* 165, 201 (1968).
- [16] R. Zwanzig, *J. Chem. Phys.* 22, 1420 (1954).
- [17] J. A. Barker and D. Henderson, *J. Chem. Phys.* 47, 2856 (1967).
- [18] G. A. Mansoori and F. B. Canfield, *J. Chem. Phys.* 51, 4958 (1969).
- [19] J. A. Barker and D. Henderson, *Ann. Rev. Phys. Chem.* 23, 439 (1972).
- [20] G. A. Mansoori and F. B. Canfield, *J. Chem. Phys.* 51, 4967 (1969); 51 5295 (1969); 53, 1618 (1970).
- [21] J. D. Weeks, D. Chandler and H. C. Andersen, *J. Chem. Phys.* 54, 5237 (1971); 55, 5422 (1971).

- [22] J. L. Bretonnet and C. Regnaut, *Phys. Rev. B* 31, 5071 (1985).
- [23] M. Ginoza, *Mol. Phys.* 71, 145 (1990).
- [24] W. H. Shih and D. Stroud, *J. Chem. Phys.* 79, 6254 (1983). D. M. Duh and L. Mier-Y-Teran, *Mol. Phys.* 90, 373 (1997). S. Zhou, *Phys. Rev. E* 74, 031119 (2006). J. Torres-Arenas, L. A. Cervantes, A. L. Benavides, G. A. Chapela and F. del Rio, *J. Chem. Phys.* 132, 034501 (2010).
- [25] J. M. Bomont, *Advances in Chemical Physics*, 139, 1 (2008).
- [26] E. Lomba and N. G Almarza, *J. Chem. Phys.* 100, 8367 (1994). C. Caccamo, G. Pellicane, D. Costa, D. Pini and G. Stell, *Phys. Rev. E* 60, 5533 (1999). J. M. Caillol, F. Lo Verso, E. Scholl-Paschinger and J. J. Weis, *Mol. Phys.* 105, 1813 (2007). A. Reiner and J. S. Hoye, *J. Chem. Phys.* 128, 114507 (2008). E. B. El Mendoub, J. F. Wax and N. Jakse, *J. Chem. Phys.* 132, 164503 (2010).
- [27] M. Shimoji, *Liquid Metals*, Acad. Press, London (1977). D. Stroud and N. W. Ashcroft, *Solid State Physics*, 33, 1, 1978.

# Probing Solution Thermodynamics by Microcalorimetry

Gregory M. K. Poon

*Department of Pharmaceutical Sciences, Washington State University  
U.S.A.*

## 1. Introduction

Solution microcalorimetry has entrenched itself as a major technique in laboratories concerned with studying the thermodynamics of chemical systems. Recent developments in the calorimeter marketplace will undoubtedly continue to popularize microcalorimeters as mainstream instruments. The technology of microcalorimetry has in turn benefited from this trend in terms of enhanced sensitivity, signal stability, physical footprint and user-friendliness. As the popularity of solution microcalorimeters has grown, so has an impressive body of literature on various aspects of microcalorimetry, particularly with respect to biophysical characterizations. The focus of this chapter is on experimental and analytical aspects of solution microcalorimetry that are novel or represent potential pitfalls. It is hoped that this information will aid bench scientists in the formulation and numerical analysis of models that describe their particular experimental systems. This is a valuable skill, since frustrations often arise from uninformed reliance on turnkey software that accompany contemporary instruments. This chapter will cover both differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC). It targets physical chemists, biochemists, and chemical engineers who have some experience in calorimetric techniques as well as nonlinear regression (least-square analysis), and are interested in quantitative thermodynamic characterizations of noncovalent interactions in solution.

## 2. Differential scanning calorimetry

DSC measures the heat capacity ( $C_p$ ) of a sample as the instrument “scans” up or down in temperature. For reversible systems, direct interpretation of the data in terms of thermodynamic parameters requires that chemical equilibrium be re-established much more rapidly than the scan rate. This can be verified by comparing data obtained at different scan rates. For transitions involving a change in molecularity (*e.g.*, self-association/dissociation, ligand binding/unbinding), reversibility can also be confirmed by the lack of hysteresis between heating and cooling experiments. The optimal scan rate is ultimately a compromise between the requirement for reversibility and the desire for reasonable throughput; typically this falls between 0.5 to 1.0 °C/min for most systems in dilute aqueous solutions.

### 2.1 Experimental conditions for DSC

The observed or apparent thermodynamics of solution systems generally include linked contributions from other solutes in addition to the species of interest. They include buffers,

salts, neutral cosolutes, and cosolvents. Of these, the choice of buffer, or any ionizable species in general, must take into account the change in  $pK_a$  with respect to temperature *i.e.*, the enthalpy of ionization ( $\Delta H_{ion}$ ). Unless a buffer's  $\Delta H_{ion}$  is negligibly small, its  $pK_a$  will exhibit a temperature dependence, leading to a change in pH of the solution upon heating or cooling. Failure to take this fact into account may introduce significant artifacts into the observed melting behavior. Such changes in pH represent a different issue from any coupled ionization enthalpy arising from the release or uptake of protons associated with the transition of interest.

The direction and extent of the temperature of pH for a buffered solution depends on the sign and magnitude of  $\Delta H_{ion}$  as well as the concentration of the buffering species. Consider the ionization of a buffer  $A^z$  in the direction of deprotonation to produce its conjugate base  $B^{z-1}$ :



According to the van't Hoff equation,

$$\frac{dK_{ion}}{d(1/T)} = \frac{d10^{-pK_a}}{d(1/T)} = -\frac{\Delta H_{ion}}{R}, \quad (2)$$

where  $R$  is the gas constant,  $T$  is absolute temperature, and

$$K_{ion} = \frac{[H^+][B^{z-1}]}{[A^z]}. \quad (3)$$

Thus, for a buffer with a positive (endothermic)  $\Delta H_{ion}$ , its ionization equilibrium shifts towards deprotonation as temperature increases, leading to a drop in pH. Conversely, the pH of a solution buffered by an exothermic buffer rises with increasing pH.

Table 1 lists several common buffers for aqueous solutions (King 1969; Disteché 1972; Lo Surdo et al. 1979; Kitamura and Itoh 1987; Goldberg et al. 2002). As a group, substituted ammonium compounds exhibit substantial positive values of  $\Delta H_{ion}$ , making them poor choices for DSC experiments. These compounds include the so-called "Good buffers" (Good et al. 1966) that are prevalent in biochemistry. Among these, Tris, is a particular offender: the pH of a 25 mM solution initially buffered at pH 9.0 drops by more than one pH unit from 0 to 37°C (Poon et al. 2002). In contrast, the ionization of carboxylic acids and their analogues is far less sensitive to temperature. Generally, buffers based on acetate, cacodylate, and phosphate, for example, are preferred choices for DSC experiments.

Another important note relates to polyprotic species such as phosphates, citrates, and borates, whose  $pK_a$  also depends markedly on ionic strength. This relationship is quantitatively given by the volume changes of ionization ( $\Delta V^\circ$ ) and interpreted in terms of electrostriction of solvent water molecules. Thus, the addition of salts such as NaCl or guanidinium salts (the latter commonly used to denature proteins) will systematically reduce the pH of a solution buffered by polyprotic acids. The pH of a 0.1 M phosphate buffer at pH 7.2, for example, can fall by 0.5 pH unit upon addition of 0.5 M of NaCl. Molar concentrations of guanidinium hydrochloride will produce an even greater drop. On the other hand, inorganic cosolvents have the opposite effect by affecting the solution dielectric. Of course, once the pH of these buffers is adjusted to a value that is compatible with the apparent  $pK_a$ , it will be stable with respect to temperature. As seen in Table 1, it is generally the case that a buffer is either sensitive to temperature or ionic strength in aqueous solution.

Buffer	pK <sub>a</sub>	$\Delta H^\circ$ , kJ mol <sup>-1</sup>	$\Delta C_p^\circ$ , J K <sup>-1</sup> mol <sup>-1</sup>	$\Delta V^\circ$ , mL mol <sup>-1</sup> <sup>a</sup>
Acetate	4.756	-0.41	-142	-10.6
Bicine	8.334	26.34	0	-2.0
Bis-tris	6.484	28.4	27	3.1
Cacodylate	6.28	-3	-86	-13.3
	3.128	4.07	-131	-10.7
Citrate	4.761	2.23	-178	-12.3
	6.396	-3.38	-254	-22.3
Glycine	2.351	4	-139	-6.8
	9.780	44.2	-57	
HEPES	7.564	20.4	47	4.8
Imidazole	6.993	36.64	-9	1.8
MES	6.27	14.8	5	3.9
MOPS	7.184	21.1	25	4.7
	2.148	-8	-141	-16.3
Phosphate	7.198	3.6	-230	-25.9
	12.35	16	-242	-36.0
Succinate	4.207	3.0	-121	
	5.636	-0.5	-217	
Tris	8.072	47.45	-142	4.3

<sup>a</sup> Ionization volume at atmospheric pressure at infinite dilution.

Table 1. Thermodynamic properties of ionization by common aqueous buffers

## 2.2 Analysis of DSC data

A complete DSC experiment consists of matched scans of a sample and a sample-free reference solution. Blank-subtracted data can be empirically analyzed to obtain model-independent thermodynamic parameters. The difference between pre- and post-transition baselines gives the change in heat capacity,  $\Delta C_p$ . After subtracting a suitable baseline across the transition range, the arithmetic integration of the  $C_p$  vs.  $T$  trace yields the so-called calorimetric enthalpy:

$$\Delta H_{\text{cal}} = \int_{T_i}^{T_f} C_p dT . \quad (4)$$

$\Delta H_{\text{cal}}$  is the value of the transition enthalpy at the transition temperature,  $T_m$ . The entropy at  $T_m$  is

$$\Delta S(T_m) = \frac{\Delta H_{\text{cal}}}{T_m} . \quad (5)$$

Thus, a single DSC experiment yields the complete thermodynamics of a transition. Model-free determination of thermodynamics, including the direct measurement of  $\Delta C_p$ , is a unique feature of DSC not possible with optical techniques (such as absorption and fluorescence spectroscopy) which access  $\Delta H$  via the van't Hoff equation. However, model

fitting by least-square analysis can extract considerably more useful information and facilitate quantitative hypothesis testing.

In general, the reference-subtracted DSC data represent the heat capacity of the initial state 0 ( $C_{p,0}$ ) as well as the excess heat capacity function,  $\langle \Delta C_p(T) \rangle$ :

$$C_p(T) = C_{p,0} + \langle \Delta C_p(T) \rangle. \quad (6)$$

Consider a general model in which the sample undergoes a transition from initial state 0 through intermediates 1, 2, ...,  $i$  to the final state  $n$ . (One can readily envisage extensions of this model in which a heterotypic complex dissociates into subunits which then go on to further, independent transitions.) The excess heat capacity function is (Privalov and Potekhin 1986)

$$\begin{aligned} \langle \Delta C_p(T) \rangle &= \frac{d\langle \Delta H(T) \rangle}{dT} = \frac{d}{dT} \left[ \sum_{i=1}^n \Delta H_i(T) \alpha_i(T) \right] \\ &= \sum_{i=1}^n \Delta H_i \frac{d\alpha_i(T)}{dT} + \sum_{i=1}^n \alpha_i(T) \Delta C_{p,i} \quad , \\ &= \sum_{i=1}^n \langle \delta C_{p,i}^{\text{tr}}(T) \rangle + \sum_{i=1}^n \langle \delta C_{p,i}^{\text{int}}(T) \rangle \end{aligned} \quad (7)$$

where  $\Delta H_i$  and  $\Delta C_{p,i}$  are the enthalpy and heat capacity change, respectively, from state 0 to state  $i$ , and  $\alpha_i(T)$  is the fractional conversion at state  $i$ .  $\langle \delta C_{p,i}^{\text{tr}}(T) \rangle$  and  $\langle \delta C_{p,i}^{\text{int}}(T) \rangle$  are called "transition" and "intrinsic" heat capacities, respectively. The intrinsic heat capacity, which represents the summed heat capacities of the various species present at  $T$ , is the baseline function of the observed DSC trace (Figure 1). Some analytical protocols invite the user to

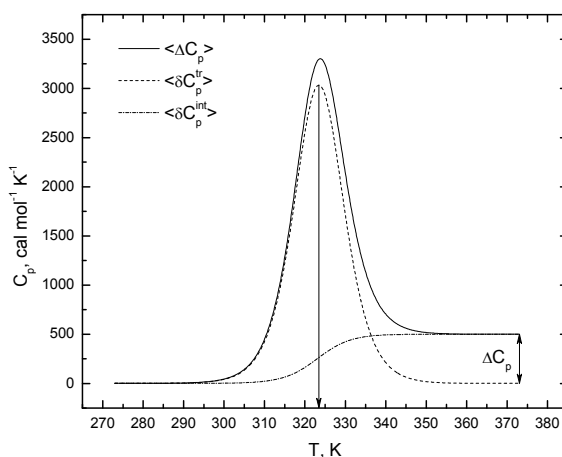


Fig. 1. An excess heat capacity function and its constituent transition and intrinsic heat capacities. Integration of the transition heat capacity,  $\langle \delta C_{p,i}^{\text{tr}}(T) \rangle$ , gives  $\Delta H_{\text{cal}}$  at the transition temperature (50°C here).

perform manual baseline subtraction before fitting a excess heat capacity function. This is intended to eliminate  $\langle \delta C_{p,i}^{\text{int}}(T) \rangle$  from the fitting function. In the transition region, manual baseline subtraction requires either heuristic or semi-empirical criteria to connect the pre- and post-transitional states. This is both unnecessary and questionable practice, since manual editing may (and probably do) bias the data. The most appropriate approach is to fit both the excess and intrinsic heat capacities directly according to Eq (7). Since both terms are functions of  $\alpha_i(T)$ , the fitted baseline will objectively track the progress of each transition. Note that  $C_{p,0}$  and  $\Delta C_{p,i}$  are taken to be constants in Eqs (6) and (7) since their temperature dependence is generally weak over the experimental range. They can be formulated, if desired, as polynomials to define nonlinear baselines. Care must be taken, however, to ensure that such curvature is not masking some low-enthalpy transition such as conformation changes of proteins in the native state (Privalov and Dragan 2007).

### 2.2.1 Formulation of DSC models

The principal task in formulating DSC models is deriving expressions for  $\alpha_i(T)$  from the relevant equilibrium expressions and equations of state. Implicit in this task is the computation of  $K_i$  from its corresponding thermodynamic parameters. This in turn requires the choice of a reference temperature, the most convenient of which is the characteristic temperature  $T^\circ$  at which  $\Delta G(T^\circ) = 0$ :

$$\Delta G(T) = \Delta H(T^\circ) \left( 1 - \frac{T}{T^\circ} \right) + \Delta C_p \left( T - T^\circ + T \ln \frac{T}{T^\circ} \right). \quad (8)$$

Again  $\Delta C_p$  is taken to be independent of temperature in the experimental range. The simplest DSC model involves the isomeric conversion of a species in a strictly two-state manner (*i.e.*, no intermediate state is populated at equilibrium). The denaturation of many single-domain proteins exemplifies this model. This excess heat capacity function is

$$\langle \Delta C_p(T) \rangle = \frac{K}{(K+1)^2} \frac{\Delta H^2}{RT^2} + \frac{K}{K+1} \Delta C_p. \quad (9)$$

The two terms on the right side represent  $\langle \delta C_{p,i}^{\text{tr}}(T) \rangle$  and  $\langle \delta C_{p,i}^{\text{int}}(T) \rangle$ , respectively. The DSC traces in Figure 1 are simulated using Eq (9) with  $\Delta H = 50$  kcal/mol,  $\Delta C_p = 500$  cal mol<sup>-1</sup> K<sup>-1</sup>, and  $T^\circ = 50^\circ\text{C}$  (1 cal = 4.184 J). In this model,  $T^\circ$  is the midpoint of the transition (*i.e.*,  $K = 1$  and  $\alpha = \frac{K}{K+1} = 0.5$ ) and marks the maximum of the  $\langle \delta C_{p,i}^{\text{int}}(T) \rangle$  function.

For transitions involving changes in molecularity,  $\alpha_i(T)$  includes total sample concentration,  $c_i$  in addition to equilibrium constants. While the mechanics of formulating such models is not different, a potential source of inconsistency arises from the choice of unit in thermodynamic parameters. Specifically, every intensive thermodynamic parameter can be defined either per unit of monomer or oligomer. Either choice is correct, of course, but the resultant differences may lead to some confusion. Take for example a two-state homooligomeric transition (Privalov and Potekhin 1986; Freire 1989):



where  $K$  is the equilibrium dissociation constant. Table 2 shows the subtle differences in accounting arising out of the two definitions.

Variable/parameter	Per unit monomer	Per unit oligomer
$c_t$	$[X] + n[X_n]$	$[X]/n + [X_n]$
$\alpha$	$\frac{[X]}{c_t}$	$\frac{[X]}{nc_t}$
$\frac{d\alpha}{dT}$	$\frac{n\alpha(1-\alpha)}{n-\alpha(n-1)} \frac{\Delta H^2}{RT^2}$	$\frac{\alpha(1-\alpha)}{n-\alpha(n-1)} \frac{\Delta H^2}{RT^2}$
$K$	$nc_t^{n-1} \frac{\alpha^n}{1-\alpha}$	$n^n c_t^{n-1} \frac{\alpha^n}{1-\alpha}$
$\Delta G$	$-nRT \ln K$	$-RT \ln K$

Table 2. Equivalent formulations of a two-state homo-oligomeric transition

In the author's experience (Poon et al. 2007), the choice of per unit monomer is more convenient, particularly when oligomers of different molecularities are compared. In addition, it can be seen that  $K$  is a polynomial in  $\alpha$  of order  $n$ . Even in cases where  $\alpha$  can be solved explicitly in terms of  $K$  and  $c_t$  ( $n \leq 4$ ), it is advisable to use numerical procedures such as Newton's method instead to minimize potential algebraic errors and avoid a loss of significance in the fitting procedure.

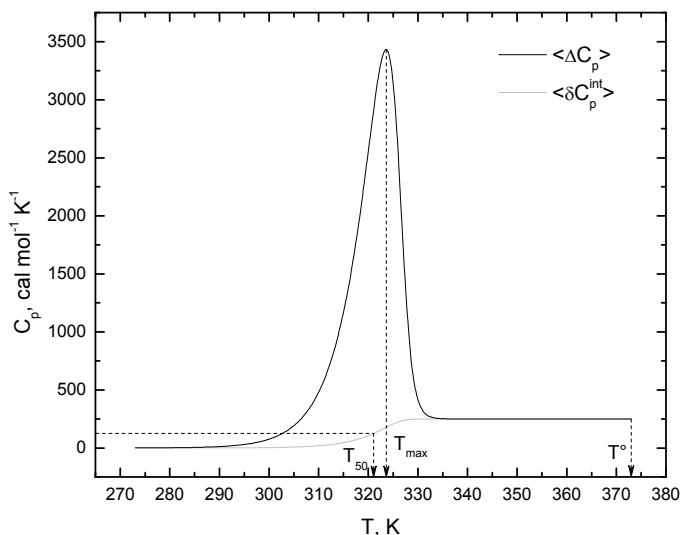


Fig. 2. Two-state dissociation of an homo-oligomer. The traces are simulated for a pentamer ( $n = 5$ ) where  $\Delta H = 50 \text{ kcal mol}^{-1}$ ,  $\Delta C_p = 250 \text{ cal mol}^{-1} \text{ K}^{-1}$  and  $T^o = 100^\circ\text{C}$ . All thermodynamics parameters are per unit *monomer*. Note the asymmetry in both heat capacity functions.



An additional consideration for transitions involving changes in molecularity concerns the choice and interpretation of the reference temperature. In contrast with isomeric transitions,  $T^\circ$  is neither the midpoint of a transition (in the context of concentrations) nor does it mark the maximum of the transition heat capacity function. Both of the latter temperatures are lower than  $T^\circ$ . In addition, the midpoint of the transition,  $T_{50}$ , is below the temperature of the transition heat capacity maximum,  $T_m$ . These relationships are illustrated for the two-state dissociation model in Figure 2. The non-equivalence of  $T_m$  and  $T_{50}$  also introduces a systematic difference between the calorimetric and van't Hoff enthalpies ( $\Delta H_{\text{vH}}$ , reported at  $T_m$ ) (Freire 1989; Freire 1995). Moreover, the actual values of  $T_{50}$  and  $T_m$  are concentration-dependent, and this serves as a diagnostic for a change in molecularity in the transition. For data fitting purposes,  $T^\circ$  remains the most efficient choice because it is independent of concentration. After data fitting, estimates of  $T_m$  and  $T_{50}$  can also be easily obtained from the fitted curve.

Extension of the foregoing discussion applies readily to multi-state transitions. However, an explicit, statistical thermodynamic approach is generally used to derive the required equations for each state as a function of the partition function (Freire and Biltonen 1978). Details in deriving these models have been discussed extensively by Privalov's and Freire's groups (Privalov and Potekhin 1986; Freire 1994). From the standpoint of numerical analysis, it is worth noting that the excess enthalpy is the summed contributions from each state:

$$\langle \Delta C_p \rangle = \frac{d\langle \Delta H(T) \rangle}{dT} = \frac{d}{dT} \left[ \sum_{i=1}^n \alpha_i(T) \Delta H_i(T) \right] \quad (11)$$

Depending on the number of states considered, the expansion of the derivative on the right side of Eq (11) can be formidable. Commercial programs such as Mathematica (Wolfram Research, Champaign, IL, USA) are thus recommended for symbolic manipulation for all but the most trivial derivations. Less preferably, one can numerically integrate the raw  $C_p$  vs.  $T$  data and fit  $\langle \Delta H(T) \rangle$  directly. There are generally enough data points (at 0.1°C resolution) in an experiment that any loss of resolution should be negligibly small.

### 3. Isothermal titration calorimetry

As its name indicates, ITC measures the heat change accompanying the injection of a titrant into titrate at constant temperature. In contemporary instruments, this is accomplished by compensating for any temperature difference between the sample and reference cells (the latter lacking titrate, usually just water). The raw ITC signal is therefore power  $P$ , a time-dependent variable. Integration with respect to time therefore yields heat  $q$  which is the primary dependent variable that tracks the progress of the titration of interest:

$$q(\Delta t) = \int_0^{\Delta t} P dt . \quad (12)$$

Typically, ITC is operated in incremental mode in which the titrant is injected in preset aliquots after successive re-equilibration periods. A feature of ITC that distinguishes it from most titration techniques is that the measured heat does not accumulate from one injection to the next, but dissipates as the instrument measures the heat signal by returning the sample and reference cells to isothermal conditions. ITC is therefore a differential technique

with respect to the concentration of the titrant ( $X$ ) *i.e.*, the derivative of  $q$  with respect to the total titrant concentration,  $[X]_t$ :

$$\frac{dq}{d[X]_t} = V \sum_{i=1}^n \Delta H_i \frac{d[X]_b}{d[X]_t} \quad (13)$$

where  $[X]_b$  is the concentration of  $X$  in the  $i$ -th bound state. This contrasts with most other physical binding signals (*e.g.*, absorbance, fluorescence intensity or anisotropy, pH) which are integrative in nature.

### 3.1 Experimental conditions for ITC

As a thermodynamic tool for studying molecular interactions, the singular strength of ITC is the direct measurement of binding enthalpies. Model-based analysis of ITC data, the subject of Section 3.2, allows the extraction of binding affinity and additional parameters in complex systems. As a label-free technique, ITC compares favorably with other titration techniques such as fluorescence and radioactivity. However, despite much-improved sensitivity (minimum detectable thermal energy  $<0.1 \mu\text{J}$ ), baseline stability, and titrant control found in contemporary instruments, sensitivity of ITC is relatively limited. The actual limit of detection depends primarily on the intrinsic enthalpy of the binding system at the temperature of interest, and to a lesser extent, the physical configuration of the instrument. Roughly speaking, a typical ITC experiment requires at least  $10^{-6}$  M of titrate in a 1-mL volume and  $10^{-4}$  M of titrant in a 100- $\mu\text{L}$  syringe. (Recently, so-called "low volume" instruments equipped with 200- $\mu\text{L}$  cells and 50- $\mu\text{L}$  syringes have become available.)

The sensitivity of ITC is helped considerably by the differential nature of its signal (which is proportional to  $\frac{dq}{d[X]_t}$ ) with respect to titrant concentration (Figure 3). In practice, the

requirement for sufficient concentration in the sample cell to produce sufficient heat signals poses a direct limit on the tightest binding that may be reliably determined. Specifically, depletion of titrant dominates at titrate concentrations that are high relative to the dissociation constant  $K$ . Under this condition, the titration approaches a discontinuity in the first derivative at the equivalence point (Poon 2010). In the case of simple 1:1 binding, an empirical relationship that the product  $1 < c_{\text{titrate}}/K < 1000$ , where  $c_{\text{titrate}}$  is the titrate concentration, is optimal for reliable estimation of  $K$  (Wiseman et al. 1989). One way of getting around this problem for very tight binding is to lower the apparent value of  $K$  by including a suitable, fixed concentration of competitor in the cell. Another possibility is continuous ITC, which will be discussed in Section 3.1.2. It should be noted that even under conditions where  $K$  cannot be determined,  $\Delta H$  can still be determined by integration of the measured heat:

$$\Delta H = \frac{1}{c_{\text{titrate}}} \int_0^v \frac{q(v)}{v} dv, \quad (14)$$

where  $V$  is the volume of the (fixed) sample cell. (A volume correction is generally necessary; see Section 3.2.1.3.) A requirement for Eq (14) is, of course, that the titration is complete. Whichever the case, repeating the titration at different temperatures provides an estimate of the change in heat capacity by Kirchoff's relation:

$$\Delta C_p = \frac{d\Delta H(T)}{dT} . \quad (15)$$

If the interaction under investigation occurs in a buffered solution, the earlier discussion on the temperature dependence of  $pK_a$  would again be relevant. More generally, binding that involves coupled uptake or release of protons will contain the buffer's enthalpy of ionization in the apparent binding enthalpy (Fisher and Singh 1995).

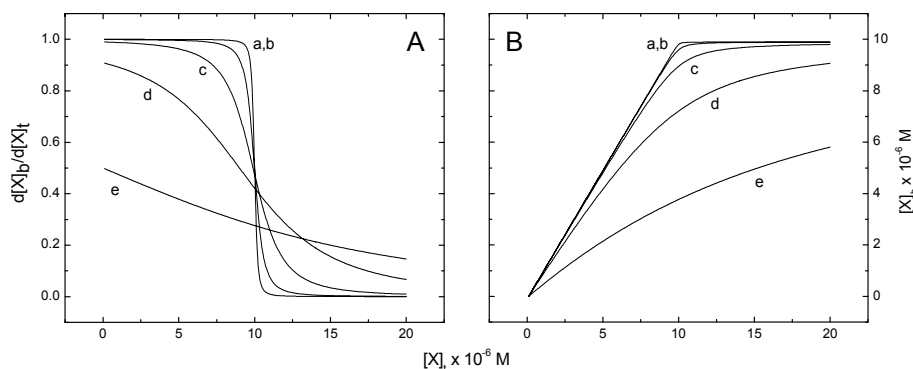


Fig. 3. The differential nature of ITC data. The volume and enthalpy are intentionally omitted to illustrate the differential nature of ITC data with respect to total titrant concentration. Simulated 1:1 binding to 10  $\mu\text{M}$  titrate is shown in Panel A, with the corresponding integrated binding curves in Panel B. The values of  $K$  are as follows: a,  $10^{-9}$  M; b,  $10^{-8}$  M; c,  $10^{-7}$  M; d,  $10^{-6}$  M; and e,  $10^{-5}$  M. The values of  $C_{\text{titrate}}/K$  range from  $10^4$  to 1.

Finally, ITC has been used as an “analog” of DSC for studying the stability of complexes. A concentrated solution of complex in the syringe is titrated into pure water or buffer in the cell. The resulting dilution drives complex dissociation and the attendant enthalpy is measured. This method has been used to characterize complexes through the spectrum of stoichiometries, from dimers (Burrows et al. 1994; Lovatt et al. 1996) to higher oligomers (Lassalle et al. 1998; Luke et al. 2005) to polymeric species (Stoesser and Gill 1967; Arnaud and Bouteiller 2004). Again, given the sensitivity of ITC, relatively high concentrations are required, so this technique is limited to the measurement of relatively weak complexes. It has been shown (Poon 2010) that the ITC data can be used to diagnose a dimeric or higher-order complex based on the presence of an inflection point in the latter.

### 3.1.1 Baseline signals

Two types of baselines are operative in an ITC experiment. One is instrument noise. Drifts on the order of 0.02  $\mu\text{W}/\text{h}$  are routinely achievable in contemporary instruments. Another source of baseline arises from the injection of titrant. At the very least, viscous mixing makes a measurable if small exothermic contribution to the observed heat. This effect can be observed in a blank-to-blank injection (Figure 4), and serves as a casual useful indicator of the cleanliness of the cell and syringe between sample runs. Moreover, any mismatch in the matrices of the titrant and titrate will be manifest as a dilution enthalpy with each injection.

For small molecules, solids or lyophilized samples are usually dissolved in water or buffer. To complicate matters, hydrophobic solutes often require a cosolvent such as DMSO or DMF to achieve initial solubility before addition of the aqueous solvent; dilution of the cosolvent will therefore contribute to the observed heat at each injection. In other cases, ionizable solutes can perturb solution pH due to their substantial ( $>10^{-3}$  M) concentration in the syringe. Unless the solvent is adequately buffered, the pH in the cell and syringe will differ significantly and neutralization heats will contribute to the observed signal.

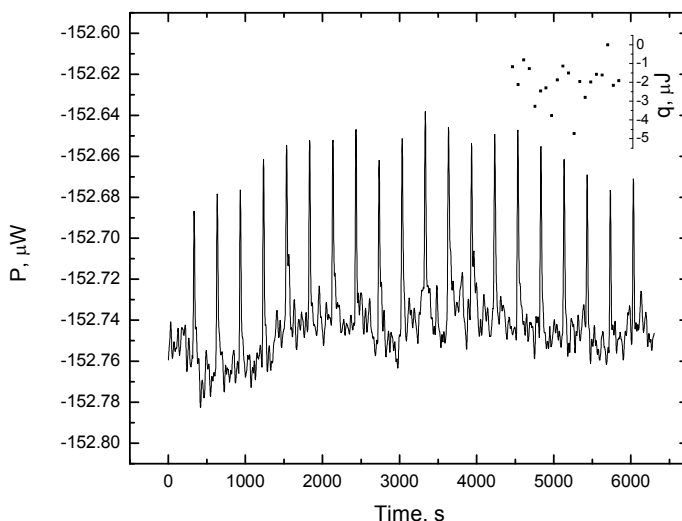


Fig. 4. Typical water-water control "titration." *Inset*, integrated areas of the water peaks. Given the low signal-to-noise ratios, the values must be considered very crude. However, the scatter also attests to the cleanliness of the cell and syringe.

In the case of macromolecular titrant and titrates, the solution matrices can be closely matched by extensive co-dialysis in the same solution. Care must be taken, however, with ionic polymers such as nucleic acids. These solutes can alter the distribution of ions in their compartment during dialysis due to the Donnan effect. Specifically, a non-diffusibile polyionic solute excludes diffusible ions of the same charge from their compartment and therefore induces an asymmetric distribution of diffusible ions across the semipermeable membrane at equilibrium. To illustrate, for a simple system consisting only of the non-diffusibile polyion M and a monovalent salt AB, the ratio concentrations of  $A^+$  or  $B^-$  across the membrane is (Cantor and Schimmel 1980):

$$r_D = \frac{z[M]}{2[AB]_i} + \sqrt{1 + \left(\frac{z[M]}{2[AB]_i}\right)^2} \equiv \Psi + \sqrt{1 + \Psi^2}, \quad (16)$$

where  $z$  is the charge on the polyion (shown simply as M for brevity). At the concentrations required for the syringe,  $z[M]$  can be substantial. For example, a duplex oligonucleotide consisting of 20 base-pairs represents at 100  $\mu$ M contributes 4 mM in total anions

(phosphates). The so-called Donnan ratio,  $r_D$ , is significantly above unity even at low polyion:salt ratios ( $\Psi$ ). Eq (16) shows that a ten-fold excess of salt concentrations ( $\Psi=0.1$ ) leads to a 10% exclusion of A or B from the compartment occupied by polyion. If AB is a buffer salt, the result is also a change of pH. Such asymmetry will modify the heat observed by ITC. As is well known, and illustrated by Eq (16), Donnan effects can be suppressed by ensuring a sufficient excess of diffusible salt in the dialysate. Unfortunately, the required salt concentrations may interfere with the investigation of interactions mediated substantially by electrostatic effects (as is usually the case for polyions). If possible, therefore, it is advisable to arrange for the polyionic species to be in the cell, where concentrations are lower.

In practical terms, none of these baselines effects are significant if the heat generated by the interaction of interest is sufficiently strong. This is not always achievable, however, for several reasons. Availability or solubility of the sample, particularly biological samples, may be limiting. It may also be desirable, for example, to perform titrations at a range of cell concentrations for binding to oligomerizing systems. Characterization of binding to a polyion at low salt concentration may require a reduced concentration. Thus, strategies for handling relatively low signal-to-noise situations are helpful in many situations. The most basic of these involve inspection and, where necessary, manual editing of the power baseline to mitigate the occasion excursion due to instrumental noise. To this end, an increase in the time between injection may be indicated to unambiguously identify the restoration of baseline. In addition, the residual heats (which may be up to 10  $\mu$ J) after the equivalence point are unlikely to be negligible. Subtraction with data from a titrant-to-blank run would likely introduce more noise into the data and be no more helpful than taking a simple average of the final post-equivalence heats. If the data is to be fitted to models, a more appropriate solution is to add a constant parameter  $B$  to the fitting equation:

$$\frac{dq_{\text{obsd}}}{d[X]_i} = V \sum_{i=1}^n \Delta H_i \frac{d[X]_b}{d[X]_i} + B. \quad (17)$$

### 3.1.2 Continuous ITC titrations

While incremental titrations most commonly performed in ITC, an alternative mode of operation is a continuous titration (cITC) (Markova and Hallén 2004). In cITC, the titrant is continuously into injected the cell at a low rate ( $\sim 0.1 \mu\text{L/s}$ ). The primary advantage of cITC is throughput. An incremental ITC experiment typically requiring 20 injections of 5  $\mu\text{L}$  at intervals of 300 s takes (neglecting time for baseline stabilization) 100 min; at 0.1  $\mu\text{L/s}$ , cITC would require approximately 17 min. Another potential motivation for cITC is increased resolution in terms of model-dependent analysis. In incremental ITC, peak-by-peak integration of thermal power is performed to obtain  $\frac{dq}{d[X]_i}$ . This step reduces the number of

collected data points (typically 1  $\text{s}^{-1}$  over  $10^3$  s, or about one to two hours) for nonlinear regression to the number of injections (usually 10 to 30). By using the thermal power data directly for fitting, cITC can in principle discriminate the curvature required to define tight binding.

Maintenance of quasi-equilibrium conditions throughout the titration is essential to correct interpretation of derived thermodynamic parameters and is a major concern for cITC. To this end, the stirring rate in cITC must be considerably increased (up to 700 rpm) relative to incremental ITC to facilitate mixing of titrant into the titrate solution. Additionally, the

instrument must be able to provide sufficient thermal compensation during the titration to maintain isothermal conditions. Finally, the kinetics of the interaction of interest must be fast relative to the injection rate. Generally speaking, these criteria are most likely met by relatively high-affinity interactions with moderate binding enthalpies.

### 3.2 Analysis of ITC data

The direct measurement of  $\Delta H$  is considered a significant advantage over non-calorimetric binding methods since the latter access the binding thermodynamics indirectly in terms of the equilibrium constant  $K$  via the van't Hoff equation:

$$\frac{d \ln K}{d(1/T)} = -\frac{\Delta H_{\text{vH}}}{R}. \quad (18)$$

Estimation of  $\Delta C_p$  by non-calorimetric methods, therefore, involves taking a second derivative of the measured data. Another potential source of difficulty is the interpretation of  $K$ , which is model-dependent. As in DSC, however, ITC becomes a considerably more useful analytical technique when model fitting is used for parameter estimation and hypothesis testing. Moreover, agreement between the directly-fitted, calorimetric  $\Delta H$  and the van't Hoff  $\Delta H_{\text{vH}}$  obtained from Eq (18) is a strong indication of the physical correctness of the model at hand.

The most fundamental concept in the analysis of ITC data is the differential nature of the heat signal with respect to titrant concentration (Figure 3). Recalling Eq (13),

$$\frac{dq}{d[X]_t} = V \sum_{i=1}^n \Delta H_i \frac{d[X_i]_b}{d[X]_t}.$$

Integration of this fundamental equation gives the is the functional form of conventional binding models:

$$q([X]_t) = V \sum_{i=1}^n \Delta H_i [X_i]_b([X]_t) \quad (19)$$

The ideal approach to fitting ITC data is to directly fit Eq (13) (Poon 2010). For simple models, it is possible to write analytical expressions for  $\frac{d[X_i]_b}{d[X]_t}$  explicitly in terms of  $[X]_t$  i.e., the functional form of Eq (19). Take, for example, simple 1:1 binding of  $X + Y \rightleftharpoons XY$ :

$$\begin{aligned} [X]_t [Y]_t - ([X]_t + [Y]_t + K)[X]_b + [X]_b^2 &= 0 \\ [X]_b &= \frac{[X]_t + [Y]_t + K - \sqrt{([X]_t + [Y]_t + K)^2 - 4[X]_t [Y]_t}}{2}, \end{aligned} \quad (20)$$

where  $K = \frac{[X][Y]}{[XY]} = \frac{[X][Y]}{[X]_b}$  is the equilibrium dissociation constant. Its derivative is

$$\frac{d[X]_b}{d[X]_t} = \frac{1}{2} + \frac{1 - \Phi - r}{2\sqrt{(\Phi + 1 + r)^2 - 4\Phi}}, \quad (21)$$

where  $r = K/[X]_t$  and  $\Phi = [X]_t/[Y]_t$ . Eq (21) is sometimes referred to as the Wiseman isotherm (Wiseman et al. 1989). As complexity of the model increases, however, the algebra involved rapidly becomes prohibitive. In any case, numerical methods provide the means for generating solutions for Eq (13).

### 3.2.1 Numerical aspects of ITC data analysis

Given the usual practice of formulating models in terms of total or unbound concentrations, rather than their derivatives, conventional ITC data analysis has handled the differential nature of calorimetric data by fitting to a finite-difference version of Eq (13). Thus, for the  $j$ -th injection:

$$\Delta q_j = q_j - q_{j-1} - \delta q_j, \quad (22)$$

where  $\delta q_j$  is a volume-correction factor. This approach has the apparent advantage that models which have been formulated in terms of  $[X]_b$  vs.  $[X]_t$  can be used directly. However, Eq (22) represents poor practice in nonlinear regression. Since each computation of  $\Delta q_j$  requires evaluation of  $q$  corresponding to two consecutive injections,  $j$  and  $j-1$ , the data points (and their errors, albeit small) are not independent. Independence of observations constitutes a major assumption of nonlinear regression, one which the recursive form of Eq (22) clearly violates. Specifically, the residual in  $\Delta q_j$  during fitting becomes increasing correlated with increasing  $j$ . In addition, the resultant propagation of error likely violates the assumption of homoscedasticity (uniform variance) as well. Unless specialized regression techniques are employed (such as correlated least-squares), violations of these assumptions potentially calls the errors of the parameters, and possibly the parameters themselves, into question.

From the perspective of numerical analysis, Eq (22) is also entirely unnecessary. As stated previously, the most appropriate approach to handling ITC data is to fit Eq (13) directly.

Posed in the form of Eq (13), ITC models (more specifically, the solution of  $\frac{d[X]_b}{d[X]_t}$ ) represent

classic initial value problems (IVPs). IVPs are first-order ordinary differential equations (ODEs) with a specific initial condition ( $[X]_b = 0$  at  $[X]_t = 0$ ) for which numerical methods for their solution are well-established. In addition avoiding the statistical pitfalls of Eq (22), formulating ITC models as differential equations simplify the algebra considerably. This is because implicit differentiation offers a welcome shortcut that obviates the need for an explicit solution for  $[X]_b$ . This is illustrated for  $X + Y \rightleftharpoons XY$  with Eq (20):

$$\begin{aligned} [Y]_t - ([Y]_t + [X]_t + K) \frac{d[X]_b}{d[X]_t} - [X]_b + 2[X]_b \frac{d[X]_b}{d[X]_t} &= 0 \\ \frac{d[X]_b}{d[X]_t} (2[X]_b - [Y]_t - [X]_t - K) &= [X]_b - [Y]_t \quad (23) \\ \frac{d[X]_b}{d[X]_t} &= \frac{[X]_b - [Y]_t}{2[X]_b - [Y]_t - [X]_t - K} \end{aligned}$$

Recognizing that  $[X]_b$ ,  $[Y]_t$  are constants (and  $K$  being the parameter to be estimated), Eq (23) takes on a noticeable simpler (but equivalent) form compared to the Wiseman isotherm, Eq (21). More importantly, implicit differentiation always yields an explicit ODE even when no

analytical expression for  $[X]_b$  exists *e.g.*, polynomials of order  $>4$ . Direct substitution of Eq (23) into Eq (13) directly yields the ODE needed to generate values for nonlinear regression. This approach of formulating explicit titration models for ITC has been demonstrated for many empirical models in common use, including the multi-site model, homotropic cooperativity, and two-state self-association (Poon 2010). To further illustrate this approach, a competitive binding model will be examined. This is a method of measuring tight binding by ITC by adding a competitive species in the cell to reduce the apparent affinity (Sigurskjold 2000). The mechanism is an example of the general multi-site model involving two ligands competing for a single site (Wells 1992). The equilibrium distribution of bound titrate is a function of the affinities of the titrant  $X_1$  and competitor  $X_2$  for the titrate  $Y$  as well as the total concentrations of all three species. For the titrant-titrate complex,  $X_1Y$ :

$$a_0 + a_1[X_1Y] + a_2[X_1Y]^2 + a_3[X_1Y]^3 = 0, \quad (24)$$

where

$$\begin{aligned} a_0 &= -K_2[X_1]_t^2[Y]_t \\ a_1 &= [X_1]_t \{K_1(K_2 + [X_2]_t - [Y]_t) + K_2([X_1]_t + 2[Y]_t)\} \\ a_2 &= K_1^2 - K_2(2[X_1]_t + [Y]_t) - K_1(K_2 - [X_1]_t + [X_2]_t - [Y]_t) \\ a_3 &= K_2 - K_1 \end{aligned}$$

and  $K_1$  and  $K_2$  are the equilibrium dissociation constants of  $X_1$  and  $X_2$  for the titrate, respectively. The corresponding expression for the competitor-titrate complex,  $X_2Y$ , can be obtained from symmetry arguments:

$$b_0 + b_1[X_2Y] + b_2[X_2Y]^2 + b_3[X_2Y]^3 = 0, \quad (25)$$

where

$$\begin{aligned} b_0 &= -K_1[X_2]_t^2[Y]_t \\ b_1 &= [X_2]_t \{K_2([X_1]_t - [Y]_t) + K_1(K_2 + [X_2]_t + 2[Y]_t)\} \\ b_2 &= K_2(K_2 - [X_1]_t + [X_2]_t + [Y]_t) - K_1(K_2 + 2[X_2]_t + [Y]_t) \\ b_3 &= K_1 - K_2 \end{aligned}$$

It is possible to solve the cubic equation in Eqs (24) and (25) explicitly, followed by differentiation of the solutions to obtain  $\frac{d[X_1Y]}{d[X_1]_t}$  and  $\frac{d[X_2Y]}{d[X_1]_t}$  as was done with the Wiseman isotherm. All of this is avoided, however, by implicit differentiation with respect to  $[X_1]_t$  which directly yields the required derivatives:

$$\begin{aligned} \frac{d[X_1Y]}{d[X_1]_t} &= \frac{-\left(\frac{da_2}{d[X_1]_t}[X_1Y]^2 + \frac{da_1}{d[X_1]_t}[X_1Y] + \frac{da_0}{d[X_1]_t}\right)}{3a_3[X_1Y]^2 + 2a_2[X_1Y] + a_1} \\ \frac{d[X_2Y]}{d[X_1]_t} &= \frac{-\left(\frac{db_2}{d[X_1]_t}[X_2Y]^2 + \frac{db_1}{d[X_1]_t}[X_2Y] + \frac{db_0}{d[X_1]_t}\right)}{3b_3[X_2Y]^2 + 2b_2[X_2Y] + b_1} \end{aligned} \quad (26)$$



The observed heat is now the sum of the enthalpy of unbinding of  $X_2$  from  $Y$  and the binding of  $X_1$  to  $Y$ :

$$\frac{dq}{d[X_1]_t} = V \left( \Delta H_{X_1Y} \frac{d[X_1Y]}{d[X_1]_t} + \Delta H_{X_2Y} \frac{d[X_2Y]}{d[X_1]_t} \right). \quad (27)$$

As shown in Figure 5, a judicious choice of competitor reduces the apparent affinity of the titrant to a range more amenable for regression. In this case, the initial condition for  $[X_2Y]$  is *not* zero because the titrate is essentially pre-equilibrated with competitor before any titrant has been injected. The initial value for numerical integration is supplied by Eq (20):

$$[X_2Y] = \frac{[X_2]_t + [Y]_t + K_2 - \sqrt{([X_2]_t + [Y]_t + K_2)^2 - 4[X_2]_t[Y]_t}}{2}.$$

Thus, the competitive model requires prior knowledge of both the concentration and affinity of the competitor. Since the competitor would therefore require its own characterization, a weak competitor is preferred, which means it would need to be present at significant concentrations (i.e.,  $[Y]_t \geq K$ ). Since binding of the titrate involves the unbinding of the competitor, the enthalpy of competitor unbinding may be substantially convoluted in the observed heat as indicated by Eq (27).

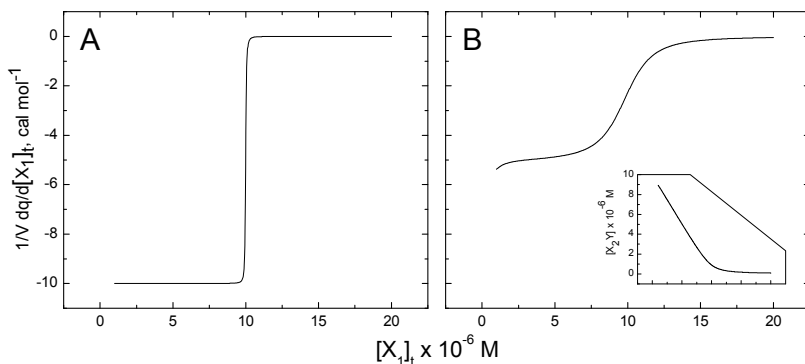


Fig. 5. Titrant-titrate binding in the presence of a competitor. *A*, simulated 1:1 titrant-titrate binding in the absence of competitor. The parameters are  $K_1 = 10^{-10}$  M,  $\Delta H_1 = -10$  kcal mol $^{-1}$ , and  $[Y]_t = 10$   $\mu$ M. Data representing this level of affinity would be unfit for model-fitting;  $K_1/[Y]_t = 10^4$ . *B*, in the presence a competitor,  $X_2$ , at 10  $\mu$ M ( $K_2 = 10^{-8}$  M,  $\Delta H_2 = -5$  kcal mol $^{-1}$ ). *Inset*, titration of the competitor.

Explicit titration models are also amenable to formulating models for continuous ITC. In the case of cITC, thermal power  $P$  is directly used in model fitting, so Eq (19) needs to be differentiated with respect to *time*:

$$P = \frac{dq(t)}{dt} = \frac{d}{dt} \left\{ V \sum_{i=1}^n \Delta H_i [X_i]_b(t) \right\} = V \left[ \sum_{i=1}^n \Delta H_i \frac{d[X_i]_b(t)}{dt} \right]. \quad (28)$$

Applying the chain rule and the relation  $[X_i]_{\text{tot}} = \frac{c_{\text{syr}} v_{\text{inj}}}{V}$ , where  $c_{\text{syr}}$  is the concentration of titrant in the syringe and  $v_{\text{inj}}$  is the injection volume,

$$\frac{d[X_i]_{\text{b}}}{dt} = \frac{d[X_i]_{\text{b}}}{d[X_i]_{\text{tot}}} \frac{d[X_i]_{\text{tot}}}{dt} = \frac{d[X_i]_{\text{b}}}{d[X_i]_{\text{tot}}} \left( \frac{c_{\text{syr}}}{V} \frac{dv_{\text{inj}}}{dt} \right). \quad (29)$$

(The subscript "tot" has been used to denote total concentration to avoid ambiguity with the variable  $t$  for time.) Note that  $\frac{dv_{\text{inj}}}{dt}$  is the (constant) injection rate. Substituting into Eq (28) gives

$$P = c_{\text{syr}} \frac{dv_{\text{inj}}}{dt} \left[ \sum_{i=1}^n \Delta H_i \frac{d[X_i]_{\text{b}}}{d[X_i]_{\text{tot}}} \right]. \quad (30)$$

Thus explicit expressions of  $\frac{d[X_i]_{\text{b}}}{d[X_i]_{\text{tot}}}$  can be directly used as in incremental ITC. At a sufficiently low injection rate, Eq (30) has the potential of "flattening out" the titration by transforming it in the time domain. This feature, in addition to the higher density of data points available for regression, may allow cITC to characterize much tighter binding than is practicable with incremental ITC (Markova and Hallén 2004).

In the foregoing discussion, the need to correct for various displacement and dilution effects due to the injection process has not been considered. In the author's experience, this is best handled during preliminary data reduction, before nonlinear regression. This aspect will be discussed in Section 3.2.1.3.

### 3.2.1.1 Choice of dependent variable

Another benefit of formulating ITC models as explicit ODEs is the flexibility in the choice of dependent variable for implicit differentiation, as long as it is a function in  $[X]_t$ . For the 1:1 binding model, formulation in terms of the unbound titrant  $X$  gives the (and simple) functional form of the familiar Langmuir isotherm:

$$[X]_{\text{b}} = \frac{[X][Y]_t}{[X] + K}. \quad (31)$$

Applying the chain rule of calculus,

$$\frac{d[X]_{\text{b}}}{d[X]_t} = \frac{d[X]_{\text{b}}}{d[X]} \frac{d[X]}{d[X]_t} = \frac{[X]K}{([X] + K)^2} \frac{d[X]}{d[X]_t}. \quad (32)$$

At the same time, using the equation of state  $[X] = [X]_t - [X]_{\text{b}}$ ,

$$\frac{d[X]}{d[X]_t} = \frac{d([X]_t - [X]_{\text{b}})}{d[X]_t} = 1 - \frac{d[X]_{\text{b}}}{d[X]_t}. \quad (33)$$

Substituting the results from Eq (23) into Eq (33),

$$\frac{d[X]}{d[X]_t} = 1 - \frac{[X]_b - [X]_t}{2[X]_b - [X]_t - [Y]_t - K} = \frac{[X] + K}{2[X] - [X]_t + [Y]_t + K} \quad (34)$$

The simultaneous equations (32) and (34) represent another formulation of the same model, except now  $[X]$  is the explicit dependent variable instead of  $[X]_b$ . (The initial condition is  $[X] = 0$  at  $[X]_t = 0$ .) Of course, we have previously derived Eq (23) directly, so this approach is regressive for this simple model. However, The flexibility to use any dependent variable of  $[X]_t$  is useful, for example, for models that are formulated in terms of the binding polynomial which is based on  $[L]$  (Schellman 1975; Freire et al. 2009).

### 3.2.1.2 Practical considerations in implementing explicit ITC models

Successful implementation of Eq (13) requires numerical procedures for solving IVPs. The explicit, closed-form ODEs encountered in most models are typically ratios of polynomials. These functions are generally amenable any of the standard Runge-Kutta methods, which are widely available. A fast CPU is helpful, but not required. To this end, the tolerance for iteration should not be unnecessarily stringent in relation to the concentrations used and the value of  $K$  expected. Generally, a value of  $10^{-6}$  will suffice. Even average single-core CPUs will handle ODEs of practical complexity with reasonable dispatch. Any significant delay is almost always related to input/output issues and can be alleviated by suppressing intermediate output. Numerical failures, if they occur, usually does so when there are two or more very different scales of the independent variable on which the dependent variable is changing *e.g.*, extremely tight binding to one site in the multi-site model. One such example is competitive binding in the presence of a very strong competitor in Eq (27). These pathological scenarios are associated with “stiff” differential equations that require more specialized algorithms for numerical solution. Fortunately, these situations are unlikely to be encountered as they are usually incompatible with experimental data in the first place.

Several technical software suites, such as Mathematica (Wolfram Research, Champaign, IL, USA), MATLAB (the MathWorks, Natick, MA USA), Maple (Maplesoft, Waterloo, Ontario, Canada), and IgorPro (WaveMetrics, OR, USA) which have built-in numerical ODE and least-square minimization capabilities, represent full-featured, integrated solutions. Alternatively, pre-compiled libraries containing optimized algorithms for numerical ODEs and least-square minimization are available commercially (the Numerical Algorithm Group Library; Numerical Algorithms Group [NAG], Oxford, UK) or free (GNU Scientific Library [GSL]) for most computing platforms. Functions from these libraries can be called under standard programming environments (*e.g.*, C/C++, FORTRAN) to perform the required procedures. Some commercial data analysis software such as Origin (Northampton, MA, USA) can interface with external libraries such as the NAG Library or GSL to perform numerical ODEs within their least-squares routines. If access to external libraries is not available, an adequate alternative is to code a numerical ODE algorithm (such as Runge-Kutta-Fehlberg) as part of the target function within the data fitting routines of the analysis program. “Cookbook” recipes for a variety of ODE solvers are straightforward and can be found in most texts of numerical analysis (Press 2007).

### 3.2.1.3 Volume correction

The sample and reference cells are typically overfilled for both DSC and ITC. Overfilling maximizes heat transfer between the solution and the wall of the cell as air is a poor thermal conductor. In the case of ITC, overfilling also minimizes stray signal arising from mechanical agitation of the solution-air-cell interface caused by the stirring paddle.

However, the introduction of titrant into an overfilled ITC sample cell leads to displacement effects that need to be taken into account. Specifically, each injected volume of titrant simultaneously displaces an equal volume of titrate and any previously injected titrant out of the sample cell (into the access tube). The accounting for these displaced volumes and their effect on titrant and titrate concentration is made on the assumption that the displaced material is immediately and completely excluded from the titration. This implies that no mixing occurs between the injected and displaced materials at the time injection. The concentrations of the titrant X after the  $i$ -th injection in the cell is therefore

$$[X]_{t,i} = \frac{v_{inj}c_{syr} + [X]_{t,i-1}(V - v_{inj})}{V}, \quad [X]_{t,0} = 0, \quad (35)$$

where  $V$  is the cell volume and  $v_{inj}$  is the injection volume. The corresponding concentrations of the titrate Y is

$$[Y]_{t,i} = \frac{[Y]_{t,i-1}(V - v_{inj})}{V}. \quad (36)$$

There are two ways to handle volume corrections. One is to incorporate Eqs (35) and (36) as additional terms in the fitting equation. In the author's experience (Poon 2010), it is more efficient instead to perform the volume corrections on the dataset at the outset, and simply treat  $[Y]_{t,i}$  as an additional dependent variable in the least-square procedure.

#### 3.2.1.4 Error analysis in ITC

Compared to other titrations, particularly in the biochemical laboratory, that requires extensive manual manipulation *e.g.*, electrophoretic mobility shift, filter binding, ITC instrumentation offers a greatly reduced level of statistical error in the measured data. Nonetheless, detailed theoretical and experimental studies of the nature and magnitude of statistical errors in ITC have offered insight into how instrumental errors can be minimized in fitted parameters. Although such studies have so far only focused on 1:1 binding, it has become clear that at least two régimes of instrumental errors exist (Tellinghuisen 2003; 2005b). Specifically, for titrations associated with relatively large heats ( $>300 \mu\text{cal}$  or "high- $q$ "), proportional errors in the injected volumes dominate, and the inclusion of statistical weights is indicated to optimize fitted parameters. For "low- $q$ " titrations, which typify low-concentration titrations needed for tight binding interactions, constant errors in thermal detection and compensation dominate, and unweighted fitting does not adversely affect parameter optimization. In either case, the optimal number of injections is considerably lower than the norm of 20 to 25 injections, especially if variable volume procedures are employed (Tellinghuisen 2005a). This approach could substantially reduce titration time and increase experimental throughput. It may be pointed out, however, that in some applications, sample-to-sample variation may be greater than any statistical error inherent in the analytical technique. In these cases, standard errors in parameters estimated from replicate experiments will be higher but more representative indicators of experimental uncertainty than the fitting error extracted from the variance-covariance matrix.

## 4. Conclusion

Commercial development of microcalorimetry has greatly increased the accessibility of this technique for the thermodynamic characterization of chemical systems in solution.

Unfortunately, the "black-box" nature of commercial software has engendered unwarranted reliance by many users on the turnkey software accompanying their instruments, and an attendant tendency to fit data to models of questionable relevance to the actual chemistry. This chapter discusses several novel aspects and potential pitfalls in the experimental practice and analysis of both DSC and ITC. This information should enable users to tailor their experiments and model-dependent analysis to the particular requirements.

## 5. Acknowledgement

Financial support by the College of Pharmacy at Washington State University is acknowledged.

## 6. References

- Arnaud, A., and Bouteiller, L. (2004). Isothermal titration calorimetry of supramolecular polymers. *Langmuir* 20: 6858-6863
- Burrows, S.D., Doyle, M.L., Murphy, K.P., Franklin, S.G., White, J.R., Brooks, I., McNulty, D.E., Scott, M.O., Knutson, J.R., and Porter, D. (1994). Determination of the monomer-dimer equilibrium of interleukin-8 reveals it is a monomer at physiological concentrations. *Biochemistry* 33: 12741-12745
- Cantor, C.R., and Schimmel, P.R. (1980). *Biophysical Chemistry: the behavior of biological macromolecules*. W. H. Freeman, 0716711915, San Francisco, USA
- Disteche, A. (1972). Effects of pressure on the dissociation of weak acids. *Symp Soc Exp Biol* 26: 27-60
- Fisher, H.F., and Singh, N. (1995). Calorimetric methods for interpreting protein-ligand interactions. *Methods Enzymol* 259: 194-221
- Freire, E. (1989). Statistical thermodynamic analysis of the heat capacity function associated with protein folding-unfolding transitions. *Comments Mol Cell Biophys* 6: 123-140
- Freire, E. (1994). Statistical thermodynamic analysis of differential scanning calorimetry data: Structural deconvolution of heat capacity function of proteins. *Methods Enzymol* 240: 502-530
- Freire, E. (1995). Thermal denaturation methods in the study of protein folding. *Methods Enzymol* 259: 144-168
- Freire, E., and Biltonen, R.L. (1978). Statistical mechanical deconvolution of thermal transitions in macromolecules. I. Theory and application to homogeneous systems. *Biopolymers* 17: 463-479
- Freire, E., Schön, A., and Velazquez-Campoy, A. (2009). Isothermal Titration Calorimetry: General Formalism Using Binding Polynomials. *Methods Enzymol* 455: 127-155
- Goldberg, R.N., Kishore, N., and Lennen, R.M. (2002). Thermodynamic Quantities for the Ionization Reactions of Buffers. *J Phys Chem Ref Data* 31: 231-370
- Good, N.E., Winget, G.D., Winter, W., Connolly, T.N., Izawa, S., and Singh, R.M.M. (1966). Hydrogen Ion Buffers for Biological Research. *Biochemistry* 5: 467-477
- King, E.J. (1969). Volume changes for ionization of formic, acetic, and butyric acids and the glycinium ion in aqueous solution at 25°C. *J Phys Chem* 73: 1220-1232
- Kitamura, Y., and Itoh, T. (1987). Reaction volume of protonic ionization for buffering agents. Prediction of pressure dependence of pH and pOH. *J Solution Chem* 16: 715-725

- Lassalle, M.W., Hinz, H.J., Wenzel, H., Vlasi, M., Kokkinidis, M., and Cesareni, G. (1998). Dimer-to-tetramer transformation: loop excision dramatically alters structure and stability of the ROP four alpha-helix bundle protein. *J Mol Biol* 279: 987-1000
- Lo Surdo, A., Bernstrom, K., Jonsson, C.A., and Millero, F.J. (1979). Molal volume and adiabatic compressibility of aqueous phosphate solutions at 25.degree.C. *J Phys Chem* 83: 1255-1262
- Lovatt, M., Cooper, A., and Camilleri, P. (1996). Energetics of cyclodextrin-induced dissociation of insulin. *Eur Biophys J* 24: 354-357
- Luke, K., Apiyo, D., and Wittung-Stafshede, P. (2005). Dissecting homo-heptamer thermodynamics by isothermal titration calorimetry: entropy-driven assembly of co-chaperonin protein 10. *Biophys J* 89: 3332-3336
- Markova, N., and Hallén, D. (2004). The development of a continuous isothermal titration calorimetric method for equilibrium studies. *Anal Biochem* 331: 77-88
- Poon, G.M. (2010). Explicit formulation of titration models for isothermal titration calorimetry. *Anal Biochem* 400: 229-236
- Poon, G.M., Broxk, R.D., Sung, M., and Gariépy, J. (2007). Tandem Dimerization of the Human p53 Tetramerization Domain Stabilizes a Primary Dimer Intermediate and Dramatically Enhances its Oligomeric Stability. *J Mol Biol* 365: 1217-1231
- Poon, G.M., Gross, P., and Macgregor, R.B., Jr. (2002). The sequence-specific association of the ETS domain of murine PU.1 with DNA exhibits unusual energetics. *Biochemistry* 41: 2361-2371
- Press, W.H. (2007). *Numerical recipes : the art of scientific computing*, 3rd ed. Cambridge University Press, 0521880688, Cambridge, UK ; New York, USA
- Privalov, P.L., and Dragan, A.I. (2007). Microcalorimetry of biological macromolecules. *Biophys Chem* 126: 16-24
- Privalov, P.L., and Potekhin, S.A. (1986). Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods Enzymol* 131: 4-51
- Schellman, J.A. (1975). Macromolecular binding. *Biopolymers* 14: 999-1018
- Sigurskjold, B.W. (2000). Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal Biochem* 277: 260-266
- Stoesser, P.R., and Gill, S.J. (1967). Calorimetric study of self-association of 6-methylpurine in water. *J Phys Chem* 71: 564-567
- Tellinghuisen, J. (2003). A study of statistical error in isothermal titration calorimetry. *Anal Biochem* 321: 79-88
- Tellinghuisen, J. (2005a). Optimizing experimental parameters in isothermal titration calorimetry. *J Phys Chem B* 109: 20027-20035
- Tellinghuisen, J. (2005b). Statistical error in isothermal titration calorimetry: variance function estimation from generalized least squares. *Anal Biochem* 343: 106-115
- Wells, J.W. (1992). Analysis and interpretation of binding at equilibrium. In: *Receptor-Ligand Interactions: a Practical Approach*. E.C. Hulme(Ed., pp. 289-395. IRL Press at Oxford University Press, 0199630909, Oxford, England; New York, USA
- Wiseman, T., Williston, S., Brandts, J.F., and Lin, L.N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* 179: 131-137

# Thermodynamics of Metal Hydrides: Tailoring Reaction Enthalpies of Hydrogen Storage Materials

Martin Dornheim

*Institute of Materials Research, Department of Nanotechnology,  
Helmholtz-Zentrum Geesthacht  
Germany*

## 1. Introduction

Considering the increasing pollution and exploitation of fossil energy resources, the implementation of new energy concepts is essential for our future industrialized society. Renewable sources have to replace current energy technologies. This shift, however, will not be an easy task. In contrast to current nuclear or fossil power plants renewable energy sources in general do not offer a constant energy supply, resulting in a growing demand of energy storage. Furthermore, fossil fuels are both, energy source as well as energy carrier. This is of special importance for all mobile applications. Alternative energy carriers have to be found. The hydrogen technology is considered to play a crucial role in this respect. In fact it is the ideal means of energy storage for transportation and conversion of energy in a comprehensive clean-energy concept. Hydrogen can be produced from different feedstocks, ideally from water using regenerative energy sources. Water splitting can be achieved by electrolysis, solar thermo-chemical, photoelectrochemical or photobiological processes. Upon reconversion into energy, by using a fuel cell only water vapour is produced, leading to a closed energy cycle without any harmful emissions. Besides stationary applications, hydrogen is designated for mobile applications, e.g. for the zero-emission vehicle. In comparison to batteries hydrogen storage tanks offer the opportunity of fast recharging within a few minutes only and of higher storage densities by an order of magnitude. Hydrogen can be produced from renewable energies in times when feed-in into the electricity grid is not possible. It can be stored in large caverns underground and be utilized either to produce electricity and be fed into the electricity grid again or directly for mobile applications.

However, due to the very low boiling point of hydrogen (20.4 K at 1 atm) and its low density in the gaseous state ( $90 \text{ g/m}^3$ ) dense hydrogen storage, both for stationary and mobile applications, remains a challenging task. There are three major alternatives for hydrogen storage: compressed gas tanks, liquid hydrogen tanks as well as solid state hydrogen storage such as metal hydride hydrogen tanks. All of these three main techniques have their special advantages and disadvantages and are currently used for different applications. However, so far none of the respective tanks fulfils all the demanded technical requirements in terms of gravimetric storage density, volumetric storage density, safety,

free-form, ability to store hydrogen for longer times without any hydrogen losses, cyclability as well as recyclability and costs. Further research and development is strongly required. One major advantage of hydrogen storage in metal hydrides is the ability to store hydrogen in a very energy efficient way enabling hydrogen storage at rather low pressures without further need for liquefaction or compression. Many metals and alloys are able to absorb large amounts of hydrogen. The metal-hydrogen bond offers the advantage of a very high volumetric hydrogen density under moderate pressures, which is up to 60% higher than that of liquid hydrogen (Reilly & Sandrock, 1980).

Depending on the hydrogen reaction enthalpy of the specific storage material during hydrogen uptake a huge amount of heat (equivalent to 15% or more of the energy stored in hydrogen) is generated and has to be removed in a rather short time, ideally to be recovered and used as process heat for different applications depending on quantity and temperature. On the other side, during desorption the same amount of heat has to be applied to facilitate the endothermic hydrogen desorption process – however, generally at a much longer time scale. On one side this allows an inherent safety of such a tank system. Without external heat supply hydrogen release would lead to cooling of the tank and finally hydrogen desorption necessarily stops. On the other side it implies further restrictions for the choice of suitable storage materials. Highest energy efficiencies of the whole tank to fuel combustion or fuel cell system can only be achieved if in case of desorption the energy required for hydrogen release can be supplied by the waste heat generated in case of mobile applications on-board by the hydrogen combustion process and the fuel cell respectively.

## 2. Basics of hydrogen storage in metal hydrides

Many metals and alloys react reversibly with hydrogen to form metal hydrides according to the reaction (1):



Here, Me is a metal, a solid solution, or an intermetallic compound,  $\text{MeH}_x$  is the respective hydride and  $x$  the ratio of hydrogen to metal,  $x=c_{\text{H}}$  [H/Me], Q the heat of reaction. Since the entropy of the hydride is lowered in comparison to the metal and the gaseous hydrogen phase, at ambient and elevated temperatures the hydride formation is exothermic and the reverse reaction of hydrogen release accordingly endothermic. Therefore, for hydrogen release/desorption heat supply is required.

Metals can be charged with hydrogen using molecular hydrogen gas or hydrogen atoms from an electrolyte. In case of gas phase loading, several reaction stages of hydrogen with the metal in order to form the hydride need to be considered. Fig. 1 shows the process schematically.

The first attractive interaction of the hydrogen molecule approaching the metal surface is the Van der Waals force, leading to a physisorbed state. The physisorption energy is typically of the order  $E_{\text{Phys}} \approx 6 \text{ kJ/mol H}_2$ . In this process, a gas molecule interacts with several atoms at the surface of a solid. The interaction is composed of an attractive term, which diminishes with the distance of the hydrogen molecule and the solid metal by the power of 6, and a repulsive term diminishing with distance by the power of 12. Therefore, the potential energy of the molecule shows a minimum at approximately one molecular radius. In addition to hydrogen storage in metal hydrides molecular hydrogen adsorption is a second technique to store hydrogen. The storage capacity is strongly related to the temperature and the specific



surface areas of the chosen materials. Experiments reveal for carbon-based nanostructures storage capacities of less than 8 wt.% at 77 K and less than 1wt.% at RT and pressures below 100 bar (Panella et al., 2005; Schmitz et al., 2008).

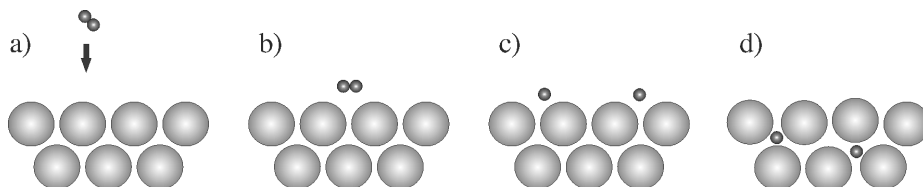


Fig. 1. Reaction of a  $H_2$  molecule with a storage material: a)  $H_2$  molecule approaching the metal surface. b) Interaction of the  $H_2$  molecule by Van der Waals forces (physisorbed state). c) Chemisorbed hydrogen after dissociation. d) Occupation of subsurface sites and diffusion into bulk lattice sites.

In the next step of the hydrogen-metal interaction, the hydrogen has to overcome an activation barrier for the formation of the hydrogen metal bond and for dissociation, see Fig. 1c and 2. This process is called dissociation and chemisorption. The chemisorption energy is typically in the range of  $E_{\text{Chem}} \approx 20 - 150 \text{ kJ/mol } H_2$  and thus significantly higher than the respective energy for physisorption which is in the order of  $4\text{-}6 \text{ kJ/mol } H_2$  for carbon based high surface materials (Schmitz et al., 2008).

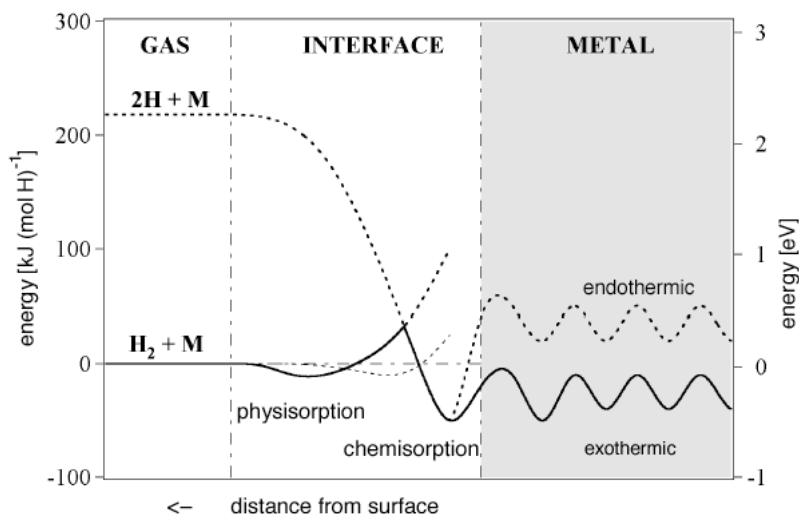


Fig. 2. Schematic of potential energy curves of hydrogen in molecular and atomic form approaching a metal. The hydrogen molecule is attracted by Van der Waals forces and forms a physisorbed state. Before diffusion into the bulk metal, the molecule has to dissociate forming a chemisorbed state at the surface of the metal (according to Züttel, 2003).

After dissociation on the metal surface, the H atoms have to diffuse into the bulk to form a M-H solid solution commonly referred to as  $\alpha$ -phase. In conventional room temperature metals / metal hydrides, hydrogen occupies interstitial sites - tetrahedral or octahedral - in the metal host lattice. While in the first, the hydrogen atom is located inside a tetrahedron formed by four metal atoms, in the latter, the hydrogen atom is surrounded by six metal atoms forming an octahedron, see Fig. 3.

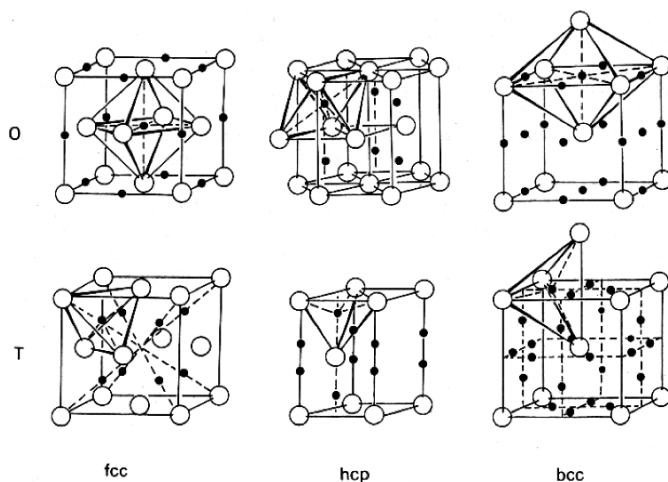


Fig. 3. Octahedral (O) and tetrahedral (T) interstitial sites in fcc-, hcp- and bcc-type metals. (Fukai, 1993).

In general, the dissolution of hydrogen atoms leads to an expansion of the host metal lattice of 2 to 3  $\text{\AA}^3$  per hydrogen atom, see Fig. 4. Exceptions of this rule are possible, e.g. several dihydride phases of the rare earth metals, which show a contraction during hydrogen loading for electronic reasons.

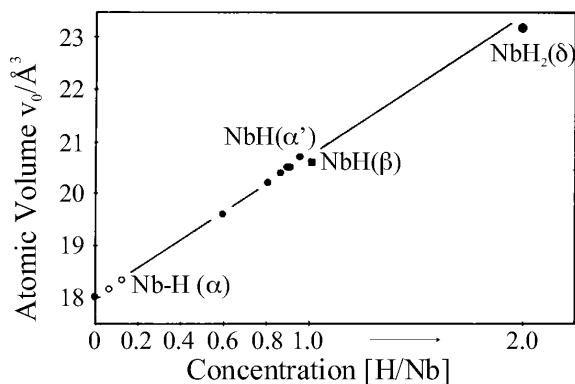


Fig. 4. Volume expansion of the Nb host metal with increasing H content. (Schober & Wenzl, 1978)

In the equilibrium the chemical potentials of the hydrogen in the gas phase and the hydrogen absorbed in the metal are the same:

$$\frac{1}{2} \mu_{gas} = \mu_{metal} \quad (2)$$

Since the internal energy of a hydrogen molecule is  $7/2 kT$  the enthalpy and entropy of a hydrogen molecule are

$$h_{gas} = \frac{7}{2} \cdot k \cdot T - E_{Diss} \quad (3)$$

and

$$s_{gas} = \frac{7}{2} \cdot k - k \cdot \ln \frac{p}{p_0(T)} \quad \text{with} \quad p_0(T) = \frac{8(\pi k T)^{\frac{7}{2}} \cdot M_{H-H}^{\frac{5}{2}} \cdot r_{H-H}^2}{h^5} \quad (4)$$

Here  $k$  is the Boltzmann constant,  $T$  the temperature,  $p$  the applied pressure,  $E_{Diss}$  the dissociation energy for hydrogen ( $E_{Diss} = 4.52 \text{ eV}$   $\text{eV}/\text{H}_2$ ),  $M_{H-H}$  the mass of the  $\text{H}_2$  molecule,  $r_{H-H}$  the interatomic distance of the two hydrogen atoms in the  $\text{H}_2$  molecule.

Consequently the chemical potential of the hydrogen gas is given by

$$\mu_{gas} = k \cdot T \cdot \ln \frac{p}{p_0(T)} - E_{Diss} = k \cdot T \cdot \ln \frac{p}{p_0} + \mu_{gas_0} \quad (5)$$

with  $p_0 = 1.01325 \cdot 10^5 \text{ Pa}$ .

In the solid solution ( $\alpha$ -phase) the chemical potential is accordingly

$$\mu_{\alpha} = h_{\alpha} - T s_{\alpha} \quad \text{mit} \quad s_{\alpha} = s_{\alpha,conf} + s_{\alpha,vibr,electr} \quad (6)$$

Here,  $s_{\alpha,conf}$  is the configuration entropy which is originating in the possible allocations of  $N_H$  hydrogen atoms on  $N_{is}$  different interstitial sites:

$$S_{\alpha,conf} = k \cdot \ln \frac{N_{is}!}{N_H!(N_{is} - N_H)!} \quad (7)$$

and accordingly for small  $c_H$  using the Stirling approximation we get

$$s_{\alpha,conf} = -k \cdot \ln \frac{c_H}{n_{is} - c_H} \quad (8)$$

with  $n_{is}$  being the number of interstitial sites per metal atom:  $n_{is} = N_{is}/N_{Me}$  and  $c_H$  the number of hydrogen atoms per metal atom:  $c_H = N_H/N_{Me}$ .

Therefore the chemical potential of hydrogen in the solid solution ( $\alpha$ -phase) is given by

$$\mu_{\alpha} = h_{\alpha} - T \cdot \left( s_{\alpha,vibr,electr} - k \cdot \ln \frac{c_H}{n_{is} - c_H} \right) \quad (9)$$

Taking into account the equilibrium condition (2) the hydrogen concentration  $c_H$  can be determined via

$$\frac{c_H}{n_{is} - c_H} = \sqrt{\frac{p}{p_0(T)}} \cdot e^{-\frac{\Delta G_s}{k \cdot T}} \quad \text{with} \quad \Delta G_s = h_\alpha - T \cdot s_{\alpha_{vibr}} + \frac{1}{2} \mu_{g_0} \quad (10)$$

or

$$\frac{c_H}{n_{is} - c_H} = \sqrt{\frac{p}{p_0(T)}} \cdot e^{-\frac{\Delta G_s}{R \cdot T}} \quad \text{with} \quad \Delta G_s = \Delta H_s - T \Delta S. \quad (11)$$

Here  $\mu_{g_0}$  is the chemical potential of the hydrogen molecule at standard conditions and  $R$  being the molar gas constant.

For very small hydrogen concentrations  $c_H|_{cH} \ll n_{is}$  in the solid solution phase  $\alpha$  the hydrogen concentration is directly proportional to the square root of the hydrogen pressure in the gas phase. This equation is also known as the **Sievert's law**, i.e.

$$c_H = \frac{1}{K_S} \sqrt{p} \quad (12)$$

with  $K_S$  being a temperature dependent constant. As the hydrogen pressure is increased, saturation occurs and the metal hydride phase  $MeH_{c_\beta}$  starts to form.

For higher hydrogen pressures/concentrations metal hydride formation occurs.

The conversion from the saturated solution phase to the hydride phase takes place at constant pressure  $p$  according to:



In the equilibrium the chemical potentials of the gas phase, the solid solution phase  $\alpha$  and the hydride phase  $\beta$  coincide:

$$\mu_\alpha(p, T, c_\alpha) = \mu_\beta(p, T, c_\beta) = \frac{1}{2} \mu_{gas}(p, T) = \frac{1}{2} \cdot k \cdot T \cdot \ln\left(\frac{p_{eq}(T)}{p_0}\right) + \frac{1}{2} \mu_{gas_0}. \quad (14)$$

Following Gibb's Phase Rule  $f=c-p+2$  with  $f$  being the degree of freedom,  $k$  being the number of components and  $p$  the number of different phases only one out of the four variables  $p, T, c_\alpha, c_\beta$  is to be considered as independent. Therefore for a given temperature all the other variables are fixed.

Therefore the change in the chemical potential or the Gibbs free energy is just a function of one parameter, i.e. the temperature  $T$ :

$$\Delta G = \frac{1}{2} \cdot R \cdot T \cdot \ln\left(\frac{p(T)}{p_0}\right). \quad (15)$$

From this equation follows the frequently-used **Van't Hoff equation** (16):

$$\frac{1}{2} \cdot \ln \frac{p}{p_0} = \frac{\Delta H}{RT} - \frac{\Delta S}{R} \quad (16)$$

The temperature dependent plateau pressure of this two phase field is the equilibrium dissociation pressure of the hydride and is a measure of the stability of the hydride, which commonly is referred to as  $\beta$ -phase.

After complete conversion to the hydride phase, further dissolution of hydrogen takes place as the pressure increases, see Fig. 5.

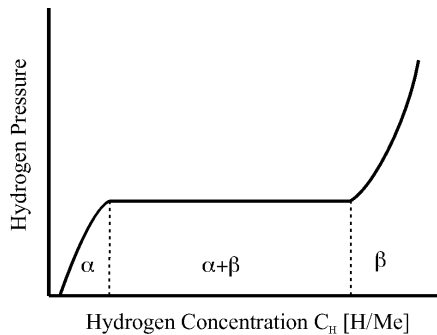


Fig. 5. Schematic Pressure/Composition Isotherm. The precipitation of the hydride phase  $\beta$  starts when the terminal solubility of the  $\alpha$ -phase is reached at the plateau pressure.

Multiple plateaus are possible and frequently observed in composite materials consisting of two hydride forming metals or alloys. The equilibrium dissociation pressure is one of the most important properties of a hydride storage material.

If the logarithm of the plateau pressure is plotted vs  $1/T$ , a straight line is obtained (van't Hoff plot) as seen in Fig. 6.

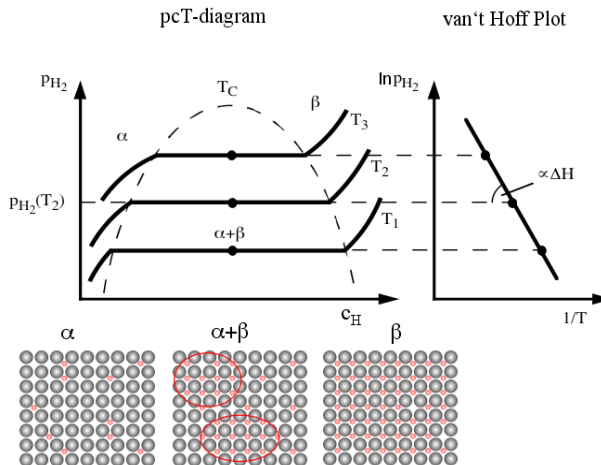


Fig. 6. Schematic  $pcT$ -diagram and van't Hoff plot. The  $\alpha$ -phase is the solid solution phase, the  $\beta$ -phase the hydride phase. Within the  $(\alpha - \beta)$  two phase region both the metal-hydrogen solution and the hydride phase coexist.

## 2.1 Conventional metal hydrides

Fig. 7 shows the Van't Hoff plots of some selected binary hydrides. The formation enthalpy of these hydrides  $H^0_f$  determines the amount of heat which is released during hydrogen absorption and consequently is to be supplied again in case of desorption. To keep the heat management system simple and to reach highest possible energy efficiencies it is necessary to store the heat of absorption or to get by the waste heat of the accompanying hydrogen utilizing process, e.g. energy conversion by fuel cell or internal combustion system. Therefore the reaction enthalpy has to be as low as possible. The enthalpy and entropy of the hydrides determine the working temperatures and the respective plateau pressures of the storage materials. For most applications, especially for mobile applications, working temperatures below 100°C or at least below 150°C are favoured. To minimize safety risks and avoid the use of high pressure composite tanks the favourable working pressures should be between 1 and 100 bar.

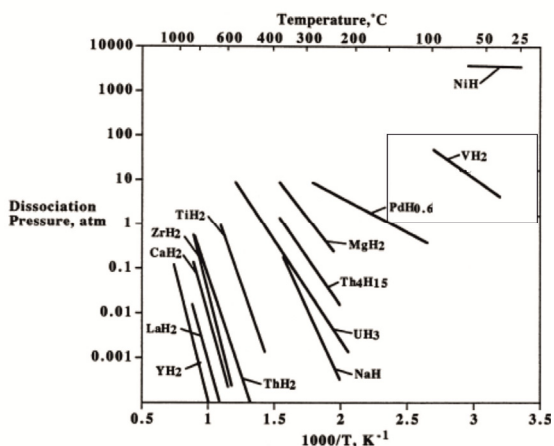


Fig. 7. Van't Hoff lines (desorption) for binary hydrides. Box indicates 1-100 atm, 0-100 °C ranges, taken from Sandrock et al. (Sandrock, 1999).

However, the Van't Hoff plots shown in Fig. 7 indicate that most binary hydrides do not have the desired thermodynamic properties. Most of them have rather high thermodynamic stabilities and thus release hydrogen at the minimum required pressure of 1 bar only at rather high temperatures ( $T > 300^\circ\text{C}$ ). The values of their respective reaction enthalpies are in the range of 75 kJ/(mol  $\text{H}_2$ ) ( $\text{MgH}_2$ ) or even higher. Typical examples are the hydrides of alkaline metals, alkaline earth metals, rare earth metals as well as transition metals of the Sc-, Ti- and V-group. The strongly electropositive alkaline metals like LiH and NaH and  $\text{CaH}_2$  form saline hydrides, i.e. they have ionic bonds with hydrogen.  $\text{MgH}_2$  marks the transition between these predominantly ionic hydrides and the covalent hydrides of the other elements in the first two periods.

Examples for high temperature hydrides releasing the hydrogen at pressures of 1 bar at extremely high temperatures ( $T > 700^\circ\text{C}$ ) are  $\text{ZrH}_2$  and  $\text{LaH}_2$  (Dornheim & Klassen, 2009).  $\text{ZrH}_2$  for example is characterized by a high volumetric storage density  $N_{\text{H}}$ .  $N_{\text{H}}$  values larger than  $7 \times 10^{22}$  hydrogen atoms per cubic centimetre are achievable. This value corresponds to

58 mol H<sub>2</sub>/l or 116 g/l and has to be compared with the hydrogen density in liquid hydrogen (20 K):  $4.2 \times 10^{22}$  (35 mol H<sub>2</sub>/l or 70 g/l) and in compressed hydrogen (350 bar / 700 bar):  $1.3 / 2.3 \times 10^{22}$  atoms/cm<sup>3</sup> ( 11 mol H<sub>2</sub>/l or 21 g/l and 19 mol H<sub>2</sub>/l or 38 g/l respectively) . The hydrogen density varies a lot between different hydrides. VH<sub>2</sub> for example has an even higher hydrogen density which amounts to  $11.4 \times 10^{22}$  hydrogen atoms per cubic centimetre and accordingly 95 mol H<sub>2</sub>/l or 190 g/l. As in the case of many other transition metal hydrides Zr has a number of different hydride phases ZrH<sub>2-x</sub> with a wide variation in the stoichiometry (Hägg, 1931). Their compositions extend from about ZrH<sub>1.33</sub> up to the saturated hydride ZrH<sub>2</sub>. Because of the limited gravimetric storage density of only about 2 wt.% and the negligibly low plateau pressure within the temperature range of 0 - 150 °C Zr as well as Ti and Hf are not suitable at all as a reversible hydrogen storage material. Thus, they are not useful for reversible hydrogen storage if only the pure binary hydrides are considered (Dornheim & Klassen, 2009). Libowitz et al. (Libowitz et al., 1958) could achieve a breakthrough in the development of hydrogen storage materials by discovering the class of reversible intermetallic hydrides. In 1958 they discovered that the intermetallic compound ZrNi reacts reversibly with gaseous hydrogen to form the ternary hydride ZrNiH<sub>3</sub>. This hydride has a thermodynamic stability which is just in between the stable high temperature hydride ZrH<sub>2</sub> ( $\Delta_f H^\circ = -169$  kJ/mol H<sub>2</sub>) and the rather unstable NiH ( $\Delta_f H^\circ = -8.8$  kJmol<sup>-1</sup>H<sub>2</sub>). Thus, the intermetallic Zr-Ni bond exerts a strong destabilizing effect on the Zr-hydrogen bond so that at 300°C a plateau pressure of 1bar is achieved which has to be compared to 900°C in case of the pure binary hydride ZrH<sub>2</sub>. This opened up a completely new research field. In the following years hundreds of new storage materials with different thermodynamic properties were discovered which generally follow the well-known semi-empirical rule of Miedema (Van Mal et al., 1974):

$$\Delta H(A_n B_m H_{x+y}) = \Delta H(A_n H_x) + \Delta H(B_m H_y) - \Delta H(A_n B_m) \quad (17)$$

Around 1970, hydrides with significantly lowered values of hydrogen reaction enthalpies, such as LaNi<sub>5</sub> and FeTi but also Mg<sub>2</sub>Ni were discovered. While 1300 °C are necessary to reach a desorption pressure of 2 bar in case of the pure high temperature hydride LaH<sub>2</sub>, in case of LaNi<sub>5</sub>H<sub>6</sub> a plateau pressure of 2 bar is already reached at 20 °C only. The value of the hydrogen reaction enthalpy is lowered to  $|\Delta H_{\text{LaNi}_5\text{H}_6}| = 30.9$  kJmol<sup>-1</sup>H<sub>2</sub>. The respective values for NiH are  $|\Delta H_{\text{NiH}}| = 8.8$  kJmol<sup>-1</sup>H<sub>2</sub> and  $P_{\text{diss,NiH,RT}}=3400$  bar.

In the meantime, several hundred other intermetallic hydrides have been reported and a number of interesting compositional types identified (table 1). Generally, they consist of a high temperature hydride forming element A and a non hydride forming element B, see fig. 8.

COMPOSITION	A	B	COMPOUNDS
A <sub>2</sub> B	Mg, Zr	Ni, Fe, Co	Mg <sub>2</sub> Ni, Mg <sub>2</sub> Co, Zr <sub>2</sub> Fe
AB	Ti, Zr	Ni, Fe	TiNi, TiFe, ZrNi
AB <sub>2</sub>	Zr, Ti, Y, La	V, Cr, Mn, Fe, Ni	LaNi <sub>2</sub> , YNi <sub>2</sub> , YMn <sub>2</sub> , ZrCr <sub>2</sub> , ZrMn <sub>2</sub> , ZrV <sub>2</sub> , TiMn <sub>2</sub>
AB <sub>3</sub>	La, Y, Mg	Ni, Co	LaCo <sub>3</sub> , YNi <sub>3</sub> , LaMg <sub>2</sub> Ni <sub>9</sub>
AB <sub>5</sub>	Ca, La, Rare Earth	Ni, Cu, Co, Pt, Fe	CaNi <sub>5</sub> , LaNi <sub>5</sub> , CeNi <sub>5</sub> , LaCu <sub>5</sub> , LaPt <sub>5</sub> , LaFe <sub>5</sub>

Table 1. Examples of intermetallic hydrides, taken from Dornheim et al. (Dornheim, 2010).

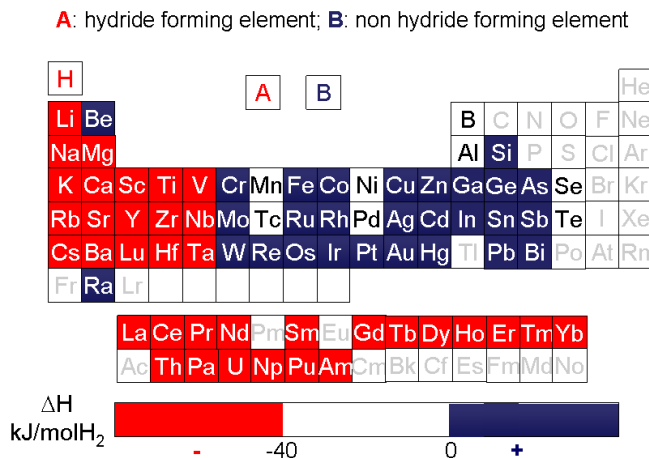


Fig. 8. Hydride and non hydride forming elements in the periodic system of elements.

Even better agreement with experimental results than by use of Miedema's rule of reversed stability is obtained by applying the semi-empirical band structure model of Griessen and Driessen (Griessen & Driessen, 1984) which was shown to be applicable to binary and ternary hydrides. They found a linear relationship of the heat of formation  $\Delta H = H^0_f$  of a metal hydride and a characteristic energy  $\Delta E$  of the electronic band structure of the host metal which can be applied to simple metals, noble metals, transition metals, actinides and rare earths:

$$\Delta H = \alpha \cdot \Delta E + \beta \quad (18)$$

with  $\Delta E = E_F - E_S$  ( $E_F$  being the Fermi energy and  $E_S$  the center of the lowest band of the host metal,  $\alpha = 59.24 \text{ kJ (eV mol H}_2\text{)}^{-1}$  and  $\beta = -270 \text{ kJ (mol H}_2\text{)}^{-1}$  and  $\Delta E$  in eV.

As described above, most materials experience an expansion during hydrogen absorption, wherefore structural effects in interstitial metal hydrides play an important role as well. This can be and is taken as another guideline to tailor the thermodynamic properties of interstitial metal hydrides. Among others Pourarian et al. (Pourarian, 1982), Fujitani et al. (Fujitani, 1991) and Yoshida & Akiba (Yoshida, 1995) report about this relationship of lattice parameter or unit cell volume and the respective plateau pressures in different material classes.

Intensive studies led to the discovery of a huge number of different multinary hydrides with a large variety of different reaction enthalpies and accordingly working temperatures. They are not only attractive for hydrogen storage but also for rechargeable metal hydride electrodes and are produced and sold in more than a billion metal hydride batteries per year. Because of the high volumetric density, intermetallic hydrides are utilized as hydrogen storage materials in advanced fuel cell driven submarines, prototype passenger ships, forklifts and hydrogen automobiles as well as auxiliary power units.



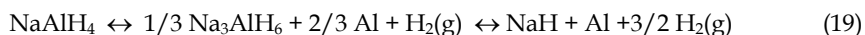
## 2.2 Hydrogen storage in light weight hydrides

Novel light weight hydrides show much higher gravimetric storage capacities than the conventional room temperature metal hydrides. However, currently only a very limited number of materials show satisfying sorption kinetics and cycling behaviour. The most prominent ones are magnesium hydride ( $\text{MgH}_2$ ) and sodium alanate ( $\text{NaAlH}_4$ ). In both cases a breakthrough in kinetics could be attained in the late 90s of the last century / the early 21<sup>st</sup> century.

Magnesium hydride is among the most important and most comprehensively investigated light weight hydrides.  $\text{MgH}_2$  itself has a high reversible storage capacity, which amounts to 7.6 wt.%. Furthermore, magnesium is the eighth most frequent element on the earth and thus comparably inexpensive. Its potential usage initially was hindered because of rather sluggish sorption properties and unfavourable reaction enthalpies. The overall hydrogen sorption kinetics of magnesium-based hydrides is as in case of all hydrides mainly determined by the slowest step in the reaction chain, which can often be deduced e.g. by modelling the sorption kinetics (Barkhordarian et al, 2006; Dornheim et al., 2006). Different measures can be taken to accelerate kinetics. One important factor for the sorption kinetics is the micro- or nanostructure of the material, e.g. the grain or crystallite size. Because of the lower packing density of the atoms, diffusion along grain boundaries is usually faster than through the lattice. Furthermore, grain boundaries are favourable nucleation sites for the formation and decomposition of the hydride phase. A second important parameter is the outer dimension of the material, e.g. in case of powdered material, its particle size. The particle size (a) determines the surface area, which is proportional to the rate of the surface reaction with the hydrogen, and (b) is related to the length of the diffusion path of the hydrogen into and out of the volume of the material. A third major factor by which hydrogen sorption is improved in many hydrogen absorbing systems is the use of suitable additives or catalysts. In case of  $\text{MgH}_2$  it was shown by Oelerich et al. (Oelerich et al., 2001; Dornheim et al., 2007) that already tiny amounts of transition metal oxides have a huge impact on the kinetics of hydrogen sorption. Using such additives Hanada et al. (Hanada et al., 2007) could show that by using such additives hydrogen uptake in Mg is possible already at room temperature within less than 1 min. The additives often do not just have one single function but multiple functions. Suitable additives can catalyze the surface reaction between solid and gas. Dispersions in the magnesium-based matrix can act as nucleation centres for the hydride or the dehydrogenated phase. Furthermore, different additives, such as liquid milling agents and hard particles like oxides, borides, etc. , can positively influence the particle size evolution during the milling process (Pranzas et al., 2006; Pranzas et al., 2007; Dornheim et al, 2007) and prevent grain i.e. crystallite growth. More detailed information about the function of such additives in  $\text{MgH}_2$  is given in (Dornheim et al., 2007). Beyond that, a preparation technique like high-energy ball milling affects both the evolution of certain particle sizes as well as very fine crystallite sizes in the nm range and is also used to intermix the hydride and the additives/catalysts. Thus, good interfacial contact with the light metal hydride as well as a fine dispersion of the additives can be achieved.

As in case of  $\text{MgH}_2$  dopants play also an important role in the sorption of Na-Al-hydride, the so-called Na-alanate. While hydrogen liberation is thermodynamically favorable at moderate temperatures, hydrogen uptake had not been possible until in 1997 Bogdanovic et al. demonstrated that mixing of  $\text{NaAlH}_4$  with a Ti-based catalyst leads to a material, which can be reversibly charged with hydrogen (Bogdanovic, 1997). By using a tube vibration mill

of Siebtechnik GmbH Eigen et al. (Eigen et al., 2007; Eigen et al., 2008) showed that upscaling of material synthesis is possible: After only 30 min milling under optimised process conditions in such a tube vibration mill in kg scale, fast absorption and desorption kinetics with charging/discharging times of less than 10 min can be obtained. The operation temperatures of this complex hydride are much lower than compared to  $\text{MgH}_2$  and other light weight hydrides. Fast kinetics is achieved at 100 °C to 150 °C which is much less than what is required in case of  $\text{MgH}_2$ , however, still significantly higher than in case of the conventional hydrides which show only a very limited storage capacity. Such hydride working temperatures offer the possibility for combinations of metal hydride tanks based on these complex hydrides with e. g. combustion engines, high temperature PEM fuel cells or other medium to high temperature fuel cells. However, compared to  $\text{MgH}_2$  the gravimetric hydrogen storage capacity is significantly reduced. Having a maximum theoretical storage capacity of about 5.6 wt. %  $\text{NaAlH}_4$  exhibits a long term practical storage capacity of 3.5-4.5 wt. %  $\text{H}_2$  only. Furthermore, in difference to  $\text{MgH}_2$   $\text{NaAlH}_4$  decomposes in two reaction steps upon dehydrogenation which implies two different pressure plateaus instead of just one:



The first decomposition step has an equilibrium pressure of 0.1 MPa at 30 °C, the second step at about 100 °C (Schüth et al., 2004). A maximum of 3.7 wt.%  $\text{H}_2$  can be released during the first desorption step, 5.6 wt.% in total. The remaining hydrogen bonded to Na is technically not exploitable due to the high stability of the respective hydride.

While the reaction kinetics was optimized significantly, the desorption enthalpy of  $\text{NaAlH}_4$  of 37 kJ/mol $\text{H}_2$  and  $\text{Na}_3\text{AlH}_6$  of 47 kJ/mol  $\text{H}_2$  respectively remains a challenge. For many applications even this value which is much below that of  $\text{MgH}_2$  is still too large.

### 3. Tailoring thermodynamics of light weight metal hydrides

While there are plenty of known hydrides with suitable thermodynamics for hydrogen uptake and release at ambient conditions (several bar equilibrium pressure at or nearby room temperature) currently no hydride is known which combines suitable thermodynamics and kinetics with such a high gravimetric storage capacity that a hydrogen storage tank based on such a material could compete with a 700 bar compressed composite vessel in regard to weight. Depending on the working temperature and pressure as well as the reversible gravimetric storage capacity of the selected hydride the achievable capacity of a metal hydride based storage tank is usually better than half of the capacity of the metal hydride bed itself (Buchner & Povel, 1982). Since modern composite pressurized gas tanks meanwhile show gravimetric hydrogen storage capacities of around 4 wt.% according to conservative extrapolations the possible choice of hydrides should be limited to those having the ability to reversibly store at least 6 wt.% $\text{H}_2$ . All currently known high capacity hydrides, however, show either too small values of the respective reaction enthalpy and are therefore not reversible or would require several thousand bar hydrogen pressure or alternatively electrochemical loading or on the other hand are too stable and have an equilibrium pressure which around room temperature is much below the required pressures. The value of reaction enthalpy aimed at is between 20 and 30 kJ/mol  $\text{H}_2$ . Fig. 9 shows the potentially available hydrogen content of some well known hydrides plotted against their hydrogen reaction enthalpies.

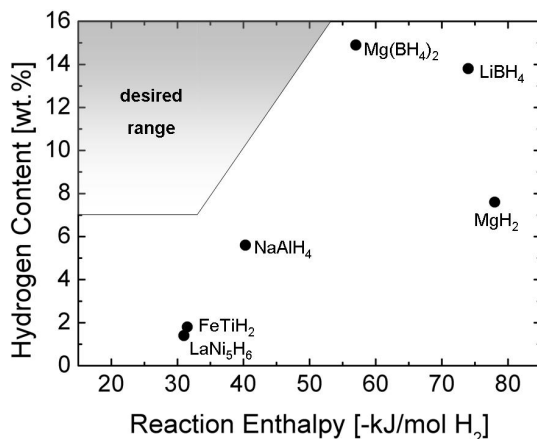


Fig. 9. Theoretically achievable reversible storage capacities and reaction enthalpies of selected hydrides.  $\text{LaNi}_5\text{H}_6$  and  $\text{FeTiH}_2$  are taken as examples for conventional room temperature hydrides. The reaction enthalpies and achievable hydrogen storage capacities are  $\Delta H = -31$  kJ/mol  $\text{H}_2$ ,  $C_{\text{H,max}} = 1.4$  wt.% for  $\text{LaNi}_5\text{H}_6$  and for the Fe-Ti system  $\Delta H = -31.5$  kJ/mol  $\text{H}_2$ ,  $C_{\text{H,max}} = 1.8$  wt.% (average over two reaction steps with  $\Delta H(\text{FeTiH}_2) = -28$  kJ/mol  $\text{H}_2$  and  $\Delta H(\text{FeTiH}) = -35$  kJ/mol  $\text{H}_2$  respectively) (Buchner, 1982). The respective values for  $\text{NaAlH}_4$  are  $\Delta H = -40.5$  kJ/mol  $\text{H}_2$ ,  $C_{\text{H,max}} = 5.6$  wt.% (average over two reaction steps with  $\Delta H(\text{NaAlH}_4) = -37$  kJ/mol  $\text{H}_2$  and  $\Delta H(\text{NaAl}_3\text{H}_6) = -47$  kJ/mol  $\text{H}_2$  (Bogdanovic et al., 2009)), for  $\text{MgH}_2$ :  $\Delta H = -78$  kJ/mol  $\text{H}_2$  (Oelerich, 2000) and  $C_{\text{H,max}} = 7.6$  wt.%, for  $\text{LiBH}_4$ :  $\Delta H = -74$  kJ/mol  $\text{H}_2$  (Mauron, 2008) and  $C_{\text{H,max}} = 7.6$  wt.%, for  $\text{Mg}(\text{BH}_4)_2$ :  $\Delta H = -57$  kJ/mol  $\text{H}_2$  (Li, 2008) and  $C_{\text{H,max}} = 14.9$  wt.%.

As shown in Fig. 9 none of the plotted hydrides, neither the conventional room temperature hydrides with their rather low gravimetric capacity nor the sophisticated novel chemical hydrides with their unsuitable reaction enthalpy, show the desired combination of properties. Therefore the tailoring of the thermodynamic properties of high capacity light weight and complex hydrides is a key issue, an imperative for future research in the area of hydrides as hydrogen storage materials.

### 3.1 Thermodynamic tuning of single phase light weight hydrides

The traditional way of tailoring the thermodynamic properties of metal hydrides is by formation of alloys with different stabilities as described in chapter 2.1. Thereby the value of reaction enthalpy can be reduced by stabilising the dehydrogenated state and/or destabilising of the hydride state, see Fig. 10 a. Accordingly, the total amount of reaction enthalpy is increased by destabilising the dehydrogenated state and/or stabilising the hydride, see Fig. 10 b.

This approach has been successfully applied to light weight metal hydrides also.

#### Mg-based hydrides

One of the first examples using this approach for tuning the thermodynamic properties of light weight metal hydrides was the discovery of the Mg-Ni -system as potential hydrogen storage system by Reilly and Wiswall (Reilly & Wiswall, 1968).  $\text{Mg}_2\text{Ni}$  has a negative heat of

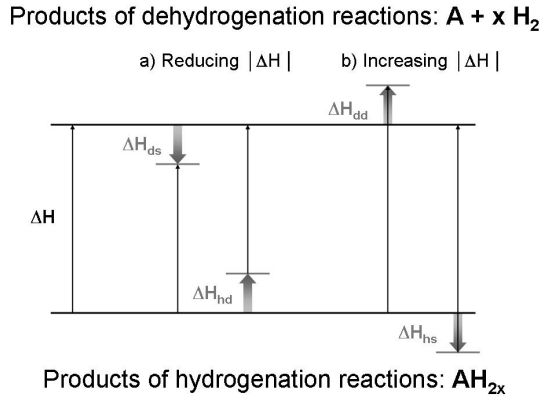


Fig. 10. Tailoring of the reaction enthalpy by altering the stability of the hydrogenated or dehydrogenated state of the metal hydrides: a) Reduction of total reaction enthalpy by stabilising the dehydrogenated phase by  $\Delta H_{ds}$  or destabilising the hydride phase by  $\Delta H_{hd}$ . b) Increase of total reaction enthalpy by destabilising the dehydrogenated state by  $\Delta H_{dd}$  or stabilising the hydrogenated state by  $\Delta H_{hs}$ .

formation of  $H^0(Mg_2Ni) = -42$  kJ/mol. Therefore, compared to pure Mg the dehydrogenated state is stabilised by  $\Delta H_{ds} = -21$  kJ/(mol Mg). The enthalpy of formation of  $Mg_2NiH_4$  is  $H^0_f(Mg_2NiH_4) = -176$  kJ/mol (=  $-88$  kJ/(mol Mg)), wherefore the hydride phase is stabilised by  $\Delta H_{hd} = -10$  kJ/(mol Mg) if compared to pure  $MgH_2$ . In total the hydrogen reaction enthalpy of  $Mg_2Ni$

$$|\Delta H_{Mg_2Ni-H}| = |\Delta H_{Mg-H}| + \Delta H_{ds} - \Delta H_{hd} \quad (20)$$

is reduced by 11 kJ/mol  $H_2$  to about  $|\Delta H(Mg_2Ni-H)| = 67$  kJ/mol  $H_2$ . While pure  $MgH_2$  exhibits a hydrogen plateau pressure of 1 bar around 300 °C, in case of  $Mg_2NiH_4$  such a plateau pressure is reached already at around 240 °C and in case of further alloying and substituting Ni by Cu at around 230 °C in  $Mg_2Ni_{0.5}Cu_{0.5}$  (Klassen et al., 1998). Unfortunately, the gravimetric storage capacity of  $Mg_2NiH_4$  is reduced to 3.6 wt.%  $H_2$  only and thus is less than half the respective value in the  $MgH_2$  system. Darnaudery et al. (Darnaudery et al., 1983) were successful to form several quaternary hydrides by hydrogenating  $Mg_2Ni_{0.75}M_{0.25}$  with different 3d elements  $M \in \{V, Cr, Fe, Co \text{ and } Zn\}$  showing stabilities very similar  $Mg_2NiH_4$ .

Increasing the amount of 3d metals Tsushio et al. (Tsushio et al., 1998) investigated the hydrogenation of  $MgNi_{0.86}M_{0.03}$  with  $M \in \{Cr, Fe, Co, Mn\}$ . Consequently, they observed a dramatic decrease in hydrogen storage capacity to 0.9 wt.% and in hydrogen reaction enthalpy which amounts to 50 kJ/(mol  $H_2$ ) for  $MgNi_{0.86}Cr_{0.03}$ . This reaction enthalpy value is in very good agreement with the value 54 kJ/(mol  $H_2$ ) given by Orimo et al. for amorphous  $MgNi$  (Orimo et al., 1998).

Lowering even more the content of Mg Terashita et al. (Terashita et al., 2001) found  $(Mg_{1-x}Ca_x)Ni_2$  based alloys desorbing hydrogen at room temperature. They determined the hydride formation enthalpy and entropy of  $(Mg_{0.68}Ca_{0.32})Ni_2$  to be  $H = -37$  kJ/(mol  $H_2$ ) and  $S = -94$  J/(mol  $H_2$  K) respectively, which is already quite near to the envisioned target.

Unfortunately, with lowering the Mg content the hydrogen storage capacity dropped down to 1.4 wt.% only.

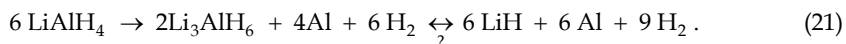
On the other side, as schematically shown in Fig. 10b the absolute value of reaction enthalpy can be increased by either stabilising the hydride phase or destabilising the dehydrogenated phase. In case of Mg-based hydrogen absorbing alloys this is not at all of interest for hydrogen storage itself since  $\text{MgH}_2$  is too stable for most hydrogen storage applications, however, this is of interest for other applications like the storage of thermal energy (Dornheim & Klassen, 2009).  $\text{Mg}_2\text{FeH}_6$  is an example of such materials with increased amount of reaction enthalpy. Furthermore, it is the one with the highest known volumetric hydrogen density which amounts to  $150 \text{ kg m}^{-3}$ . This enormously high hydrogen density is more than double the value found in case of liquid hydrogen at 20 K and moderate pressures of up to 20 bar (Klell, 2010). The gravimetric storage capacity is 5.6 wt.% and thus still rather high. Since Mg and Fe are immiscible the dehydrogenated state is destabilised compared to pure Mg:  $\Delta H_{\text{dd}} > 0 \text{ kJ}/(\text{mol H}_2)$ . Accordingly the hydride phase is more difficult to be synthesised and reversibility as well as long term stability is more difficult to be accomplished.

Nevertheless, hydrogenation is possible at hydrogen pressures of at least 90 bar and temperatures of at least  $450 \text{ }^\circ\text{C}$  (Selvam & Yvon, 1991). Bogdanovic et al. (Bogdanovic et al., 2002) achieved very good reversibility and cycling stability with the hydrogen storage capacities remaining unchanged throughout 550-600 cycles at a level of 5-5.2 wt.%  $\text{H}_2$ . The reaction enthalpy value is reported to be in between  $77 \text{ kJ}/(\text{mol H}_2)$  and  $98 \text{ kJ}/(\text{mol H}_2)$  (Bogdanovic et al., 2002), (Konstanchuk et al, 1987), (Puszkiewicz et al., 2008), (Didisheim et al., 1984).

The large reaction enthalpies of  $\text{MgH}_2$  and  $\text{Mg}_2\text{FeH}_6$  lead to weight and volume related heat storage densities in the temperature range of  $500 \text{ }^\circ\text{C}$  which are many times higher than that of the possible sensible or latent heat storage materials (Bogdanovic et al., 2002). The calculated and experimental heat storage densities to weight given by Bogdanovic et al. are  $2814 \text{ kJ}/\text{kg}$  and  $2204 \text{ kJ}/\text{kg}$  for the  $\text{MgH}_2$ -Mg system and  $2106$  and  $1921 \text{ kJ}/\text{kg}$  for the  $\text{Mg}_2\text{FeH}_6 - 2\text{Mg} + \text{Fe}$  system respectively. The corresponding calculated and experimental values for the volumetric thermal energy storage density are  $3996 \text{ kJ}/\text{dm}^3$  and  $1763 \text{ kJ}/\text{dm}^3$  for the  $\text{MgH}_2$ -Mg system and  $5758 \text{ kJ}/\text{dm}^3$  and  $2344 \text{ kJ}/\text{dm}^3$  respectively (Bogdanovic et al., 2002). These thermal energy densities ought not to be mistaken with the energy stored in the hydrogen (lower heating value) which is more than a factor of three larger.

### Aluminum-based complex hydrides

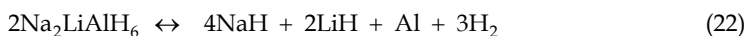
As  $\text{Mg}_2\text{FeH}_6$  decomposes during hydrogen release into 2 Mg, Fe and 3  $\text{H}_2$   $\text{LiAlH}_4$  decomposes during hydrogen release in  $1/3 \text{ Na}_3\text{AlH}_6 + 2/3 \text{ Al} + \text{H}_2$  and finally  $\text{NaH} + \text{Al} + 3/2 \text{ H}_2$ . As written in chapter 2.2 while much lower than those of the Mg-based hydrides the reaction enthalpies of  $|\Delta H| = 37 \text{ kJ}/(\text{mol H}_2)$  and  $|\Delta H| = 47 \text{ kJ}/(\text{mol H}_2)$  are still too high for many applications especially for the usage in combination with low temperature PEM fuel cells.  $\text{LiAlH}_4$  on the other hand is much less stable. It decomposes in two steps as is the case of the  $\text{NaAlH}_4$ :



The first reaction step, however, the decomposition of  $\text{LiAlH}_4$  is found to be exothermic with  $\Delta H_{\text{decomposition}} = -10 \text{ kJ}/(\text{mol H}_2)$ . Since the entropy of decomposition is positive.

Rehydrogenation is not possible at all. The second reaction step, the decomposition of  $\text{Li}_3\text{AlH}_6$  is endothermic with  $\Delta H_{\text{decomposition}} = 25 \text{ kJ}/(\text{mol H}_2)$ . The decomposition of  $\text{LiH}$  itself takes place at much higher temperatures with  $\Delta H = 140 \text{ kJ}/(\text{mol H}_2)$  (Orimo et al., 2007). While the second reaction step, the decomposition of  $\text{Li}_3\text{AlH}_6$  and rehydrogenation of  $\text{LiH} + \text{Al}$  shows rather suitable thermodynamic properties, sluggish kinetics prevent this system so far from being used for hydrogen storage.

To increase the storage capacity and tailor the reaction enthalpy of the  $\text{NaAlH}_4$  system it is a comprehensible approach to replace some of the Na by Li. Indeed Huot et al. (Huot et al., 1999) proved the existence of  $\text{Na}_2\text{LiAlH}_6$  and the possible formation by high energy ball-milling of  $\text{NaH} + \text{LiH} + \text{NaAlH}_4$ . Reversible hydrogen sorption is found to be possible in the Na-Li-Al-H system according to the following reaction:



As in case of the pure Na-Al-H system and the Li-Al-H system kinetics can be improved by the addition of transition metal compounds like metal oxides, chlorides and fluorides, see (Ares Fernandez et al., 2007), (Ma et al., 2005) and (Martinez-Franco et al., 2010). However, due to the lack of any stable compound in the dehydrogenated state and the formation of a rather stable hydride the value of reaction enthalpy isn't decreased but increased if compared to the original single Na and Li based aluminium hydrides. Fossdal et al. (Fossdal et al., 2005) has determined the pressure-composition isotherms of  $\text{TiF}_3$ -doped  $\text{Na}_2\text{LiAlH}_6$  in the temperature range of  $170 \text{ }^\circ\text{C} - 250 \text{ }^\circ\text{C}$ . They determined the dissociation enthalpy and the corresponding entropy from the Van't Hoff plot:  $|\Delta H| = 56 \text{ kJ}/(\text{mol H}_2)$  and  $S = 138 \text{ J}/(\text{K mol H}_2)$ . Therefore, instead of a lowering the heat of reaction the opposite is observed. The heat of reaction of the hexa-hydride phase is increased by about  $10 \text{ kJ}/(\text{mol H}_2)$  if compared to the pure  $\text{Na}_3\text{AlH}_6$  hydride phase.

In 2007 Yin et al. (Yin et al., 2007) presented DFT calculations about the doping effects of  $\text{TiF}_3$  on  $\text{Na}_3\text{AlH}_6$ . Their calculations suggested F- substitution for the H-anion leading to a reduction of the desorption enthalpy and therefore for a favourable thermodynamic modification of the  $\text{Na}_3\text{AlH}_6$  system which was experimentally confirmed by Brinks et al. (Brinks et al., 2008) and Eigen et al. (Eigen et al., 2009).

### Borohydrides

Only a very few hydrides show a higher gravimetric storage capacity than  $\text{MgH}_2$ . For this they must be composed from very light elements. Knowing that Al-containing compounds can form reversible complex metal hydrides it is a reasonable approach to look for Boron-containing compounds as reversible hydrogen storage materials with even higher storage capacity. Borohydrides are known since 1940 when Schlesinger and Brown report about the successful synthesis of  $\text{LiBH}_4$  by reaction of  $\text{LiEt}$  and diborane (Schlesinger & Brown, 1940). Despite the early patent from Goerrig in 1958 (Goerrig, 1960) direct synthesis from gaseous  $\text{H}_2$  was not possible for long times. Until in 2004 three different groups from the USA (Vajo et al., 2005), South Korea (Cho et al., 2006) and Germany (Barkhordarian et al., 2007) independently discovered that by using  $\text{MgB}_2$  instead of pure Boron as starting material formation of the respective borohydrides occurs at rather moderate conditions of  $5 \text{ MPa H}_2$  pressure. Orimo et al. (Orimo et al., 2005) reports on the rehydrogenation of previously dehydrogenated  $\text{LiBH}_4$  at  $35 \text{ MPa H}_2$  pressure at  $600 \text{ }^\circ\text{C}$ . Mauron et al. (Mauron et al., 2008) report that rehydrogenation is also possible at  $15 \text{ MPa}$ . As in case of the Mg-based alloys

and the aluminum hydrides the reaction enthalpy of many borohydrides is rather unsuitable for most applications.  $\text{LiBH}_4$  as one of the most investigated borohydrides with a very high gravimetric hydrogen density of 18.5 wt.% shows an endothermic desorption enthalpy of  $|\text{DH}| = 74 \text{ kJ}/(\text{mol H}_2)$  (Mauron et al., 2008) which is almost the same as in  $\text{MgH}_2$ . Therefore the tailoring of the reaction enthalpy by substitution is a key issue for these materials as well. As in case of the aluminium hydrides there are two different possibilities for substitution in these complex hydrides: cation substitution and anion substitution. Nakamori et al. (Nakamori et al., 2006) reports about a linear relationship between the heat of formation  $\Delta H_{\text{boro}}$  of  $\text{M}(\text{BH}_4)_n$  determined by first principle methods and the Pauling electronegativity of the cation  $\chi_P$ :

$$\frac{\Delta H_{\text{boro}}}{\text{kJ}(\text{mol BH}_4)^{-1}} = 248.7 \chi_P - 390.8 \quad (23)$$

Aiming to confirm their theoretical results the same group performed hydrogen desorption experiments which show that the experimentally determined desorption temperature  $T_d$  shows correlates with the Pauling electronegativity  $\chi_P$  as well, see Fig. 11.

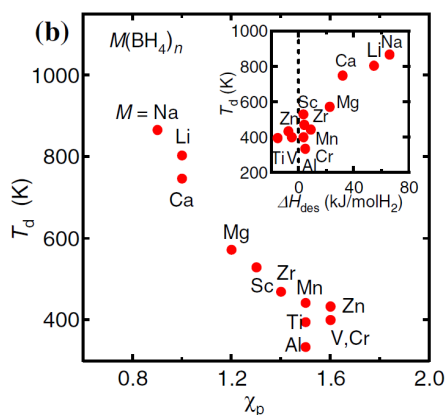


Fig. 11. The desorption temperature  $T_d$  as a function of the Pauling electronegativity  $\chi_P$  and estimated desorption enthalpies  $\Delta H_{\text{des}}$  (Nakamori et al., 2007).

Based on these encouraging results several research groups started to investigate the partial substitution of one cation by another studying several bialkali metal borohydrides. The decomposition temperature of the bialkali metal borohydrides like  $\text{LiK}(\text{BH}_4)_2$  is approximately the average of the decomposition temperature of the mono alkali borohydrides (Rude et al., 2011). Investigations of Li et al. (Li et al., 2007) and Seballos et al. (Seballos et al., 2009) confirmed that this correlation between desorption enthalpy / observed  $T_d$  holds true for many double cation  $\text{MM}'(\text{BH}_4)_n$  systems, see Fig. 12.

Several experiments are indicating that transition metal fluorides are among the best additives for borohydrides (Bonatto Minella et al., 2011). While for some cases the function of the transition metal part as additive is understood (Bösenberg et al., 2009; Bösenberg et al., 2010; Deprez et al., 2010; Deprez et al., 2011), the function of F so far remained unclear. DFT calculations performed by Yin et al. (Yin et al., 2008) suggest a favourable modification

of hydrogen reaction enthalpy in the  $\text{LiBH}_4$  system by substitution of the H-ion with the F-ion. However, no clear indicative experimental results for F- substitution in borohydrides are found yet. In contrast to the F the heavier and larger halides Cl, Br, I are found to readily substitute in some borohydrides for the  $\text{BH}_4$ -ion and form solid solutions or stoichiometric compounds and are so far reported to stabilize the hydride phase leading to an increase of the desorption enthalpy  $|\Delta H|$  (Rude et al., 2011).

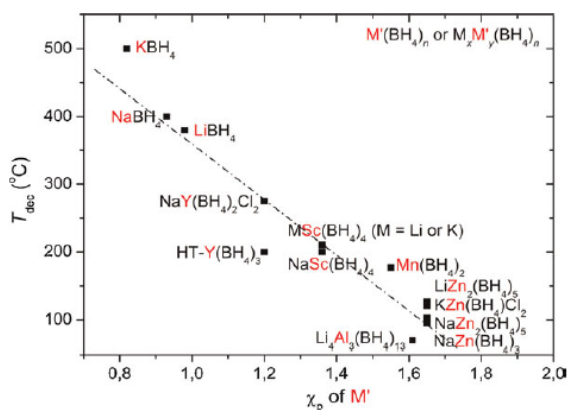
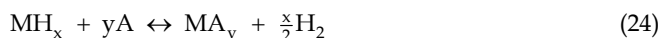


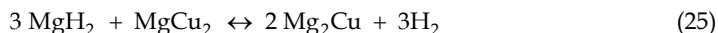
Fig. 12. Decomposition temperatures,  $T_{\text{dec}}$  for metal borohydrides plotted as a function of the electronegativity of the metal,  $M'$ . (Rude et al., 2011)

### 3.2 Thermodynamic tuning using multicomponent systems: reactive additives and reactive hydride composites

In 1967 Reilly and Wiswall (Reilly & Wiswall, 1967) found another promising approach to tailor reaction enthalpies of hydrides ( $\text{MH}_x$ ) by mixing them with suitable reactants (A):

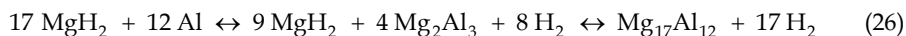


They investigated the system  $\text{MgH}_2/\text{MgCu}_2$  which reversibly reacts with hydrogen according to:



The formation of  $\text{MgCu}_2$  from  $\text{Mg}_2\text{Cu}$  and Cu is exothermic and thus counteracts the endothermic release of hydrogen. Thereby, the total amount of hydrogen reaction enthalpy is reduced to roughly  $|\Delta H| = 73 \text{ kJ}/(\text{mol H}_2)$  (Wiswall, 1978). The equilibrium temperature for 1 bar hydrogen pressure is reduced to about 240 °C. In spite of the lower driving force for rehydrogenation,  $\text{Mg}_2\text{Cu}$  is much more easily hydrogenated than pure Mg. A fact found in many other systems like the Reactive Hydride Composites as well.

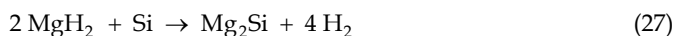
Aluminum is another example of a reactive additive for  $\text{MgH}_2$ . The reaction occurs via two steps (Bouaricha et al., 2000):





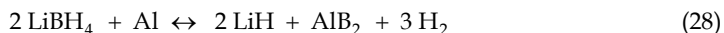
The system can reversibly store 4.4 wt.% H<sub>2</sub>. Since the formation enthalpy  $\Delta H_{\text{Form}}$  of Mg<sub>17</sub>Al<sub>12</sub> is -102 kJ/mol the total value of reaction enthalpy of reaction (26) is reduced by ~ 6 kJ/(mol H<sub>2</sub>) if compared to pure MgH<sub>2</sub>. An equilibrium pressure of 1 bar is reached at around 240 °C again.

To further decrease the reaction enthalpy of a Mg-based system a much more stable compound would have to be formed during dehydrogenation. A system investigated by many groups is the MgH<sub>2</sub>-Si system. Mg<sub>2</sub>Si has an enthalpy of formation of  $\Delta H_{\text{Form}} = -79$  kJ/mol. Due to the formation of Mg<sub>2</sub>Si the value of reaction enthalpy of MgH<sub>2</sub>/Si should therefore be reduced by 37 kJ/(mol H<sub>2</sub>) to about  $|\Delta H| = 41$  kJ/(mol H<sub>2</sub>) (Dornheim, 2010). Theoretically 5 wt.% H<sub>2</sub> can be stored via the reaction



The thermodynamic data indicate a very favourable equilibrium pressure of about 1 bar at 20 °C and 50 bar at 120 °C (Vajo, 2004). While so far rehydrogenation of Mg<sub>2</sub>Si was not shown to be possible the system LiH-Si turned out to be reversible. The enthalpy of dehydrogenation of LiH being 190 kJ/(mol H<sub>2</sub>) an equilibrium H<sub>2</sub> pressure of 1 bar is reached at 910 °C (Sangster, 2000; Dornheim, 2010). LiH reversibly reacts with Si via a two step reaction with the equilibrium pressure being more than 10<sup>4</sup> times higher and the dehydrogenation enthalpy being reduced by 70 kJ/(mol H<sub>2</sub>) (Vajo, 2004).

This approach has recently also been applied to borohydrides. According to Cho et al. (Cho et al., 2006) the decomposition temperature of pure LiBH<sub>4</sub> is determined by CALPHAD to 1 bar H<sub>2</sub> pressure at 403 °C while the corresponding equilibrium temperature for the reaction

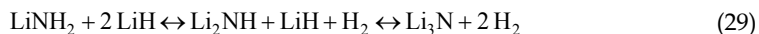


is reduced to 188 °C. Kang et al. (Kang et al., 2007) and Jin et al. (Jin et al., 2008) could show that this system indeed is reversible if suitable additives are used.

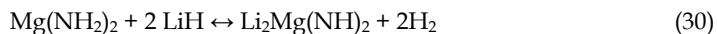
The only disadvantage of this approach is that the total reversible storage capacity per weight is reduced if something is added to the hydrogen storing material which contains no hydrogen.

The problem of reduced hydrogen capacity by using reactive additives has recently overcome by the approach of the Reactive Hydride Composites (Dornheim, 2006). Thereby, different high capacity hydrogen storage materials are combined which react exothermically with each other during decomposition, see Fig. 13.

One of the first examples of such a system is the LiNH<sub>2</sub>-LiH system which was discovered by Chen et al. (Chen et al., 2002):



However, the value of reaction enthalpy is  $|\Delta H| = 80$  kJ/(mol H<sub>2</sub>) and therefore for most applications still much too high. In contrast the system



shows a much more suitable desorption enthalpy of  $|\Delta H| \sim 40$  kJ/(mol H<sub>2</sub>) with an expected equilibrium pressure of 1 bar at approximately 90 °C (Xiong et al., 2005; Dornheim, 2010).

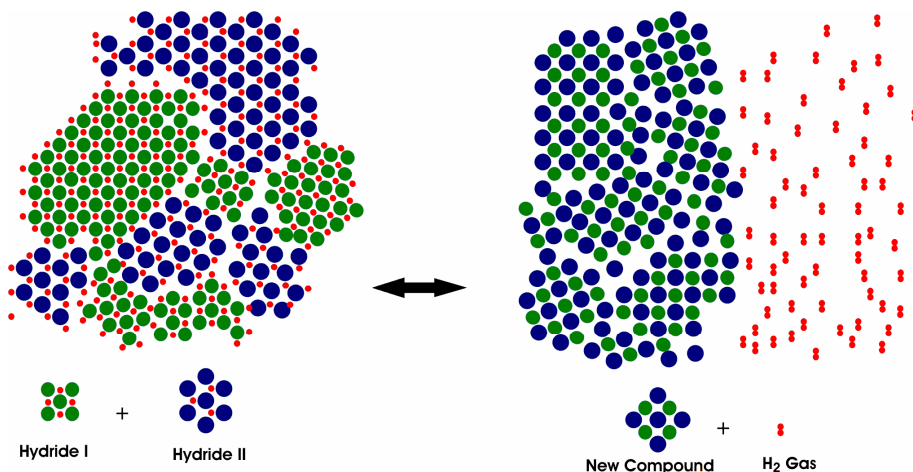
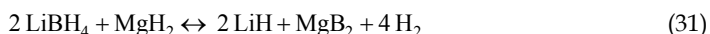


Fig. 13. Schematic of the reaction mechanism in Reactive Hydride Composite.

In 2004 Vajo et al. (Vajo et al., 2005), Cho et al. (Cho et al., 2006) and Barkhordarian et al. (Barkhordarian et al., 2007) independently discovered that the usage of borides especially  $\text{MgB}_2$  as a starting material facilitates the formation of different borohydrides. This finding initiated the development and investigation of several new reversible systems with high storage capacities of 8 – 12 wt.%  $\text{H}_2$  and improved thermodynamic and kinetic properties such as  $2 \text{LiBH}_4 + \text{MgH}_2$  (Bösenberg et al., 2009; 2010; 2010b),  $2 \text{NaBH}_4 + \text{MgH}_2$  (Garroni et al., 2010; Pistidda et al., 2010; 2011; Pottmaier et al., 2011),  $\text{Ca}(\text{BH}_4)_2 + \text{MgH}_2$  (Barkhordarian et al., 2008),  $6 \text{LiBH}_4 + \text{CeH}_2$ ,  $6 \text{LiBH}_4 + \text{CaH}_2$  (Jin et al., 2008b),  $\text{LiBH}_4 / \text{Ca}(\text{BH}_4)_2$  (Lee et al., 2009). One of the most intensely studied systems hereof is the  $2 \text{LiBH}_4 + \text{MgH}_2$  system. The intended reaction pathway is:



However, several other reaction pathways are possible leading to products such as  $\text{LiB}_2$ , amorphous B,  $\text{Li}_2\text{B}_{12}\text{H}_{12}$  or  $\text{Li}_2\text{B}_{10}\text{H}_{10}$ . Bösenberg et al. (Bösenberg et al., 2010b) could show that due to a higher thermodynamic driving force for the favoured reaction the competing reactions can be suppressed by applying a hydrogen back pressure and limiting the dehydrogenation temperature. Nevertheless, since long-range diffusion of metal atoms containing species is required, see Fig. 13, in bulk ball-milled samples dehydrogenation so far occurs only at temperatures higher than 350 °C, hydrogenation at temperatures higher than 250 °C.

The dehydrogenation temperatures of this Reactive Hydride Composite, however, can be significantly reduced by using nanoconfined  $2 \text{LiBH}_4 + \text{MgH}_2$  stabilised in inert nanoporous aerogel scaffold materials whereby long-range phase separation is hindered and thus the diffusion path length reduced (Gosalawit-Utke, 2011).

#### 4. Conclusion

Metal hydrides offer a safe and compact alternative for hydrogen storage. The thermodynamic properties of them determine both their reaction heat as well as hydrogen

equilibrium pressure at given temperature and, therefore, are important parameters to be taken into account. Optimised system integration for a given application is not possible without selecting a hydride with suitable thermodynamic properties. To achieve highest possible energy efficiencies the heat of reaction and temperature of operation of the metal hydride should be adapted to the waste heat and temperature of operation of the fuel cell / fuel combustion system. It has been found that the thermodynamic properties of metal hydrides can be tailored in a wide range. Unfortunately, so far all the known conventional metal hydrides with more or less ideal reaction enthalpies and hydrogen equilibrium pressures above 5 bar at room temperature suffer from a rather limited reversible hydrogen storage capacity of less than 2.5 wt.%. With such a material it is not possible to realise a solid storage hydrogen tank with a total hydrogen storage density of more than 1.8 wt.% H<sub>2</sub>. Such tank systems still have advantages for the storage of small quantities of hydrogen for larger quantities, however, modern high pressure composite tank shells have a clear advantage in respect of gravimetric storage density. To realise a solid storage tank for hydrogen with a comparable gravimetric storage density it is required that novel hydrogen storage materials based on light weight elements are developed. There are several promising systems with high gravimetric storage densities in the range of 8 - 12 wt.% H<sub>2</sub>. For the applications of these novel material systems it is important to further adapt thermodynamic properties as well as the temperatures of operation towards the practical requirements of the system.

The discovery of the approach of combining different hydrides which react with each other during hydrogen release by forming a stable compound, the so-called Reactive Hydride Composites, show a great promise for the development of novel suitable hydrogen storage material systems with elevated gravimetric storage densities. However, so far, the ideal storage material with low reaction temperatures, a reaction heat in the range of  $|\Delta H| = 20\text{--}30$  kJ/(mol H<sub>2</sub>) and a on-board reversible hydrogen storage density of more than 6 wt.% H<sub>2</sub> has not been found.

## 5. References

- Ares Fernandez, J.R.; Aguey-Zinsou, F.; Elsaesser, M.; Ma, X.Z.; Dornheim, M.; Klassen, T.; Bormann, R. (2007). Mechanical and thermal decomposition of LiAlH<sub>4</sub> with metal halides. *International Journal of Hydrogen Energy*, Vol. 32, No. 8, pp. (1033-1040), ISSN: 0360-3199
- Barkhordarian, G.; Klassen, T.; Bormann, R. (2006). Kinetic investigation of the effect of milling time on the hydrogen sorption reaction of magnesium catalyzed with different Nb<sub>2</sub>O<sub>5</sub> contents. *Journal of Alloys and Compounds*, Vol. 407, No. 1-2, pp. (249-255), ISSN: 0925-8388
- Barkhordarian, G.; Klassen, T.; Dornheim, M.; Bormann, R. (2007). Unexpected kinetic effect of MgB<sub>2</sub> in reactive hydride composites containing complex borohydrides. *Journal of Alloys and Compounds*, Vol. 440, No. 1-2, pp. (L18-L21), ISSN: 0925-8388
- Barkhordarian, G.; Jensen, T.R.; Doppiu, S.; Bösenberg, U.; Borgschulte, A.; Gremaud, R.; Cerenius, Y.; Dornheim, M.; Klassen, T.; Bormann, R. (2008). Formation of Ca(BH<sub>4</sub>)<sub>2</sub> from Hydrogenation of CaH<sub>2</sub>+MgB<sub>2</sub> Composite. *Journal of Physical Chemistry C*, Vol. 112, No. 7, pp. (2743-2749), ISSN: 1932-7447
- Bösenberg, U.; Vainio, U.; Pranzas, P.K.; Bellosta von Colbe, J.M.; Goerigk, G.; Welter, E.; Dornheim, M.; Schreyer, A.; Bormann, R. (2009). On the chemical state and

- distribution of Zr- and V-based additives in Reactive Hydride Composites. *Nanotechnology*, Vol. 20; No. 20, pp. (204003/1-204003/9) ISSN: 1361-6528
- Bösenberg, U.; Kim, J.W.; Gossler, D.; Eigen, N.; Jensen, T.R.; Bellosta von Colbe, J.M.; Zhou, Y.; Dahms, M.; Kim, D.H.; Guenther, R.; cho, Y.W.; Oh, K.H.; Klassen, T.; Bormann, R.; Dornheim, M. (2010). Role of additives in LiBH<sub>4</sub>-MgH<sub>2</sub> Reactive Hydride Composites for sorption kinetics. *Acta Materialia*, Vol. 58, No. 9; pp. (3381-3389), ISSN: 1359-6454
- Bösenberg, U.; Ravnsbaek, D. B.; Hagemann, H.; D'Anna, V.; Bonatto Minella, C.; Pistidda, C.; van Beek, W.; Jensen, T.R.; Bormann, R.; Dornheim, M. (2010b). Pressure and Temperature Influence on the Desorption Pathway of the LiBH<sub>4</sub>-MgH<sub>2</sub> Composite System. *Journal of Physical Chemistry C*, Vol. 114, No. 35, pp. (15212-15217)
- Bogdanovic, B.; Schwickardi, M. (1997). Ti-doped alkali metal aluminum hydrides as potential novel reversible hydrogen storage materials. *Journal of Alloys and Compounds*, Vol. 253-254, pp. (1-9), ISSN: 0925-8388
- Bogdanovic, B.; Reiser, A.; Schlichte, K.; Spliethoff, B.; Tesche, B. (2002). Thermodynamics and dynamics of the Mg-Fe-H system and its potential for thermochemical thermal energy storage. *Journal of Alloys and Compounds*, Vol. 345, No. 1-2, pp. (77-89), ISSN: 0925-8388
- Bogdanovic, B.; Felderhoff, M.; Streukens, G. (2009). Hydrogen storage in complex metal hydrides. *Journal of the Serbian Chemical Society*, Vol. 74, No. 2, pp. (183-196), ISSN: 0352-5139
- Bonatto Minella, Christian; Garroni, Sbastiano; Pistidda, Claudio; Goslawit-Utke, R.; Barkhordarian, G.; Rongeat, C.; Lindeman, I.; Gutfleisch, O.; Jensen, T.R.; Cerenius, Y.; Christnsen, J.; Baro, M.D.; Bormann, R.; Klassen, T.; Dornheim, M. (2011). Effect of Transition Metal Fluorides on the Sorption Properties and Reversible Formation of Ca(BH<sub>4</sub>)<sub>2</sub>. *Journal of Physical Chemistry C*, Vol. 115, No. 5, pp (2497-2504), ISSN: 1932-7447
- Bouaricha, S.; Dodelet, J.P.; Guay, D.; Huot, J. Boily, S.; Schulz, R. (2000). Hydriding behaviour of Mg-Al and leached Mg-Al compounds prepared by high-energy ball-milling. *Journal of Alloys and Compounds*, Vol. 297, pp. (282-293)
- Brinks, H.; Fossdal, A.; Hauback, B. (2008). Adjustment of the stability of complex hydrides by anion substitution. *Journal of Physical Chemistry C*, Vol. 112, No. 14; pp. (5658-5661), ISSN: 1932-7447
- Buchner, H. (1982). *Energiespeicherung in Metallhydriden*, Springer-Verlag, 3-211-81703-4, Wien
- Buchner, H.; Povel, R. (1982). The Daimler-Benz Hydride Vehicle Project. *International Journal of Hydrogen Energy*, Vol. 7, No. 3, pp. (259-266), ISSN: 0360-3199/82/030259-08
- Chen, P.; Xiong, Z.T.; Luo, J.Z.; Lin, J; Tan, K.L. (2002). Interaction of hydrogen with metal nitrides and imides. *Nature*, Vol. 420, pp. (302-304)
- Cho, Y.W.; Shim, J.-H.; Lee, B.-J. (2006). Thermal destabilization of binary and complex metal hydrides by chemical reaction: A thermodynamic analysis. *CALPHAD: Computer Coupling of Phase Diagrams and Thermochemistry*, Vol. 30, No. 1, pp. (65-69), ISSN: 0364-5916

- Darnaudry, J.P.; Darriet, B.; Pezat, M. (1983). The  $Mg_2Ni_{0.75}M_{0.25}$  alloys (M = 3d element): their application to hydrogen storage. *International Journal of Hydrogen Energy*, Vol. 8, pp. (705-708)
- Deprez, E.; Justo, A.; Rojas, T.C.; Lopez, Cartes, C.; Bonatto Minella, C.; Bösenberg, U.; Dornheim, M.; Bormann, R.; Fernandez, A. (2010) Microstructural study of the LiBH<sub>4</sub>-MgH<sub>2</sub> Reactive Hydride Composite with and without Ti isopropoxide additive. *Acta Materialia*, Vol. 58, No. 17, pp. (5683-5694), ISSN: 1359-6454
- Deprez, E.; Munoz-Marquez, M.A.; Jimenez de Haro, M.C.; Palomares, F.J.; Foria, F.; Dornheim, M.; Bormann, R.; Fernandez, A. (2011). Combined x-ray photoelectron spectroscopy and scanning electron microscopy studies of the LiBH<sub>4</sub>-MgH<sub>2</sub> Reactive Hydride Composite with and without a Ti-based additive. *Journal of Applied Physics*, Vol. 109, No. 1, pp. (014913/1-014913/10), ISSN: 0021-8979
- Didisheim, J.-J.; Zolliker, P.; Yvon, K.; Fischer, P.; Schefer, J.; Gubelmann, M.; Williams, A.F. (1984). Dimagnesium iron(II) hydride; Mg<sub>2</sub>FeH<sub>6</sub>, containing octahedral FeH<sub>6</sub><sup>4-</sup> anions. *Inorganic Chemistry*, Vol. 23, No. 13, pp. (1953-1957), ISSN: 0020-1669
- Dornheim, M.; Eigen, N.; Barkhordarian, G.; Klassen, T.; Bormann, R. (2006). Tailoring Hydrogen Storage Materials Towards Application. *Advanced Engineering Materials*, Vol. 8, No. 5, pp. (377-385), ISSN: 1438-1656
- Dornheim, M.; Doppiu, S.; Barkhordarian, G.; Boesenberg, U.; Klassen, T.; Gutfleisch, O.; Bormann, R. (2007) Hydrogen storage in magnesium-based hydrides and hydride composites. Viewpoint paper in: *Scripta Materialia*, Vol. 56, pp. (841-846), ISSN: 1359-6462
- Dornheim, M.; Klassen, T. (2009). High Temperature Hydrides, In: *Encyclopedia of Electrochemical Power Sources*, Vol. 3, J. Garche, C. Dyer, P. Moseley, Z. Ogumi, D. Rand, B. Scrosati, pp. (459-472), Elsevier, ISBN 10: 0-444-52093-7, Amsterdam
- Dornheim, M. (2010). Tailoring Reaction Enthalpies of Hydrides, In: *Handbook of Hydrogen Storage*, Michael Hirscher, pp. (187-214), Wiley-VCH Verlag GmbH & Co, ISBN: 978-3-527-32273-2, Weinheim
- Eigen, N.; Keller, C.; Dornheim, M.; Klassen, T.; Bormann, R. (2007). Industrial production of light metal hydrides for hydrogen storage. Viewpoint Set in *Scripta Materialia*, Vol. 56, No. 10, pp. (847-851), ISSN: 1359-6462
- Eigen, N.; Gosch, F.; Dornheim, M.; Klassen, T.; Bormann, R. (2008). Improved hydrogen sorption of sodium alanate by optimized processing. *Journal of Alloys and Compounds*, Vol. 465, No. 1-2, pp. (310-316), ISSN: 0925-8388
- Eigen, N.; Bösenberg, U.; Bellosta von Colbe, J.; Jensen, T.R.; Cerenius, Y.; Dornheim, M.; Klassen, T.; Bormann, R. (2009). Reversible hydrogen storage in NaF-Al composites. *Journal of Alloys and Compounds*. Vol. 477, No. 1-2, pp. (76-80), ISSN: 0925-8388
- Fossdal, A.; Brinks, H.W.; Fonneloep, J.E.; Hauback, B.C. (2005). Pressure-composition isotherms and thermodynamic properties of TiF<sub>3</sub>-enhanced Na<sub>2</sub>LiAlH<sub>6</sub>. *Journal of Alloys and Compounds*, Vol. 397, No. 1-2, pp. (135-139), ISSN:0925-8388
- Fujitani, S.; Yonezu, I.; Saito, T.; Furukawa, N.; Akiba, E.; Hayakawa, H.; Ono, S. (1991). Relation between equilibrium hydrogen pressure and lattice parameters in pseudobinary Zr–Mn alloy systems. *Journal of the Less Common Metals*, vol. 172-174, No. 1, pp. (220-230)
- Fukai, Y. (1993). The Metal-Hydrogen System. *Springer Series in Materials Science*, Vol. 21, Springer, Berlin

- Garroni, S.; Milanese, C.; Girella, A.; Marini, A.; Mulas, G.; Menendez, E.; Pistidda, C.; Dornheim, M.; Surinach, S.; Baro, M. D. (2010). Sorption properties of NaBH<sub>4</sub>/MH<sub>2</sub> (M = Mg, Ti) powder systems. *International Journal of Hydrogen Energy*, Vol. 35, No. 11, pp. (5434-5441)
- Goerrig, D. (1960). Borohydrides of alkali and alkaline earth metals. German Patent 1077644, F27373 IVa/12i, Application No. DE 1958-F27373
- Gosalawit-Utke, R.; Nielsen, T.K.; Saldan, I.; Laipple, D.; Cerenius, Y.; Jensen, T.R.; Klassen, T.; Dornheim, M. (2011). Nanoconfined 2LiBH<sub>4</sub>-MgH<sub>2</sub> prepared by direct melt infiltration into nanoporous materials. *Journal of Physical Chemistry C*, Vol. 115, No. 21, pp. (10903-10910)
- Griessen, R.; Driessen, A. (1984). Heat of formation and band structure of binary and ternary metal hydrides. *Phys. Rev. B*, Vol. 30, No. 8, pp. (4372-81), ISSN: 0163-1829
- Hägg, G. (1931). Röntgen investigations on the hydrides of titanium zirconium, vanadium and tantalum. *Zeitschrift für physikalische Chemie B*, Vol. 11, pp. (433-445)
- Hanada, N.; Ichikawa, T.; Fujii, H. (2007). Hydrogen absorption kinetics of the catalyzed MgH<sub>2</sub> by niobium oxide. *Journal of Alloys and Compounds*, Vol. 446-447, pp. (67-71), ISSN: 0925-8388
- Huot, J.; Boily, S.; Güther, V.; Schulz, R. (1999). Synthesis of Na<sub>3</sub>AlH<sub>6</sub> and Na<sub>2</sub>LiAlH<sub>6</sub> by mechanical alloying. *Journal of Alloys and Compounds*, Vol. 283, No. 1-2, pp. (304-306), ISSN: 0925-8388
- Jin, S.-A.; Shim, J.-H.; Cho, Y.W.; Yi, K.-W.; Zabara, O.; Fichtner, M. (2008). Reversible hydrogen storage in LiBH<sub>4</sub>-Al-LiH composite powder. *Scripta Materialia*, Vol. 58, No. 11, pp. (963-965), ISSN: 1359-6462
- Jin, S.-A.; Lee, Y.-S.; Shim, J.-H.; Cho, Y.W. (2008b). Reversible Hydrogen Storage in LiBH<sub>4</sub>-MH<sub>2</sub> (M = Ce, Ca) Composites. *Journal of Physical Chemistry C*, Vol. 112, No. 25, pp. (9520-9524), ISSN: 1932-7447
- Kang, X.-D.; Wang, P.; Ma, L.-P.; Cheng, H.-M. (2007). Reversible hydrogen storage in LiBH<sub>4</sub> destabilized by milling with Al. *Applied Physics A: Materials Science & Processing*, Vol. 89, No. 4, pp. (963-966)
- Klassen, T.; Oelerich, W.; Zeng, K.; Bormann, R.; Huot, J. (1998). Nanocrystalline Mg-based alloys for hydrogen storage, In: *Magnesium Alloys and their Applications*, B.L. Mordike and K. U. Kainer, pp. (308-311), Werkstoff-Informationsgesellschaft mbH Frankfurt, CODEN: 68TSA9, Wolfsburg, Germany
- Klell, M. (2010). Storage of Hydrogen in the Pure Form, In: *Handbook of Hydrogen Storage*, Michael Hirscher, pp. (187-214), Wiley-VCH Verlag GmbH & Co, ISBN: 978-3-527-32273-2, Weinheim
- Konstanchuk, I.G.; Ivanov, E.Y.; Pezat, M.; Darriet, B.; Bodyrev, V.V.; Hagenmüller, P. (1987). The hydriding properties of a mechanical alloy with composition Mg-25% Fe. *J. Less-Common Met.*, Vol. 131, pp. (181-189)
- Lee, J.Y.; Ravnsbaek, D.; Lee, Y.-S.; Kim, Y.; Cerenius, Y.; Shim, J.-H.; Jensen, T.R.; Hur, N.H.; Cho, Y.W. (2009). Decomposition Reactions and Reversibility of the LiBH<sub>4</sub>-Ca(BH<sub>4</sub>)<sub>2</sub> Composite. *Journal of Physical Chemistry C*, Vol. 113, No. 33, pp. (15080-15086), ISSN: 1932-7447
- Libowitz, G.G.; Hayes, H.F.; Gibb, T.R.G.Jr. (1958). The system zirconium-nickel and hydrogen. *Journal of Physical Chemistry*, Vol. 62 pp. (76-79), ISSN: 0022-3654

- Li, H.-W.; Orimo, S.; Nakamori, Y.; Miwa, K.; Ohba, N.; Towata, S.; Züttel, A. (2007). Materials designing of metal borohydrides: Viewpoints from thermodynamical stabilities. *Journal of Alloys and Compounds*, Vol. 446-447, pp. (315-318)
- Li, H.-W.; Kikuchi, K.; Nakamori, Y.; Ohba, N.; Miwa, K.; Towata, S.; Orimo, S. (2008) Dehydrogenating and Rehydrogenating Processes of Well-Crystallized  $\text{Mg}(\text{BH}_4)_2$  Accompanying with Formation of Intermediate Compounds. *Acta Materialia*, Vol. 56, pp. (1342-1347)
- Ma, X.Z.; Martinez-Franco, E.; Dornheim, M.; Klassen, T.; Bormann, R. (2005). Catalyzed  $\text{Na}_2\text{LiAlH}_6$  for hydrogen storage. *Journal of Alloys and Compounds*, Vol. 404-406, pp. (771-774), ISSN: 0925-8388
- Martinez-Franco, E.; Klassen, T.; Dornheim, M.; Bormann, R.; Jaramillo-Vigueras, D. (2010). Hydrogen sorption properties of Ti-oxide/chloride catalyzed  $\text{Na}_2\text{LiAlH}_6$ . *Ceramic Transactions*, Vol. 209, No. 13-20, pp. (13-20), ISSN: 1042-1122
- Mauron, P., Buchter, F., Friedrichs, O., Remhof, A., Biemann, M., Zwicky, C.N., Züttel, A. (2008) Stability and Reversibility of  $\text{LiBH}_4$ . *J. Phys. Chem. B*, Vol. 112, pp. (906-910)
- Nakamori, Y.; Miwa, K.; Ninomiya, A.; Li, H.; Ohba, N.; Towata, S.; Züttel, A.; Orimo, S. (2006). Correlation between thermodynamical stabilities of metal borohydrides and cation electronegativities: First-principles calculations and experiments. *Physical Review B*, Vol. 74, pp. (045126 - 1-9)
- Nakamori, Y.; Miwa, K.; Li, H.; Ohba, N.; Towata, S.; Orimo, S. (2007). Tailoring of Metal Borohydrides for Hydrogen Storage Applications. *Mater. Res. Soc. Symp. Proc.* Vol. 971, 0971-Z02-01
- Oelerich, W. (2000). *Sorptionseigenschaften von nanokristallinen Metallhydriden für die Wasserstoffspeicherung*, Helmholtz-Zentrum Geesthacht Zentrum für Material- und Küstenforschung GmbH (formerly GKSS-Forschungszentrum Geesthacht GmbH), ISSN 0344-9626, Geesthacht, Germany
- Oelerich, W. (2001). Metal oxides as catalysts for improved hydrogen sorption in nanocrystalline Mg-based materials. *Journal of Alloys and Compounds*, Vol.315, No. 1-2, pp. (237-242), ISSN: 0925-8388
- Orimo, S.; Ikeda, K.; Fujii, H.; Saruki, S.; Fukunaga, T.; Züttel, A.; Schlapbach, L. (1998). Structural and hydriding properties of  $(\text{Mg}_{1-x}\text{Al}_x)\text{Ni-H}(\text{D})$  with amorphous or CsCl-type cubic structure ( $x=0-0.5$ ), *Acta Materialia*, Vol. 46, No. 13, pp. (4519-4525), ISSN: 1359-6454
- Orimo, S.; Nakamori, Y.; Kitahara, G.; Miwa, K.; Ohba, N.; Towata, S.; Züttel, A. (2005). Dehydrogenating and rehydrogenating reactions of  $\text{LiBH}_4$ . *Journal of Alloys and Compounds*, Vol. 404-430, pp. (0925-8388), ISSN: 0925-8388
- Orimo, S.; Nakamori, Y.; Eliseo, J.R.; Züttel, A.; Jensen, C.M. (2007). Complex Hydrides for Hydrogen Storage. *Chemical Reviews*, Vol. 107, No. 10, pp. (4111-4132). ISSN:0009-2665
- Panella, B.; Hirscher, M.; Roth, S. (2005). Hydrogen adsorption in different carbon nanostructures. *Carbon*, Vol. 43, pp. (2209-2214)
- Pistidda, C.; Garroni, S.; Bonatto Minella, C.; Coci, F.; Jensen, T.R.; Nolis, P.; Bösenberg, U.; Cerenius, Y.; Lohstroh, W.; Fichtner, M.; Baro, M.D.; Bormann, R.; Dornheim, M.

- (2010). Pressure Effect on the  $2\text{NaH} + \text{MgB}_2$  Hydrogen Absorption Reaction. *Journal of Physical Chemistry C*, Vol. 114, No. 49, pp. (21816-21823)
- Pistidda, C. Barkhordarian, G.; Rzeszutek, A.; Garroni, S.; Minella, C. Bonatto; Baro, M. D.; Nolis, P.; Bormann, Ruediger; Klassen, T.; Dornheim, M. (2011). Activation of the reactive hydride composite  $2\text{NaBH}_4 + \text{MgH}_2$ . *Scripta Materialia*, Vol. 64, No. 11, pp. (1035-1038)
- Pottmaier, D.; Pistidda, C.; Groppo, E.; Bordiga, S.; Spoto, G.; Dornheim, M.; Baricco, M. (2011). Dehydrogenation reactions of  $2\text{NaBH}_4 + \text{MgH}_2$  system. *International Journal of Hydrogen Energy*. Vol. 36, No. 13, pp. (7891-7896)
- Pourarian, F.; Shinha, V.K.; Wallace, W.E; Smith, H.K. (1982) Kinetics and thermodynamics of  $\text{ZrMn}_2$ -based hydrides. *Journal of the Less Common Metals*, Vol. 88, No. 2, pp. (451-458)
- Pranzas, P.K.; Dornheim, M.; Bellmann, D.; Aguey-Zinsou, K.-F., Klassen, T.; Schreyer, A. (2006). SANS/USANS investigations of nanocrystalline  $\text{MgH}_2$  for reversible storage of hydrogen. *Physica B: Condensed Matter*, Vol. 385-386, No. 1, pp. (630-632), ISSN: 0921-4526
- Pranzas, P.K.; Dornheim, M.; Boesenberg, U.; Ares Fernandez, J.R.; Goerigk, G.; Roth, S.V.; Gehrke, R.; Schreyer, A. (2007). Small-angle scattering investigations of magnesium hydride used as a hydrogen storage material. *Journal of Applied Crystallography*, Vol 40, No. S1, pp. (383-387), ISSN: 0021-8898
- Puszkiel, J.A.; Larochette, P.A.; Gennari, F.C. (2008). Thermodynamic-kinetic characterization of the synthesized  $\text{Mg}_2\text{FeH}_6$ - $\text{MgH}_2$  hydrides mixture. *International Journal of Hydrogen Energy*, Vol. 33, No. 13, pp. (3555-3560), ISSN: 0360-3199
- Reilly, J.J.; Wiswall, R. H. (1967). Reaction of hydrogen with alloys of magnesium and copper. *Inorganic Chemistry*, Vol. 6, pp. (2220-2223), ISSN: 0020-1669
- Reilly, J.J.; Wiswall, R.H. (1968). Reaction of hydrogen with alloys of magnesium and nickel and the formation of  $\text{Mg}_2\text{NiH}_4$ . *Inorganic Chemistry*, Vol. 7, pp. (2254-2256), ISSN: 0020-1669
- Reilly, J.J., Sandrock, G.D. (1980). Hydrogen storage in metal hydrides. *Scientific American*, Vol. 242, No2, pp. ( 5118- 5129), ISSN: 00368733
- Rude, L.H.; Nielsen, T. K.; Ravnsbaek, D.B.; Bösenberg, U.; Ley, M.B.; Richter, B.; Arnbjerg, L.M.; Dornheim, M.; Filinchuk, Y.; Besenbacher, F.; Jensen, T.R. (2011). Tailoring properties of borohydrides for hydrogen storage: A review. *Phys. Status Solidi A*, DOI: 10.1002/pssa.201001214
- Sandrock, G. (1999). A panoramic overview of hydrogen storage alloys from a gas reaction point of view. *Journal of Alloys and Compounds*, Vol. 293-295, pp. (877-888)
- Sangster, J.J.; Pelteon, A. D. (2000). In: *Phase Diagrams of Binary Hydrogen Alloys*; Manchester, F.D. ; ASM International: Materials Park, OH, pp. 74
- Schlesinger, H.I.; Brown, H.C. (1940). Metallo borohydrides. III. Lithium borohydride. *Journal of the American Chemical Society*, Vol. 62, pp. (3429-3435), ISSN: 0002-7863
- Schmitz, B.; Mueller, U.; Trukhan, N.; Schubert, M.; Ferey, G.; Hirscher, M. (2008). Heat of adsorption in microporous high-surface-area materials. *Chem Phys Chem*, Vol. 9, pp. (2181-2184)



- Schober, T.; Wenzl, H. (1978). The systems niobium hydride (deuteride), tantalum hydride (deuteride), vanadium hydride (deuteride): structures, phase diagrams, morphologies, methods of preparation, In: *Topics in Applied Physics Vol. 29, Hydrogen in Metals, Vol. 2*, G. Alefeld and J. Völkl, pp. (11-71), Springer, ISSN: 0303-4216, Berlin
- Schüth, F.; Bogdanovic, B., Felderhoff, M. (2004). Light metal hydrides and complex hydrides for hydrogen storage. *Chemical Communications*, No. 20, pp. (2249-2258), ISSN: 1359-7345
- Seballos, L.; Zhang, J.Z.; Rönnebro, E.; Herberg, J.L.; Majzoub, E.H. (2009). Metastability and crystal structure of the bialkali complex metal borohydrides  $\text{NaK}(\text{BH}_4)_2$ . *Journal of Alloys and Compounds*, Vol. 476, pp. (446-450)
- Selvam, P.; Yvon, K. (1991). Synthesis of magnesium iron hydride ( $\text{Mg}_2\text{FeH}_6$ ), magnesium cobalt hydride ( $\text{Mg}_2\text{CoH}_5$ ) and magnesium nickel hydride ( $\text{Mg}_2\text{NiH}_4$ ) by high-pressure sintering of the elements, *International Journal of Hydrogen Energy*, Vol. 16, No. 9, pp. (615-617), ISSN: 0360-3199
- Terashita, N.; Kobayashi, K.; Sasai, T.; Akiba, E. (2001). Structural and hydriding properties of  $(\text{Mg}_{1-x}\text{Ca}_x)\text{Ni}_2$  Laves phase alloys. *Journal of Alloys and Compounds*, Vol. 327, No. 1-2, pp. (275-280), ISSN: 0925-8388
- Tsushio, Y.; Enoki, H.; Akiba, E. (1998). Hydrogenation properties of  $\text{MgNi}_{0.86}\text{M}_{10.03}$  ( $\text{M}=\text{Cr, Fe, Co, Mn}$ ) alloys. *Journal of Alloys and Compounds*, Vol. 281, pp. (301-305).
- Vajo, J.J.; Mertens, F.; Ahn, C.C.; Bowman, R.C.Jr.; Fultz, B. (2004). Altering Hydrogen Storage Properties by Hydride Destabilization through Alloy Formation:  $\text{LiH}$  and  $\text{MgH}_2$  Destabilized with Si. *Journal of Physical Chemistry B*, Vol. 108, No. 37, pp. (13977-13983), ISSN: 1520-6106
- Vajo, J.J.; Skeith, S.L.; Mertens, F. (2005). Reversible Storage of Hydrogen in Destabilized  $\text{LiBH}_4$ . *Journal of Physical Chemistry B*, Vol. 109, No. 9, pp. (3719-3722), ISSN: 1520-6106
- Van Mal, H.H.; Buschow, K.H.H.; Miedema, A.R. (1974). Hydrogen absorption in lanthanum-nickel ( $\text{LaNi}_5$ ) and related compounds. Experimental observations and their explanation. *Journal of the Less-Common Metals*, Vol. 35, No. 1, pp. (65-76), ISSN: 0022-5088
- Wiswall, R. (1978). Hydrogen Storage in Metals. In: *Topics of Appl. Phys., Vol. 29 - Hydrogen in Metals II*, G. Alefeld and J. Völkl, pp. (201-242), Springer-Verlag, ISSN: 0303-4216, Berlin, Heidelberg, New York
- Xiong, Z.T.; Hu, J.J.; Wu, G.T.; Chen, P.; Luo, W.; Gross, K.; Wang, J. (2005). Thermodynamic and kinetic investigations of the hydrogen storage in the Li-Mg-N-H system. *Journal of Alloys and Compounds*, Vol. 398, No. 1-2, pp. (235-239), ISSN: 0925-8388
- Yin, L.-C.; Wang, P.; Kang, X.-D.; Sun, C.-H.; Cheng, H.-M. (2007). Functional anion concept: Effect of fluorine anion on hydrogen storage of sodium alanate. *Physical Chemistry Chemical Physics*, Vol. 9, No. 12, pp. (1499-1502), ISSN: 1463-9076

- Yin, L.; Wang, P.; Fang, Z.; Cheng, H. (2008). Thermodynamically tuning LiBH<sub>4</sub> by fluorine anion doping for hydrogen storage: A density functional study. *Chemical Physics Letters*, Vol. 450, No. 4-6, pp. (318-321), ISSN: 0009-2614
- Yoshida, M.; Akiba, E. (1995). Hydrogen absorbing-desorbing properties and crystal structure of the Zr-Ti-Ni-Mn-V AB<sub>2</sub> Laves phase alloys. *Journal of Alloys and Compds.*, Vol. 224, pp. (121-126)
- Züttel, A. (2003). Materials for Hydrogen Storage. *Materials Today*, Vol. 6, pp. (24-33), ISSN : 1369 7021