IntechOpen

Some Applications of Quantum Mechanics

Edited by Mohammad Reza Pahlavani

SOME APPLICATIONS OF QUANTUM MECHANICS

Edited by **Mohammad Reza Pahlavani**

Some Applications of Quantum Mechanics
http://dx.doi.org/10.5772/2540
Edited by Mohammad Reza Pahlavani

Contributors

Nelson Flores-Gallegos, Abdennaceur Karoui, Faouzia Sahtout Karoui, Mikhail Krivoruchenko, Bjorn Jensen, Toru Matsui, Hideaki Miyachi, Yasuteru Shigeta, Kimihiko Hirao, Juan Manuel López, Rodolfo Esquivel, Norton G De Almeida, Miloš Vaclav Lokajíček, Kenji Mishima, Mohammad Reza Pahlavani, Seyed Mohammad Motevalli, Masaru Tateno, Jioung Kang, Omar Morandi, Peng Tao, Joseph Larkin, Bernard Brooks, Shigeaki Ono

Notice

First published in Croatia, 2012 by INTECH d.o.o.
eBook (PDF) Published by IN TECH d.o.o.
Place and year of publication of eBook (PDF): Rijeka, 2019.
IntechOpen is the global imprint of IN TECH d.o.o.
Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Some Applications of Quantum Mechanics
Edited by Mohammad Reza Pahlavani

p. cm.

ISBN 978-953-51-0059-1

eBook (PDF) ISBN 978-953-51-4953-8

Meet the editor

Dr Mohammad Reza Pahlavani, was born in 1958 at north-east of Iran. He has been obtained his Bs.c from Mashhad University (Mashhad-Iran), Ms.c from Tehran University (Tehran-Iran) and Ph.D from Indian Institute of technology Bombay (Mumbaei-India) in nuclear physics at 2001. He was working as professor of physics in department of physics Mazandaran University since, and currently occupied as head of nuclear physics department. During last ten years, he published more than fifty papers in international journals, mostly in PRC, MPLB, IJMPE and presented more than fifty papers in national and international conferences. He has guided more than 20 Ms.c student, 5 Ph.D student and taught several courses of Bs.c, Ms.c and Ph.D students of nuclear physics, mostly quantum mechanics, nuclear physics and classical mechanics for under graduate and graduate students.

Contents

Preface XI

Chapter 1 **Quantum Phase-Space Transport and Applications to the Solid State Physics 1**
Omar Morandi

Chapter 2 **Reaction Path Optimization and Sampling Methods and Their Applications for Rare Events 27**
Peng Tao, Joseph D. Larkin and Bernard R. Brooks

Chapter 3 **Semiclassical Methods of Deformation Quantisation in Transport Theory 67**
M. I. Krivoruchenko

Chapter 4 **Synergy Between First-Principles Computation and Experiment in Study of Earth Science 91**
Shigeaki Ono

Chapter 5 **Quantum Mechanical Three-Body Systems and Its Application in Muon Catalyzed Fusion 109**
S. M. Motevalli and M. R. Pahlavani

Chapter 6 **Application of Quantum Mechanics for Computing the Vibrational Spectra of Nitrogen Complexes in Silicon Nanomaterials 131**
Faouzia Sahtout Karoui and Abdennaceur Karoui

Chapter 7 **Metal-Assisted Proton Transfer in Guanine-Cytosine Pair: An Approach from Quantum Chemistry 167**
Toru Matsui, Hideaki Miyachi, Yasuteru Shigeta and Kimihiko Hirao

Chapter 8 **Quantum Mechanics on Surfaces 189**
Bjørn Jensen

Chapter 9 **Quantum Statistics and Coherent Access Hypothesis 215**
Norton G. de Almeida

Chapter 10 **Flows of Information and Informational Trajectories in Chemical Processes 233**
Nelson Flores-Gallegos and Carmen Salazar-Hernández

Chapter 11 **Quantum Mechanics Design of Two Photon Processes Based Solar Cells 257**
Abdennaceur Karoui and Ara Kechiantz

Chapter 12 **Quantum Information-Theoretical Analyses of Systems and Processes of Chemical and Nanotechnological Interest 297**
Rodolfo O. Esquivel, Edmundo M. Carrera, Cristina Iuga, Moyocoyani Molina-Espíritu, Juan Carlos Angulo, Jesús S. Dehesa, Sheila López-Rosa, Juan Antolín and Catalina Soriano-Correa

Chapter 13 **Quantum Computing and Optimal Control Theory 335**
Kenji Mishima

Chapter 14 **Recent Applications of Hybrid *Ab Initio* Quantum Mechanics – Molecular Mechanics Simulations to Biological Macromolecules 359**
Jiyoung Kang and Masaru Tateno

Chapter 15 **Battle of the Sexes: A Quantum Games Theory Approach 385**
Juan Manuel López R.

Chapter 16 **Einstein-Bohr Controversy After 75 Years, Its Actual Solution and Consequences 409**
Miloš V. Lokajiček

Preface

The volume Some Applications of Quantum Mechanics is intended to serve as a reference for Graduate level students as well as researchers from all fields of science. Quantum mechanics has been extremely successful in explaining microscopic phenomena in all branches of physics. Quantum mechanics is used on a daily basis by thousands of physicists, chemists and engineers. There were two revolutions in the way we viewed the physical world in the twentieth century: relativity and quantum mechanics. In quantum mechanics, the revolution was both profound, requiring a dramatic revision in the structure of the laws of mechanics that govern the behavior of all particles, be they electrons or photons, and determining the stability of matter itself, shaping the interactions of particles on the atomic, nuclear, and particle physics level, and leading to macroscopic quantum effects ranging from lasers and superconductivity to neutron stars and radiation from the black holes. We have always had a great deal of difficulty understanding the worldview that quantum mechanics represents. Quantum mechanics is often thought of as being the physics of the very small, as seen through its successes in describing the structure and properties of atoms and molecules (the chemical properties of matter), the structure of atomic nuclei and the properties of elementary particles. But this is true only insofar as the fact that peculiarly quantum elects are most readily observed at the atomic level. Beyond that, quantum mechanics is needed to explain radioactivity, how semiconducting devices (the backbone of modern high technology) work, and the origin of superconductivity, what makes a laser function. Although this book does not cover all areas of application of quantum mechanics, it is nevertheless a valuable effort by an international group of invited authors. I believed that it is necessary to publish at least one volume for each type of the enormous applications of quantum mechanics. This book is contains sixteen chapters and its brief outline is as follows:

Chapters one to five provide some methods to solve the Schrodinger equation in different areas of science. Chapter six describes the application of quantum mechanics in three-body systems, which are mostly used in fusion phenomena as an attractive part of nuclear physics. Applications of quantum mechanics in solid-state physics and nanotechnology are described well in chapter seven. Chapter eight covers the applications of quantum mechanics in biotechnology, for analyzing Ciplatin bounds in DNA. A study of a different surface in non-relativistic and relativistic reference frame using quantum mechanics is presented in chapter nine. Quantum Hall effect,

superconductivity and related subjects using fractional statistic in quantum mechanics are covered in chapter ten. Chemical processes and quantum chemistry are discussed in chapter eleven. The application of quantum mechanics in photo electronic properties of semiconductors to study the effect of two-photon absorption in solar cells is discussed in chapter twelve. Chapter thirteen is related to quantum mechanical study of multi electronic systems and their relation to information theory and thermodynamical properties of Microsystems. Quantum computing and quantum information science are presented as a fresh and attractive research area of applied science in chapter fourteen. Chapter fifteen describes the hybrid ab initio quantum mechanics applied to investigate the molecular structure of biological macromolecules. The final chapter, chapter sixteen, deals with the application of game theory to predict the battle of sex using matrix representation of quantum mechanics, accompanied with related statistics.

This collection is written by an international group of invited scientists and researchers and I gratefully acknowledge their collaboration in this project. I would like to thank Ms. Maja Bozicevic for her valuable assistance in different stages of the project, and the InTech publishing team for creating this opportunity for scientists and researchers to communicate and publish this book.

Mohammad Reza Pahlavani
Head of Nuclear Physics Department,
Mazandaran University, Mazandaran, Babolsar,
Iran

1

Quantum Phase-Space Transport and Applications to the Solid State Physics

Omar Morandi
Institute of Theoretical and Computational Physics, Graz University of Technology
Austria

1. Introduction

Quantum modeling is becoming a crucial aspect in nanoelectronics research in perspective of analog and digital applications. Devices like resonant tunneling diodes or graphene sheets are examples of solid state structures that are receiving great importance in the modern nanotechnology for high-speed and miniaturized systems. Differing from the usual transport where the electronic current flows in a single band, the remarkable feature of this new solid state structures is the possibility to achieve a sharp coupling among states belonging to different bands. Under some conditions, a non negligible contribution to the particle transport induced by interband tunneling can be observed and, consequently, the single band transport or the classical phase-space description of the charge motion based on the Boltzmann equation are no longer accurate. Different approaches have been proposed for the full quantum description of the electron transport with the inclusion of the interband processes. Among them, the phase-space formulation of quantum mechanics offers a framework in which the quantum phenomena can be described with a classical language and the question of the quantum-classical correspondence can be directly investigated. In particular, the visual representation of the quantum mechanical motion by quantum-corrected phase-plane trajectories is a valuable instrument for the investigation of the particle-particle quantum coherence. However, due to the non-commutativity of quantum mechanical operators, there is no unique way to describe a quantum system by a phase-space distribution function. Among all the possible definitions of quantum phase-space distribution functions, the Wigner function, the Glauber-Sudarshan P and Q functions, the Kirkwood and the Husimi distribution have attained a considerable interest (Lee, 1995). The Glauber-Sudarshan distribution function has turned out to be particularly useful in quantum optics and in the field of solid state physics and the Wigner formalism represents a natural choice for including quantum corrections in the classical phase-space motion (see, for example (Jüngel, 2009)).
This Chapter is intended to present different approaches for modeling the quantum transport in nano-structures based on the Wigner, or more generally, on the quantum phase-space formalism. Our discussion will be focused on the application of the Weyl quantization procedure to various problems. In particular, we show the existence of a quite general multiband formalism and we discuss its application to some relevant cases. In accordance with the Schrödinger representation, where a physical system can be characterized by a set of projectors, we extend the original Wigner approach by considering a wider class of representations. The applications of this formalism span among different subjects: the

multi-band transport and its applications to nano-devices, quasi classical approximations of the motion and the characterization of a system in terms of Berry phases or, more generally, the representation of a quantum system by means of a Riemann manifold with a suitable connection. We discuss some results obtained in this contexts by presenting the major lines of the derivation of the models and their applications. Particular emphasis is devoted to present the methods used for the approximation of the solution. The latter is a particularly important aspect of the theory, but often underestimated: the description of a system in the quantum phase-space usually involves a very complex mathematical formulation and the solution of the equation of motion is only available by numerical approximations. Furthermore, the approximation of the quantum phase-space solution in some cases is not merely a technical trick to depict the solution, but could reveal itself to be a valuable basis for a further methodological investigation of the properties of a system. In the multiband case, some asymptotic procedures devised for the approximation of the quantum Wigner solution have shown a very attractive connection with the Dyson theory of the particle interaction, which allows us to describe the interband quantum transition by means of an effective scattering process (Morandi & Demeio, 2008). Furthermore, the formal connection between the Wigner formalism and the classical Boltzmann approach suggests some direct and general approximations where scattering and relaxation mechanisms can be included in the quantum mechanical framework.
The chapter is organized as follows. In sec. 2 an elementary derivation of the Wigner formalism is introduced. The Wigner function is the basis element of a more general theory denoted by Wigner-Weyl quantization procedure. This is explained in section 3.4 and in sections 3.1. The sections 3.2 and 3.4 are devoted to the application of the Wigner-Weyl formalism to the particle transport in semiconductor structures and in graphene. In section 4 an interesting connection between the diagonalization procedure exposed in section 3.1 and the Berry phase theory is presented. In section 5 a general approximation procedure of the pseudo-differential force operator is proposed. This leads to the definition of an effective force field. Its application in some quantum corrected transport model is discussed. Finally, in section 6, the inclusion of phonon collisions in a quantum corrected kinetic model is addressed and the current evolution in graphene is numerically investigated.

2. Definition of the Wigner function

The quantum mechanical motion of a statistical ensemble of electrons is usually characterized by a trace class function denoted as density matrix. For some practical and theoretical reasons, as an alternative to the use of the density matrix, the system is often described by the so-called quasi-density Wigner function, or equivalently, by using the quantum phase-space formalism. The Wigner formalism, for example, has found application in different areas of theoretical and applied physics. For the simulation of out-of-equilibrium systems in solid state physics, the Wigner formalism is generally preferred to the well investigated density matrix framework, because the quantum phase-space approach offers the possibility to describe various relaxation processes in an simple and intuitive form. Although the relaxation processes are ubiquitous in virtually all the real systems involving many particles or interactions with the environment, from the the microscopical point of view, they are sometime extremely difficult to characterize. The description of a system where the quantum mechanical coherence of the particle wave function is only partially lost or the understanding of how a pure quantum state evolves into a classical object, still constitutes an open challenge for the modern theoretical solid state physics (see for example (Giulini et al., 2003)). On

the contrary, when the particles experience many collisions and their coherence length is smaller than the De Broglie distance, an ensemble of particles can easily be described at the macroscopic level, by using for example diffusion equations (the mathematical literature refers to the "diffusive limit" of a particle gas). A strongly-interacting gas becomes essentially an ensemble of "classical particles" for which position and momentum are well defined function (and no longer operators) of time. The phase-space formalism, reveal itself to be a valuable instrument to fill the gap between this two opposite situations. The microscopic evolution of the system can be described exactly and the close analogy with the classical mechanics can be exploited in order to formulate some reasonable approximations to cope with the relaxation effects. Scattering phenomena can be included at different levels of approximation. The simplest approach is constituted by the Wigner-BGK model, where a relaxation-time term is added to the equation of motion. A more sophisticated model is obtained by the Wigner-Fokker-Plank theory, where the collision are included via diffusive terms. Finally, we mention the Wigner-Boltzmann equation where the particle-particle collisions are modeled by the Boltzmann scattering operator (see i. e. Jüngel (2009) for a general introduction to this methods). Furthermore, systems constituted by a gas where the particles are continuously exchanged with the environment ("open systems") are easily described by the quantum phase-space formalism. It results in special boundary conditions for the quasi-distribution function. In this paragraph, we give an elementary introduction to the Wigner quasi-distribution function and we illustrate some of the properties of the quantum phase-space formalism. A more general discussion will be given in sec. 3. For the sake of simplicity, we consider a spinless particle gas, described by the density matrix $\rho(\mathbf{x}_1, \mathbf{x}_2)$, in the presence of a static potential $V(\mathbf{r})$. Following (Wigner, 1932), we define the quasi-distribution function

$$f(\mathbf{r}, \mathbf{p}, t) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d_\eta} \rho\left(\mathbf{r} + \hbar\frac{\boldsymbol{\eta}}{2}, \mathbf{r} - \hbar\frac{\boldsymbol{\eta}}{2}, t\right) e^{-i\mathbf{p}\cdot\boldsymbol{\eta}} \, \mathrm{d}\boldsymbol{\eta} \, . \tag{1}$$

Here, d denotes the dimension of the space. The Wigner description of the quantum motion provides a framework that preserves many properties of the classical description of the particle motion. The equation of motion for the Wigner function writes (explicit calculation can be found for example in (Markowich, 1990))

$$\frac{\partial f}{\partial t} = -\frac{\mathbf{p}}{m} \cdot \nabla_{\mathbf{r}} f + \theta[f] \, , \tag{2}$$

where m is the particle mass and the pseudo-differential operator $\theta[f]$ is

$$\theta[f] = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d_\eta} \int_{\mathbb{R}^d_{p'}} \mathcal{D}\left(\mathbf{r}, \boldsymbol{\eta}\right) e^{i(\mathbf{p}-\mathbf{p}')\cdot\boldsymbol{\eta}} f(\mathbf{r}, \mathbf{p}') \, \mathrm{d}\boldsymbol{\eta} \, \mathrm{d}\mathbf{p}' \tag{3}$$

$$= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d_\eta} \mathcal{D}\left(\mathbf{r}, \boldsymbol{\eta}\right) \widetilde{f}(\mathbf{r}, \boldsymbol{\eta}) e^{i\mathbf{p}\cdot\boldsymbol{\eta}} \, \mathrm{d}\boldsymbol{\eta} \, , \tag{4}$$

with

$$\mathcal{D}\left(\mathbf{r}, \boldsymbol{\eta}\right) = \frac{i}{\hbar}\left[U\left(\mathbf{r} + \frac{\hbar}{2}\boldsymbol{\eta}\right) - U\left(\mathbf{r} - \frac{\hbar}{2}\boldsymbol{\eta}\right)\right] \, . \tag{5}$$

Equation (4) shows that the pseudo-differential operator acts just as a multiplication operator in the Fourier transformed space $\mathbf{r} - \boldsymbol{\eta}$. We used the following definition of Fourier transform

$\widetilde{f} = \mathcal{F}_{p\to\eta}[f]$:

$$\widetilde{f} = \int_{\mathbb{R}^d_p} f(\mathbf{r},\mathbf{p}) e^{-i\mathbf{p}\cdot\boldsymbol{\eta}}\, \mathrm{d}\mathbf{p}$$
$$f = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d_\eta} \widetilde{f}(\boldsymbol{\eta},\mathbf{p}) e^{i\mathbf{p}\cdot\boldsymbol{\eta}}\, \mathrm{d}\boldsymbol{\eta}\,.$$

The remarkable difference between the quantum phase-space equation of motion and the classical analogous (Liouville equation)

$$\frac{\partial f}{\partial t} = -\frac{\mathbf{p}}{m}\cdot\nabla_{\mathbf{r}} f - \mathbf{E}(\mathbf{r})\cdot\nabla_{\mathbf{p}} f\,, \tag{6}$$

is constituted by the presence of the pseudo-differential operator $\theta[f]$ that substitutes the classical force $\mathbf{E} = -\nabla_{\mathbf{r}} U$. The increasing of the complexity encountered when passing from Eq. (6) to Eq. (2) is justified by the possibility to describe all the phase-interference effects occurring between two different classical paths, and thus characterizing completely the particle motion at the atomic scale. The analogies and the differences between the Wigner transport equation and the classical Liouville equation have been the subject of many study and reports (see for example Markowich (1990)). In particular, we can convince ourselves that in the classical limit $\hbar \to 0$, Eq. (2) becomes Eq. (6), by noting that, formally, we have

$$\begin{aligned}\lim_{\hbar\to 0}\theta[f] &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d_\eta}\int_{\mathbb{R}^d_{p'}} i\boldsymbol{\eta}\cdot\nabla_{\mathbf{r}} U(\mathbf{r})\, e^{i(\mathbf{p}-\mathbf{p}')\cdot\boldsymbol{\eta}} f(\mathbf{r},\mathbf{p}')\, \mathrm{d}\boldsymbol{\eta}\, \mathrm{d}\mathbf{p}' \\ &= \frac{1}{(2\pi)^d}\nabla_{\mathbf{r}} U\cdot\frac{\partial}{\partial\mathbf{p}}\int_{\mathbb{R}^d_\eta}\int_{\mathbb{R}^d_{p'}} e^{i(\mathbf{p}-\mathbf{p}')\cdot\boldsymbol{\eta}} f(\mathbf{r},\mathbf{p}')\, \mathrm{d}\boldsymbol{\eta}\, \mathrm{d}\mathbf{p}' = \nabla_{\mathbf{r}} U\cdot\frac{\partial}{\partial\mathbf{p}} f(\mathbf{r},\mathbf{p})\,.\end{aligned}$$

This limit was rigorously proved in (Lions & Paul, 1993) and in (Markowich & Ringhofer, 1989), for sufficiently smooth potentials. From the definition of the Wigner function given by Eq. (1), we see that the $\mathbf{L}^2(\mathbb{R}^d_r \times \mathbb{R}^d_p)$ space constitutes the natural functional space where the theoretical study of the quantum phace-space motion can be addressed (Arnold, 2008).
The key properties through which the connection between the Wigner formulation of the quantum mechanics and the classical kinetic theory becomes evident, are the relationship between the Wigner function and the macroscopic thermodynamical quantities of the particle ensemble. In particular, the first two momenta of the Wigner distribution, taken with respect to the $\mathbf{p}$ variable, are

$$n(\mathbf{r},t) = \int_{\mathbb{R}^d_p} f(\mathbf{r},\mathbf{p},t)\ \mathrm{d}\mathbf{p} \tag{7}$$

and

$$J(\mathbf{r},t) = -\frac{q}{m}\int_{\mathbb{R}^d_p} \mathbf{p} f(\mathbf{r},\mathbf{p},t)\ \mathrm{d}\mathbf{p} \tag{8}$$

where n and J denote the particle and the current density, respectively. More generally, the expectation value of a physical quantity described classically by a function of the phase-space $\mathcal{A}(\mathbf{r},\mathbf{p},t)$ (relevant cases are for example the total Energy $\frac{p^2}{2m} + V(\mathbf{r})$ or the linear momentum $\mathbf{p}$), is given by

$$\langle\mathcal{A}\rangle = \int_{\mathbb{R}^d_p} \mathcal{A}(\mathbf{r},\mathbf{p},t) f(\mathbf{r},\mathbf{p},t)\ \mathrm{d}\mathbf{p}\, \mathrm{d}\mathbf{r}\,. \tag{9}$$

This equation reminds the ensemble average of a Gibbs system and coincides with the analogous classical formula.

3. Wigner-Weyl theory

The definition of the Wigner function given in Eq. (1) was introduced in 1932. It appears as a simple transformation of the density matrix. The spatial variable $\mathbf{r}$ of the Winger quasi-distribution function is the mean of the two points $(\mathbf{x}_1, \mathbf{x}_2)$ where the corresponding density matrix is evaluated (for this reason sometime is pictorially defined by "center of mass") and the momentum variable is the Fourier transform of the difference between the same points. The Wigner transform is a simple rotation in the plane $\mathbf{x}_1 - \mathbf{x}_2$, followed by a Fourier transform. Despite the apparently easy and straightforward form displayed by the Wigner transformation, its deep investigation, performed by Moyal (1949), revealed an unexpected connection with the former pioneering work of Weyl (1927), where the correspondence between quantum-mechanical operators in Hilbert space and ordinary functions was analyzed. Furthermore, when the Wigner framework was considered as an autonomous starting point for representing the quantum world, the presence of an internal logic or algebra, becomes evident. The Lie algebra of the quantum phase-space framework is defined in terms of the so-called Moyal $\star-$product, that becomes the key tool of this formalism. The noncommutative nature of the $\star-$product reflects the analogous property of the quantum Hilbert operators. In this context, following Weyl, by the term "quantization procedure" is intended a general correspondence principle between a function $\mathcal{A}(\mathbf{r}, \mathbf{p})$, defined on the classical phase-space, and some well-defined quantum operator $\widehat{\mathcal{A}}(\mathfrak{r}, \mathfrak{p})$ acting on the physical Hilbert space (here, in order to avoid confusion, we indicate by $\mathfrak{r}$ and $\mathfrak{p}$ the quantum mechanical position and the momentum operators, respectively). In quantum mechanics, observables are defined by Hilbert operators. We are interested in deriving a systematical and physically based extension of the concept of measurable quantities like energy, linear and orbital momentum. Due to the non-commutativity of the quantum operators $\mathfrak{r}$ and $\mathfrak{p}$, different choices are possible. In particular, based on the correspondence $\mathcal{A}(r, p) \rightarrow \widehat{\mathcal{A}}(\mathfrak{r}, \mathfrak{p})$, any other operator that differs from $\widehat{\mathcal{A}}(\mathfrak{r}, \mathfrak{p})$ in the order in which the operators $\mathfrak{r}$ and $\mathfrak{p}$ appear, can in principle been used equally well to define a new quantum operator. More specifically, at the Schrödinger level, the "position" and the "momentum" representations are alternative mathematical descriptions of the system, where the position and momentum operators $(\mathfrak{r}, \mathfrak{p})$ are formally substituted by the operators $(\mathbf{r}, -i\hbar\nabla_{\mathbf{r}})$ and $(i\hbar\nabla_{\mathbf{p}}, \mathbf{p})$, respectively. From a mathematical point of view, a clear distinction is made between position and momentum degrees of freedom of a particle (and which are represented by multiplicative or derivative operators). This is in contrast to the classical motion described in the phase-space, where the position and the momentum of a particle are treated equally, and they can be interpreted just as two different degrees of freedom of the system. As it will be clear in the following, the Weyl quantization procedure maintains this peculiarity and, from the mathematical point of view, position and momentum share the same properties.

The most common quantization procedures are the standard (anti-standard) Kirkwood ordering, the Weyl (symmetrical) ordering, and the normal (anti-normal) ordering. In particular, standard (anti-standard) ordering refers to a quantization procedure where, given a function $\mathcal{A}$ admitting a Taylor expansion, all of the $\mathfrak{p}$ operators appearing in the expansion of $\widehat{\mathcal{A}}(\mathfrak{r}, \mathfrak{p})$ follow (precede) the $\mathfrak{r}$ operators. A different choice is made in the Weyl ordering rule where each polynomial of the $\mathbf{p}$ and $\mathbf{r}$ variables is mapped, term by term, in a completely

ordered expression of $\mathfrak{r}$ and $\mathfrak{p}$. The generic binomial $\mathbf{p}^m\mathbf{r}^n$ becomes (see i. e. (Zachos et al., 2005))

$$\mathbf{p}^m\mathbf{r}^n \to \frac{1}{2^n}\sum_{r=0}^{n}\binom{n}{r}\mathfrak{r}^r\mathfrak{p}^m\mathfrak{r}^{n-r} = \frac{1}{2^m}\sum_{r=0}^{m}\binom{m}{r}\mathfrak{p}^r\mathfrak{r}^n\mathfrak{p}^{m-r}\,. \tag{10}$$

Following Cohen, (Cohen, 1966), one can consider a general class of quantization procedures defined in terms of an auxiliary function $\chi(r,p)$. The invertible map (for avoiding cumbersome expressions, the symbol of the integral indicates the integration over the whole space for all the variables)

$$\begin{aligned}\mathcal{A}(\mathbf{r},\mathbf{p}) &\equiv \mathrm{Tr}\left\{\widehat{\mathcal{A}}(\mathfrak{r},\mathfrak{p})\, e^{i(\mathbf{p}\mathfrak{r}+\mathbf{r}\mathfrak{p})}\chi(\mathbf{r},\mathbf{p})\right\}\\ &= \left(\frac{\hbar}{2\pi}\right)^d \int \left\langle \mathbf{r}' + \frac{\boldsymbol{\eta}\hbar}{2}\,\middle|\,\widehat{\mathcal{A}}\,\middle|\,\mathbf{r}' - \frac{\boldsymbol{\eta}\hbar}{2}\right\rangle \chi(\boldsymbol{\mu},\boldsymbol{\eta}) e^{i(\mathbf{r}-\mathbf{r}')\cdot\boldsymbol{\mu} - i\mathbf{p}\cdot\boldsymbol{\eta}}\,\mathrm{d}\boldsymbol{\mu}\,\mathrm{d}\boldsymbol{\eta}\,\mathrm{d}\mathbf{r}'\end{aligned} \tag{11}$$

defines the correspondence $\widehat{\mathcal{A}}(\mathfrak{r},\mathfrak{p}) \to \mathcal{A}(\mathbf{r},\mathbf{p})$. Different choices of the function χ describe different rules of association. In particular, if $\widehat{\mathcal{A}}$ is the density operator $\widehat{\rho}$ (representing a state of the system), from Eq. (11) we obtain the quantum distribution function f^χ. One of the main advantages in the application of the definition (11) is that the expectation value of the operator $\widehat{\mathcal{A}}(\mathfrak{r},\mathfrak{p})$ can be obtained by the mean value of the function $\mathcal{A}(\mathbf{r},\mathbf{p})$ under the "measure" f^χ

$$\mathrm{Tr}\left\{\widehat{\mathcal{A}}(\mathfrak{r},\mathfrak{p})\,\widehat{\rho}(\mathfrak{r},\mathfrak{p},t)\right\} = \int \mathcal{A}^\chi(\mathbf{r},\mathbf{p})\, f^\chi(\mathbf{r},\mathbf{p},t)\;\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{r}\,.$$

As particular cases, it is possible to recover the definition of the most common quasi-probability distribution functions (classification scheme of Cohen). For example for $\chi = e^{\mp i\frac{\hbar}{2}\mu\eta}$ we obtain the standard $(-)$ or anti-standard $(+)$ ordered Kirkwood distribution function. Hereafter, we limit ourselves to consider the case $\chi = 1$, which gives the Weyl ordering rules. The function f^χ becomes the Wigner quasi-distribution

$$f(\mathbf{r},\mathbf{p}) = \frac{1}{(2\pi)^d}\int \left\langle \mathbf{r} + \frac{\boldsymbol{\eta}\hbar}{2}\,\middle|\,\widehat{\rho}\,\middle|\,\mathbf{r} - \frac{\boldsymbol{\eta}\hbar}{2}\right\rangle e^{-ip\cdot\boldsymbol{\eta}}\,\mathrm{d}\boldsymbol{\eta}\,. \tag{12}$$

The Weyl-Moyal theory provides the mathematical ground and a rigorous link between a phase-space function and a symmetrically ordered operator. More into detail, the correspondence between $\widehat{\mathcal{A}}$ and the function $\mathcal{A}(\mathbf{r},\mathbf{p})$ (called the symbol of the operator) is provided by the map $\mathcal{W}[\mathcal{A}] = \widehat{\mathcal{A}}$ (Folland, 1989)

$$\left(\widehat{\mathcal{A}}h\right)(\mathbf{x}) = \mathcal{W}[\mathcal{A}]\,h = \frac{1}{(2\pi\hbar)^d}\int \mathcal{A}\left(\frac{\mathbf{x}+\mathbf{y}}{2},\mathbf{p}\right) h(\mathbf{y})\, e^{\frac{i}{\hbar}(\mathbf{x}-\mathbf{y})\cdot\mathbf{p}}\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{p}\,. \tag{13}$$

Here, h is a generic function. The inverse of $\mathcal{W}$ is given by the Wigner transform

$$\mathcal{A}(\mathbf{r},\mathbf{p}) = \mathcal{W}^{-1}\left[\widehat{\mathcal{A}}\right](\mathbf{r},\mathbf{p}) = \int \mathcal{K}_{\mathcal{A}}\left(\mathbf{r}+\frac{\boldsymbol{\eta}}{2},\mathbf{r}-\frac{\boldsymbol{\eta}}{2}\right) e^{-\frac{i}{\hbar}\mathbf{p}\cdot\boldsymbol{\eta}}\,\mathrm{d}\boldsymbol{\eta}\,, \tag{14}$$

where $\mathcal{K}_{\mathcal{A}}(\mathbf{x},\mathbf{y})$ is the kernel of the operator $\widehat{\mathcal{A}}$. Let us now fix an orthonormal basis $\psi = \{\psi_i \mid i = 1,2,\ldots\}$. A mixed state is defined by the density operator $\hat{\mathcal{S}}_\psi$

$$\left(\widehat{\mathcal{S}}_\psi\, h\right)(\mathbf{x}) = \int \rho_\psi(\mathbf{x},\mathbf{x}')h(\mathbf{x}')\,\mathrm{d}\mathbf{x}'$$

whose kernel is the density matrix. In the basis $\{\psi_i\}$

$$\rho_\psi(\mathbf{x},\mathbf{x}') = \sum_{i,j} \rho_{ij}\ \psi_i(\mathbf{x})\overline{\psi}_j(\mathbf{x}')\ , \tag{15}$$

where the overbar means conjugation. The von Neumann equation gives the evolution of the density operator $\widehat{\mathcal{S}}_\psi = \widehat{\mathcal{S}}_\psi(t)$ in the presence of the Hamiltonian $\widehat{\mathcal{H}}$:

$$i\hbar\frac{\partial\widehat{\mathcal{S}}_\psi}{\partial t} = \left[\widehat{\mathcal{H}},\widehat{\mathcal{S}}_\psi\right] \tag{16}$$

where, as usual, the brackets denote the commutator. The equivalent quantum phase-space evolution equation can be obtained by applying the Wigner transform. We obtain

$$i\hbar\frac{\partial f_\psi}{\partial t} = \left[\mathcal{H}, f_\psi\right]_\star = \mathcal{H}\star f_\psi - f_\psi\star\mathcal{H} \tag{17}$$

where the symbol $(2\pi\hbar)^d f_\psi(\mathbf{r},\mathbf{p}) \equiv \mathcal{S}_\psi = \mathcal{W}^{-1}\left[\hat{\mathcal{S}}_\psi\right]$ is the Wigner transform of $\rho_\psi(\mathbf{x},\mathbf{x}')$ (see Eq. (1) and Eq. (12)) and we used the following fundamental property

$$\mathcal{W}^{-1}\left[\widehat{\mathcal{A}}\,\widehat{\mathcal{B}}\right] = \mathcal{A}\star\mathcal{B}. \tag{18}$$

For symbols sufficiently regular, the star-Moyal product $\star$ is defined as

$$\begin{aligned}
\mathcal{A}\star\mathcal{B} &\equiv \mathcal{A}\, e^{\frac{i\hbar}{2}\left(\overleftarrow{\nabla}_\mathbf{r}\cdot\overrightarrow{\nabla}_\mathbf{p} - \overleftarrow{\nabla}_\mathbf{p}\cdot\overrightarrow{\nabla}_\mathbf{r}\right)}\ \mathcal{B} \\
&= \sum_n \left(\frac{i\hbar}{2}\right)^n \frac{1}{n!}\mathcal{A}(\mathbf{r},\mathbf{p})\left[\overleftarrow{\nabla}_\mathbf{r}\cdot\overrightarrow{\nabla}_\mathbf{p} - \overleftarrow{\nabla}_\mathbf{p}\cdot\overrightarrow{\nabla}_\mathbf{r}\right]^n \mathcal{B}(\mathbf{r},\mathbf{p}) \\
&= \sum_n \sum_{k=0}^{n} \left(\frac{i\hbar}{2}\right)^n \frac{(-1)^k}{n!}\binom{n}{k}\mathcal{A}(\mathbf{r},\mathbf{p})\left(\overleftarrow{\nabla}_\mathbf{r}\cdot\overrightarrow{\nabla}_\mathbf{p}\right)^{n-k}\left(\overleftarrow{\nabla}_\mathbf{p}\cdot\overrightarrow{\nabla}_\mathbf{r}\right)^k \mathcal{B}(\mathbf{r},\mathbf{p}),
\end{aligned} \tag{19}$$

where the arrows indicate on which operator the gradients act. The Moyal product can be expressed also in integral form (that extends the definition (19) to simply $\mathbf{L}^2$ symbols):

$$\begin{aligned}
\mathcal{A}\star\mathcal{B} &= \frac{1}{(2\pi)^{2d}}\int \mathcal{A}\left(\mathbf{r}-\frac{\hbar}{2}\boldsymbol{\eta},\mathbf{p}+\frac{\hbar}{2}\boldsymbol{\mu}\right)\mathcal{B}\left(\mathbf{r}',\mathbf{p}'\right)e^{i(\mathbf{r}-\mathbf{r}')\cdot\boldsymbol{\mu}+i(\mathbf{p}-\mathbf{p}')\cdot\boldsymbol{\eta}}\,\mathrm{d}\boldsymbol{\mu}\,\mathrm{d}\mathbf{r}'\,\mathrm{d}\boldsymbol{\eta}\,\mathrm{d}\mathbf{p}' \\
&= \frac{1}{(2\pi)^{2d}}\int \mathcal{A}\left(\mathbf{r}',\mathbf{p}'\right)\mathcal{B}\left(\mathbf{r}+\frac{\hbar}{2}\boldsymbol{\eta},\mathbf{p}-\frac{\hbar}{2}\boldsymbol{\mu}\right)e^{i(\mathbf{r}-\mathbf{r}')\cdot\boldsymbol{\mu}+i(\mathbf{p}-\mathbf{p}')\cdot\boldsymbol{\eta}}\,\mathrm{d}\boldsymbol{\mu}\,\mathrm{d}\mathbf{r}'\,\mathrm{d}\boldsymbol{\eta}\,\mathrm{d}\mathbf{p}'\ .
\end{aligned}$$

In particular, if both operators depend only on one variable ($\mathbf{r}$ or $\mathbf{p}$), the Moyal product becomes the ordinary product. For a one-dimensional system the Moyal product simplifies

$$\mathcal{A}\star\mathcal{B} = \sum_{k=0}^{\infty}\frac{\hbar^k}{(2i)^k}\sum_{|\alpha|+|\beta|=k}\frac{(-1)^{|\alpha|}}{\alpha!\beta!}\left(\partial_r^\alpha\partial_p^\beta\mathcal{A}\right)\left(\partial_p^\alpha\partial_r^\beta\mathcal{B}\right) \tag{20}$$

and

$$\begin{aligned}
[\mathcal{A},\mathcal{B}]_\star = \sum_{k=1,3,5,\ldots}^{\infty}\frac{\hbar^k}{(2i)^k}\sum_{0<\beta<k/2}\frac{2(-1)^{\beta+1}}{(k-\beta)!\beta!}\Big[&\left(\partial_r^{k-\beta}\partial_p^\beta\mathcal{A}\right)\left(\partial_p^{k-\beta}\partial_r^\beta\mathcal{B}\right) \\
&-\left(\partial_r^{k-\beta}\partial_p^\beta\mathcal{B}\right)\left(\partial_p^{k-\beta}\partial_r^\beta\mathcal{A}\right)\Big].
\end{aligned}$$

3.1 Generalization of the Wigner-Moyal map

A separable Hilbert space can be characterized by a complete set of basis elements ψ_i or, equivalently, by a unitary transformation Θ (defined in terms of the projection of the ψ_i set on a reference basis). The class of unitary operators $\mathcal{C}(\Theta)$ defines all the alternative sets of basis elements or "representations" of the Hilbert space. Once a representation is defined, the relevant physical variables and the quantum operator can be explicitly addressed. Unitary transformations are a simple and powerful instrument for investigating different and equivalent mathematical formulations of a given physical situation. We study the modification of the explicit form of the Hamiltonian $\mathcal{H}$ (and thus of the equation of motion (17)), induced by a unitary transformation. We consider a unitary operator $\widehat{\Theta}$ and the "rotated" orthonormal basis $\varphi = \{\varphi_i \mid i = 1, 2, \ldots\}$, where $\varphi_i = \widehat{\Theta}\ \psi_i$. It is easy to verify that the following property

$$\Theta^{-1}(\mathbf{r}, \mathbf{p}) = \overline{\Theta}(\mathbf{r}, \mathbf{p}), \tag{21}$$

holds true, where, according to Eq. (14), Θ (Θ^{-1}) is the Weyl symbol of $\widehat{\Theta}$ ($\widehat{\Theta}^{-1}$). The phase-space representation of the state under the unitary transformation $\widehat{\Theta}$ will be denoted by
$(2\pi\hbar)^d f_\varphi \equiv \mathcal{W}^{-1}\left[\widehat{\mathcal{S}}_\varphi\right]$, where

$$\widehat{\mathcal{S}}_\varphi = \widehat{\Theta}\ \widehat{\mathcal{S}}_\psi\ \widehat{\Theta}^\dagger. \tag{22}$$

is the new density operator of the system. Here, the dagger denotes the adjoint operator. By using Eq. (21) it is immediate to verify that the equation of motion for f_φ is still expressed by Eq. (17) with the Hamiltonian $\mathcal{H}' = \Theta \star \mathcal{H} \star \Theta^{-1}$. Explicitly, $\mathcal{H}' \equiv \mathcal{W}^{-1}\left[\widehat{\Theta}\ \widehat{\mathcal{H}}\ \widehat{\Theta}^\dagger\right]$ is given by

$$\mathcal{H}'(\mathbf{r}, \mathbf{p}) = \frac{1}{(2\pi\hbar)^{2d}} \int \Theta\left(\frac{\mathbf{r} + \mathbf{r}' + \mathbf{r}''}{2}, \frac{\mathbf{p} + \mathbf{p}' + \mathbf{p}''}{2}\right) \Theta^{-1}\left(\frac{\mathbf{r} + \mathbf{r}' - \mathbf{r}''}{2}, \frac{\mathbf{p} + \mathbf{p}' - \mathbf{p}''}{2}\right) \times$$
$$\mathcal{H}(\mathbf{r}', \mathbf{p}') e^{\frac{i}{\hbar}[(\mathbf{r} - \mathbf{r}') \cdot \mathbf{p}'' - (\mathbf{p} - \mathbf{p}') \cdot \mathbf{r}'']}\, d\mathbf{r}'\, d\mathbf{p}'\, d\mathbf{r}''\, d\mathbf{p}''. \tag{23}$$

When passing from the position representation (where the basis elements in the Schrödinger formalism are the Dirac delta distributions and where $\widehat{\Theta}$ is the identity operator), to another possible representation, the Hamiltonian operator modifies according to formula (23). Although the mathematical structure of the equation of motion can be strongly affected by such a basis rotation, the distribution function f_φ is always defined in terms of the classical conjugated variables of position and momentum. The generality of this approach is ensured by the bijective correspondence between a generical unitary transformation (describing all the physical relevant basis transformation) and a framework where the description of the problem is a priori in the phase-space.

3.2 Application to multiband structures: graphene

The previous formalism is particularly convenient for the description of quantum particles with discrete degrees of freedom like spin, pseudo-spin or semiconductor band index. The mathematical structure, emerged in sec. 3.1, can be used in order to define a suitable set of r-p-dependent eigenspaces (with a consequent set of projectors) of the "classical-like" Hamiltonian matrix (that in our case is just the symbol of the Hamiltonian operator). Consequently, a "quasi-diagonalized" matrix representation of the Wigner dynamics can be obtained. This special starting point of the phase-space representation, aids to obtain information on the particle transitions among this countable set of eigenspaces. From a

Physical point of view, these transitions could represent, case by case, spin flip, jumping of a particle from conduction to valence band or particle-antiparticle conversion. The analysis performed in sections 3-3.1 providing Eqs. (16)-(23), maintains its validity when $\widehat{\Theta}$, $\widehat{\mathcal{H}}$ are $n \times n$ matrices of operators (and, consequently the symbols Θ, $\mathcal{H}$ are matrices of functions). This for example, is the standard situation for the Schrödinger-Hilbert space of the form $\mathbf{L}^2\left(\mathbb{R}^d_{\mathbf{x}};\mathbb{C}^n\right)$. The only new prescription is to maintain the order in which the operators and symbols appear in the formulae. To concretize to our exposition, we apply the phase-space formalism to graphene and we present the explicit form of the equation of motion.

Graphene is the two-dimensional honeycomb-lattice allotropic form of carbon. Its discovery stimulated a great interest in the scientific community. In fact, this novel functional material displays some unique electronic properties (see for example (Neto et al., 2009) for a general introduction to graphene). In a quite wide range of energy around the Dirac point, electrons and holes propagate as massless Fermions and the Hamiltonian writes (Beenakker et al., 2008)

$$\widehat{\mathcal{H}} = \widehat{\mathcal{H}}_0 + \sigma_0 U(\mathbf{r}) \,, \tag{24}$$

$$\widehat{\mathcal{H}}_0 = -i\, v_F \hbar\, \boldsymbol{\sigma} \cdot \nabla_{\mathbf{r}} = v_F \hbar \begin{pmatrix} 0 & -i\frac{\partial}{\partial x} - \frac{\partial}{\partial y} \\ -i\frac{\partial}{\partial x} + \frac{\partial}{\partial y} & 0 \end{pmatrix} \,, \tag{25}$$

which describes the motion of an electron-hole pair in a graphene sheet in the presence of an external potential $U(\mathbf{r})$. Here, v_F is the Fermi velocity, $\boldsymbol{\sigma} = \left(\sigma_x, \sigma_y, \sigma_z\right)$ indicate the Pauli vector-matrix and σ_0 denotes the identity 2×2 matrix. The upper and lower bands are sometimes denoted by pseudo-spin components of the particle, since the Hamiltonian can be interpreted as an effective momentum-dependent magnetic field $\mathbf{h} \propto \boldsymbol{\sigma} \cdot \nabla_{\mathbf{r}}$.

The application of the theory exposed in sec. 3.1 leads us to consider the density operator $\widehat{\mathcal{S}'} \equiv \widehat{\Theta}\ \widehat{\mathcal{S}}\ \widehat{\Theta}^\dagger$ where $\widehat{\Theta}\left(\mathbf{r}, \nabla_{\mathbf{r}}\right)$ is a unitary 2×2 matrix operator. The approach generally adopted for simplifying the description of a quantum system, is the use of a coordinate framework where the Hamiltonian is diagonal. The graphene Hamiltonian contains off-diagonal terms proportional to the momentum. Since position and momentum are non-commuting quantities, it is not possible to diagonalize $\widehat{\mathcal{H}}$ simultaneously in the position and in the momentum space. Anyway, up to the zero order in $\hbar$, an approximate (r-p)-diagonalization of the Hamiltonian can be obtained. We take advantage of the Weyl correspondence principle and consider the symbol $\Theta\left(\mathbf{r}, \mathbf{p}\right) \equiv \mathcal{W}^{-1}\left[\widehat{\Theta}\right]$. Here, $\Theta\left(\mathbf{r}, \mathbf{p}\right)$ is a unitary matrix parametrized by the $\mathbf{r} - \mathbf{p}$ coordinates. It can be used in order to diagonalize the Hamiltonian symbol $\mathcal{H} = v_F\, \boldsymbol{\sigma} \cdot \mathbf{p} + \sigma_0 U(\mathbf{r})$. With

$$\Theta(\mathbf{p}) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \frac{p_x - i p_y}{\sqrt{p_x^2 + p_y^2}} \\ \frac{p_x + i p_y}{\sqrt{p_x^2 + p_y^2}} & -1 \end{pmatrix} \tag{26}$$

we have

$$\Theta \mathcal{H} \Theta^\dagger = \Lambda \tag{27}$$

where $\Lambda(\mathbf{p}) = \sigma_z v_F |\mathbf{p}| + U(\mathbf{r})$ is the relativistic-like spectrum of the graphene sheet. The equation of motion for the new Wigner symbol $\mathcal{S}'$ becomes (see (Morandi & Schürrer, 2011)

for the details of the calculation)

$$i\hbar\frac{\partial\mathcal{S}'}{\partial t} = \left[\mathcal{U}' + \Lambda(\mathbf{p}), \mathcal{S}'\right]_\star . \tag{28}$$

The symbol $\mathcal{U}'(\mathbf{r},\mathbf{p})$ is given by

$$\mathcal{U}'(\mathbf{r},\mathbf{p}) = \Theta \star U(\mathbf{r}) \star \Theta^\dagger \tag{29}$$

and writes explicitly as

$$\mathcal{U}'(\mathbf{r},\mathbf{p}) = \frac{1}{(2\pi)^2}\int \Theta\left(\mathbf{p}+\frac{\hbar}{2}\boldsymbol{\mu}\right)\Theta^\dagger\left(\mathbf{p}-\frac{\hbar}{2}\boldsymbol{\mu}\right)U(\mathbf{r}')e^{i(\mathbf{r}-\mathbf{r}')\cdot\boldsymbol{\mu}}\,\mathrm{d}\boldsymbol{\mu}\,\mathrm{d}\mathbf{r}' .$$

We address explicitly the components of $\mathcal{S}'$, by denoting

$$\mathcal{S}' \equiv (2\pi\hbar)^2\begin{pmatrix} f^+(\mathbf{r},\mathbf{p}) & f^i(\mathbf{r},\mathbf{p}) \\ \overline{f^i}(\mathbf{r},\mathbf{p}) & f^-(\mathbf{r},\mathbf{p})\end{pmatrix} . \tag{30}$$

Equation (28) is written in terms of the Moyal commutator and defines implicitly a non-local evolution operator for the matrix-Wigner function $\mathcal{S}'$. It requires the evaluation of infinite-order derivatives with respect to the variables $\mathbf{r}$ and $\mathbf{p}$. The commutators appearing in Eq. (28) can be written in integral form as

$$\left[\Lambda,\mathcal{S}'\right]_\star = \frac{1}{(2\pi)^2}\int\left[\Lambda\left(\mathbf{p}+\frac{\hbar}{2}\boldsymbol{\mu}\right)\mathcal{S}'(\mathbf{r}',\mathbf{p}) - \mathcal{S}'(\mathbf{r}',\mathbf{p})\,\Lambda\left(\mathbf{p}-\frac{\hbar}{2}\boldsymbol{\mu}\right)\right]e^{i(\mathbf{r}-\mathbf{r}')\cdot\boldsymbol{\mu}}\,\mathrm{d}\boldsymbol{\mu}\,\mathrm{d}\mathbf{r}' \tag{31}$$

$$\left[\mathcal{U}',\mathcal{S}'\right]_\star = \frac{1}{(2\pi)^4}\int\left[\mathcal{U}'\left(\mathbf{r}-\frac{\hbar}{2}\boldsymbol{\eta},\mathbf{p}+\frac{\hbar}{2}\boldsymbol{\mu}\right)\mathcal{S}'(\mathbf{r}',\mathbf{p}') - \mathcal{S}'(\mathbf{r}',\mathbf{p}')\,\mathcal{U}'\left(\mathbf{r}+\frac{\hbar}{2}\boldsymbol{\eta},\mathbf{p}-\frac{\hbar}{2}\boldsymbol{\mu}\right)\right] \times e^{i(\mathbf{r}-\mathbf{r}')\cdot\boldsymbol{\mu}+i(\mathbf{p}-\mathbf{p}')\cdot\boldsymbol{\eta}}\,\mathrm{d}\boldsymbol{\mu}\,\mathrm{d}\mathbf{r}'\,\mathrm{d}\boldsymbol{\eta}\,\mathrm{d}\mathbf{p}' . \tag{32}$$

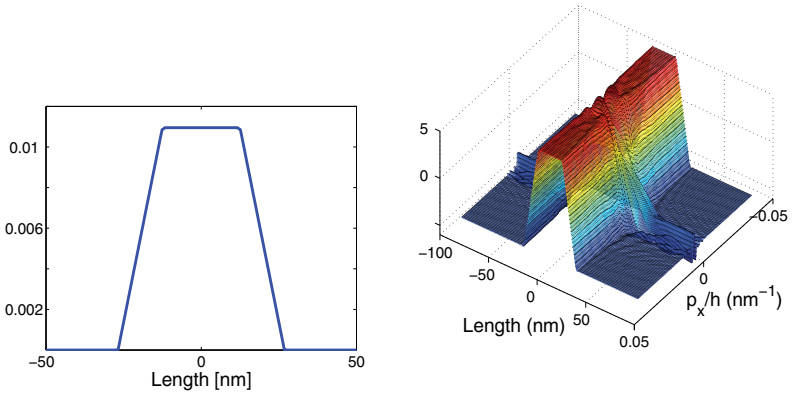
The commutator of Eq. (31) describes the free motion of the electron-hole pairs in the upper and lower conically shaped energy surfaces. When we discard the external potential U, the evolution of the particles f^+ (f^-) belonging to the upper (lower) part of the spectrum is described by

$$\frac{\partial f^\pm}{\partial t} = \pm\frac{1}{(2\pi)^2}\int\left[E\left(\mathbf{p}+\frac{\hbar}{2}\boldsymbol{\mu}\right) - E\left(\mathbf{p}-\frac{\hbar}{2}\boldsymbol{\mu}\right)\right]f^\pm(\mathbf{r}',\mathbf{p})\,e^{i(\mathbf{r}-\mathbf{r}')\cdot\boldsymbol{\mu}}\,\mathrm{d}\boldsymbol{\mu}\,\mathrm{d}\mathbf{r}' . \tag{33}$$

By expanding up to the leading order in $\hbar$, the previous equation reduces to

$$\frac{\partial f^\pm}{\partial t} \simeq \pm v_F\frac{\mathbf{p}}{|\mathbf{p}|}\cdot\nabla_{\mathbf{r}}f^\pm \tag{34}$$

which is equal to the semi-classical free evolution of the two-particle system in the graphene band structure. We emphasize that the usual semi-classical prescription $v_g = \nabla_{\mathbf{p}}E = v_F\frac{\mathbf{p}}{|\mathbf{p}|}$, where v_g is the group velocity, is automatically fulfilled. As expected from a physical point of view, the coupling between the bands arises from the presence of an external field $U(\mathbf{r})$ which



(a) External potential $U(\mathbf{r})$ (b) First component of $\mathcal{U}'(\mathbf{r},\mathbf{p})$

Fig. 1. Comparison between the classical potential U and the momentum dependent pseudo-potential $\mathcal{U}'$.

perturbs the periodic crystal potential. This is described by Eq. (32). In order to illustrate the main characteristics of the pseudo-potential $\mathcal{U}'(\mathbf{r},\mathbf{p})$, in fig. 1 we depict the first component $[\mathcal{U}']_{++}$ of the matrix $\mathcal{U}'$, when the external potential $U(\mathbf{r})$ (represented in the sub-plot 1-(a)) is a single barrier. Equation (29) shows that the elements of the 2×2 matrix $\mathcal{U}'$ depend both on the position $\mathbf{r}$ and the momentum $\mathbf{p}$. The main corrections to the potential arise around $p_x = 0$, whereas $[\mathcal{U}']_{++}$ stays practically identical to U for high values of the momentum p_x. This reflects the presence of the singular behavior of the particle-hole motion in the proximity of the Dirac point (see the discussion concerning this point given in (Morandi & Schürrer, 2011)). The effective potential $[\mathcal{U}']_{++}$ represents the potential "seen" by the particles located in the upper Dirac cone. For small values of p_x, the original squared shape of the potential changes dramatically. The effective potential $[\mathcal{U}']_{++}$ becomes smooth and a long range effective electric field (the gradient of $[\mathcal{U}']_{++}$) is produced. Around $\mathbf{p} = 0$, a barrier or, equivalently, a trap potential becomes highly non-local. It is somehow "spread over the sheet" and, in the case of a trap, its localization effect is greatly reduced.
The equation of motion (28) reproduces the full quantum ballistic motion of the particle-hole gas. In the numerical study presented in (Morandi & Schürrer, 2011b), one of the main quantum transport effects, namely the Klein tunneling, is investigated. The numerical study of the full Wigner system in the presence of a discontinuous potential is presented in (Morandi & Schürrer, 2011). The high computational effort required for solving the full ballistic motion and the need of developing appropriate numerical schemes, limits the practical application of the exact theory. This becomes particularly constraining in view of the simulation of real devices containing dissipative effects like, for example, electron-phonon collisions, that further increase the complexity of the problem. The Wigner formalism is well suited for the inclusion of weak dissipative effects. The overall theoretical and computational complexity displayed by the pseudo-spinorial Wigner dynamics, can be reduced by exploiting some general properties of the system that characterize the application of the multiband Wigner system to real structures (typically, the presence of fast and slow time scaling can be exploited). Approximated models or iterative methods can be derived (see (Morandi & Schürrer, 2011) and (Morandi, 2009) for the application to graphene and to interband diodes, and (Morandi, 2010) for the WKB method in semiconductors).

3.3 Application to multiband structures: correction to the classical trajectory in semiconductors.

We investigate the application of the multiband Wigner formalism to the semiconductor structures. The study of the particle motion in semiconductors has attracted the scientific community, e. g., to the sometime anti-intuitive properties of Bloch waves (especially compared with the classical counterpart). Moreover, the interest has been renewed by the discovery of the unipolar and bipolar junctions and the final impulse to the semiconductor research was given by the unrestrainable progress of the modern industry of electronic devices. An important branch of the semiconductor research is now constituted by the numerical simulation applied to the particle transport. In particular, the continuous miniaturization of field effect transistors (length of a MOS channel approaches the ten nm) imposes the use of a full quantum mechanical (or at least a quantum-correct) model for the correct reproduction of the device characteristic. Beside the Green function formalism and the direct application of the Schrödinger approach, the Wigner framework is a widely employed tool for device simulation. Anyway, most attention is usually devoted to the interband motion since it is often implicitly assumed that electron motion is supported only by one single band. This approximation is based on the assumption that the band-to-band transition probability vanishes exponentially with increasing band gaps (that, for example, in silicon is around one eV), so that under normal conditions all the multiband effects can be discarded. However, this assumption is violated in many heterostructures (devices obtained by connecting semiconductors with different chemical compounds), or when a strong electric field is applied to a normal diode. In both cases electrons are free to flow from one band to another. Beside the evident modification of how the device operates (a new channel for the particle transport becomes available), there is also a more subtle consequence. The application of a strong electric field for example, is able to provide a strong local modification of the electronic spectrum. Since high electric fields could induce a strong mixing of the bands, the Bloch band theory becomes inadequate to describe the particle transport. Even when the particle does not undergo a complete band transition, its motion becomes affected by the interference of the other bands. In the following, we show how these problem can be attacked with the use of the multiband Wigner formalism.

A multiband transport model, based on the Wigner-function approach, was introduced in (Demeio et al., 2006) and in (Unlu et al., 2004) the multiband equation of motion is derived by using the generalized Kadanoff-Baym non-equilibrium Green's function formalism. The model equations there derived are still too hard to be solved numerically. In order to maintain easily the discussion of the problem, we consider a simple model, where only two bands, namely one conduction and one valence band, are retained. We adopt the multiband envelope function model (MEF) described in Ref. (Morandi & Modugno, 2005). This model is derived within the $k \cdot p$ framework and is so far very general. In particular, this approach is focused on the description of the electron transport in devices where tunneling mechanisms between different bands are induced by an external applied bias U. It has been recently applied to some resonant diodes showing self-sustained oscillations (Alvaro & Bonilla, 2010). Under this hypothesis the MEF model furnishes the following Hamiltonian

$$\widehat{\mathcal{H}} = \begin{pmatrix} E_c + U(\mathbf{r}) - \dfrac{\hbar^2}{2m^*}\Delta_{\mathbf{r}} & -\dfrac{\hbar}{m_0}\dfrac{\mathbf{p}_K \cdot \mathcal{E}(\mathbf{r})}{E_g} \\ -\dfrac{\hbar}{m_0}\dfrac{\mathbf{p}_K \cdot \mathcal{E}(\mathbf{r})}{E_g} & E_v + U(\mathbf{r}) + \dfrac{\hbar^2}{2m^*}\Delta_{\mathbf{r}} \end{pmatrix} . \tag{35}$$

Here, E_c (E_v) is the minimum (maximum) of the conduction (valence) energy band, $\mathbf{p}_k$ is the Kane momentum, m_0, m^* are the bare and the effective mass of the electron and U ($\mathcal{E} = \nabla_{\mathbf{r}} U$) is the "external" potential which takes into account different effects, like the bias voltage applied across the device, the contribution from the doping impurities and from the self-consistent field produced by the mobile electronic charge. According to Eq. (27), the multiband system is characterized by the matrix

$$\Theta = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{1+\sigma} & \sqrt{1-\sigma} \\ -\sqrt{1-\sigma} & \sqrt{1+\sigma} \end{pmatrix} , \tag{36}$$

where

$$\sigma = \frac{\Omega}{\sqrt{P_R^2 + \Omega^2}} ,$$

with $\Omega(\mathbf{p}) = \frac{E_g}{2} + \frac{|\mathbf{p}|^2}{2m^*}$ $P_R(\mathbf{r}) = -\hbar \frac{\mathbf{p}_k \cdot \mathcal{E}(\mathbf{r})}{E_g m_0}$ and $E_g = E_c - E_v$ is the band gap. The eigenvalues of the Hamiltonian are $\mathcal{H}^{\pm}(\mathbf{r}, \mathbf{p}) = \pm\sqrt{P_R^2 + \Omega^2} + U$. Here we limit ourselves to discuss the system obtained by expanding the full quantum equation of motion given in Eq. (28) up to the first order in $\hbar$ (the study of the full quantum system is addressed in (Morandi, 2009)). With the definition (in order to avoid confusion with the graphene Wigner functions defined in Eq. (30), we changed the name of the various components of the matrix)

$$\mathcal{S}' \equiv (2\pi\hbar)^3 \begin{pmatrix} h_c(\mathbf{r}, \mathbf{p}) & h_{cv}(\mathbf{r}, \mathbf{p}) \\ \overline{h_{cv}}(\mathbf{r}, \mathbf{p}) & h_v(\mathbf{r}, \mathbf{p}) \end{pmatrix} . \tag{37}$$

We obtain the following equations of motion

$$\frac{\partial h_c}{\partial t} = -\nabla_{\mathbf{p}} \mathcal{H}^+ \cdot \nabla_{\mathbf{r}} h_c + \nabla_{\mathbf{r}} \mathcal{H}^+ \cdot \nabla_{\mathbf{p}} h_c - 2\xi\, \Re(h_{cv}) \tag{38}$$

$$\frac{\partial h_v}{\partial t} = -\nabla_{\mathbf{p}} \mathcal{H}^- \cdot \nabla_{\mathbf{r}} h_v + \nabla_{\mathbf{r}} \mathcal{H}^- \cdot \nabla_{\mathbf{p}} h_v + 2\xi\, \Re(h_{cv}) \tag{39}$$

$$\frac{\partial h_{cv}}{\partial t} = -\frac{i}{\hbar} \left(\mathcal{H}^+ - \mathcal{H}^- \right) h_{cv} + \mathcal{E} \cdot \nabla_{\mathbf{p}} h_{cv} + \xi \left(h_c - h_v \right) \tag{40}$$

where $\Re$ denotes the real part and

$$\xi = \frac{P_R}{P_R^2 + \Omega^2} \frac{\mathcal{E} \cdot \mathbf{p}}{m^*} . \tag{41}$$

Here, h_c and h_v represent the Wigner quasi-distribution functions of particles in a regime of strong band-to-band coupling. They differ from the analogous functions based on a direct application of the projection of the particle motion in the Bloch basis. The system of Eq. (38)-(40) shows that, up to the zero order in $\hbar$, the Wigner functions h_c (h_v) follows the Hamiltonian flux generated by $\mathcal{H}^+$ ($\mathcal{H}^-$). Furthermore, the term $\mathcal{H}^+ - \mathcal{H}^- = 2\sqrt{{P_R}^2 + \Omega^2}$ in Eq. (40) induces fast-in-time oscillations (whose frequency is of the order of $E_g/\hbar$) which, up to zero order in $\hbar$, decouple h_{cv} from the slowly varying functions h_c and h_v. This aspect is examined in sec. 3.4. We explore the single band limit of Eqs. (38)-(40). From the physical point of view, we expect that when the electric field goes to zero or the band gap goes to

infinity, all the multiband corrections become negligible and the dynamics of the electrons in the conduction band decouples from those in the valence band. It is convenient to define the parameter $\mathrm{Y} = \frac{P_R}{\Omega}$ that vanishes in the single band limits $\mathcal{E}, 1/E_g \to 0$. When $\mathrm{Y} \to 0$ the evolution of h_c and h_v is described by two Liouville equations (one for each band) with the Hamiltonian

$$\begin{aligned} \mathcal{H}^{\pm} &= \pm\sqrt{{P_R}^2 + \Omega^2} + U(\mathbf{r}) \\ &= \pm\frac{|\mathbf{p}|^2}{2m^*} + U(\mathbf{r}) \pm \frac{E_g}{2}\left(1 + \frac{\Omega}{E_g}\mathrm{Y}^2\right) + o\left(\mathrm{Y}^2\right) . \end{aligned} \tag{42}$$

Equation (42) shows that the eigenvalues of the Hamiltonian symbol, provide a simple quantum correction to the classical single band Hamiltonian $\mathcal{H}_{sb} = U(\mathbf{r}) \pm \frac{|\mathbf{p}|^2}{2m^*}$. The particles follow a new trajectory defined by

$$\begin{cases} \dot{\mathbf{r}} = & \dfrac{1}{\sqrt{1 + \left(\frac{P_R}{\Omega}\right)^2}} \dfrac{\mathbf{p}}{m^*} \\ \dot{\mathbf{p}} = -\mathcal{E} - & \dfrac{1}{\sqrt{1 + \left(\frac{\Omega}{P_R}\right)^2}} \dfrac{\hbar}{m_0 E_g} \nabla_{\mathbf{r}} \left(\mathcal{E} \cdot \mathbf{p}_k\right) \end{cases} . \tag{43}$$

Similar results can be obtained with $\mathcal{H}^-$. Due to the term $\sqrt{1 + \left(\frac{P_R}{\Omega}\right)^2}$, the particles move with a slightly larger effective mass. The mass correction depends on the classical position and momentum. This effect could partially compensate the small effective mass values predicted by the $k \cdot p$ theory in semiconductors with a small band gap like InAs or InSb.

3.4 Study of the band transition, an iterative solution Wigner function

The quasi-diagonal Wigner formalism suggests an interesting analogy between band transition induced by a constant electric field (usually denoted as Zener transition (Zener, 1934)) and the scattering processes. In sec. 3.3 the analysis of the equation of motion was restricted to the single band dynamics. In this section, the full many-band dynamics is treated by means of an iterative procedure. For the sake of simplicity, we consider the two-band system in the presence of a uniform electric field. We introduce the new momentum variable $\mathbf{p}' = \mathbf{p} + \mathcal{E}t$ and we apply the Fourier transformation with respect to the $\mathbf{r}$ variable. The Eqs. (38)-(40) become

$$\frac{\partial g_c}{\partial t} = i\boldsymbol{\mu} \cdot \nabla_{\mathbf{p}} \mathcal{H}^+(t) g_c - \xi(t)\ (g_{cv} + g_{vc}) \tag{44}$$

$$\frac{\partial g_v}{\partial t} = i\boldsymbol{\mu} \cdot \nabla_{\mathbf{p}} \mathcal{H}^-(t) g_v + \xi(t)\ (g_{cv} + g_{vc}) \tag{45}$$

$$\frac{\partial g_{cv}}{\partial t} = -\frac{i}{\hbar} 2\sqrt{P_R^2 + \Omega^2(t)}\, g_{cv} + \xi(t)\,(g_c - g_v) \tag{46}$$

$$\frac{\partial g_{vc}}{\partial t} = \frac{i}{\hbar} 2\sqrt{P_R^2 + \Omega^2(t)}\, g_{cv} + \xi(t)\,(g_c - g_v)\ , \tag{47}$$

where, in order to avoid confusion, we defined the new unknowns $g_i = \mathfrak{F}_{\mathbf{r}\to\boldsymbol{\mu}}\left[h_i(\mathbf{r}, \mathbf{p} + \mathcal{E}t, t)\right]$ with $i = c, v, cv, vc$. The time dependence of the coefficients is originated by the definition of

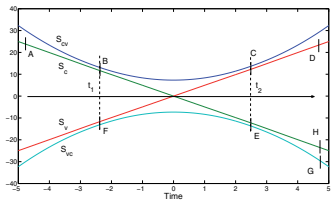


Fig. 2. Eigenvalues of the $\mathfrak{L}$ matrix

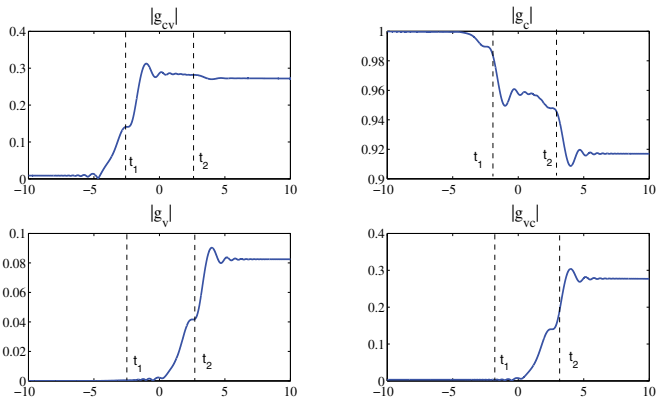


Fig. 3. Modulus of the functions g_i with $i = cv, c, v, vc$.

the $\mathbf{p}'$ variable. Explicitly, $\nabla_{\mathbf{p}}\mathcal{H}^-(t) \equiv \nabla_{\mathbf{p}}\mathcal{H}^-\big|_{\mathbf{p}=\mathbf{p}'-\mathcal{E}t}$, $\xi(t) \equiv \xi(\mathbf{p} = \mathbf{p}' - \mathcal{E}t)$, and similar for the other coefficients.

The system of Eqs. (44)-(47) is a time-dependent eigenvalue problem with perturbation. In fact, if we define the four component vector $\mathbf{G} = (g_c, g_v, g_{cv}, g_{vc})^t$, Eqs. (44)-(47) can be rewritten as $\frac{\partial \mathbf{G}}{\partial t} = i\mathfrak{L}(t)\mathbf{G} + \mathfrak{T}(t)\mathbf{G}$, where $\mathfrak{L}$ is a diagonal time-dependent matrix and $\mathfrak{T}$ is the perturbation. In order to make the subsequent discussion easier, we define the elements of $\mathfrak{L}$ by $\lambda_c = \boldsymbol{\mu} \cdot \nabla_{\mathbf{p}}\mathcal{H}^+(t)$, $\lambda_v = \boldsymbol{\mu} \cdot \nabla_{\mathbf{p}}\mathcal{H}^-(t)$, $\lambda_{cv} = -\frac{2}{\hbar}\sqrt{P_R^2 + \Omega^2(t)}$ and $\lambda_{vc} = -\lambda_{cv}$ (the coefficients of $\mathfrak{T}$ can be obtained by comparison with Eqs. (44)-(47)). Each function g_i can be identified by the component of $\mathbf{G}$ of the unperturbed eigenvector basis (in this case, the simple canonical basis). The eigenvalues of the matrix $\mathfrak{L}$ are shown in fig. 2.

If we assume that $\mathfrak{L}(t)$ and $\mathfrak{T}$ vary slowly in time, according to well known results of adiabatic perturbation theory, eigenspaces belonging to different eigenvalues are decoupled as long as the difference among the eigenvalues is large. In this case, the projections of the solution on the different eigenspaces evolve independently. Only when the eigenvalues become closer, a coupling is possible and a transition from one eigenspace to another can be performed. In our case, a coupling of the eigenspaces is can be observed only around $t \approx t_1$ and $t \approx t_2$ (see fig. 2).

For the sake of concreteness, we consider a tunneling transition from the conduction band to the valence band. This can be described by setting initially all the functions to zero, with the exception of g_c. As it is customary in the time-dependent perturbation theory, we fix the initial time equal to $-\infty$. The value of g_v for $t \to +\infty$ gives the measure of the interband tunneling

induced by $\mathcal{E}$. We write the solution in terms of the Dyson expansion

$$\mathbf{G}^i = e^{i\int_{-\infty}^{t}\mathfrak{L}(\tau)\,\mathrm{d}\tau}\mathbf{G}_0 + \int_{-\infty}^{t} e^{i\int_{t'}^{t}\mathfrak{L}(\tau)\,\mathrm{d}\tau}\mathfrak{T}(t')\mathbf{G}^{i-1}(t')\,\mathrm{d}t' \tag{48}$$

The integral equation (48) can be easily approximated order by order. The second order writes (we discuss only the valence component of $\mathbf{G}$)

$$g_v^1 = \mathcal{J}_{v,cv} + \mathcal{J}_{v,vc}\ , \tag{49}$$

with

$$\mathcal{J}_{v,vc}(t) = e^{\int_{-\infty}^{t}\lambda_v(\tau)\,\mathrm{d}\tau}\int_{-\infty}^{t}\xi(t')\,e^{\int_{-\infty}^{t'}\lambda_{cv}(\tau)-\lambda_v(\tau)\,\mathrm{d}\tau}g_{cv}^1(t')\ \,\mathrm{d}t' \tag{50}$$

$$g_{cv}^1(t') = g_c^0\ e^{\int_{-\infty}^{t'}\lambda_c\,\mathrm{d}\tau}\int_{-\infty}^{t'}\xi(t'')\ e^{-\int_{t''}^{t'}(\lambda_c-\lambda_{cv})\,\mathrm{d}\tau}\,\mathrm{d}t''\ . \tag{51}$$

where g_c^0 is the initial condition of g_c. Similar formula holds for $\mathcal{J}_{v,cv}$. We have that $|\lambda_{cv}(t) - \lambda_c(t)| > E_g$, therefore, in a wide gap semiconductor, the only relevant contribution to the integral is generated in the neighborhood of the minimum of the oscillation frequency (for $t \approx t_1$, see fig. 2). Consequently, at $t = t_1$ the g_{cv} function increases sharply (see fig. 3). The integral in Eq. (50) can be approximated in the same manner. Since the minimum of $|\lambda_{cv}(t) - \lambda_v(t)|$ occurs for $t = t_2 > t_1$, g_{cv}^1 can be considered as constant around t_2 and the integral can be estimated by using the stationary phase approximation.

According to these considerations, the time evolution of the system can be described as follows. For $t < t_1$ the solution, which initially belongs to the $\mathcal{S}_c$ eigenspace (we denote with $\mathcal{S}_i$ the eigenspace spanned by the i-th component of $\mathbf{G}$), evolves adiabatically remaining in $\mathcal{S}_c$. As shown in fig. 3, g_c is the only non-vanishing component of the solution $\mathbf{G}$ until $t = t_1$. At $t = t_1$, a very sudden drop of the value of g_c is observed, and, correspondingly, the g_{cv} distribution function increases. This can be interpreted as the creation of an excited state in the g_{cv} band (visualized with the B point in fig. 2). This excited state "moves" in the S_{cv} band until, at $t = t_2$, it generates an S_v state, which is described by the g_v distribution function. The term $\mathcal{J}_{v,cv}$ of Eq. (50) is thus associated with the path $A - B - C - D$ indicated in fig. 2. The particle is initially in the conduction band (represented by the point A) and in B an excited state is created. It moves towards the point C. There it generates a particle in the valence band which moves adiabatically (point D). The inverse of the difference of the eigenvalues ($\lambda_c - \lambda_{cv}$ in t_1 and $\lambda_{cv} - \lambda_v$ in t_2) quantifies the strength of the coupling (or the probability of a transition). The behavior of the function g_c can be described with similar arguments. The g_c function describes the states that move from A (initial time) to H (final time). This distribution undergoes two scattering events, in B (at $t = t_1$) and in E (at $t = t_2$). We note that, at $t = 0$, no scattering phenomena can be observed, since the eigenspaces S_c and S_v are always decoupled. This represents the analogous of the selection rules for the ordinary scattering phenomena.

This iterative procedure resemble very closely the formalism used for the description of the electron scattering phenomena in semiconductors. In our study of interband transitions, this analogy used for the description of the Zener phenomenon in term of a tunneling process where a particle "disappears" from the band where it was initially located, and it "appears" in a different branch of the band diagram. This behaves similarly to the generation of an electron-hole pair induced by the absorption of a photon. This procedure has been exposed more into details in (Morandi & Demeio, 2008). The field dependent case is treated in (Morandi, 2009).

4. Berry phase and Wigner-Weyl formalism

In a crystal where the effective Hamiltonian is expressed by a partially diagonalized basis (e. g. in graphene or in semiconductors), the major particle operators have off-diagonal elements and the usual definitions of the macroscopic quantities, like for example the mean velocity or the particle density, no longer apply. The theory of Berry phases offers an elegant explanation of this effect in terms of the intrinsic curvature of the perturbed band (Bohm et al., 2008; Xiao et al., 2010). We discuss how it is possible to characterize the Berry phase in a multiband system by using our kinetic description of the quantum dynamics.
The Berry phase theory cannot be directly applied to the particle evolution in a graphene sheet for the obvious reason that the Hamiltonian given in Eq. (25) does not contain any adiabatic variable. Anyway, a Berry-like procedure can be developed if we renounce to treat rigorously the particle dynamics and some approximations are retained. From the physical point of view, one of the most interesting properties of the particle-hole pair in graphene is its pseudo-spinorial character and its connection with the orbital motion. In the momentum representation, the unperturbed graphene Hamiltonian writes $v_F\ \boldsymbol{\sigma}\cdot\mathbf{p}$. If we assume that the particle wave function is represented by a non-spreading wave packet centred around the position $\mathbf{r}$ and the momentum $\mathbf{p}$, we expect due to the Ehrenfest theorem that, in the presence of a gentle potential U (sufficiently smooth), the center of mass of such wave function will describe a trajectory $\mathbf{r}(t),\mathbf{p}(t)$. Sometime this is pictorially visualized by saying that the particle is confined in a small box located at a certain position $\mathbf{r}$ and that the wave packet moves without spreading along a certain trajectory $\mathbf{r}(t)$. If we now *assume* that in such situation the graphene Hamiltonian can be approximated by $v_F\ \boldsymbol{\sigma}\cdot\mathbf{p}(t)$, we can treat the momentum trajectory like an external adiabatic variable. It should be noted that since a non-trivial trajectory is always generated by a potential U, this term should be explicitly included in the graphene Hamiltonian as we did in Eq. (24). Anyway, when included, the Hamiltonian in the momentum space would loose the easy expression $v_F\ \boldsymbol{\sigma}\cdot\mathbf{p}$ (the potential U generates a sum over all the possible momenta). In the following, we will show that the multiband Wigner procedure suggests a natural way to treat the Berry phases of the system for which there is no need to identify in the particle trajectory the "external parameter" of the Hamiltonian, as indicated by the previous artificial procedure. We define by $u_\pm$ the orthonormal eigenvectors of $\mathcal{H}_{g,a}=v_F\ \boldsymbol{\sigma}\cdot\mathbf{p}$. With the Dirac notation

$$\mathcal{H}_{g,a}(\mathbf{p})\left|u_\pm(\mathbf{p})\right\rangle=\pm v_F|\mathbf{p}|\left|u_\pm(\mathbf{p})\right\rangle \tag{52}$$

we write the solution of the Schrödinger problem

$$i\hbar\frac{\partial\left|\psi\right\rangle}{\partial t}=\mathcal{H}_{g,a}\left|\psi\right\rangle \tag{53}$$

as $\left|\psi\right\rangle=c_+(t)\left|u_+(\mathbf{p})\right\rangle+c_-(t)\left|u_-(\mathbf{p})\right\rangle$. A straightforward calculation gives (similar equation hold true for c_-)

$$i\hbar\frac{\partial c_+(t)}{\partial t}=-c_+(t)\left\langle u_+(\mathbf{p})\right|\frac{\partial u_+(\mathbf{p})}{\partial t}\left\rangle-c_-(t)\right\langle u_+(\mathbf{p})\left|\frac{\partial u_-(\mathbf{p})}{\partial t}\right\rangle+c_+(t)v_F|\mathbf{p}|\,. \tag{54}$$

The adiabatic theory ensures that the second term on the right side of the equation becomes arbitrarily small in the limit of sufficiently slow-in-time evolution of the momentum $\mathbf{p}$ (quasi-static or adiabatic hypothesis). An introduction to the adiabatic theory containing a

rigorous proof of this statement presented in a general context, can be found in (Teufel, 2003). If we assume the initial condition $|\psi(t_0)\rangle = |u_+(\mathbf{p})\rangle$, in the adiabatic limit Eq. (54) gives

$$|\psi(t)\rangle = |u_+(\mathbf{p}(t))\rangle \; e^{i\gamma_+(t) - \frac{i}{\hbar}\int_{t_0}^{t} v_F|\mathbf{p}(t')|\,\mathrm{d}t'} \;, \tag{55}$$

where the term γ_+ is denoted as dynamical phase factor. It can be evaluated by the path integral, along the $\mathbf{p}(t)$-trajectory of the Berry connection $\mathbf{A}(\boldsymbol{\xi})$

$$\gamma_+ = \int \mathbf{A}_{++}(\mathbf{p}) \cdot \,\mathrm{d}\mathbf{p} \,. \tag{56}$$

The Berry connection is given by

$$\mathbf{A}_{rs}(\mathbf{p}) = i\left\langle u_r(\mathbf{p})|\nabla_{\mathbf{p}} u_s(\mathbf{p})\right\rangle \quad ; \quad r,s = +,- \,. \tag{57}$$

According to the discussion presented in sec. 3.2, the multiband Wigner-Weyl formalism describes the particle motion by the set of equations (28)-(31)-(32). In order to see the connection with the Berry phase theory, it is useful to explore the classical limit, or $\hbar$-expansion of the Wigner-Weyl system. According to Eq. (19), if the external electric potential $U(\mathbf{r})$ is sufficiently regular, we have

$$\left[\Lambda(\mathbf{p}), \mathcal{S}'\right]_\star + \left[\mathcal{U}', \mathcal{S}'\right]_\star = \left[\mathfrak{A}, \mathcal{S}'\right] - \frac{i\hbar}{2}\left\{\nabla_{\mathbf{p}}\Lambda, \nabla_{\mathbf{r}}\mathcal{S}'\right\} + i\hbar\nabla_{\mathbf{r}}U\cdot\nabla_{\mathbf{p}}\mathcal{S}' + o(\hbar^2) \,, \tag{58}$$

where curly brackets denote the anti-commutators. We focus our attention to the first term of Eq. (58). In particular, $\mathfrak{A}$ groups all the terms that, originated from the $\hbar$-expansion procedure, are simple matrix multiplications acting on $\mathcal{S}'$ (all the other are differential operators):

$$\begin{aligned}\mathfrak{A} &= \Lambda + \frac{i\hbar}{2}\left[\Theta, \nabla_{\mathbf{p}}\Theta\right]\cdot\nabla_{\mathbf{r}}U \\ &= \begin{pmatrix} v_F|\mathbf{p}| & 0 \\ 0 & -v_F|\mathbf{p}| \end{pmatrix} + \frac{i\hbar}{2}\begin{pmatrix} \mathbf{A}_{++} & \mathbf{A}_{+-} \\ \mathbf{A}_{-+} & \mathbf{A}_{--} \end{pmatrix}\cdot\nabla_{\mathbf{r}}U\end{aligned} \tag{59}$$

In Eq. (59) we used that the columns of Θ are the eigenvectors of $\mathcal{H}$ (Eq. (24)) and by applying the definition of Eq. (57), we obtain $\mathbf{A}_{ij}(\mathbf{p}) = \left[\Theta(\mathbf{p})\nabla_{\mathbf{p}}\Theta(\mathbf{p})\right]_{ij} = \sum_k \Theta_{ik}\nabla_{\mathbf{p}}\Theta_{kj}$. Equation (59) emphasizes the role played by the Berry connection in the kinetic description of the particle-hole motion. In our formalism, the Berry connection leads to the first correction (in terms of an $\hbar$ expansion) of the classical of motion. Up to the first order in $\hbar$, the equations of motion (28) become (the components of $\mathcal{S}'$ are defined in Eq. (30))

$$\frac{\partial f^{\pm}}{\partial t} = \pm v_F\frac{\mathbf{p}}{|\mathbf{p}|}\cdot\nabla_{\mathbf{r}}f^{\pm} + \nabla_{\mathbf{r}}U\cdot\nabla_{\mathbf{p}}f^{\pm} \pm i\left(\mathcal{B}f^i - \overline{\mathcal{B}f^i}\right), \tag{60}$$

$$\frac{\partial f^i}{\partial t} = i\mathcal{A}f^i + \nabla_{\mathbf{r}}U\cdot\nabla_{\mathbf{p}}f^i + i\overline{\mathcal{B}}\left(f^+ - f^-\right), \tag{61}$$

where overbar means conjugation and

$$\mathcal{A} = -\frac{2v_F}{\hbar}|\mathbf{p}| + \frac{1}{2}\left(\mathbf{A}_{++} - \mathbf{A}_{--}\right)\cdot\nabla_{\mathbf{r}}U = -\frac{2v_F}{\hbar}|\mathbf{p}| + \frac{1}{|\mathbf{p}|^2}\left(\mathbf{p}\wedge\nabla_{\mathbf{r}}U\right)_z \,, \tag{62}$$

$$\mathcal{B} = \frac{1}{2}\mathbf{A}_{-+}\cdot\nabla_{\mathbf{r}}U = \frac{1}{2}\frac{p_x + ip_y}{|\mathbf{p}|^3}\left(\mathbf{p}\wedge\nabla_{\mathbf{r}}U\right)_z \,. \tag{63}$$

Here, $(\mathbf{p} \wedge \nabla_{\mathbf{r}} U)_z$ denotes the out-of-plane component (z-coordinate) of the vector $\mathbf{p} \wedge \nabla_{\mathbf{r}} U$. We remark that we use a slightly generalized definition of Berry connection. The standard Berry theory limits itself to consider the "in band" evolution of the system. This is a direct consequence of the adiabatic approximation that forbids band transitions. The Wigner-Weyl formalism, being more general, is not limited to any adiabatic hypothesis and band transition are allowed. For that reason, besides the diagonal Berry connections $\mathbf{A}_{++}$ and $\mathbf{A}_{--}$, the terms $\mathbf{A}_{+-}$ and $\mathbf{A}_{-+}$ appear. They are responsible for the particle band transitions (see the discussion of this point in (Morandi & Schürrer, 2011)).

5. Approximated model for the Wigner dynamics

The numerical solution to the equation of motion for the Wigner quasi-distribution function has been the subject of many studies (see i.e. (Frensley, 1990)). Often, a strong the similarity of the shape of the Wigner function with the classical counterpart can be observed. This is especially true in situations where strong quantum interference effects are not expected, but sometime also in the presence of sharp barriers and resonant structures. This consideration is often invoked for justifying the approximation of the θ operator appearing in Eq. (2) with the classical force term (leading term in the $\hbar$-expansion). Although the $\hbar$-expansion appears to be the most natural way to proceed, its application encounter many difficulties when approximations beyond the classical term are concerned. In fact, when applied to realistic problems, this procedure could generate a proliferation of corrective terms. Their number could be quite large and, furthermore, it is usually very difficult (sometime impossible) to ascribe to each term a clear physical meaning. Moreover, the range of validity of such an expansion, when truncated at a certain order, is questionable. The reason is that, at the microscopic level, the particle motion is characterized by complex phase-interference phenomena, which cannot be viewed as a simple refinement of the classical dynamics. Here, we present a slightly different strategy for approximating the Wigner equation of motion. The idea is to replace the θ operator, which is the source of the difference between classical and quantum dynamics, with a more tractable term. The similitude with the classical motion is exploited by approximating the Wigner evolution equation with a Liouville-like equation, where the force operator is the "best classical" approximation of the θ operator in the sense of the $\mathbf{L}^2$ norm. We consider the functional

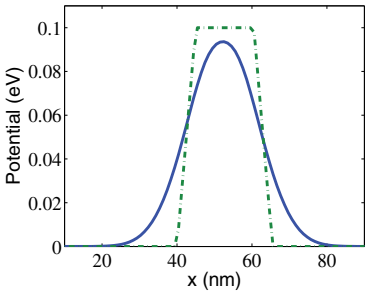
$$\mathcal{N}[f] = \left\| \theta[f] - \mathbf{F}(\mathbf{r}) \frac{\partial f}{\partial \mathbf{p}} \right\|_{\mathbf{L}^2(\mathbb{R}^d_{\mathbf{r}} \times \mathbb{R}^d_{\mathbf{p}})} .$$

Here, the Wigner function is considered as a given function and the pseudo-field $\mathbf{F}$ is the unknown. We choose $\mathbf{F}$ such that the previous functional reaches the minimum value. The function $\mathbf{F}(\mathbf{r}, \mathbf{p})$ thus provides the closest approximation of $\theta[f]$ in the $\mathbf{L}^2(\mathbb{R}^d_{\mathbf{r}} \times \mathbb{R}^d_{\mathbf{p}})$ norm, for each function f sufficiently regular. The minimization of $\mathcal{N}[f]$ is obtained by solving the variational problem

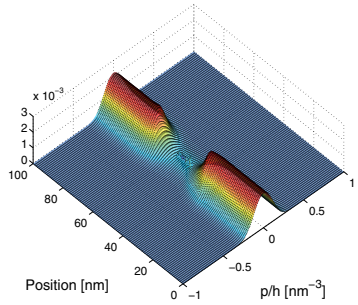
$$\delta_{\mathbf{F}} \mathcal{N}[f] = 0 .$$

Straightforward calculations show that the minimizing function $\mathbf{F}$ is given by

$$F_j(\mathbf{r}) = -i \frac{\int_{\mathbb{R}^d_\eta} \eta_j \mathcal{D}(\mathbf{r}, \boldsymbol{\eta}) \left| \widetilde{f}(\mathbf{r}, \boldsymbol{\eta}) \right|^2 \mathrm{d}\boldsymbol{\eta}}{\int_{\mathbb{R}^d_\eta} \eta_j^2 \left| \widetilde{f}(\mathbf{r}, \boldsymbol{\eta}) \right|^2 \mathrm{d}\boldsymbol{\eta}} . \tag{64}$$



(a) Green dot dashed line: external potential. Blue continuous line: pseudo-potential



(b) Quantum-corrected distribution.

Fig. 4. Comparison between the external potential U and the pseudo-potential $U^* = \int \mathbf{F}\, \mathrm{d}\mathbf{r}$.

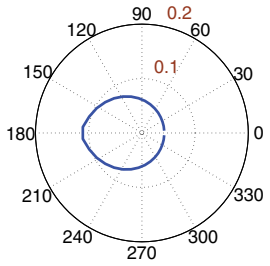
Equation (64) reveals that the calculation of the pseudo-force field in a certain position requires the knowledge of the potential in the overall $\mathbf{r}$ space (via the term $\mathcal{D}$). By computing the integral, the potential U is evaluated at the positions $\mathbf{r} \pm \frac{\hbar}{2}\boldsymbol{\eta}$ and has a measure proportional to $\left|\widetilde{f}\right|^2$, the spectral power of f in the $\boldsymbol{\eta}-$space. As a consequence, the more the $\mathbf{p}$-gradient of the solution $f(\mathbf{r}, \mathbf{p})$ increases, the more the force $\mathbf{F}$ becomes non-local and the values of the potential faraway from $\mathbf{r}$ are important. This can be expressed pictorially by saying that, as compared with a smoother distribution function, an irregular profile of the solution "sees" a larger spatial region. The approximated quantum-Wigner evolution equation becomes

$$\frac{\partial f}{\partial t} = -\frac{\mathbf{p}}{m} \cdot \nabla_{\mathbf{r}} f - \mathbf{F}(\mathbf{r}) \cdot \nabla_{\mathbf{p}} f \,. \tag{65}$$

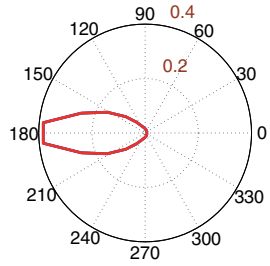
This is a nonlinear system where the pseudo-electric field $\mathbf{F}$ depends on the solution itself. In some situations, the nonlinearity can be eliminated and a good approximation of the field $\mathbf{F}$ can be obtained by replacing in Eq. (64) the solution f with the classical Boltzmann equilibrium distribution at the temperature T

$$f^{eq} = \sqrt{\frac{mkT}{\pi}} e^{-\frac{1}{kT}\left(\frac{|\mathbf{p}|^2}{2m} - \mu\right)} \,,$$

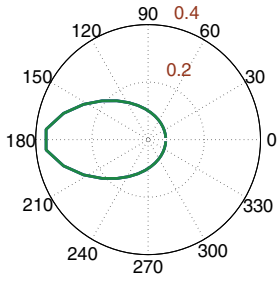
where μ is the chemical potential of the particle gas and k the Boltzmann constant. In fig. 4 the comparison of the classical and the pseudo electric field obtained by using the Boltzmann distribution function is presented. A glance at the figure reveals that, compared with the bare potential U, the effective pseudo-potential is smoother and extends beyond the support of U. As a consequence, the particle in the presence of the quantum corrected potential are decelerated or accelerated before they reach the classical force field $-\nabla_{\mathbf{r}} U$, making evident the non-local action of the quantum potential. Furthermore, the snapshot fig. 4-(a) shows that the maximum value of the effective potential is smaller than the classical one. As a consequence, particles with energy smaller than the maximum of the potential (but greater than the maximum of the pseudo-potential) are not reflected by the barrier. This simple example illustrates how quantum tunneling can be approximatively described by a classical formalism. Furthermore, in fig. 4-(b) we depict the solution of Eq. (65) in the presence of the potential U. At the boundary, the Boltzmann distribution is imposed.



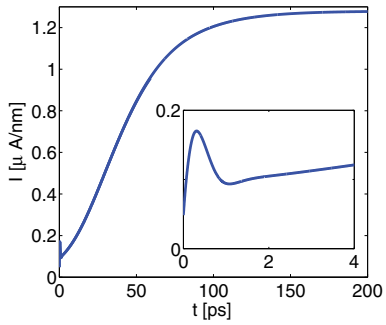
(a) Initial condition.



(b) $t = 0.3$ ps.



(c) $t = 10$ ps.



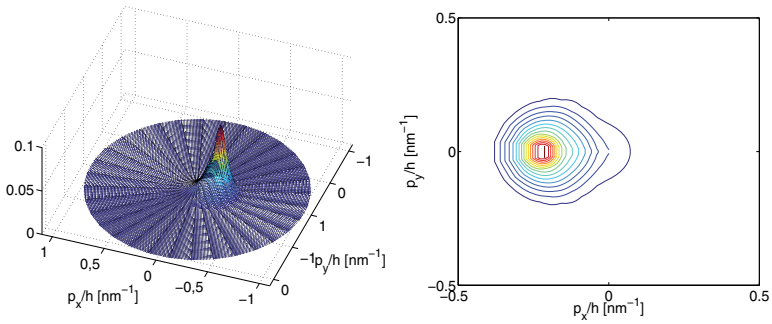
(d) Time evolution of the total current.

Fig. 5. (a)-(c) Polar plot of the density for current in graphene. (d) Total current.

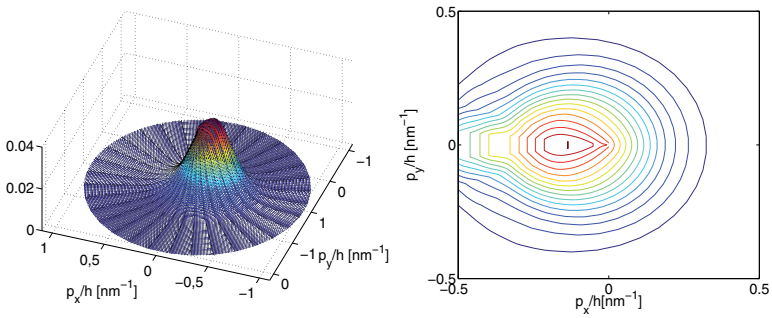
6. Dissipative effects in the Wigner formalism: electron-phonon collisions in graphene

As described in the introduction, one of the major advantages of the Wigner formalism is the possibility to include in a quantum mechanical treatment also some dissipative effects, or (in the opposite limit) to derive some quantum corrected models for the simulation of quasi-classical systems. As an example, we apply the results obtained in sec. 5 for studying the particles evolution in graphene and we include a detailed description of the electron-phonon scattering phenomena, via a Boltzmann scattering collision operator. An important property of the pseudo-electric field approximation is the preservation of the positivity of the quantum-corrected distribution function. Since the Boltzmann collision operator is defined only for positive functions, positivity preservation becomes a fundamental property for any Boltzmann quantum-corrected kinetic model (anyway, despite the lack of theoretical support, some Wigner-Boltzmann solver have been numerically tested (Kosina & Nedjalkov, 2006)). A semiclassical Boltzmann model with quantum corrections, allows the study of the relaxation processes dynamically, providing information on the time scale on which the equilibrium is established.

The phonon system of graphene has already been thoroughly investigated by means of density functional theory (DFT) and Raman spectroscopy (Piscanec et al., 2004). The phonon dispersion relations and electron-phonon coupling matrix elements are essential ingredients for kinetic models of carrier transport in graphene. Results of DFT calculations show that longitudinal optical (LO) and transversal optical (TO) phonons modes contribute significantly

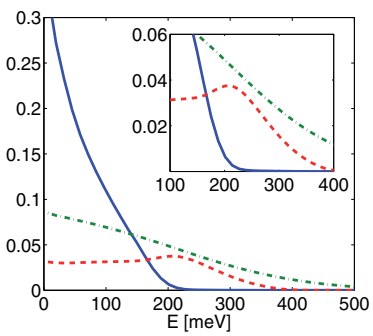


(a) $t = 0.3$ ps.

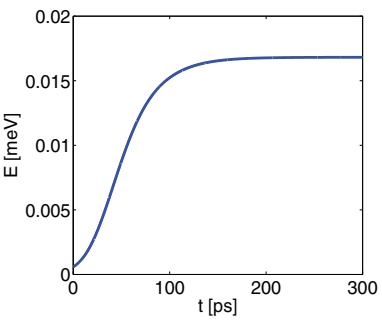


(b) $t = 1$ ps.

Fig. 6. 3D and contour plot representation of the f^+ distribution for different times.



(a) Evolution of the energy density for $t = 0$ (continuous blue line), $t = 0.3$ ps (dashed red line) $t = 10$ ps (dot-dashed green line).



(b) Total energy of the system.

Fig. 7. Evolution of the energy density and the total energy of the particle gas.

to inelastic scattering of electrons in graphene. Because of their short wave vectors these phonons scatter electrons within one valley. In addition, zone boundary phonons close to the **K**-point are responsible for intervalley processes. The Boltzmann equation of motion

including optical phonon scattering writes

$$\frac{\partial f^{\pm}}{\partial t} \mp v_F \frac{\mathbf{p}}{|\mathbf{p}|} \cdot \nabla_{\mathbf{r}} f^{\pm} - \mathbf{F} \cdot \nabla_{\mathbf{p}} f^{\pm} = \sum_{\eta,j} \mathcal{C}^{\eta}_{\pm j} \,. \tag{66}$$

The collision kernel, containing emission and absorbtion processes, is given by

$$\mathcal{C}^{\eta}_{ij}(\mathbf{p}) = \frac{1}{(2\pi)^2} \int_{\mathcal{B}} \left\{ W^{\eta}_{\mathbf{p}'_j \mathbf{p}_i} f^j(\mathbf{p}')[1 - f^i(\mathbf{p})] - W^{\eta}_{\mathbf{p}_i \mathbf{p}'_j}[1 - f^j(\mathbf{p}')] f^i(\mathbf{p}) \right\} \, \mathrm{d}\mathbf{p}' \,, \quad i,j = \pm \tag{67}$$

where η labels the specific scattering processes, $\mathcal{B}$ denotes the first Brillouin zone of graphene and

$$W^{\eta}_{\mathbf{p}_i \mathbf{p}'_j} = s^{\eta}_{\mathbf{p}'_j \mathbf{p}_i} \, \left[1 + g_\eta(\boldsymbol{p} - \boldsymbol{p}')\right] \, \delta(\varepsilon_i - \varepsilon'_j - \hbar\omega_\eta) + s^{\eta}_{\mathbf{p}_i \mathbf{p}'_j} \, g_\eta(\boldsymbol{p}' - \boldsymbol{p}) \, \delta(\varepsilon_i - \varepsilon'_j + \hbar\omega_\eta) \,. \tag{68}$$

The delta functions (where we adopt the simplified notation $\varepsilon_i = \varepsilon_i(\boldsymbol{p})$, $\varepsilon'_j = \varepsilon_j(\boldsymbol{p}')$ and ω_η denotes the energy of the η-th mode) ensure the conservation of the energy during the scattering processes. The explicit expression of the scattering elements $s^{\eta}_{\mathbf{p}'_j \mathbf{p}_i}$, can be found in (Lichtenberger et al., 2011; Piscanec et al., 2004). According to sec. 3.2, the functions f^+ and f^- represent the particle distribution in the upper and in the lower Dirac cone, respectively. Finally, the g_η are the phonon equilibrium distribution functions related to the η-th mode. Here, for sake of simplicity, we assume that the phonon system is an infinite reservoir at a constant temperature T. In this hypothesis, the g_η can be approximated by the Bose-Einstein distributions $g^0_\eta = [\exp(\hbar\omega_\eta / k_B T) - 1]^{-1}$. The study of the coupled electron-phonon system is presented in (Lichtenberger et al., 2011). It has been shown that the optical phonons are in equilibrium only for a low-bias polarization (around 0.1 eV), otherwise hot phonon effects should be included.

We apply the Boltzmann system given in Eq. (66) to study the transient evolution of the electron-hole and phonon gas in response on the abrupt change of the applied bias. As initial datum, for $t = 0$, we assume that the graphene sheet is in the stationary state for an applied voltage U equal to 0.01 V. For $t > 0$ we impose $U = 0.1$ V. In Fig. 5-d we show the evolution of the total current at the drain contact for the intrinsic graphene. The simulations reveal the presence of a current overshoot (approximatively one picosecond after the potential change) and a subsequent approach to the equilibrium value. The further approach to the equilibrium is a quite slower process of approximatively 200 picoseconds.

The detailed explanation of the transient current overshoot observed during the first picosecond requires a deeper analysis of the high non-equilibrium motion of the hot carriers. The presence of an overshoot in the current evolution is an unexpected phenomenon in graphene. It is well known that, in this material, the carrier velocity is independent from the modulus of the momentum. For this reason, we expect that even if some transient phenomena are able to move the hot carriers toward high values of the momentum, this should not significantly affect their velocity and consequently the total current of the system. The overshoot can be explained by analyzing the following two-step process: initially the particles are ballistically accelerated by the strong external field (the temperature of the particles gas stays essentially constant). However, after some picoseconds, the scattering processes are able to transform the kinetic energy of the carriers into thermal energy. During the first picosecond, the component of the momentum parallel to the external field increases. As a consequence, the direction of the momentum (and thus the velocity) is turned toward the direction of the

electric field. In this first part of the dynamics, the motion is essentially ballistic, the particles share similar momentum direction and group together in the velocity space. This behaviour is evident from fig. 5 where we depict the polar plot of the angular density of the current. For $t < 0.3$ ps the drift term dominates the Boltzmann collision operator. The latter is a nonlinear operator and its effects on the distribution function depend on the shape of the function itself. On the contrary, the ballistic operator translates the distribution function over the phase plane along the Hamiltonian flux and is independent of the distribution. During the overshoot of the current, the Boltzmann operator is not able to balance the effect of the ballistic term. This can be seen in fig. 6 where we depict the evolution of the electron distribution function f^+ for different times. The first part of the dynamics (fig. 6(a)) is just a rigid translation of f^+ towards higher values of the momentum variable. After one picosecond, an enlargement of the distribution function around its center of mass can be observed. This is a clear signature of the temperature increase of the system. The friction process occurs only by dissipating the kinetic energy of the particles by phonon emission. This requires a certain time delay. A closer look at the density of energy of the carriers explains the reason why the particle gas need a delay before starting to emit phonons. In fig. 7 we plot the evolution of the energy density and the total energy of the particles. We see that a peak of high energy particles is present after 0.3 ps. This peak represents the particles accelerated by the field. Their kinetic energy increases until they are able to emit optical phonons (whose energy is 196 and 161 eV for Γ and $\mathbf{K}$ phonons respectively). Around an energy of 200 meV, the kinetic energy can be efficiently dissipated and the distribution reaches a new thermal-like state characterized by a smaller total current.

6.1 Conclusions

In this Chapter, various approaches based on the Wigner-Weyl formalism, are presented. In particular, we highlight the existence of a general formalism where in analogy with the Schrödinger formalism, we use the class of unitary operators in order to define a class of equivalent quasi-distribution functions. The applications of this formalism span among different subjects: the multi-band transport in nano-devices, the infinite-order $\hbar$-approximations of the motion and the characterization of a system in terms of Berry phases or, more generally, the representation of a quantum system by means of a Riemann manifold with a suitable connection. The exposition of the theory is completed with some numerical test and applications to real devices.

7. Acknowledgment

The research was founded by the Austrian Science found (FWF): P21326-N16.

8. References

Alvaro, M. & Bonilla, L. L., (2010). Two miniband model for self-sustained oscillations of the current through resonant-tunneling semiconductor superlattices, *Phys. Rev. B* Vol. 82, 035305.

Arnold, A, (2008). Mathematical properties of quantum evolution equations. In: G. Allaire, A. Arnold, P. Degond, and T. Hou (eds.), *Quantum Transport - Modelling, Analysis and Asymptotics, Lecture Notes Math.* 1946, 45-110. Springer, Berlin.

Beenakker, C. W. J., Akhmerov, A. R., Recher, P. & Tworzydło, J., (2008). Correspondence between Andreev reflection and Klein tunneling in bipolar graphene, *Phys. Rev. B* Vol. 77, 075409.

Bohm, A., Mostafazadeh, A., Koizumi, H., Niu, Q. & Zwanziger, J., (2008). *The Geometric Phase in Quantum Systems: Foundations, Mathematical Concepts, and Applications in Molecular and Condensed Matter Physics*, Sringer-Verlag (Berlin, Heideberg).

Cohen, L., (1966). Generalized phase-space distribution functions, *Journal of Mathematical Physics* Vol. 7, 781.

Demeio, L., Bordone, P. & Jacoboni, C., (2006). Multi-band, non-parabolic Wigner-function approach to electron transport in semiconductors, *Transp. Theory Stat. Phys.*, Vol. 34, 1.

Folland, G. B., (1989). *Harmonic Analysis in Phase Space*, Princeton University Press, Princeton.

Frensley, W. R. (1990). Boundary conditions for open quantum driven far from equilibrium, *Rev. Mod. Phys.* Vol. 62 (No. 3), 745.

Giulini, W., Joos, E., Kiefer, C., Kupsch, J., Stamatescu, I.-O., Zeh, H. D., (2003). *Decoherence and the appearance of a classical world in quantum theory* Springer-Verlag Berlin.

Jüngel, A., (2009). *Transport Equations for Semiconductors*, Springer-Verlag, Berlin.

Kosina H. & Nedjalkov, M., (2006). Wigner function-based device modeling. In: M. Rieth and W. Schommers (eds), Handbook of Theoretical and Computational Nanotechnology 10, 731.

Lee, H.-W. (1995). Theory and application of the quantum phase-space distribution functions, *Physics Reports*, Vol. 259, 147.

Lions, P.-L. & Paul, T. (1993). Sur les mesures de Wigner. *Rev. Mat. Iberoamer*, Vol. 9, 553.

Lichtenberger, P., Morandi, O. & Schuerrer, F., (2011). High field transport and optical phonon scattering in graphene, *Phys. Rew. B*, Vol. 84, 045406.

Markowich, P., Ringhofer, C. & Schmeiser, C. (1990). *Semiconductor Equations.* Wien: Springer Verlag.

Markowich, P. & Ringhofer, C. (1989). An analysis of the quantum Liouville equation. *Z. Angew. Math. Mech.*, Vol. 69, 121.

Morandi, O., Modugno, M. (2005). A multiband envelope function model for quantum transport in a tunneling diode. *Phys. Rev. B*, Vol. 71, 235331.

Morandi, O. & Demeio, L., (2008). A Wigner-function approach to interband transitions based on the multiband-envelope-function model, *Transp. Theory Stat. Phys.*, Vol. 37, 437.

Morandi, O. (2009). Multiband Wigner-function formalism applied to the Zener band transition in a semiconductor, *Phys. Rev. B* Vol. 80, 024301.

Morandi, O., (2010). A WKB approach to the quantum multiband electron dynamics in the kinetic formalism, *Communications in Applied and Industrial Mathematics*, Vol. 1, 474.

Morandi, O., (2010). Effective classical Liouville-like evolution equation for the quantum phase space dynamics, *J. Phys. A: Math. Theor.* Vol. 43, 365302.

Morandi, O. & Schuerrer, F. (2011). Wigner model for quantum transport in graphene, *J. Phys. A: Math. Theor*, Vol. 44, 265301.

Morandi, O. & Schuerrer, Wigner model for Klein tunneling in graphene, *Communications in Applied and Industrial Mathematics*, in publication.

Morandi, O., Barletti, L. & Frosali, G. Perturbation Theory in terms of a generalized phase Quantization procedure, *Boll. Unione Mat. Ital.*, Vol. 4 (No. 1), 1.

Castro Neto, A. H., Guinea, F.,Peres, N. M. R., Novoselov K. S., & Geim, A. K. (2009). The electronic properties of graphene, *Rev. Mod. Phys.*, Vol. 81, 109.

Piscanec, S., Lazzeri, M., Mauri, F., Ferrari, A. C. and Robertson, J., (2004). Kohn Anomalies and Electron-Phonon Interactions in Graphite, *Phys. Rev. Lett.*, Vol. 93, 185503.

Teufel, S., (2003). *Adiabatic Perturbation Theory in Quantum Dynamics*. Springer-Verlag Berlin.

Unlu, M. B., Rosen B., Cui, H.-L. & Zhao, P., (2004). Multi-band Wigner function formulation of quantum transport, *Phys. Lett. A* Vol. 327, 230.

Wigner, E. (1932). On the quantum correction for thermodynamic equilibrium, *Phys. Rev.*, Vol 40, 749.

Xiao, D., Chang, M.-C., & Niu, Q., Berry phase effects on electronic properties, *Rev. Mod. Phys.*, Vol. 82, 1959.

Zachos, C. K., Fairlie, D. B., & Curtright T. L. (editors), (2005). *Quantum mechanics in phase space. An overview with selected papers*. World Scientific Publishing: Hackensack (NJ).

Zener, C. (1934). A Theory of the Electrical Breakdown of Solid Dielectrics, *Proc. R. Soc. London, Ser. A* Vol. 145, 523.

2

Reaction Path Optimization and Sampling Methods and Their Applications for Rare Events

Peng Tao*, Joseph D. Larkin and Bernard R. Brooks
Laboratory of Computational Biology,
National Heart, Lung and Blood Institute,
National Institutes of Health, Bethesda, Maryland
USA

1. Introduction

Reaction mechanisms are an important tool for chemists in the determination of thermodynamic and kinetic properties of chemical reactions.(Hänggi & Borkovec 1990; Heidrich 1995; March 1992; Tolman 1925) The mechanisms are integral in the understanding of detailed molecular or chemical transitions from one equilibrium state (reactant) to another equilibrium state (product). In computational chemistry, the reaction mechanism is often represented as a reaction path on the Born-Oppenheimer potential energy surface (PES) of the system of interest through construction of a potential energy function of the nuclear coordinates.(Bader & Gangi 1975; Lewars 2011; Mezey 1987; Truhlar 2001; Wales 2003) The PES serves as an important theoretical construct to provide a framework to describe the transition between different states in detail. The equilibrium states correspond to local minimum on the PES with zero first order derivatives (gradient) in all directions and all positive eigenvalues of the second order derivative (Hessian) matrix, excluding rotation and translation degrees of freedom. The transition states (TSs), based on the transition state theory (TST),(Doll 2005; Eyring 1935; Laidler & King 1983; Pechukas 1981; Truhlar et al. 1983; Wigner 1938; Yamamoto 1960) are the first order saddle points with zero gradient and only one negative eigenvalue of the Hessian matrix. The equilibrium states are often easy to identify through experimental or computational studies. Understanding the detailed transition process between equilibrium states is of more interest in research, but unfortunately is very difficult to study experimentally. On a given PES, one can imagine that there could exist an infinite number of possible routes connecting two predefined states on that surface. However, not every route has the same weight in elucidating of reaction mechanisms. In the static point of view, the minimum energy path (MEP) is the route that needs the least amount of potential energy for the system to undergo the transition. The MEP connecting two local minima must go through one or more TSs, and is identified as a representative reaction path. In the statistical point of view, the minimum free energy path (MFEP) is the most probable transition path connects two metastable states. The simulation of the systems either through molecular dynamics (MD) or Monte Carlo (MC) sampling on the PES could generate an ensemble of transition paths, from which the MFEP can be identified. Both an MEP and an MFEP can be used to predict important properties, such as a reaction's kinetic isotope effect.

Although one could generate a reduced PES for complex systems with selected reaction coordinates,(Klähn et al. 2005; Shi et al. 2008; Tao et al. 2009a) it is rarely practical to construct a reduced PES of a system of interest to identify a reaction path connecting two minima. Moreover, reducing a complex multi-dimensional system to a simplified pathway is a form of data reduction. This reduction is non-unique and a choice imposed in this reduction will affect the quality and applicability of the results. It is more feasible to search for the MEP or MFEP directly on a given PES. Given the complexity and high degree of freedom of most systems of interest in chemistry, molecular biology and materials science, there has been a rapid development in methodologies for reaction pathway identification in large systems. As an attempt to reflect the current development and to better understand the consequences of specific methodological choices, this chapter reviews the recent progress of various methods to identify reaction paths with or without knowing the TS(s) *a priori*. Specifically, it emphasizes the applications of these methods in macromolecular systems.
It should be noted that identifying saddle points on a PES alone is not the focus of this report. This report does not cover the geometry optimization including equilibrium and transition structures,(Farkas & Schlegel 2003; Henkelman et al. 2000a; Olsen et al. 2004; Schlegel 1982, 2003, 2011) conformational sampling,(Beusen 1996; Leach 1991; Parish 2002) or global minimum search methodologies,(Floudas & Pardalos 1996; Horst 1995, 2000; Torn 1989) which are all important for the studies of computational chemistry and biology. Other related topics, including enhanced sampling methods,(Earl & Deem 2005; Hamelberg et al. 2004; Lei & Duan 2007; Okur et al. 2006; Sugita 1999; Swendsen & Wang 1986; Thomas et al. 2005; Wen et al. 2004) simulation of nonequilibrium states,(Bair et al. 2002; Cummings & Evans 1992; Hoover 1983; Hoover & Hoover 2005; Kjelstrup & Hafskjold 1996; Li et al. 2008; Mundy et al. 2000) and minimization methods,(Bonnans 2003; Dennis & Schnabel 1996; Fletcher 2000; Gill 1982; Haslinger & Mäkinen 2003; Nocedal 2006; Scales 1985) are not covered in this chapter either. The curious readers are welcome to read cited references for more information.

2. PES walking methods

Without the intention to generate a complete PES, it is logical to develop methods to explore the PES by walking along the surface from certain starting points using local information of the PES, such as the energy, gradient and even the Hessian.(Hratchian & Schlegel 2005a; Schlegel 2003, 2011) In this way, one can start from somewhere on the PES, either reactant, product, or TS, and walk uphill or downhill, depending on the starting points to reach the adjacent stationary points. The walking trajectories, after successfully reaching these stationary points, are the reaction pathways that describe the mechanism of transitions. For smaller systems, walking methods may be sufficient to fully understand a given reaction mechanism. However, for larger and more complex systems, pathways explored by such walking mechanisms are often not reversible, and can show significant hysteresis that results in a poor representation of the reaction.

2.1 Reaction path following methods

When using mass-weighted Cartesian coordinates, a steepest descent path from the TS down to the reactant and product is referred to as the intrinsic reaction coordinate (IRC) path.(Fukui 1981; Quapp & Heidrich 1984; Tachibana & Fukui 1980; Yamashita et al. 1981) The steepest descent pathway is given by the differential equation

$$\frac{d\mathbf{x}(s)}{ds} = -\frac{\mathbf{g}(\mathbf{x})}{|\mathbf{g}(\mathbf{x})|} \tag{1}$$

where **x** is the vector of Cartesian coordinates, s is the step size of the path, and **g** is the energy gradient at **x**. The path obtained by solving this equation is the IRC, when **x** is mass-weighted. Numerous methods were developed to locate the TS on a PES.(Baker 1986; Banerjee et al. 1985; Bell & Crighton 1984; Cerjan 1981; Ionova & Carter 1993, 1995; Jensen 1983; Müller & Brown 1979; Peng et al. 1996; Simons & Nichols 1990) The IRC could be optimized based on its variational nature.(Bofill & Quapp 2011; Quapp 2008) However, it is more applicable for many systems to construct the IRC by solving Eq. (1) from a TS.

Gonzalez and Schlegel developed the implicit trapezoid method (GS−IRC) for reaction path following at second order accuracy.(Gonzalez & Schlegel 1989, 1990) In their initial development, the points along the target reaction path are constructed by constrained optimization using internal degrees of freedom of the molecules. For each step of optimization along the path, the new point is constructed and optimized so that the gradient at each point is tangent to the path. Therefore, the resulting path is both continuous and differentiable. This initial method is correct to second order in the limit of small step size. The same method was later developed up to sixth order accuracy.(Gonzalez & Schlegel 1991) The GS−IRC method is generally efficient for small systems.

To improve the computational efficiency, the velocity Verlet algorithm (Verlet 1967) to propagate a classical dynamics trajectory was applied to integrate the IRC with a magnitude of the velocity damping for each step.(Hratchian & Schlegel 2002) This method is referred as the damped velocity Verlet (DVV) algorithm. The time step for each integration step is adjusted to ensure that the damped trajectory stays close to the IRC. The DVV-IRC method can be considered as running downhill along the PES from TS in a slow motion (by damping the velocity at each step). It enjoys the stability of the Verlet integrator and low cost of computation since the Hessian does not need to be calculated.

In their later work, Hratchian and Schlegel introduced an approach using a Hessian based predictor-corrector (HPC) integrator to solve Eq. (1).(Hratchian & Schlegel 2004) The HPC integrator comprises two steps: the predictor step and the corrector step. The gradient **g** and Hessian **H** of the system PES are used to calculate the predictor step with second order accuracy. Then, the correction of the predicted step is calculated through a modified Bulirsch-Stoer algorithm based on the gradient information at the predicted step.(Bulirsch & Stoer 1964, 1966a, b) Although, the HPC-IRC method is comparable to GS-IRC with fourth order accuracy, calculation of the Hessian at each step can be rather expensive for large systems. This bottleneck was resolved by applying a Hessian updating scheme in their later development.(Hratchian & Schlegel 2005b) For each step of an IRC calculation, the Hessian is not calculated *de novo*, but updated from the Hessian of the previous step and the change of the gradient and step size between two steps. With this scheme, the Hessian only needs to be calculated once at the TS, and then is updated at each step of the IRC calculation. This HPC-IRC method with Hessian updating has been applied successfully in large protein systems using a combined quantum mechanical and molecular mechanical (QM/MM) method.(Tao et al. 2010; Tao et al. 2009b; Zhou et al. 2010) In these studies, the inhibition mechanism of matrix metalloproteinase 2 (MMP2) by its potent inhibitor, was elucidated in great detail using QM/MM methods. The TS of the key reaction in the active site of MMP2 was identified. The IRC of the reaction including the protein environment was calculated to confirm that the reactant and product are connected through the identified TS (Fig. 1).

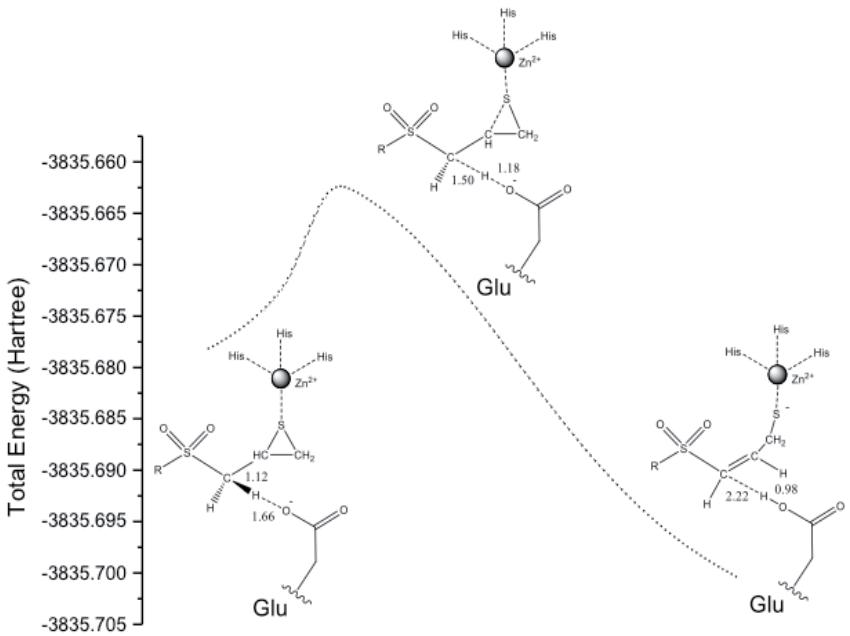


Fig. 1. IRC profile for SB-3CT in the MMP2 active site at the ONIOM(B3LYP/6-31G(d):AMBER) level of theory. Key bond lengths are in angstroms. (Reprinted with permission from ref. (Zhou et al. 2010). Copyright 2010 American Chemical Society.)

Very recently, Hratchian and Schlegel applied the Euler (first-order) predictor and corrector method (EulerPC) using an Euler explicit integrator in the calculation of predicted step. (Hratchian et al. 2010) This method avoids the expensive Hessian calculation at the TS and updating afterwards. By repeating the evaluation and correction several steps after the prediction, the error of the calculation is greatly reduced. The newly developed EulerPC method shows comparable accuracy with HPC but with much less computational cost, and is tested on several rather large enzymatic systems.(Hratchian & Frisch 2011)

As a summary, the IRC calculations are becoming practical even for large enzyme systems using the QM/MM approach. However, to apply any of the IRC methods listed in this section, a well-defined TS structure is necessary to serve as the starting point. For a large system, e.g. an enzymatic reaction system, using QM/MM methods may take substantial effort in identifying a TS.

2.2 Uphill walking methods

By walking uphill from a minimum, one could reach an adjacent TS. Applying a reaction path following method on the obtained TS could yield another minimum corresponding to a product or intermediate state and a complete reaction pathway could be formed. Simons and coworkers developed methods that walk on the PES toward the selected direction (either uphill or downhill) using local gradient and Hessian information.(Nichols et al. 1990; Simons & Nichols 1990; Taylor & Simons 1985) By applying a local quadratic approximation, the PES close to a starting structure $\mathbf{x}_0$ can be written as

$$E(\mathbf{x}) = E_0 + \mathbf{x}\mathbf{F}_0 + \tfrac{1}{2}\mathbf{x}\mathbf{H}_0\mathbf{x} \,. \tag{2}$$

where E_0, $\mathbf{F_0}$ and $\mathbf{H_0}$ are the energy, gradient and Hessian at $\mathbf{x_0}$, respectively. Vector $\mathbf{x}$ is in the region around $\mathbf{x_0}$, in which the local quadratic approximation is valid. In the attempt to walk uphill, the vector with the lowest Hessian eigenvalue will be chosen to define the moving direction. The potential energy increases along the chosen direction, but remains at minima along the other eigenvectors. For the best computational efficiency, the step size for each round of searching is controlled using eigenvalues of the Hessian matrix. After leaving the region in which the local quadratic approximation is valid, the Hessian matrix $\mathbf{H}$ can be recomputed or updated for further calculations. Along the walk, the Hessian eigenvalue of the eigenvector, which is being followed may cross with other eigenvalues. This is likely to occur when the starting geometry is not within the quadratic approximation region of the TS. In such situations, a decision needs to be made to either keep track of the original eigenvector, or follow the eigenvector with the lowest eigenvalue after crossing. This decision may significantly affect the final TS.

An algorithm was developed by Ohno and Maeda to find reaction pathways on a PES systemically.(Maeda 2003; Ohno 2004; Ohno & Maeda 2006; Yang et al. 2005) In their method, the PES around the equilibrium structure (ES) is expanded using reduced normal coordinates in terms of normal coordinates $\mathbf{Q}_i$ with eigenvalues λ_i,

$$\mathbf{q}_i = \lambda_i^{\frac{1}{2}} \mathbf{Q}_i \,. \tag{3}$$

The degrees of freedom for translation and rotation are projected out from the normal coordinates. In such representation, any constant energy around the ES within the limit of harmonic potential gives spherical hypersurface (hypersphere) (Fig. 2). Any structure represented on a hypersphere is somewhat distorted from the ES around which the hypersphere is constructed. This common feature is referred to as anharmonic downward distortions following (ADDF), and is used for a reaction path search by walking uphill toward the direction with the local maxima of ADDF. Through a series of hyperspheres with different sizes but common origin, different reaction paths may be identified by following the local maxima of ADDF on each hypersphere. TS regions or dissociation channels (DC) can be identified through the variation of the first order derivatives along these reaction paths. Further calculations need to be carried out to precisely determine the real TS structures, which may not be on any hypersphere. The reaction path following methods can be applied to the TSs to find new ESs. The whole procedure can be repeated for new TSs. This is referred as scaled hypersphere search (SHS) method. Theoretically, the SHS method can be repeated until all the stationary points and reaction channels are identified for a given system. This method provides a means to systematically explore the PES and is referred to as global reaction route mapping (GRRM). By applying GRRM, the PESs of several small organic molecules were explored with numerous ESs and TSs identified for each system.(Maeda & Ohno 2007; Yang et al. 2005)

Recently, a new GRRM method was developed to search reaction pathways for large flexible systems using a microiteration-ADDF (μ-ADDF) technique.(Maeda et al. 2009) The microiteration scheme was originally developed for the QM/MM method.(Svensson et al. 1996; Vreven et al. 2006b; Vreven et al. 2006a; Vreven et al. 2003) For large systems, it may not be practical to follow all the ADDF maxima. Instead, two other methods were developed by the same authors to follow only large ADDF pathways (lADDF),(Maeda & Ohno 2007) or to follow the reverse direction from a point on a very large hypershpere to the sphere center with decreasing hypershere radius (double-ended ADDF, or dADDF).(Maeda et al. 2009)

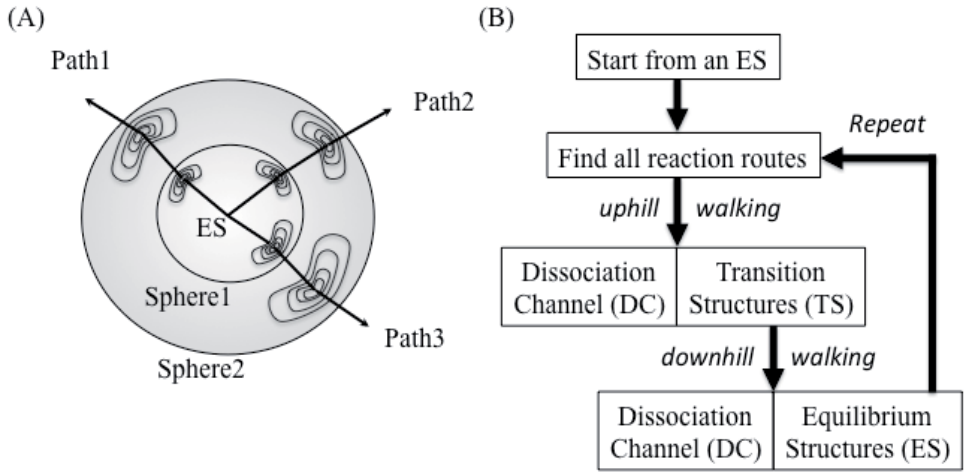


Fig. 2. Schematic illustrations of computational procedures for GRRM by the SHS method: (A) Though the reference harmonic potential has a constant energy on the scaled hypersphere surface, the real potential has some minima on the same surface, which correspond to the anharmonic downward distortions indicating the symptoms of chemical reactions. Following those minima on different sizes of scaled hypersphere (spheres 1 and 2), reaction routes (paths 1-3) can be traced from the equilibrium structure (ES) by the SHS method as shown by arrows. (B) Starting from an ES, find all reaction routes as energy minima on the scaled hypersphere (maxima of anharmonic downward distortions), then continue uphill walking to reach DC or TS, and then downhill walking from the TS region to DC or another ES. From each new ES, the above procedures for finding DC, TS, and ES should be repeated until no new ES is found. This one-after-another approach in the SHS method can be automated, and it enables us to perform GRRM within finite processes. (Reprinted with permission from ref. (Ohno & Maeda 2006). Copyright 2006 American Chemical Society.)

Both methods were implemented with the μ-ADDF technique. In this new GRRM method, a large system is divided into reaction-center and nonreaction-center parts. All the path following calculations, including minimizations on scaled hyperspheres, ES and TS optimization, and IRC following, are carried out in macroiteration steps. The positions of nonreaction-center atoms are optimized during microiteration steps after macroiteration. All movements of reaction-center atoms are treated by the GRRM method as in the case without microiterations with either exact or updated Hessian. $(H_2CO)(H_2O)_{100}$ and $(Si_6)(C_{21}H_{17})_6$ were examples used to test the stability and efficiency of the new method. Multiple reaction pathways were identified in both cases. Thus, the GRRM method with μ-ADDF could serve as a powerful tool to explore PESs of reactions in large molecular systems. However, the GRRM has not been reported to be applied on protein systems.

2.3 Combined method for determining reaction path, minima, and TSs

It is worth noting that Schlegel *et al.* developed a method to determine TSs, minima and the reaction path in a single procedure without calculating the Hessian matrix.(Ayala & Schlegel 1997) In this method, a starting approximate path is constructed as several (5 to 7) structures on PES, and iteratively relaxed until two endpoints reach minima, and one of the middle points reaches a TS. The final path is a second order approximation of the steepest descent

path. However, this method became impractical for large biological systems, such as proteins, because of the computational cost of eigenvector following searches for TSs. Nevertheless, this method provides a rigorous means of identifying reaction paths and TSs simultaneously.

2.4 Reduced Hessian methods

Hessian calculations are important for TS calculations, as well as reaction path following methods. However it may not be practical or necessary to calculate a full Hessian for macromolecular systems, since a large amount of degrees of freedom are unrelated to the reaction of interest. Therefore, these degrees of freedom can be somewhat disregarded. The partial Hessian vibrational analysis (PHVA),(Li & Jensen 2002) was developed to diagonalize only a subblock of the Hessian matrix to yield vibrational frequencies for partially optimized systems. In a recent development of vibrational subsystem analysis (VSA), the complexity of Hessian calculation can be reduced by separating a large system into an active "subsystem" with the remainder of the system defining the "environment". The environment is kept at a minimum energy with respect to the motion of the subsystem, thus an effective Hessian involving only the subsystem needs to be considered.(Woodcock et al. 2008) The VSA is an improvement over PHVA, but does entail higher computational costs. These reduced Hessian approaches can only work if an adequate subsystem including all important atoms for the reaction can be identified.

By applying the methods introduced in this section, one could locate an IRC on any given PES with well-defined TSs. There are certain limitations of these methods, especially for large systems with high degrees of freedom. For large systems, the uphill walking methods should not be one's first choice since PESs of such systems are rather complicated and rugged, with numerous local saddle points. The downhill walking methods worked very well for the listed studies. However, a TS needs to be identified before applying IRC calculations. There is always a possibility that the IRC calculated from identified TS does not reach the desired reactant or product. It is even more likely that there are multiple TSs that exist between the reactant or product. In consideration of these factors, there might be more interest to obtain information about reaction pathway rather than TSs, especially for large biomolecular systems. Accordingly, the so-called chain-of-states methods were developed to obtain a reaction pathway without identifying TSs.

3. Chain-of-states methods

In chain-of-states methods, a number of replicas (i.e. states) of a system are used to connect two endpoints, and are subject to minimization simultaneously. The first and the last replicas usually correspond to the reactant and product, and are often fixed during the minimization. For large complex systems, chain-of-states methods can be used to address issues relating to hysteresis, free energy, reaction rates, and multiple pathways. In this section, various chain-of-states methods to build reaction paths are surveyed.

3.1 Line integral methods

Elber and Karplus (EK) developed a method using a line integral representation of a discretized path subject to optimization.(Elber 1987) In their method, the objective function subject to optimization reads as

$$S^{EK}(\boldsymbol{R}_0,...,\boldsymbol{R}_M)_L = \frac{1}{\sqrt{\sum_{j=1}^{M}\left(\Delta l_j\right)^2}}\sum_{j=1}^{M}V(\boldsymbol{R}_j)\Delta l_j + \sum_{j=1}^{M}\lambda\left(\Delta l_j - \sqrt{\sum_{j=1}^{M}\frac{(\Delta l_j)^2}{M}}\right)^2 \quad (4)$$

where $\boldsymbol{R}_j$ is the coordinates of replica *j*, M is the total number of steps from the starting replica $\boldsymbol{R}_0$ to the final replica $\boldsymbol{R}_M$, $\Delta l_j = \sqrt{[(\boldsymbol{R}_j - \boldsymbol{R}_{j-1})^2]}$, and $V(\boldsymbol{R}_j)$ is the potential energy of the system at replica $\boldsymbol{R}_j$. The degrees of freedom of the rigid body, i.e. translation and rotation, are projected out from the minimization for replicas with reference to the end points of the path. The objective function is subject to non-linear optimization for the final reaction path. The method was applied to several systems including the conformational change of myoglobin.(Elber 1987)

Czerminski and Elber then developed the self-penalty walk (SPW) method (Czerminski & Elber 1990a) based on original EK formulation. The main development of SPW is the addition of repulsion terms for each replica *j*:

$$\rho\sum_{i>j+1}^{M}\exp\left(-\frac{\Delta l_{i,j}^2}{\left(\lambda\Delta l\right)^2}\right) \quad (5)$$

where $\Delta l_{i,j}$= | $\mathbf{R}_i$-$\mathbf{R}_j$ | , Δl= $\sqrt{\sum_{j=1}^{M}\frac{(\Delta l_j)^2}{M}}$, λ and ρ determine the range and maximal value of the repulsion between replica *i* and *j*. These repulsive terms can help to prevent the aggregation of replicas in the neighborhood of two endpoints where the energies of replicas are lower than those close to the TS region. As discussed in their paper, the repulsive terms reflect the stiffness of the reaction path and mimic the effect of kinetic energy on the classical trajectory. The reaction path calculation of alanine dipeptide isomerization using SPW displayed a convergence rate that is 10 times faster than the one using the EK method. The conformational change of isobutyryl-ala$_3$-NH-methyl (IAN) between the helix and extended chain was also studied. By using the optimal values of parameters in this method, a reaction path that is very close to the MEP was obtained for IAN from the calculations with a straight line as initial path. In another study, the SPW method was applied to study the diffusion of carbon monoxide through leghemoglobin.(Nowak et al. 1991) Three similar but distinct diffusion pathways were identified and compared. The barrier heights calculated for the three pathways were in the agreement with the proposed model.

Ulitsky and Elber (UE) proposed a locally updated planes (LUP) method to calculate steepest descent paths (SDP) in flexible polyatomic systems.(Ulitsky & Elber 1990) For a series of replicas $\{\mathbf{r}^k\}_{k=1,M}$, $\mathbf{s}^k$ is the unit vector along the gradient for replica $\mathbf{r}^k$. The SPD satisfies that $\nabla V_{proj} = \nabla V - \left(\nabla V \bullet \mathbf{s}^k\right)\mathbf{s}^k = 0$, where V is the potential energy. For a discretized path, the vector $\mathbf{s}^k$ is approximated as $(\mathbf{r}^{k+1} - \mathbf{r}^{k-1})/ | \mathbf{r}^{k+1} - \mathbf{r}^{k-1} |$. To refine the path of each round, the coupled differential equations of all the replicas $\{(\partial/\partial t)\mathbf{r}^k(t)=\nabla \boldsymbol{V}_{\text{proj}}\}_{k=1,M}$ were solved by a fifth order Adams predictor-corrector algorithm.(Gear 1971) The SDP could be reached in the limit of $t\to\infty$. Choi and Elber later improved the LUP method by a

gradient updating scheme.(Choi & Elber 1991) The local gradient vector $\mathbf{s}^k$ is calculated based on the initial path, and updated based on new reaction path after every M steps of optimization. The authors found that M=20 to be efficient for the helix formation of tetrapeptides under their studies. As Choi and Elber pointed out, the final results depend on the initial guess. If multiple MEPs exist between two endpoints and the initial guess is not in the radius of convergence of a single path, the result may be discontinuous and contain segments from different MEPs.

3.2 Nudged elastic band (NEB) methods

As pointed by Jónsson, Mills and Jacobesen,(Jónsson et al. 1998) the line integral methods suffer from the "corner cutting" problems in which the final paths bypass the TS region, leading to overestimated barriers. This problem originates from the fact that elastic forces added to replicas have non-zero components perpendicular to the path. The optimization of objective functions that include elastic forces will have a tendency to pull replicas off from the MEP. In addition, replicas along the final path tend to aggregate around endpoints where the potential energies are smaller than the TS region, which is underrepresented in the chain. This is because the actual forces of the path pull the replicas downhill against elastic forces. To solve these problems, the NEB methods were developed to project out perpendicular components of elastic forces and parallel components of the true force with respect to the path under minimization.(Henkelman et al. 2000a; Jónsson et al. 1998)

For a reaction path with N+1 replicas [$\mathbf{R}_0$, $\mathbf{R}_1$,…,$\mathbf{R}_N$], the NEB method is implemented as the following: The tangent at the replica i, $\boldsymbol{\tau}_i$, can be estimated based on adjacent replicas i-1 and i+1:

$$\hat{\boldsymbol{\tau}}_i = \frac{\mathbf{R}_{i+1} - \mathbf{R}_{i-1}}{\left|\mathbf{R}_{i+1} - \mathbf{R}_{i-1}\right|} \tag{6}$$

The force acting on a replica subject to optimization is

$$\mathbf{F}_i^{NEB} = \mathbf{F}_i^{\perp} + \mathbf{F}_i^{s//} \tag{7}$$

where $\mathbf{F}_i^{\perp}$ is the sum of the true force perpendicular to the tangent:

$$\mathbf{F}_i^{\perp} = -\nabla V(\mathbf{R}_i) + \nabla V(\mathbf{R}_i) \bullet \hat{\boldsymbol{\tau}}_i \hat{\boldsymbol{\tau}}_i \tag{8}$$

and $\mathbf{F}_i^{s//}$ is the elastic force along the tangent,

$$\mathbf{F}_i^{s//} = k\left(\left|\mathbf{R}_{i+1} - \mathbf{R}_i\right| - \left|\mathbf{R}_i - \mathbf{R}_{i-1}\right|\right) \bullet \hat{\boldsymbol{\tau}}_i \hat{\boldsymbol{\tau}}_i \tag{9}$$

V is the potential energy of the system, k is the elastic force constant. By projecting out the elastic force perpendicular to the path, the final path should relax to the MEP in principle.

One potential problem that the NEB method may encounter is producing kinks along the path, mainly in regions where the parallel component of force is large compared with the perpendicular component. A new NEB method was developed by Henkelman and Jónsson to eliminate the kinks along the path.(Henkelman & Jónsson 2000) In the new implementation, the tangent is defined as

$$\boldsymbol{\tau}_i = \begin{cases} \mathbf{R}_{i+1} - \mathbf{R}_i & if\ V_{i+1} > V_i > V_{i-1} \\ \mathbf{R}_i - \mathbf{R}_{i-1} & if\ V_{i+1} < V_i < V_{i-1} \end{cases} \tag{10}$$

where V_i is the potential energy of replica i, $V(\mathbf{R}_i)$. However, when the replica i is at a minimum ($V_{i+1}>V_i<V_{i-1}$) or at a maximum ($V_{i+1}<V_i>V_{i-1}$), the tangent is estimated as

$$\boldsymbol{\tau}_i = \begin{cases} (\mathbf{R}_{i+1} - \mathbf{R}_i)\Delta V_i^{\max} + (\mathbf{R}_i - \mathbf{R}_{i-1})\Delta V_i^{\min} & if\ \ V_{i+1} > V_{i-1} \\ (\mathbf{R}_{i+1} - \mathbf{R}_i)\Delta V_i^{\min} + (\mathbf{R}_i - \mathbf{R}_{i-1})\Delta V_i^{\max} & if\ \ V_{i+1} < V_{i-1} \end{cases} \tag{11}$$

where $\Delta V_i^{\max} = \max\left(|V_{i+1} - V_i|, |V_{i-1} - V_i|\right)$, and $\Delta V_i^{\min} = \min\left(|V_{i+1} - V_i|, |V_{i-1} - V_i|\right)$. This implementation was motivated by the calculations of finding the MEP from a TS. It is always easier to find a MEP by following the PES downhill from a TS rather than going uphill from a minimum. Therefore, the local tangent of the path is always defined by the higher energy replicas nearby to improve the stability of the calculation. The updated NEB method is reported to behave well and remove kinks from the reaction paths.(Henkelman & Jónsson 2000)

Henkelman, Uberuaga and Jónsson (Henkelman et al. 2000b) developed climbing image NEB (CI-NEB) by defining the force of the replica with the highest energy $i_{\max}$ as

$$\mathbf{F}_{i\max} = -\nabla V(\mathbf{R}_{i\max}) + 2\nabla V(\mathbf{R}_{i\max}) \bullet \hat{\boldsymbol{\tau}}_{i\max}\hat{\boldsymbol{\tau}}_{i\max} \tag{12}$$

The basic idea is to invert the component of force along the path. Therefore, the minimization movement based on a given force will lead this replica toward the saddle point, which has the maximum energy along the path but minimum in all other directions. In addition, the elastic force constant along the path is scaled linearly on the energy of the replicas. Therefore, the replicas around the TS region would be connected through stronger elastic bands than those closer to the endpoints. Compared with the regular NEB method, the CI-NEB method generates reaction pathways with TS regions better represented, and leads to more accurate estimates of reaction barriers.

Maragakis *et al.* (Maragakis et al. 2002) proposed the adaptive nudged elastic band approach (ANEBA) to search for saddle points. Instead of starting with a large number of replicas, three movable replicas are added for the initial round of the NEB calculation. After obtaining reasonable convergence, the two replicas that are adjacent to the one with the highest energy will serve as endpoints for the next round of NEB calculations after inserting new movable replicas in between. This process could be repeated until a well-defined TS is identified. The major goal of ANEBA is TS searching, but a reaction path that approximates the MEP will be obtained after finding the TS.

Trygubenko and Wales proposed a doubly nudged elastic band (DNEB) method by retaining a portion of perpendicular component of elastic force.(Trygubenko & Wales 2004b; Trygubenko & Wales 2004a) The limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) quasi-Newton optimizer was implemented with the DNEB method. One of the strategies adapted is gradually projecting the perpendicular component of elastic force until the full path is reasonably stable. Then replicas with potential energies above their adjacent replicas are subject to an eigenvector-following calculation for TSs.

Chu *et al.* (Chu et al. 2003) developed the first superlinear minimizer for the NEB method that was based on expanding the adopted basis Newton-Raphsion (ABNR) method (Brooks

et al. 1983), and this is available in CHARMM.(Brooks et al. 2009) Other improved optimizers,(Alfonso & Jordan 2003; Bergonzo et al. 2009; Galván & Field 2008) and other software packages, such as AMBER,(Case et al. 2005) have efficient implementations of NEB as well.

In general, the NEB methods have been applied in numerous studies, including diffusion processes on surfaces,(Ágoston & Albe 2010; Sørensen et al. 1996; Uberuaga et al. 2000; Villarba & Jonsson 1994; Yang et al. 2009) nucleation process,(Lutsko 2008) stability of nanoparticle,(Vélez et al. 2008) and dissociative adsorption of a molecule on a surface.(Mills 1995)

Xie, Liu and Yang adapted the NEB method for enzymatic reactions.(Xie et al. 2004) The major development of their method is carefully choosing the degrees of freedom that are essential for the reaction. The chosen degrees of freedom were subject for the calculation of distances between replicas. The large number of soft and floppy spectator degrees of freedoms are, therefore, excluded from the elastic force computation in NEB. The potential problem of such selection is discontinuity of contributions from spectator degrees of freedoms to the total potential energy. One solution is starting from relatively rigid reference systems and keeping position restraints to the environmental atoms throughout the NEB calculations. The third alteration is cutting off the elastic band for intermediate states from multiple reaction steps to allow them to relax to the minimum. They applied this modified NEB method to study the mechanism of enzyme class A β-lactamase, and obtained the MEP of the reaction at the active site. With further improvement,(Cisneros et al. 2005; Liu et al. 2004) the NEB methods were applied in more enzyme systems to map out the detailed MEP of their mechanisms.(Cisneros et al. 2009; Zhao & Liu 2008)

3.3 Zero temperature string (ZTS) methods

Similar to NEB, the ZTS method was developed to find the reaction pathway connecting two minima on a PES.(Cameron et al. 2010; E et al. 2002, 2007; Ren 2003) In the ZTS method, a series of evenly distributed initial states between two endpoints are minimized towards a MEP using path gradients and tangent information. Instead of adding spring forces on the path to maintain appropriate distances between states, a step is added to evenly redistribute the states along the path after each step of minimization. The basics of the ZTS method can be summarized as the following. For transitions between two metastable states A and B, the MEP is a smooth curve φ^* between A and B which satisfies

$$(\nabla V)^{\perp}(\varphi^*) = 0 \tag{13}$$

where $(\nabla V)^{\perp}$ is the component of ∇V perpendicular to φ^*. From an arbitrary string φ connecting A and B, searching MEP can be realized by evolving φ through

$$u^{\perp} = -(\nabla V)^{\perp}(\varphi) \tag{14}$$

where $u^{\perp}$ is the perpendicular component of force with reference to φ. For simplicity, φ is parametrized by normalized arc length a, for which a=0 at A and a=1 at B. The Eq. (13) can be rewritten as

$$\varphi_{\tau} = -(\nabla V)^{\perp} + r\hat{\tau} \tag{15}$$

where

$$(\nabla V)^{\perp} = \nabla V - (\nabla V, \hat{\tau})\hat{\tau} \tag{16}$$

$\hat{\tau}$ is the unit tangent factor of φ and scalar field $r(\alpha,t)$ is a Lagrange multiplier determined by parameterization. In real applications, the string is discretized into a series of replicas, which move according to the first term of the right hand side of Eq. (15). Eq. (15) can be solved using ordinary differential equation (ODE) solvers to evolve the string towards a MEP. The reparametrization can be carried out after certain number of evolving steps of the string to redistribute the replicas to enforce the equal arc length (distance) between adjacent replicas.
In a further development,(E et al. 2007) the Eq. (15) is rewritten as

$$\varphi_{\tau} = -(\nabla V) + \bar{r}\hat{\tau} \tag{17}$$

where $\bar{r} = r + (\nabla V, \hat{\tau})$. Eq. (17) is equivalent to Eq. (15), but avoids a force projection operation in the string evolving step for better numerical stability. It is noted that the ZTS method does not require the location of minima beforehand to generate an initial string. The final converged string should connect two minima, as long as the two endpoints of the initial string lie in the attraction basins of the minima (Fig. 3). Based on the framework of the ZTS method, the transition path theory was developed to sample the minimum free energy path. This theory will be introduced in section 4.3.

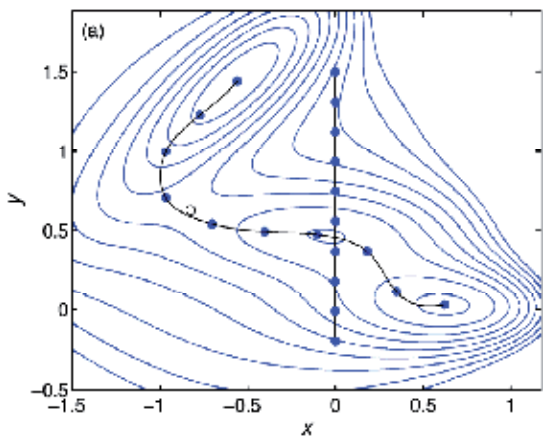


Fig. 3. Initial string and calculated MEP using the string method with ten images. The empty circle indicates the saddle point identified by combining the string method with the climbing image technique. (Reprinted with permission from ref. (E et al. 2007). Copyright 2007, American Institute of Physics.)

3.4 Growing string methods

So far, the chain-of-states algorithms introduced in section 3.1−3.3 require an initial path composed by series of replicas. The initial paths are usually constructed by linear interpolation between two minima. There are two potential drawbacks of these methods when applying them in studies involving expensive quantum mechanics (QM) calculations. First, the expensive QM calculation needs to be carried out for every replica in each round of

optimization from the very beginning of the reaction path calculation until convergence. Second, the initial linear interpolation path may include structures with severely overlapping atoms, which can cause failure of the QM calculations. The first drawback means intensive computation of reaction path optimization. The second drawback simply means immediate failure of the chain-of-states calculation. Growing string methods (GSM) were proposed (Peters et al. 2004) and are under continuous development to address these issues.(Goodrow et al. 2008, 2009, 2010; Quapp 2005, 2009)

The basic idea is to gradually grow two strings separately from two minima by continuously adding replicas to each string until both strings merge. The GSM generally comprises five steps.(Goodrow et al. 2008) In step one, two replicas are added along the linear synchronous transit path, and placed close to reactant and product, which defines the linear path. The reactant and product plus each added replica close to them serve as starting segments of two separate strings. In step two, the two string segments are minimized until the norm of the orthogonal force of each replica converges to a specified tolerance. After step two, the two string segments lie on the steepest descent direction from the frontier replica of each segment. In step three, replicas on both string segments are redistributed uniformly in terms of arc length, similar to the ZTS. The arc length of the string can be calculated by integrating the cubic spline fitted to all the replicas. The rigid body movement needs to be projected out for all the replicas with reference to the reactant and product after each round of the minimization. In step four, a new replica is added to each string segment along the fitted cubic spline. Steps two through four are repeated until the arc length between two frontier replicas from the string segments is small enough so that the two string segments can be considered a united string. In step five, the joined string is minimized and reparametrized with a fixed number of replicas until the sum of the norm of the orthogonal forces on all nodes falls under predefined tolerance. Step five is analogous to the ZTS method. One or more replicas with maximum energies among their neighbors can be used for TS search calculations. The general GSM is illustrated in Fig. 4.

In various developments, Newton projector,(Quapp 2005) internal coordinate,(Goodrow et al. 2008) conjugate gradient method,(Goodrow et al. 2008) and different TS search strategies (Goodrow et al. 2009, 2010) are implemented in GSM for better performance.

The major goal of GSM is improving the calculation efficiency of the ZTS framework for systems involving expensive QM calculations. By growing strings stepwise from two endpoints instead of minimizing from an initial path with potentially bad geometries, a large amount of calculations can be saved from carrying out QM calculations to move the fixed number of replicas from initial path to MEP iteratively. The GSM has been applied in one study using a model system representation of an enzymatic system (Maresh et al. 2008) and some other studies using small models.(Goodrow & Bell 2008; Zheng & Bell 2008a, b) The application of GSM directly on protein system has yet to be reported.

3.5 Conjugate peak refinement (CPR) method

The CPR method (Fischer 1992; Noé et al. 2006) focuses on finding saddle points along a path that connects a predefined reactant (**r**) and product (**p**). At the beginning of the CPR calculation, the replica with maximum potential energy, $\mathbf{y}_1$, is searched along an initial path, e.g. linear interpolation between **r** and **p** (Fig. 5). Then, the energy is minimized along the direction of each conjugate vector to reach the next replica, $\mathbf{x}_1$. This replica is added to the path, which is represented as (**r**, $\mathbf{x}_1$, **p**). The procedure is repeated for the partial path (**r**, $\mathbf{x}_1$) and ($\mathbf{x}_1$, **p**). If there is only one TS that exists between **r** and **p**, the maximum may not

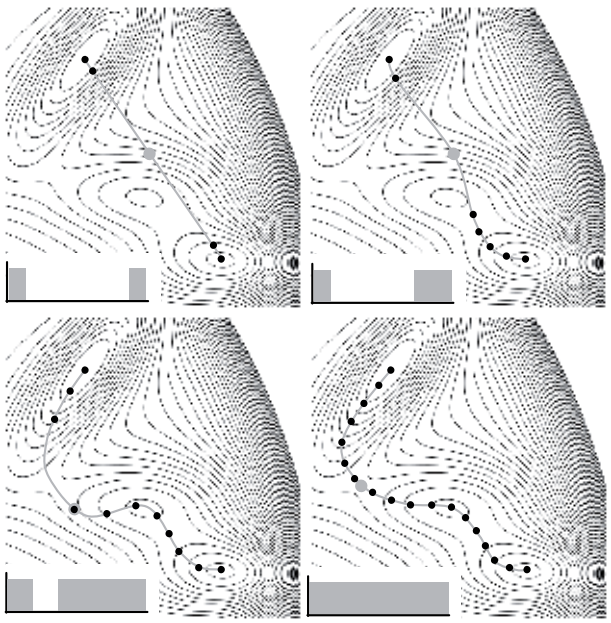


Fig. 4. Four snapshots of a growing string on the Muller–Brown potential energy surface. (Reprinted with permission from ref. (Peters et al. 2004). Copyright 2004, American Institute of Physics.)

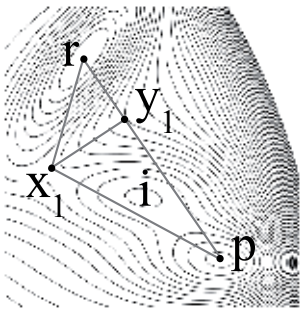


Fig. 5. Illustration of the conjugate peak refinement (CPR) method. The reactant, product and intermediate wells are indicated with r, p and i. (Reprinted from ref. (Fischer 2004), Copyright 2004, with permission from Elsevier.)

exist for one path segment. Therefore the procedure will not be carried out for this specific path segment. The procedure is carried out recursively on each new path segment defined by newly added replicas, until the root mean square (RMS) of the gradient falls under a predefined tolerance for the saddle points by refinement along conjugate vectors. In practice, the linear interpolation often produces very poor structures with severe overlapping atoms. This could result in extremely high energies and gradients. Therefore, certain limitations need to be imposed on atomic movements during refinement. The CPR method can possibly identify one or more TSs between **r** and **p** through recursive searches and refinement. However, a smooth path will be not a direct result from the CPR procedure. From identified TSs, steepest descent calculations can be carried out to generate a final MEP

between **r** and **p**. Implicitly, it is assumed that TSs identified through the CPR method lie on the same MEP. Otherwise, the steepest descent calculations from TSs do not necessarily generate a continuous path between **r** and **p**. The CPR method has been applied to study the ligand-binding pathways in myolobins,(Golden & Olsen 2008) several small molecule systems (Fischer et al. 1994; Fischer et al. 1995; Verma et al. 1996) as well as large molecules, such as proteins.(Blondel et al. 1999; Caflisch et al. 1997; Dutzler et al. 2002; Fischer et al. 1993; Gruia et al. 2005; Santos et al. 2000)

3.6 Reference path methods

In some applications, especially for macromolecules, the convergence to the MEP using chain-of-states methods could require a large number of iterations. Reference path methods were developed to generate reference reaction paths that are a good approximation to true MEPs with fast convergence rates and evenly distributed replicas along the path. Such reference paths can be used as an expansion to calculate free energies, reaction rates, or kinetic isotope effects.

3.6.1 Hyperplane projection methods

Czerminski and Elber (Czerminski & Elber 1989) proposed applying holonomic constraints on replica **R** along a reaction path connecting replicas $\mathbf{R_1}$ and $\mathbf{R_2}$:

$$(\mathbf{R}-\mathbf{R}_\alpha)\bullet(\mathbf{R_1}-\mathbf{R_2})=0 \tag{18}$$

where $\mathbf{R}_\alpha=\left[(1-\alpha)\mathbf{R_1}+\alpha\mathbf{R_2}\right]$, and α is a parameter that varies from 0 (reactant) to 1 (product) with small steps. For each α value, the potential energy of the system $V(\mathbf{R}_\alpha)$ is minimized while keeping the system in the hyperplane defined by Eq. (18). It is necessary to project out the rigid body motions from the optimization using either $\mathbf{R_1}$ or $\mathbf{R_2}$ as reference structures. The optimized path satisfies the condition that at any point along the path, only one direction may have negative energy curvature, because energy is minimized for all degrees of freedom except the direction perpendicular to the hyperplane defined by Eq. (18). The Powell conjugate gradient algorithm was applied to optimize the reaction path with linear constraints in this method. The conformational change of IAN between the helix and extended chain was studied using this method. The lowest energy path of the transition identified by the calculation was presented. In addition, the 257 direct transitions between 112 minima of IAN were also identified and subject to statistical analysis. The same authors later introduced another constraint to the path using parameter η:

$$\frac{\left|\mathbf{R}-\mathbf{R_1}\right|}{\left|\mathbf{R}-\mathbf{R_1}\right|+\left|\mathbf{R}-\mathbf{R_2}\right|}-\eta=0 \tag{19}$$

where η is the parameter between 0 (reactant) and 1 (product).(Czerminski & Elber 1990b) Combining application of α and η constraints, more comprehensive results were presented for conformational changes of IAN.

3.6.2 RPATH/restraint method

Woodcock *et al.* implemented the reaction path (RPATH) method with restraints of both best-fit root mean square distances (RMSD) and angles on replicas along the pathway. (Woodcock et al. 2003) The RMSD restraint forces are defined as

$$E_{rms} = \sum_{i=1}^{N} \frac{1}{2} k_{rms} (r_i - \bar{r})^2 \tag{20}$$

where N is the number of replicas, k_{rms} is the force constant used to restrain distance between adjacent replicas along the reaction pathway, r_i is the best-fit RMSD between replica i and i+1, and $\bar{r}$ is the average distance between adjacent replicas. The angle energetic penalty term reads

$$\begin{aligned} E_{angle} &= \sum_{i=1}^{N} \frac{1}{2} k_{ang} (COSMAX - \cos(\theta)_i)^2 \quad COSMAX > \cos(\theta)_i, \\ E_{angle} &= 0 \quad COSMAX \le \cos(\theta)_i \end{aligned} \tag{21}$$

The angle θ, illustrated in Fig. 6, defines the deviation of the pathway from linearity. The force constant k_{ang} controls the rigidity of the pathway. Constant COSMAX determines the value of cos(θ) subject to the angle forces.

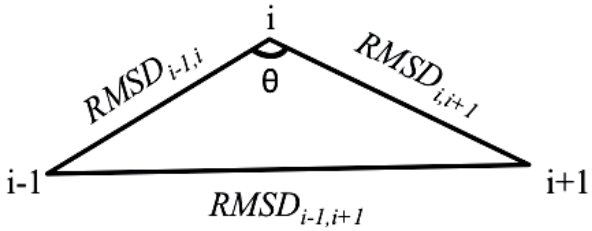


Fig. 6. Definition of angle θ for replica i in RPATH calculation. $RMSD_{i-1,i}$ is the distance between replica i-1 and i. It is similar to $RMSD_{i,i+1}$ and $RMSD_{i-1,i+1}$.

The RPATH/restraint method was applied to study the mechanism of chorismate mutase. (Woodcock et al. 2003) All protein residues or water molecules that had any atom within 6 Å of the substrate were replicated 21 times to build the reaction path. All other atoms within the system served as the environment of all the replicas. Starting from a linear interpolation pathway, the optimization of the RPATH/restraint calculation was converged within 400 steps. The reaction barrier obtained from RPATH/restraint calculations was also in good agreement with experiments.

3.6.3 RPATH/constraint method

Recently, Chu and coworkers (Brokaw et al. 2009) applied an equal distance holonomic constraint in the RPATH framework. Given two states of a molecular system with N atoms, $\boldsymbol{r}^0$ and $\boldsymbol{r}^k$, a chain of K+1 replicas can be constructed to connect these two states. The distance between each pair of adjacent replicas is set to be equal to each other:

$$\Delta l^0 = \bullet\bullet\bullet = \Delta l^i = \bullet\bullet\bullet = \Delta l^{K-1} = \overline{\Delta l} \tag{22}$$

Here, Δl^i is the distance between replica i and i+1. $\overline{\Delta l}$ is the average distance between adjacent replicas. The distance (Δl) can be in any form, including best-fit RMS distance. To apply the equal distance constraint defined in Eq. (22), the following scheme is used to propagate the reaction path with two ends, $\boldsymbol{r}^0$ and $\boldsymbol{r}^k$, fixed at the initial coordinates.

i. Set up and calculate initial average distance, $\overline{\Delta l}$, for replicas $r^{0(0)}$ through $r^{k(0)}$. The superscript "(0)" indicates the optimization iteration step.
ii. To maintain the equal distance, a set of K coefficients, $(\lambda^i)^{(n)}(i=0,K-1)$, are used to update the coordinates of each replica i:

$$\left(\mathbf{r}^i\right)^{(n+1)} = \left(\mathbf{r}^i\right)^{(n)} + \left(\lambda^{i-1}\right)^{(n)}\left(\frac{\partial \Delta l^{i-1}}{\partial \mathbf{r}^i}\right)^{(n)} + \left(\lambda^i\right)^{(n)}\left(\frac{\partial \Delta l^i}{\partial \mathbf{r}^i}\right)^{(n)} \tag{23}$$

iii. Solve $(\lambda^i)^{(n)}$ by setting the first order Taylor expansion of each of $((\lambda^i)^{(n+1)} - \overline{\Delta l})(i=0,K-1)$ with respect to $(\lambda j)^{(n)}$ to zero:

$$-\left(\left(\Delta l^i\right)^{(n)} - \overline{\Delta l}\right) = \sum_{j=i-1}^{i+1}\left(\frac{\partial \Delta l^i}{\partial \lambda^j}\right)^{(n)}_{\lambda^j}\left(\lambda^j\right)^{(n)} \tag{24}$$

iv. If any of the values of $|(\Delta l^i)^{(n+1)} - \overline{\Delta l}|$ ($i=0,K-1$) is greater than a selected tolerance, then repeat steps (ii) and (iii).

After convergence, the RPATH calculation leads to a reaction path composed by K+1 equal distance replicas connecting states $\boldsymbol{r^0}$ and $\boldsymbol{r^k}$.

In the framework of using constraints with RPATH, a kinetic energy potential can be added to the potential energy with overall objective functions as a Hamiltonian for path optimization. Therefore the optimized path is the so-called minimum Hamiltonian path (MHP) instead of MEP. The kinetic energy component in the potential prevents kinks, therefore helping maintain the smoothness of the path. This smoothness comes at the cost of deviating from the MEP, resulting in higher reaction barriers, but with the benefit as a better starting point for free energy studies.

The RPATH/constraint was applied to study the helix-to-sheet transition of a GNNQQNY heptapeptide.(Brokaw et al. 2009) An initial reaction path with 100 replicas was generated. Then a stable intermediate was chosen to divide the pathway into two segments in which the number of replicas was doubled. Using this divide-and-conquer strategy, a smooth reaction pathway with 464 replicas was obtained to describe the transition (Fig. 7).

4. Free energy sampling of reaction paths for large systems

The methods introduced in Section **3** provide a framework for determining a minimum potential energy pathway connecting two equilibrium states. For biological processes the minimum *free* energy pathway (MFEP) is often more desirable to compute, ensuring that the results obtained for the PES are converged, in a thermodynamic sense. Experimental observables can be transformed more readily to free energies of the reactant and product states.(Hu et al. 2008) Biased sampling methods, such as umbrella sampling (Bartels & Karplus 1997; Kästner 2011; Torrie & Valleau 1977; Torrie & Valleau 1974) and metadynamics (Bussi et al. 2006; Laio 2002; Laio & Gervasio 2008; Raiteri et al. 2006) have been widely applied to sample free energies of the transition path for large systems. Due to their high computational cost, the applications of these methods have been limited in studies of enzymatic reaction mechanisms using QM/MM methods. However, powerful modern computing systems and advanced methodologies make the dynamical simulation of enzymatic reactions for free energy profiles much more feasible than a decade ago. In this

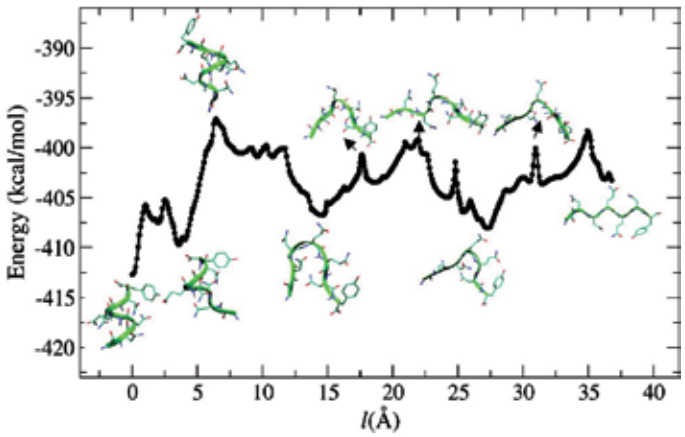


Fig. 7. An optimized path of the helix-to-sheet transition of a solvated GNNQQNY heptapeptide RPATH/constraint. The path contains a total of 464 replicas. Structures of selected replicas are shown to illustrate the nature of the transition. (Reprinted with permission from ref. (Brokaw et al. 2009). Copyright 2009 American Chemical Society.)

section, methods for direct simulation of protein systems using QM/MM methods are surveyed.

4.1 QM/MM free energy path (FE) and minimum free energy (MFEP) path methods

QM/MM-FE methods have been pioneered by Yang and coworkers (Zhang et al. 2000) and are sequential based approaches for first optimizing an equilibrium geometry on a reaction potential energy surface, followed by the calculation of free energies between two states, or along a path. (Hu & Yang 2008) The first innovation of their approach involves separating the QM and MM parts of the system and optimizing them independently to allow for faster convergence. There is no concurrent optimization of the QM and MM energy functions making it effective in reducing the number of expensive QM energy and gradient evaluations. Starting from a given structure, the coordinates of the MM portion are frozen while the QM coordinates are optimized. The MM portion is subsequently minimized while the QM portion is held fixed and the process is repeated until convergence is reached for each partition. Further simplification is made by using an approximate QM/MM energy function for the MM minimization where only Coulombic interactions between the point charges of the MM atoms and the QM electrostatic potential (ESP) fitted charges are accounted for:

$$\left\langle \Psi \left| -\sum_{1}^{N_{off}} \sum_{\beta \in MM} \frac{q_\beta}{r_{\beta i}} + \sum_{\alpha \in QM,\ \beta \in MM} \frac{Z_\alpha^* q_\beta}{r_{\alpha\beta}} \right| \Psi \right\rangle = \sum_{\alpha \in QM,\ \beta \in MM} \frac{Q_\alpha^* q_\beta}{r_{\alpha\beta}} \tag{25}$$

In order for the approximation to yield a monotonic convergence of the QM/MM energy, the MM minimization must be able to approximate the PES closely. It is important to fully optimize the reactant and product structure at the same level of theory applied on QM/MM-FE calculations.

From the converged PES obtained through the sequential optimizations, a free energy of perturbation is performed:

$$\Delta F_{QM/MM}^{(A\to B)} = F_{QM/MM}(R_c^B) - F_{QM/MM}(R_c^A)$$
$$= -\frac{1}{\beta}\ln\left\langle \exp\left\{-\beta\left[E_{QM/MM}\left(r_{QM}^{\min}\left(R_c^B\right)\right) - E_{QM/MM}\left(r_{QM}^{\min}\left(R_c^A\right)\right)\right]\right\}\right\rangle_{MM,A} \quad (26)$$

R_c^A and R_c^B are different points along the reaction coordinate and $\langle ... \rangle_{MM,A}$ represents an ensemble average over the MM sub-system with the QM region frozen to the optimized coordinates.(Zhang et al. 2000) A major advantage of the QM/MM-FE method over previous QM-FE approaches (Chandrasekhar & Jorgensen 1985; Chandrasekhar et al. 1985; Jorgensen 1989) is the inclusion of the enzymatic environment in the QM optimization.

Yang and coworkers (Hu & Yang 2008) have extended the QM/MM-FE method to overcome the challenge of having the optimized path be influenced by the initial starting conformation. This influence is problematic when solvent effects may be prominent in the reaction system, e.g. for solvent exposed active sites. Instead of using MM minimization in the sequential optimization scheme, MD sampling is used for obtaining the MM subsystem minima. Optimization of the QM region is then performed with a fixed MM conformational ensemble. Another key distinction of this so-called QM/MM-MFEP over the original QM/MM-FE method is that the optimization function for the QM region is the QM potential of mean force

$$A^{(n)}(\mathbf{r}_{QM}) = A_{ref} - \frac{1}{\beta}\ln\left\{\frac{1}{N}\sum_{\tau=1}^{N}\exp\{-\beta\Delta E\}\right\} \quad (27)$$

instead of the total energy surface. The associated gradient with respect to the *i*-th QM coordinate

$$\frac{\partial A^{(n)}(\mathbf{r}_{QM})}{\partial \mathbf{r}_{QM,i}} = \frac{\sum_{\tau=1}^{N}\partial\tilde{E}\left(\mathbf{r}_{QM}, \mathbf{r}_{MM}^{(n)}(\tau)/\partial\mathbf{r}_{QM,i}\exp\{-\beta\Delta E\}\right)}{\sum_{\tau=1}^{N}\exp\{-\beta\Delta E\}} \quad (28)$$

provides a convenient way to compute the free energy as an ensemble average over the QM atoms. The efficiency of the method derives from having a fixed size ensemble of MM conformations instead of repetitive sampling at each step. This is achieved through the updated QM reference structures of each iteration cycle. A precise potential of mean force (PMF) is obtained through optimization with classical numerical schemes. Convergence is equally quick, often achieved within 10 steps.

4.2 Transition path sampling (TPS) methods

First pioneered by Pratt (Pratt 1986) and developed much further by Chandler and his collaborators,(Bolhuis et al. 2002; Bolhuis et al. 2000; Dellago & Bolhuis 2004; Dellago & Bolhuis 2009; Dellago et al. 2002) TPS methods are a set of computational techniques to determine the transition process between two metastable states of complex systems. TPS does not require a reaction path *a priori*, because no specific or "typical" reaction pathway is pursued in TPS methods at all. Instead, a large number of transitions between two metastable states (say *A* and *B* in Fig. 8) need to be sampled to represent a transition path

ensemble (TPE) between these states. The TPE comprises all the dynamical trajectories that start from one state and end up in the other, which are called reactive trajectories. It is impractical to sample the TPE by starting a dynamics simulation from either *A* or *B* in many trials and hoping that enough number of trials can overcome the barrier and reach the other state. This is simply because the transition between *A* and *B* is a rare event, and can only be observed in an extremely limited number of trajectories if at all. Instead of this direct sampling strategy, the importance sampling methods, such as generalized MC procedures, were applied to enhance the sampling efficiency and generate a collection of reactive trajectories with a frequency proportional to their probability weight in the TPE.
Starting from a reactive trajectory, new trajectories can be generated through so-called *shooting moves* (Fig. 8). From initial trajectory, a random frame is selected as the *shooting point*. Then, the velocity of each particle in the shooting point is perturbed. From the shooting point with perturbed velocity, the equations of motion are integrated forward and backward to obtain a complete trajectory. In a simplified formalism, the new trajectory is accepted based on the acceptance probability:

$$P_{acc}[x^{(o)}(T) \to x^{(n)}(T)] = h_A[x_0^{(n)}]h_B[x_T^{(n)}] \tag{29}$$

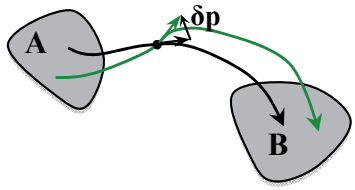


Fig. 8. In the shooting algorithm for deterministic dynamics a new path (green) is generated from an old one (black) by first randomly selecting one point on the old path, the *shooting point*. Then, the particle momenta at that point are modified by addition of a small perturbation δp. From the point with perturbed momenta the equations of motion are integrated forward and backward to obtain a complete trajectory. For small perturbations, the new trajectory will be close to the old one near the shooting point but will then rapidly diverge from it due to the chaoticity of the underlying dynamics. (With kind permission from Springer Science+Business Media: Fig. 8 in ref. (Dellago & Bolhuis 2009).)

Function $h_A[x_0^{(n)}]$ indicates that whether the system resides in state A at time 0 in the new trajectory: $h_A[x_0^{(n)}]$ is 1 for $x_0 \in A$, and 0 otherwise. Function $h_B[x_T^{(n)}]$ is defined analogously to indicate whether the system resides in state B at time T in the new trajectory. According to this expression, any new trajectory that is reactive will be accepted. The perturbation added to the shooting point can be adjusted to control the acceptance probability and optimize the TPE sampling efficiency.(Dellago et al. 1999) To begin path sampling, an initial reactive trajectory is required as the starting point. In most studies of rare events, this may not be achieved by simply running a long molecular dynamics trajectory starting from either metastable state. In practice, the system can be driven from one state to the other artificially. Such a trajectory may not carry large statistical weight in the TPE, but will be sufficient as a starting point for the sampling procedure.
A key factor for a successful TPS calculation is the sampling efficiency in the path space. For a simple system with only one MEP and one TS in between, the path sampling could

converge quickly. However, the PES of a complex system may have multiple reaction pathways, with one or more stable intermediate states presenting on each pathway. To enhance the sampling efficiency for such a complex system, the TPS with path replica exchange (Bolhuis 2008) and multi states transition TPS (Rogal & Bolhuis 2008) were developed. Based on the TPS framework, a transition interface sampling (TIS) method (Moroni et al. 2004; Van Erp et al. 2003; Vanerp & Bolhuis 2005) was developed to measure the positive flux through a series of hypersurfaces in phase space for the calculation of reaction rate constants. The TPS and TIS methods were applied to simulate the β-hairpin folding,(Bolhuis 2003) base-pair binding of DNA,(Hagan 2003) Trp-cage folding in explicit solvent,(Juraszek & Bolhuis 2006) and the reaction mechanism of lactate dehydrogenase. (Quaytman & Schwartz 2007)

4.3 Transition path theory (TPT)

Most of the reaction path theories are based on the TST,(Eyring 1935; Wigner 1938; Yamamoto 1960) with major emphasis on identifying a first order saddle point on the PES as the TS. The TS concept is also the foundation for the methods introduced in section **2** and **3**. The TPS method bypasses the TS identification by sampling the reactive trajectories directly. (Bolhuis et al. 2002) Vanden-Eijnden and his coworkers advanced further by proposing TPT. (E et al. 2005b; E et al. 2005a; E & Vanden-Eijnden 2006; Maragliano et al. 2006; Metzner et al. 2006; Ren et al. 2005; Vanden-Eijnden ; Vanden-Eijnden & Venturoli 2009) The TPT is a framework to study the transition trajectories between two metastable states and the probability density functions of these reactive trajectories. In the TPT framework, the reactive trajectories are viewed as a portion of a long trajectory that oscillates between the metastable states *A* and *B* (Fig. 9). Two important quantities in TPT are probability density function and probability current function, which can be defined for a hypothetical long trajectory of an overdamped system with friction γ_i and white noise η_i on compoment x_i, as the following. For a given position x (neither in A nor in B) and t, the function $q(x)$ defines the probability that the trajectory reaches first *B* rather than *A*. The probability density to observe a reactive trajectory at point x and time t is

$$Z^{-1}e^{-\beta V(x)}q(x)(1-q(x)) \tag{30}$$

where Z is the partition function of the whole trajectory, V(x) is the potential function of the system at x. Another quantity of interest is the probability flux of reactive trajectories across an isoprobability surface, or isocommittor surface, dividing *A* and *B* through x. The flux is the vector field with ith component reads

$$J_{AB,i}(x) = Z^{-1}e^{-\beta V(x)}k_B T\gamma_i^{-1}\frac{\partial q(x)}{\partial x_i} \tag{31}$$

where k_B is the Boltzmann factor, T is the temperature. It should be noted that for a system in thermodynamic equilibrium, the net flux through any surface is zero. The overall reaction rate from A to B can be calculated by integrating the reactive flux through an isocommittor surface *S* dividing *A* and *B*:

$$v_R = \int_S \hat{\mathbf{n}}_S(z) \bullet J_R(z) d\sigma_S(z) \tag{32}$$

where $\hat{\mathbf{n}}_S(z)$ is the unit vector perpendicular to *S* toward *B*. The TPT is a general theory about the rare transition events between metastable states based on reactive trajectories

but in configurational space. Similar to the TPS, the concept of reactive trajectories does not require the identification of the TS, which is not well defined for many complex systems.

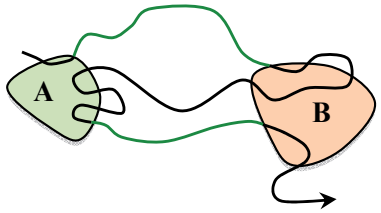


Fig. 9. Schematic illustration of a long trajectory oscillating between the reactant state A and the product state B. The reactive pieces of this trajectory, during which the system travels from A to B, are shown in green. (Reprinted with permission from ref. (E & Vanden-Eijnden 2010). Copyright 2010 Annual Reviews.)

The method to apply TPT in practice is finite temperature string (FTS) method,(E et al. 2005b) which is a finite temperature generalization of the ZTS method introduced in section 3.3. In the FTS method, the concept of string ensemble $\varphi^{\omega}(a)$ is introduced to have a mean value as the string $\varphi(a)$, where a is a parameter between [0,1]. The string ensemble can evolve by the equation

$$\varphi_t^{\omega} = -\left(\nabla V\left(\varphi^{\omega}\right)\right)^{\perp} + \left(\eta^{\omega}\right)^{\perp} + r\hat{\tau} \tag{33}$$

where $\hat{\tau}$ is the unit tangent vector along φ, η^{ω} is a white noise with covariance

$$\left\langle \eta_j^{\omega}(\alpha,t)\eta_k^{\omega}(\alpha',0)\right\rangle = \begin{cases} 2k_B T\delta_{jk}\delta(t) & \text{if } \alpha = \alpha' \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

Using isocommitter surface, a transition tube can be defined as certain region around the most probable reactive trajectory through isocommitter surfaces between A and B. These regions have a significant probability to be visited by reactive trajectories. Of course, there could exist more than one transition tube between two metastable states. These may correspond to the multiple reaction pathways or mechanisms observed on the PES.
The free energy associated with isocommitter surface $S(a)$ is given as

$$F(\alpha) = -k_B T \ln \int_{S(\alpha)} e^{-\beta V(q)} d\sigma \tag{35}$$

In this presentation, these hyperplanes serve as reaction coordinates of the transition between metastable states. The free energies are a minimum at A and B. There may exist one or more maxima between the two metastable states.
The FTS method has been applied in a couple of biomolecular systems very recently. In one study,(Rosta et al. 2011) the mechanism of cleavage of the RNA backbone catalyzed by ribonuclease H is elucidated through QM/MM simulations. The converged strings for the entire reaction are reported with estimated free energies along the path (Fig. 10). In another study,(Ovchinnikov et al. 2011) the conformational change between the prepowerstroke and rigor states of myosin VI was simulated using a replica exchange umbrella sampling

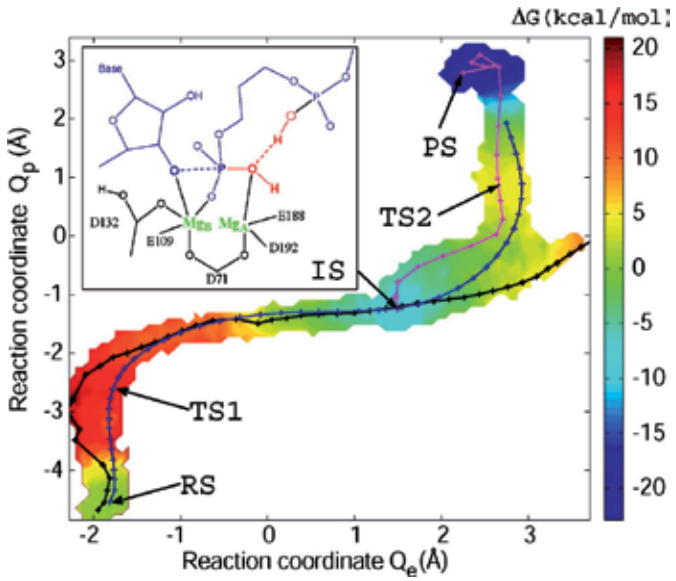


Fig. 10. 2D free energy surface and reaction strings for the complete catalytic reaction with a protonated Asp132, obtained by projection onto coordinates Qe = ET probing bond formation and breaking and Qp = (PT1a+PT1b)/2+PT2 probing PT steps. The black curve shows the initial string; the blue and purple curves show the converged strings for the entire reaction and for the final step, respectively. The inset shows a schematic of the active-site coordination in the intermediate state. (Reprinted with permission from ref. (Rosta et al. 2011). Copyright 2011 American Chemical Society.)

algorithm in a string method framework. The free energy and reaction rate of the transition were calculated and found to be consistent with the experimental observation.(De La Cruz et al. 2001)

The harmonic Fourier bead (HFB) approximation proposed by Khavrutskii and his coworkers (Khavrutskii et al. 2006; Khavrutskii et al. 2008a; Khavrutskii & Mccammon 2007) possesses many similarities with FTS method. The main difference is the representation and parametrization of the path by

$$q_i(\alpha) = q_i(0) + (q_i(1) - q_i(0))\alpha + \sum_{n=1}^{P} b_n^i \sin(n\pi\alpha) \tag{36}$$

where a is single progress parameter between 0 and 1, q_i is the ith bead, which is the ith component of the configuration vector R=(q_1,…,q_{3N}), b_n^i are the amplitudes of the Fourier basis functions. The string or reaction path in HFB can evolve under certain constraints or reparametrization, similar to ZTS and FTS, to obtain MEP and MFEP. The HFB method was applied to study the conformational transition of the signature peptide of the KcsA ion selectivity filter.(Khavrutskii et al. 2008b) A novel hypothesis of the ion selectivity mechanism was proposed based on the HFB simulation results.

4.4 Action based methods

There is a group of methods that are based on minimization of action functional of transition paths. According to the Onsager-Machlup (Onsager & Machlup 1953) and Freidlin-Wentzell

theories,(E et al. 2004; Freidlin 1984) the most probable path for the random dynamical system is the *minimum action path* (MAP). The MAP is a more general definition than MEP, for MEP is undefined for nongradient systems. For gradient systems with energy landscapes, the MAP degenerates with MEP in the high friction limit. Therefore, MEP can be obtained through optimization of path action. Olender and Elber (Olender 1997) first proposed to minimize a functional, which is called "scalar work" for a trajectory:

$$W_{scalar} = \int |\nabla V||dl| \tag{37}$$

where V is the potential energy of the system, *dl* is a segment of the trajectory. For the steepest descent path, W_{scalar} has a minimum value, where ∇V is parallel to *dl* along the path. More recently, the geometric minimum action method (Vanden-Eijnden & Heymann 2008) and adaptive minimum action method (Zhou et al. 2008) were developed as two action based methods for MEP optimization.

Some action-based methods are developed to simulate the reaction path. Doniach and his co-workers (Eastman et al. 2001) proposed reaction path annealing methods to simulate protein folding trajectories which are distributed according to the Onsager-Machlup action functional:

$$P[\mathbf{X}(t)] \propto \exp\left[\frac{-S[\mathbf{X}(t)]}{k_B T}\right] \tag{38}$$

where the action S[**X**(*t*)] can be calculated for an overdamped system with friction γ in discretization form as

$$S[\mathbf{X}(t)] = \frac{1}{2\gamma}\sum_i \left(\gamma \mathbf{M}\left(\frac{\mathbf{X}_i - \mathbf{X}_{i-1}}{\Delta t}\right) + \nabla V(\mathbf{X}_i)\right)^2 \bullet \Delta t \tag{39}$$

M is the mass matrix, V is the potential energy function, and Δt is the time step of the trajectories. The reaction path annealing method was employed to simulate the conformational changes among the α-helix, β-sheet and the random coil of a seven residue peptide from prion-like protein.(Lipfert et al. 2005)

Faccioli and co-workers developed the dominant reaction path (DRP) method, and applied it on numerous systems.(Autieri et al. 2009; Faccioli 2008; Faccioli et al. 2006; Mazzola et al. 2011; Sega et al. 2007) Their method is based on the Fokker-Planck equation, which defines the probability of finding a particle at position *x* and time *t*:

$$\frac{\partial}{\partial t}P(\mathbf{X},t) = D\frac{\partial}{\partial \mathbf{X}}\left(\frac{1}{k_B T}\frac{\partial U(\mathbf{X})}{\partial \mathbf{X}}P(\mathbf{X},t)\right) + D\frac{\partial^2}{\partial \mathbf{X}^2}P(\mathbf{X},t) \tag{40}$$

where D is the diffusion constant, *U* is the potential energy function for the particle. The Boltzmann distribution of probability $P(\mathbf{X}) \sim \exp(-V(\mathbf{X})/k_B T)$ is a stationary solution of Eq. (40). Using a substitution

$$P(\mathbf{X},t) = e^{-(\frac{1}{2}k_B T)U(\mathbf{X})}\psi(\mathbf{X},t) \tag{41}$$

the Fokker-Planck equation can be written in the form of a time dependent Schrödinger equation,

$$-\frac{\partial}{\partial t}\psi(\mathbf{X},t)=\hat{H}_{eff}\psi(\mathbf{X},t) \tag{42}$$

where the effective Hamiltonian operator is defined as

$$\hat{H}_{eff}=-D\nabla^2+V_{eff}(\mathbf{X}) \tag{43}$$

and the effective potential is defined as

$$V_{eff}(\mathbf{X})=\frac{D}{4(k_BT)^2}\left[(\nabla U(\mathbf{X}))^2-2k_BT\nabla^2U(\mathbf{X})\right] \tag{44}$$

A path-integral representation can be obtained as the solution of Eq. (42) subject to the boundary conditions as that the system is in state *A* at time 0, and state *B* at time *T*. this presentation also allows a switch from the time-dependent Newtonian description to the energy-dependent Hamilton-Jacobi (HJ) description. In the HJ framework, the most probable trajectory can be obtained by minimizing the target HJ functional

$$S_{HJ}(\mathbf{X})=\int_A^B dl\sqrt{\frac{\gamma}{k_BT}\left[E_{eff}+V_{eff}(\mathbf{X})\right]} \tag{45}$$

where $dl=\sqrt{d\mathbf{X}^2}$, γ is the friction, and E_{eff} is an effective energy parameter, which determines the total time of trajectory. In very recent studies, the simulation of reaction pathways connecting two boundary states were carried out to search for the dominant reaction pathway for folding of a small peptide,(Faccioli et al. 2010) an organic reaction from cyclobutene to butadiene,(Beccara et al. 2010) and the folding of tetraalanine at semiempirical level of theory.(Beccara et al. 2011) Although the DPR method is still in its early stage of development, the current progress demonstrates the efficiency of this method in sampling the reaction pathways for rare transition events.

5. Summary

The development of reaction path methods to study the mechanism of rare events of high dimensional systems has been an active research field for more than two decades. Many efficient methods have been developed and applied in numerous applications. The IRC is a well defined reaction pathway of the MEP. It plays an important role in explaining reactions in complicated systems, especially when the reaction coordinates are not simple order parameters. However, one can only generate an IRC after obtaining a TS. In many studies, identifying a TS is the bottleneck of the research. In some cases, the IRC generated from a TS may not connect with the target reactant and product. For large systems, there may exist multiple reaction pathways between the reactant and product. To map out multiple reaction pathways through IRC, it is required to identify multiple TSs *a priori*, which is a very challenging task. The chain-of-states methods were developed to simultaneously optimize a series of replicas as the representation of the reaction pathway connecting two equilibrium states. Starting from an initial guess, the path optimization could lead to a reaction pathway that closely resembles the MEP between two equilibrium states. One advantage of the chain-of-states methods is that no TS is required to generate a reaction path. The replicas with

highest energy from reaction path calculations are often subject to a TS search for better estimation of the reaction barriers.
For large systems, special caution is needed when carrying out the IRC calculations or chain-of-states calculations. To ensure the continuity of the reaction pathways that connect the reactant and product, large amounts of spectator degrees of freedom need to be controlled, or frozen, during the calculations. This treatment can make the working PES cleaner, but has the drawback of ignoring important degrees of freedom that contribute to the free energies, which can be directly compared with the experimental observations. Recently, more research has been reported to develop methods that sample the reaction pathway directly in the transition path space, rather than the phase space. With these developments, multiple reaction pathways could potentially be sampled without obtaining TSs for each pathway *a priori*. For rare events with high thermodynamic or kinetic barriers, the efficiency of sampling in the reaction path space is higher than the one of sampling in the phase space. The free energies could be estimated for each pathway with the sampled ensembles. Currently, the reaction path sampling method is still in its early stage of development for both theory and sampling techniques. It is clear that advanced computational power, theories and sampling techniques are in sight to make the computational study of the mechanisms of rare events for large systems more practical and appealing.

6. Abbreviations

μ-ADDF, microiteration-ADDF;
ABNR, adopted basis Newton-Raphsion;
ADDF, anharmonic downward distortions following;
ANEBA, adaptive nudged elastic band approach;
CI-NEB, climbing image NEB;
CPR, conjugate peak refinement;
dADDF, double-ended ADDF;
DC, dissociation channels;
DNEB, doubly nudged elastic band;
DRP, dominant reaction path;
DVV, damped velocity Verlet;
EK, Elber and Karplus;
ES, equilibrium structure;
ESP, electrostatic potential;
EulerPC, Euler predictor and corrector method;
FTS, finite temperature string;
GRRM, global reaction route mapping;
GS−IRC, Gonzalez and Schlegel IRC;
GSM, Growing string method;
HFB, harmonic Fourier bead;
HJ, Hamilton-Jacobi;
HPC, Hessian based predictor-corrector;
IAN, isobutyryl-ala$_3$-NH-methyl;
IRC, intrinsic reaction coordinate;
lADDF, large ADDF;
L-BFGS, limited-memory Broyden-Fletcher-Goldfarb-Shanno;

LUP, locally updated planes;
MAP, minimum action path;
MC, Monte Carlo;
MD, molecular dynamics;
MEP, minimum energy path;
MFEP, minimum free energy path;
MHP, minimum Hamiltonian path;
MMP2, matrix metalloproteinase 2;
NEB, nudged elastic band;
ODE, ordinary differential equation;
PES, potential energy surface;
PHVA, partial Hessian vibrational analysis;
PMF, potential of mean force;
QM, quantum mechanics;
QM/MM, combined quantum mechanical and molecular mechanical method;
RMS, root mean square;
RMSD, root mean square distances;
RPATH, reaction path;
SDP, steepest descent path;
SHS, scaled hypersphere search;
SPW, self-penalty walk;
TPE, transition path ensemble;
TPS, transition path sampling;
TPT, transition path theory;
TS, transition state;
TST, transition state theory;
UE, Ulitsky and Elber;
VSA, vibrational subsystem analysis;
ZTS, zero temperature string.

7. References

Ágoston, P., Albe, K. (2010). Ab initio modeling of diffusion in indium oxide. *Phys. Rev. B*, Vol. 81, No. 19, pp. 195205.

Alfonso, D. R., Jordan, K. D. (2003). A flexible nudged elastic band program for optimization of minimum energy pathways using ab initio electronic structure methods. *J. Comput. Chem.*, Vol. 24, No. 8, pp. 990-996.

Autieri, E., Faccioli, P., Sega, M., Pederiva, F., Orland, H. (2009). Dominant reaction pathways in high-dimensional systems. *J. Chem. Phys.*, Vol. 130, No. 6, pp. 064106.

Ayala, P. Y., Schlegel, H. B. (1997). A combined method for determining reaction paths, minima, and transition state geometries. *J. Chem. Phys.*, Vol. 107, No. 2, pp. 375-384.

Bader, R. F. W., Gangi, R. A. (1975). Ab initio calculation of potential energy surfaces, In: *Theoretical Chemistry*. Dixon RN, Thomson C ed. pp. 1-65, Royal Society of Chemistry, ISBN 9780851867649, Cambridge.

Bair, S., Mccabe, C., Cummings, P. (2002). Comparison of Nonequilibrium Molecular Dynamics with Experimental Measurements in the Nonlinear Shear-Thinning Regime. *Phys. Rev. Lett.*, Vol. 88, No., pp. 058302.

Baker, J. (1986). An algorithm for the location of transition states. *J. Comput. Chem.*, Vol. 7, No. 4, pp. 385-395.

Banerjee, A., Adams, N., Simons, J., Shepard, R. (1985). Search for stationary points on surfaces. *J. Phys. Chem.*, Vol. 89, No. 1, pp. 52-57.

Bartels, C., Karplus, M. (1997). Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. *J. Comput. Chem.*, Vol. 18, No. 12, pp. 1450-1462.

Beccara, A. S., Faccioli, P., Sega, M., Pederiva, F., Garberoglio, G., Orland, H. (2011). Dominant folding pathways of a peptide chain from ab initio quantum-mechanical simulations. *J. Chem. Phys.*, Vol. 134, No. 2, pp. 024501.

Beccara, S. A., Garberoglio, G., Faccioli, P., Pederiva, F. (2010). Communications: Ab initio dynamics of rare thermally activated reactions. *J. Chem. Phys.*, Vol. 132, No. 11, pp. 111102.

Bell, S., Crighton, J. S. (1984). Locating transition states. *J. Chem. Phys.*, Vol. 80, No. 6, pp. 2464-2475.

Bergonzo, C., Campbell, A. J., Walker, R. C., Simmerling, C. (2009). A Partial Nudged Elastic Band Implementation for Use With Large or Explicitly Solvated Systems. *Int. J. Quantum Chem.*, Vol. 109, No. 15, pp. 3781-3790.

Beusen, D. (1996). Systematic search strategies in conformational analysis. *Drug Discovery Today*, Vol. 1, No. 10, pp. 429-437.

Blondel, A., Renaud, J.-P., Fischer, S., Moras, D., Karplus, M. (1999). Retinoic acid receptor: a simulation analysis of retinoic acid binding and the resulting conformational changes. *J. Mol. Biol.*, Vol. 291, No. 1, pp. 101-115.

Bofill, J. M., Quapp, W. (2011). Variational nature, integration, and properties of Newton reaction path. *J. Chem. Phys.*, Vol. 134, No. 7, pp. 074101.

Bolhuis, P. G. (2003). Transition-path sampling of β-hairpin folding. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 100, No. 21, pp. 12129-12134.

Bolhuis, P. G. (2008). Rare events via multiple reaction channels sampled by path replica exchange. *J. Chem. Phys.*, Vol. 129, No. 11, pp. 114108.

Bolhuis, P. G., Dellago, C., Geissler, P. L., Chandler, D. (2000). Transition path sampling: throwing ropes over mountains in the dark. *J. Phys.: Condens. Matter*, Vol. 12, No. 8A, pp. A147-A152.

Bolhuis, P. G., Chandler, D., Dellago, C., Geissler, P. L. (2002). Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, Vol. 53, No. 1, pp. 291-318.

Bonnans, J. (2003). *Numerical optimization: theoretical and practical aspects*. Springer, ISBN 9783540001911, Berlin.

Brokaw, J. B., Haas, K. R., Chu, J.-W. (2009). Reaction Path Optimization with Holonomic Constraints and Kinetic Energy Potentials. *J. Chem. Theory Comput.*, Vol. 5, No. 8, pp. 2050-2061.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, Vol. 4, No. 2, pp. 187-217.

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S. and others (2009). CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.*, Vol. 30, No. 10, pp. 1545-1614.

Bulirsch, R., Stoer, J. (1964). Fehlerabschätzungen und Extrapolation mit rationalen Funktionen bei Verfahren vom Richardson-Typus. *Numer. Math.*, Vol. 6, No. 1, pp. 413-427.

Bulirsch, R., Stoer, J. (1966a). Numerical treatment of ordinary differential equations by extrapolation methods. *Numer. Math.*, Vol. 8, No. 1, pp. 1-13.

Bulirsch, R., Stoer, J. (1966b). Asymptotic upper and lower bounds for results of extrapolation methods. *Numer. Math.*, Vol. 8, No. 2, pp. 93-104.

Bussi, G., Gervasio, F. L., Laio, A., Parrinello, M. (2006). Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.*, Vol. 128, No. 41, pp. 13435-13441.

Caflisch, A., Fischer, S., Karplus, M. (1997). Docking by Monte Carlo minimization with a solvation correction: Application to an FKBP-substrate complex. *J. Comput. Chem.*, Vol. 18, No. 6, pp. 723-743.

Cameron, M., Kohn, R. V., Vanden-Eijnden, E. (2010). The String Method as a Dynamical System. *J. Nonlinear Sci.*, Vol. 21, No. 2, pp. 193-230.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., Woods, R. J. (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.*, Vol. 26, No. 16, pp. 1668-1688.

Cerjan, C. J. (1981). On finding transition states. *J. Chem. Phys.*, Vol. 75, No. 6, pp. 2800-2806.

Chandrasekhar, J., Jorgensen, W. L. (1985). Energy profile for a nonconcerted SN2 reaction in solution. *J. Am. Chem. Soc.*, Vol. 107, No. 10, pp. 2974-2975.

Chandrasekhar, J., Smith, S. F., Jorgensen, W. L. (1985). Theoretical examination of the SN2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J. Am. Chem. Soc.*, Vol. 107, No. 1, pp. 154-163.

Choi, C., Elber, R. (1991). Reaction path study of helix formation in tetrapeptides: Effect of side chains. *J. Chem. Phys.*, Vol. 94, No. 1, pp. 751-760.

Chu, J.-W., Trout, B. L., Brooks, B. R. (2003). A super-linear minimization scheme for the nudged elastic band method. *J. Chem. Phys.*, Vol. 119, No. 24, pp. 12708-12717.

Cisneros, G. A., Liu, H. Y., Lu, Z. Y., Yang, W. T. (2005). Reaction path determination for quantum mechanical/molecular mechanical modeling of enzyme reactions by combining first order and second order "chain-of-replicas" methods. *J. Chem. Phys.*, Vol. 122, No. 11, pp. 114502.

Cisneros, G. a. S., Perera, L., Schaaper, R. M., Pedersen, L. C., London, R. E., Pedersen, L. G., Darden, T. A. (2009). Reaction Mechanism of the ε Subunit of E. coli DNA Polymerase III: Insights into Active Site Metal Coordination and Catalytically Significant Residues. *J. Am. Chem. Soc.*, Vol. 131, No. 4, pp. 1550-1556.

Cummings, P. T., Evans, D. J. (1992). Nonequilibrium molecular dynamics approaches to transport properties and non-Newtonian fluid rheology. *Ind. Eng. Chem. Res.*, Vol. 31, No. 5, pp. 1237-1252.

Czerminski, R., Elber, R. (1989). Reaction path study of conformational transitions and helix formation in a tetrapeptide. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 86, No. 18, pp. 6963-6967.

Czerminski, R., Elber, R. (1990a). Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. *Int. J. Quantum Chem.*, Vol. 38, No. S24, pp. 167-185.

Czerminski, R., Elber, R. (1990b). Reaction path study of conformational transitions in flexible systems: Applications to peptides. *J. Chem. Phys.*, Vol. 92, No. 9, pp. 5580-5601.

De La Cruz, E. M., Ostap, E. M., Sweeney, H. L. (2001). Kinetic Mechanism and Regulation of Myosin VI. *J. Biol. Chem.*, Vol. 276, No. 34, pp. 32373-32381.

Dellago, C., Bolhuis, P. (2004). Activation Energies from Transition Path Sampling Simulations. *Mol. Simul.*, Vol. 30, No. 11, pp. 795-799.

Dellago, C., Bolhuis, P. G. (2009). Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events, In: *Advanced Computer Simulation Approaches for Soft Matter Sciences III*. Holm C, Kremer K ed. pp. 167-233, Springer, ISBN 9783540877059, Berlin, Heidelberg.

Dellago, C., Bolhuis, P. G., Chandler, D. (1999). On the calculation of reaction rate constants in the transition path ensemble. *J. Chem. Phys.*, Vol. 110, No. 14, pp. 6617-6625.

Dellago, C., Bolhuis, P. G., Geissler, P. L. (2002). Transition path sampling, In: *Advances in Chemical Physics, Vol 123*. pp. 1-78, John Wiley & Sons Inc, ISBN 9780471231509, New York.

Dennis, J., Schnabel, R. B. (1996). *Numerical methods for unconstrained optimization and nonlinear equations*. Society for Industrial and Applied Mathematics, ISBN 9780898713640, Philadelphia.

Doll, J. D. (2005). A Modern Perspective on Transition State Theory, In: *Handbook of Materials Modeling*. Yip S ed. pp. 1573-1583, Springer Netherlands, ISBN 9781402032875, Dordrecht.

Dutzler, R., Schirmer, T., Karplus, M., Fischer, S. (2002). Translocation Mechanism of Long Sugar Chains across the Maltoporin Membrane Channel. *Structure*, Vol. 10, No. 9, pp. 1273-1284.

E, W., Vanden-Eijnden, E. (2006). Towards a Theory of Transition Paths. *J. Stat. Phys.*, Vol. 123, No. 3, pp. 503-523.

E, W., Vanden-Eijnden, E. (2010). Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annu. Rev. Phys. Chem.*, Vol. 61, No. 1, pp. 391-420.

E, W., Ren, W., Vanden-Eijnden, E. (2002). String method for the study of rare events. *Phys. Rev. B*, Vol. 66, No. 5, pp. 052301.

E, W., Ren, W., Vanden-Eijnden, E. (2004). Minimum action method for the study of rare events. *Comm. Pure Appl. Math.*, Vol. 57, No. 5, pp. 637-656.

E, W., Ren, W., Vandeneijnden, E. (2005a). Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chem. Phys. Lett.*, Vol. 413, No. 1-3, pp. 242-247.

E, W., Ren, W., Vanden-Eijnden, E. (2005b). Finite Temperature String Method for the Study of Rare Events. *J. Phys. Chem. B*, Vol. 109, No. 14, pp. 6688-6693.

E, W., Ren, W., Vanden-Eijnden, E. (2007). Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.*, Vol. 126, No. 16, pp. 164103.

Earl, D. J., Deem, M. W. (2005). Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, Vol. 7, No. 23, pp. 3910-3916.

Eastman, P., Grønbech-Jensen, N., Doniach, S. (2001). Simulation of protein folding by reaction path annealing. *J. Chem. Phys.*, Vol. 114, No. 8, pp. 3823-3841.

Elber, R. (1987). A method for determining reaction paths in large molecules: Application to myoglobin. *Chem. Phys. Lett.*, Vol. 139, No. 5, pp. 375-380.

Eyring, H. (1935). The Activated Complex in Chemical Reactions. *J. Chem. Phys.*, Vol. 3, No. 2, pp. 107-115.

Faccioli, P. (2008). Characterization of Protein Folding by Dominant Reaction Pathways. *J. Phys. Chem. B*, Vol. 112, No. 44, pp. 13756-13764.

Faccioli, P., Lonardi, A., Orland, H. (2010). Dominant reaction pathways in protein folding: A direct validation against molecular dynamics simulations. *J. Chem. Phys.*, Vol. 133, No. 4, pp. 045104.

Faccioli, P., Sega, M., Pederiva, F., Orland, H. (2006). Dominant Pathways in Protein Folding. *Phys. Rev. Lett.*, Vol. 97, No. 10, pp. 108101.

Farkas, Ö., Schlegel, H. B. (2003). Geometry optimization methods for modeling large molecules. *THEOCHEM*, Vol. 666-667, No. 1, pp. 31-39.

Fischer, S. (1992). Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem. Phys. Lett.*, Vol. 194, No. 3, pp. 252-261.

Fischer, S., Michnick, S., Karplus, M. (1993). A mechanism for rotamase catalysis by the FK506 binding protein (FKBP). *Biochemistry*, Vol. 32, No. 50, pp. 13830-13837.

Fischer, S., Dunbrack, R. L., Karplus, M. (1994). Cis-Trans Imide Isomerization of the Proline Dipeptide. *J. Am. Chem. Soc.*, Vol. 116, No. 26, pp. 11931-11937.

Fischer, S., Grootenhuis, P. D. J., Groenen, L. C., Van Hoorn, W. P., Van Veggel, F. C. J. M. V., Reinhoudt, D. N., Karplus, M. (1995). Pathways for Conformational Interconversion of Calix[4]arenes. *J. Am. Chem. Soc.*, Vol. 117, No. 5, pp. 1611-1620.

Fletcher, R. (2000). *Practical methods of optimization*. Wiley, ISBN 9780471494638, Chichester.

Floudas, C. A., Pardalos, P. M. (1996). *State of the art in global optimization: computational methods and applications*. Kluwer Academic, ISBN 9780792338383.

Freidlin, M. (1984). *Random perturbations of dynamical systems*. Springer, ISBN 9780387908588, New York.

Fukui, K. (1981). The path of chemical reactions - The IRC approach. *Acc. Chem. Res.*, Vol. 14, No. 12, pp. 363-368.

Galván, I. F., Field, M. J. (2008). Improving the efficiency of the NEB reaction path finding algorithm. *J. Comput. Chem.*, Vol. 29, No. 1, pp. 139-143.

Gear, C. (1971). In: *Numerical initial value problems in ordinary differential equations*., Prentice-Hall, ISBN 9780136266068, Englewood Cliffs N.J.

Gill, P. (1982). *Practical optimization*. Academic, ISBN 9780122839528, London.

Golden, S., Olsen, K. (2008). Use of the Conjugate Peak Refinement Algorithm for Identification of Ligand-Binding Pathways in Globins. *Methods in Enzymology*, Vol. 437, No. 1, pp. 417-437.

Gonzalez, C., Schlegel, H. B. (1989). An improved algorithm for reaction path following. *J. Chem. Phys.*, Vol. 90, No. 4, pp. 2154-2161.

Gonzalez, C., Schlegel, H. B. (1990). Reaction path following in mass-weighted internal coordinates. *J. Phys. Chem.*, Vol. 94, No. 14, pp. 5523-5527.

Gonzalez, C., Schlegel, H. B. (1991). Improved algorithms for reaction path following: Higher-order implicit algorithms. *J. Chem. Phys.*, Vol. 95, No. 8, pp. 5853-5860.

Goodrow, A., Bell, A. T. (2008). A Theoretical Investigation of the Selective Oxidation of Methanol to Formaldehyde on Isolated Vanadate Species Supported on Titania. *J. Phys. Chem. C*, Vol. 112, No. 34, pp. 13204-13214.

Goodrow, A., Bell, A. T., Head-Gordon, M. (2008). Development and application of a hybrid method involving interpolation and ab initio calculations for the determination of transition states. *J. Chem. Phys.*, Vol. 129, No. 17, pp. 174109.

Goodrow, A., Bell, A. T., Head-Gordon, M. (2009). Transition state-finding strategies for use with the growing string method. *J. Chem. Phys.*, Vol. 130, No. 24, pp. 244108.

Goodrow, A., Bell, A. T., Head-Gordon, M. (2010). A strategy for obtaining a more accurate transition state estimate using the growing string method. *Chem. Phys. Lett.*, Vol. 484, No. 4-6, pp. 392-398.

Gruia, A. D., Bondar, A.-N., Smith, J. C., Fischer, S. (2005). Mechanism of a Molecular Valve in the Halorhodopsin Chloride Pump. *Structure*, Vol. 13, No. 4, pp. 617-627.

Hagan, M. F. (2003). Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 100, No. 24, pp. 13922-13927.

Hamelberg, D., Mongan, J., Mccammon, J. A. (2004). Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, Vol. 120, No. 24, pp. 11919-11929.

Hänggi, P., Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.*, Vol. 62, No. 2, pp. 251-341.

Haslinger, J., Mäkinen, R. a. E. (2003). *Introduction to Shape Optimization: Theory, Approximation, and Computation*. SIAM, ISBN 9780898715361, Philadelphia, PA.

Heidrich, D. (1995). *The reaction path in chemistry: current approaches and perspectives*. Kluwer Academic Publishers, ISBN 9780792335894, Boston, MA.

Henkelman, G., Jónsson, H. (2000). Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, Vol. 113, No. 22, pp. 9978-9985.

Henkelman, G., Jóhannesson, G., Jónsson, H. (2000a). Methods for Finding Saddle Points and Minimum Energy Paths, In: *Theoretical Methods in Condensed Phase Chemistry*, Schwartz SD ed. pp. 269-302, Kluwer Academic Publishers, ISBN 9780792366874, New York.

Henkelman, G., Uberuaga, B. P., Jónsson, H. (2000b). A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.*, Vol. 113, No. 22, pp. 9901-9904.

Hoover, W. G. (1983). Nonequilibrium Molecular Dynamics. *Annu. Rev. Phys. Chem.*, Vol. 34, No. 1, pp. 103-127.

Hoover, W. G., Hoover, C. G. (2005). Nonequilibrium molecular dynamics. *Condensed Matter Physics*, Vol. 8, No. 2, pp. 247-260.

Horst, R. (1995). *Handbook of global optimization*. Kluwer Academic Publishers, ISBN 9780792331209, Dordrecht; Boston.

Horst, R. (2000). *Introduction to global optimization*. Kluwer Academic Publishers, ISBN 9780792365747, Dordrecht; Boston.

Hratchian, H. P., Schlegel, H. B. (2002). Following Reaction Pathways Using a Damped Classical Trajectory Algorithm. *J. Phys. Chem. A*, Vol. 106, No. 1, pp. 165-169.

Hratchian, H. P., Schlegel, H. B. (2004). Accurate reaction paths using a Hessian based predictor-corrector integrator. *J. Chem. Phys.*, Vol. 120, No. 21, pp. 9918-9924.

Hratchian, H., Schlegel, H. B. (2005a). Finding minima, transition states, and following reaction pathways on ab initio potential energy surfaces, In: *Theory and Applications of Computational Chemistry*. pp. 195-249, Elsevier, ISBN 9780444517197.

Hratchian, H. P., Schlegel, H. B. (2005b). Using Hessian Updating To Increase the Efficiency of a Hessian Based Predictor-Corrector Reaction Path Following Method. *J. Chem. Theory Comput.*, Vol. 1, No. 1, pp. 61-69.

Hratchian, H. P., Frisch, M. J. (2011). Integrating steepest-descent reaction pathways for large molecules. *J. Chem. Phys.*, Vol. 134, No. 20, pp. 204103.

Hratchian, H. P., Frisch, M. J., Schlegel, H. B. (2010). Steepest descent reaction path integration using a first-order predictor-corrector method. *J. Chem. Phys.*, Vol. 133, No. 22, pp. 224101.

Hu, H., Yang, W. (2008). Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods. *Annu. Rev. Phys. Chem.*, Vol. 59, No. 1, pp. 573-601.

Hu, H., Lu, Z., Parks, J. M., Burger, S. K., Yang, W. (2008). Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: Sequential sampling and optimization on the potential of mean force surface. *J. Chem. Phys.*, Vol. 128, No. 3, pp. 034105.

Ionova, I. V., Carter, E. A. (1993). Ridge method for finding saddle points on potential energy surfaces. *J. Chem. Phys.*, Vol. 98, No. 8, pp. 6377-6386.

Ionova, I. V., Carter, E. A. (1995). Direct inversion in the iterative subspace-induced acceleration of the ridge method for finding transition states. *J. Chem. Phys.*, Vol. 103, No. 13, pp. 5437-5441.

Jensen, A. (1983). Location of transition states and stable intermediates by MINIMAX/MINIMI optimization of synchronous transit pathways. *Theor. Chim. Acta*, Vol. 63, No. 4, pp. 269-290.

Jónsson, H., Mills, G., Jacobsen, K. W. (1998). Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions, In: *Classical and Quantum Dynamics in Condensed Phase Simulations*. Berne BJ, Ciccotti G, Coker DF ed. pp. 385-404, World Scientific, ISBN 9789810234980, Singapore.

Jorgensen, W. L. (1989). Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.*, Vol. 22, No. 5, pp. 184-189.

Juraszek, J., Bolhuis, P. G. (2006). Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 103, No. 43, pp. 15859-15864.

Kästner, J. (2011). Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science, Vol. 1, No. 6, pp. 932-942.*

Khavrutskii, I. V., Mccammon, J. A. (2007). Generalized gradient-augmented harmonic Fourier beads method with multiple atomic and/or center-of-mass positional restraints. *J. Chem. Phys.*, Vol. 127, No. 12, pp. 124901.

Khavrutskii, I. V., Arora, K., Brooks, C. L. (2006). Harmonic Fourier beads method for studying rare events on rugged energy surfaces. *J. Chem. Phys.*, Vol. 125, No. 17, pp. 174108.

Khavrutskii, I. V., Dzubiella, J., Mccammon, J. A. (2008a). Computing accurate potentials of mean force in electrolyte solutions with the generalized gradient-augmented harmonic Fourier beads method. *J. Chem. Phys.*, Vol. 128, No. 4, pp. 044106.

Khavrutskii, I. V., Fajer, M., Mccammon, J. A. (2008b). Intrinsic Free Energy of the Conformational Transition of the KcsA Signature Peptide from Conducting to Nonconducting State. *J. Chem. Theory Comput.*, Vol. 4, No. 9, pp. 1541-1554.

Kjelstrup, S., Hafskjold, B. R. (1996). Nonequilibrium Molecular Dynamics Simulations of Steady-State Heat and Mass Transport in Distillation. *Ind. Eng. Chem. Res.*, Vol. 35, No. 11, pp. 4203-4213.

Klähn, M., Braun-Sand, S., Rosta, E., Warshel, A. (2005). On Possible Pitfalls in ab Initio Quantum Mechanics/Molecular Mechanics Minimization Approaches for Studies of Enzymatic Reactions. *J. Phys. Chem. B*, Vol. 109, No. 32, pp. 15645-15650.

Laidler, K. J., King, M. C. (1983). Development of transition-state theory. *J. Phys. Chem.*, Vol. 87, No. 15, pp. 2657-2664.

Laio, A. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 99, No. 20, pp. 12562-12566.

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.*, Vol. 71, No. 12, pp. 126601.

Leach, A. R. (1991). A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules, In: *Reviews in Computational Chemistry, Vol 2*. Lipkowitz KB, Boyd DB ed. pp. 1-55, John Wiley & Sons, Inc., ISBN 9780470125793, Hoboken, NJ, USA.

Lei, H., Duan, Y. (2007). Improved sampling methods for molecular simulation. *Curr. Opin. Struct. Biol.*, Vol. 17, No. 2, pp. 187-191.

Lewars, E. G. (2011). The Concept of the Potential Energy Surface, In: *Computational Chemistry*. Lewars EG ed. pp. 9-43, Springer Netherlands, ISBN 9789048138609, Dordrecht.

Li, H., Jensen, J. H. (2002). Partial Hessian vibrational analysis: the localization of the molecular vibrational energy and entropy. *Theor. Chem. Acc.*, Vol. 107, No. 4, pp. 211-219.

Li, S., Sheng, N., Liu, X. (2008). A non-equilibrium multiscale simulation paradigm. *Chem. Phys. Lett.*, Vol. 451, No. 4-6, pp. 293-300.

Lipfert, J., Franklin, J., Wu, F., Doniach, S. (2005). Protein Misfolding and Amyloid Formation for the Peptide GNNQQNY from Yeast Prion Protein Sup35: Simulation by Reaction Path Annealing. *J. Mol. Biol.*, Vol. 349, No. 3, pp. 648-658.

Liu, H., Lu, Z., Cisneros, G. a. S., Yang, W. (2004). Parallel iterative reaction path optimization in ab initio quantum mechanical/molecular mechanical modeling of enzyme reactions. *J. Chem. Phys.*, Vol. 121, No. 2, pp. 697-706.

Lutsko, J. F. (2008). Density functional theory of inhomogeneous liquids. III. Liquid-vapor nucleation. *J. Chem. Phys.*, Vol. 129, No. 24, pp. 244501.

Maeda, S. (2003). A new method for constructing multidimensional potential energy surfaces by a polar coordinate interpolation technique. *Chem. Phys. Lett.*, Vol. 381, No. 1-2, pp. 177-186.

Maeda, S., Ohno, K. (2007). Structures of Water Octamers $(H_2O)_8$: Exploration on Ab Initio Potential Energy Surfaces by the Scaled Hypersphere Search Method. *J. Phys. Chem. A*, Vol. 111, No. 20, pp. 4527-4534.

Maeda, S., Ohno, K., Morokuma, K. (2009). An Automated and Systematic Transition Structure Explorer in Large Flexible Molecular Systems Based on Combined Global Reaction Route Mapping and Microiteration Methods. *J. Chem. Theory Comput.*, Vol. 5, No. 10, pp. 2734-2743.

Maragakis, P., Andreev, S. A., Brumer, Y., Reichman, D. R., Kaxiras, E. (2002). Adaptive nudged elastic band approach for transition state calculation. *J. Chem. Phys.*, Vol. 117, No. 10, pp. 4651-4658.

Maragliano, L., Fischer, A., Vanden-Eijnden, E., Ciccotti, G. (2006). String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.*, Vol. 125, No. 2, pp. 024106.

March, J. (1992). *Advanced organic chemistry: Reactions, mechanisms and structures*. Wiley, ISBN 9780471601807, New York N.Y.

Maresh, J. J., Giddings, L.-A., Friedrich, A., Loris, E. A., Panjikar, S., Trout, B. L., Stöckigt, J., Peters, B., O'connor, S. E. (2008). Strictosidine Synthase: Mechanism of a Pictet−Spengler Catalyzing Enzyme. *J. Am. Chem. Soc.*, Vol. 130, No. 2, pp. 710-723.

Mazzola, G., Beccara, S. A., Faccioli, P., Orland, H. (2011). Fluctuations in the ensemble of reaction pathways. *J. Chem. Phys.*, Vol. 134, No. 16, pp. 164109.

Metzner, P., Schütte, C., Vanden-Eijnden, E. (2006). Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.*, Vol. 125, No. 8, pp. 084110.

Mezey, P. (1987). *Potential energy hypersurfaces*. Elsevier, ISBN 9780444428875, Amsterdam; New York.

Mills, G. (1995). Reversible work transition state theory: application to dissociative adsorption of hydrogen. *Surf. Sci.*, Vol. 324, No. 2-3, pp. 305-337.

Moroni, D., Vanerp, T., Bolhuis, P. (2004). Investigating rare events by transition interface sampling. *Physica. A*, Vol. 340, No. 1-3, pp. 395-401.

Müller, K., Brown, L. D. (1979). Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta*, Vol. 53, No. 1, pp. 75-93.

Mundy, C. J., Balasubramanian, S., Bagchi, K., Tuckerman, M. E., Martyna, G. J., Klein, M. L. (2000). Nonequilibrium Molecular Dynamics, In: *Reviews in Computational Chemistry, Vol. 14*. Lipkowitz KB, Boyd DB ed. pp. 291-397, John Wiley & Sons, Inc., ISBN 9780470125915, Hoboken, NJ, USA.

Nichols, J., Taylor, H., Schmidt, P., Simons, J. (1990). Walking on potential energy surfaces. *J. Chem. Phys.*, Vol. 92, No. 1, pp. 340-346.

Nocedal, J. (2006). *Numerical optimization*. Springer, ISBN 9780387400655, New York.

Noé, F., Krachtus, D., Smith, J. C., Fischer, S. (2006). Transition Networks for the Comprehensive Characterization of Complex Conformational Change in Proteins. *J. Chem. Theory Comput.*, Vol. 2, No. 3, pp. 840-857.

Nowak, W., Czerminski, R., Elber, R. (1991). Reaction path study of ligand diffusion in proteins: application of the self penalty walk (SPW) method to calculate reaction coordinates for the motion of CO through leghemoglobin. *J. Am. Chem. Soc.*, Vol. 113, No. 15, pp. 5627-5637.

Ohno, K. (2004). A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface. *Chem. Phys. Lett.*, Vol. 384, No. 4-6, pp. 277-282.

Ohno, K., Maeda, S. (2006). Global Reaction Route Mapping on Potential Energy Surfaces of Formaldehyde, Formic Acid, and Their Metal-Substituted Analogues. *J. Phys. Chem. A*, Vol. 110, No. 28, pp. 8933-8941.

Okur, A., Wickstrom, L., Layten, M., Geney, R. E., Song, K., Hornak, V., Simmerling, C. (2006). Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model. *J. Chem. Theory Comput.*, Vol. 2, No. 2, pp. 420-433.

Olender, R. (1997). Yet another look at the steepest descent path. *THEOCHEM*, Vol. 398-399, No. 1, pp. 63-71.

Olsen, R. A., Kroes, G. J., Henkelman, G., Arnaldsson, A., Jónsson, H. (2004). Comparison of methods for finding saddle points without knowledge of the final states. *J. Chem. Phys.*, Vol. 121, No. 20, pp. 9776-9792.

Onsager, L., Machlup, S. (1953). Fluctuations and Irreversible Processes. *Phys. Rev.*, Vol. 91, No. 6, pp. 1505-1512.

Ovchinnikov, V., Karplus, M., Vanden-Eijnden, E. (2011). Free energy of conformational transition paths in biomolecules: The string method and its application to myosin VI. *J. Chem. Phys.*, Vol. 134, No. 8, pp. 085103.

Parish, C. (2002). A comparison of the Low Mode and Monte Carlo conformational search methods. *J. Mol. Graphics Modell.*, Vol. 21, No. 2, pp. 129-150.

Pechukas, P. (1981). Transition State Theory. *Annu. Rev. Phys. Chem.*, Vol. 32, No. 1, pp. 159-177.

Peng, C., Ayala, P. Y., Schlegel, H. B., Frisch, M. J. (1996). Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.*, Vol. 17, No. 1, pp. 49-56.

Peters, B., Heyden, A., Bell, A. T., Chakraborty, A. (2004). A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *J. Chem. Phys.*, Vol. 120, No. 17, pp. 7877-7886.

Pratt, L. R. (1986). A statistical method for identifying transition states in high dimensional problems. *J. Chem. Phys.*, Vol. 85, No. 9, pp. 5045-5048.

Quapp, W. (2005). A growing string method for the reaction pathway defined by a Newton trajectory. *J. Chem. Phys.*, Vol. 122, No. 17, pp. 174106.

Quapp, W. (2008). Chemical reaction paths and calculus of variations. *Theor. Chem. Acc.*, Vol. 121, No. 5-6, pp. 227-237.

Quapp, W. (2009). The growing string method for flows of Newton trajectories by a second-order method. *J. Theor. Comput. Chem.*, Vol. 8, No. 1, pp. 101-117.

Quapp, W., Heidrich, D. (1984). Analysis of the concept of minimum energy path on the potential energy surface of chemically reacting systems. *Theor. Chim. Acta*, Vol. 66, No. 3-4, pp. 245-260.

Quaytman, S. L., Schwartz, S. D. (2007). Reaction coordinate of an enzymatic reaction revealed by transition path sampling. *Proc. Natl. Acad. Sci. U. S. A.*, Vol. 104, No. 30, pp. 12253-12258.

Raiteri, P., Laio, A., Gervasio, F. L., Micheletti, C., Parrinello, M. (2006). Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B*, Vol. 110, No. 8, pp. 3533-3539.

Ren, W. (2003). Higher order string method for finding minimum energy paths *COMM. MATH. SCI.*, Vol. 1, No. 2, pp. 377-384.

Ren, W., Vanden-Eijnden, E., Maragakis, P., E, W. (2005). Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.*, Vol. 123, No. 13, pp. 134109.

Rogal, J., Bolhuis, P. G. (2008). Multiple state transition path sampling. *J. Chem. Phys.*, Vol. 129, No. 22, pp. 224107.

Rosta, E., Nowotny, M., Yang, W., Hummer, G. (2011). Catalytic Mechanism of RNA Backbone Cleavage by Ribonuclease H from Quantum Mechanics/Molecular Mechanics Simulations. *J. Am. Chem. Soc.*, Vol. 133, No. 23, pp. 8934-8941.

Santos, J. S.-D. O., Fischer, S., Guilbert, C., Lewit-Bentley, A., Smith, J. C. (2000). Pathway for Large-Scale Conformational Change in Annexin V. *Biochemistry*, Vol. 39, No. 46, pp. 14065-14074.

Scales, L. (1985). *Introduction to non-linear optimization*. Springer-Verlag, ISBN 9780387912523, New York.

Schlegel, H. B. (1982). Optimization of equilibrium geometries and transition structures. *J. Comput. Chem.*, Vol. 3, No. 2, pp. 214-218.

Schlegel, H. B. (2003). Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *J. Comput. Chem.*, Vol. 24, No. 12, pp. 1514-1527.

Schlegel, H. B. (2011). Geometry optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Vol. 1, No. 5, pp. 790-809.

Sega, M., Faccioli, P., Pederiva, F., Garberoglio, G., Orland, H. (2007). Quantitative Protein Dynamics from Dominant Folding Pathways. *Phys. Rev. Lett.*, Vol. 99, No. 11, pp. 118102.

Shi, Q., Meroueh, S. O., Fisher, J. F., Mobashery, S. (2008). Investigation of the Mechanism of the Cell Wall DD-Carboxypeptidase Reaction of Penicillin-Binding Protein 5 of Escherichia coli by Quantum Mechanics/Molecular Mechanics Calculations. *J. Am. Chem. Soc.*, Vol. 130, No. 29, pp. 9293-9303.

Simons, J., Nichols, J. (1990). Strategies for walking on potential energy surfaces using local quadratic approximations. *Int. J. Quantum Chem.*, Vol. 38, No. S24, pp. 263-276.

Sørensen, M., Jacobsen, K., Jónsson, H. (1996). Thermal Diffusion Processes in Metal-Tip-Surface Interactions: Contact Formation and Adatom Mobility. *Phys. Rev. Lett.*, Vol. 77, No. 25, pp. 5067-5070.

Sugita, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, Vol. 314, No. 1-2, pp. 141-151.

Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., Morokuma, K. (1996). ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition. *J. Phys. Chem.*, Vol. 100, No. 50, pp. 19357-19363.

Swendsen, R. H., Wang, J.-S. (1986). Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.*, Vol. 57, No. 21, pp. 2607-2609.

Tachibana, A., Fukui, K. (1980). Novel variational principles of chemical reaction. *Theor. Chim. Acta*, Vol. 57, No. 1, pp. 81-94.

Tao, P., Gatti, D. L., Schlegel, H. B. (2009a). The Energy Landscape of 3-Deoxy-D-manno-octulosonate 8-Phosphate Synthase. *Biochemistry*, Vol. 48, No. 49, pp. 11706-11714.

Tao, P., Fisher, J. F., Shi, Q., Mobashery, S., Schlegel, H. B. (2010). Matrix Metalloproteinase 2 (MMP2) Inhibition: DFT and QM/MM Studies of the Deprotonation-Initialized Ring-Opening Reaction of the Sulfoxide Analogue of SB-3CT. *J. Phys. Chem. B*, Vol. 114, No. 2, pp. 1030-1037.

Tao, P., Fisher, J. F., Shi, Q., Vreven, T., Mobashery, S., Schlegel, H. B. (2009b). Matrix Metalloproteinase 2 Inhibition: Combined Quantum Mechanics and Molecular Mechanics Studies of the Inhibition Mechanism of (4-Phenoxyphenylsulfonyl)methylthiirane and Its Oxirane Analogue. *Biochemistry*, Vol. 48, No. 41, pp. 9839-9847.

Taylor, H., Simons, J. (1985). Imposition of geometrical constraints on potential energy surface walking procedures. *J. Phys. Chem.*, Vol. 89, No. 4, pp. 684-688.

Thomas, G. L., Sessions, R. B., Parker, M. J. (2005). Density guided importance sampling: application to a reduced model of protein folding. *Bioinformatics*, Vol. 21, No. 12, pp. 2839-2843.

Tolman, R. C. (1925). The mechanism of chemical reaction. *J. Am. Chem. Soc.*, Vol. 47, No. 6, pp. 1524-1553.

Torn, A. (1989). *Global optimization*. Springer-Verlag, ISBN 9783540508717, Berlin; New York.

Torrie, G., Valleau, J. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, Vol. 23, No. 2, pp. 187-199.

Torrie, G. M., Valleau, J. P. (1974). Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.*, Vol. 28, No. 4, pp. 578-581.

Truhlar, D. G. (2001). Potential Energy Surfaces, In: *Encyclopedia of Physical Science and Technology*, 3rd ed. Meyers RA ed. pp. 9-17, Academic Press, ISBN 9780122274107, New York.

Truhlar, D. G., Hase, W. L., Hynes, J. T. (1983). Current status of transition-state theory. *J. Phys. Chem.*, Vol. 87, No. 15, pp. 2664-2682.

Trygubenko, S. A., Wales, D. J. (2004a). Erratum: A doubly nudged elastic band method for finding transition states [J. Chem. Phys. 120, 2082 (2004)]. *J. Chem. Phys.*, Vol. 120, No. 16, pp. 7820.

Trygubenko, S. A., Wales, D. J. (2004b). A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.*, Vol. 120, No. 5, pp. 2082-2094.

Uberuaga, B., Leskovar, M., Smith, A., Jónsson, H., Olmstead, M. (2000). Diffusion of Ge below the Si(100) Surface: Theory and Experiment. *Phys. Rev. Lett.*, Vol. 84, No. 11, pp. 2441-2444.

Ulitsky, A., Elber, R. (1990). A new technique to calculate steepest descent paths in flexible polyatomic systems. *J. Chem. Phys.*, Vol. 92, No. 2, pp. 1510-1511.

Van Erp, T. S., Moroni, D., Bolhuis, P. G. (2003). A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.*, Vol. 118, No. 17, pp. 7762-7774.

Vanden-Eijnden, E. (2006). Transition Path Theory, In: *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Vol. 1*. Ferrario M, Ciccotti G, Binder K ed. pp. 453-493, Springer, ISBN 9783540352709, Berlin, Heidelberg.

Vanden-Eijnden, E., Heymann, M. (2008). The geometric minimum action method for computing minimum energy paths. *J. Chem. Phys.*, Vol. 128, No. 6, pp. 061103.

Vanden-Eijnden, E., Venturoli, M. (2009). Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.*, Vol. 130, No. 19, pp. 194103.

Vanerp, T., Bolhuis, P. (2005). Elaborating transition interface sampling methods. *J. Comput. Phys.*, Vol. 205, No. 1, pp. 157-181.

Vélez, P., Dassie, S. A., Leiva, E. P. M. (2008). When do nanowires break? A model for the theoretical study of the long-term stability of monoatomic nanowires. *Chem. Phys. Lett.*, Vol. 460, No. 1-3, pp. 261-265.

Verlet, L. (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, Vol. 159, No. 1, pp. 98-103.

Verma, C. S., Fischer, S., Caves, L. S. D., Roberts, G. C. K., Hubbard, R. E. (1996). Calculation of the Reaction Pathway for the Aromatic Ring Flip in Methotrexate Complexed to Dihydrofolate Reductase. *J. Phys. Chem.*, Vol. 100, No. 7, pp. 2510-2518.

Villarba, M., Jonsson, H. (1994). Diffusion mechanisms relevant to metal crystal growth: Pt/Pt(111). *Surf. Sci.*, Vol. 317, No. 1-2, pp. 15-36.

Vreven, T., Morokuma, K., Farkas, O., Schlegel, H. B., Frisch, M. J. (2003). Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *J. Comput. Chem.*, Vol. 24, No. 6, pp. 760-769.

Vreven, T., Frisch, M. J., Kudin, K. N., Schlegel, H. B., Morokuma, K. (2006a). Geometry optimization with QM/MM methods II: Explicit quadratic coupling. *Mol. Phys.*, Vol. 104, No. 5-7, pp. 701-714.

Vreven, T., Byun, K. S., Komaromi, I., Dapprich, S., Montgomery, J. A., Jr., Morokuma, K., Frisch, M. J. (2006b). Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.*, Vol. 2, No. 3, pp. 815-826.

Wales, D. (2003). *Energy landscapes*. Cambridge University Press, ISBN 9780521814157, Cambridge, UK; New York.

Wen, E. Z., Hsieh, M.-J., Kollman, P. A., Luo, R. (2004). Enhanced ab initio protein folding simulations in Poisson-Boltzmann molecular dynamics with self-guiding forces. *J. Mol. Graphics Modell.*, Vol. 22, No. 5, pp. 415-424.

Wigner, E. (1938). The transition state method. *Trans. Faraday Soc.*, Vol. 34, No. 1, pp. 29-41.

Woodcock, H. L., Hodošček, M., Sherwood, P., Lee, Y. S., Schaefer III, H. F., Brooks, B. R. (2003). Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theor. Chem. Acc.*, Vol. 109, No. 3, pp. 140-148.

Woodcock, H. L., Zheng, W., Ghysels, A., Shao, Y., Kong, J., Brooks, B. R. (2008). Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit. *J. Chem. Phys.*, Vol. 129, No. 21, pp. 214109.

Xie, L., Liu, H., Yang, W. (2004). Adapting the nudged elastic band method for determining minimum-energy paths of chemical reactions in enzymes. *J. Chem. Phys.*, Vol. 120, No. 17, pp. 8039-8052.

Yamamoto, T. (1960). Quantum Statistical Mechanical Theory of the Rate of Exchange Chemical Reactions in the Gas Phase. *J. Chem. Phys.*, Vol. 33, No. 1, pp. 281-289.

Yamashita, K., Yamabe, T., Fukui, K. (1981). IRC approach to chemical dynamics: toward mode-selective chemical reactions. *Chem. Phys. Lett.*, Vol. 84, No. 1, pp. 123-126.

Yang, J., Hu, W., Chen, S., Tang, J. (2009). Surface Self-Diffusion Behavior of a Pt Adatom on Wulff Polyhedral Clusters. *J. Phys. Chem. C*, Vol. 113, No. 52, pp. 21501-21505.

Yang, X., Maeda, S., Ohno, K. (2005). Global Investigation on the Potential Energy Surface of CH_3CN: Application of the Scaled Hypersphere Search Method. *J. Phys. Chem. A*, Vol. 109, No. 32, pp. 7319-7328.

Zhang, Y., Liu, H., Yang, W. (2000). Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.*, Vol. 112, No. 8, pp. 3483-3492.

Zhao, Z., Liu, H. (2008). A Quantum Mechanical/Molecular Mechanical Study on the Catalysis of the Pyridoxal 5′-Phosphate-Dependent Enzyme L-Serine Dehydratase. *J. Phys. Chem. B*, Vol. 112, No. 41, pp. 13091-13100.

Zheng, X., Bell, A. T. (2008a). Theoretical Analysis of the Mechanism for the Oxidative Carbonylation of Toluene to p-Toluic Acid by Rhodium Complexes. *J. Phys. Chem. C*, Vol. 112, No. 6, pp. 2129-2136.

Zheng, X., Bell, A. T. (2008b). A Theoretical Investigation of Dimethyl Carbonate Synthesis on Cu-Y Zeolite. *J. Phys. Chem. C*, Vol. 112, No. 13, pp. 5043-5047.

Zhou, J., Tao, P., Fisher, J. F., Shi, Q., Mobashery, S., Schlegel, H. B. (2010). QM/MM Studies of the Matrix Metalloproteinase 2 (MMP2) Inhibition Mechanism of (S)-SB-3CT and its Oxirane Analogue. *J. Chem. Theory Comput.*, Vol. 6, No. 11, pp. 3580-3587.

Zhou, X., Ren, W., E, W. (2008). Adaptive minimum action method for the study of rare events. *J. Chem. Phys.*, Vol. 128, No. 10, pp. 104111.

3

Semiclassical Methods of Deformation Quantisation in Transport Theory

M. I. Krivoruchenko
Institute for Theoretical and Experimental Physics, Moscow
Russia

1. Introduction

Among the equivalent formulations of quantum mechanics, Heisenberg's matrix mechanics and quantum theory, based on the Schrödinger wave equation, are the most technically advanced. Since the advent of quantum mechanics, these two schemes have provided the mathematical tools and the primary basis for the description of quantum phenomena. The path integral method developed by Dirac and Feynman has important advantages in the quantisation of gauge theories. We discuss here the formulation of quantum mechanics in phase space, known as deformation quantisation.

Deformation quantisation uses the Wigner-Weyl association rule (Weyl, 1927, 1929, 1931; Wigner, 1932) to establish a one-to-one correspondence between the functions in the phase space and the operators in the Hilbert space. Wigner's function appears as the Weyl symbol of the density matrix. A consistent dynamical description of the systems with the help of the Wigner function leads to deformation quantization. A useful formulation of the Wigner-Weyl association rule was proposed by Groenewold (1946) and Stratonovich (1957).

Groenewold introduced a non-commutative associative $\star$-product (star-product) of the functions in the phase space (Groenewold, 1946). The evolution of the quantum systems is determined by the antisymmetric part of the $\star$-product (Groenewold, 1946; Moyal, 1949), known as the Moyal bracket. The Moyal bracket represents the quantum deformation of the Poisson bracket. Deformation quantisation preserves many features of classical Hamiltonian dynamics.

The formulation of deformation quantisation is based on the Wigner function and the Moyal bracket (i.e., the $\star$-product). The Wigner-Weyl association rule is necessary to prove the equivalence of deformation quantisation and the standard formalisms of quantum mechanics.

Extensive literature has reported on the formulation of quantum mechanics in the phase space and the $\star$-product. We refer the reader to excellent reviews by Bayen et al. (1978a, 1978b), Carruthers & Zachariasen (1983), Balazs & Jennings (1984), Hillery et al. (1983), Karasev & Maslov (1991), and Osborn & Molzahn (1995), where one may find additional references. Wigner's function, as a fundamental object of deformation quantisation, has numerous applications in many-body physics, kinetic theory, collision theory, and quantum chemistry. Transport models, originally created to simulate chemical reactions, have been modified and are widely used to describe heavy-ion collisions.

Deformation quantisation does not have many recognised successful applications in quantum theory. Recently, attempts have been made to use specific properties of the formalism to investigate semiclassical expansion (Osborn & Molzahn, 1995; McQuarrie et al., 1998) and to calculate the determinants of one-loop operators in field theory (Pletnev & Banin, 1999; Banin et al., 2001) and the high-order corrections to the Bohr-Sommerfeld quantisation rule (Gracia-Saz, 2004; Cargo et al., 2005). Potential applications of deformation quantisation in transport models are presented in this review.

Transport models for heavy-ion physics are designed for phenomenological descriptions of the complex dynamics of nuclear collisions. Several types of advanced transport models are based on the Boltzmann-Uhlenbeck-Ueling equations (BUU) (Blaettel et al., 1993), (relativistic) quantum molecular dynamics (QMD/RQMD) (Sorge et al., 1989; Aichelin, 1991; Faessler, 1992), or antisymmetrised molecular dynamics (AQMD) (Feldmeier & Schnack, 1997). These approaches have the correct classical limit and they contain special plug-ins and quantum-mechanical attributes, such as Pauli blocking for binary collisions of fermions. Numerical solutions are implemented through the distribution of test particles (BUU) or centroids of wave packets (QMD, AQMD) for classical trajectories in the phase space. For AQMD, the wave packets are antisymmetrised in their parameters. The transport models provide a solid basis for a phenomenological description of a variety of complex nuclear collisions phenomena. However, quantum coherence effects and non-localities are beyond the scope of these models. The internal consistencies of the approximations in the models remain a subject of debate, promoting further developments (see, for example, papers by Kohler (1995) and Feldmeier & Schnack (1997) and references therein).

The most striking feature of the transport models is the depiction of the trajectories in the phase space, which are test particles or centroids of wave packets. The evolution of the system of classical particles can be calculated using standard programs to solve first-order ordinary differential equations (ODEs). At the same time, the evolution of wave functions of many-body systems is a field-theoretic problem with an infinite number of degrees of freedom that cannot be solved either analytically or numerically.

Any simulation of many-body quantum dynamics must be based on a concept of trajectories, which is the only attribute allowing access to an approximate description of complex quantum systems.

The concept of phase-space trajectories arises naturally in the formalism of deformation quantisation through the Wigner transformation of the operators of the canonical coordinates and momenta in the Heisenberg representation. These trajectories satisfy the quantum version of Hamilton's equations (Osborn & Molzahn, 1995; Krivoruchenko & Faessler, 2006b) and are the characteristics by which the time-dependent Weyl's symbols for the other operators can be determined (Krivoruchenko & Faessler, 2006b; Krivoruchenko et al., 2006c, 2007). In the classical limit, quantum characteristics reduce to classical trajectories. Knowledge of the quantum phase flow, i.e., the quantum trajectories, is equivalent to a complete knowledge of the quantum dynamics.

In this chapter, we provide an introduction to deformation quantisation and demonstrate the usefulness of the formalism in solving the evolution problem for many-body systems in terms of semiclassical expansion. We show that, in any fixed order of expansion over the Planck's constant, the evolution problem can be reduced to a statistical-mechanics problem of calculating an ensemble of quantum characteristics in the phase space and their Jacobi fields. In comparison with the corresponding rules of classical statistical mechanics, the rules for computing the probabilities and time-dependent averages of observables are

modified. The evolution equations represent a finite system of first-order ODEs for quantum trajectories in the phase space and the associated Jacobi fields (Krivoruchenko et al., 2006c).
The method of quantum characteristics allows the consistent inclusion of specific quantum effects, such as non-localities and coherence, in the description of the propagation of particles in the transport models.
In the next section, the Wigner-Weyl association rule is described, and the concept of the ★-product is introduced using the Groenewold method (Groenewold, 1946). Section 3 is devoted to the properties of the quantum characteristics. We explore the transformation properties of the canonical variables and functions in phase space under unitary transformations in Hilbert space using the Wigner-Weyl correspondence rule. The role of quantum characteristics coincides with the role of characteristics in a solution of the classical Liouville equation. Section 4 starts from the semiclassical expansion of ★-functions around the normal functions. The results are then applied to the decomposition of the functions of the quantum characteristics. Quantum characteristics, when expanded in a power series of Planck's constant, are found by solving a coupled system of ODEs for quantum characteristics and the associated Jacobi fields. The numerical methods for solving many-body scattering problem, including the rule for the calculation of the average values of physical observables, are discussed in Section 5. Special features of the scattering problem are discussed in Section 6.
In the field of deformation quantisation, the terminology is not well established. Deformation quantisation is synonymous with Weyl-Groenewold quantisation and ★-product quantisation. The ★-product is synonymous with the Moyal product. Wigner's image of an operator is Weyl's symbol. The Moyal bracket is also known as the sine bracket. The quantum Liouville equation is synonymous with the Groenewold equation and is the Wigner image of the von Neumann equation.

2. The Wigner-Weyl association rule, the ★-product and the Wigner function

In Hamiltonian formalism, classical systems with n degrees of freedom are described by $2n$ canonical coordinates and momenta

$$\xi^i = (q^1,...,q^n,p_1,...,p_n) \in \mathsf{R}^{2n}.$$

The Poisson bracket for these variables takes the simple form

$$\{\xi^k,\xi^l\} = -I^{kl}. \tag{1}$$

The matrix

$$\|I\| = \begin{Vmatrix} 0 & -E_n \\ E_n & 0 \end{Vmatrix},$$

where E_n is the identity matrix, endows the phase space with a symplectic structure. In the following, we use the form I^{ij} to lower and raise the indices, e.g., $A^{\mathrm{i}} = A_{\mathrm{j}} I^{\mathrm{ji}}$, $A_{\mathrm{i}} = I_{\mathrm{ij}} A^{\mathrm{j}}$, where $I_{\mathrm{ij}} = I^{\mathrm{ji}}$ and A is a vector in the phase space.
In quantum mechanics, the canonical variables are mapped into operators of the canonical coordinates and momenta in a Hilbert space:

$$\mathfrak{x}^{\mathrm{i}} = (\mathfrak{q}^1, ..., \mathfrak{q}^{\mathrm{n}}, \mathfrak{p}_1, ..., \mathfrak{p}_{\mathrm{n}}) \in Op(L^2(\mathbb{R}^n)) .$$

These operators obey the commutation relations

$$[\mathfrak{x}^{\mathrm{k}}, \mathfrak{x}^{\mathrm{l}}] = -i\hbar I^{kl} . \tag{2}$$

The operators $\mathfrak{f} \in Op(L^2(\mathbb{R}^n))$ are denoted by Gothic letters, and the functions in phase space $\mathbb{R}^{2n}$ are denoted by Latin letters.

The Wigner-Weyl correspondence $\xi^i \leftrightarrow \mathfrak{x}^{\mathrm{i}}$ extends to arbitrary functions and operators. A set of operators $\mathfrak{f} \in Op(L^2(\mathbb{R}^n))$ in the Hilbert space is a set closed under multiplication by *c*-numbers and summation. This set forms a vector space $\mathbb{V}$. The elements of its basis can be numbered by the phase space coordinates $\xi^i \in \mathbb{R}^{2n}$. Typically, the Weyl-Groenewold basis is used:

$$\mathfrak{B}(\xi) = (2\pi\hbar)^n \delta^{2n}(\xi - \mathfrak{x}) = \int \frac{d^{2n}\eta}{(2\pi\hbar)^n} \exp(-\frac{i}{\hbar}\eta_k(\xi - \mathfrak{x})^{\mathrm{k}}) . \tag{3}$$

The vectors $\mathfrak{B}(\xi) \in \mathbb{V}$ satisfy the following properties:

$$\begin{aligned}
&\mathfrak{B}(\xi)^+ = \mathfrak{B}(\xi), \\
&\int \frac{d^{2n}\xi}{(2\pi\hbar)^n} \mathfrak{B}(\xi) = 1, \\
&\int \frac{d^{2n}\xi}{(2\pi\hbar)^n} \mathfrak{B}(\xi) \mathrm{Tr}[\mathfrak{B}(\xi)\mathfrak{f}] = \mathfrak{f}, \\
&Tr[\mathfrak{B}(\xi)] = 1, \\
&Tr[\mathfrak{B}(\xi)\mathfrak{B}(\xi')] = (2\pi\hbar)^n \delta^{2\mathrm{n}}(\xi - \xi'), \\
&Tr[\mathfrak{B}(\xi)\mathfrak{B}(\xi')\mathfrak{B}(\xi'')] = (2\pi\hbar)^n \delta^{2\mathrm{n}}(\xi' - \xi) \exp(\frac{i\hbar}{2}\mathcal{P}_{\xi\xi})(2\pi\hbar)^n \delta^{2\mathrm{n}}(\xi - \xi''), \\
&\mathfrak{B}(\xi) \exp(-\frac{i\hbar}{2}\mathcal{P}_{\xi\xi'})\mathfrak{B}(\xi') = (2\pi\hbar)^n \delta^{2\mathrm{n}}(\xi - \xi')\mathfrak{B}(\xi'),
\end{aligned} \tag{4}$$

where $\mathcal{P}_{\xi\xi'}$ is the Poisson operator

$$\mathcal{P}_{\xi\xi'} = -I^{kl} \frac{\overleftarrow{\partial}}{\partial\xi^{\mathrm{k}}} \frac{\overrightarrow{\partial}}{\partial\xi'^{\mathrm{l}}} .$$

In Equation (4), the first line is obvious. The equations in lines two to five are equivalent to Equations (4.15) – (4.18) in Groenewold (1946). The last equation can be found, e.g., in Krivoruchenko et al. (2006a). The equation in the 6th line is a consequence of the 5th and 7th equations.

The Wigner-Weyl association rule $f(\xi) \leftrightarrow \mathfrak{f}$ takes, in the basis $\mathfrak{B}(\xi)$, the simple form (Groenewold, 1946; Stratonovich, 1957):

$$f(\xi) = Tr[\mathfrak{B}(\xi)\mathfrak{f}], \tag{5}$$

$$\mathfrak{f} = \int \frac{d^{2n}\xi}{(2\pi\hbar)^n} f(\xi)\mathfrak{B}(\xi). \tag{6}$$

In particular, $\mathfrak{r}^i \to \xi^i = Tr[\mathfrak{B}(\xi)\mathfrak{r}^i]$. The reciprocal relation $\xi^i \to \mathfrak{r}^i$, defined by the second line, also holds.

The function $f(\xi)$ can be interpreted as a coordinate of the operator $\mathfrak{f}$ in the basis $\mathfrak{B}(\xi)$, and $Tr[\mathfrak{B}(\xi)\mathfrak{f}]$ can be viewed as the scalar product of $\mathfrak{B}(\xi)$ and $\mathfrak{f}$. Other operator bases are also discussed (Balazs & Jennings, 1984).

The set $Op(L^2(\mathrm{R}^n))$ is closed under the addition and the multiplication of the operators. Thus, vector space V acquires the structure of associative algebra. For any two functions $f(\xi)$ and $g(\xi)$, a third function can be constructed (Groenewold, 1946):

$$f(\xi) \star g(\xi) \equiv Tr[\mathfrak{B}(\xi)\mathfrak{f}\mathfrak{g}]. \tag{7}$$

The operation is called star-product ($\star$-product) of $f(\xi)$ and $g(\xi)$. The explicit form of $\star$-product is as follows

$$f(\xi) \star g(\xi) = f(\xi)\exp(\frac{i\hbar}{2}\mathcal{P})g(\xi), \tag{8}$$

where $\mathcal{P} = \mathcal{P}_{\xi\xi}$. The $\star$-product splits into symmetric and skew-symmetric parts

$$f \star g = f \circ g + \frac{i\hbar}{2} f \wedge g. \tag{9}$$

The skew-symmetric component is known as the Moyal bracket (Groenewold, 1946; Moyal, 1949). In the classical limit, the Moyal bracket $f \wedge g$ turns into the Poisson bracket $\{f, g\} \equiv f(\xi)\mathcal{P}g(\xi)$.

Weyl's symbol of the symmetrised product of the operators of the canonical coordinates and momenta $\mathfrak{x}^{(i_1}\mathfrak{x}^{i_2}...\mathfrak{x}^{i_s)}$ coincides with the dot (ordinary) product of the associated canonical variables

$$Tr[\mathfrak{B}(\xi)\mathfrak{x}^{(i_1}\mathfrak{x}^{i_2}...\mathfrak{x}^{i_s)}] = \xi^{i_1}\xi^{i_2}...\xi^{i_s},$$

which is explicitly symmetric for permutations of the indices. The symmetrised products of the Hermitian operators $\mathfrak{u}^i$ correspond to the symmetrised $\star$-products of the associated real functions $u^i(\xi) = Tr[\mathfrak{B}(\xi)\mathfrak{u}^i]$:

$$Tr[\mathfrak{B}(\xi)\mathfrak{u}^{(i_1}\mathfrak{u}^{i_2}...\mathfrak{u}^{i_s)}] = u^{(i_1}(\xi) \circ u^{i_2}(\xi) \circ ... \circ u^{i_s)}(\xi).$$

The $\circ$-product is not associative. Its order here is not important, because the indices are symmetrised.

Weil's map from functions to operators was originally formulated in terms of the Taylor expansion. Consider the decomposition of a function near zero

$$f(\xi) = \sum_{s=0}^{\infty} \frac{1}{s!} \frac{\partial^s f(0)}{\partial\xi^{i_1}...\partial\xi^{i_s}} \xi^{i_1}...\xi^{i_s}.$$

According to the Weyl rule (Weyl, 1927, 1929, 1931), the function $f(\xi)$ maps into the operator $\mathfrak{f}_\Gamma$

$$\mathfrak{f}_\Gamma \equiv f(\mathfrak{x}) = \sum_{s=0}^{\infty} \frac{1}{s!} \frac{\partial^s f(0)}{\partial \xi^{i_1} \ldots \partial \xi^{i_s}} \mathfrak{x}^{i_1} \mathfrak{x}^{i_2} \ldots \mathfrak{x}^{i_s} .$$

Note that the indices of summation over the coordinate components are automatically symmetrised. A simple calculation,

$$\begin{aligned} f_T(\xi) &= Tr[\mathfrak{B}(\xi)\mathfrak{f}_\Gamma] \\ &= \sum_{s=0}^{\infty} \frac{1}{s!} \frac{\partial^s f(0)}{\partial \xi^{i_1} \ldots \partial \xi^{i_s}} \xi^{i_1} \star \xi^{i_2} \star \ldots \star \xi^{i_s} \\ &= \sum_{s=0}^{\infty} \frac{1}{s!} \frac{\partial^s f(0)}{\partial \xi^{i_1} \ldots \partial \xi^{i_s}} \xi^{i_1} \xi^{i_2} \ldots \xi^{i_s} , \end{aligned}$$

shows that the Taylor expansion of the product of the operators of the canonical coordinates and momenta provides the correspondence rule, completely equivalent to Equation (6). Thus, we obtain $f_T(\xi) = f(\xi) \leftrightarrow \mathfrak{f}_\Gamma = \mathfrak{f}$.

For any operator $\mathfrak{f} \in Op(L^2(\mathbb{R}^n))$, one may find a function $f(\xi)$ such that $\mathfrak{f} = f(\mathfrak{x})$. This property shows the completeness of the set of operators of canonical coordinates and momenta in $Op(L^2(\mathbb{R}^n))$.

The average value of a physical observable $\mathfrak{f}$ is determined by the trace of $\mathfrak{f}\mathfrak{r}$, where $\mathfrak{r}$ is the density matrix, or by averaging the function $f(\xi)$ over the Wigner function

$$W(\xi) = Tr[\mathfrak{B}(\xi)\mathfrak{r}]. \tag{10}$$

Because $Tr[\mathfrak{r}] = 1$, the Wigner function is normalised to unity

$$\int \frac{d^{2n}\xi}{(2\pi\hbar)^n} W(\xi) = 1 .$$

The average value of $\mathfrak{f}$ is given by

$$\langle \mathfrak{f} \rangle = Tr[\mathfrak{f}\mathfrak{r}] = \int \frac{d^{2n}\xi}{(2\pi\hbar)^n} f(\xi) \star W(\xi) = \int \frac{d^{2n}\xi}{(2\pi\hbar)^n} f(\xi) W(\xi) . \tag{11}$$

In this case, the $\star$-product of $f(\xi)$ and $W(\xi)$ can be replaced by a dot product, because the derivatives of the Poisson operator are reduced to surface integrals and can be omitted.

Not every normalised function in phase space can be interpreted as the Wigner function. The eigenvalues of the density matrix are nonnegative. For each density matrix, one may then find a Hermitian matrix $\mathfrak{r}^{1/2}$, such that $\mathfrak{r}^{1/2}\mathfrak{r}^{1/2} = \mathfrak{r}$. For any Wigner function $W(\xi)$ a function $W_{1/2}(\xi)$ exists, such that $W_{1/2}(\xi) \star W_{1/2}(\xi) = W(\xi)$.

For a pure state, $\mathfrak{r} = |\psi><\psi|$, and the Wigner function equals $W(\xi) = <\psi|\mathfrak{B}(\xi)|\psi>$. To show that the Wigner map, which is described in standard textbooks, is equivalent to Equation (5), we first find the mixed matrix elements of the basis vectors of V:

$$\langle x|\mathfrak{B}(q,p)|k\rangle = 2^{n}\exp\left(-\frac{2i}{\hbar}(p-k)\cdot(q-x)\right)\langle x|k\rangle ,$$

where x and q are the canonical coordinates and k and p are the canonical momenta. After simple transformations, the usual expression then follows:

$$\begin{aligned} W(q,p) &= \langle\psi|\mathfrak{B}(q,p)|\psi\rangle \\ &= \int\frac{d^n x d^n k}{(2\pi\hbar)^n}\langle\psi|x\rangle\langle x|\mathfrak{B}(q,p)|k\rangle\langle k|\psi\rangle \\ &= \int d^n x\langle\psi|q+\frac{x}{2}\rangle\exp(\frac{i}{\hbar}xp)\langle q-\frac{x}{2}|\psi\rangle. \end{aligned}$$

Applying the Schwartz inequality to the integral and taking the normalisation condition $\langle\psi|\psi\rangle = 1$ into account, one obtains the constraint $-2^n \le W(\xi) \le 2^n$ (Baker, 1958). The value of $W(\xi)$ is bounded, provided $|\psi>$ has a finite norm, which is the case for bound states of discrete spectrum and wave packets in the continuum.

3. Quantum trajectories in phase space as characteristics

One-parameter unitary transformations acting on the operators of the canonical coordinates and momenta generate the trajectories in the phase space by Wigner's association rule. Knowledge of these trajectories is equivalent to knowledge of the quantum dynamics. The time-dependent symbols of the operators are functions of the trajectories. In this sense, the phase-space trajectories play a special role, similar to the role of classical trajectories in solving the Liouville equation.

The Liouville equation is a partial differential equation (PDE). Its general solution can be represented by its characteristics. The characteristics of the classical Liouville equation are the classical trajectories of the particles. Quantum trajectories solve the Groenewold evolution equation. For this reason, we call them "quantum characteristics."

3.1 Wigner map of unitary transformation

Consider a unitary transformation acting on the operators $\mathfrak{f} \to \mathfrak{f}' = \mathfrak{U}^{+}\mathfrak{f}\mathfrak{U}$, where $\mathfrak{U}^{+}\mathfrak{U} = \mathfrak{U}\mathfrak{U}^{+} = 1$. The operators of the canonical coordinates and momenta are transformed according to the rule $\mathfrak{x}^{i} \to \mathfrak{x}'^{i} = \mathfrak{U}^{+}\mathfrak{x}^{i}\mathfrak{U}$, while their Weyl's symbols are transformed according to the rule

$$\xi^{i} \to \xi'^{i} = u^{i}(\xi) \equiv Tr[\mathfrak{B}(\xi)\mathfrak{U}^{+}\mathfrak{x}^{i}\mathfrak{U}]. \tag{12}$$

Thus, unitary transformations in the space $Op(L^2(\mathsf{R}^n))$ generate, through the Wigner association rule, a coordinate transformation in the phase space R^{2n}. Such transformations

are not canonical (see below), and we call them "unitary transformations". The transformation law of functions under the unitary transformation takes the form

$$\begin{aligned} f(\xi) \to f'(\xi) &= Tr[\mathfrak{B}(\xi)\mathfrak{f}'] = Tr[\mathfrak{B}(\xi)\mathfrak{U}^{+}\mathfrak{f}\mathfrak{U}] \\ &= \sum_{s=0}^{\infty} \frac{1}{s!} \frac{\partial^s f(0)}{\partial\xi^{i_1}\partial\xi^{i_2}...\partial\xi^{i_s}} Tr[\mathfrak{B}(\xi)\mathfrak{U}^{+}\mathfrak{x}^{i_1}\mathfrak{x}^{i_2}...\mathfrak{x}^{i_s}\mathfrak{U}] \\ &= \sum_{s=0}^{\infty} \frac{1}{s!} \frac{\partial^s f(0)}{\partial\xi^{i_1}\partial\xi^{i_2}...\partial\xi^{i_s}} Tr[\mathfrak{B}(\xi)\mathfrak{x}'^{i_1}\mathfrak{x}'^{i_2}...\mathfrak{x}'^{i_s}] \\ &= \sum_{s=0}^{\infty} \frac{1}{s!} \frac{\partial^s f(0)}{\partial\xi^{i_1}\partial\xi^{i_2}...\partial\xi^{i_s}} u^{i_1}(\xi) \star u^{i_2}(\xi) \star ... \star u^{i_s}(\xi) \\ &\equiv f(\star u(\xi)). \end{aligned} \tag{13}$$

This expression defines a composite ★-function. The ★-product here may be substituted by the ○-product. The ○-product contains even powers of the Planck's constant in its decomposition. Consequently, the expansion around $f(u(\xi))$ contains even powers of $\hbar$. Provided that $u(\xi)$ is a linear function, $f(\star u(\xi)) = f(u(\xi))$. In the general case, the composition law of two functions is not local: $f(\star u(\xi)) \equiv f(\circ u(\xi)) \neq f(u(\xi))$.

3.2 Conservation of Moyal bracket

The antisymmetrised products of an even number of operators of canonical coordinates and momenta are *c*-numbers. These products are invariant under unitary transformations:

$$\mathfrak{U}^{+}\mathfrak{x}^{[i_1}\mathfrak{x}^{i_2}...\mathfrak{x}^{i_{2s}]}\mathfrak{U} = \mathfrak{x}^{[i_1}\mathfrak{x}^{i_2}...\mathfrak{x}^{i_{2s}]}. \tag{14}$$

In the phase space, this equation is expressed as:

$$\begin{aligned} u^{[i_1}(\xi) \star u^{i_2}(\xi) \star ... \star u^{i_{2s}]}(\xi) &= \xi^{[i_1} \star \xi^{i_2} \star ... \star \xi^{i_{2s}]} \\ &= \left(\frac{-i\hbar}{2}\right)^s \frac{1}{(2s)!} \sum_{\sigma} (-)^{\sigma} I^{i'_1 i'_2} ... I^{i'_{2s-1} i'_{2s}}. \end{aligned}$$

The summation is over all of the permutations of the indices. The sign is plus or minus, depending on whether the sequence σ is an even or odd permutation of (i_1,i_2,...,i_{2n}). The invariance of the antisymmetrised products of even numbers of operators of canonical coordinates and momenta constitutes a quantum analogue of the Poincaré theorem on the conservation of 2*n* forms in classical Hamiltonian dynamics (Krivoruchenko et al, 2006c). In particular,

$$u^i(\xi) \wedge u^j(\xi) = \xi^i \wedge \xi^j = -I^{ij}. \tag{15}$$

The real functions $u^i(\xi)$ are associated, by virtue of Equation (6), with the Hermitian operators $\mathfrak{x}'^{i}$. If $u^i(\xi)$ satisfies Equation (15), then the operators $\mathfrak{x}'^{i}$ obey the commutation relations

$$[u^i(\mathfrak{x}), u^j(\mathfrak{x})] = [\mathfrak{x}^i, \mathfrak{x}^j] = -i\hbar I^{ij}.$$

We are mainly interested in the case in which $\mathfrak{U}$ is the evolution operator. Applying a unitary transformation to the product $\mathfrak{fg}$, we obtain the function $f(\zeta)\star g(\zeta)|_{\zeta=\star u(\xi,\tau)}$ that is associated with the expression $\mathfrak{U}^{+}\mathfrak{fgU}$ and the function $f(\star u(\xi))\star g(\star u(\xi))$ associated with the expression $\left(\mathfrak{U}^{+}\mathfrak{fU}\right)\left(\mathfrak{U}^{+}\mathfrak{gU}\right)$. These operators coincide, so their symbols coincide:

$$f(\zeta)\star g(\zeta)|_{\zeta=\star u(\xi,\tau)}= f(\star u(\xi))\star g(\star u(\xi)) . \tag{16}$$

In the first case, the ★-product is calculated with respect to ζ^{i}, and it is calculated with respect to ξ^{i} in the second case. Equation (16) shows that the ★-product can be calculated in the original coordinate system prior to the change of the variables, or we can first change the variables and then compute the ★-product. Equation (16) holds separately for the symmetric and antisymmetric parts of the ★-product.

Thus, we can calculate the ★-product for any of the unitary equivalent coordinate systems. Dynamic equations constructed using the summation and multiplication of ★-functions are generally covariant under unitary transformations in phase space. The ★-product is not invariant under canonical transformations.

Example: The classical Liouville equation is covariant under canonical transformations. This equation, however, is not covariant under unitary transformations. The quantum Liouville equation, i.e., the Wigner map of the von Neumann equation, is covariant under unitary transformations and is not covariant under canonical transformations.

3.3 Phase flow generated by an evolution operator

The one-parameter family of the unitary transformations describe the evolution of quantum systems, and it is usually parameterised in the form

$$\mathfrak{U}(\tau) = \exp(-\frac{i}{\hbar}\mathfrak{H}\tau) ,$$

where $\mathfrak{H}$ is the Hamiltonian operator. The functions $u^{i}(\xi)$, defined in Equation (12), acquire the dependence on the parameter τ, so that we can write $u^{i}(\xi,\tau)$. These functions determine the quantum phase flow, which is a quantum-mechanical analogue of the phase flow in the Hamiltonian formalism of classical mechanics.

Equation (13) shows that the evolution of the symbols of the operators in the Heisenberg representation is completely determined by the functions $u^{i}(\xi,\tau)$.

We use the term "canonical transformation" in the conventional sense to refer to the coordinate transformations that preserve the Poisson bracket. The transformations preserving the Moyal bracket are unitary transformations. These transformations correspond to the action of a unitary operator in the Hilbert space. The unitary transformation in the phase space represents the quantum deformation of the canonical transformation.

Quantum characteristics arise in Heisenberg's matrix mechanics. Suppose that we have solved the evolution equations for the operators of the canonical coordinates and momenta in the Heisenberg representation. These operators evolve according to $\mathfrak{x}^{i}\rightarrow\mathfrak{x}^{i}(\tau)=\mathfrak{U}^{+}(\tau)\mathfrak{x}^{i}\mathfrak{U}(\tau)$. We use the earlier assertion that, for any operator $\mathfrak{f}$, one can find a function $f(\xi)$ through which $\mathfrak{f}$ is represented in the form $f(\mathfrak{x})$. The same operator $\mathfrak{f}$ at time τ is equal to

$$\mathfrak{f}(\tau) \equiv \mathfrak{U}^{+}(\tau)\mathfrak{f}\mathfrak{U}(\tau) = \mathfrak{U}^{+}(\tau) f(\mathfrak{x})\mathfrak{U}(\tau) = f(\mathfrak{U}^{+}(\tau)\mathfrak{x}\mathfrak{U}(\tau)) = f(\mathfrak{x}(\tau)) .$$

This equation shows that the operators of the canonical coordinates and momenta are characteristics that determine the evolution for all of the operators in $Op(L^2(\mathbb{R}^n))$. This property is fully transferred to the phase space upon deformation quantisation.

3.4 Energy conservation and composition law for trajectories

Energy conservation in the process of evolution means

$$H(\xi) = H(\star u(\xi,\tau)) , \tag{17}$$

where $H(\xi) = Tr[\mathfrak{B}(\xi)\mathfrak{H}]$ is the Hamiltonian function of the quantum system. We see that energy is conserved along quantum characteristics, but not in the geometric sense. The ★-product sign in the argument indicates the non-local nature of the conservation law.
The law of composition of the particle trajectories also has a non-local character:

$$u(\xi,\tau_1+\tau_2) = u(\star u(\xi,\tau_1),\tau_2) .$$

Such compositions, but without the ★-product, are valid for the trajectories of classical particles. The ★-product does not allow considering the motion of particles as movement along a certain trajectory in the geometrical sense.

3.5 Quantum Hamilton equations

Quantum trajectories can be found by solving Hamilton's equations, which can be written in one of four equivalent forms:

$$\begin{aligned} \frac{\partial}{\partial\tau} u^i(\xi,\tau) &= \{\zeta^i, H(\zeta)\}\,|_{\zeta=\star u(\xi,\tau)} \\ &= \zeta^i \wedge H(\zeta)|_{\zeta=\star u(\xi,\tau)} \\ &= u^i(\xi,\tau) \wedge H(\star u^i(\xi,\tau)) \\ &= u^i(\xi,\tau) \wedge H(\xi), \end{aligned} \tag{18}$$

with the initial conditions

$$u^i(\xi,0) = \xi^i .$$

These equations appear as Wigner's image of the evolution equations for operators of the canonical coordinates and momenta in the Heisenberg representation. The equivalence of the different records of the right-hand side can be verified with the help of the above-described properties of the ★-product, the rules of substitution (Equation (16)), and the condition of energy conservation (Equation (17)). Note that $\{\zeta^i, H(\zeta)\} = \zeta^i \wedge H(\zeta) = H(\zeta)^{,i}$.

The substitution $\zeta = \star u(\xi,\tau)$ in the first line of Equation (18) leads to a modification of the classical expression for the right-hand side and, correspondingly, to quantum deformation of the classical phase flow. The value of $\partial u^i(\xi,\tau)/\partial\tau$ depends on the phase space coordinate $u^i(\xi,\tau)$, as in classical mechanics, and on the infinite number of partial derivatives of $u^i(\xi,\tau)$ as a specific manifestation of the quantum non-locality.

An equivalent form of Equation (18), using a cluster expansion of the ★-exponentials, was given by Osborn & Molzahn (1995). Equation (18) was found independently by Krivoruchenko & Faessler (2006b).

3.6 Quantum Liouville equation

The functions corresponding to physical observables evolve in the Heisenberg representation according to the equation

$$f(\xi,\tau) = \mathrm{Tr}[\mathfrak{B}(\xi)\mathfrak{U}^{+}(\tau)\mathfrak{f}\mathfrak{U}(\tau)] = f(\star u(\xi,\tau),0)\,, \tag{19}$$

while the Wigner function remains constant. The evolution law can be expressed in terms of Green's function in the phase space as

$$f(\xi,\tau) = \int \frac{d^{2n}\eta}{(2\pi\hbar)^n} G(\xi,\eta,\tau) f(\eta,0)\,.$$

With the help of quantum characteristics, a compact expression for Green's function can be written as

$$G(\xi,\eta,\tau) = (2\pi\hbar)^n \delta^{2n}(\star u(\xi,\tau) - \eta)\,.$$

The function $f(\xi,\tau)$ obeys the Groenewold evolution equation (Groenewold, 1946)

$$\frac{\partial}{\partial\tau} f(\xi,\tau) = f(\xi,\tau) \wedge H(\xi)\,, \tag{20}$$

which is the Wigner map of the evolution equation of the operator $\mathfrak{f}$ in the Heisenberg representation. The right-hand side can be replaced by the equivalent expressions $f(\xi,\tau) \wedge H(\star u(\xi,\tau))$ or $f(\zeta,0) \wedge H(\zeta)|_{\zeta=\star u(\xi,\tau)}$.

The solutions to the evolution equations for the quantum characteristics and functions $f(\xi,\tau)$ can be represented as a formal power series in the parameter τ:

$$\begin{aligned} u^i(\xi,\tau) &= \sum_{s=0}^{\infty} \frac{\tau^s}{s!} (\underbrace{...((}_{s} \xi^i \wedge H(\xi)) \wedge H(\xi)) \wedge ... H(\xi)), \\ f(\star u(\xi,\tau)) &= \sum_{s=0}^{\infty} \frac{\tau^s}{s!} (\underbrace{...((}_{s} f(\xi) \wedge H(\xi)) \wedge H(\xi)) \wedge ... H(\xi)). \end{aligned} \tag{21}$$

Unitary coordinate transformations are canonical to the first order in τ (Dirac, 1930; Weyl, 1931).

For higher orders, deviations from the canonicity arise (Krivoruchenko & Faessler, 2006b). Until recently, these deviations were not well understood.[1] The infinitesimal transformations generate canonical or unitary global transformations depending on how we

[1] In the papers by B. Leaf, J. Math. Phys. **9**, 769 (1968) and T. Curtright and C. Zachos, J. Phys. A 32, 771 (1999), erroneous conclusions about the entire coincidence of classical and quantum trajectories can be found.

define the multiplication. If this is the usual dot product, then we obtain the canonical transformations. If this is the ★-product, then we obtain unitary transformations.
If an operator $\mathfrak{A}$ commutes with $\mathfrak{H}$, then its symbol is preserved in the sense of $A(\xi) = A(\star u(\xi,\tau))$. The density matrix of a stationary state commutes with $\mathfrak{H}$; therefore, in the Schrödinger representation $W_S(\xi) = W_S(\star u(\xi,-\tau))$, the Wigner function does not evolve. In the harmonic oscillator, the quantum trajectory depends linearly on ξ^i and coincides with the classical trajectory. In this case, the ★-symbol in the argument of the Wigner function can be omitted, and we can write $W_S(\xi) = W_S(u(\xi,-\tau))$ for a stationary state and $W_S(\xi,\tau) = W_S(u(\xi,-\tau),0)$ for an arbitrary state.

4. Semiclassical expansion of quantum characteristics

The methods of this section apply to the evolution problem. The most promising applications appear to be connected to many-body scattering and transport models.
The problem of evolution can by divided into two parts. First, we seek a solution to the quantum Hamilton equations. Second, we use Equation (19) to find the time-dependent symbols of the operators. The key issue is an efficient algorithm to calculate the ★-functions. These functions arise in the quantum Hamilton equations and in solving the evolution problem for functions.

4.1 Semiclassical expansion of ★-functions

We consider the semiclassical expansion of $f(\star u(\xi,\tau))$ around $f(u(\xi,\tau))$. The function $f(\xi)$ can be represented through its Fourier transform

$$f(\xi) = \int \frac{d^{2n}\eta}{(2\pi\hbar)^n} \exp\left(\frac{i}{\hbar}\eta_k \xi^k\right) f(\eta) . \tag{22}$$

Determining how to calculate $\exp(\star U)$, where $U = \frac{i}{\hbar}\eta_k u^k(\xi,\tau)$, is sufficient. With the help of Equation (13), we find

$$\exp(\star U) = \left(1 + \hbar^2 c_2 + \hbar^4 c_4 + O(\hbar^6)\right)\exp(U),$$

where

$$\begin{aligned} c_2 &= -\frac{1}{48}(2UUP^2U + 3UP^2U), \\ c_4 &= \frac{1}{23040}(90(UP^2U)UP^2U + 60(UP^2U)P^2U + 48(UUP^2U)UP^2U + 45(UP^2U)(UP^2U) \\ &\quad + 60(UUP^2U)(UP^2U) + 20(UUP^2U)(UUP^2U) + 30(\mathrm{UUP^2UP^2U}) \\ &\quad + \frac{1}{11520}(6UUUUP^4U + 45UUUP^4U + 30(UUP^4U)_1 + 40(UUP^4U)_2 + 15UP^4U). \end{aligned} \tag{23}$$

The operator P acts as follows:

$$AP^2B = A_{,kl}B^{,kl}, \quad ABP^2C = A_{,k}B_{,l}C^{,kl}, \quad AP^4B = A_{,ijkl}B^{,ijkl},$$
$$(ABP^4C)_1 = A_{,ij}B_{,kl}C^{,ijkl}, \quad (ABP^4C)_2 = A_{,i}B_{,jkl}C^{,ijkl},$$
$$ABCP^4D = A_{,i}B_{,j}C_{,kl}D^{,ijkl}, \quad ABCDP^4E = A_{,i}B_{,j}C_{,k}D_{,l}E^{,ijkl},$$

where

$$A(\xi)_{,i_1\ldots i_s} = \frac{\partial^s A(\xi)}{\partial\xi^{i_1}\ldots\partial\xi^{i_s}}, \quad A(\xi)^{,i_1\ldots i_s} = A(\xi)_{,j_1\ldots j_s} I^{j_1 i_1}\ldots I^{j_s i_s}. \tag{24}$$

Osborn & Molzahn (1995) and Gracia-Saz (2004) developed a diagram technique to calculate Weyl's symbols of composite operators for higher orders of the $\hbar$-expansion. Equation (23) from Krivoruchenko et al. (2006c), obtained with the use of MAPLE, agrees with the calculation of Gracia-Saz (2004).

The expansion of $f(\star u(\xi,\tau))$ is now straightforward. We replace $\eta_i \to -i\hbar\partial/\partial u^i$ and $U \to u^i(\xi)\partial/\partial u^i$ to obtain

$$f(\star u(\xi,\tau)) = f(u(\xi,\tau)) - \frac{\hbar^2}{24}u^i(\xi,\tau)_{,l}u^j(\xi,\tau)_{,m}u^k(\xi,\tau)^{,lm}f(u(\xi,\tau))_{,ijk}$$
$$- \frac{\hbar^2}{16}u^i(\xi,\tau)_{,kl}u^j(\xi,\tau)^{,kl}f(u(\xi,\tau))_{,ij} + O(\hbar^4).$$

The derivatives of $f(u(\xi,\tau))$ are calculated with respect to u.

As a simple application, one can find a semiclassical expansion of the Weyl symbol for the finite-temperature density matrix (Wigner, 1932):

$$W(\xi) \propto \exp(-\frac{\star H(\xi)}{T})$$
$$= \exp(-\frac{H(\xi)}{T})\left(1 + \frac{\hbar^2}{24T^3}H(\xi)_{,l}H(\xi)_{,m}H(\xi)^{,lm} - \frac{\hbar^2}{16T^2}H(\xi)_{,kl}H(\xi)^{,kl} + O(\hbar^4)\right).$$

After the replacement $1/T \to i\tau/\hbar$, the expression for the Weyl symbol of the evolution operator $\mathfrak{U}(\tau) = \exp(-\frac{i}{\hbar}\mathfrak{H}\tau)$ is derived.

The evolution operator and its Weyl's symbol are singular as $\hbar \to 0$. The expansion of the semiclassically admissible $\star$-functions starts, however, with a classical expression, which is independent of $\hbar$. The question of what happened to the singularity arises. The answer is obtained by considering the time-dependent operator $\mathfrak{f}(\tau) = \mathfrak{U}^+(\tau)\mathfrak{f}\mathfrak{U}(\tau)$. The derivatives of $\mathfrak{f}(\tau)$ of order k are expressed in terms of k commutators. Each commutator generates $\hbar$. The commutator $[\frac{i}{\hbar}\mathfrak{H},\mathfrak{f}]$ and the high-order terms $[\frac{i}{\hbar}\mathfrak{H},[\frac{i}{\hbar}\mathfrak{H},\ldots,[\frac{i}{\hbar}\mathfrak{H},\mathfrak{f}]\ldots]]$ are regular as $\hbar \to 0$. Therefore the Weyl symbol of $\mathfrak{f}(\tau)$ has a classical limit.

In transport models, the solutions to the evolution problem are based on solving systems of ODEs. We now turn our attention to the construction of the ODEs.

4.2 Semiclassical expansion of quantum Hamilton equations

Suppose we have a system of N particles. The interaction of the particles is described by some potential. The initial-state wave function is assumed to be known. Consequently, the initial-state Wigner function is known.

The first step in solving the evolution problem consists of finding the quantum characteristics, using Equation (18). We expand the solution in powers of Planck's constant

$$u^i(\xi,\tau) = \sum_{r=0}^{\infty} \hbar^{2r} u_r^i(\xi,\tau)\,. \tag{25}$$

Here, $u_0^i(\xi,\tau)$ is the classical trajectory starting at time $\tau = 0$ at $\xi^i \in \mathrm{R}^{2n}$. The initial conditions for the quantum corrections $r \geq 1$ at $\tau = 0$ are then set equal to zero. As a result,

$$\begin{aligned} u^{\mathrm{i}}{}_0(\xi,0) &= \xi^{\mathrm{i}}, \quad r = 0, \\ u_{\mathrm{r}}(\xi,0) &= 0, \quad r \geq 1. \end{aligned} \tag{26}$$

The right side of Equation (18) is the ★-function, so we use the decomposition

$$F^i(\star u(\xi,\tau)) \equiv \{\zeta^i, H(\zeta)\}\,|_{\zeta=\star u(\xi,\tau)} = \sum_{r=0}^{\infty} \hbar^{2r} F_r^i(u_0(\xi,\tau),\ldots,u_r(\xi,\tau))\,. \tag{27}$$

If the functions $u^i(\xi,\tau)$ are known, then the functions $F_r^i(u_0(\xi,\tau),\ldots,u_r(\xi,\tau))$ are completely determined. $F_r^i(u_0(\xi,\tau),\ldots,u_r(\xi,\tau))$ also depends on the derivatives of $u_0(\xi,\tau),\ldots,u_r(\xi,\tau)$ with respect to ξ^{i}. In particular,

$$\begin{aligned} F_0^i(u_0) &= F^i(u_0), \\ F_1^i(u_0,u_1) &= u_1^j(\xi,\tau) F^i(u_0)_{,j} - \frac{1}{24} u_0^j(\xi,\tau)_{,m} u_0^k(\xi,\tau)_{,n} u_0^l(\xi,\tau)^{,mn} F^i(u_0)_{,jkl} \\ &\quad - \frac{1}{16} u_0^j(\xi,\tau)_{,lm} u_0^k(\xi,\tau)^{,lm} F^i(u_0)_{,jk}. \end{aligned} \tag{28}$$

4.3 Jacobi fields

The second line of Equation (28) contains the first- and second-order derivatives of $u^i{}_0(\xi,\tau)$ with respect to ξ^{i}. Therefore, we should monitor the evolution of the trajectories and their derivatives with respect to the initial coordinates

$$J^{\mathrm{i}}{}_{\mathrm{r},\mathrm{j}_1\ldots\mathrm{j}_{\mathrm{t}}}(\xi,\tau) = \frac{\partial^{\mathrm{t}} u_{\mathrm{r}}{}^{\mathrm{i}}(\xi,\tau)}{\partial\xi^{\mathrm{j}_1}\ldots\partial\xi^{\mathrm{j}_{\mathrm{t}}}}\,. \tag{29}$$

These values determine the decomposition of $f(\star u(\xi,\tau))$ and determine the high-order quantum corrections to the phase-space trajectories.

We call these values "Jacobi fields". This term is adopted in Hamiltonian mechanics for the first-order derivatives of $u^i{}_0(\xi,\tau)$, that determine the stability of the systems. The value of

$J^{\mathrm{i}}{}_{0,\mathrm{j}}(\xi,\tau)=\partial u^{\mathrm{i}}{}_0(\xi,\tau)/\partial\xi^{\mathrm{j}}$ is known as the Jacobi matrix, and its determinant is called the Jacobian. In quantum mechanics, the derivatives of higher orders are involved. In the following, the number *r* is called the order of the Jacobi field, the number of lower indices *t* is called the degree of Jacobi field.
In the notation of Equation (29), the first pair of evolution equations becomes

$$\begin{aligned}
&\frac{\partial}{\partial\tau}u_0^i(\xi,\tau)=F^i(u_0),\\
&\frac{\partial}{\partial\tau}u_1^i(\xi,\tau)=u_1^j(\xi,\tau)F^i(u_0)_{,j}-\frac{1}{24}J^j_{0,m}(\xi,\tau)J^k_{0,n}(\xi,\tau)J_0^{l,mn}(\xi,\tau)F^i(u_0)_{,jkl}\\
&\qquad\qquad -\frac{1}{16}J^j_{0,lm}(\xi,\tau)J_0^{k,lm}(\xi,\tau)F^i(u_0)_{,jk}.
\end{aligned}\tag{30}$$

In the first line, the classical Hamilton equations are recognisable. The second and third lines determine the lowest-order quantum correction to the classical trajectory.
The system in Equation (30) is not yet closed. It needs to be supplemented by the equations of motion of $J^i_{0,j}(\xi,\tau)$ and $J^i_{0,jk}(\xi,\tau)$. These equations are obtained by taking the first- and second-order derivatives of the first equation with respect to the initial coordinates:

$$\begin{aligned}
&\frac{\partial}{\partial\tau}J^i_{0,j}(\xi,\tau)=F^i(u_0)_{,\mathrm{k}}J^k_{0,j}(\xi,\tau),\\
&\frac{\partial}{\partial\tau}J^i_{0,jk}(\xi,\tau)=F^i(u_0)_{,lm}J^l_{0,j}(\xi,\tau)J^m_{0,k}(\xi,\tau)+F^i(u_0)_{,l}J^l_{0,jk}(\xi,\tau).
\end{aligned}\tag{31}$$

Differentiating Equation (26) on ξ^{i}, we obtain the initial conditions for the Jacobi fields. In general,

$$\begin{aligned}
&J^{\mathrm{i}}{}_{\mathrm{r},\mathrm{j}}(\xi,0)=\delta^{\mathrm{i}}{}_{\mathrm{j}},\quad r=0,\\
&J^{\mathrm{i}}{}_{\mathrm{r},\mathrm{j}_1\ldots\mathrm{j}_\mathrm{t}}(\xi,0)=0,\quad r=0,t\geq 2\ \text{ or }\ r\geq 1,t\geq 1.
\end{aligned}\tag{32}$$

The coordinates ξ^{i} enter Equations (30) and (31) as parameters. Thus, we meet a typical case of first-order ODEs, for the variables $u_0^i(\xi,\tau)$, $u_1^i(\xi,\tau)$, $J^i_{0,j}(\xi,\tau)$ and $J^i_{0,jk}(\xi,\tau)$ with the initial conditions given in Equations (26) and (32).
Here, we show a proof of the statement (Krivoruchenko et al., 2006c) that, in any fixed order of $\hbar$, we still have a finite system of first-order ODEs for the variables $u_r^i(\xi,\tau)$ and the associated Jacobi fields.

4.4 Reduction of the quantum Hamilton equations to a finite system of first-order ODEs

Let us consider the effect of the ★-product. According to Equation (8), each power of $\hbar$ is accompanied by differentiation. For order $\hbar^{2s}$, expansion of $F^i(\star u(\xi,\tau))$ contains the derivatives of order 2*s* at most. Therefore, the Jacobi fields have the highest degree 2*s* ($1\leq t\leq 2\mathrm{s}$).

This assertion can be strengthened. The number of indices t in reality also depends on the order r of the Jacobi fields $J^{i}_{r,j_1\ldots j_t}(\xi,\tau)$. In the expansion in Equation (13), the derivatives of the trajectory $u^{i}_{0}(\xi,\tau)$ of order $\le 2s$, the derivatives of the trajectory $u^{i}_{1}(\xi,\tau)$ of order $\le 2s - 2$, ..., and the derivatives of the trajectory $u^{i}_{s-1}(\xi,\tau)$ of order ≤ 2 survive among all of the derivatives of orders $\le 2s$. The highest-order correction $u^{i}_{s}(\xi,\tau)$ has no derivatives. Thus, the maximum number of lower indices of $J^{i}_{r,j_1\ldots j_t}(\xi,\tau)$, involved in the expansion to order $\hbar^{2s}$, depends on r and equals $2s-2r$ $(1\le t\le 2s-2r)$.

Let us consider in more detail the equations of evolution to order $\hbar^{2s}$ for a fixed $r\le s$.

The first part of the ODE system can be written as

$$\frac{\partial}{\partial\tau}u^{i}_{r} = F^{i}_{r}(u_0,u_1,\ldots,u_r,J_0,J_1,\ldots,J_{r-1}), \tag{33}$$

where the index r takes the values 1 ... s. In the argument for the function on the right-hand side, we dropped the indices of the trajectories and the Jacobi fields. Note that the time derivative depends on the Jacobi fields of order $r - 1$ at most.

The higher-order corrections depend on the lower-order corrections. In the right-hand side of Equation (33), the Jacobi fields have the following degrees: $J^{i}_{0,j_1\ldots j_t}(\xi,\tau)$ - not more than $2r$, $J^{i}_{1,j_1\ldots j_t}(\xi,\tau)$ - not more than $2r$ - 2, and so on. In the highest order $\hbar^{2s}$-term, the maximum degree of $J^{i}_{r-1,j_1\ldots j_t}(\xi,\tau)$, entering the right-hand side, is equal to $2s - 2r + 2$. The functions $F^{i}_{r}(u_0,u_1,\ldots,u_r,J_0,J_1,\ldots,J_{r-1})$ do not depend on the variables with r' for $r < r'$. Equation (33) clearly allows the determination of the trajectories $u_0(\xi,\tau),\ldots,u_s(\xi,\tau)$, provided that the Jacobi fields are known.

We now supplement the resulting system (Equation (33)) with the equations of evolution of the Jacobi fields.

Consider first Equation (33) for $r = 0$, i.e., the classical Hamilton equations. The right-hand side depends only on $u^{i}_{0}(\xi,\tau)$. Differentiating this equation from one to $2s$ times, we obtain evolution equations for the zero-order Jacobi fields of degrees $t = 1 \ldots 2s$. Equation (33) for $r = 0$ and the $2s$ of these equations form a closed system of ODEs whose solutions are well defined.

As a next step we consider Equations (33) for $r = 1$. The right-hand side depends on the trajectories $u^{i}_{0}(\xi,\tau)$ and $u^{i}_{1}(\xi,\tau)$ and the Jacobi fields $J^{i}_{0,j_1\ldots j_t}(\xi,\tau)$ for $t = 1, 2$. Differentiating this equation with respect to ξ^{i} from one to $2s$ - 2 times, we obtain the evolution equations for the first-order Jacobi fields $J^{i}_{1,j_1\ldots j_t}(\xi,\tau)$ of degrees $t = 1 \ldots 2s$ - 2. After differentiating, the right-hand side depends on the Jacobi fields $J^{i}_{0,j_1\ldots j_t}(\xi,\tau)$ of degrees $\le 2s$ $(= 2 + 2s - 2)$, while the Jacobi fields $J^{i}_{1,j_1\ldots j_t}(\xi,\tau)$ arising from differentiating $u_1{}^{i}(\xi,\tau)$ do not have degrees that are higher than the number of derivatives taken, i.e., not above $2s$ - 2. Thus, to determine $J^{i}_{1,j_1\ldots j_t}(\xi,\tau)$, additional information about the zero-order Jacobi fields is not required. From these equations, we can find $u^{i}_{1}(\xi,\tau)$ and $J^{i}_{1,j_1\ldots j_t}(\xi,\tau)$ of degrees $t \le 2s$ - 2.

Further arguments are fairly obvious. We are moving in the direction of increasing the order of Jacobi fields. Consider the general case. We take the derivatives of Equation (33) from one to $2s - 2r$ times and obtain the evolution equations for the order-r Jacobi fields of degrees $t = 1 \ldots 2s - 2r$.

$$\frac{\partial}{\partial\tau} J^{i}{}_{\mathrm{r},\mathrm{j}_1\ldots\mathrm{j}_\mathrm{t}}(\xi,\tau) = G^{i}_{\mathrm{r}}(u_0, u_1, \ldots, u_\mathrm{r}, J_0, J_1, \ldots, J_r) \,. \tag{34}$$

Consider the right-hand side of Equation (33). It depends on the Jacobi fields $J^{\mathrm{i}}{}_{0,\mathrm{j}_1\ldots\mathrm{j}_\mathrm{t}}(\xi,\tau)$ of degrees $\leq 2r$. After differentiating one to $2s - 2r$ times, a dependence on the Jacobi fields $J^{\mathrm{i}}{}_{0,\mathrm{j}_1\ldots\mathrm{j}_\mathrm{t}}(\xi,\tau)$ of degrees $\leq 2s$ $(= 2r + 2s - 2r)$ is acquired. Consequently, for any r, the right-hand side of Equations (34) depends on the zero-order Jacobi fields of degree $2s$ at most. Furthermore, the right-hand side of Equation (33) depends on the first-order Jacobi fields of degrees $\leq 2r - 2$. After the differentiation, a dependence on the first-order Jacobi fields of degrees $\leq 2s - 2$ $(= 2r - 2 + 2s - 2r)$ occurs in Equation (34). The upper value is also independent of r.

For a fixed r, Equation (33) depends on the h–order Jacobi fields ($h < r$) of degrees $t \leq 2r - 2h$. Thus, the evolution equations for the Jacobi fields $J^{\mathrm{i}}{}_{\mathrm{r},\mathrm{j}_1\ldots\mathrm{j}_\mathrm{t}}(\xi,\tau)$ of degrees $t = 1 \ldots 2s - 2r$ contain the trajectory functions $u_0(\xi,\tau), \ldots, u_\mathrm{r}(\xi,\tau)$ and the Jacobi fields $J^{\mathrm{i}}{}_{\mathrm{h},\mathrm{j}_1\ldots\mathrm{j}_\mathrm{t}}(\xi,\tau)$ of orders $h = 0 \ldots r$ and degrees $t = 1 \ldots 2s - 2h$.

The truncation of the expansion at any s provides us with the complete system of first-order ODEs for the trajectories and the Jacobi fields. The system is determined by Equations (33) and (34) with the initial conditions given in Equations (26) and (32).

In terms of the mathematical induction, the above arguments indicate that Equations (33) and (34) are sufficient to determine the trajectories and the Jacobi fields for some r, provided that the trajectories and the Jacobi fields of lower orders $< r$ are determined. We have seen that this is true for $r = 0, 1$; therefore, it is true for all r.

4.5 Perspectives of transport models

We show the lists of the dynamical variables and the numbers of the independent degrees of freedom in the decomposition of the quantum Hamilton equations up to the fourth order in $\hbar$:

$$\begin{aligned}
\hbar^0 :\ & u^i_0(\xi,\tau) \\
& \to 2n \\
\hbar^2 :\ & u^i_0(\xi,\tau) \,\&\, J^i_{0,j}(\xi,\tau), J^i_{0,jk}(\xi,\tau) \\
& u^i_1(\xi,\tau) \\
& \to 2n(2 + 3n + 2n^2) \\
\hbar^4 :\ & u^i_0(\xi,\tau) \,\&\, J^i_{0,j}(\xi,\tau), J^i_{0,jk}(\xi,\tau), J^i_{0,jkl}(\xi,\tau), J^i_{0,jklm}(\xi,\tau) \\
& u^i_1(\xi,\tau) \,\&\, J^i_{1,j}(\xi,\tau), J^i_{1,jk}(\xi,\tau) \\
& u^i_2(\xi,\tau) \\
& \to n(18 + 43n + 47n^2 + 20n^3 + 4n^4) / 3
\end{aligned}$$

Here, $n = 3N$, where N is the number of particles. In higher orders, lists of each line are extended by two units at the expense of the Jacobi fields of higher degrees. A new line containing the next order correction to quantum trajectory is also added.

Table 1 shows the number of dynamic degrees of freedom for orders $\hbar^0$, $\hbar^2$, and $\hbar^4$ for the potential scattering of a proton (a nucleus) with nuclei. The spin degrees of freedom of the nucleons are disregarded.

Reaction	$\hbar^0$	$\hbar^2$	$\hbar^4$
^{2}H + ^{2}H	24	7 824	499 224
^{9}Be + ^{9}Be	108	647 568	671 416 128
p + ^{238}U	1 434	1 477 494 654	254 426 548 725 264
^{238}U + ^{238}U	2 856	11 660 059 824	7 945 116 177 770 184

Table 1. The number of dynamical degrees of freedom for the potential scattering of protons (nuclei) with nuclei of orders $\hbar^0$, $\hbar^2$, and $\hbar^4$.

As the number of Jacobi fields involved in the dynamics increases rapidly with the increasing order of the expansion, limitations due to the semiclassical expansion being restricted by computing power must be considered. Since the mid 1960's, the performance of computers has doubled approximately every 15 months. Due to technical peculiarities of processor manufacturing, this regime is expected to continue for 5 - 15 years.

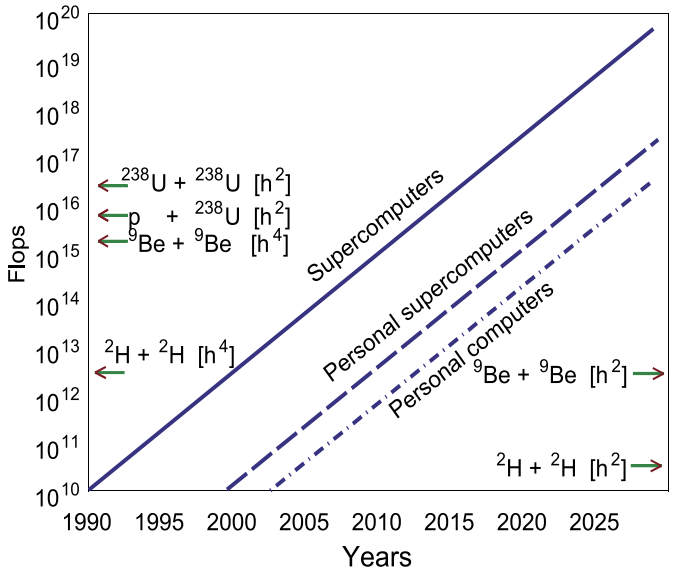


Fig. 1. Computer power in flops (floating points operations) required to simulate the proton-nucleus and nucleus-nucleus collisions listed in Table 1 as compared with the growth in power of supercomputers (solid line), personal supercomputers (dashed line) and personal computers (dot-dashed line) starting from 1990. The orders $\hbar^2$ and $\hbar^4$ of the simulations are shown in square brackets.

The theoretical foundations of the transport models used to simulate heavy-ion collisions were created in the late 1980s - early 1990s (Sorge et al., 1989; Aichelin, 1991; Faessler, 1992;

Blaettel et al., 1993). Many uncertainties exist in the estimates of the times of the simulations. We assume that simulations of ^{238}U + ^{238}U collisions with a supercomputer required about one week in 1990. The calculations are restricted by zero order in the Planck's constant; hence, Jacobi fields are not involved. We are dealing with 2856 dynamical degrees of freedom, as indicated in Table 1, plus the spin degrees of freedom of the nucleons and, potentially, the number of meson degrees of freedom which depends on the beam energy. The calculation time appears to grow linearly with the number of degrees of freedom. Hence, we obtain the estimates shown in Figure 1.

^{2}H + ^{2}H reactions of order $\hbar^4$ and ^{9}Be + ^{9}Be reactions of order $\hbar^2$ could be simulated with supercomputers in 1999. ^{9}Be + ^{9}Be reactions of order $\hbar^4$, p + ^{238}U and ^{238}U + ^{238}U reactions of order $\hbar^2$ should be simulated with supercomputers by 2011 - 2016. A delay will occur, if computing power is limited to the use of personal supercomputers and computers.

Since the early 1990s, computing power has increased by about *five* orders of magnitude. This dramatic rise in computing power makes it possible to include Jacobi fields in the collision dynamics, to extend beyond the purely classical treatment of phase-space trajectories, currently adopted in all of the transport models.

5. Averaging over the Wigner function using the Monte Carlo method

The reduction of the evolution problem to the search for quantum trajectories and the associated Jacobi fields makes it possible to calculate averages using the Monte Carlo method. The problem becomes a task of statistical physics with the modified rules of calculations of probabilities and average values.

The average of the observable associated with an operator $\mathfrak{f}$ at time τ can be found from the evolution equation (Equation (19)) for the associated function $f(\xi)$. We then use the decomposition (Equation (13)). In the Heisenberg representation, the average is determined by the integral of $f(\xi)$ multiplied by the Wigner function, given at the initial time $W(\xi,0) \equiv W(\xi)$

$$\langle f(\xi,\tau)\rangle = \int \frac{d^{2n}\xi}{(2\pi\hbar)^n} f(\star u(\xi,\tau)) W(\xi)\,. \tag{35}$$

We partition the phase space R^{2n} into two regions Ω_+ and Ω_-, so that $\Omega_+ \cup \Omega_- = \mathrm{R}^{2n}$, in which the Wigner function is positive and negative, respectively. Therefore, we have $W(\xi) = W_+(\xi) - W_-(\xi)$ with $W_\pm(\xi) \geq 0$ in $\Omega_\pm$. Outside these regions, the functions $W_\pm(\xi)$ vanish. As a next step, we generate events in $\Omega_\pm$ (i.e., we select the points $\xi^{\mathrm{i}} \in \Omega_\pm \subset \mathrm{R}^{2n}$), distributed according to the normalised probability densities $W_\pm(\xi) / W_\pm$, with

$$W_\pm = \int \frac{d^{2n}\xi}{(2\pi\hbar)^n} W_\pm(\xi)\,.$$

First consider the region Ω_+. We generate $2n + 1$ numbers $(\xi^{\mathrm{i}}, \gamma)$ with the values of ξ^{i} uniformly distributed in Ω_+ and the value of γ uniformly distributed in the interval

$(0, W_{+\max} / W_+)$, where $W_{+\max} = \max_{\xi}(W_+(\xi))$. If the joint probability density of the variables $(\xi^{\mathrm{i}}, \gamma)$ is given by $\theta(W_+(\xi) / W_+ - \gamma)$, then the marginal probability density of ξ^{i} equals

$$W_+(\xi) / W_+ = \int_0^{W_{+\max}/W_+} \theta(W_+(\xi) / W_+ - \gamma) d\gamma .$$

To obtain a sample $\left\{\xi^{\mathrm{i}}_{\mathrm{a}}\right\}$ of events $a = 1,\ldots,N_+$ distributed with the probability density $W_+(\xi) / W_+$, it is straightforward to discard those generated numbers that do not satisfy the condition $W_+(\xi) / W_+ > \gamma$.

Suppose we generated the numbers $(\xi^{\mathrm{i}}, \gamma)$ at some step. If $W_+(\xi) / W_+ > \gamma$ holds, then we shift the number N_+ of successful tests by one unit and assign $\xi^{\mathrm{i}}_{+\mathrm{a}} = \xi^{\mathrm{i}}$ for $a = N_+$. Next, we calculate the quantum trajectory $u^i(\xi_{+a}, \tau)$ and the associated Jacobi fields, find the value of $f(\star u(\xi_{+a}, \tau))$ in the required order of the $\hbar$-expansion and store the information. If the inequality $W_+(\xi) / W_+ > \gamma$ is not satisfied, then the event is simply discarded, and we generate the next set of numbers $(\xi^{\mathrm{i}}, \gamma)$. The saved values $\left\{\xi^{\mathrm{i}}_{\mathrm{a}}\right\}$ are distributed with the probability density $W_+(\xi) / W_+$. A similar procedure applies to the Ω_- region.

Suppose we have generated N_+ and N_- successful events $\xi_{\pm a} \in \Omega_{\pm}$. To find the average value of $f(\xi, \tau)$, the values $f(\star u(\xi_{\pm a}, \tau))$ should be multiplied by $W_{\pm}$, divided by the number of successful tests and summed to give the following:

$$\left\langle f(\xi, \tau) \right\rangle \approx \frac{W_+}{N_+} \sum_{\mathrm{a}=1}^{\mathrm{N}_+} f(\star u(\xi_{+a}, \tau)) - \frac{W_-}{N_-} \sum_{\mathrm{a}=1}^{\mathrm{N}_-} f(\star u(\xi_{-a}, \tau)) . \tag{36}$$

This equation completes the reduction of the quantum evolution problem to the problem of calculating the statistical averages over an ensemble of quantum characteristics and associated Jacobi fields.

6. The scattering problem

In the scattering problem, elementary particles and bound states are considered on an equal footing. We fix the in- and out- scattering states at $\tau = \pm\infty$ and define clusters α of elementary particles and bound states with momenta $\boldsymbol{p}_\alpha{}'$ and $\boldsymbol{p}_\alpha{}''$ in the initial and final asymptotic states, respectively. Wigner functions have the form

$$\begin{aligned} W_{in}(\xi) &= \prod_a (2\pi\hbar)^3 \delta(\boldsymbol{p}_a{}' - \sum_{i\in a} \boldsymbol{p}_i) W_a{}'(\xi_a), \\ W_{out}(\xi) &= \prod_a (2\pi\hbar)^3 \delta(\boldsymbol{p}_a{}'' - \sum_{i\in a} \boldsymbol{p}_i) W_a{}''(\xi_a). \end{aligned} \tag{37}$$

Here, $\xi^i = (q^1,\ldots,q^n, p_1,\ldots,p_n)$, ξ_α are the variables of the particles in cluster α. Each cluster contains $\sum_{\mathrm{i}\in\alpha} 1$ particles, in total $\sum_\alpha \sum_{\mathrm{i}\in\alpha} 1 = \mathrm{N}$. A similar situation holds for the partition of particles in the final state.

The Wigner functions $W_{in}(\xi)$ and $W_{out}(\xi)$ are constructed as products of the asymptotic Wigner functions of non-interacting elementary particles and bound states. On the right

sides of Equation (37), $W_\alpha'(\xi_\alpha) = W_\alpha''(\xi_\alpha) = 1$ for elementary particles, while Wigner functions of bound states must be constructed on the basis of the wave functions of the bound states of the clusters.

The transition probability is the square modulus of the *S*-matrix element $w_{fi} = |\langle out \,|\, in \rangle|^2$. In terms of the Wigner function,

$$w_{fi} = \int \frac{d^{2n}\xi}{(2\pi\hbar)^n} W_{out}(\star u(\xi,\tau)) W_{in}(\xi) \,, \tag{38}$$

where $n = 3N$. The technique described in the previous sections is fully applicable to Equation (38), and it applies to the scattering problem.

7. Conclusion

In this chapter, we discussed the basic properties of the formalism of deformation quantisation and its applications to the description of the evolution of many-body systems in terms of the expansion in powers of the Planck's constant.

We described the dynamics of quantum systems in phase space. In our presentation, a special role is assigned to quantum trajectories $u^i(\xi,\tau)$, which appear as the Weyl symbols of the operators of the canonical coordinates and momenta in the Heisenberg representation. These trajectories differ from the classical trajectories and from the de Broglie - Bohm trajectories. The transformation of the coordinate system in phase space, associated with quantum trajectories, preserves the Moyal bracket and does not preserve the Poisson bracket. In this sense, the quantum trajectories and phase flow, which they define, can be regarded as a quantum deformation of classical trajectories and phase flow in the formalism of classical Hamiltonian mechanics. The quantum trajectories satisfy the quantum Hamilton equations that are infinite-order PDEs.

Deformation quantisation preserves many features of classical Hamiltonian mechanics. The classical Hamilton's equations are the characteristics equations of the classical Liouville equation for particle distributions in phase space. Accordingly, the solutions of the Hamilton equations contain all of the dynamic information needed to determine the time dependence of all of the observables, including the distribution function.

The same situation occurs in quantum physics. Solutions to the quantum Hamilton equations define quantum trajectories, which possess all of the properties of characteristics. As a rule, characteristics satisfy a system of first-order ODEs, for example, the system of Hamilton's equations. Characteristics are used further to construct solutions of first-order PDEs, such as the classical Liouville equation. The peculiarity of quantum mechanics is that quantum trajectories obey infinite-order PDEs, and they also solve evolution equations that are infinite-order PDEs.

In the Heisenberg representation, the evolution of the Weyl symbol of operator can be written as

$$f(\xi,\tau) = f(\star u(\xi,\tau), 0) \,.$$

The relationships of fundamental interest, such as the one shown above, are formulated in terms of the ★-functions that are not local and form a special class of functionals. To date,

no effective methods exist for calculating ★-functions, with the possible exception of the expansion in powers of the Planck's constant. We have outlined recipes to eliminate the ★-symbols from the arguments of composite functions using the semiclassical expansion.
For any fixed order of the semiclassical expansion, the quantum characteristics are constructed by solving a finite system of first-order ODEs. This important circumstance makes it possible to approach the problem of quantum evolution of complex systems using numerically efficient ODE integrators. The evolution problem thereby reduces to a statistical-mechanics problem of constructing an ensemble of the quantum characteristics and the associated Jacobi fields. After constructing the quantum characteristics, the physical observables can be found without further recourse to quantum dynamics.
A clear gap exists between the classical dynamics of a particle and its quantum dynamics. In the first case, we are dealing with a finite number of degrees of freedom. In the second case, we are dealing *per se* with field theory and an infinite number of dynamical degrees of freedom. We see that this gap is filled with the Jacobi fields of higher orders. By increasing the order of $\hbar$ -expansion, the number of Jacobi fields is growing rapidly. This provides, in principle, a smooth transition from classical dynamics to quantum dynamics and from mechanics to field theory.
Quantum characteristics are useful for calculating the evolution of complex quantum systems - atoms, molecules and nuclei. The main advantage of deformation quantisation is its proximity to the classical picture of evolution in phase space. Specific quantum effects, such as coherence and non-localities, appear due to the increase in the number of dynamical degrees of freedom: Jacobi fields. The method of quantum characteristics allows for the consistent inclusion of non-localities and coherence in the transport models.

8. Acknowledgment

This work was supported by Grant No. 09-02-91341 of the Russian Foundation for Basic Research.

9. References

Aichelin, J., (1991). 'Quantum' molecular dynamics: A Dynamical microscopic n body approach to investigate fragment formation and the nuclear equation of state in heavy ion collisions. *Physics Reports*, Vol. 202 (April 1991), pp. 233 - 360. ISSN 0370-1573.

Balazs, N. L., Jennings, B. K. (1984). Wigner's function and other distribution functions in mock phase spaces. *Physics Reports*, Vol. 104, No. 6 (February 1984), pp. 347-391. ISSN 0370-1573.

Baker, G. A. Jr. (1958). Formulation of Quantum Mechanics Based on the Quasi-Probability Distribution Induced on Phase Space. *Physical Review*, Vol. 109, No. 6 (March 1958), pp. 2198 - 2206. ISSN 0031-899X.

Banin, A. T., Buchbinder, I. L., Pletnev, N. G. (2001). Low-Energy Effective Action of N=2 Gauge Multiplet Induced by Hypermultiplet Matter. *Nuclear Physics B*, Vol. 598, pp. 371-399. ISSN 0556-2813.

Bayen, F., Flato, M., Fronsdal, C., Lichnerowicz, A., Sternheimer, D. (1978a). Deformation theory and quantization. I. Deformations of symplectic structures. *Annals of Physics (N.Y.),* Vol. 111, No. 1 (March 1978), pp. 61-110. ISSN 0003-4916.

Bayen, F., Flato, M., Fronsdal, C., Lichnerowicz, A., Sternheimer, D. (1978b). Deformation theory and quantization. II. Physical applications. *Annals of Physics (N.Y.)*, Vol. 111, No. 1 (March 1978), pp. 111-151. ISSN 0003-4916.

Blaettel, B., Koch, V., Mosel, U. (1993). Transport theoretical analysis of relativistic heavy ion collisions. *Reports on Progress in Physics*, Vol. 56, No. 1 (June 1993), pp. 1-62. ISSN 0034-4885.

Cargo, M., Gracia-Saz, A., Littlejohn, R. G., Reinsch, M. W. and de M Rios, P. (2005). Quantum normal forms, Moyal star product and Bohr–Sommerfeld approximation. *Journal of Physics A*, Vol. 38 (February 2005), pp. 1977–2004. ISSN 1751-8113.

Carruthers P., Zachariasen, F. (1983). Quantum Collision Theory with Phase Space Distribution Functions. *Review of Modern Physics,* Vol. 55, pp. 245–285. ISSN 00346861.

Dirac, P. A. M. (1930). The Principles of Quantum Mechanics, (Oxford: Clarendon Press).

Faessler, A., (1992). Description of heavy ion collisions with medium dependent forces. *Progress in Particle and Nuclear Physics*, Vol. 30 (June 1993), pp. 229 - 246.

Feldmeier, H., Schnack, J. (1997). Fermionic molecular dynamics. *Progress in Particle and Nuclear Physics*, Vol. 39 (March 1997), pp. 393 - 442. ISSN 0146-6410.

Gracia-Saz, A., (2004). The symbol of a function of an operator. *arXiv: math.QA/0411163*. Available from http://arxiv.org/abs/math/0411163.

Groenewold, H. J. (1946). On the Principles of elementary quantum mechanics. *Physica,* Vol. 12 (1946), pp. 405 - 460.

Hillery, M. , O'Connell, R. F., Scully, M. O. , Wigner, E. P. (1983). Distribution functions in physics: Fundamentals. *Physics Reports*, Vol. 106, No. 3 (April 1984), pp. 121-167. ISSN 0370-1573.

Sorge, H., Stöcker, H., Greiner, W. (1989). Poincaré invariant Hamiltonian dynamics: Modelling multi-hadronic interactions in a phase space approach. *Annals of Physics* (*N.Y.*), Vol. 191 (January 1989), pp. 266 - 306. ISSN 0003-4916.

Stratonovich, R. L. (1957). On distributions in representation space. *Soviet Physics JETP,* Vol. 4 (1957), pp. 891–898. ISSN 0038-5646.

Kohler, H. S., (1995). Memory and correlation effects in nuclear collisions. *Physical Review,* Vol. C51 (June 1995), pp. 3232-3239. ISSN 0556-2813.

Karasev, M. V., Maslov, V. P. (1993), Nonlinear Poisson Brackets: Geometry and Quantization, Translations of Mathematical Monographs No. 119 (American Mathematical Society, Providence), ISSN 0065-9282.

Krivoruchenko, M. I., Raduta, A. A., Faessler, A. (2006a). Quantum deformation of the Dirac bracket. *Physical Review*, Vol. D73, No. 2 (Jauary 2006), pp. 025008. ISSN 1550-7998.

Krivoruchenko, M. I., Faessler, A. (2006b). Weyl's symbols of Heisenberg operators of canonical coordinates and momenta as quantum characteristics. *e-Print: quant-ph/0604075* (April 2006). Available from http://arxiv.org/abs/quant-ph/0604075; *Journal of Mathematical Physics*, Vol. 48 (May 2007), pp. 052107. ISSN 0022-2488.

Krivoruchenko, M. I., Fuchs, C., Faessler, A. (2006c). Deformation quantization and semiclassical expansion in many-body potential scattering problem. *e-Print: arXiv:nucl-th/0605015 [nucl-th]* (May 2006). Available from http://arxiv.org/abs/arXiv:nucl-th/0605015; *Annalen der Physik*, Vol. 16 (August 2007), pp. 587 - 614. ISSN 1521-3889.

Krivoruchenko, M. I., Martemyanov, B. V., Fuchs, C. (2007). Comment on 'Dynamics of nuclear fluid. VIII. Time-dependent Hartree-Fock approximation from a classical point of view'. *Physical Review*, Vol. C76 (November 2007), pp. 059801. ISSN 0556-2813.

McQuarrie, B. R., Osborn, T. A., Tabisz, G. C. (1998). Semiclassical Moyal quantum mechanics for atomic systems. *Physical Review A*, Vol. 58, No. 4, (October 1998), pp. 2944 - 2961. ISSN 1050-2947.

Moyal, J. E. (1949). Quantum mechanics as a statistical theory. *Proc. Cambridge Phil. Soc.* Vol. 45 (1949), pp. 99-124. ISSN 0305-0041.

Osborn, T. A., Molzahn, F. H. (1995). Moyal Quantum Mechanics: The Semiclassical Heisenberg Dynamics. *Annals of Physics* (N.Y.), Vol. 241 (January 1995), pp. 79-127. ISSN 0003-4916.

Pletnev, N. G., Banin, A. T. (1999). Covariant technique of derivative expansion of the one-loop effective action. *Physical Review D*, Vol. 60 (October 1999), 105017. ISSN 1550-7998.

Weyl, H. (1927). Quantenmechanik und Gruppentheorie, *Zeitschrift für Physik A,* Vol. 46, No. 1 – 2 (October 1927), pp. 1 – 46. ISSN 0340-2347.

Weyl, H. (1928). Gruppentheorie und Quantenmechanik (Leipzig: Hirzel).

Weyl, H. (1931). The Theory of Groups and Quantum Mechanics (Dover Publications, New York Inc.).

Wigner, E. P. (1932). On the quantum correction for thermodynamic equilibrium, *Physical Review*, Vol. 40 (March 1932), pp. 749–759. ISSN 0031-899X.

4

Synergy Between First-Principles Computation and Experiment in Study of Earth Science

Shigeaki Ono
Japan Agency for Marine-Earth Science and Technology
Japan

1. Introduction

Atomic-scale modeling of materials based on first-principles quantum mechanics is playing an important role in a broad range of sciences, such as chemistry, biology, and geophysics. Part of our aim in this article is to introduce the usefulness of atomic-scale modeling of materials to the geoscience field. The structure, dynamics, and evolution of the Earth and other planets depend on processes that take place deep in their interiors. However, these interiors are inaccessible to direct observation due to extremely high pressures and temperatures. Laboratory experiments at high pressures and temperatures can provide important information on the physical properties of materials that constitute the Earth's and planets' interiors. However, recently, a new tool for materials modeling based on first-principles quantum mechanics has come into use for probing the properties of planetary interiors. This method has advanced to the point where it can provide reliable data for conditions of extreme high pressure and high temperature that experiments cannot achieve. High-pressure experimental and materials modeling studies regarding the physical properties of materials were generally presented in separate papers and often with a publication gap of a couple of years. A recent approach using both high-pressure experimental data and first-principles materials modeling was successful in discovering new minerals and predicting the physical properties of many materials at high pressures and temperatures.

This article provides examples of the synergy between first-principles computation and high-pressure experiments. First, the argument for the determination of equations of state for materials used as pressure calibrants in high-pressure experiments is discussed. In this discussion, the advantages and disadvantages of first-principles materials modeling or high-pressure experiments are described. It is hoped that the reader can understand the reliability of both the first-principles materials modeling and the high-pressure experiment. Second, some approaches to the discovery of new materials and the exploration of their physical properties at extremely high pressures and temperatures are described.

2. Pressure scale

It is particularly important to know precisely the temperature and pressure conditions of a system in all scientific fields. However, it is often difficult to determine the temperature and pressure when they are extremely higher than ambient conditions. The uncertainty of the

pressure scale at extremely high temperatures has recently become an open question in earth science. Although "Pressure" is an important index in all science fields, the problem of the uncertainty of a standard scale for pressure is debated in the small earth science community. Extreme pressures and temperatures cause various problems for the earth science discipline. It is known that the Earth's deep interior reaches high temperatures and high pressures. Therefore, when we want to know the behavior of materials that constitute the Earth's interior, it is necessary to investigate materials at temperatures of up to ~6000 K and pressures of up to ~400 GPa. This is greatly different from other scientific fields. For instance, studies at low temperatures are performed quite often in the field of physics, because generally it is easy to observe the essence of the physical phenomenon. Furthermore, the Earth's interior is probed by observing seismic wave velocities, and a detailed 3D map can be used to investigate it. This is a similar to obtaining a X-ray computed tomography (CT) scan of the human body. This approach obtains the change in the elastic properties of the Earth's interior as a 3D map. Thus, we can determine the elastic properties of the Earth's interior as a function of depth (equal to pressure). Our next aim is to know what type of material corresponds to the elastic property determined by the mapping of seismic wave velocities. However, earth scientists face a problem at this step. It is thought that some vast changes in seismic velocities have taken place in the Earth's interior because of phase transitions of materials that constitute the Earth. Therefore, the hypothesized phase transitions have been investigated to determine the pressure at which they occur (equal to depths) in the laboratory, and this hypothesis has been verified. This attempt led to the conclusion that the accuracy of the pressure standards used by earth scientists was not satisfactory. This indicates that the accuracy of the observed mapping of elastic properties of the Earth's interior probed by seismic wave velocities is much better than that of the pressure determined in laboratory experiments. In the past 10 years, this "Pressure" problem has been debated by not only earth scientists, but also by physicists.

Here, the attempt to research a reliable pressure scale is performed by both an experimental and theoretical approach. The combination of both approaches leads to an understanding of the uncertainty of each method, and establishes a new pressure scale that is more reliable than previous scales. We use a molecular dynamics method based on first-principles quantum mechanics for the theoretical approach, and use a method of combining a high-pressure diamond anvil cell apparatus with synchrotron X-ray powder diffraction for the experimental approach. In particular, the first-principles computation approach has developed rapidly in recent years and can reliably predict the interesting physical properties of materials that cannot be investigated by high-pressure experiments. This chapter introduces the reliability of first-principles computation and a new pressure scale obtained by combining computational and experimental methods. Moreover, the essential problem of previous pressure scales is discussed and has been clarified by using the data from our first-principles computations.

2.1 Overview of first-principles techniques

To begin, a brief introduction of the first-principles quantum mechanics calculation is described. The first-principles method can predict various physical properties of materials, which is independent of experimental observations. Although it is best if the Schrödinger equation can be strictly solved based on quantum mechanics, this approach is impossible because most materials have many atoms that contain many electrons. Therefore, simplifications and approximations are needed to solve the interaction with many electrons

in atoms. Examples based on density functional theory (DFT) are shown. DFT is assumed to calculate the physical properties of materials using calculated density of electrons. The reader who wants to achieve a detailed understanding of DFT can turn to several comprehensive papers (Parr & Yang, 1994; Martin, 2004; Kohanoff, 2006). The reliability of DFT calculations depends on the approximation used. A pseudopotential method has been used in classical DFT calculations. The computational speed of this technique is very fast, although the first-principles calculations need huge computer power. In the case of the pseudopotential method, it is assumed that the core electrons play little or no role in the energetics of the material, which depends almost entirely on the valence electrons. The core electrons are not handled directly, and these are replaced by the potential function of the valence electrons. Good results have been obtained in predicting physical properties, because the influence of core electrons is small. However, there is a significant problem in estimating physical properties under extremely high-pressure conditions. The material is compressed at high pressures, and the distance between atoms is shortened. In a word, the overlap of the adjacent electrons is nonnegligible. It is expected that the influence of not only valence electrons, but also the core electrons, cannot be disregarded when this overlap increases. Therefore, it is expected that the pseudopotential method is unsuitable for high-pressure studies. Recently, the all-electron method has been used to estimate the physical properties of materials under high pressures instead of pseudopotential methods. Generally, the all-electron method needs huge computing time, and it is difficult to handle materials with many atoms and/or complicated chemical composition. We use the Projector Augmented Wave (PAW) method, because this is a comparatively fast calculation method in the all-electron approach. In most studies of first-principles calculation, the physical properties of materials have been estimated in the ground state (at 0 K). However, earth scientists want to know the behavior of materials at not only high pressures, but also high temperatures, such as those in the interior of the Earth and other planets. Although it is possible to extrapolate the physical properties at high temperatures from the ground state, this approach has a significant error. Therefore, it is necessary to combine a molecular dynamics method that actually deals with the thermal vibration of atoms to estimate physical properties at high temperatures. We use a calculation code (Vienna *Ab initio* Simulation Package) that combines the first-principles calculation of the all-electron approach with a molecular dynamics method (Kresse & Hafner, 1993; Kresse & Furthmuller, 1996). It is possible to estimate precisely the physical properties of a material under high temperatures and pressures using this method. However, there is a problem in that the maximum number of atoms that can be considered is about several hundred, which is much smaller than Avogadro's constant, because this first-principles molecular dynamics method needs huge computer resources. Moreover, it is almost impossible to calculate the duration of a second in real time; however, the duration of tens of picoseconds can be treated. Therefore, it is necessary to estimate thermodynamic parameters under high temperatures and pressures carefully.

2.2 Reliability of the first-principles calculation

The first-principles calculation is used in various fields, such as physics, chemistry, and material science, and huge efforts to increase the accuracy of the calculation have been attempted. Therefore, developments concerning the first-principles technique are very rapid. However, it is important to determine whether its reliability satisfies our purpose to estimate precisely the physical properties of materials at high temperatures and pressures.

For example, it is generally known that the values of the lattice parameters of most crystals estimated by first-principles calculations have about 1–2% errors compared with experimental values at ambient conditions. Next, we will consider the reliability of first-principles calculations under high pressures, which is our subject of interest.
In Figure 1, the relationship between the volume and the pressure of a high-pressure phase of B2-type sodium chloride (NaCl) is shown. In the case of a crystal that is stable at ambient conditions, it is better to obtain physical properties by an experimental method compared with estimations by first-principles calculations. However, there is an advantage of first-principles calculations to estimate the physical properties of materials that cannot be recovered at ambient conditions by high-pressure experiments. As the high-pressure phase of NaCl cannot be quenched at ambient pressures, a complicated method and/or much time is necessary to obtain good experimental data. In such a case, first-principles calculations have a significant advantage in estimating the physical properties of materials compared with experiments. In this chapter, the reliability of first-principles calculations is considered using the example of the high-pressure phase of NaCl.

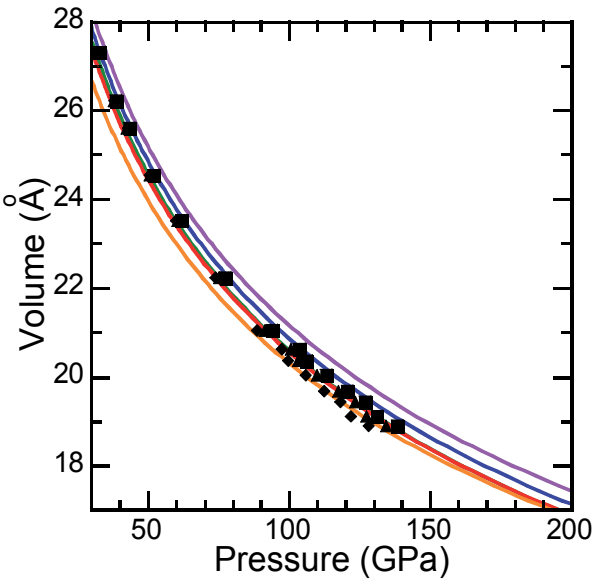


Fig. 1. Comparison of pressure–volume data of B2-type NaCl between high-pressure experiments and first-principles calculations. Symbols represent the experimental data calculated by different gold pressure scales. Diamonds, Jamieson et al. (1982); triangles, Anderson (1989); squares, Takemura (2007) corrected by the ruby scale of Dorogokupets and Oganov (2007). Lines represent the calculated data by different approximations: orange, Local Density Approximation (LDA); purple, PW91 (Wang & Perdew, 1991); blue, HSE06 (Paier et al., 2006); green, AM05 (Armiento & Mattsson, 2005); red, PBEsol (Perdew et al., 2008).

The volume data for the high-pressure phase of NaCl observed by high-pressure experiments at pressures higher than 30 GPa are shown in Figure 1. A high-pressure diamond anvil cell apparatus and a synchrotron X-ray diffraction technique were used to measure volume data at each pressure increment. NaCl powder was mixed with gold powder where the latter was

used as the pressure standard (Ono et al., 2006a). The sample was compressed into a thin pellet, and was sandwiched in the pressure-transmitting medium of magnesium oxide (MgO). This composite sample was placed into the sample chamber of the high-pressure apparatus. It is known that the high-pressure diamond anvil cell-type apparatus has the disadvantage that a significant differential stress is accumulated in the sample chamber during the compression of the sample. Even if a rare gas is used as the pressure-transmitting medium, the influence of differential stress is nonnegligible at pressures higher than ~50 GPa (Takemura, 2007). An alternative approach is annealing the sample using an infrared laser to reduce the differential stress in the sample chamber at each pressure increment. In Figure 1, three experimental data sets are plotted at each volume, because the experimental pressure is calculated using the three proposed pressure scales based on the equation of state for gold (Jamieson et al., 1982; Anderson, 1989; Takemura, 2007). The colored lines are the volume–pressure curves calculated by the first-principles approach. Some types of approximation for the exchange–correlation functional have been proposed. It is necessary to know the exchange–correlation energy between electrons in the materials when the physical properties of the materials are calculated based on the DFT method. As it is difficult to strictly determine the exchange–correlation energy, an approximation must be used. The improvement of the exchange–correlation functional is important for the reliability of the DFT method. Therefore, the results from the different approximations for the exchange–correlation functional are shown in Figure 1. One of the classical approximations is the Local Density Approximation (LDA), where the result is shown as the orange line. Historically, the LDA was partially successful in predicting the physical properties of materials. However, a small difference between the LDA results and the experimental data for the volume–pressure curve of B2-type NaCl is confirmed. For instance, it is known that the volumes of most crystals calculated by the LDA are slightly smaller than those measured experimentally. As the estimations for other physical properties have nonnegligible uncertainties in the LDA method, other approximations were proposed to improve the exchange–correlation functional. One of the other classical approximations is the generalized gradient approximation established by Wang & Perdew (1991) (PW91). The results from the PW91 approximation are also shown as the purple line in Figure 1. In the case of the PW91 approximation, the calculated volumes are larger than are those from experiments. The experimental values of B2-type NaCl are plotted in the intermediate region between the LDA and PW91 calculations. According to previous studies, we understand that the physical properties of materials can be predicted qualitatively using classical approximations. However, it should be noted that further improvements are needed to achieve quantitative predictions. For the last decade, several new approximations have been suggested to improve the accuracy of the DFT calculation. The volume–pressure data calculated by several new approximations are also shown in Figure 1. The results from the new approximations of HSE06 (Paier et al., 2006), AM05 (Armiento & Mattsson, 2005), and PBEsol (Perdew et al., 2008) are in excellent agreement with experimental values, especially at low pressures. This indicates that the computations using the new approximations are likely to produce predictions that are more reliable for the various physical properties of materials.

2.3 Reliability of high-pressure experiments

Next, we will discuss the reliability of data obtained by high-pressure experiments. The three experimental data sets using different pressure standards are plotted in Figure 1. The discrepancy between the different pressure standards increases as the pressure increases.

Thus, the uncertainty of experimental pressures is significant at extremely high pressures. The experimental pressure standards using the equations of state for materials have been determined by a dynamic compression experiment (shock-wave compression experiment) or a static compression experiment (e.g., diamond anvil cell experiment). Shock-wave experimental data have frequently been used to establish and verify a pressure standard. It is useful that the three independent parameters (length, density, and time) can be obtained from each shock-wave experiment and these parameters can be converted to volume, temperature, and pressure, which constitute the equation of state of materials. However, there are a couple of problems in the construction of the equation of state used as the pressure standard by the shock-wave compression data. One of the problems is that the error of the experimental data is great, compared with that of static compression experiments. When the equation of state of a solid is determined and used as the pressure standard, the data from the shock-wave experiments are used sometimes to analyze small changes in thermoelastic parameters, whose changes are smaller than the error in the experimental data. This is quite misleading. Another problem is that there is a significant problem in that an uncertain parameter, such as the Grüneisen constant, is used to estimate the experimental pressure. The dependence of the Grüneisen constant on temperature and pressure has to be assumed with great uncertainty.

On the other hand, static compression experiments also provide invaluable data to investigate a pressure standard. Here, we consider the problem of the pressure standards proposed from previous studies by using the data of the high-pressure phase of NaCl shown in Figure 1. The experimental data in Figure 1 are plotted by using the pressure standards that have been frequently used in previous experimental studies. The difference in pressure data among the different pressure standards indicates the rough size of error for the pressure standard. A comparison of the first-principles calculation and the high-pressure experiment data leads to a valuable discussion. It has already been pointed out that the difference in pressure values obtained by the different approximations used in the first-principles calculation indicates the rough size of the error of the calculations. At low pressures, it is clear that the error in the numerical results is remarkably large compared with the error associated with the experimental data. However, this relationship changes at high pressures. The magnitude of the error in the experimental data increases as the pressure increases, and at pressures higher than ~100 GPa, it becomes almost the same size as the error in the calculation. In addition, this relation is reversed at extremely high pressures. It is known that the difficulty in obtaining reliable measurements increases as the pressure increases, and the error in experimental data becomes significant at high pressures. In the case of the first-principles calculation, the error does not increase as highly when compared with experiments conducted at extremely high pressures. Although significant uncertainties at ambient conditions of the first-principles calculations compared with those of experiments are recognized, the first-principles calculations can provide invaluable data in high-pressure science, where there is great difficulty with experimental approaches.

There is another interesting relationship between the calculated and experimental data. At low pressures, the experimental data are located in the middle of the calculated data. As the pressure increases, the experimental data based on two pressure standards (Jamieson et al., 1982; Anderson, 1989) shift to the low-volume side compared with the calculated values. This indicates that these pressure standards give pressures lower than those from another standard (Takemura, 2007) corrected by the new ruby scale of Dorogokupets & Oganov (2007). The pressures from Takemura's standard are in good agreement with those from the

first-principles calculations using modern approximations. It is believed that the compression data for gold reported recently by Takemura (2007) are among the most reliable. The correction of Takemura's pressure standard is done by one of the reliable ruby standards (Dorogokupets & Oganov, 2007). On the other hand, the experimental data that significantly deviated from the first-principles calculations are based on the old ruby pressure standards (e.g., Mao et al., 1978; Mao et al., 1986). Their pressure standards based on the pressure shift in the fluorescence line of the ruby crystal has been used frequently in many high-pressure studies, especially at ambient and low temperatures. According to the comparison between our numerical results and experimental data, these old ruby standards may underestimate the experimental pressure. Recently, the ruby pressure standard has been revised using experimental data or numerical results by other research groups (Holzapfel, 2003; Kunc et al, 2003; Dewaele et al., 2004; Chijioke et al., 2005; Dorogokupets & Oganov, 2007). These studies reported that the old ruby standards have a significant uncertainty, which is consistent with our study. The underestimation of the old ruby standards seems to be of the order of 5–10%. Finally, we can confirm that the volume–pressure data of B2-type NaCl based on the modern pressure standard are in excellent agreement with those of the first-principles calculations using modern approximations.

2.4 The problem with the Grüneisen constant

Some formulas can be used to describe the equation of state for crystals. For instance, the Mie–Grüneisen–Debye formula has been frequently used because it is simple and easy to apply to various types of systems. Is not there a problem in using this formula? In the Mie–Grüneisen–Debye formula, a thermal pressure (the increase in pressure by heating a solid at constant volume) is expressed as

$$P_{th} = \frac{\gamma}{V}\Delta E_{th} \tag{1}$$

where ΔE_{th}, V, and γ are the internal energy, the volume, and the Grüneisen constant, respectively. The internal energy is given by the Debye's model, and the Grüneisen constant is expressed as a function of the volume.

$$\gamma = \gamma_0 \left(\frac{V}{V_0} \right)^q \tag{2}$$

This expression is the Mie–Grüneisen–Debye equation of state. The Grüneisen constant used in the Mie–Grüneisen–Debye equation of state has a significant problem. This equation indicates that the pressure value might have a large error if the Grüneisen constant has a significant uncertainty. In the case of the Mie–Grüneisen–Debye equation of state, the temperature dependence of the Grüneisen constant is assumed to be negligible. If the temperature dependence of the Grüneisen constant is nonnegligible, the Mie–Grüneisen–Debye equation of state gives misleading values of pressure in the study of solid crystals. Therefore, we verified the dependence of the temperature and the pressure on the Grüneisen constant using the first-principles molecular dynamics method (Ono et al., 2008).

The first-principles molecular dynamics method can directly calculate the internal energy, the volume, and the thermal pressure defined in equation (1), and the dependence of the temperature and the pressure on the Grüneisen constant can also be calculated by using these values. The pressure dependence of the Grüneisen constant included in the Mie–

Grüneisen–Debye equation of state was confirmed and is demonstrated in Figure 2. On the other hand, the temperature dependence neglected in the Mie–Grüneisen–Debye equation

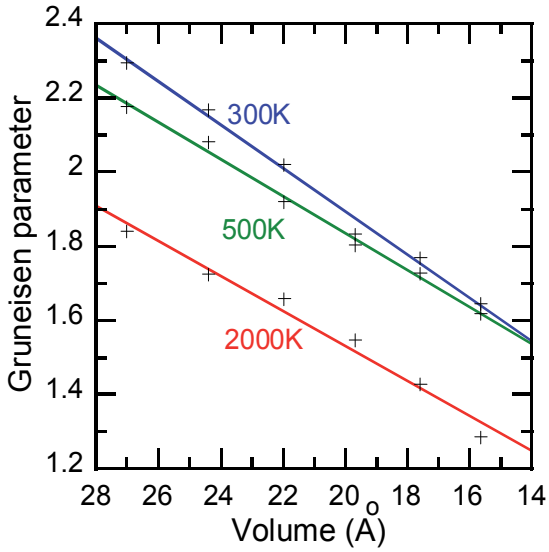


Fig. 2. Calculated Grüneisen parameter. The crosses represent the calculated Grüneisen parameters at 300, 500 and 2000 K using first-principles molecular dynamics calculations. The solid lines denote the linear fit using the least-squares method.

of state is small at temperatures of up to several hundred degrees. However, the difference in the dependence between low and high temperatures could not be neglected especially at temperatures higher than 1000 K. According to this behavior, the Mie–Grüneisen–Debye equation of state can be applied at several hundred degrees at the highest, when it is constructed using experimental data obtained at low temperatures. In other words, the Mie–Grüneisen–Debye equation of state is not suitable for constructing the equation of state of solids applied to a wide range of temperatures, and the pressure standard based on the Mie–Grüneisen–Debye equation of state may involve significant uncertainty. According to Maxwell's relations in thermodynamics, the temperature dependence of the Grüneisen constant at constant volume is given by

$$\left(\frac{\partial \gamma}{\partial T}\right)_V = \frac{1}{T}\left(\frac{\partial \ln C_V}{\partial \ln V}\right)_S \tag{3}$$

When the temperature is much higher than the Debye temperature, the Dulong–Petit law suggests that the specific heat (C_V) is almost constant in solids. That is, the temperature dependence of the Grüneisen constant is extremely small at extremely high temperatures. As the temperature dependence is not small at middle temperatures, it is possible that an obvious difference in the temperature dependence of the Grüneisen constant between low and high temperatures can be confirmed. It can explain the change in the Grüneisen constant calculated by the first-principles molecular dynamic method. Here, the high-pressure phase of NaCl, which is a typical crystal constituted by an ionic bond, has been investigated. It is thought that similar studies for other materials with different bond types,

such as metals or oxides, should be carried out to understand a general rule for the temperature dependence of the Grüneisen constant of solids at high pressures. It is extremely dangerous to use the Grüneisen constant without understanding its behavior at high temperatures, because it is one of the most important parameters to establish the equations of state of solids for not only static compression, but also shock-wave compression experiments. Indeed, most studies on constructing pressure standards have not considered this influence of the Grüneisen constant.

2.5 New pressure standard obtained by first-principles calculations combined with high-pressure experimental data

It is difficult to construct a new pressure standard that is more reliable than those proposed by previous studies using experimental data at high pressures by the first-principles method, because it is clear that the errors are not negligible in the current first-principles computations. On the other hand, we have understood that there are serious problems in determining experimental pressure standards under high pressure because of the uncertainties of the ruby pressure standard or the Grüneisen constant.

Recent approaches to establish a reliable pressure standard employed much previously reported experimental data obtained with static and the shock-wave compression methods to determine the parameters of the equation of state. However, a serious problem is associated with this approach. The previous static compression data used to construct the equation of state are affected directly or indirectly by the old ruby pressure standard. As described above, a significant error is confirmed in the old ruby pressure standards (Mao et al., 1978; Mao et al., 1986), which have been widely used in previous studies. Thus, it is thought that most static compression experimental data used in the construction of the equation of state for the pressure standard have nonnegligible errors. Therefore, it should only be used as fundamental data to construct the equation of state after the problem of the old ruby pressure standard is solved. In Figure 1, one of the experimental data of B2-type NaCl is plotted using the new gold pressure values corrected by the new ruby pressure standard. This is one of the approaches that can be used to avoid the uncertainty of the old ruby pressure standard. Moreover, it is also clear that the problem with the Grüneisen constant must be solved before previous shock-wave compression experimental data are used to establish the equation of state for a reliable pressure standard.

Recently, an attempt to understand an internal consistency among some pressure standards using high-pressure experiments has been reported (Fei et al., 2007), as it is not easy to construct a reliable pressure standard. This approach in constructing a reliable pressure standard also has a major problem. Each experimental data set has a different uncertainty that is dependent on the experimental conditions, such as skill and/or the method. When these experimental data are combined to analyze the pressure standard, the uncertainties of each data are accumulated. The most reliable gold pressure standard used in Figure 1 is established using the compression data of gold reported by Takemura (2007). This compressibility of gold was obtained using the correction of the nonhydrostatic pressure conditions for the experimental data of Takemura (2001) and Dewaele et al. (2004). Both research groups have published many papers concerning the equations of state for solids, and their experimental data have been cited repeatedly. Although the reliability of their experimental skills is recognized in the community of high-pressure science, it is necessary to correct their raw data when we establish a precise pressure standard. This indicates that it is quite difficult to compare the experimental data reported by different research groups that

have different levels of experimental skills and use different methods. In other words, the study of consistency between different pressure standards is difficult without considering experimental errors, such as the nonhydrostatic condition. Indeed, investigation into the consistency of pressure standards while considering the error of each experiment is rare.
In order to overcome the difficulties mentioned above, we have performed an alternative approach of combining high-pressure experimental data and first-principles calculations to make up for each disadvantage. According to our study, the volume–pressure relation of B2-type NaCl estimated by first-principles calculations may have nonnegligible errors. The reliability of high-pressure experimental data obtained at ambient temperatures is better than that of first-principles calculations if an appropriate correction for the experimental pressures is done. Therefore, the reliability of the numerical results is improved considerably if the pressure–volume–temperature data estimated by the first-principles molecular dynamics calculations are corrected based on the experimental data.
On the other hand, it is known that there is significant uncertainty of the temperature in the study of the equation of state of solids reported by previous high-pressure experiments. It is difficult experimentally to determine precisely the thermal pressure and the coefficient of thermal expansion at high temperatures. In the case of the static compression experiments using a large volume press apparatus, a thermocouple is used to measure the sample temperature. The thermocouple indicates the sample temperature using a voltage (EMF) proportional to the temperature difference between two different metal conductors. Although the method has been established at ambient pressures, it is thought that a significant uncertainty is introduced at high pressures. The influence of pressure on the EMF has to be considered in high-pressure experiments. Indeed, most experimental studies have not taken into account this pressure effect on EMF. At higher pressures performed using large volume press experiments, laser-heated diamond anvil cell experiments have often been performed. In this experimental method, the temperature of the sample is estimated using the radiation from the sample during heating. It is known that this temperature measurement has an extremely large error. In addition to the large fluctuation of temperature (~several hundred K) on heating, the conversion from the spectrum of the radiation from the sample to sample temperature has a significant uncertainty that concerns thermal emission. If the sample is an ideal black body, it is simple in that the emissivity is 1. However, the sample is not a black body. Therefore, it is necessary to know the dependence of the emissivity on the temperature, pressure, and wavelength of each material to estimate an accurate temperature from the radiation of the sample. These dependencies under high pressures and temperatures are uncertain. In the case of the shock-wave compression experiments, the experimental data might contain large errors because of the problem concerning the Grüneisen constant mentioned above. Thus, there is a significant advantage in the use of data calculated by first-principles molecular dynamics to know the dependence of the temperature concerning the equations of state for materials.
From the viewpoint mentioned above, an attempt to establish a reliable equation of state for a material to construct a new pressure standard has been performed by combining the data from high-pressure experiments at room temperature with the data from first-principles molecular dynamics calculations at high temperatures. We have already discussed the synergy between the high-pressure experiments and the first-principles computations concerning the equation of state for the high-pressure phase of B2-type NaCl. Recently, a study on a pressure standard of B2-type NaCl combining both experiments and computations has been reported (Ono, 2010a). A study on Ta metal has also been reported

(Ono, 2009). The equations of state of B2-type NaCl and Ta have been revised considerably by these studies. In contrast, a certain amount of error in the parameters of the equation of state concerning the temperature was confirmed in the first-principles molecular dynamics calculations. The error in the calculations is thought to be related to the estimation of the interaction factor between the electrons. As the first-principles computations clarified the interesting features of the ruby pressure standard and the Grüneisen constant, which are important factors to determine the precise equation of state for materials, synergy between first-principles computations and high-pressure experiments is necessary to establish more reliable pressure standards in future studies.

3. Implications for earth science

It is believed that the Earth's and terrestrial planets' interiors consist of oxides and iron compounds. The physical properties of these minerals are important keys to understand their structure, composition, and evolution. Recently, the approach of combining high-pressure experiments with first-principles computations was successful in discovering new minerals and predicting physical properties of minerals at high pressures and temperatures. In this section, some interesting topics are introduced.

3.1 High-pressure phase of iron

The stable structure of iron under ambient conditions is body-centered cubic (bcc) Fe (α-Fe). The phase transition from bcc to the face-centered cubic (fcc) structure (γ-Fe) has been confirmed to occur at a temperature of 1185 K. Under high pressure, the bcc-Fe transforms into the hexagonal close-packed (hcp) structure (ε-Fe) (Takahashi and Bassett, 1964), and this structure seems to be stable over a wide range of pressures and temperatures approaching those existing in the Earth's core. It is known that the magnetic and spin states of iron have a major influence on the physical properties of iron. Although the magnetic structure of hcp-Fe has been investigated for over four decades, there is an inconsistency between experimental and theoretical studies. Mössbauer experiments have been interpreted to show the absence of magnetism in hcp-Fe (Williamson et al., 1972; Nasu et al., 2002). In contrast, the theoretical study based on DFT has shown that the antiferromagnetic state is stable at pressures below 50 GPa (Steinle-Neumann et al., 1999). However, it has not been clearly explained why it is difficult to identify the antiferromagnetic state experimentally. A first possibility is that the significant hysteresis of the bcc–hcp transition may disturb the magnetic ordering in hcp-Fe; this is because most previous experiments were performed at low temperatures where the accumulated differential stress in the sample could not be released. Second, the experimental errors of the Mössbauer technique used in previous studies were not negligible, because the antiferromagnetic moment predicted by first-principles calculations is small. Third, the quantum spin fluctuation in hcp-Fe (Mazin et al., 2002) is too fast for the time scale of the Mössbauer measurement, thereby inhibiting detection of a hyperfine field.

We investigated the magnetic properties of hcp-Fe using first-principles calculations to determine the change in the cell parameters. The antiferromagnetic type II structure (space group = *Pmma*) was used in our calculations. The antiferromagnetic type II structure has two different Fe sites, and their magnetic moments are opposed to each other. Figure 3 shows the change in the ratio of the cell parameters of hcp-Fe as a function of pressure. In the case of the antiferromagnetic type II structure, the calculated c/a ratio decreases up to

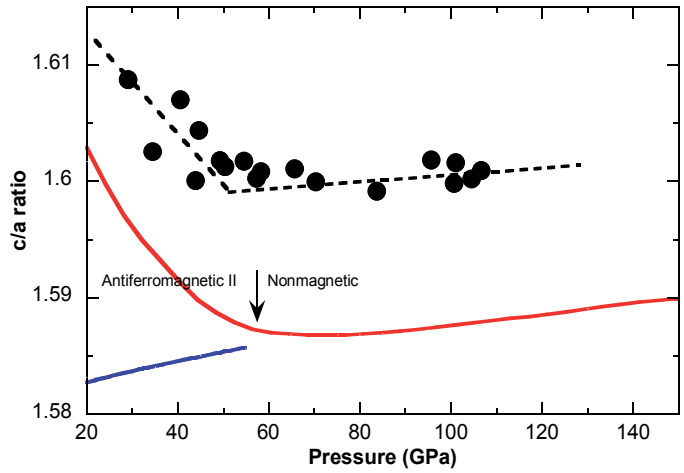


Fig. 3. Changes in the ratio of the cell parameters of hcp-Fe. Black circles represent the c/a ratio from experimental data at room temperature. The red line represents the change in c/a ratio calculated by the first-principles computations including the spin effect at 0 K. The blue line represents the calculated ratio without the spin configuration.

approximately 55 GPa, and then increases slightly with increasing pressure. In contrast, the nonmagnetic type structure without the spin effect shows that the calculated c/a ratio increases simply with increasing pressure. Experimental data (Ono et al., 2010b) are also shown in Figure 3. Although the ratio of the cell parameters from our first-principles calculations is approximately 1% greater than that observed in our experiments, the change in the ratio from the calculations for the antiferromagnetic type II structure is in good agreement with the experimental data. The change in the ratio of the metastable nonmagnetic state calculated without the spin effect is apparently inconsistent with that observed in experiments. The magnetic moment of the antiferromagnetic state decreased gradually, with increasing pressure, and reached zero at approximately 55 GPa (Ono et al., 2010b). The disappearance of the magnetic moment indicates that a magnetic transition occurred from the antiferromagnetic to the nonmagnetic state. As the spin directions of the antiferromagnetic state are orientated perpendicular to the c axis in the hexagonal symmetry cell, and the a axis is less compressible than the c axis, it is therefore clear that a magnetic transition would induce a change in the rate at which the c/a ratio responds to compression. Finally, the disagreement between previous experiments and computations was reconciled by our approach using both first-principles calculations and high-pressure experimental data. This study indicates that the low-spin hcp-Fe is the most stable phase in the Earth's core at pressures of ~350 GPa.

3.2 High-pressure phases of $MgSiO_3$

Seismic measurements infer that the Earth's interior has two layers, the mantle and the core. The core is much denser than the mantle, and consists of iron-rich materials. The chemistry of the mantle can be estimated from information obtained from meteorites and cosmochemistry. This indicates that the composition of the mantle is close to that of the universe as a whole, but with a strong depletion of volatile elements, such as hydrogen,

carbon, and rare gases. According to mineralogy studies, the upper mantle must be dominated by Mg_2SiO_4 compounds (Ono, 2008). In contrast, the lower mantle may consist of $MgSiO_3$ and MgO. High-pressure experiments and first-principles calculations show that $MgSiO_3$ in the lower mantle has a perovskite structure. The nature of the D″ layer at the base of the lower mantle is unique. It has quite a variable thickness and a significant seismic anisotropy. The most plausible explanation is that $MgSiO_3$ transforms from a perovskite to a $CaIrO_3$-type (post-perovskite) structure in this region (Oganov & Ono, 2004). First-principles computations contributed to the discovery of this new mineral.

At first, the $CaIrO_3$-type structure at high pressures was reported by an experimental study of iron oxide (Fe_2O_3) (Ono et al., 2004; Ono & Ohishi, 2005). It is known that one of the high-pressure phases of $MgSiO_3$ has an ilmenite-type structure. The ilmenite-type structure is the same as that of hematite, which is the stable structure of Fe_2O_3 at ambient conditions. Thus, $MgSiO_3$ may have a transition sequence similar to Fe_2O_3, and the $CaIrO_3$-type structure observed in Fe_2O_3 may appear in $MgSiO_3$ at high pressures. This assumption has been confirmed by first-principles computations. Finally, a new mineral of $MgSiO_3$ has been confirmed by both first-principles calculations and high-pressure experiments (Oganov & Ono, 2004). The structure of $CaIrO_3$-type $MgSiO_3$ determined experimentally is shown in Figure 4.

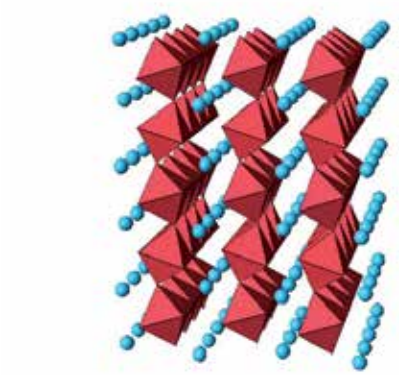


Fig. 4. Crystal structure of $CaIrO_3$-type (post-perovskite) $MgSiO_3$ (JCPDS-0580689). Blue spheres are Mg atoms. Polyhedra indicate SiO_6 octahedra. The precise structure was determined by Rietveld refinement (Ono et al., 2006b).

The existence of $CaIrO_3$-type $MgSiO_3$ can explain the unique features at the base of the Earth's lower mantle (Ono & Oganov, 2005). Although at present the $CaIrO_3$-type $MgSiO_3$ phase exists in the Earth's deep interior, the condition of the Earth in the past was quite different from that of present Earth. The Earth's temperature just after its formation was much higher than it is presently. At higher temperatures, the $CaIrO_3$-type $MgSiO_3$ phase in the D″ layer did not exist at the base of the lower mantle, because perovskite-type $MgSiO_3$ is stable deep in the Earth's mantle (Figure 5). Then, the $CaIrO_3$-type $MgSiO_3$ phase appeared during the cooling of the Earth (Ono & Oganov, 2005). The solid inner core, which consists of iron compounds, also appeared during the cooling stage of the Earth.

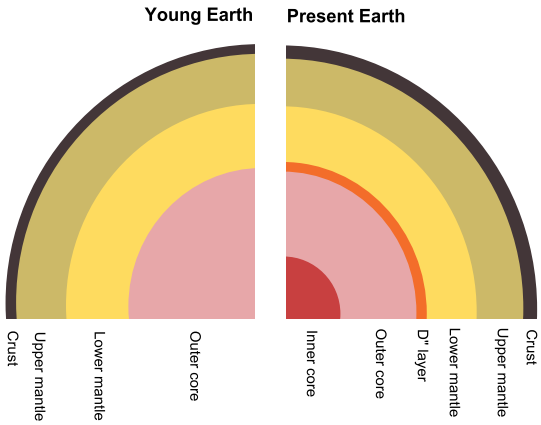


Fig. 5. Structure of the Earth's interior. The temperature of the young Earth was higher than that of the present Earth (Ono & Oganov, 2005).

3.3 High-pressure phases of carbonates

High-pressure polymorphs of carbon-bearing minerals are important to understand the circulation of carbon in the Earth's interior. Therefore, phase transitions and physical properties of high-pressure phases related to carbon have been the object of intense experimental investigation. High-pressure phases of carbonates are likely to be one of the host minerals for carbon that are present deep in the mantle. However, significant discrepancies in the crystal structures of high-pressure carbonates have been reported by previous high-pressure studies because of experimental difficulties. Recent first-principles computations solved some of these experimental discrepancies and contributed to the discovery of several new high-pressure structures.

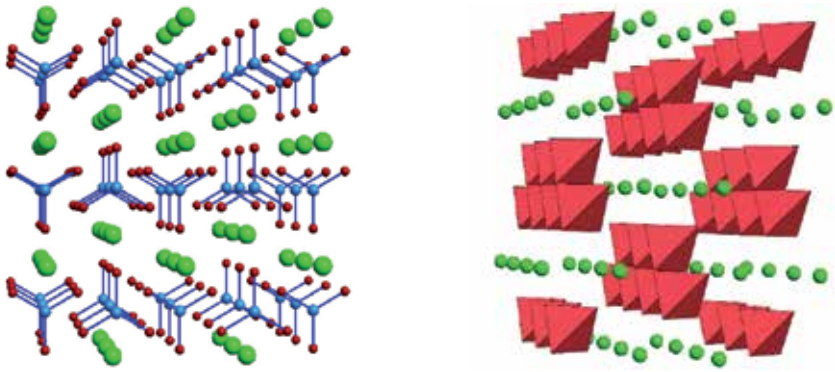


Fig. 6. Crystal structures of high-pressure polymorphs of $CaCO_3$. Left, post-aragonite $CaCO_3$; right, pyroxene-like phase. Green, blue, and red spheres are Ca, C, and O atoms, respectively. Blue bonds indicate triangular CO_3. Red polyhedra indicate CO_4 tetrahedra. The structure was predicted by first-principles calculations (Oganov et al., 2006).

Calcium carbonate ($CaCO_3$) is believed to be a major mineral containing carbon that exists in the deep interior of the Earth. It is generally known that calcite, which is the stable structure of $CaCO_3$ at ambient conditions, transforms to aragonite, which often occurs in

metamorphic rocks, formed at high pressures and temperatures that correspond to the Earth's crust and the uppermost upper mantle. It was unknown whether aragonite transforms to a new high-pressure phase deep in the mantle. Recently, experimental observations indicated that aragonite $CaCO_3$ transformed into a new structure (post-aragonite phase) (Ono et al., 2005). The estimation of the crystal structure of this new $CaCO_3$ had significant uncertainty because of the poor quality of the experimental data. In contrast, the recent approach to determine the unknown crystal structure based on first-principles computation can provide a powerful tool for high-pressure mineralogy. One of the successful codes used in this approach is "USPEX" developed by Oganov & Glass (2008). The computations using the USPEX code solved the unknown structure of the high-pressure phase of $CaCO_3$ (Oganov et al., 2006). This new phase (post-aragonite $CaCO_3$) has an orthorhombic symmetry with space group *Pmmn*, as shown in Figure 6. This study also predicted the existence of another unknown structure with $C222_1$ symmetry, which had not been observed in previous experimental studies. One of the interesting features of this new structure is that a change in the coordination number of the carbon atom is predicted. All $CaCO_3$ polymorphs reported in previous studies are composed of triangular CO_3 units. In contrast, the new structure has CO_4 tetrahedron units (Figure 6). After this prediction, high-pressure experiments confirmed the existence of the $C222_1$ structure (pyroxene-like phase) (Ono et al., 2007). The transition pressure predicted by first-principles calculations (Oganov et al., 2006) is in excellent agreement with that observed experimentally (Ono et al., 2007). This indicates that the prediction based on first-principles computations is a powerful tool to investigate the behavior of high-pressure phases.

4. Conclusion

In conclusion, we considered uncertainties in the computations based on first-principles quantum mechanics and high-pressure experiments, and introduced a new attempt to construct a pressure standard based on the equations of state of materials using the synergy between the first-principles computations and high-pressure experiments. In the present situation of first-principles calculations, it is difficult for the computation approach to establish the equation of state for materials that are more reliable than that established by experimental data. However, the speed of the development in the field of first-principles calculation is extremely rapid because of the very high demand from many fields, such as physics, chemistry, life science, medicine, and engineering. Therefore, computations based on first-principles quantum mechanics will make an important contribution to the problem of establishing a pressure standard in the near future. Thus, the importance of the combination of experimental and theoretical approaches increases further, and the possibility of a significant breakthrough using this synergy can be expected. Indeed, our approach combining first-principles computations and high-pressure experiments led to a new insight into the interesting behavior of the high-pressure phase of iron. Furthermore, the discovery of a new type of iron oxide (Fe_2O_3) inspired the discovery of a new phase of $MgSiO_3$, which contributes to understanding the properties of the bottom of the Earth's mantle. In the case of carbonate, first-principles computations predicted the structure of new phases that had not previously been identified experimentally, and helped to identify new phases in later experiments. Our investigations into carbonates also contributed to understanding the carbon cycle in the Earth's interior.

5. References

Anderson, O.L., Isaak, D.G. & Yamamoto, S. (1989) Anharmonicity and the equation of state for gold, *Journal of Applied Physics*, (February 1989), Vol.65, No.4, pp.1534-1543, ISSN 0021-8979.

Armiento, R. &Mattsson, A. E. (2005) Functional designed to include surface effects in self-consistent density functional theory, *Physical Review B*, (August 2005), Vol.72, No.8, pp.085108, ISSN 0163-1829.

Chijioke, A.D., Nellis, W.J., Soldatov, A. & Silvera, I.F. (2005) The ruby pressure standard to 150 GPa, *Journal of Applied Physics*, (December 2005), Vol.98, No.11, pp.073526 ISSN 0021-8979.

Dewaele, A., Loubeyre, P. & Mezouar, M. (2004) Equations of state of six metals above 94 GPa, *Physical Review B*, (September 2004),Vol.70, No.9, pp.094112, ISSN 0163-1829.

Dorogokupets, P.I. & Oganov, A.R. (2007) Ruby, metals, and MgO as alternative pressure scales: A semiempirical description of shock-wave, ultrasonic, x-ray, and thermochemical data at high temperatures and pressures, *Physical Review B*, (January 2007),Vol.75, No.2, pp.024115, ISSN 1098-0121.

Fei, Y., Ricolleau, A., Frank, M., Mibe, K., Shen, G. & Prakapenka, V. (2007) Toward an internally consistent pressure scale, *Proc. Natl. Acad. Sci. U. S. A.*, (May 2007), Vol.104, No.22, pp.9182-9186, ISSN 0027-8424.

Holzapfel, W.B. (2003) Refinement of the ruby luminescence pressure scale, *Journal of Applied Physics*, (February 2003), Vol.93, No.3, pp.1813-1818, ISSN 0021-8979.

Jamieson, J.C., Fritz, J.N. & Manghnani, M.H. (1982). Pressure measurement at high temperature in x-ray diffraction studies: gold as a primary standard, In: *High-Pressure Research in Geophysics*, S. Akimoto & M.H. Manghnani, (Ed.), 27-48, Center for Academic Publishing, ISBN 978-902-7714-39-8, Tokyo, Japan

Kohanoff, J. (2006), *Electoinic Structure Calculations for Solids and Molecules : Theory and Computaional Method,* Cambridge University Press, ISBN 978-052-1815-91-8, Cambridge, UK

Kresse, G. & Hafner, J. (1993) *Ab initio* molecular-dynamics for liquid-metals, *Physical Review B*, Vol.47, No.1 (January 1993), pp. 558-561, ISSN 0163-1829.

Kresse, G. & Furthmuller, J. (1996) Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, *Physical Review B*, Vol.54, No.16, (October 1996), pp. 11169-11186, ISSN 0163-1829.

Kunc, K., Loa, I. & Syassen, K. (2004) Diamond under pressure: Ab-initio calculations of the equation of state and optical phonon frequency revisited, *High Pressure Research*, Vol.24, No.1, pp.101-110, ISSN 0895-7959.

Mao, H-K., Bell, P.M., Shaner, J.W. & Steinberg, D.J. (1978) Specific volume measurements of Cu, Mo, Pd, and Ag and calibration of ruby R1 fluorescence pressure gauge from 0.06 to 1 mbar, *Journal of Applied Physics*, (June 1978) Vol.49, No.6, pp.3276-3283, ISSN 0021-8979.

Mao, H-K., Xu, J.P. & Bell, P. J. (1986) Calibration of the ruby pressure gauge to 800-kbar under quasi-hydrostatic conditions, *Journal of Geophysical Research-Solid Earth and Planets*, (April 1986) Vol.91, No.B5, pp. 4673-4676, ISSN 0148-0227.

Mazin, I.I., Papaconstantopoulos, D.A., & Mehl, M.J. (2002) Superconductivity in compressed iron: role of spin fluctuations. *Physical Review B*, (March 2002), Vol.65, No.10, pp.100511, ISSN 1098-0121.

Martin, R.M. (2004). *Electoinic Structure,* Cambridge University Press, ISBN 978-052-1534-40-6, Cambridge, UK

Nasu, S., Sasaki, T., Kawakami, T., Tsutsui, T., & Endo, S. (2002) Mossbauer study of epsilon-Fe under an external magnetic field, *Journal of Physics Condensed Matter*, (November 2002), Vol.14, No.44, pp.11167-11171, ISSN 0953-8984.

Oganov, A.R. & Ono, S. (2004) Theoretical and experimental evidence for a post-perovskite phase of MgSiO3 in Earth's D" layer, *Nature*, (July 2004), Vol.430, No.6998, pp.445-448, ISSN 0028-0836.

Oganov, A.R., Glass, C.W. & Ono, S. (2006) High-pressure phases of $CaCO_3$: crystal structure prediction and experiment, *Earth and Planetary Science Letters*, (January 2006) Vol.241, No.1-2, pp.95-103, ISSN 0012-821X.

Oganov, A.R. & Glass, C.W. (2008) Evolutionary crystal structure prediction as a tool in materials design, *Journal of Physics Condensed Matter*, (February 2008) Vol.20, No.6, pp.064210, ISSN0953-8984.

Ono, S., Kikegawa, T. & Ohishi, Y. (2004) High-pressure phase transition of hematite, Fe_2O_3, *Journal of Physics and Chemistry of Solids*, (August-September 2004) Vol.65, No.8-9, pp. 1527-1530, ISSN 0022-3697.

Ono, S. & Ohishi, Y. (2005) In situ X-ray observation of phase transformation in Fe_2O_3 at high pressures and high temperatures, *Journal of Physics and Chemistry of Solids*, (October 2005) Vol.66, No.10, pp.1714-1720, ISSN 0022-3697.

Ono, S. & Oganov, A.R. (2005) In situ observations of phase transition between perovskite and $CaIrO_3$-type phase in $MgSiO_3$ and pyrolitic mantle composition, *Earth and Planetary Science Letters*, (august 2005), Vol.236, No.3-4, pp.914-932, ISSN 0012-821X.

Ono, S., Kikegawa, T., Ohishi, Y. & Tsuchiya, J. (2005) Post-aragonite phase transformation in $CaCO_3$ at 40 GPa, Am. Mineral., *American Mineralogist*, (April 2005), Vol.90, No.4, pp.667-671, ISSN 0003-004X.

Ono, S., Kikegawa, T. & Ohishi, Y. (2006b) Structural properties of $CaIrO_3$-type $MgSiO_3$ up to 144 GPa, *American Mineralogist*, (February-March 2006)Vol.91, No.2-3, pp.475-478, ISSN 0003-004X.

Ono, S., Kikegawa, T. & Ohishi, Y. (2006a) Structural property of CsCl-type sodium chloride under pressure, *Solid State Communications*, Vol.137, No.10, (March 2006), pp. 517-521, ISSN 0038-1098.

Ono, S., Kikegawa, T. & Ohishi, Y. (2007) High-Pressure Transition of $CaCO_3$, *American Mineralogist*, (July 2007), Vol.92, No.7, pp.1246-1249, ISSN 0003-004X.

Ono, S. (2008) Experimental constraints on the temperature profile in the lower mantle, *Physics of the Earth and Planetary Interiors*, (November 2008), Vol.170, No.3-4, pp.267-273, ISSN 0031-9201.

Ono, S., Brodholt, J.P., Alfè, D., Alfredsson, M. & Price, G.D. (2008) Ab initio molecular dynamics simulations for thermal equation of state of B2-type NaCl, *Journal of Applied Physics*, (January 2008), Vol.103, No.2, pp.023510, ISSN 0021-8979.

Ono, S. (2009) First-principles molecular dynamics calculations of the equation of state for tantalum, *International Journal of Molecular Sciences*, (October 2009), Vol.10, No.10, pp.4342-4351, ISSN 1422-0067.

Ono, S. (2010a) The equation of state of B2-type NaCl, *Journal of Physics: Conference Series*, Vol.215, (March 2010), pp. 012196, ISSN 1742-6588.

Ono, S., Kikegawa, T., Hirao, N. & Mibe, K. (2010b) High-pressure magnetic transition in hcp-Fe, *American Mineralogist*, (May-June 2010), Vol.95, No.5-6, pp.880-883, ISSN 0003-004X.

Paier, J., Hirschl, R., Marsman, M. & Kresse, G. (2005) The Perdew-Burke-Ernzerhof exchange-correlation functional applied to the G2-1 test set using a plane-wave basis set, *Journal of Chemical Physics*, (June 2005), Vol.122, No.23, pp.234102, ISSN 0021-9606.

Parr, R.G. & Yang, W. (1994). *Density-Functional Theory of Atoms and Molecules,* Oxford University Press, ISBN 978-019-5092-76-9, Oxford, UK

Perdew, J.P., Ruzsinszky, A., Csonka, G.I., Vydrov, O.A., Scuseria, G.E., Constantin, L.A., Zhou, X. & Burke, K. (2008) Restoring the density-gradient expansion for exchange in solids and surfaces, *Physical Review Letters*, (April 2008), Vol.100, No.13, pp.136406 ISSN 0031-9007.

Steinle-Neumann, G., Stixrude, L. & Cohen, R.E. (1999) First-principles elastic constants for the hcp transition metals Fe, Co, and Re at high pressure. *Physical Review B*, (July 1999), Vol.60, No.2, pp.791-799, ISSN 0163-1829.

Takahashi, T. & Bassett, W.A. (1964) High-pressure polymorph of iron, *Science*, (July 1964), Vol.145, No.3631, pp.483-486, ISSN 0036-8075.

Takemura, K. (2001) Evaluation of the hydrostaticity of a helium-pressure medium with powder x-ray diffraction techniques, *Journal of Applied Physics*, (Juanuary 2001), Vol.89, No.1, pp.662-668, ISSN 0021-8979.

Takemura, K. (2007) Pressure scales and hydrostaticity, *High Pressure Research*, (December 2007), Vol.27, No.4, pp.465-472, ISSN 0895-7959.

Wang, Y. & Perdew, J.P. (1991) Correlation hole of the spin-polarized electron gas, with exact small-wave-vector and high-density scaling, *Physical Review B*, (December 1991), Vol.44, No.24, pp.13298-13307, ISSN 0163-1829.

Williamson, D.J., Bukshpan, S. & Ingalls, R. (1972) Search for magnetic ordering in hcp iron. *Physical Review B*, (December 1972), Vol.6, No.11, pp.4194-4206, ISSN 0163-1829.

5

Quantum Mechanical Three-Body Systems and Its Application in Muon Catalyzed Fusion

S. M. Motevalli and M. R. Pahlavani
Department of Physics, Faculty of Sciences,
University of Mazandaran, Babolsar
Iran

1. Introduction

A negative muon is a lepton of the second generation with mass number about times heavier than that of electrons, and has a finite lifetime of $\tau_\mu = 2.197 \times 10^{-6}$ sec . This lifetime is amply long for most experiments. Muon catalyzed fusion (μCF) is a physical phenomenon in which the negative muon is able to cause fusion at room temperature and thereby eliminating the need for high temperature plasmas or powerful lasers (Owski, 2007; Imo et al., 2006; Filchenkov et al., 2005; Filipowicz et al., 2008; Pahlavani & Motevalli, 2008, 2009; Marshal, 2001; Bystritsky et al., 2006, Nagamine et al., 1987; Nagamine, 2001; Ponomarev, 2001). In comparison with (μCF), hot fusion schemes are made difficulty by the electrostatic (Coulomb) repulsion between positively charged nuclei. In the two conventional approaches to control fusion namely, Magnetic Confinement Fusion (MCF) and Inertial Confinement Fusion (ICF), barrier is partially surmounted by energetic collisions. The particle densities, n and confinement times, τ in the plasma, ($T \geq 10^8 K$) are typically more than ten orders of magnitude difference for these two schemes but the product of these quantities required for d−t fusion is $n\tau \geq 10^{14}\,\mathrm{sec}/\,cm^3$. For the μCF, effectively $n\tau \approx 10^{25}\,\mathrm{sec}/\,cm^3$, but this criterion does not tell the whole truth because, in μCF the objective is to tunnel through the barrier without the benefit of kinetic energy. It is known that the $d-t$ fusion by the usual magnetic or inertia confinement suffering a lot of difficulties and problems causing from tritium handling, neutron damage to materials and neutron-induced radioactivity, etc.

Study of the muon catalyzed fusion reactions is of great interest and carried out in many laboratories of the world recently (Ishida et al., 1999; Petitjean et al., 1992, 1993; Bystritsky et al., 2005; Pahlavani & Motevalli, 2008, Bystritsky et al., 2000; Matsuzaki et al., 2001). Muons can be created by the decay of pion which is generated in the collision of intermediate-energy proton with target nuclei. In the muon catalyzed fusion process, after injection of muon in to deuterium and tritium mixture, either $d\mu$ or a $t\mu$ atom is formed, with a probability proportional to the relative concentrations of D and T in the mixture. These atoms are formed in exited states (Breunlich et al., 1989; Korenman, 1996) and then, due to cascade processes, de-excite to ground states. The following reactions illustrate direct formation of muonic dμ and tμ atoms

$$\mu^- + D \rightarrow d\mu + e^- \quad (\lambda_d) \tag{1}$$

$$\mu^- + T \rightarrow t\mu + e^- \quad (\lambda_t) \tag{2}$$

where e^- denotes an electron and λ_d and λ_t are the rate of reactions (1) and (2). The probability of formation of the $d\mu$ atom that will reach its *1s* ground state is quantified by the parameter q_{1s}, which is a function of target density, φ and tritium concentration, C_t. Also it is very sensitive to the $d\mu$ kinetic energy distribution (Menshikov & Ponomarev, 1984; Czaplinski et al., 1994). The difference between binding energies of $t\mu$ and $d\mu$ is about *48.1 eV* (Bom et al., 2005). Therefore, the transfer of a muon from $d\mu$ to a triton is favorable for all temperatures in the given processes

$$d\mu + t \rightarrow t\mu + d + 48.1eV \ (\lambda_{dt}) \tag{3}$$

with a rate of $\lambda_{dt} = 2.8 \times 10^8 \phi$ (Caffery et al., 1987; Jones et al., 1987; Bystritsky et al., 1980; Breunlich et al., 1987). The muon mass is about *206.77* times larger than the mass of electron. Consequently, the size of a muonic hydrogen atom is smaller than the one of the electronic hydrogen by the same rate approximately. These small muonic atoms can approach other hydrogen nuclei experiencing reduced Coulomb barrier and then induce *d–t* fusions. The process in which a muonic molecule is formed is the most important step in the μCF. The formation of muonic molecules of hydrogen isotopes and their nuclear reactions have been the subject of many experimental and theoretical studies (Caffery et al., 1987; Jones et al., 1987; Bystritsky et al., 1980). In collisions of $t\mu$ muonic atoms with D_2 and DT molecules, the muonic molecules $dt\mu$ are formed during a time interval $\tau_{dt\mu} \leq 10^{-8}$ sec (Jones et al., 1983; Eliezer & Henis, 1994) according to the following resonance reactions

$$t\mu + D_2 \rightarrow \left[(dt\mu)_{Jv}\, d2e \right] \ (\lambda_{dt\mu-d}) \tag{4}$$

$$t\mu + DT \rightarrow \left[(dt\mu)_{Jv}\, t2e \right] \ (\lambda_{dt\mu-t}) \tag{5}$$

$$\lambda_{dt\mu} = \lambda_{dt\mu-d} C_d + \lambda_{dt\mu-t} C_t \tag{6}$$

in the excited rotational–vibrational (Jv) state with quantum number $J=v=1$, where C_d and C_t are concentrations of deuterium and tritium nuclei, respectively. A strong resonance effect appears due to degeneracy in the excited state of the $dt\mu$ and the electron molecule complex. The rate of formation of the $dt\mu$ molecules has been found to depend strongly on temperature, density and on whether collision of the $t\mu$ atom occurs with a D_2 or a DT molecule (Bom et al., 2005; Faifman et al., 1996; Ackerbauer et al., 1999).

In fact, the radius of a muonic hydrogen ion ($dt\mu$) is much smaller (about ≈ 200 times) than a usual electron molecule, therefore the nuclei may tunnel the coulomb barrier with a high probability and fuse with a rate of ≈ 10^{12} sec^{-1} (Bogdanova et al., 1982). Resonant formation of the $dd\mu$ molecule at very low temperatures was observed in solid and liquid D_2 targets (Bogdanova et al., 1982).

Developed methods in this field are based on detailed three-body equations which provide a correct description of the quantum mechanical three-body systems (Takahashi & Takatsuka, 2006; Kilic, Karr & Hilico, 2004; Nielsen et al., 2001; Pahlavani, 2010). Theoretical study of muonic three-body system comprises different theoretical methods, e.g. variational

methods (Viviani et al. 1998; Frolov, 1993), Born-Oppenheimer approximation (Beckel et al., 1970; Kilic et al., 2004) and adiabatic expansion (Fano, 1981; Lin, 1995).

The Born-Oppenheimer approach assumes the nuclei to be infinitely heavy with respect to the negatively charged particle. It should be kept in mind that the following Born-Oppenheimer approach is the simplest solution to the three-body coulomb system. This approach is expected to be a poor approximation for calculations of muonic molecule eigenvalues. In this work, we calculate binding energies of the bound states of the *ddμ* muonic three-body system molecule using the adiabatic expansion method.

2. Adiabatic expansion approximation for the three-body system

The exact Hamiltonian that describes muonic three-body system can be shown by following relation:

$$H = -\frac{1}{2m_1}\nabla^2_{R_1} - \frac{1}{2m_2}\nabla^2_{R_2} - \frac{1}{2m_\mu}\nabla_{r_\mu} - \frac{z_1 z_\mu}{\left|\vec{r}_\mu - \vec{R}_1\right|} - \frac{z_2 z_\mu}{\left|\vec{r}_\mu - \vec{R}_2\right|} + \frac{z_1 z_2}{\left|\vec{R}_1 - \vec{R}_2\right|} \tag{7}$$

where *1* and *2* denote the two nuclei, their position is given by R_1 and R_2, and the muon coordinate is r_μ. The center of mass coordinate R_{CM} is given by

$$\vec{R}_{CM} = \frac{m_1\vec{R}_1 + m_2\vec{R}_2 + m_\mu\vec{r}_\mu}{m_1 + m_2 + m_\mu} \tag{8}$$

It is convenient to define Jacobi coordinate *r* and *R* as follow:

$$\vec{r} = \vec{r}_\mu - \frac{\vec{R}_1 + \vec{R}_2}{2} \tag{9}$$

$$\vec{R} = \vec{R}_2 - \vec{R}_1 \tag{10}$$

where *R* is the internuclear coordinate and *r* is the muon coordinate to midpoint between the two nuclei. In these coordinates (*R; r*), the Hamiltonian denoted by equation (7) is change to the following operator:

$$H = -\frac{1}{2M_T}\nabla^2_{R_{CM}} - \frac{1}{2M}(\nabla_R + \lambda / 2\nabla_r)^2 - \frac{1}{2m}\nabla^2_r - \frac{z_1 z_\mu}{\left|\vec{r}_\mu - \vec{R}_1\right|} - \frac{z_2 z_\mu}{\left|\vec{r}_\mu - \vec{R}_2\right|} + \frac{z_1 z_2}{\left|\vec{R}_1 - \vec{R}_2\right|} \tag{11}$$

where

$$M_T = m_1 + m_2 + m_\mu \tag{12}$$

$$\frac{1}{M} = \frac{1}{m_1} + \frac{1}{m_2} \tag{13}$$

$$\frac{1}{m} = \frac{1}{m_1 + m_2} + \frac{1}{m_\mu} \tag{14}$$

$$\lambda = \frac{m_1 - m_2}{m_1 + m_2} \tag{15}$$

After separation of variables, the non-relativistic Hamiltonian in units of $e = \hbar = m_\mu = 1$, can be given by

$$H(\hat{o}, R) = -\frac{1}{2M}(\nabla_R + \frac{\lambda}{2}\nabla_r)^2 - H_1 + \frac{z_1 z_2}{R} \tag{16}$$

where

$$H_1 = -\frac{1}{2m}\nabla^2{}_r - \frac{z_1}{r_1} - \frac{z_2}{r_2} \tag{17}$$

where $\hat{o}$ represent the five dimensional variable. We use the set $\hat{o} = (\Theta, \Phi, \xi, \eta, \varphi)$ where (Θ, Φ, φ) define the Euler rotation specifying the body-fixed frame with its unit vectors to coincide with the principal axes of the inertia tensor of a three-body system. The hyper-spheroidal coordinates ξ and η are easily expressed by the muon-nucleus distances r_1, r_2 and the internuclear distance R,

$$\eta = \frac{r_1 - r_2}{R} \quad (-1 \le \eta \le 1) \tag{18}$$

$$\xi = \frac{r_1 + r_2}{R} \quad (1 \le \xi < \infty) \tag{19}$$

The three-body Hamiltonian (16) commutes with the total angular momentum operator for the three particle system, J, its projection on z-axis, J_z, and the total parity operator, $P(R \to -R, r \to -r)$. Eigenfunctions of the Hamiltonian in the total angular momentum representation reads:

$$\Psi^p_{J,M}(R, r) = \sum_{m=-J}^{J} F^{Jp}_m(R, \xi, \eta) D^{Jp}_{Mm}(\Phi, \Theta, \varphi) \tag{20}$$

Adiabatic expansion of radial function, $F^{Jp}_m(R, \xi, \eta)$ is usually written in the form:

$$F^{Jp}_m(R, \xi, \eta) = \sum_{N=1}^{\infty} \sum_{l=0}^{N-1} \psi_{Nlm}(R; \xi, \eta) \chi^{Jp}_{Nlm}(R) R^{-1} + \sum_{l=0}^{\infty} \int_0^{\infty} dk \psi_{klm}(R; \xi, \eta) \chi^{Jp}_{klm}(R) R^{-1} \tag{21}$$

where $\chi^{Jp}_i(R)$ describe relative motion of the nuclei. Let us consider the Wigner function, $D^{Jp}_{Mm}(\Phi, \Theta, \varphi)$ which is the eigenstates of J^2, J_z and $R.J/R$ with the eigenvalues $J(J + 1)$, M and m (Davydov, 1973). It can be transformed under the inversion as follow:

$$PD^J_{Mm}(\Phi, \Theta, \varphi) = D^J_{Mm}(\Phi + \pi, \pi - \Theta, \pi - \varphi) = (-1)^{J-m} D^J_{M,-m}(\Phi, \Theta, \varphi) \tag{22}$$

If $m \neq 0$, the resultant Wigner functions would be different, and the angular functions consist both even and odd combinations. It is convenient to specify these combinations as follows:

$$D^{Jp}_{Mm}(\Phi,\Theta,\varphi) = \frac{\sqrt{2J+1}}{4\pi}\left[(-1)^m D^{J}_{Mm}(\Phi,\Theta,\varphi) + p(-1)^J D^{J}_{M,-m}(\Phi,\Theta,\varphi)\right] \tag{23}$$

where $p = \pm(-1)^J$ is the eigenvalue of the parity operator:

$$PD^{Jp}_{Mm} = pD^{Jp}_{Mm} \tag{24}$$

The functions presented in equation (23) (in bracket) are satisfying the following orthonormality condition:

$$\int_0^{\pi} \sin\Theta\, d\Theta \int_0^{2\pi} d\Phi \int_0^{2\pi} d\varphi\, [D^{Jp}_{Mm}(\Phi,\Theta,\varphi)]^* D^{J'p'}_{M'm'}(\Phi,\Theta,\varphi) = \delta_{JJ'}\delta_{pp'}\delta_{MM'}\delta_{mm'} \tag{25}$$

If $m = 0$, both the Wigner functions in (22) are reduced to the ordinary spherical function $Y_{JM}(\Theta,\Phi)$ so that the dependence of ' disappears and the angular functions satisfying the conditions (24) and (25) are:

$$D^{Jp}_{M,m=0}(\Phi,\Theta,\varphi) = \frac{Y_{JM}(\Theta,\Phi)}{\sqrt{2\pi}} \tag{26}$$

In this case the parity is unambiguously specified by the quantum number J: $p=+(-1)^J$. So, our basis functions have the following structure:

$$\Psi^{Jp}_{Mjm}(R,\Theta,\Phi,\xi,\eta,\varphi) = D^{Jp}_{Mm}(\Phi,\Theta,\varphi)\psi_{jm}(\xi,\eta;R)\frac{\chi^{Jp}_{jm}(R)}{R} \tag{27}$$

The wave functions $\Psi^{Jp}_{Mjm}(R,\Theta,\Phi,\xi,\eta,\varphi)$ describing reactions $h\mu+h$, h = (p, d, t) can be decomposed over the solutions $\psi_{jm}(\xi,\eta;R)$ of the Coulomb two-center problem. $\psi_{jm}(\xi,\eta;R)$ is the complete set of solutions of the Coulomb two-center problem, therefore

$$H_1\psi_i(\xi,\eta;R)F(\varphi) = E_i(R)\psi_i(\xi,\eta;R)F(\varphi) \tag{28}$$

describing the muon motion around fixed nuclei separated by a distance R. $E_i(R)$ is the energy of a muon in the state i as a function of R. Here we show how to separate the variables through the use of the ellipsoidal (or, prolate spheroidal) coordinates

$$x = \frac{R}{2}\sqrt{(\xi^2-1)(1-\eta^2)}\cos\varphi \tag{29}$$

$$y = \frac{R}{2}\sqrt{(\xi^2-1)(1-\eta^2)}\sin\varphi \tag{30}$$

$$z = \frac{R}{2}\xi\eta \tag{31}$$

Note that the coordinates ξ , η and φ are orthogonal, and we have the first fundamental form

$$ds^2 = dx^2 + dy^2 + dz^2 = h_\xi^2 d\xi^2 + h_\eta^2 d\eta^2 + h_\varphi^2 d\varphi^2 \tag{32}$$

where

$$h_\xi^2 = \left(\frac{\partial x}{\partial \xi}\right)^2 + \left(\frac{\partial y}{\partial \xi}\right)^2 + \left(\frac{\partial z}{\partial \xi}\right)^2 = \frac{R^2}{4}\left(\frac{1-\eta^2}{\xi^2-1}\right) \tag{33}$$

$$h_\eta^2 = \left(\frac{\partial x}{\partial \eta}\right)^2 + \left(\frac{\partial y}{\partial \eta}\right)^2 + \left(\frac{\partial z}{\partial \eta}\right)^2 = \frac{R^2}{4}\left(\frac{\xi^2-1}{1-\eta^2}\right) \tag{34}$$

$$h_\varphi^2 = \left(\frac{\partial x}{\partial \varphi}\right)^2 + \left(\frac{\partial y}{\partial \varphi}\right)^2 + \left(\frac{\partial z}{\partial \varphi}\right)^2 = \frac{R^2}{4}\left(\xi^2-1\right)\left(1-\eta^2\right) \tag{35}$$

Thus

$$\nabla^2 = \frac{1}{h_\xi h_\eta h_\varphi}\left[\frac{\partial}{\partial \xi}\left(\frac{h_\eta h_\varphi}{h_\xi}\frac{\partial}{\partial \xi}\right) + \frac{\partial}{\partial \eta}\left(\frac{h_\xi h_\varphi}{h_\eta}\frac{\partial}{\partial \eta}\right) + \frac{\partial}{\partial \varphi}\left(\frac{h_\xi h_\eta}{h_\varphi}\frac{\partial}{\partial \varphi}\right)\right]$$

$$= \frac{4}{R^2(\xi^2-\eta^2)}\left\{\frac{\partial}{\partial \xi}\left[(\xi^2-1)\frac{\partial}{\partial \xi}\right] + \frac{\partial}{\partial \eta}\left[(1-\eta^2)\frac{\partial}{\partial \eta}\right] + \frac{\xi^2-\eta^2}{(\xi^2-1)(1-\eta^2)}\frac{\partial^2}{\partial \varphi^2}\right\} \tag{36}$$

Note that through the coordinate transformation (29-31), we have

$$r_1 = \frac{R}{2}(\xi + \eta) \tag{37}$$

$$r_2 = \frac{R}{2}(\xi - \eta) \tag{38}$$

Writing the wave function as $\psi_i(\xi,\eta;R)F(\varphi) = G(\xi)H(\eta)F(\varphi)$ and changing the variable to spheroidal coordinates, equation (28) can be separated into following three one-dimensional equations:

$$\frac{d^2F(\varphi)}{d\varphi^2} + m^2F(\varphi) = 0 \tag{39}$$

$$\frac{d}{d\xi}\left[(\xi^2-1)\frac{dG(\xi)}{d\xi}\right] + \left(-A + \alpha q\xi - q^2\xi^2 - \frac{m^2}{\xi^2-1}\right)G(\xi) = 0 \tag{40}$$

$$\frac{d}{d\eta}\left[(1-\eta^2)\frac{dH(\eta)}{d\eta}\right] + \left(A - \beta q\eta - q^2\eta^2 - \frac{m^2}{1-\eta^2}\right)H(\eta) = 0 \tag{41}$$

where

$$\alpha = \frac{R}{q(z_1 + z_2)} \tag{42}$$

$$\beta = \frac{R}{q(z_1 - z_2)} \tag{43}$$

Note that A and q are unknown parameters and should be obtained from (40) and (41) as eigenvalues of the coupled system. Once A and q are obtained, then E can be obtained from q^2=-R^2E/2. By substitution of expression (8) into the Schrödinger equation with Hamiltonian (16) and after averaging over spherical angles (Θ,Φ) and the muon state, one obtains the radial equation

$$\frac{1}{2M}\frac{d^2}{dR^2}\chi_i^J(R) + \left[\varepsilon - V_{B-O}(R) - U_i(R) - \frac{J(J+1)}{2MR^2}\right]\chi_i^J(R) = 0 \tag{44}$$

where $\varepsilon = E - E_i(\infty)$ is the collision energy and E is the total energy of the system and $E_i(\infty)$ is the ground state energy of muonic atom. $V_{B-O} = E - E(\infty) + \frac{z_1 z_2}{R}$ is the potential corresponding to the Born-Oppenheimer approximation and $U_i = \left\langle i \middle| (H - H_1 - \frac{z_1 z_2}{R} \middle| i \right\rangle$ is the adiabatic correction. The adiabatic potential $V_{Ad}(R)$ is:

$$V_{Ad}(R) = V_{B-O}(R) + U_i(R) \tag{45}$$

The adiabatic potential $V_{Ad}(R)$ for the $(dd\mu)$ muonic three-body molecule is calculated in the adiabatic expansion method. The adiabatic potential curves and qualitatively similar for each of muonic molecules and are displayed for the $(dd\mu)$ muonic molecule in Figure 1. Results of the calculations of binding energies of the bound states (J,υ) of the $(dd\mu)$ muonic molecule are compared with the results of the other methods used in (Korobov et al., 1992; Kilic, Karr & Hilico, 2004) and are given in Table 1.

States (J,ν)	**Ad** (Pahlavani&Motevalli, 2008)	(Korobov et al., 1992)	(Kilic, Karr & Hilico, 2004)
(0,0)	325.06	325.0735	325.070540
(0,1)	35.79	35.8436	35.844227
(1,0)	226.62	226.6815	226.679792
(1,1)	1.73	1.97475	1.974985
(2,0)	86.20	86.4936	---

Table 1. Binding energies (eV) of the states (J,ν) for the $dd\mu$ muonic molecule.

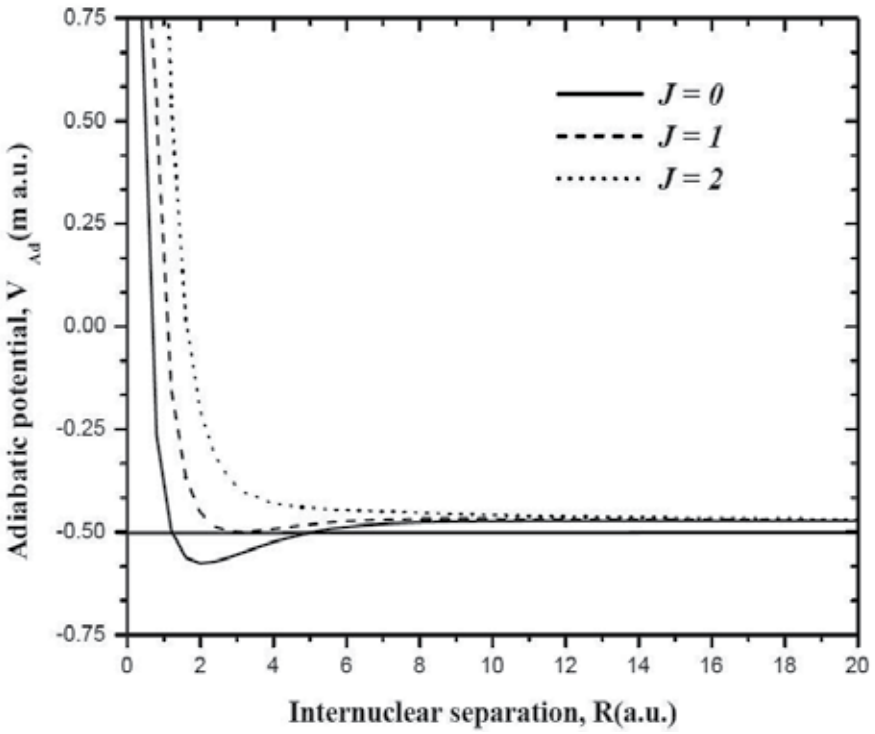


Fig. 1. Adiabatic potential curves, $V_{Ad}(R)$, corresponding to $dd\mu$ system (Pahlavani & Motevalli, 2008).

The calculated binding energies are in good agreement with the previous calculations by other authors using different methods.

3. Charge-asymmetric three-body system in hyper-spherical elliptic coordinate system

Study of the nuclear synthesis reaction $d-^3He$ at low collision energies (below *1 keV*) is of interest for its applications in nuclear and astrophysics (Belyaev et al.,1995). The relatively large energy gain as well as the lack of tritons in the initial and neutrons in the final channel makes this reaction a very attractive source of thermonuclear fusion energy.

The negatively charged energetic muons, after stopping in the $D-^3He$ mixture, fuse to d or ^3He in order to form the mesic atoms in excited states. After a sequence of cascade transitions lasting about 10^{-11} *sec* at Liquid Hydrogen Density (LHD), mesic atoms are formed in the ground state (Ponomarev, 1991; Breunlich et al., 1989; Czaplinski et al., 1996). The three-body molecules, $\left(^3He\mu d\right)_J^{++}$, are formed in collision of $(d\mu)$ atoms in ground state with helium atoms via so-called electron conversion process,

$$(d\mu)_{1s} + {}^3He \xrightarrow{\lambda_m^J} \left[\left(^3He\mu d\right)_{J\nu}^{++}, e\right]^+ + e \tag{46}$$

The molecule dissociates quickly with a rate of about $10^{12}\ sec^{-1}$ to the unbound ground state either by a well-known predissociation mechanism, via Auger transition or γ-emission processes

$$\left({}^3He\mu d\right)_J \xrightarrow{\lambda_p^J} d + \left(He\mu\right)_{1s} \tag{47}$$

$$\left({}^3He\mu d\right)_J \xrightarrow{\lambda_A^J} d + \left({}^3He\mu\right)_{1s} + e \tag{48}$$

$$\left({}^3He\mu d\right)_J \xrightarrow{\lambda_\gamma^J} d + \left({}^3He\mu\right)_{1s} + \gamma \tag{49}$$

resulting a hydrogen nucleus and a mesic helium atom. This mechanism leads to transfer rates of the order $10^{8}\ sec^{-1}$. The asymmetric-charged ${}^3He\mu d$ molecule undergo nuclear fusion via two different channels,

$$\left({}^3He\mu d\right)_J \rightarrow \alpha + p + \mu + 14.64\ MeV \tag{50}$$

$$\left({}^3He\mu d\right)_J \rightarrow {}^5Li\mu + \gamma + 16.7\ MeV \tag{51}$$

The muon is released after the fusion and can proceed to cause another fusion. Thus the muon works as a catalyst and this cycle can be repeated many times during its lifetime.

To test the stability of the mesic ${}^3He\mu d$ system, we consider only Coulomb interaction between particles. As a starting point, we employ the aim of hyperspherical method to solve the multi-dimensional Schrödinger equation numerically for this three-body system.

$$H\Psi = (T + V)\Psi = E\Psi \tag{52}$$

The wave function, Ψ, can be constructed explicitly by exploiting a specific representation, namely, the hyper-spherical adiabatic expansion method. Here, T is the kinetic energy in its enter-of-mass coordinate frame, V is the potential energy, and E is the total energy of the system. We briefly discuss the general structure of the method and formulate its basic equations for a three-body system in hyper-spherical elliptic coordinates. The Hamiltonian of this molecule in Jacobian coordinates *(R, r)* can be shown by the following equation (Gusev et al., 1990; Stuchi et al., 2000)

$$H = T + V = \left[\frac{1}{2M_i}\Delta(\vec{R}_i) - \frac{1}{2m_i}\Delta(\vec{r}_i)\right] + V \tag{53}$$

where M_i and m_i are reduced masses. It is convenient to define mass-scaled Jacobian vectors, ($\vec{x}_i$ and $\vec{y}_i$),

$$\vec{x}_i = \sqrt{\frac{M_i}{\mu}}\vec{R}_i \tag{54}$$

$$\vec{y}_i = \sqrt{\frac{m_i}{\mu}}\vec{r}_i \tag{55}$$

Therefore the kinetic energy of the system can be rewritten as,

$$T = -\frac{1}{2\mu}\left[\Delta(\vec{x}_i) - \Delta(\vec{y}_i)\right] \tag{56}$$

In this relation, $\mu = (m_d\, m_{^3He}\, m_\mu / m_{tot})^{1/2}$ is an arbitrary coefficient with dimension of mass and index *i=1,2,3* refer to a set of Jacobian coordinates ($\vec{x}_i$, $\vec{y}_i$). The transformation of Jacobian coordinates (e.g. ($\vec{x}_1$, $\vec{y}_1$)) in to another set (e.g. ($\vec{x}_2$, $\vec{y}_2$)), can be done as follows:

$$\vec{x}_2 = -\vec{x}_1 \cos\gamma - \vec{y}_1 \sin\gamma \tag{57}$$

$$\vec{y}_2 = \vec{x}_1 \sin\gamma - \vec{y}_1 \cos\gamma \tag{58}$$

where can be regarded as a rotational parameter which is shown by,

$$\gamma = Arc\tan\left[\sqrt{\frac{m_\mu(m_d + m_{^3He} + m_\mu)}{m_d\, m_{^3He}}}\right]; \qquad 0 \le \gamma \le \frac{\pi}{2} \tag{59}$$

It is convenient to calculate these sets of Jacobian coordinates ($\vec{x}_i$, $\vec{y}_i$) for mesic three-body, $^3He\mu d$ system. These three sets should be used as coordinates in configuration space. Therefore this system contains six dimensions (*d=6*). In hyperspherical coordinates, (ρ,Ω), $\rho = \sqrt{x^2 + y^2}$ represents the size of the system and $\Omega = (\Omega_s, \Omega_0)$, consist of five variables, where Ω_s denote a set of two angles defining the shape of the system and Ω_0 refer to a set of three angles defining the overall orientation of the three-body system. The Hamiltonian, in this coordinate system, will take the following form:

$$H = -\frac{1}{2\mu}\left[\left(\rho^{-5}\frac{\partial}{\partial\rho}\rho^5\frac{\partial}{\partial\rho}\right) - \frac{\Lambda^2}{\rho^2}\right] + V \tag{60}$$

where Λ^2 is regarded as the square of general angular momentum operator. Our aim is to solve the eigenvalue equation $H\Psi(\rho,\Omega) = E\Psi(\rho,\Omega)$ in the adiabatic expansion method. The idea of adiabatic separability between the hyper-radius ρ and the hyper-angular variables Ω in three-body systems was first exploited by Macek (Macek, 1968) for studying doubly excited states of the Helium atom. The wave equation of the system in this method can be defined by:

$$\Psi(\rho,\Omega) = \rho^{-\frac{5}{2}}\sum_\upsilon F_\upsilon(\rho)\Phi_\upsilon(\rho,\Omega) \tag{61}$$

Here the quantum number, *j* characterizes a channel function, the radial functions, $F_j(\rho)$ satisfy the system of coupled ordinary differential equations and $\phi_j(\Omega;\rho)$ are angular functions. For any value of ρ, these functions form a set of complete orthogonal basis which satisfy the following relation:

$$\left[H_{ad} - U_\nu(\rho)\right]\phi_\nu(\Omega;\rho) = 0 \tag{62}$$

In this relation, H_{ad} is the adiabatic Hamiltonian which is defined by,

$$H_{ad} = \frac{1}{2}\Lambda^2 + \rho C \tag{63}$$

where $C = \rho V$ is the effective charge of the system. In the first step we try to solve the differential equation (62), which contains coordinate ρ as a parameter. The hyper-spherical elliptic coordinates (η,ξ) on S (projection of the hyper-sphere $\rho = const.$ onto shape space) are induced by conical coordinates on its 3D image. The hyper-spherical elliptic coordinates (η,ξ) are defined in the following intervals (Tolstikhin et al., 1995)

$$-2\gamma \le \eta \le 2\gamma$$

$$2\gamma \le \zeta \le 2\pi - 2\gamma \tag{64}$$

The definition of (η,ξ) resembles the representation of plane elliptic coordinates. In order to rewrite Eq. (62) in a new set of coordinates, (η,ξ), it is necessary to define the square of general angular momentum operator, Λ^2 in this set of coordinates,

$$\Lambda^2 = \frac{-16}{\cos\eta - \cos\xi}\left[\frac{\partial}{\partial\eta}\left(\cos\eta - \cos 2\gamma\right)\frac{\partial}{\partial\eta} + \frac{\partial}{\partial\xi}\left(\cos 2\gamma - \cos\xi\right)\frac{\partial}{\partial\xi}\right]$$

$$+ \frac{4m^2\sin^2 2\gamma}{\cos\eta - \cos\xi}\left[\frac{1}{\cos\eta - \cos 2\gamma} + \frac{1}{\cos 2\gamma - \cos\xi}\right] \tag{65}$$

where m is azimuthal quantum number which is the projection of the general angular momentum along body-fixed axis. The potential energy, V, of the system is the sum of three inter-particle Coulombian interactions potential,

$$V = \frac{Z_1 Z_2 e^2}{r_{12}} + \frac{Z_2 Z_3 e^2}{r_{23}} + \frac{Z_3 Z_1 e^2}{r_{31}} = \frac{C(\eta,\xi)}{\rho} \tag{66}$$

As it was mentioned earlier, $C(\eta,\xi) = \rho V$, is the effective charge. The inter-particle distances r_{12}, r_{23} and r_{31} are simply defined by the following relations:

$$r_{12} = \frac{\rho}{\sqrt{2\mu_3}}\sqrt{1 + d^+\cos\left(\frac{\eta}{2}\right)\cos\left(\frac{\xi}{2}\right) - d^-\sin\left(\frac{\eta}{2}\right)\sin\left(\frac{\xi}{2}\right)}$$

$$r_{23} = \frac{\rho}{\sqrt{\mu_1}}\sin\left(\frac{\eta+\xi}{4}\right) \tag{67}$$

$$r_{31} = \frac{\rho}{\sqrt{\mu_2}}\sin\left(\frac{\eta-\xi}{4}\right)$$

where $d^+ = 1 + \frac{2m_\mu}{m_d + m_{^3He}}$ and $d^- = \frac{m_d - m_{^3He}}{m_d + m_{^3He}}$ are mass related constants respectively. The effective charge *C* of $^3He\mu d$ molecule as a function of variables η and ξ is shown in Fig. 2. The steep spike at $(\eta,\xi) = (-2\gamma, 2\gamma), (2\gamma, 2\gamma)$, corresponds to the strong attractive coulomb singularities of effective charge *C* and associated to collisions in the pairs $d\mu$ and $^3He\mu$, when muon is very close to the nucleus. The singular Coulomb repulsion between two positively charged particles, are represented by the repulsive wall at the neighborhoods of the $(\eta,\xi) = (0, 2\pi - 2\gamma)$.

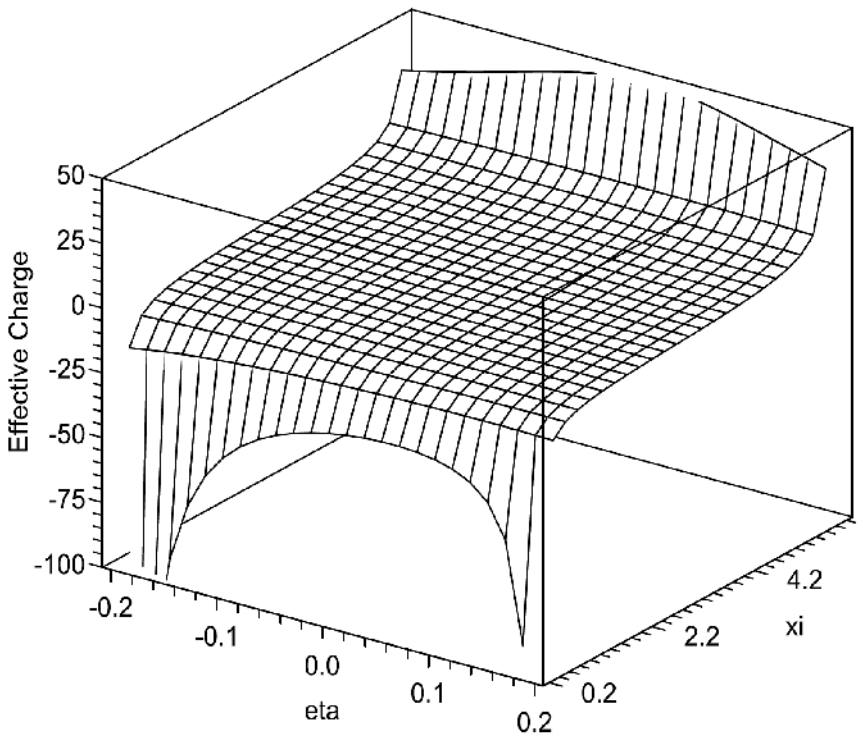


Fig. 2. Variation of effective charge *C* as a function of hyper-angular variables $-2\gamma \le \eta \le 2\gamma$ and $2\gamma \le \zeta \le 2\pi - 2\gamma$ for the $^3He\mu d$ molecule (Pahlavani, Sadeghi, Motevalli & Aqabaei, 2010).

By substituting Eqs. (65) and (66) for Λ^2 and *V* into Eq. (62), we obtain a differential equation for adiabatic Hamiltonian that should be solve with appropriate boundary conditions. In the case, for infinite small values of ρ, the solutions of adiabatic Hamiltonian (62) can be constructed in the following form (Pahlavani & Motevalli, 2008):

$$\Phi(\eta,\xi) = N\,X(\eta)Y(\xi) \tag{68}$$

where *N* is the normalization parameter. With some mathematical simplification, one obtains the following set of ordinary differential equations:

$$\left[L^{\eta}-\frac{2m^2\cos^2 2\gamma}{\cos\gamma-\cos 2\gamma}+U(\cos\eta-\cos 2\gamma)+A\right]+X(\eta)=0 \tag{69}$$

$$\left[L^{\xi}-\frac{2m^2\cos^2 2\gamma}{\cos 2\gamma-\cos\xi}+U(\cos 2\eta-\cos\xi)-A\right]+Y(\xi)=0 \tag{70}$$

where A is the separation constant and the one-dimensional derivative operators, L^{η} and L^{ξ} are defined as,

$$L^{\eta}=8\frac{d}{d\eta}(\cos\eta-\cos 2\gamma)\frac{d}{d\eta} \tag{71}$$

$$L^{\xi}=8\frac{d}{d\xi}(\cos 2\gamma-\cos\xi)\frac{d}{d\xi} \tag{72}$$

The resultant equations are subject to the regularity of boundary conditions and can be satisfied only for certain values of A and U. The method for solving these set of differential equations, are very similar to those equations which we presented in our previous work (Pahlavani & Motevalli, 2008), when we have studied the motion of muon in the two-center Coulomb problem in prolate spheroidal coordinate system for the symmetric mesic system $dd\mu$. By solving these equations, one obtains the functions $\Phi_{n(\eta)n(\xi)}(\eta,\xi)$, where $n(\eta)$ and $n(\xi)$ are quantum numbers correspond to number of zeros of the functions $X(\eta)$ and $Y(\xi)$ that appeared in Eqs. (69) and (70). These functions form a set of complete bases and satisfy the following normalization condition,

$$\int_{-2\gamma}^{2\gamma}\int_{2\gamma}^{2\pi-2\gamma}\Phi^{*}_{n(\eta)n(\xi)}(\eta,\xi)\Phi_{n'(\eta)n'(\xi)}(\eta,\xi)\,ds=\delta_{n(\eta)n(\xi)}\,\delta_{n'(\eta)n'(\xi)} \tag{73}$$

where ds is the surface element that can be defined by,

$$ds=\frac{\pi^2}{4\cos 2\gamma}(\cos\eta-\cos\xi)\,d\eta\,d\xi \tag{74}$$

The results of the calculations are displayed graphically in Fig. 3. The normalization factor $N_{n(\eta)n(\xi)}$, is a function of the rotational parameter, at different quantum numbers $n(\eta)$ and $n(\xi)$. Calculated values of adiabatic potential $U(\rho)$ as a function of hyper-radius ρ have been shown in Fig. 4. By substituting Eq. (61) into Eq. (52), one can obtain the following set of ordinary differential equations for radial functions $F_j(\rho)$:

$$\left\{\frac{d^2}{d\rho^2}+2\mu\left[E-U(\rho)-\frac{15}{8\mu\rho^2}\right]\right\}F_{\nu}(\rho)+\sum_{\kappa}W_{\nu\kappa}F_{\kappa}(\rho)=0 \tag{75}$$

where $U_j(\rho)$ is the adiabatic potential and the operator $W_{j\kappa}(\rho)$ has the following form:

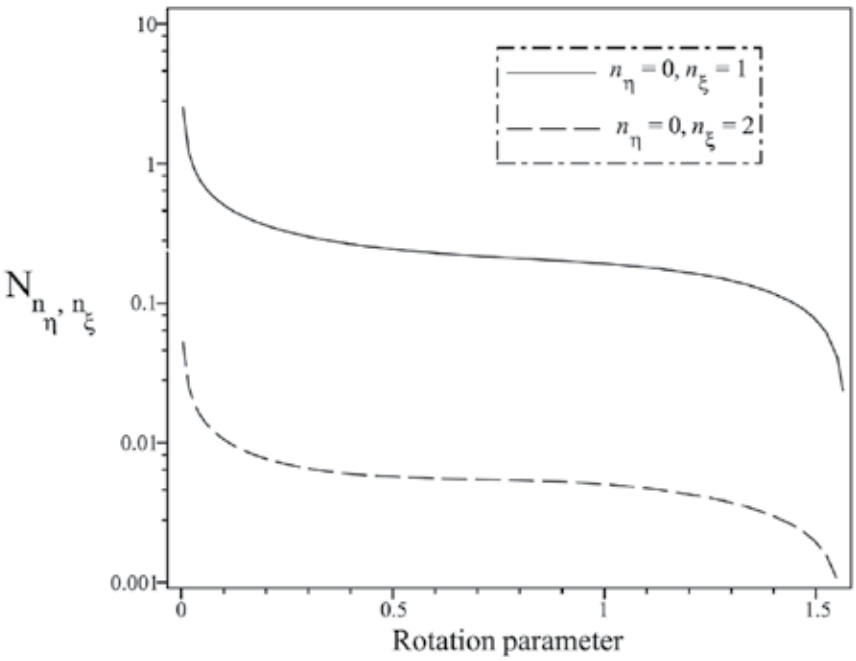


Fig. 3. Variation of normalization parameter $N_{n(\eta)n(\xi)}$ for the cases: $(n_\eta, n_\eta) = (0,1)$ and $(n_\eta, n_\eta) = (0,2)$ for the $^3He\mu d$ molecule (Pahlavani, Sadeghi, Motevalli & Aqabaei, 2010).

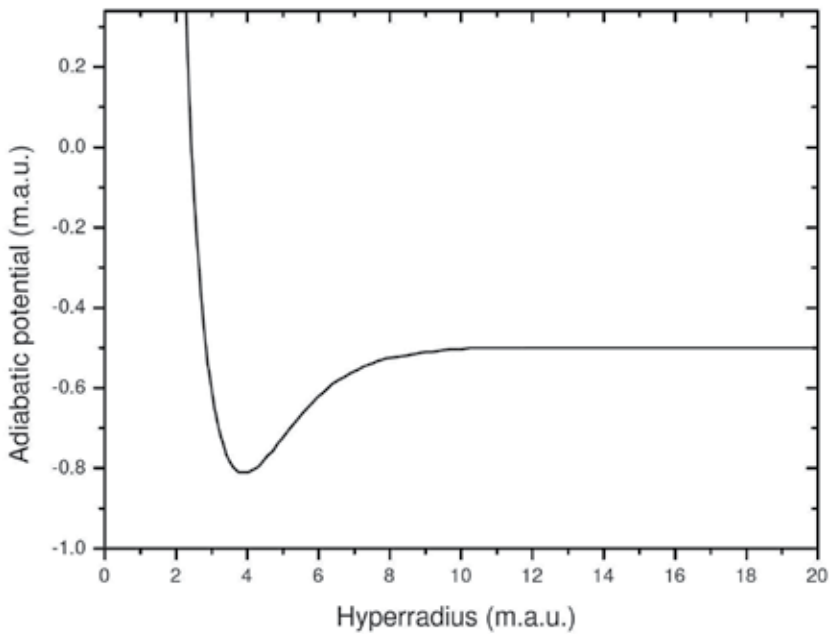


Fig. 4. The adiabatic potential as a function of hyper-radius coordinate for the $^3He\mu d$ molecule (Pahlavani, Sadeghi, Motevalli & Aqabaei, 2010).

$$W_{\nu\kappa}(\rho) = 2R_{\nu\kappa}(\rho)\frac{d}{d\rho} + S_{\nu\kappa}(\rho) \tag{76}$$

The above equation is a set of differential equations coupled by the following nonadiabatic terms:

$$R_{\nu\kappa}(\rho) = \left\langle \phi_\nu(\Omega;\rho) \middle| \frac{\partial}{\partial\rho}\phi_\kappa(\Omega;\rho) \right\rangle \tag{77}$$

$$S_{\nu\kappa}(\rho) = \left\langle \phi_\nu(\Omega;\rho) \middle| \frac{\partial^2}{\partial\rho^2}\phi_\kappa(\Omega;\rho) \right\rangle \tag{78}$$

where the brackets represent integration over the angular variables Ω. The hyperspherical adiabatic approximation amounts to retaining only one term in Eq. (61) (Macek, 1968). Then the radial function, $F_j(\rho)$ satisfies the following differential equation

$$\left\{ \frac{d^2}{d\rho^2} + 2\mu\left[E - U(\rho) - \frac{15}{8\mu\rho^2} \right] + W_{\nu\nu} \right\} F_\nu(\rho) = 0 \tag{79}$$

This approximation turns out to be surprisingly accurate in the sense that in many situations the non adiabatic couplings in Eq. (75) are rather weak. Subject to this reality and considering appropriate boundary conditions, we obtain solutions of the differential equation (79) numerically. Finally, the calculated values of the binding energies of the bound states (J,υ) for the $^3He\mu d$ system are compared with available data obtained using other methods in Table 2.

States (J,ν)	(Gershtein & Gusev, 1993)	(Kravtsov et al., 1993)	(Hara & Ishihara, 1989)	(Pahlavani et al., 2010)
(0,0)	69.96	70.6	70.74	70.879
(1,0)	46.75	48.2	47.90	48.391
(2,0)	---	9.6	---	9.346

Table 2. Binding energy E_B *(eV)* of the bound states (J,ν) (the quantum numbers of rotational-vibrational state) for the $^3He\mu d$ molecule.

The Born–Oppenheimer approach assumes the nuclei to be infinitely heavy with respect to the negatively charged particle. It should be kept in mind that Born–Oppenheimer approach is the simplest solution to the three-body Coulomb system. Usually, the most accurate results for the ground state energy levels of mesic three-body molecule were obtained from variational calculations. Comparison of our result for $J = 0$ with the ones obtained by the available variational calculation (Bogdanova et al., 1982) indicates difference that dose not exceed *0.2%*. One can conclude that this fact supports the validity of adiabatic expansion in hyper-spherical elliptic coordinates method which have been used.

4. Muon stripping in the muon catalyzed fusion

The sticking of muons to alpha particles after fusion is an unwanted process and eliminates muons from the chain of fusion reactions. This process is the main loss mechanism in the μCF. The probability of forming a muonic helium ion is called initial sticking probability $\omega_S^0 (= 0.912\%)$ (Hu, Hale & Cohen, 1994). Where muonic helium ions are formed with an energy of $E_{\alpha\mu}^{in} = 3.47\,MeV$ $(v_{\alpha\mu}^{in} = 5.83\,a.u.)$ then are slowed down toward thermal energy by collision with the surrounding D_2 and DT molecules (Jones, 1986). During the same time, as long as the kinetic energy exceeds the appropriate threshold $(E_{\alpha\mu}^{th} \approx 10\,KeV)$, the $a\mu$ ion can be stripped as a result of collisions. This process is referred to as reactivation and final sticking fraction, ω_s that conventionally related to the initial sticking fraction by $\omega_S = \omega_S^0 (1 - R)$. The reactivation coefficient, R depends upon the stopping power of the media and several important cross sections. Stripping process can occur through several channels. Collisions of the $(a\mu)_{1s}$ ions with the surrounding D_2 and DT molecules during the slowing down process can result in $a\mu$ charge transfer, ionization or excitation of the discrete $a\mu$ levels. Stripping (charge transfer plus ionization) can also happen from the $a\mu$ which is the results of the sticking or collisional excitation processes.

The kinetic of reactivation is described by the various rates in a set of coupled differential equations. The fraction of stripped muonic helium ions in terms of population probabilities can be written as

$$\frac{dP_{strip}(t)}{dt} = \sum_i \lambda_{strip}^{(i)}\left(v(t)\right) P_i(t)\,. \tag{80}$$

where $\lambda_{strip}^{(i)}\left(v(t)\right)$ are velocity-dependent stripping rates from the individual energy levels and $P_i(t)$ are the time dependent population probabilities for the state i of muonic helium ion. The time-dependent population probabilities for the state i of the muonic helium ion are determined by

$$\frac{dP_i(t)}{dt} = \lambda_{pop}^{(i)} - P_i(t)\,\lambda_{depop}^{(i)} \tag{81}$$

where $\lambda_{pop}^{(i)}$ and $\lambda_{depop}^{(i)}$ are the rates of populating and de-populating probability of state i, respectively. These rates can be given by the following relations:

$$\begin{aligned}\lambda_{pop}^{(i)} = &\sum_{i'(n_{i'} > n_i)} \left(\lambda_{Au}^{(i' \to i)} + \lambda_{ra}^{(i' \to i)} + \lambda_{de-ex}^{(i' \to i)}\right) P_{i'}(t) \\ &+ \sum_{i'(n_{i'} < n_i)} \lambda_{ex}^{(i' \to i)} P_{i'}(t) + \sum_{i'(n_{i'} = n_i)} \lambda_{Stark}^{(i' \to i)} P_{i'}(t)\end{aligned} \tag{82}$$

$$\lambda_{depop}^{(i)} = \lambda_{strip}^{(i)} + \sum_{i'(n_{i'} < n_i)} \left(\lambda_{Au}^{(i \to i')} + \lambda_{ra}^{(i \to i')} + \lambda_{de-ex}^{(i \to i')}\right) + \sum_{i'(n_{i'} > n_i)} \lambda_{ex}^{(i \to i')} + \sum_{i'(n_{i'} = n_i)} \lambda_{Stark}^{(i \to i')} \tag{83}$$

where λ_{Au}, λ_{ra}, λ_{de-ex}, λ_{ex}, λ_{Stark} and λ_{strip} are the Auger de-excitation, radiative, Coulomb de-excitation, Coulomb excitation, Stark mixing and striping rates, respectively. In general, λ is given by

$$\lambda = N\sigma v \quad (\sec^{-1}) \tag{84}$$

where N, v and σ are density of surrounded media, relative velocity and cross section for all processes under consideration, respectively. The time and velocity dependence in Eq. (80) are coupled through the energy-loss equation for muonic helium ion given by

$$\frac{dE_{\alpha\mu}}{dx} = -v_{\alpha\mu} S\left(E_{\alpha\mu}\right) = -\left(\frac{2E_{\alpha\mu}}{m_{\alpha\mu}}\right)^{1/2} S\left(E_{\alpha\mu}\right) \tag{85}$$

where $S = -dE/dx$ is the stopping power of the surrounding media and $m_{\alpha\mu}$ is the mass of muonic helium ion. The initial conditions are: $E_{\alpha\mu}(0) = E_{\alpha\mu}^{in} = 3.47\,MeV$, $P_{st}(0) = 0$ and the initial values of populated levels are determined by the initial sticking, $P_i(0) = \omega_s^0(i) / \omega_s^0$. The populations $P_i(t)$ for $n = 1,2,...,6$ and the l sublevels are treated in detail for $n < 4$. The reactivation coefficient R is equivalent to the stripping fraction $P_{st}(t)$ at $t \to \infty$. The intensity of X-ray transition in muonic helium ion is another quantity which can be measured experimentally and calculated along with reactivation coefficient (R). Muons in excited levels of the $\alpha\mu^+$ may de-excite under X-ray emission. The X-ray spectrum depends not only on the initial sticking in the atomic levels and the reactivation of the muon but also on intra-atomic transitions due to inelastic collisions, internal and external Auger effect and Stark mixing. The photon intensity per sticking event is calculated using

$$\frac{d\gamma_{n'\to n}}{dt} = \sum_{i'(n_{i'}=n')} \sum_{i(n_i=n)} \lambda_{ra}^{(i'\to i)} P_{i'}(t) \tag{86}$$

The number of X-ray photons emitted per fusion is the most useful quantity that can be measured experimentally. The X-ray yields for the $n' \to n$ transition is given by

$$Y(n' \to n) = \gamma_{n'\to n}\omega_s^{\circ} \tag{87}$$

The calculation for muon stripping probability from $a\mu^+$ and the intensity of X-ray transitions have been done by solving a set of coupled differential equations numerically. The time-dependent population probabilities $P_i(t)$ for *1s, 2s, 2p, 3s, 3p, 3d* are shown in Fig. 5 for a deuterium–tritium target at density φ=*1.2 L.H.D* (*L.H.D*≡ *Liquid Hydrogen Density = 4.25* × 10^{22} *atoms/cm*^{3}). The initial populations of all excited states are seen to drop to *0* during the stopping time, and only *1s* orbital stays occupied.

The time-dependent stripping fraction, $P_{st}(t)$ and surviving fraction of the initial kinetic energy, E/E_0 are shown in Fig. 6. Slowing down of $a\mu^+$ from $v_{\alpha\mu} = 5.83\,a.u.$ to $v_{\alpha\mu} \approx 1\,a.u.$ takes about $t_{stop} \approx 4\times10^{-11}$ (sec). This time is longer than the lifetime of the excited $a\mu^+$ states so that the cascade of $a\mu^+$ actually takes place during the slowing down process. The calculated reactivation coefficient, final sticking and the average number of X-

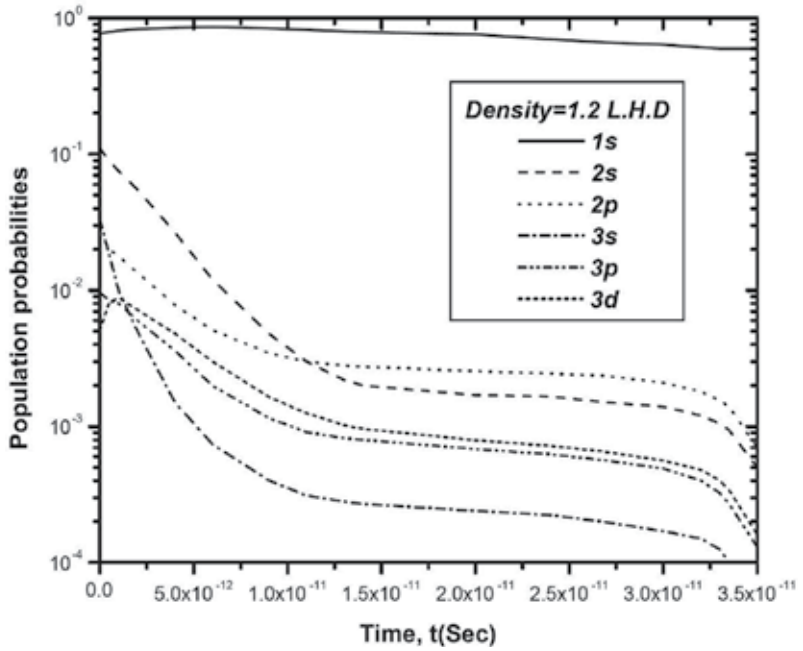


Fig. 5. The population probabilities $P_i(t)$ as a function of time in a *D–T* target at density φ=*1.2 L.H.D* (Pahlavani & Motevalli, 2008).

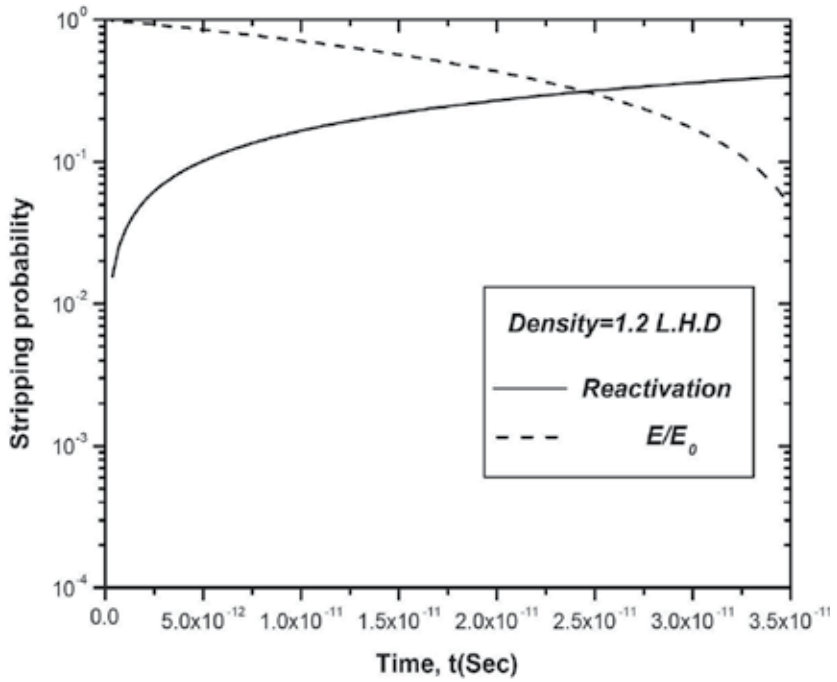


Fig. 6. Stripping fraction, *R* (heavy solid curve), surviving fraction of initial kinetic energy, E/E_0 (dashed curve) in a *D–T* target at density φ = *1.2 L.H.D* (Pahlavani & Motevalli, 2008).

rays per sticking $(K_\alpha, K_\beta, K_\gamma)$ as a function of density are shown in Fig. 7 for $\phi < 4LHD$. The most K_α radiation actually emitted by $a\mu^+$ atoms that formed in the ground state. If $a\mu^+$ is formed in the $2p$ state more than one $K_\alpha(2p \to 1s)$ X-ray expected per sticking. Our theoretical results for stripping are compared in Table 3 with other theoretical and experimental data. It is evident that experimental results of the effective sticking probability are smaller than the theoretical calculations, however, our results agree well with experiment.

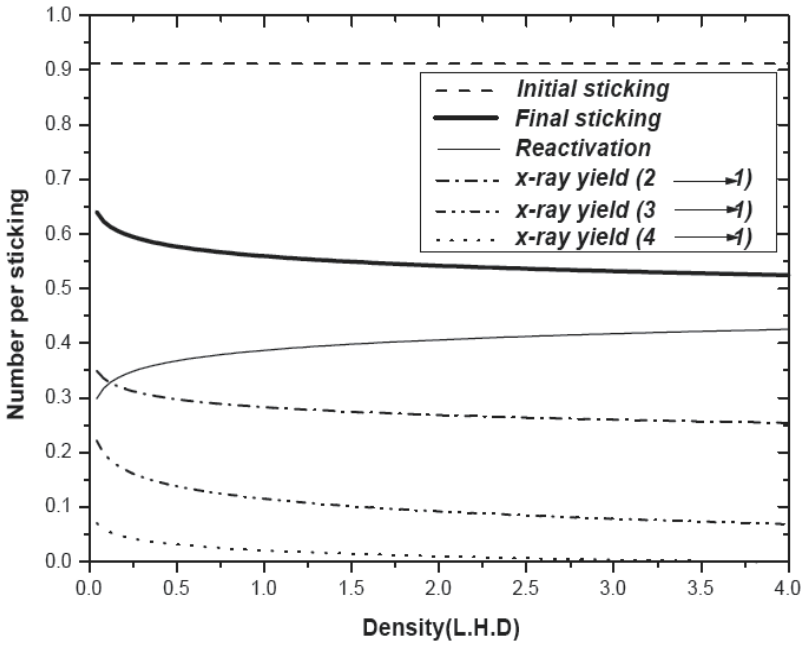


Fig. 7. The density dependence of initial sticking, $\omega_s^0(\%)$, final sticking, $\omega_s(\%)$, reactivation coefficient, R and K-series X-ray per sticking $(K_\alpha, K_\beta, K_\gamma)$ for dtμ fusion (K_β and K_γ multiplied by factor 3) (Pahlavani, Motevalli, 2008).

The density dependence of probability of muon reactivation, final sticking coefficient and intensity of X-rays emitted by muonic helium ion have been studied numerically. In order to do this, we consider all reactions that separate muon from muonic helium ion, namely coulomb excitation and de-excitation, ionization, charge transfer, Stark mixing, radiative transitions and Auger de-excitation. Using a set of coupled differential equations, the time dependence of muon reactivation coefficient (R) and surviving fraction of the initial kinematic energy of $a\mu^+$ (E/E_0) in the D–T mixture for different fuel density have been calculated. The measurement of muonic helium ion X-ray provides an independent method to test our knowledge about muon reactivation and sticking. Results based on our calculation shown that the muon reactivation increases when the average number of X-rays per sticking reduces with increasing density. Our calculated results are in good agreement with available experimental data (Ishida, Nagamine et al., 1999; Petitjean et al., 1993; Breunlich et al., 1987; Bossy et al., 1987; Jones, Taylor & Andeson, 1993; Nagamine et al., 1993; Ishida et al., 2001; Petitjean, 2001) at all.

Source	Reactivation	Final sticking (%)
$Density = 1.2\,LHD$		
(Pahlavani & Motevalli, 2008)	0.391	0.555
(Markushin, 1988)	---	0.57 ± 0.07
(Refelski et al.,1989)	0.36	0.57
(Takahashi, 1987)	0.248	0.664
(Cohen, Hale, Hu, 1996)	---	0.59
Experimets		
PSI (Bossy et al., 1987)	---	0.39 ± 0.10
PSI (Breunlich et al., 1987)	---	0.45 ± 0.05
PSI (Petitjean et al., 1993)	---	$0.48 \pm 0.02 \pm 0.04$
LAMPF (Jones, Taylor & Andeson, 1993)	---	$0.43 \pm 0.05 \pm 0.06$
KEK (Nagamine et al., 1993)	---	0.51 ± 0.004
RIKEN-RAL, Liquid (Ishida, Nagamine et al., 1999)	---	0.434 ± 0.030
RIKEN-RAL, Solid (Ishida, Nagamine et al., 1999)		0.421 ± 0.030
RIKEN (Ishida et al., 2001)	---	0.532 ± 0.030
$Density = 1.45\,LHD$		
(Pahlavani, Motevalli, 2008)	0.395	0.551
Experiment		
PSI (Petitjean, 2001)	---	0.505 ± 0.029

Table 3. The reactivation coefficient, R and final sticking, $\omega_s(\%)$ for muonic helium ion in different densities.

5. Conclusions

The quantum-mechanical three-body problem plays an important role in modern physics by providing an appropriate description of three-particle systems in presence of Coulomb and nuclear forces. Developed methods in this field are based on detailed three-body equations which provide a correct description of the quantum mechanical three-body systems (Takahashi & Takatsuka, 2006; Kilic, Karr & Hilico, 2004; Nielsen et al., 2001; Pahlavani, 2010). Theoretical study of muonic three-body system comprises different theoretical methods, e.g. variational methods (Viviani et al. 1998; Frolov, 1993), Born-Oppenheimer approximation (Beckel et al., 1970; Kilic et al., 2004) and adiabatic expansion (Fano, 1981; Lin, 1995). In this investigation, we presented an appropriate method that enables us to study the solutions of Schrodinger equation for $^3He\mu d$ system. The adiabatic expansion in hyper-spherical elliptic coordinates has shown a good approach for calculating the adiabatic potential. Fast convergent of this method led us to obtain precise results for the existence of the bound states in $^3He\mu d$ three-body molecule. The obtained results for the adiabatic potential of this system are comparable with results gathered from other approximation methods.

The corresponding eigenvalue problem has been solved and the binding energy of this system is calculated. The obtained results agreed with the expected values of various theoretical methods. This approach can be applied for other three-body systems with variety of masses and charges. The obtained results are of significant importance for experimental and theoretical investigation of $d-{}^3He$ nuclear fusion especially at low collision energies.

In section 4, the obtained results show that the muon cycle coefficient increases almost slowly with the density of deuterium and tritium mixture. The energy required to produce a muon estimated to be about *5000 MeV*. Since each deuterium and tritium fusion generates *17.6 MeV*, we see that the number of catalysis reactions by a muon should be about 285 to reach the scientific break-even (*1/3* of the commercial break-even). The break-even point is reached when the fusion process generates as much energy as was initially put in (i.e., the energy output equals the energy input). The output energy of the number of catalysis reactions by a muon in it's lifetime (τ_μ = *2.197 μsec*), is much smaller than the input energy required to produce a muon. Therefore, a fusion energy system based on the muon catalyzed fusion in deuterium and tritium fuel seems to be viable at plasma conditions with fuel densities about *100* times of L.H.D.

6. References

Ackerbauer, P. et al. (1999). *Nucl. Phys. A*, Vol. 652, pp. 311
Beckel, C.L. et al. (1970). *Chem. Phys.*, Vol. 53, pp. 3681
Belyaev, V.B. et al. (1995). *Nucleonica,* Vol. 40, pp. 3
Bogdanova, L. et al. (1982). *Zh. Eksp. Teor. Fiz.* Vol. 83, pp. 1615
Bom, V.R. et al. (2005). *J. Exper. Theor. Phys.,* Vol. 100, pp. 663
Bossy, H. et al. (1987). *Phys. Rev. Lett.*, Vol. 59, pp. 2864
Breunlich, W.H. et al. (1989). *Annu. Rev. Nucl. Part. Sci.*, Vol. 39, pp. 311
Breunlich, W.H. et al. (1987). *Phys. Rev. Lett.,* Vol. 58, pp. 329
Bystritsky, V.M. et al. (2006). *Eur. Phys. J. D,* Vol. 38, pp. 455
Bystritsky, V.M. et al. (2005). *Phys. Rev. A,* Vol. 71, pp. 032723
Bystritsky, V.M. et al. (2000). *Eur. Phys. J. D,* Vol. 8, pp. 75
Bystritsky, V.M. et al. (1980). *Phys. Lett. B* Vol. 94, pp. 476
Caffery, A.J. et al. (1987). *Muon Catal. Fusion*, Vol. 1, pp. 53
Cohen, J.S.; Hale, G. M. & Hu, C-Y. (1996). *Hyp. Interact.*, Vol. 101/102, pp. 349
Czaplinski, W. et al. (1994). *Phys. Rev. A*, Vol. 50, pp. 525
Czaplinski, W. et al. (1996). *Z. Phys. D*, Vol. 37, pp. 283
Davydov, A.S., *Quantum Mechanics*, 2nd edn. (Science, Moscow, 1973), Sects. 43, 44, 109, 118 (in Russian)
Demin, D.L. et al. (1996). *Hyp. Interact.* Vol. 101/102, pp. 13
Eliezer, S. & Henis, Z. (1994). *Fusion Technology*, Vol. 26, pp. 46
Faifman, M.P. et al. (1996). *Hyp. Interact.*, Vol. 101/102, pp. 179
Fano, U. (1981). *Phys. Rev. A*, Vol. 24, pp. 2402
Filchenkov, V.V. et al. (2005). *Hyp. Interact.,* Vol. 163, pp. 143
Filipowicz, M. et al. (2008). *Eur. Phys. J. D,* Vol. 47, pp. 157
Frolov, A.M. (1993). *J. Phys. B*, Vol. 26, pp. L845
Gershtein, S.S. & Gusev, V.V. (1993). *Hyp. Interact.*, Vol. 82, pp. 185
Gusev, V.V. et al. (1990). *Few-Body Syst.*, Vol. 9, pp. 137

Hara, S. & Ishihara, T. (1989). *Phys. Rev. A*, Vol. 39, pp. 5633
Hu, C-Y. ; Hale, G.M. & Cohen, J.S. (1994). *Phys. Rev. A*, Vol. 49, pp. 4481
Imo, H. et al. (2006). *Phys. Lett. B*, Vol. 632, pp. 192
Ishida, K. & Nagamine, K. et al. (1999). *RIKEN Rev.,* Vol. 20, pp. 3
Ishida, K. et al. (2001). *Hyp. Interact.*, Vol. 138, pp. 225
Jones, S.E. et al. (1987). *Muon Catal. Fusion* Vol. 1, pp. 21
Jones, S.E. et al. (1983). *Phys. Rev. Lett.* Vol. 51, pp. 1757
Jones, S.E. (1986). *Nature,* Vol. 321, pp. 127
Jones, S. E.; Taylor, S.F. & Andeson, A.N. (1993). *Hyp. Interact.,* Vol. 82, pp. 303
Kilic, S. ; Karr, J.P. & Hilico, L. (2004). *Phys. Rev. A*, Vol. 70, pp. 042506
Knowles, P.E. et al. (1996). *Hyp. Interact.* Vol. 101/102, pp. 21
Korobov, V.I. et al. (1992). *Muon Catal. Fusion,* Vol. 7, pp. 63
Korenman, G.Y. (1996). *Hyp. Interact.* Vol. 101/102, pp. 81
Kravtsov, A.V.; Mikhailov, A.I. & Savichev, V.I. (1993). *Hyp. Interact.*, Vol. 82, pp. 205
Lin, C.D. (1995). *Phys. Rep.*, Vol. 257, pp. 1
Macek, J. (1968). *J. Phys. B*, Vol. 1, pp. 831
Marshal, G.M. (2001). *Hyp. Interact.,* Vol. 138, pp. 203
Markushin, V.E. (1988). *Muon Catal. Fusion*, Vol. 3, pp. 395
Menshikov, L.I. & Ponomarev, L.I. (1984). *JETP Lett.* Vol. 39, pp. 663
Matsuzaki, T. et al. (2003). *Phys. Lett. B,* Vol. 557, pp. 176
Matsuzaki, T. et al. (2004). *Prog. Theor. Phys. Suppl.* Vol. 154, pp. 225
Matsuzaki, T. et al. (2001). *Nucl. Instrum. Meth. Phys. Res. A,* Vol. 465, pp. 365
Nagamine, K. et al. (1987). *Muon Catal. Fusion,* Vol. 1, pp. 137
Nagamine, K. (2001). *Hyp. Interact.,* Vol. 138, pp. 5
Nagamine, K. et al. (1993). *Hyp. Interact.,* Vol. 82, pp. 343
Nielsen, E. et al. (2001) *Phys. Rep.,* Vol. 347, pp. 373
Owski, A.A. (2007). *Eur. Phys. J. D*, Vol. 41, pp. 483
Pahlavani, M.R. & Motevalli, S.M. (2008). *Acta Phys. Pol. B,* Vol. 39, pp. 683
Pahlavani, M.R. & Motevalli, S.M. (2009). *Acta Phys. Pol. B,* Vol. 40, pp. 319
Pahlavani, M.R. & Motevalli, S.M. (2008). *Appl. Sci.,* Vol. 10, pp. 199
Pahlavani, M.R.; Sadeghi, M.; Motevalli, S.M. & Aqabaei, Y. (2010). *Mod. Phys. Lett. A*, Vol. 25, No. 5, pp. 389
Petitjean, C. et al. (1993). *Hyp. Interact.,* Vol. 82, pp. 273
Petitjean, C. (1992). *Nucl. Phys. A,* Vol. 543, pp. 79c
Petitjean, C. (2001). *Hyp. Interact.*, Vol. 138, pp. 191
Ponomarev, L.I. (2001). *Hyp. Interact.,* Vol. 138, pp. 15
Ponomarev, L.I. (1991). *Contemp. Phys.,* Vol. 31, pp. 219
Refelski, H. et al. (1989). *Prog. Part. Nucl. Phys.*, Vol. 22, pp. 297
Stuchi, T.J. et al. (2000). *Phys. Rev. A*, Vol. 62, pp. 6
Takahashi, H. (1987). *Muon Catal. Fusion,* Vol. 1, pp. 237
Takahashi, S. & Takatsuka, K. (2006). *J. Chem. Phys.*, Vol. 124, pp. 144101
Tolstikhin, O.I. et al. (1995). *Phys. Rev. Lett.*, Vol. 74, pp. 3573
Viviani, M. et al. (1998). *Nucl. Phys. A*, Vol. 631, pp. 111c

6

Application of Quantum Mechanics for Computing the Vibrational Spectra of Nitrogen Complexes in Silicon Nanomaterials

Faouzia Sahtout Karoui[1] and Abdennaceur Karoui[2]
[1]*Department of Computer Science, ISCAE, University of Manouba, Manouba*
[2]*Department of Natural Science and Mathematics, Shaw University, Raleigh NC*
[1]*Tunisia*
[2]*USA*

1. Introduction

Nitrogen is a key dopant in silicon for modern electronics including nanoscale devices and third generation solar cells. Even at concentration levels as low as 10^{15} cm^{-3} nitrogen doping can change drastically the physical properties of silicon wafers. For instance, large Czochralski silicon (CZ Si) wafers, as well as float zone silicon (FZ Si) wafers for photovoltaic applications benefit from nitrogen in silicon by increasing wafer toughness. An exceptional hardening due to nitrogen doping enabled the growth of wider silicon crystals, in excess of 300 mm in diameter. Nitrogen doped silicon appeared tougher than its oxygen doped counterpart, which enabled thinner and lighter wafers, thus easier to handle. The hardness is induced by dislocation locking effect (Sumino et al., 1983; Chiou et al., 1984; Abe et al., 1884; Murphy et al., 2006), and an increase of the density of as-grown precipitates which originates from nitrogen-oxygen clusters (Karoui et al., 2004; Karoui & Rozgonyi, 2004; Nakai et al., 2001; Karoui et. al., 2002). Nitrogen interacts with point defects such as Si vacancy (V) or Si self-interstitial (I), as well as light impurities affecting the formation of micro-defects, thereby significantly reducing swirl defects as well as vacancy related defects known as D-defects, COPs and voids, and improving the gate oxide integrity (GOI) (von Ammon et al., 1996; Tamatsuka et al., 1999; Ikari et al., 1999). Nitrogen also dramatically enhances oxygen precipitation by interacting with oxygen, achieving strong gettering of metallic impurities in the bulk (Ikari et al., 1999; von Ammon et al., 2001; Shimura & Hockett, 1986; Sun et al., 1992; Aihara et al., 2000). Fourier Transform Infrared Spectroscopy (FTIR) has been extensively used to identify the atomic structure of N-related defects and to determine nitrogen concentration in nitrogen doped FZ (N-FZ) and CZ (N-CZ) Si wafers (Stein, 1983, 1986; Wagner, 1988; Qi et al., 1991; Yang et al., 1998; Qi et al., 1992). FTIR measurements on N-FZ Si wafers shows that 80% of nitrogen atoms are paired (N-pairs) and bonded to silicon at concentration much larger than the solid solubility limit (Stein, 1983). Most nitrogen atoms are coupled by pair and

are fully coordinated with the Si atoms removing any electrical activity (Brower, 1982; Stein, 1987). The possible atomic structures for a N-pairs is either in interstitial split arrangement as suggested by Jones et al. (Jones et al., 1994), or in substitutional position, occupying either a vacancy (V) or a divacancy (V_2), forming nitrogen-vacancy (N-V) complexes (Stein, 1983, 1985). N-V complexes have been identified by DLTS measurement (Fuma et al., 1996), platinum diffusion (Quast et al., 2000) and positron annihilation (Shaik Adam et al., 2001).

As shown in Table 1, the FTIR absorption bands 771 cm^{-1}, 967 cm^{-1} at low temperature (< 15 K) and 766 cm^{-1}, 963 cm^{-1} at room temperature (RT) relate to the localized modes of N-pairs (Stein, 1983; Wagner, 1988; Qi et al., 1991). Two additional FTIR lines, 551 cm^{-1} and 653 cm^{-1} (RT), have been detected after laser annealing of N-implanted FZ Si and have been attributed to N subtitutional (Stein, 1985). The absorption coefficient of line 963 cm^{-1} is often used in the calibration curve derived by Itoh et al (Itoh et al., 1985) to measure nitrogen concentration in N-FZ and N-CZ Si wafers: $(1.83\pm0.24)x10^{17}x\ \alpha_{963}$ at/cm^{-3}.

In N-CZ Si or O-rich N-FZ Si, some of the grown-in N-pairs interact with oxygen forming nitrogen-oxygen or nitrogen-vacancy-oxygen complexes (that we will refer later as N-O complex) hence, reducing the number of N-N centers (Stein, 1986; Wagner, 1988; Qi et al., 1991; Yang et al., 1998; Qi et al., 1992). N-O complexes form between 400 ^oC and 700 ^oC. Beyond 700^o C these complexes dissociate, emitting the oxygen interstitial atom and leaving the N-pair intact. During subsequent cooling the N-O complexes form again (Wagner, 1988; Qi et al., 1991; Qi et al., 1992; Berg Rasmussen et al., 1996). Both N-N and N-O complexes FTIR response anneal out above 1000^{o}C. N-O complexes are believed to strongly control the mechanisms of formation of oxygen precipitates and voids in N-doped silicon (Karoui et al., 2004; Karoui & Rozgonyi, 2004; Nakai et al., 2001; Karoui et. al., 2002; Von Ammon et al., 2001; Shimura & Hockett, 1986; Stein, 1986; Hara et al., 1989; Rozgonyi et al., 2002). As shown in Table 1, the formation of N-O defects results in several additional infrared absorption bands (Wagner, 1988; Qi et al., 1991; Qi et al., 1992). FTIR absorption lines for N-O defects (at T < 15K) are 806, 815, 1000, 1021, and 1031 cm^{-1}. An additional weak line at 739 cm^{-1} has been observed at low temperature in FZ Si samples implanted with nitrogen and oxygen (Berg Rasmussen et al., 1996). The occurrence of these additional infrared (IR) lines affects the measurement of nitrogen concentration in N-CZ Si. The calibration relationship derived by Itoh has been revised by Qi et al (Qi et al., 1992) based on FTIR measurements as follow: $(1.83\pm0.24)\ x\ 10^{17}x\ (\alpha_{963} + 1.4\alpha_{801})$ at/cm^{-3} (300K) which take into consideration the N-O complexes to whose have been assigned the 801 cm^{-1} absorption line. Despite the technological importance of N-doped Si, little is known about the atomistic structure of N-O complexes often resulting in an inaccurate evaluation of the nitrogen content in silicon. The mechanisms by which nitrogen affects O-precipitation and vacancy aggregation in N-doped silicon remain unclear and direct experimental evidence are still needed. Although, several papers report on the electronic and atomic structure of N-pairs complexes (Jones et al., 1994; Ewels, C., 1997; Sawada & Kawakami, 2000; Kageshima & al., 2000; Goss et al., 2003; Karoui et al., 2003), theoretical studies on N-O complexes atomic structure, stability and vibrational spectra remain scarce (Ewels, C., 1997). Few of them report on the vibrational spectra of N-pair (Ewels, C., 1997; Goss et al., 2003; Jones et al., 1994).

Studying the atomic structure and vibrational spectra of nitrogen-oxygen-vacancy complexes will help us to comprehend how nitrogen, oxygen, and vacancies interact, and how nitrogen effects oxygen precipitation and void formation during crystal growth and

wafer processing. Therefore, to correctly assess nitrogen concentration in N-doped Si crystals. In the present work, we have investigated the formation energy and vibrational spectra of several structures of major grown-in nitrogen-vacancy-oxygen, using quantum mechanics Density Functional Theory (DFT) as implemented in DMol[3] package (Delley, 1990; 2000) and, the semi-empirical Modified Neglect of Diatomic Overlap Parametric Method (MNDO) in the restricted Hartree-Fock approximation (UniChem; Dewar et al., 1985). We will start by presenting the theory behind the quantum mechanics computation of vibrational spectra (Bernath, 1995; Harris & Bertolucci, 1985; Atkins & Friedman, 2011). Then we will detail our study followed by results and discussion.

	FTIR Measurement (cm^{-1})	
	$T < 15K$	RT
N-pair	771, 967 (551, 653, 782, 790)*	766, 963
N-O	806, 815, 1000, 1021, 1031 739*	801, 810, 996, 1018, 1026

* Detected in N-FZ implanted wafers.

Table 1. Measured FTIR spectra for N-N and N-O defects (Stein, 1983; Wagner, 1988; Qi et al., 1991; Qi et al., 1992).

2. Experimental measurement of vibrational spectra

Spectroscopy is the study of the interaction of electromagnetic radiation with matter. Molecules, consisting of electrically charged nuclei and electrons, may interact with the oscillating electric and magnetic fields of light and absorb the energy carried by the light. The molecules does not interact with all light that comes its way, but only with light that carries the right amount of energy to promote the molecule from one discret energy level to another. The light can be absorbed and a ground state molecule can be promoted to its first excited vibrational state. When this happen we say that the molecule has made a transition between the ground state and the first excited vibrational state.

Vibrational spectra are measured by two different techniques, *Infrared* (IR) *spectroscopy* and *Raman spectroscopy*. In *IR spectroscopy*, the infrared spectrum of a sample is recorded by passing a beam of infrared light through the sample. When the frequency of the IR is the same as the vibrational frequency of a bond, absorption occurs. Examination of the transmitted light reveals how much energy was absorbed at each frequency (or wavelength). This can be achieved by scanning the wavelength range using a monochromator. Alternatively, the whole wavelength range is measured at once using a Fourier transform instrument, hence the name of Fourier Transform Infrared Spectroscopy (FTIR). Then, a transmittance or absorbance spectrum is generated using a dedicated procedure. Analysis of the position, shape and intensity of peaks in this spectrum reveals details about the molecular structure of the sample. At frequencies corresponding to vibrational energies of the sample, some light is absorbed and less light is transmitted than at frequencies which do not correspond to vibrationals energies of the molecule. In order to compensate for absorption and scattering of the light by the sample cell, the incident light is split into two beams, one of which goes through the sample, and the other is passed through

a reference cell. Transmittance is then defined as I_s/I_r where I_s is the intensity of light passing through the sample cell, and I_r is the intensity of light passing through the reference cell.

In *Raman spectroscopy* we do not observe transmitted light but light scattered by the sample. The scattered light may be observed from any convenient direction with respect to the incident light. Light of a single frequency, monochromatic light, must be used for a Raman experiment. This phenomenon in which light of frequency ν_0 is scattered in all directions is called Rayleigh scattering. A very small fraction of the scattered light is not of frequency ν_0. The process which produce light of frequency other than ν_0 is called Raman scattering. The amount of light of frequency less than ν_0 is much greater than that with frequency higher than ν_0. The former scattered light radiation is called Stokes radiation and the latter is called anti-stokes radiation. In Raman spectroscopy, light of greater value than infrared frequencies is used and we measure the difference between the frequency of the incident light and the one of the Raman scattered light. The molecular vibrations stimulated in the Raman process are not necessarily the same as those excited by the absorption of infrared light. Therefore, the IR and Raman spectra will usually look different and will complement each other.

What exactly happens at a molecular level?

Infrared Spectroscopy: Infrared absorption spectroscopy deals with vibrations of chemical bonds. Light of infrared frequencies can generally promote molecules from one vibrational energy level to another, which allows characterization of atomic bondings and enables identification of the molecule composition and its atomic structure. These capabilities make the IR spectroscopy a powerful tool. Only photons that carry the right amount of energy promote the molecule from one discrete energy level to another.

First we need to describe the permanent dipole moment of a molecule. If two particles of charges +q and -q are separated by a distance r, the permanent electric dipole moment, μ, is given by:

$$\mu = qr \tag{1}$$

Polyatomic molecules with a center of inversion will not have a dipole moment whereas noncentrosymmetric molecules will usually have one. If we consider a heteronuclear diatomic molecule vibrating at a particular frequency, the molecular dipole moment also oscillates about its quilibrium as the two atoms move back and forth. This oscillating dipole can absorb energy from an oscillating electric field only if the field oscillates at the same frequency. The absorption of energy from the light wave by the oscillating permanent dipole is a molecular explanation of IR spectroscopy.

Raman spectroscopy: If a molecule is placed in an electric field, *f*, a dipole moment, $\boldsymbol{\mu_{ind}}$, is induced in the molecule because the nuclei are attracted toward the negative pole of field, and the electrons are attracted the opposite way. The induced dipole moment is proportional to the field strength α, which is called the *polarizability* of the molecule:

$$\mu_{\mathrm{ind}} = \alpha \mathrm{f} \tag{2}$$

All atoms and molecules will have non-zero polarizability even if they have no permanent dipole moment. A light wave electric field oscillates at a certain point in space according to the equation:

$$f = f_0 cos2\pi\nu t \tag{3}$$

where f_0 is the maximum value of the field, ν the frequency, and t is time. The induced dipole moment in the oscillating field is:

$$\mu_{ind} = \alpha f_0 cos2\pi\nu t \tag{4}$$

α varies at the natural vibrational frequency of the bond:

$$\alpha = \alpha_0 + (\Delta\alpha) cos2\pi\nu_0 t \tag{5}$$

where α_0 is the equilibrium polarizability, $\Delta\alpha$ is its maximum variation, and ν_0 is the natural vibrational frequency. The induced dipole moment is then:

$$\mu_{ind} = \alpha_0 f_0 \cos2\pi\nu t + (1/2)\Delta\alpha f_0[\cos2\pi(\nu + \nu_0)t + \cos2\pi(\nu - \nu_0)t] \tag{6}$$

Eq. 6 shows that the induced dipole moment will oscillate with components of frequency ν, $\nu - \nu_0$ and $\nu + \nu_0$. The oscillating electric dipole radiates electromagnetic waves of frequency ν (Rayleigh scattering), $\nu - \nu_0$ (Stokes radiation) and $\nu + \nu_0$ (anti-Stokes radiation).

3. Symmetry point groups

3.1 Introduction

Vibrational spectroscopy and molecular orbital theory make extensive use of molecular symmetry. While it is true that most molecules considered as a whole don't possess any symmetry, many molecules do have local symmetry. In many instances, only a region within the molecule i.e. few atoms and its neighbors, needs to be considered to understand the spectroscopic behavior of this region of the molecule. Studying carefully the symmetry of the molecule reduces significantly the number of energy levels one must deal with. The more symmetric the molecule, the fewer different energy levels it has, and the greater degeneracies of those levels. Symmetry is even powerful than that, because it helps us decide which transitions between energy levels are possible. That is to say a molecule may not be able to absorb light even if that light has precisely the correct energy to span two energy levels of the molecule. The symmetries of the states must be compatible in order that the molecule may absorb light. The selection rules which tell us which transitions are possible, will be one of the most important uses of symmetry and will be explained as we proceed.

3.2 Symmetry operations and molecules

Point symmetry groups are groups whose elements are the symmetry operations of molecules. This group have all the properties of a group in mathematics. They are called point groups because the center of mass of the molecule remains unchanged under all symmetry operations and all of the symmetry elements meet at this point. To determine the symmetry point group of a molecule is very important, because all symmetry related properties are dependent on the symmetry point group of the molecule. A symmetry operation is an operation that leaves an object apparently unchanged. Every object has at least one symmetry operation: the *identity* , the operation of doing nothing. To each symmetry operation there corresponds a symmetry element, the point, line, or plane with respect to which the operation is carried. There are five types of symmetry operations that leave the object apparently unchanged and five corresponding types of symmetry element:

E : The *identity operation*, the act of doing nothing. The corresponding symmetry element is the object itself.
C_n : An *n-fold rotation*, the operation, a rotation by $2\pi/n$ around an axis of symmetry.
σ : A reflection in a mirror plane. When the mirror plane includes the principal axis of symmetry, it is termed a *vertical plane* and denoted σ_v. If the principal axis is perpendicular to the mirror plane, then the latter symetry element is called a *horizontal plane* and denoted σ_h. *A dihedral plane,* σ_d, is a vertical plane that bisects the angle between two C_2 axes that lie perpendicular to the principle axis.
i: An *inversion*, the operation through a center of symmetry. The inversion operation consists of taking each point of an object through its center and out to an equal distance on the other side.
S_n: An *n-fold improper rotation* about an axis of improper rotation. It is a composite operation consisting of an n-fold rotation followed by a horizontal reflection in a plane perpendicular to the n-fold axis. Particular cases are S_1 which is equivalent to a reflection and S_2 is equivalent to an inversion.
Lets consider the group C_{2h} which is, as we will show later, the point group associated with the N_2 molecule in Si. Point group C_{2h} has four members, {E, C_2, σ_h and i}. E is the identity operation which leave the molecule unchanged; C_2 is a n-fold (n=2 here) rotation by $2\pi/n$ (180^o for C_2) around an axis of symmetry; σ_h is a reflection in a mirror plane (here a plane perpendicular to the principal axis C_2). All the symmetry operations of a molecule as a group can be written in the form of group multification table and they obey all the properties of a group. The product of any two operations must be a member of the group. For example the product of two C_2 operations is the identity operation E which is indeed a member of the group. Also $C_2 \cdot \sigma_h = i$, *i* is also a member of the group. Table 2 shows the complete multiplication table for the point group C_{2h}.

C_{2h}	E	C_2	σ_h	i
E	E	C_2	σ_h	i
C_2	C_2	E	i	σ_h
σ_h	σ_h	i	E	C_2
i	i	σ_h	C_2	E

Table 2. Multiplication Table for the Point Group C_{2h}.

However, to further determine the symmetry properties of molecular orbitals and vibrational modes we need character tables which will be introduced next.

3.3 Characters and character tables

We can use matrices as representations of symmetry operations. Let's consider the symmetry group (C_{2h}) of N_2 defect. Consider a vector $\overrightarrow{v_1}\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$, assuming that the principal axis C_2 is the z axis, using matrices representations, $\overrightarrow{v_1}$ will transform as follow through the different operations of the group:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \tag{7}$$

$$E \quad . \overrightarrow{v_1} = \overrightarrow{v_1} \tag{8}$$

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} -x_1 \\ -y_1 \\ -z_1 \end{bmatrix} \tag{9}$$

$$i \quad . \overrightarrow{v_1} = -\overrightarrow{v_1} \tag{10}$$

σ_h leaves the x and y coordinates unchanged but changes z to -z:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ -z_1 \end{bmatrix} \tag{11}$$

$$\sigma_h \quad . \overrightarrow{v_1} = -\overrightarrow{v_2} \tag{12}$$

and C_2 which correspond to 180° rotation around the z axis leaves the z coordinate unchanged but changes the x and y coordinates as follow:

$$\begin{bmatrix} cos\pi & sin\pi & 0 \\ -sin\pi & cos\pi & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} -x_1 \\ -y_1 \\ z_1 \end{bmatrix} \tag{13}$$

$$R_\theta \quad . \overrightarrow{v_1} = \overrightarrow{v_2} \tag{14}$$

The four matrices form a mathematical group which obeys the same multiplication table as the operations. Therefore, each matrix has an inverse matrix just as each operation of a group has an inverse operation. Using a matrix and its inverse we can perform similarity transformations with matrices:

$$B = Q^{-1} \cdot A \cdot Q \tag{15}$$

A and B are said to be conjugate just as symmetry operations related by similarity transformations are said to be conjugate. Through similarity of transformation, we can define the reducible and irreducible representations of a group. If a matrix representation A can be transferred to block-factored matrix, a matrix composed of blocks (A_1, A_2, A_3) at the diagonal and zero in any other position, by similarity transformation, A is called the reducible representation of the group. If blocks (A_1, A_2, A_3) cannot be further transferred to block-factored matrix through similarity transformation, A_1, A_2, A_3 are called irreducible representations of the group. The sum of the trace of A_1, A_2, A_3 is called the characters of this representation. Reducible representations can be reduced to irreducible representations and irreducible representations cannot be reduced further. The complete list of characters of all possible irreducible representations of a group is called *a character table*. There are only a finite number of irreducible representations for group of finite order. We will see that these tables are of great importance and usefullness when analysing the vibration modes of molecules.

The members of a group can be divided into classes. Two members of a group, P and R, belong to the same class if they are conjugate to each other. As an example, all possible classes associated with the symmetry group $C_{2h}(N_2)$ defect are the following:

- E is in a class by itself since $A^{-1}EA = A^{-1}(EA) = A^{-1}A = E$ for any operation A of the group.
- C_2: $C_2^{-1}C_2C_2 = C_2^{-1}(C_2C_2) = C_2(E) = C_2$

- $i^{-1}C_2 i = i^{-1}(C_2 i) = i(\sigma_h) = C_2$
- $\sigma_h{}^{-1}C_2\sigma_h = \sigma_h{}^{-1}(C_2\sigma_h) = \sigma_h\,(i) = C_2$

Hence, for any operation A of the group we have $A^{-1}\,C_2A = A^{-1}(C_2A) = A^{-1}A = C_2$. In a same way, based on the multiplication Table 2, we can verify that $A^{-1}iA = A^{-1}(iA) = A^{-1}A = i$ and $A^{-1}\sigma_h A = A^{-1}(\sigma_h A) = A^{-1}A = \sigma_h$. Therefore, we have four classes for the group symmetry C_{2h}: {E}, {C_2}, {i}, {σ_h}. Each class correspond to an irreducible representation.

We have as much as irreducible representations as classes of operations in the group. The character of an irreducible representation is the trace, the sum of the diagonal elements, of the matrix representing the irreducible representation. The sum of the traces equal the order of the group. All characters of a group are given in a table, Table 3. This table is divided in several areas. The main part contains the characters. On the left are the *names* of the irreducible representations, known as Mulliken symbols. Conventionally, we use the letters A, B, E, and T (or F in some tables). A and B are one-dimentional. E is two-dimensional and T is three-dimensional. The dimension of an irreducible representation is the dimension of any of its matrices. Since the representation of the operation E is always the identity matrix, the character of E is always the dimension of the irreducible representation. The difference between A and B is that the character under the principal rotational operation, C_n, is always +1 for A and -1 for B representations. The subscript 1,2,3, etc.,which may be appear with A, B, E or T can be considered arbitrary label. The subscript *g* (German word *gerade* meaning *even*) means the representation is symmetric with respect to inversion and, the subscript *u* (German word *ungerade* meaning *odd*) means that the representation is antisymmetric to inversion. Any p or f orbital is transformed into minus upon inversion, is therefore a *u* function. A d orbital is transformed into itself upon inversion and is therefore, a *g* function. In a similar way, the superscripts ' and " denote irreducible representations which are respectively, symmetric and antisymmetric with respect to reflection through a horizontal mirror plane. The two columns on the right side of the table contain basis functions for the irreducible representations. The character table for the point group C_{2h} (N_2 defect) is as follow (Bernath, 1995; Harris & Bertolucci, 1985)

C_{2h}	E	C_2	i	σ_h		
A_g	1	1	1	1	R_z	x^2, y^2, z, xy^2
B_g	1	1	1	-1	R_x, R_y	xz, yz
A_u	1	-1	1	1	z	
B_u	1	-1	1	-1	x, y	

Table 3. Character Table for the Point Group C_{2h} : N_2 Defect.

3.4 Atomic orbitals and symmetry

One-electron wavefunctions in atoms are called *atomic orbitals*. Atomic orbitals with l=0 are called s-orbitals, those with l = 1 are called p-orbitals, those with l = 2 are called d-orbitals, and those with l = 3 are called f-orbitals. We are mainly interested here to s- and p-orbitals because the atoms of interest namely nitrogen, oxygen, and Si atoms are bonded to Si neighbors by sp^3 hybrid electron orbital that protrude in a tetrahedral shape. The s-orbitals are spherically symetrical; the three real orbitals p_x, p_y, p_z have the same double-lobed shape, but are aligned with the x-, y-, and z-axes, respectively; they are shown in Fig. 1.

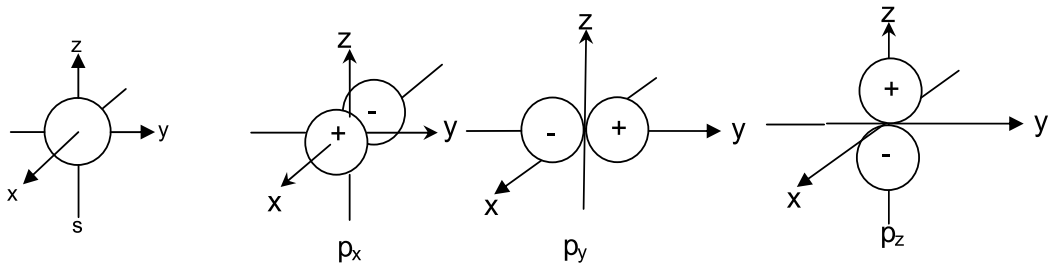


Fig. 1. The s, p_x, p_y, p_z atomic orbitals.

As stated before, the two columns on the right side of the table contain basis functions for the irreducible representations. These basis functions have the same symmetry properties as the atomic orbitals which bear the same names. To understand what a basis function is, let's go back to the matrix representations for the operations of C_{2h} (N_2 defect). The E operation does nothing; the C_2 operation about the z axis leaves the z coordinate of any point unchanged, but changes the x and y coordinates according to R_θ; the σ_h is a reflection in the mirror plane (x,y) $\perp$ to z (C_2 axis); and finally the inversion operation *i* changes each coordinate into minus itself. The atomic orbitals will obey the same multiplication table as the operations, Fig. 2:

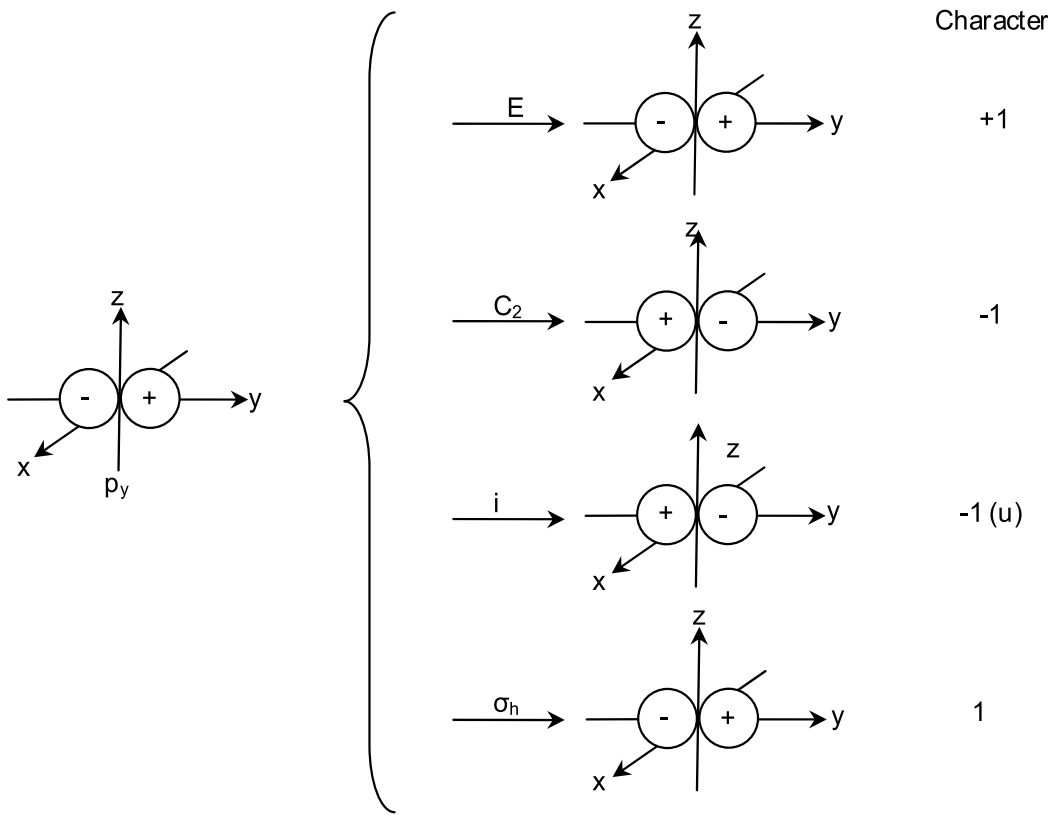


Fig. 2. Orbital p_y through the operation of symmetry point group C_{2h}.

We can see that p_y orbital changes as B_u irreducible representation through the operations of the group C_{2h} : (E, C_2, i, σ_h) : (1,-1,-1,1) ≡ {B_u}.

4. Vibrational spectroscopy modeling

4.1 Introduction

In IR spectroscopy, molecules are modeled as an assembly of oscillators which interact with the electric and magnetic fields of incident light and absorb energy of incident photons. The total energy of a single molecule, whether in free space or embedded in liquid or solid material, involves different types of molecule motions and behaviors. Hence, the molecule energy is decomposed in: (i) translational energy levels, which are related to the movement of the molecule as a whole. As these levels are very close to each other, they appear continuous, (ii) rotational energy levels, they implicate rotation of the whole molecule, (iii) vibrational energy levels, which are due to the vibration of chemical bonds within the molecule, and (iv) the electronic energy associated to the electrons of the molecule. To better comprehend the vibrational spectroscopy modeling we will start with diatomic molecules and then generalize the model to polyatomic molecules.

4.2 The vibration modeling of diatomic molecules

4.2.1 Introduction

The solution of the Schrödinger equation for a diatomic molecule plays an important role in spectroscopy. In addition, the vibrational spectra of diatomic molecules illustrate most of the fundamental principles which apply to complicated polyatomic molecules. Diatomic molecules can be simulated as shown in Fig. 3. The center of mass of a diatomic is defined such that $m_1 r_1 = m_2 r_2$. The moment of inertia of a system is defined as:

$$I = \sum_i m_i r_i^2 \tag{16}$$

where r_i is the distance of mass m_i from the center of mass. For diatomic molecules,

$$I = \frac{m_1 m_2}{m_1 + m_2} r_e^2 \equiv \mu r_e^2 \tag{17}$$

where

$$\mu \equiv \frac{m_1 m_2}{m_1 + m_2} \tag{18}$$

The quantity μ is called the reduced mass and should not be confused with the dipole moment which has the same symbol.

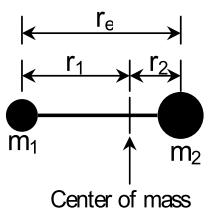


Fig. 3. Model for diatomic molecules.

The molecular potential energy of a diatomic molecule increases if the nuclei are displaced from their equilibrium positions. When the displacement is small, we can express the potential energy as the first few terms of Taylor series:

$$V(x) = V(0) + \left(\frac{dV}{dx}\right)_0 x + \frac{1}{2}\left(\frac{d^2V}{dx^2}\right)_0 x^2 + \frac{1}{6}\left(\frac{d^3V}{dx^3}\right)_0 x^3 + \cdots \tag{19}$$

Since we are not interested in the absolute potential energy of the molecule,we can set $V(0) = 0$.

4.2.2 Harmonic oscillation

The harmonic oscillator model is one of the most important models in chemical physics, and has been used extensively in molecular spectroscopy. Provided that the displacement x is small, the terms in Eq. 19 that are higher than second order may be neglected, so we may write

$$V(x) = \frac{1}{2}kx^2 \qquad k = \left(\frac{d^2V}{dx^2}\right)_0 \tag{20}$$

This means that the potential energy close to equilibrium is parabolic. It follows that the hamiltonian for the two atoms of masses m_1 and m_2 is

$$H = -\frac{\circ^2}{2m_1}\frac{d^2}{dx_1^2} - \frac{\circ^2}{2m_2}\frac{d^2}{dx_2^2} + \frac{1}{2}kx^2 \tag{21}$$

Therefore, when the potential energy depends only on the separation of the particles, the hamiltonian can be expressed as a sum, one term referring to the motion of the center of mass of the system and the other to the relative motion. The former term is of no concern here as it corresponds to the translational motion of the molecule. The latter term is

$$H = -\frac{\hbar^2}{2\mu}\frac{d^2}{dx_2^2} + \frac{1}{2}kx^2 \tag{22}$$

where μ is the reduced mass.
A hamiltonian with a parabolic potential energy as in Eq. 23, is characteristic of a *harmonic oscillator*. The solutions for the harmonic oscillator is

$$E_\nu = \left(\nu + \frac{1}{2}\right)\hbar\omega \qquad \omega = \frac{1}{2\pi}\sqrt{k/\mu} \tag{23}$$

with $\nu = 0,1,2,...$ These levels lie in a uniform ladder with separation $\hbar\omega$, see Fig. 4. The corresponding wavefunctions are bell-shaped Gaussian functions multiplied by a Hermite polynomial. In the lowest vibrational state (ν = 0), the molecule still has the *zero point energy*, $E_0 = \frac{1}{2}\hbar\omega$. The vibrational spectra of diatomic molecules usually result from excitation from the $\nu = 0$ to the $\nu = 1$ energy levels.

4.2.3 Anharmonic oscillation

The truncation of Taylor expansion of the molecular potential energy in Eq. 20 is only an approximation , and in real molecules the neglected terms are important, particularly for

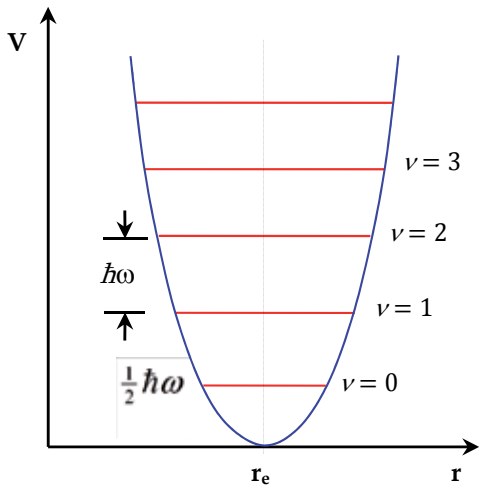


Fig. 4. Harmonic oscillator potential well and energy levels. $V = \frac{1}{2}k(r - r_e)^2$.

large displacements from equilibrium. The typical form of the potential is shown in Fig. 5 and because of high excitations it is less confining than a parabola; the energy levels converge instead of staying uniformly separated. It follows that anhamonic vibration is increasingly important as the degree of vibrational excitation of a molecule is increased. One way for coping with anharmonicities is to solve the Schrodinger equation with a potential energy term that matches the true potential energy over a wide range. One of the most useful approximation function is the *Morse potential* :

$$V(x) = \hbar c D_e (1 - e^{-ax})^2 \qquad a = \left(\frac{k}{2\hbar c D_e}\right) \tag{24}$$

The parameter D_e is the depth of of the minimum of the curve and $x = r - r_e$ the displacement. At small displacement, the Morse and harmonic oscillator potentials coincide. The quantized energy levels, solution of the Schrodinger equation with the Morse potential are

$$E_\nu = \left(\nu + \frac{1}{2}\right)\hbar\omega - \left(\nu + \frac{1}{2}\right)^2 \hbar\omega x_e \tag{25}$$

with

$$\omega x_e = \frac{a^2\hbar}{2\mu}$$

The quantity x_e is called the anharmonicity constant. The energy levels at high excitation converge as ν becomes large. The ground state of a Morse potential has a zero-point energy of

$$E_0 = \frac{1}{2}\hbar\omega\left(1 - \frac{1}{2}x_e\right) \tag{26}$$

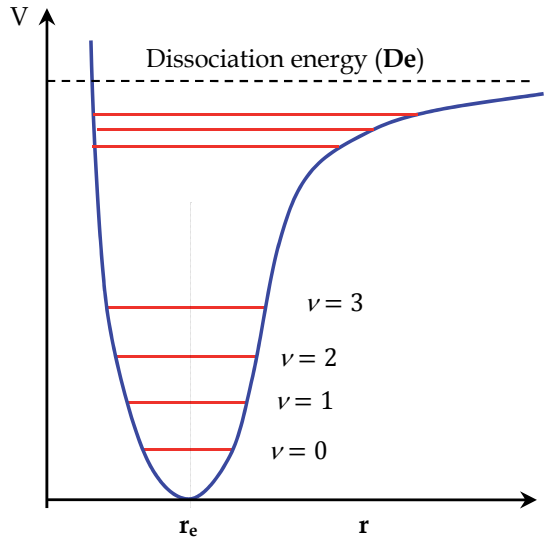


Fig. 5. The Morse potential : $\boldsymbol{V = \frac{1}{2}\hbar c D_e (1 - e^{-a(r-r_e)})^2}$

4.2.4 Vibrational selection rule

The selection rules for the vibrational transition $\nu' \leftarrow \nu$ are based on the electric dipole transition moment. The selection rules for electric-dipole transitions, specify the specific optical transitions that occur based on the examination of dipole moment transitions between the two states of interest. Because the dipole moment $\boldsymbol{\mu}$ depends on the bond length R, we can express its variation with displacement of the nuclei from equilibrium as

$$\mu = \mu_0 + \left(\frac{d\mu}{dx}\right)_0 x + \frac{1}{2}\left(\frac{d^2\mu}{dx^2}\right)_0 x^2 + \cdots \tag{27}$$

where $\boldsymbol{\mu_0}$ is the dipole moment when the displacement is zero.

To show a vibrational spectrum, a diatomic molecule must have a dipole moment that varies with extension. The selection rule for electric dipole vibrational transitions within the harmonic approximation is $\Delta\nu = \pm 1$.

The selection rule for the observation of vibrational Raman spectra of diatomic molecules is that the molecular polarizability must vary with internuclear separation. The selection rule for vibrational Raman transitions is the same, $\Delta\nu = \pm 1$, as for vibrational absorption and emission because the polarizability, like the electric dipole moment, returns to its initial value once during each oscillation. The transitions with $\Delta\nu = +1$ give rise to the Stokes lines in the spectrum, and those with $\Delta\nu = -1$ give the anti-Stokes lines. Only the Stoke lines are normally observed, because most molecules have $\nu = 0$ initially.

4.3 Vibration of polyatomic molecules

4.3.1 Normal modes of vibration and symmetry

A diatomic molecule possesses a single vibration. Even at absolute zero this vibration occurs because the molecule cannot have less than the zero point energy. Polyatomic molecules undergo much more complex vibrations. However, these motions may be resolved into a

superposition of a limited number of fundamental motions called normal modes of vibration. We are interested in the number, types, and symmetries of these modes.
The motion of a single particle in a three dimensional space can be represented by three coordinates, each one representing a translation of the particle in the x, y, or z direction. The particle is said to have three degrees of freedom. For a diatomic molecule, since we have two particles, the system as a whole has six degrees of freedom. Three are translation in the x, y, or z directions. Two degrees of freedom correspond to rotations about the center of mass. The rotation about the molecular axis of a linear molecule is undefined because it does not represent any change of the nuclear coordinates. Only one vibrational degree of freedom is left, Fig. 6.

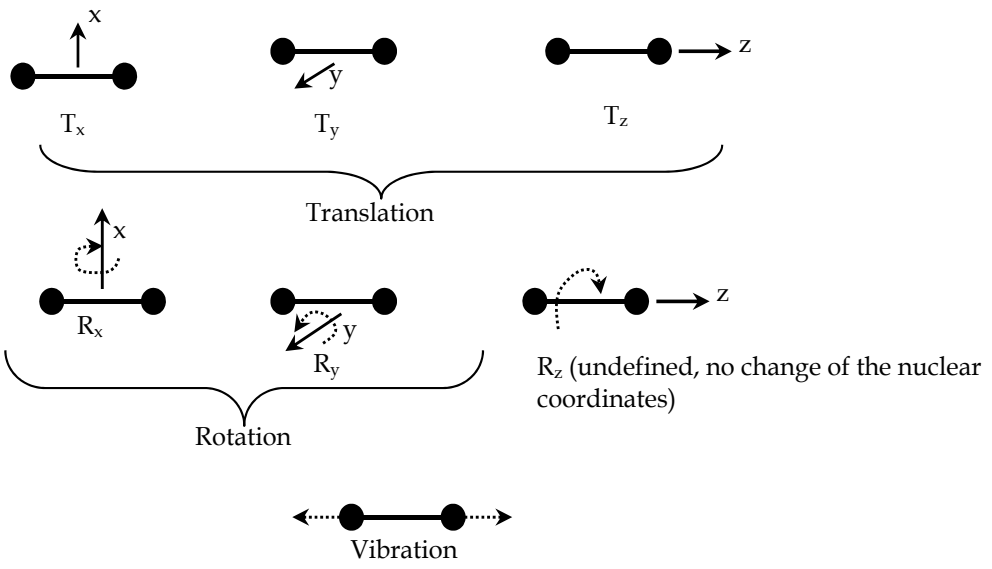


Fig. 6. Degrees of freedom (six) of a diatomic molecule; only one vibration mode (3*2-5=1).

A nonlinear molecular system containing three particles (e.g water molecule) has nine degrees of freedom. Three translations, three rotations, and three vibrational degree of freedom. these three kinds of vibration are the three normal modes of vibration of the molecule. In general, a non linear molecule with n atoms will have 3n-6 modes of vibration; a linear molecule will have 3n-5 modes of vibration because there is no rotation about the molecular axis. Within a molecule, atomic displacements occur at the same frequency and in phase. Displacement is measured from the equilibrium atomic separation in the ground state along a normal coordinate. A normal coordinate, q_i , is a single coordinate along which the progress of a single normal mode of vibration can be followed; q_i is a mass-weighted coordinate. The normal coordinates are defined such that the potential energy V and the kinetic energy K of the molecule are as follow (Harris & Bertolucci, 1985)

$$V = (1/2)\sum_i \lambda_i \, q_i^2 \tag{28}$$

$$K = (1/2)\sum_i (dq_i/dt)^2 \quad (29)$$

where λ_i is a constant. The vibrations that correspond to displacements along these normal coordinates are called the *normal modes* of the molecule. In the harmonic approximation, the ground state vibrational wavefunctions of a molecule is totally symmetric under all symmetry operations of the molecule. The ground state vibrational wavefunctions therefore spans the completely symmetric irreducible representations of the molecular point group.

Each normal mode of vibration will form a basis for an irreducible representation of the point group of the molecule. This key property which connects the symmetry of normal modes of vibration to the symmetry point group of the molecule.

Lets consider the Si_2O molecule which belongs to point group C_{2v}. As we will see subsequently, this molecule is of particular interest to our study. C_{2v} character table is given in Table 4. The operations of the group C_{2v} are (E, C_2, $\sigma_v(xz)$ and $\sigma'_v(yz)$). The three normal modes of vibration of Si_2O are given in Fig. 7 and are noted ν_1, ν_2, ν_3. We are going now to study the effect of each operation of the group on the ν_3 vibration.

C_{2v}	E	C_2	$\sigma_v(xz)$	$\sigma'_v(yz)$		
A_1	1	1	1	1	z	x^2, y^2, z^2
B_1	1	1	-1	-1	R_z	xy
A_2	1	-1	1	-1	x, R_y	xz
B_2	1	-1	-1	+1	y, R_x	yz

Table 4. Character Table for the Symmetry Point Group C_{2v} : Si_2O molecule.

Fig. 7. ν_1 (symmetric stretching, 517 cm^{-1}), ν_2 (symmetric bending, 1203 cm^{-1},), ν_3 (asymmetric stretching, 1136 cm^{-1}) normal modes of Si_2O.[78,86]

The operation E leaves the ν_3 vibration unchanged so it has the character +1. The C_2 operation changes the direction of motion of each atom when the molecule is vibrating in the normal mode ν_3. Each atom moves in the opposite direction after performing the C_2 operation (Fig. 8). Therefore, the character of C_2 is -1. Similarly, $\sigma_v(xz)$ changes the direction of motion of the atoms and $\sigma'_v(yz)$ leaves them unchanged. Hence, they have the character -1 and +1 respectively. In a similar way, we can easily show that ν_1 and ν_2 does not change through all the group operations. This leads us to the characters table given in Table 5 for the vibration modes of the Si_2O molecule.

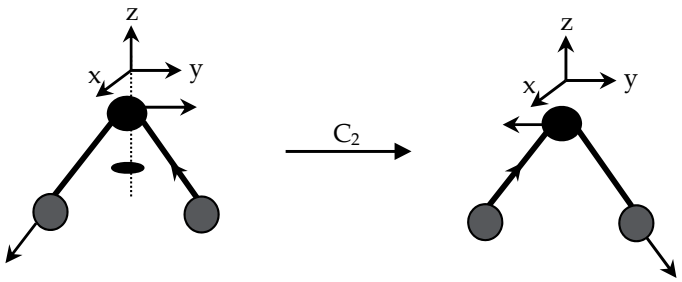


Fig. 8. Application of C_2 operation on the normal mode ν_3 of Si_2O.

C_{2v}	E	C_2	$\sigma_v(xz)$	$\sigma'_v(yz)$
ν_1	1	1	1	1
ν_2	1	1	1	1
ν_3	1	-1	-1	1

Table 5. Character table for for the normal modes of vibration of Si_2O (C_{2v}).

If we look at Table 4 and Table 5 we can see that ν_1 changes as A_1 representation through the operations of the group, ν_2 as A_1 and ν_3 as B_2. Vibration modes ν_1 and ν_2 are symmetric stretching and symmetric bending modes respectively, while ν_3 is an asymmetric stretching mode. All three normal modes are infrared active because they have the same symmetry as z and y.

4.3.2 Selection rules for polyatomic molecules

Non-zero dipole moment transition corresponds to allowed transition, and vice-versa. At this point, we are interested only in transitions within a given electronic state. The selection rules are derived from the transition matrix by expressing the matrix element in terms of, in a first approximation, the harmonic oscillator wavefunctions. The selection rule for harmonic oscillators are $\Delta\nu = \pm 1$. Each normal mode of vibration shall obeys this selection rule within the harmonic approximation. Moreover, electric dipole transitions can occur only for normal modes that correspond to a change in the electric dipole moment of the molecule. The molecular dipole moment depends on an arbitrary displacement as follows:

$$\mu = \mu_0 + \left(\frac{d\mu}{dq_i}\right)_0 q_i + \frac{1}{2}\left(\frac{d^2\mu}{dq_i^2}\right)_0 {q_i}^2 + \cdots \tag{30}$$

where q_i are the normal coordinates.

Since, electric dipole transitions occur only for normal modes that correspond to a change in the electric dipole moment of the molecule, normal modes for which $(\partial\boldsymbol{\mu}/\partial q_i)_0 \neq 0$ are said to be *infrared active* as they can contribute to a vibrational, infrared, absorption, or emission spectrum. Group theory, as shown before, greatly aids the determination of which modes are infrared active.

Normal modes for which the polarizability varies as the atoms are displaced collectively along a normal coordinate i.e. $(\partial\boldsymbol{\alpha}/\partial q_i)_0 \neq 0$, are classified as Raman active as they can contribute to a Raman spectrum.

5. Quantum mechanics computation of the equilibrium structure, energy and vibrational spectra of nitrogen-related complex in nitrogen-doped silicon

5.1 Introduction

In the present work, we have investigated the formation energy and vibrational spectra of several structures of major grown-in nitrogen-vacancy-oxygen, using quantum mechanics Density Functional Theory (DFT) as implemented in DMol^3 package and, the semi-empirical Modified Neglect of Diatomic Overlap Parametric Method (MNDO) in the restricted Hartree-Fock approximation. The defects that are of interests are N-pairs structures either in interstitial (N_2) or substitutional positions coupled to a Si vacancy or a divacancy (VN_2 and V_2N_2) and N-O complexes formed by the coupling of these N-N centers with a O_i or an oxygen dimer (O_2). MNDO calculations were solely performed to compute the IR absorption band intensities and electric charges of the IR active LVMs obtained by DMol^3-DFT, because not attainable on periodic systems. Performing ab-initio total energy calculations and normal mode analysis require powerful computational resources. Therefore, all our calculations have been carried out on various multiprocessor supercomputers: Origin2400 for COMPASS Force field and Fastructure calculations, IBM SP for the DMol^3-DFT calculations and Cray T916 for the MNDO-AM1 calculations.

We have cross-correlated the formation energy, the degree of stability, and the vibrational spectra of each complex in order to precisely identify their structure (Karoui Sahtout & A. Karoui, 2010). Calculated vibrational spectra have been compared to experimental spectra obtained by Fourier-transform infrared spectroscopy.

5.2 Computational method

5.2.1 Chemical reactions and atomic structure of N-pairs and N-O complexes

To simulate the defected crystal structure while preserving the symmetry group of the diamond structure of the host crystal, we built a periodic cubic system consisting of a supercell of 64 silicon atoms with the defect located in its center. All N-defects are in their neutral state. To avoid defect-defect interactions during relaxation, the Si atoms at the boundaries of the supercell were maintained immobile.

The chemical reactions considered in this study to produce N-pairs either in interstitial or substitutional position are shown below:

$$N_i + N_i \rightarrow N_2 \quad \text{(R1)}$$

$$N_i + N_s \rightarrow VN_2 \quad \text{(R2)}$$

$$N_2 + V \rightarrow VN_2 \quad \text{(R3)}$$

$$VN_2 + V \rightarrow V_2N_2 \quad \text{(R4)}$$

$$N_2 + V_2 \rightarrow V_2N_2 \quad \text{(R5)}$$

N_i and N_s are nitrogen atoms in interstitial and substitutional position, respectively, V is a Si vacancy and V_2 a Si divacancy.

The neutral N_2 has a C_{2h} symmetry (Fig. 9) with the axis parallel to <110> direction, see Fig. 10 (a), and the N-N center sits symmetrically off the bond center, in an anti-parallel

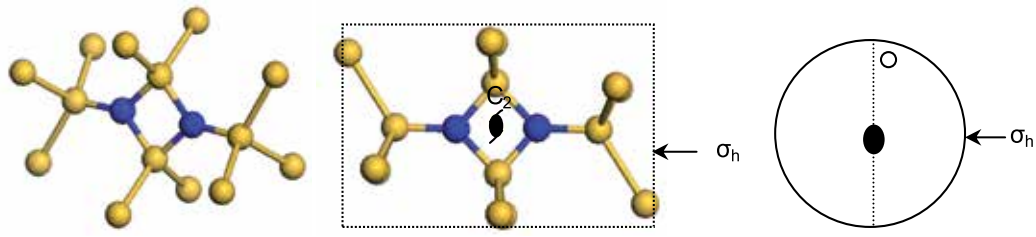


Fig. 9. N_2 Molecule Structure in Si, symmetry point group C_{2h} and stereographic projection.

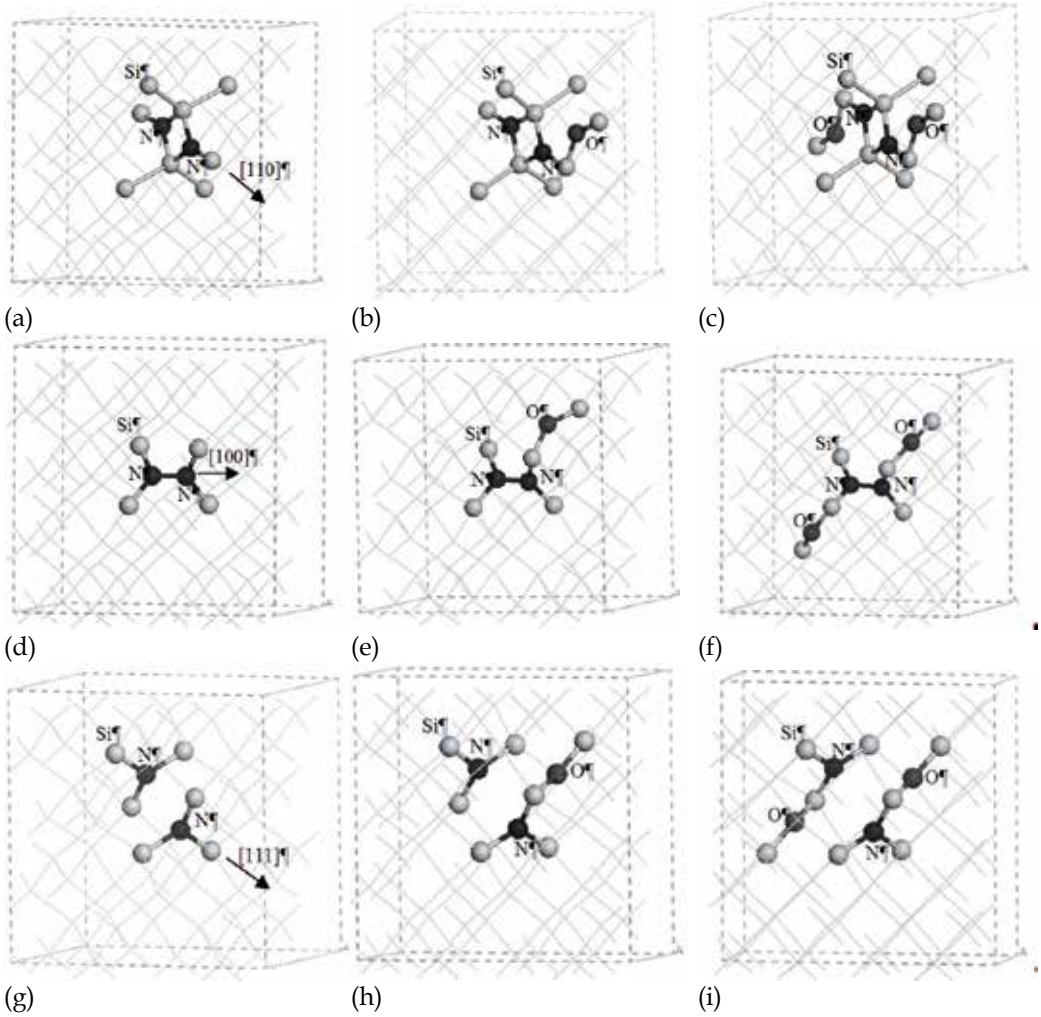


Fig. 10. N-Pair defects atomic structure (a, b, c) N_2 (C_{2h}), N_2O, N_2O_2; (d, e, f) VN_2 (D_{2d}), VN_2O, VN_2O_2; (g, h,i) V_2N_2 (D_{3d}), V_2N_2O, $V_2N_2O_2$.

configuration, as proposed by Jones et al. (Jones et al., 1994). The four Si-N bonds form a diamond shape lying in (110) plane. The bond centered interstitial configuration for N_2 has been found energetically favorable based on ion channeling, infrared absorption, and theoretical calculations (Jones et al., 1994).

As shown in Fig. 10 (d), neutral VN_2 complex is formed by inserting an N-N pair at a vacancy site in the center of a tetrahedron (Stein, 1986). The central bond of the N-N pair is aligned along <100> whereas the four N-Si bonds point to the summits of the tetrahedron, and lie in two perpendicular {110} planes which makes the symmetry group of VN_2 of D_{2d} type, Fig. 11.

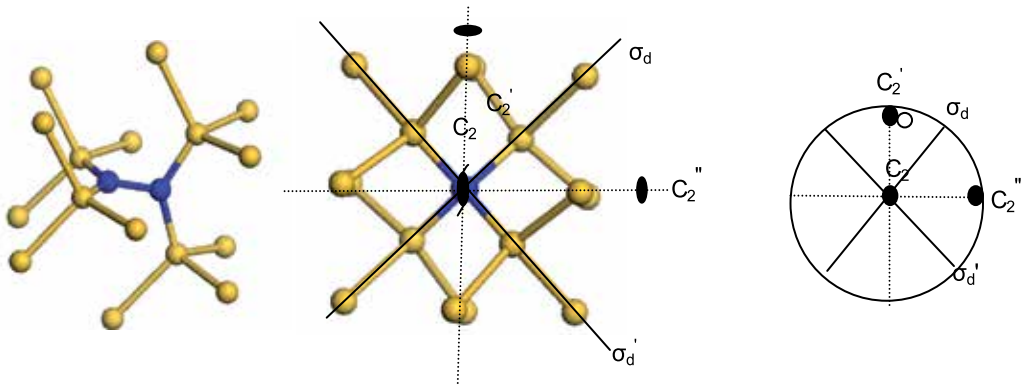


Fig. 11. VN_2 molecule Structure in Si, symmetry point group D_{2d} and stereographic projection.

The V_2N_2 complex is created by inserting two N atoms in the vacancy sites of a relaxed divacancy, see Fig. 10 (g). The divacancy six silicon dangling bonds are fully reconstructed. V_2N_2 has a D_{3d} symmetry (Fig. 13) similar to the ideal divacancy with "breathing" bonding (Coomer et al., 1999), see Fig. 12.

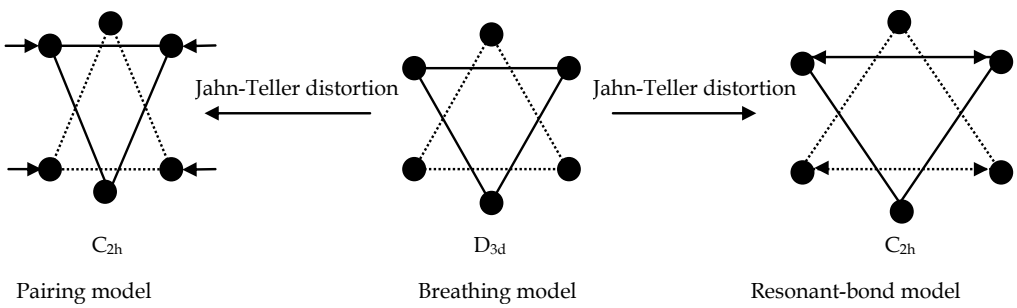


Fig. 12. Si divacancy Jahn Teller distortion (Coomer et al., 1999; Watkins & Corbett, 1965).

The N-O complexes object of this study, result from the coupling of a N-N core defect product of chemical reactions R1 thru R5, with an oxygen interstitial (O_i) or an oxygen dimer (O_2) as shown in the chemical reactions R6 thru R11:

$$N_2 + O \rightarrow N_2O \tag{R6}$$

$$N_2 + O_2 \rightarrow N_2O_2 \quad \text{(R7)}$$

$$VN_2 + O \rightarrow VN_2O \quad \text{(R8)}$$

$$VN_2 + O_2 \rightarrow VN_2O_2 \quad \text{(R9)}$$

$$V_2N_2 + O \rightarrow V_2N_2O \quad \text{(R10)}$$

$$V_2N_2 + O_2 \rightarrow V_2N_2O_2 \quad \text{(R11)}$$

The atomic structures of N_2O, VN_2O and V_2N_2O complexes are obtained by adding one O_i atom on the dilated Si-Si bond neighboring the N-N center as shown in Fig. 10 (b), (e), (h). Indeed, previous investigations on oxygen interstitial in silicon (Umerski, 1993) showed that oxygen bridges dilated Si bonds preferentially along <111> directions. Likewise, the N_2O_2, VN_2O_2 and $V_2N_2O_2$ are built by inserting two O_i atoms on Si-Si dilated bonds neighboring the N-pair, Fig. 10 (c), (f), (i).

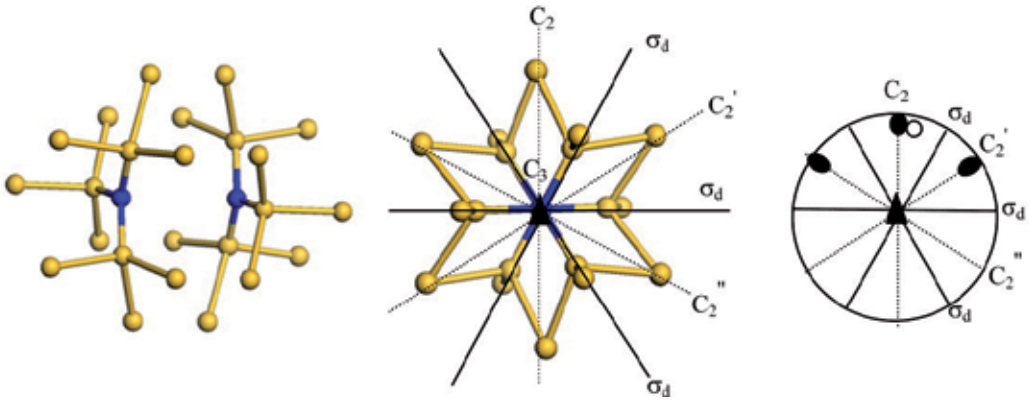


Fig. 13. V_2N_2 molecule Structure in Si, symmetry point group D_{3d} and stereographic projection.

5.2.2 Equilibrium structure and energy of formation

To cut down the computation time, the defected supercells were first relaxed using a valence force field method with COMPASS potential and then DFT Fastructure program (Fastructure). This allowed a fast cleanup and optimization of the guessed atomic structure. Faststructure determines the ground state energies and forces within the Harris functional (Harris, 1985), an approximate scheme of the DFT scheme of Kohn and Sham (Kohn & Sham, 1965). For this scheme, the exchange-correlation terms are calculated using the Vosko, Wilk, Nusair (VWN) parameterization (Vosko et al., 1980) and the radial cutoff was set at 10Å. The optimizations of the resulting structures were then performed using DFT program DMol^3. DMol^3 utilizes a basis set of numeric atomic functions that are exact solutions to the Kohn-Sham equations for the atoms (Delley, 1995). For the present study, a doubled numerical basis set with d-polarization functions, termed DND basis set, was used as it ensures an accurate description of the bonding environment. This basis set is well parameterized for nitrogen, oxygen, and Si atoms that are bonded to neighbors by sp^3

hybrid electron orbital that protrude in a tetrahedral shape. Perdew-Wang functional was used for the exchange correlation terms. To accelerate the convergence of the self-consistent-field (SCF) procedures we use the direct inversion of the iterative subspace method (DIIS) (Pulay, 1980). Integration over the Brillouin-zone and over all occupied orbital is done with the tetrahedron method(Blöchl, 1994) and equispaced Fourier meshes similar to the ones proposed by Monkhorst and Pack (Monkhorst,1976). Eight k-points and 24 tetrahedra were used to sample the Brillouin zone. The atomic positions were optimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton optimization algorithm.
For the general chemical reaction $X_A + X_B \rightarrow X_C$, the formation energy of the product X_C is defined as (Nichols, 1989):

$$E_f(X_C) \equiv E(X_C) + E_{bulk} - E(X_A) - E(X_B) \tag{31}$$

Where $E_f(X_C)$ denotes the formation energy of the complex, $E(X_i)$ the total energy of the system containing a reactant or a product, and E_{bulk} is the total energy of pure bulk Si supercell.
All reactants and products involved in reactions R1 through R11, including the Si vacancy (V) and the Si divacancy (V_2) which are important for our study, have been optimized with the same level of theory and accuracy. As reference, our calculations suggest a formation energy of 3.89eV for the Si vacancy and 5.80 eV for the divacancy in good agreement with previously published values (3.31-3.98 eV) for the vacancy (Hull, 1999; Puska et al., 1998) and 4.4-5.7 eV for the divacancy (Hull, 1999; Pesola, 2000). During relaxation, the neutral Si divacancy (V_2) are usually subject to Jahn-Teller (JT) distortion which lower the symmetry of the defect from trigonal D_{3d} to monoclinic C_{2h} resulting in the 'pairing' or 'resonant-bond' structure (Coomer, 1999; Pesola, 2000; Watkins, 1965), see Fig. 12. Our calculations exhibit a 'large pairing' JT distortion for the relaxed V_2 rather than a 'resonant bonding' distortion, in line with positron annihilation spectroscopy measurements (Nagai, 2003). The reconstructed bonds of the divacancy form two co-planar isosceles triangles, the length of the two congruent sides of the triangles is about 4.44 Å, and the base is about 3.54 Å long.

5.2.3 Computation of vibrational spectra

The absorption spectrum for each optimized atomic structure was calculated using the eigenvalue method (Dean & Martin, 1960) implemented in DMol^3 package. The eigenvalue method is sensitive to the accuracy of the calculated electronic structure; therefore vibrational analysis can be meaningful only when all atomic forces are zero. This is attainable only when the geometry is optimized at the same level of theory and with the same basis set used to generate the Hessian.
Since, there is at present no method for calculating the intensities of IR active LVMs for periodic boundary environment because the Hessian in internal coordinates can not be evaluated in that case, the LVM intensities were computed using the MNDO method and the Austin Model 1 (AM1) Hamiltonian (Dewar et al., 1985). MNDO works only on molecular systems therefore a crystal macro-molecule (CMM) has been extracted from the Dmol^3 relaxed supercell. Each CMM contains the N-N core defect in its center surrounded by 76 silicon, and 74 hydrogen atoms which saturate the dangling bonds at the surface. The N-N core is totally surrounded by as much Si atoms as possible to properly simulate the host crystal. Given the fact that the CMM is relatively small, special care was taken to ensure that the symmetry is not broken for the defect and neighboring silicon atoms.

MNDO-AM1 method considers only valence electrons in the calculation of the electronic states, treating the inner-shell electrons together with the nucleus as a core. The electron-electron, core-core, and core-electron interactions are obtained empirically. We used a restricted closed-shell wavefunctions, which constrain all molecular orbitals to be either doubly occupied or empty.

Within DMol[3] and MNDO-AM1 the LVMs are determined from the Hessian matrix using the Harmonic Oscillator Approximation. This approach is known to adequately describe the vibrational behavior of molecules and crystals at low temperature, where only the lowest vibrational levels are populated and the displacements from equilibrium are small. Usually, the harmonic vibrational frequencies produced by ab-initio calculations are larger than the experimentally fundamental lines by 5 to 10%. Both methods neglect the effect of anharmonicity, which is insignificant at the ground state.

The force constant for each pair of bonded atoms are obtained by diagonalization of the mass-weighted Hessian matrix element, defined as:

$$H_{i,j}^{m} = \frac{1}{\sqrt{m_i m_j}} \frac{\delta^2 E}{\delta x_i \delta x_j} \tag{32}$$

Only change in the dipole moment induces measurable IR transitions. For bonds which have a weak dipole moment the polarizability is usually high and the vibrational states of the bond are Raman active (RA). These two kinds of activities are not always mutually exclusive as for non-centrosymmetric molecules (or unit cells), some vibrations can be both Raman and infrared active.

The IR absorption intensity I_i and effective charge e_i of the i^{th} normal mode are evaluated from the dipole moment derivatives with respect to the vibrational coordinates q_i such that I_i is proportional to e_i^2 (Leigh & Szigetti, 1967; Whalley, 1972):

$$I_i = \frac{N_0 \pi}{3c^2 (4\pi\varepsilon_0)} e_i^2 \tag{33}$$

$$e_i = \left[\left(\frac{\partial p_x}{\partial q_i} \right)^2 + \left(\frac{\partial p_y}{\partial q_i} \right)^2 + \left(\frac{\partial p_z}{\partial q_i} \right)^2 \right]^{1/2} \tag{34}$$

where N_0 is the Avogadro number, c the speed of light, ε_0 dielectric constant in vacuum and p the electrical dipole moment:

$$p = \sum_i e_i q_i \tag{35}$$

The dipole moment derivatives are accurately calculated from the energy gradients $\frac{\partial E}{\partial q_i}$ (Galabov & Dudev, 1996)

$$\frac{\partial p}{\partial q_i} = \frac{\partial}{\partial f} \left(\frac{\partial E}{\partial q_i} \right) \qquad \text{where} \quad \frac{\partial E}{\partial q_i} = \left(\frac{\partial}{\partial q_i} \right) \langle \Psi | H - pf | \Psi \rangle \tag{36}$$

Where f is the applied electric field, ψ the wavefunction and H the Hamiltonian.
The calculated IR active modes were compared to the measured low temperature (T < 15K) FTIR absorption bands since the vibrational spectra are calculated from the ground state specifically from the minimum energy at 0 K corrected by the zero point energy.

5.3 Results and discussion

5.3.1 Formation energy of N_2, VN_2, and V_2N_2 complexes

The equilibrium geometries for N-pair defects are summarized in terms of bond length, bond angle in Table 6. The formation energies are summarized in Table 7. We found that the formation energy for N_2 (reaction R1) is about -3.95 eV at the ground state showing that N_2 is a very stable complex. Being highly exothermic, we believe that chemical reaction R1 occurs at earliest stages of point defect clustering, mainly at high temperature close to the melting point (1423°C) because of the high mobility of N_i. N_i has a low diffusion barrier of 0.4 eV (Schultz & Nelson, 2001).

	N-N (Å)	Si-N (Å)	Si-Si (Å) (*1st neighbor*)	∠Si-N-Si	∠N-Si-N	∠Si-N-N
N_2	2.45	1.73,1.76	2.32, 2.40	90.8	89.2	-
VN_2	1.43	1.80	2.33, 2.42	131.4	-	114.4
V_2N_2	3.55	1.82	2.38, 2.35	118	-	-

Table 6. N-pair relaxed geometry parameters from DFT-DMol³ calculations.

E_f(eV)	N_2	VN_2	V_2N_2	N_2O	N_2O_2	VN_2O	VN_2O_2	V_2N_2O	$V_2N_2O_2$
DFT DMol³ (this work)	-3.95	-0.21 (R2) -1.8 (R3)	-4.62 (R4) -4.42 (R5)	-0.96	-1.4	-0.15	+0.51	-0.70	-0.95
Harris functional and VWN (Karoui et al. 2003; Sahtout Karoui,2004)	- 4.1	+2.0	-5.2 -1.0	-0.78	-1.52	+0.08	+0.33	-0.62	-1.31
(Sawada et al., 2000)	-4.3	-1.4	-4.55 -5.69						
(Kagashima et al., 2000)	-3.86	+0.33	-4.07 -3.61						
(Goss et al., 2003)	-3.67	-1.3	-3.7 -3.4						
(Kagashima et al., 2003)								-0.95	

Table 7. Formation energy of N-N and N-O defects from DFT-DMol³ Calculations.

The formation energy of VN_2 complex is about - 0.21 eV when formed from N_i and N_s (reaction R2) and is about -1.8 eV thus more stable, when formed from the coupling of N_2 with a Si vacancy (reaction R3). Reaction R2 occurs mainly when N_i and N_s coexist i.e. near melting temperature of Si. Nevertheless, the small energy gain suggests that VN_2 would easily dissociate at such high temperature. The formation of VN_2 complexes is more likely to happen through reaction R3 during the interstitial-substitutional diffusion process of N_2 around the void formation temperature. This infers that VN_2 should intermittently slip back to N_2 via reaction $VN_2 + I \rightarrow N_2$ and back to VN_2 through reaction $N_2 + V \rightarrow VN_2$ as suggested in (von Ammon et al., 2001). Though, VN_2 is a metastable complex, it is also foreseen as an active complex during crystal growth as it contributes to the formation of very stable grown-in N-related microdefects such as V_2N_2 (reaction R4 and R5). Indeed, our calculations show that both reactions forming V_2N_2 complexes are highly exothermic. The energy gain is about -4.62 eV for R4 and -4.42 eV for R5 revealing the high stability of V_2N_2 complexes independently of the chemical reaction pathway. The formation of V_2N_2 from the coupling of VN_2 with a vacancy is favored over the coupling of N_2 with a divacancy because V_2 cannot form at high temperature (crystal growth temperature). However, reaction R5 might occur during crystal cooling at temperatures lower than 300°C, the survival temperature range of V_2. Indeed, V_2 is known to be immobile and stable at room temperature and to anneal out around 200-300°C. The calculated formation energy for N_2, VN_2 (R3) and V_2N_2 agree with previous work (Sawada et al., 2000; Kagashima et al., 2000; Goss et al., 2003).

5.3.2 Vibrational spectra of N_2, VN_2, and V_2N_2 complexes

As shown in Table 8, N_2 interstitial display four vibrational modes among them two are IR active asymmetric stretching, 779 cm^{-1} and 986 cm^{-1}, and two Raman active symmetric stretching (dipole-forbidden), 743 cm^{-1} and 1084 cm^{-1}. Since, the symmetry group of N_2 is C_{2h}, it has normal modes belonging to the irreducible representations A_g, B_g, A_u, and B_u; A_g and B_g being Raman active, A_u and B_u IR active. The selection rule for absorption in the IR spectrum is that the vibration must have the same symmetry as a *p*-orbital. Choosing z as the principal axis of symmetry (axis C_2 for C_{2h}), 779 cm^{-1} line transforms as B_u by the symmetry operations of the group, which has the same symmetry as p_x and p_y orbitals. Vibrational mode 986 cm^{-1} transforms as A_u which has the same symmetry as the p_z orbital. The two IR active modes relate to nitrogen. In these modes, the two N atoms are dynamically coupled and move in the same direction, along [001] and [110] respectively, see Fig. 14 (a), (b). These two lines match measured FTIR frequencies 771 cm^{-1} and 967 cm^{-1}. The absorption intensity ratio $\alpha(779)/\alpha(986)$ between the two IR modes as given by MNDO-AM1, is about 0.78. Lines 1084 and 743 cm^{-1} transform as A_g and B_g respectively and are Raman active with the same symmetry as s- and d-orbitals. For these modes, the N atoms move in opposite directions along [110] and [001], respectively. Our calculated LVMs for N_2 agree with reported values by Goss et al (Goss et al., 2003).

VN_2 complex has two IR active LVMs , 585 cm^{-1} and 781 cm^{-1}, and one Raman active line 997 cm^{-1}, see Table 9. The symmetry group of VN_2 is D_{2d} which has normal modes belonging to the irreducible representations A_1, A_2, B_1, B_2, and E. A_1 and B_1 are Raman active; B_2 and E are IR and Raman active with E representation doubly degenerate. The 585 cm^{-1} mode transforms as B_2 and involves an in-phase asymmetric stretching of the two N atoms along [001]. The 781 cm^{-1} absorption band is doubly degenerate and thus belongs to the E

	Symmetry	Vibrating atoms	Frequencies (cm^{-1}) Calculated	Measured (T<15K)	Intensity (km/mol)[1]	Effective Charge(e)[1]	Activity
N_2							
(C_{2h})	B_g	N-N	743		0	0	RA
	B_u	N-N	779		335	0.59	IR
	A_u	N-N	986		542	0.75	IR
	A_g	N-N	1084		0	0	RA
N_2O							
		O(ν_1)	666	771, 967 (551,653,782,790)*	33	0.18	IR
		N-N	750		19	0.14	IR
		N-N	814		316	0.59	IR
		N-N+ O(ν_3)	1003		975	1.02	IR
		N-N+ O(ν_3)	1029		289	0.55	IR
		N-N	1137		~0	~0	RA
N_2O_2							
		O-O(ν_1)	658	806,815,1000, 1021, 1031, 739*	105	0.32	IR
		O-O(ν_1)	665		0	0	RA
		N-N	780		0	0	RA
		N-N	825		409	0.66	IR
		N-N	945		1083	1.07	IR
		O-O (ν_3)	1016		0	0	RA
		O-O(ν_3)	1019		482	0.71	IR
		N-N	1027		0	0	RA

[1] As given by semi-empirical MNDO-AM1 quantum mechanics calculations on a macro-molecule.

Table 8. Vibrational spectra for N_2, N_2O and N_2O_2 from DFT-DMol[3] Calculations.

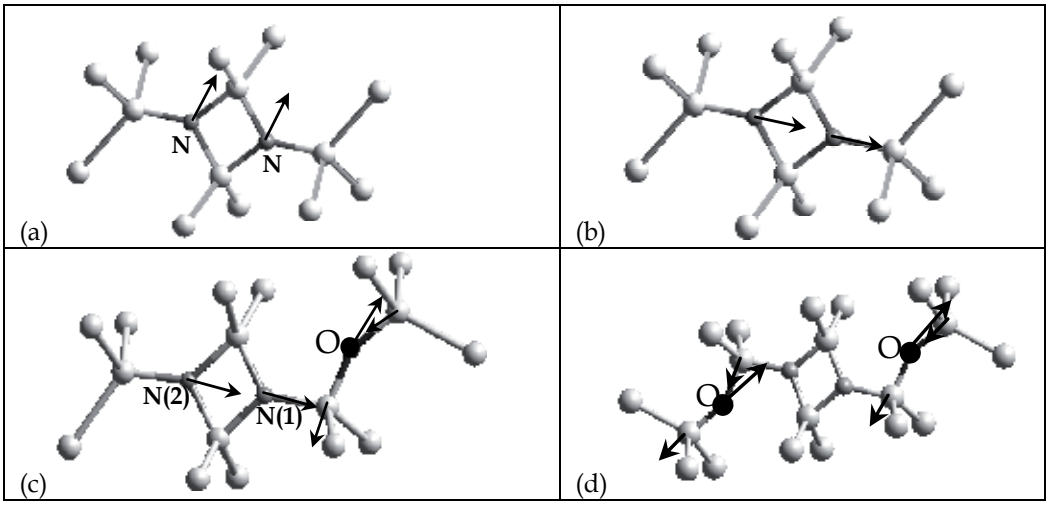


Fig. 14. (a) N_2 779 cm^{-1} N-N asymmetric stretching along [001]; (b) N_2 986 cm^{-1} N-N asymmetric stretching along [110]; (c) N_2O: 1003 cm^{-1} N-N + O (ν_3) asymmetric stretching; (d) N_2O_2 1019 cm^{-1} O (ν_3) asymmetric stretching.

	Symmetry	Vibrating atoms	Frequency (cm^{-1}) calculated	Measured (T<15K)	Intensity (km/mol)	Effective Charge (e)	Activity
VN_2 (D_{2d})	B_2	N-N	585		107	0.33	IR
	E	N(1)[1]	781		183	0.43	IR
		N(2)[1]	781		183	0.43	IR
	A_1	N-N	997		0	0	RA
VN_2O				771, 967			
		N-N	578	(551, 653, 782, 790)*	124	0.36	IR
		O(ν_1)	642		~ 0	-	IR
		N(1)	797		141	0.38	IR
		N(2)	835		141	0.38	IR
		N-N	985		0	0	RA
		O(ν_3)	1049		320	0.57	IR
VN_2O_2		N-N	567		121	0.35	IR
		O-O(ν_1)	644	806, 815, 1000, 1021, 1031, 739*	~ 0	-	IR RA
		N-N	819		208	0.44	IR
		N-N	828		208	0.44	IR
		N-N	976		0	0	RA
		O-O(ν_3)	1103		302	0.55	IR
		O-O(ν_3)	1113		0	0	RA

[1] N(1) and N(2) are respectively, the inner and outer nitrogen atom in the N-N-Si-O-Si branch.

Table 9. Vibrational spectra for VN_2, VN_2O and VN_2O_2 defects using DFT-DMol[3] Calculations.

representation. For this mode, the N atoms are dynamically decoupled moving in opposite direction along $[1\,1\,0]$ and $[\bar{1}\,1\,0]$ respectively. 781 cm^1 vibrational mode is close to measured 782 and 790 cm^{-1} lines due to unknown defects (Goss et al., 2003) which probably relate from our calculations to VN_2 defects. VN_2 IR active lines have equivalent strength. The Raman active mode 997 cm^{-1} transforms as A_1 and involves an in-phase symmetric stretching of the nitrogen atoms along the principal axis [001]. The calculated frequencies for VN_2 as well as their spectroscopy activity are in accordance with reported values for the N_i-N_s complex (Goss et al., 2003). The 585 cm^{-1} line matches measured 551 cm^{-1} band detected in N-implanted and laser annealed FZ Si wafer which was attributed to N-substitutional (Stein, 1985). We believe that 551 cm^{-1} line is due to a localized vibration mode of VN_2 defect in N-implanted FZ Si crystals where vacancies are in excess.

V_2N_2 complex has three degenerate LVMs, each one IR and RA active, 615 cm^{-1}, 625 and 637 cm^{-1}, Table 10. The 615 cm^{-1} line is weak and involves an in-phase asymmetric (IR active mode) and a symmetric stretching (RA active mode) of the N atoms along the pair axis in the [111] direction with a small deviation along [100]. In this mode, the N atoms are dynamically coupled and the Si-N bonds stretch in a same way. The 625 cm^{-1} and 637 cm^{-1} modes involve each one a ν_3 type in-phase asymmetric (IR active mode) and a symmetric (RA active mode) vibration of the N atoms in the plane perpendicular to the pair axis. V_2N_2 structure has a D_{3d} symmetry which has normal modes belonging to A_{1g}, E_g, A_{2u}, E_u irreducible representation. A_{1g} and E_g are Raman active, and A_{2u}, E_u are IR active. E-type

	Symmetry	Vibrating atoms	Frequencies (cm^{-1}) Calculated	Measured (T<15K)	Intensity (km/mol)	Effective Charge (e)	Activity
V_2N_2	D_{3d}						
	E_u	N-N	615	771, 967	19	0.04	IR, RA
	E_g			(551, 653, 782, 790)*			
	E_u	N-N	625		115	0.24	IR, RA
	E_g						
	E_u	N-N	637		236	0.49	IR, RA
	E_g						
V_2N_2O							
		O(ν_1)	651		43	0.11	IR, RA
		N(1)	696		168	0.41	R
		N(1)	709		172	0.42	IR
		N(2)	731	806, 815, 1000, 1021, 1031, 739*	201	0.51	IR
		N(2)	819		201	0.51	IR
		O(ν_3)	1068		389	0.63	IR
$V_2N_2O_2$							
		O-O(ν_1)	649		0	0	RA
		N-N	729		358	0.6	IR, RA
		N-N	810		210	0.46	IR,RA
		O-O(ν_3)	1061		702	0.84	IR,RA

Table 10. Vibrational spectra for V_2N_2, V_2N_2O and $V_2N_2O_2$ defects from DFT-DMol[3] Calculations.

modes are double degenerate. We found that V_2N_2 relaxes in D_{3d} structure, each LVM pair pertains to the (E_u, E_g) irreducible representations. This in accordance with Cunha et al.(Cunha, 1993) whose work also reported that a substitutional N_2 having a D_{3d} structure preserve the D_{3d} symmetry during relaxation, while Goss et al. calculations (Goss et al., 2003) showed that a starting C_{3v} geometry relaxed in D_{3d} structure. FTIR measured 653 cm^{-1} line detected in N-implanted FZ Si is close to calculated IR modes for V_2N_2 defect, which show that this line is generated by localized vibrational modes of N substitutional as stated by Stein (Stein, 1985). The calculated absorption spectra for N_2, VN_2 and V_2N_2 are shown in Fig. 15. We can easily see that N_2 complex bears the highest absorption intensity and that measured 771 and 967 cm^{-1} lines are IR signature for that defect.

5.3.3 Formation energy of N_2O_n, VN_2O_n, $V_2N_2O_n$ complexes (n=1, 2)

The equilibrium geometries for N-O defects are summarized in terms of bond length, bond angle in Table 11. When N_2 captures an O atom, the energy gain is about 0.96 eV for N_2O. When O_{2i} is trapped by the N-pair (N_2O_2), the energy gain is about 1.4 eV. Likewise, the energy gain is about 0.70 eV and 0.95 eV for V_2N_2O and $V_2N_2O_2$, respectively. In contrary, the chemical reaction forming VN_2O (R8) is slightly exothermic (E_f = - 0.15 eV), and when capturing an oxygen dimer (R9), the formation energy becomes positive (0.51 eV) meaning that the so-formed VN_2O_2 is unstable. Although, the chemical reactions which form N_2O_n and $V_2N_2O_n$ (n=1,2) are energetically favorable, the limited energy gain render these complexes unstable at elevated temperature. Upon heating, the oxygen atoms would break free and easily diffuse in the matrix, leaving the N-pair intact. Indeed, FTIR measurements have shown a reversible formation and dissociation mechanism between N-N and N-O complexes upon successive heating and cooling (Wagner et al., 1988; Qi et al., 1992).

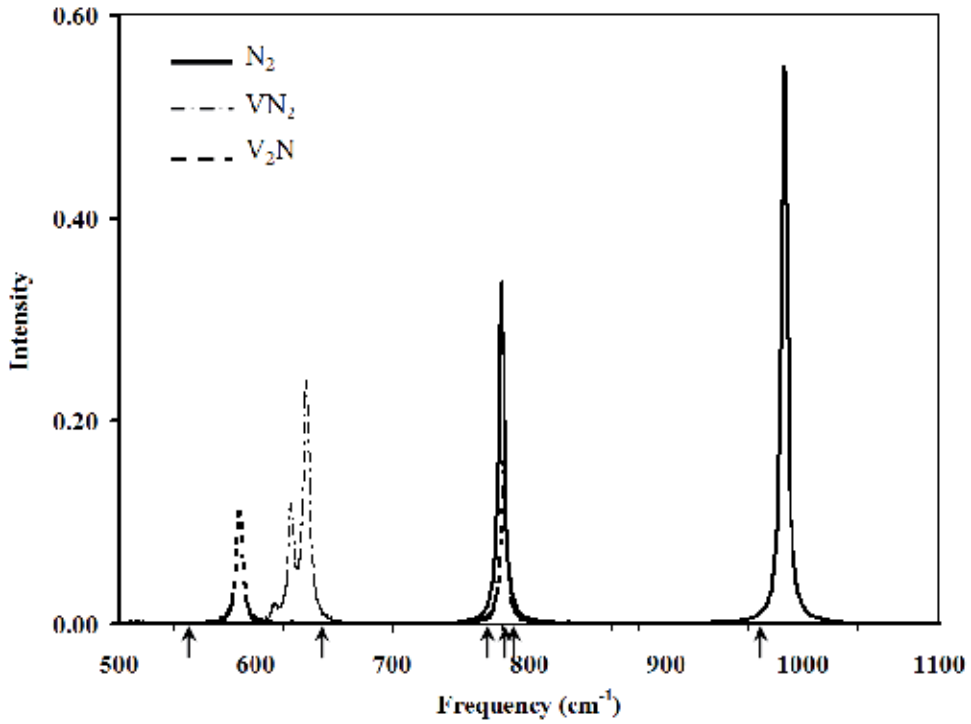


Fig. 15. Vibrational spectra of N_2, VN_2 and V_2N_2 complexes from DFT-DMol^3 Calculations. The arrows shows the FTIR measured frequencies.

	N-N (Å)	Si-N (Å)	Si-O (Å)	Si-Si (Å) (1st neighbor)	∠Si-O-Si
N_2O	2.43	1.70 (O) ,1.72 1.74 (sides of the diamond)	1.64 (N), 1.66	2.30 (O) 2.31-2.38	136
N_2O_2	2.38	1.68 1.73 (sides of the diamond structure)	1.64 (N), 1.65	2.29	136.6
VN_2O	1.44	1.77 (O), 1.80	1.62 (N), 1.64	2.29-2.42 (O) 2.33-2.44	141.5
VN_2O_2	1.45	1.77 (O), 1.80	1.61 (N), 1.63	2.29-2.42	147
V_2N_2O	3.62	1.77 (O), 1.82	1.61 (N), 1.65	2.33-2.41 (O) 2.35-2.40	143.6
$V_2N_2O_2$	3.72	1.77 (O), 1.79, 1.82	1.62 (N), 1.65	2.34-2.41	143.5

Table 11. N-O complexes relaxed geometry parameters from DFT-DMol^3 calculations. (O) means Si-N bond length in the N-Si-O-Si branch; (N) means Si-O bond length in the N-Si-O-Si branch.

5.3.4 Vibrational spectra of N_2O and N_2O_2 complexes (n=1,2)

Calculated vibrational spectra for N_2O displays five IR active LVMs 666, 750, 814, 1003, and 1029 cm^{-1} and one dipole forbidden transitions at 1137 cm^{-1}, see Table 8. These lines shift to 658, 665, 825, 945 and 1019 cm^{-1} for N_2O_2. The 779 cm^{-1} mode of N_2 shifts to 814 cm^{-1} when

capturing one O atom (N_2O) and to 825 cm^{-1} when capturing an oxygen dimer (N_2O_2). This blue shift can be explain by the shortening of the N-Si bonds as a result of the insertion of the O atom(s) in the vicinity which causes the N-Si bond stretching force constant to increase so the frequency. As for N_2, the 814 cm^{-1} and the 825 cm^{-1} modes involve N-N asymmetric stretching in the $[1\bar{1}0]$ direction, with the highest absorption strength on the nitrogen atom neighboring the O atom in the N_2O complex. These two modes are close to the measured 806 cm^{-1} and 815 cm^{-1} FTIR frequencies, ascribed respectively to NNO and NNO_2 defects (Jones et al., 1994). The 750 cm^{-1} mode is a symmetric stretching of the N atoms in [110]. This mode has a small strength because of the imbalanced mass due to the O atom in the vicinity of one N atom. The 1003 cm^{-1} mode (N_2O) combines a Si-O-Si asymmetric stretching of ν_3 type (Fig. 7) and an in-phase N-N asymmetric stretching along [110] analog to the N_2 986 cm^{-1} mode, a blue shift of 17 cm^{-1}, see Fig. 14 (c). This frequency matches the measured 1000 cm^{-1} absorption band. N_2 986 cm^{-1} wavenumber shifts down to 945 cm^{-1} when an oxygen dimer interacts with N_2 producing N_2O_2. As for N_2, this mode involves an N-N asymmetric stretching along [110]. The 945 and 1003 cm^{-1} frequencies have almost identical strength and bear the highest absorption intensity. N_2O 1029 cm^{-1} mode involves an N-N asymmetric stretching combined with a vibration of ν_3 type on the O atom. N_2O_2 1019 cm^{-1} mode is of ν_3 type and involves only the O atoms, see Fig. 14 (d).

Several low frequency local vibrational modes appeared in our calculations. These modes are of ν_1 type (Fig. 7) and are entirely due to Si-O stretch. The 666 cm^{-1} (N_2O) and 658 cm^{-1} (N_2O_2) vibrational modes are IR active, 665 cm^{-1} (N_2O_2) is Raman active. Our results show that calculated 814, 1003, 1029 cm^{-1} frequencies for N_2O fit (within a margin of less than 1%) the measured 806, 815, 1000, and 1031 cm^{-1} (low temperature) FTIR lines for N-O complexes. Alike, N_2O_2 1019 cm^{-1} mode matches the 1021 cm^{-1} FTIR line. Measured 1021 cm^{-1} is an IR signature for N_2O_2 and 1031 cm^{-1} N_2O in accordance with what has been suggested by Jones et al. (Jones et al., 1994) based on FTIR measurements.

N_2O defect has been previously theoretically investigated (Ewels, 1997; Jones et al., 1994) using AIMPRO method and H-terminated cluster of about the same size as the macro-molecule we used in the MNDO-AM1 calculations. Some discrepancies have been found in the AIMPRO calculations especially concerning the O_i vibration modes which were found between the two N modes in contrary to the observation. The O atom has to be displaced close to the neighboring Si atoms in order to fit the experimental lines.

5.3.5 Vibrational spectra of VN_2O_n and $V_2N_2O_n$ complexes (n = 1,2)

VN_2O and VN_2O_2 complexes have several LVMs ranging from 500 to 1100 cm^{-1}, see Table 9. VN_2O has five IR active modes and one Raman active mode. VN_2 585 cm^{-1} line shifts to 578 cm^{-1} for VN_2O and involves as for VN_2, an in-phase asymmetric stretching of the N atoms along [100]. This mode shifts to 567 cm^{-1} for VN_2O_2.

The degeneracy due to the D_{2d} symmetry of VN_2 is removed by the oxygen atom in the vicinity of the N-N core, which reduces the symmetry of the defect. As a result, VN_2 781 cm^{-1} frequency splits into two IR active lines, 797 cm^{-1} and 835 cm^{-1} for VN_2O defect. The N atoms are now dynamically decoupled. Each mode is of ν_3 type, one for each nitrogen atom, the N atom neighboring the oxygen bearing the highest frequency. 797 cm^{-1} and 835 cm^{-1} lines shift respectively to 819 cm^{-1} and 828 cm^{-1} for VN_2O_2 complex. These normal modes involve the N atoms which are now dynamically coupled and of ν_3 type. The dipole forbidden band 997 cm^{-1} of VN_2 remains IR inactive for VN_2O and VN_2O_2. This frequency shifts to 985 cm^{-1} for VN_2O

and to 976 cm^{-1} for VN_2O_2. We observe several additional LVMs involving exclusively the oxygen atom(s). VN_2O 642 cm^{-1} normal mode is of ν_1 type. This mode becomes doubly degenerate for VN_2O_2 and shifts to 644 cm^{-1} which is IR and Raman active. This frequency involves two ν_1 symmetric stretching, one on each O atom. The IR active mode is induced by an in-phase vibration of the two O atoms, while in the RA mode the O atoms move in opposite direction thus does not induce any change in the dipole moment. All ν_1 IR active modes are very weak. VN_2O has a high frequency mode at 1049 cm^{-1} and is caused by a ν_3 type asymmetric stretching of the O atom. This vibrational frequency shifts to 1103 cm^{-1} in VN_2O_2 complex and include an in-phase ν_3 type vibration of the O atoms. VN_2O_2 has a high frequency vibrational mode at 1113 cm^{-1} which is Raman active. For this mode, the O atoms vibrate in opposite directions, each Si-O-Si branch vibration being of ν_3 type. The 1103 and 1113 cm^{-1} frequencies are centered around the well known oxygen interstitial 1107 cm^{-1} line of nitrogen-free CZ silicon; the average shift being ± 5 cm^{-1}.

V_2N_2O defect has six IR active LVMs, Table 10. The degeneracy observed for V_2N_2 complex is removed and the degenerate levels, split into four IR active modes: 696, 709, 731, and 819 cm^{-1}. The dipole-forbidden transitions for V_2N_2 due to the D_{3d} symmetry are now allowed because of the reduced symmetry. Frequencies 696 cm^{-1} and 709 cm^{-1} involve a vibration of the outer N atom whereas the 731 cm^{-1} and 819 cm^{-1} lines involve a vibration of the inner N atom (neighboring the O atom). These LVMs are of ν_3 type and the N atoms are dynamically decoupled. The frequencies and vibrational intensities of the inner N atom are higher than those of the outer N atom because of the unbalanced mass center. The V_2N_2O complex has a high frequency mode at 1068 cm^{-1} of ν_3 type and involving exclusively the O atom. The 651 cm^{-1} absorption line is a ν_1 symmetric stretching of the O atom and is very weak.

All degenerated LVMs found for V_2N_2 core defect remain degenerated for $V_2N_2O_2$, formed by symmetrically trapping two O atoms. Three are IR and Raman active: 729 cm^{-1}, 810 cm^{-1}, and 1061 cm^{-1}; and one is Raman active: 649 cm^{-1}. The 729 cm^{-1} and 810 cm^{-1} absorption modes involve respectively, an in-phase symmetric and asymmetric N-N stretching in [111] perpendicular to the N-N center axis, similar to V_2N_2 615 cm^{-1} and 638 cm^{-1} lines. Frequency 649 cm^{-1} and 1061 cm^{-1} are exclusively oxygen atom related. The 649 cm^{-1} mode involves two dynamically coupled ν_1 LVMs, one for each Si-O-Si branch, in opposite phase. The 1061 cm^{-1} is induced by two pairs of coupled ν_3 stretching mode, one for each O atom. One pair is due to a symmetric vibration of the O atoms, therefore does not induce any change in the dipole moment, while the other pair is an asymmetric movement, and is thus IR active. This mode bears the highest strength.

Lines 729 cm^{-1} ($V_2N_2O_2$) or 731cm^{-1} (V_2N_2O) might be assigned to the measured FTIR frequency 739 cm^{-1} observed in samples implanted with both N and O atoms and which was attributed to NNO complex (Berg Rasmussen, 1996). As a matter of fact, ion implantation creates an excess of vacancies that will couple with nitrogen to form the V_2N_2 complexes which will subsequently couple with oxygen. These two modes have equivalent strength. The 810 cm^{-1} ($V_2N_2O_2$) and 819 cm^{-1} (V_2N_2O) frequencies fit the FTIR measured 806 cm^{-1} and 815 cm^{-1} absorption lines for N-O defects.

5.3.6 Nucleation of extended defects and nitrogen concentration measurement

In order to comprehend how the N-related defects in N-doped Si shape the nucleation and growth of extended defects, and to accurately assess the nitrogen concentration in N-doped

silicon, it is necessary to compare the properties of each defect obtained from DFT calculations to experimental data. This is done by studying the equilibrium structures, and comparing the calculated vibrational modes with experimental Fourier-transform infrared spectra.

The high stability of N_2 and V_2N_2 complexes should explain the strong change observed in the kinetics of oxygen precipitation and void formation in N-CZ Si as compared to N-free CZ Si. The formation of V_2N_2 complexes lower the vacancy supersaturation during crystal cooling. N_2 might equally reacts with oxygen through reactions R6 and R7 or a silicon vacancy through reaction R3. Reactions R8 and R9 show that VN_2 complexes are less able to react with oxygen but would preferentially react with a Si vacancy (reaction R4) to form the very stable V_2N_2 defects which will in turn act as nucleation sites for oxygen precipitates.

It appears from this study that V_2N_2O and $V_2N_2O_2$ complexes equally compete with N_2O and N_2O_2 as nucleation sites for O precipitation. They are the most stable N-O defects since the capture of O_i or O_2 by N_2 and V_2N_2 are all exothermic reactions, and the N-pair complexes (N_2 and V_2N_2) from which they originate are extremely stable. However, N_2O_n defect should be the dominant defect since our results strongly support the assignment of the FTIR 771 and 967 cm^{-1} (15K) local vibrational modes to $N_{2'}$ meaning they are experimentally detectable. The process formation of N-O complexes is likely to follow the subsequent chemical pathways: $N_2 \rightarrow VN_2 \rightarrow V_2N_2 \rightarrow V_2N_2O \rightarrow V_2N_2O_2$ and $N_2 \rightarrow N_2O \rightarrow N_2O_2$. Therefore, nitrogen will increase the number of O-precipitates nucleation sites by coupling with vacancies and oxygen atoms, explaining the high density of as-grown oxygen precipitates and the decrease in the vacancy supersaturation observed in N-CZ Si. Consequently, the formation of void will be hindered and their density will decrease compared to N-free CZ Si. These results combined with our previous results obtained from molecular mechanics force field calculations (Karoui et al., 2003; Sahtout Karoui et al., 2004) and experimental measurements (Wright etching, STEM, HRTEM, and Oxygen Precipitates Profiler) (Karoui et al., 2004a, 2004b, 2002) confirm that N_2 and V_2N_2 are much more likely to adsorb O atoms than to trap vacancies thus act as nucleation centers for oxygen precipitation rather than voids. However, N-O complexes might also co-exist in the oxide layer covered walls of the voids in N-doped Si. Indeed, EDS measurements on N-doped CZ Si samples showed that voids are covered with an oxide layer as in the case of undoped crystals (Takahashi et al., 2003).

All studied structures of N-O complex have IR active lines falling around measured 806 and 815 cm^{-1} (15K) lines : 814 cm^{-1} (N_2O), 825 cm^{-1} (N_2O_2), 819 cm^{-1} (V_2N_2O), 810 cm^{-1} ($V_2N_2O_2$), 797 cm^{-1} and 835 cm^{-1} (VN_2O), 819 cm^{-1} and 828 cm^{-1} (VN_2O_2). Therefore, the absorption intensity of lines 806 cm^{-1} and 815 cm^{-1} (15K) have to be taken into account when evaluating nitrogen concentration in N-CZ Si or O-rich N-FZ Si wafers. So, a more accurate calibration relationship for [N] measurement in N-CZ Si or O-rich N-FZ Si would be:

$$[N] = (1.83 \pm 0.24)x\,10^{17}x\,[\alpha_{967} + \alpha_{806} + \alpha_{815}]\ at/cm^{-3}\ (T<15K) \quad (37)$$

$$[N] = (1.83 \pm 0.24)x\,10^{17}x\,[\alpha_{963} + \alpha_{801} + \alpha_{810}]\ at/cm^{-3}\ (at\ 300K) \quad (38)$$

This is in agreement the experimental calibration curve proposed by Qi et al. (Qi et al., 1992) based on FTIR measurements. For N-implanted O-rich FZ wafer the absorbance of 653 cm^{-1} line has to be considered in the calibration formula because of the important excess of vacancy created during the implantation process. These absorption bands were found to be due to localized vibration modes of substitutional nitrogen (Stein, 1985) and appear from

this study to actually relate to V_2N_2 complexes, therefore the nitrogen calibration curve for N-implanted FZ-Si wafers would be:

$$[N] = (1.83 \pm 0.24)x\,10^{17}x\,[\alpha_{967} + \alpha_{653}]\ \text{at/cm}^{-3}\ (T<15K) \quad (39)$$

5.4 Conclusions

To comprehend the effect of nitrogen doping on vacancy aggregation, oxygen precipitation, and nitrogen concentration measurement in silicon, we have theoretically investigated using density functional theory (DMol[3]) and semi-empirical MNDO-AM1 method, the atomic structure, formation energy, and vibrational spectra of several dominant N-related complexes. The focus was on N-pairs occupying either an interstitial or a substitutional position and the associated N-O complexes consisting of bridging one or two oxygen interstitial atoms on the first Si-Si bond neighboring the N-N center. We found a good correlation between the degree of stability of the defect, the IR active energy levels foreseen by the symmetry of the defect and the calculated and measured IR active modes. We found, in agreement with an earlier theoretical study, that $N_i + N_i \rightarrow N_2$ and $VN_2 + V \rightarrow V_2N_2$ are highly exothermic chemical reaction thus are very stable complexes. Both complexes are believed to chiefly form and coexist during crystal growth. The VN_2 complex is a metastable species playing a central role in the formation of very stable V_2N_2 defects. The N-O complexes formed from N_2 and V_2N_2 defects that is N_2O_n and $V_2N_2O_n$ (n=1, 2) are the most stable among studied N-O complexes suggesting that they act as nucleation centers for oxygen precipitation. It is likely that vacancy concentration during crystal growth is affected by the following chemical reaction pathway $N_2 \rightarrow VN_2 \rightarrow V_2N_2 \rightarrow V_2N_2O \rightarrow V_2N_2O$, which decreases the vacancy supersaturation, delaying the onset of vacancy clustering and lessening the void density.

Our results strongly support the assignment of 771 cm^{-1} and 967 cm^{-1} (T<15K) absorption bands to N_2 pairs in split interstitial positions. Our calculations show that 551 cm^{-1} line and 653 cm^{-1} detected in N-implanted FZ Si crystals relates to nitrogen substitutional and are caused by the vibration of VN_2 and V_2N_2 complexes. Unexplained measured 782 and 790 cm^{-1} FTIR lines probably relate to VN_2 defects. Our DFT calculations on N-O complexes show that N_2O and N_2O_2 complexes best match the FTIR absorption bands measured for N-O complexes. N_2O_n, VN_2O_n and $V_2N_2O_n$ (n = 1, 2) have IR absorption bands around measured 806 cm^{-1} and 815 cm^{-1} lines and all relate to nitrogen vibrations. We found that FTIR 1021 cm^{-1} absorption band is an IR signature for N_2O_2 and that 1000 cm^{-1} and 1031 cm^{-1} relate to N_2O vibrations. The 739 cm^{-1} line measured in N implanted FZ Si wafers originates from V_2N_2O or $V_2N_2O_2$ vibrations.

The degree of stability and matching infrared vibrational spectra suggest that N_2O_n as well as $V_2N_2O_n$ complexes develop during crystal growth and wafer annealing and that both act as nucleation site for oxygen precipitates. The increase in the number of nucleation sites due to nitrogen-vacancy-oxygen coupling explain the high density of grown-in oxygen precipitates and the delay in void formation observed in N-doped Si crystals.

Since all studied N-O complexes have normal modes falling around 806 cm^{-1} and 815 cm^{-1} (T<15K) measured FTIR lines, these absorption bands have to be considered in the [N] calibration relationship for N-CZ and O-rich N-FZ Si, $[N] = (1.83 \pm 0.24)x\,10^{17}x\,[\alpha_{967} + \alpha_{806} + \alpha_{815}]$ at/cm^{-3} (T<15K). For N-FZ implanted wafer the absorbance of 653 cm^{-1} line which relate to V_2N_2 have to be considered in the calibration formula because of the important excess of vacancy existing in implanted wafers.

6. Acknowledgements

The high performance computing has been carried out through user grants from North Carolina supercomputing center and the high performance computing center at East Carolina State University (U.S.A.). The second author acknowledges partial support by DOD under contract No. W91CRB-10-C-0321.

7. References

Abe, T.; Masui, T.; Harada, H.; & Chikawa, J. (1985). *VLSI science and technology, Proceedings of the Second International Symposium on Very Large Scale Integration Science and Technology*, Electrochemical Society, Pennington, NJ, 1985, pp. 543-550.

Aihara, K. , Takeno, H.; Hayamizu, Y.; Tamatsuka, M.; & Masui, T. (2000). Enhanced nucleation of oxide precipitates during Czochralski silicon crystal growth with nitrogen doping. *J. Appl. Phys.*, Vol. 88, No. 6, pp. 3705-3707.

Atkins, Peter; Friedman, Ronald (Ed(s)). (2011). *Molecular Quantum Mechanics*, Oxford University Press Inc, ISBN 978-0-19-954142-3, New York, U.S.A.

Berg Rasmussen, F.; Öberg, S.; Jones, R.; Ewels, C.; Goss, J.; Miro, J.; & Deák, P. (1996). The nitrogen-pair oxygen defect in silicon. *Mat. Sc. Eng. B*, Vol. 36, pp. 91-95.

Bernath, Peter F. (Ed(s).). (1995). *Spectra of Atoms and Molecules*, Oxford University Press Inc, ISBN 0-19-507598-6, New York, U.S.A.

Blöchl, P. E.; Jepsen, O.; & Andersen, O. K. (1994). Improved tetrahedron method for Brillouin-zone integrations. *Phys. Rev. B*, Vol. 49, No. 23, pp. 16223-16233.

Brower, K.L. (1982). Deep-level nitrogen centers in laser-annealed ion-implanted silicon. *Phys. Rev. B*, Vol. 26, No. 11, pp. 6040–6052.

Chiou, H-D.; Moody, J.; Sandford, R.; and Shimura, F. (1984). Effects of Oxygen and Nitrogen on slip in Cz Wafers. *VLSI science and technology, Proceedings of the Second International Symposium on Very Large Scale Integration Science and Technology*, Vol. 84-7, pp. 59-65, Cincinnati, Ohio, 1984.

COMPASS force field potential developed by Accelrys Inc. (USA).

Coomer, B. J.; Resende, A.; Goss, J. P.; Jones, R.; Oberg, S.; & Briddon, P. R. (1999). The divacancy in silicon and diamond. *Physica B*, Vol. 273-274, pp. 520-523.

Cunha, C.; Canuto, S.; & Fazzio, A. (1993). Role played by N and N-N impurities in type-IV semiconductors, *Phys. Rev. B*, Vol. 48, No. 24, pp. 17806–17810.

Dean, P.; & Martin, J. L. (1960). A Method for Determining the Frequency Spectra of Disordered Lattices in Two-Dimensions. *Proc. Roy. Soc.*, A 259 , No. 1298, pp. 409-418.

Delley, B. J. (1990).. An all-electron numerical method for solving the local density functional for polyatomic molecules. *J. Chem. Phys.*, Vol. 92, No. 1, pp. 508-517.

Delley, B. (2000). From molecules to solids with the DMol^3 approach. *J. Chem. Phys.*, Vol. 113, Vol. 508, pp. 7756-7764.

Delley, B. (1995). Modern Density Functional Theory: A Tool for Chemistry. *Theoretical and Computational Chemistry*, Vol.2 ; Elsevier 1995.

Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; & Stewart, J. J. P. (1985). Development and use of quantum mechanical molecular models. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, 107, pp. 3902-3909.

Ewels, C. (1997). Density Functional Modelling of Point Defects in Semiconductors, *PhD. Thesis*, University of Sussex, UK.

Faststructure in Cerius2 package, Accelrys Inc. (USA).

Fuma, Naoki ; Tashiro, Koichiro; Kakumoto, Katsunori; & Yukio, Takano (1996). Diffused Nitrogen-Related Deep Level in N-Type Silicon. *Jpn. J. Appl. Phys.*, Vol. 35,No. 4A, pp. 1993-1999.

Galabov, Boris S.; & Dudev, Todor (1996). *Vibrational Spectra and Structure : Vibrational Intensities*, James R. Durig Series Editor, Elsevier, 1996, ISBN 0444814973 & 9780444814975.

Goss, J. P.; Hahn, I.; Jones, R.;. Briddon, P. R ; Öberg, S. (2003). Vibrational modes and electronic properties of nitrogen defects in silicon. *Phys. Rev. B*, Vol.67, 045206-045216.

Hara, A.; Myabo, T.; & Hirai, L. (1989). Oxygen-nitrogen complexes in silicon formed by annealing in nitrogen. *Appl. Phys. Lett.*, Vol. 54, No. 7, pp. 626-628,

Harris, Daniel C.; and Bertolucci, Michael D. (Ed(s)). (1989). *Symmetry and Spectroscopy*, Dover publications Inc, ISBN 0-486-66144-X, New York, U.S.A.

Harris, J. (1985). Simplified method for calculating the energy of weakly interacting fragments. *Phys. Rev. B*, Vol. 31, pp. 1770–1779.

Hull, Robert (1999). *Properties of Crystalline Silicon,* Institution of Engineering & Technology, 1999, ISBN 978-0-86341-556-2.

Ikari, A.; Nakai, K.; Tachikawa, Y.; Deal, H.; Hideki, Y.; Ohta, Y.; Masahashi, N.; Hayashi, S.; Hoshino, T.; & Ohashi, W. (1999). Defect Control in Nitrogen Doped Czochralski Silicon Crystals. *Solid State Phenomena*, Vol. 69-70, pp. 161-166.

Itoh, Y; Nazaki, T.; Masui, T.; Abe, T. (1985). Calibration curve for infrared spectrophotometry of nitrogen in silicon. *Applied Physics Letters*, Vol. 47, No. 5, pp. 488-489.

Jones, R. ; Oberg, S.; Rasmussen, F. B.; & Nielsen, B. B. (1994). Identification of the dominant nitrogen defect in silicon. *Phys. Rev. Lett.*, Vol. 72, pp. 1882–1885.

Jones, R.; Ewels, C.; Goss, J.; Miro, J.; Deak, P.; Öberg, S.; & Rasmussen, F. Berg (1994). Theoretical and isotopic infrared absorption investigations of nitrogen-oxygen defects in silicon. *Semicond. Sci. Technol.*, Vol. 9, No. 11, pp. 2145-2148.

Kageshima, H.; Taguchi, A.; & Wada, K. (2000). Theoretical investigation of nitrogen-doping effect on vacancy aggregation processes in Si. *Applied Physics Letters*, Vol. 76, No. 25, pp. 3718-3720.

Kageshima, H.; Taguchi, A.; Wada, K. (2003). Formation of stable N–V–O complexes in Si. *Physica B*, Vol. 340–342, pp. 626-629.

Karoui, A.; Sahtout Karoui, F.; Rozgonyi, G. A.; & Yang, D. J. (2004). Oxygen precipitation in nitrogen doped Czochralski silicon wafers. I. Formation mechanisms of near-surface and bulk defects. *J. Appl. Phys.*, Vol. 96, No. 6, pp. 3255-3263.

Karoui, A. ; & Rozgonyi, G. A. (2004). Oxygen precipitation in nitrogen doped Czochralski silicon wafers. II. Effects of nitrogen and oxygen coupling. *J. Appl. Phys. II*, Vol. 96, No. 6, pp. 3264-3270.

Karoui, A.; Sahtout Karoui, F.; Kvit, A.; Rozgonyi, G. A.; & Yang, D. (2002). Role of nitrogen related complexes in the formation of defects in silicon. *Applied Physics Letters,* Vol. 80, No. 12, pp. 2114-2116.

Karoui, A.; Sahtout Karoui, F.; Rozgonyi, G. A.; Hourai, M.; & Sueoka, K. (2003). Structure, Energetics, and Thermal Stability of Nitrogen-Vacancy-Related Defects in Nitrogen Doped Silicon. *J. of the Electrochem. Soc.*, Vol. 150, No. 12, G771-G777.

Karoui, F. Sahtout; & Karoui, A. (2010). A density functional theory study of the atomic structure, formation energy, and vibrational properties of nitrogen-vacancy-oxygen defects in silicon. J. App. Phys., Vol. 108, No. 3, pp. p033513- p033525.

Kohn W. ; & Sham, L. J. (1965). Self-Consistent Equations Including Exchange and Correlation Effects. Phys. Rev., Vol. 140, No. 4A, pp. A1133-A1138.

Leigh, R.S.; & Szigetti., B. (1967). Absorption by the Vibrations of Uncharged Atoms. *Proc. Roy. Soc.*, Vol. 301, No. 1465, pp. 211-221.

Monkhorst H. J.; & Pack, J. D. (1976). Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13, pp. 5188-5192.

Murphy, J.D.; Alpass, C.R.; Giannattasio, A.; Senkader, S.; Falster, R.J.; & Wilshaw, P.R. (2006). Nitrogen in silicon: Transport and mechanical properties. *Nuclear Instruments and Methods in Physics Research*, Vol. 253, No. 1-2, pp. 113–117.

Nagai, Y. ; Inoue, K.; Tang, Z.; Yonenaga, I.; Chiba, T.; Saito, M.; & Hasegawa, M. (2003). Jahn–Teller distortion of neutral divacancy in Si studied by positron annihilation spectroscopy. *Physica B*, Vols. 340-342, pp. 518-522.

Nakai, K.; Inoue, Y.; Yokota, H.; Ikari, A.; Takahashi, J.; Tachikawa, A.; Kitahara, K.; Ohta, Y.; Ohashi, W. (2001). Oxygen precipitation in nitrogen-doped Czochralski-grown silicon crystals. *J. Appl. Phys.*, Vol. 89, No. 8, pp. 4301-4309.

Nichols, C. S.; Van de Walle, C. G.; & Pantelides, S. T. (1989). Mechanisms of dopant impurity diffusion in silicon *Phys. Rev. B,* Vol. 40, No. 8, 5484–5496.

Pesola, Marko (2000). Electronic Structure Calculations for Vacancies and Oxygen-Related Defects in Semiconductors. *Dissertation Thesis*, Helsinki University of Technology, Finland, 18th of August, 2000, ISBN 951-22-5102-7.

Pulay, P. (1980). Convergence acceleration of iterative sequences. the case of scf iteration Chem. *Physics Letter*, Vol.73, pp. 393-398.

Puska, M. J. ; Pöykkö, S.; Pesola, M.; & Nieminen, R. M. (1998). Convergence of supercell calculations for point defects in semiconductors: Vacancy in silicon. *Phys. Rev. B*, Vol. 58, No. 3, pp. 1318–1325.

Qi, M.W.; Tan, S.S.; Zhu, B.; Cai, P.X.; Gu, W.F.; Xu, X.M.; Shi, T.S.; Que, D.L.; & Li, L.B. (1991). The evidence for interaction of the N-N pair with oxygen in Czochralski silicon. *J. Appl. Phys.*, Vol. 69, No. 6, pp. 3775-3777.

Qi, M.W; Shi, T.S.; Tan, S.S.; Zhu, B.; Cai, P.X.; Liu, L.Q.; Que D.L.; & Li, L.B. (1992). On the Determination of Nitrogen in Czochralski Silicon. *Materials Science Forum*, Vols. 83-87, pp. 263-268.

Quast, F.; Pichler, P.; Ryssel, H.; & Falster, R. (2000). Vacancy-Nitrogen complexes in float-zone silicon. *Electrochemical Society Proceedings, High Purity Silicon VI*, Vols. 2000-17, pp. 156-163.

Rozgonyi, G. A. ; Karoui, A.; Kvit, A.; & Duscher , G. (2002). Nano-scale analysis of precipitates in nitrogen-doped Czochralski silicon. *Microelectronic Engineering*, Vol. 66, No. 1-4, pp. 305-313.

Sahtout Karoui, F. ; Karoui, A.; Rozgonyi, G. A.; Hourai, M.; & Sueoka, K. (2004). Characterization of Nucleation Sites in Nitrogen Doped Czochralski Silicon by Density Functional Theory and Molecular Mechanics. *Solid State Phenomena*, Vol. 95-96, pp. 99-104.

Sawada, H.; & Kawakami, K. (2000). First-principles calculation of the interaction between nitrogen atoms and vacancies in silicon. *Phys. Rev. B*, Vol. 62, No. 3, pp. 1851–1858.

Schultz, P.A. ; & Nelson, J.S. (2001). Fast through-bond diffusion of nitrogen in silicon. *Applied Physics Letters,* Vol. 78, No. 6, pp. 736-738.

Shaik Adam, L.; Law, M. E.; Szpala, S.; Simpson, P.J.; Lawther, D.; Dokumaci, O. ; & Hegde, S. (2001). Experimental identification of nitrogen-vacancy complexes in nitrogen implanted silicon. *Applied Physics Letters*, Vol. 79, No. 5, pp. 623-625.

Shimura, F.; & Hockett, R. S. (1986). Nitrogen effect on oxygen precipitation in Czochralski silicon. *Applied Physics Letters*, Vol. 48, No. 3, pp. 224-226.

Stein, H. J. (1983). Vibrational absorption bands for implanted nitrogen in crystalline silicon. *Applied Physics Letter*, Vol. 43, No. 3, 296-298.

Stein, H. J. (1986). Nitrogen in Crystalline Si. *Mat. Res. Soc. Symp. Proc.*, Vol. 59, 523-535.

Stein, H. J. (1987). Nitrogen Related Donors in Silicon. *J. Electrochem. Soc.*, Vol. 134, No. 10, pp. 2592-2596.

Stein, H. J. (1985). Infrared absorption band for substitutional nitrogen in silicon. *Appl. Phys. Lett.*, 47 (12), pp. 1339-1341.

Sumino, K.; Yonenaga, I.; & Imai, M. (1983). Effects of nitrogen on dislocation behavior and mechanical strength in silicon crystals, *J. App. Phys.*, Vol. 54, No.4, pp. 5016-5020.

Sun, Q.; Yao, K. H.; Gatos, H. C. ; & Lagowski, J. (1992). Effects of nitrogen on oxygen precipitation in silicon. *J. Appl. Phys.*, Vol. 71, No. 8, pp. 3760-3765.

Takahashi, Jun; Nakai, Katsuhiko; Kawakami, Kazuto; Inoue, Yoshiharu; Yokota, Hideki; Tachikawa, Akiyoshi; Ikari, Atsushi; & Ohashi, Wataru (2003). Microvoid Defects in Nitrogen- and/or Carbon-doped Czochralski-grown Silicon Crystals. *Jpn. J. Appl. Phys.*, Vol. 42, No. 2A, pp. 363-370.

Tamatsuka, M.; Kobayashi, N.; Tobe, S.; & Masui, T. (1999). High performance silicon wafer with wide grown-in void free zone and high density internal gettering site achieved via rapid crystal growth with nitrogen doping and high temperature hydrogen and/or argon annealing. *ECS. Proceedings, Defects in Silicon III*, Pennington N. J., 1999, Vol. 99-1, pp. 456-467.

Umerski A.; & Jones, R. (1993). The interaction of oxygen with dislocation cores in silicon. *Philosophical Magazine A*, 67(4), 905-915.

UniChem Software, Oxford Molecular Group, Inc.

Von . Ammon, W.; Dreier, P.; Hensel, W.; Lambert, U.; & Koster, L. (1996). Influence of oxygen and nitrogen on point defect aggregation in silicon single crystals, *Materials Science and Engineering B*, Vol. 36, No. 1-3, pp. 33-41.

Von Ammon, W.; Holzl, R.; Virbulis, J.; Dornberger, E.; Schmolke, R.; & Graf, D. (2001). The impact of nitrogen on the defect aggregation in silicon, *J. of Crystal Growth*, Vol. 226, pp. 19-30.

Vosko, S. H. ; Wilk, L.; & Nusair, M. (1980). Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, Vol.58, No. 8, pp. 1200-1211.

Wagner, P. ; Oeder, R.; & Zulehner, W. (1988). Nitrogen-oxygen complexes in Czochralski-silicon. Appl. Phys. A, Vol. 46, No. 2, pp. 73-76.

Watkins G. D.; & Corbett, J. W. (1965). Defects in Irradiated Silicon: Electron Paramagnetic Resonance of the Divacancy. *Phys. Rev.*, Vol. 138, No. 2A, pp. A543–A555.

Whalley, E. (1972). The Dipole Moment Derivative of the Hydrogen Bond in Ice. *Canadian Journal of Chemistry,* Vol. 50, No. 3, pp. 310-314.

Yang, D. ; Que, D.; & Sumino, K. (1998). Nitrogen–Oxygen Complexes in Silicon. *Phys. Stat. Sol. B*, Vol. 210, No. 2, pp. 295-299.

7

Metal-Assisted Proton Transfer in Guanine-Cytosine Pair: An Approach from Quantum Chemistry

Toru Matsui, Hideaki Miyachi,
Yasuteru Shigeta and Kimihiko Hirao
Department of Applied Chemistry, School of Engineering, The University of Tokyo
Japan

1. Introduction

Since the discovery of the double helix structure by Watson and Crick, deoxyribo-nucreic acid (DNA) has been the most important substance in molecular biology. DNA consists of base, sugar and phosphoric acid. There are four kinds of base in natural DNA, adenine (A), cytosine (C), guanine (G) and thymine (T) as shown in Figure 1.

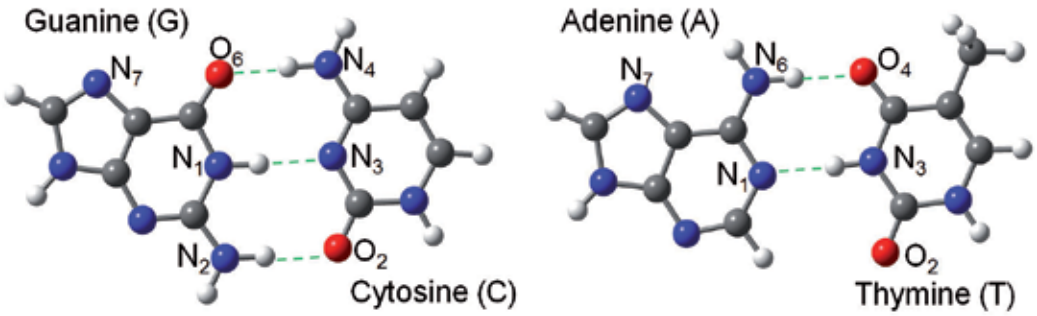


Fig. 1. The chemical structures of base pair (GC and AT pair), where the broken lines represents the hydrogen bond.

From the view of chemistry, DNA is one of the most remarkable examples for self-assembled materials existing in nature. It is said that the phosphoric acid has hydrophilic properties. This is why DNA is stable in aqueous phase. Moreover, there are two kinds of interaction which stabilizes the whole structures of DNA base pairs. These interactions are the driving force of self-assemble which leads to beautiful double helix structure. Structures of double-stranded DNA are stable in aqueous solution, because there are two kinds of interaction between Watson–Crick nucleobases. The first is the hydrogen bonding between nucleobases, where an A-T and a G-C base pairs have two and three hydrogen bonds per pair, respectively. The second is a base stacking interaction caused by π-π interaction between the base pairs. This type of interaction mainly originates from the van der Waals forces.

Hydrogen bond is important factor for interaction within DNA base pair. Many theoretical studies have computed the hydrogen bonding energy between nucleobases in a base pair. According to the most accurate computation up to date, hydrogen bonding energies of an adenine-thymine (AT) pair and a guanine-cytosine (GC) pair are about 60 and 110 kJ/mol, respectively [1, 2].

One of the most important topics in the hydrogen bonding between DNA bases is the proton-transfer (PT) reaction. In 1963, Löwdin proposed the possibility of a proton-tunneling model in the DNA base pairs [3]. He also suggested a simultaneous double proton-transfer (DPT) in AT and GC pairs, which may cause a mutation in the structure of DNA. The process of DNA mutation through proton-transfer (PT) may be estimated by computational chemistry using small molecular model systems such as several DNA base pairs. In particular, two types of the PT reaction in a GC pair (illustrated in Figure 2) have been studied extensively. One is a single proton-transfer (SPT) reaction, in which a hydrogen atom moves from N_1 of guanine (N_1 (G)) to N_3 of cytosine (N_3 (C)) as a proton. Thus, the SPT reaction causes ion pair G^-C^+ to form, where the proton donor G becomes negative and the proton acceptor C becomes positive. The other is the DPT reaction, in which two hydrogen atoms (one hydrogen atom is located at N_1 (G)-H…N_3 (C), the other is located at O_6 (G)…H-N_4 (C)) move to the other side of each hydrogen bond, which results in G^*C^* pair (G^* and C^* represent an isomer of guanine and cytosine, respectively). In the original DNA base pair, the DPT reaction does not affect the sum of charges on each base. Unfortunately, there is very little evidence to confirm the existence of proton-transferred base pairs at room temperature by experiments, because it is difficult to determine the position of each hydrogen atom in nucleobases accurately. In this circumstance, theoretical simulations can help one to understand whether or not the PT reactions occur in DNA base pair.

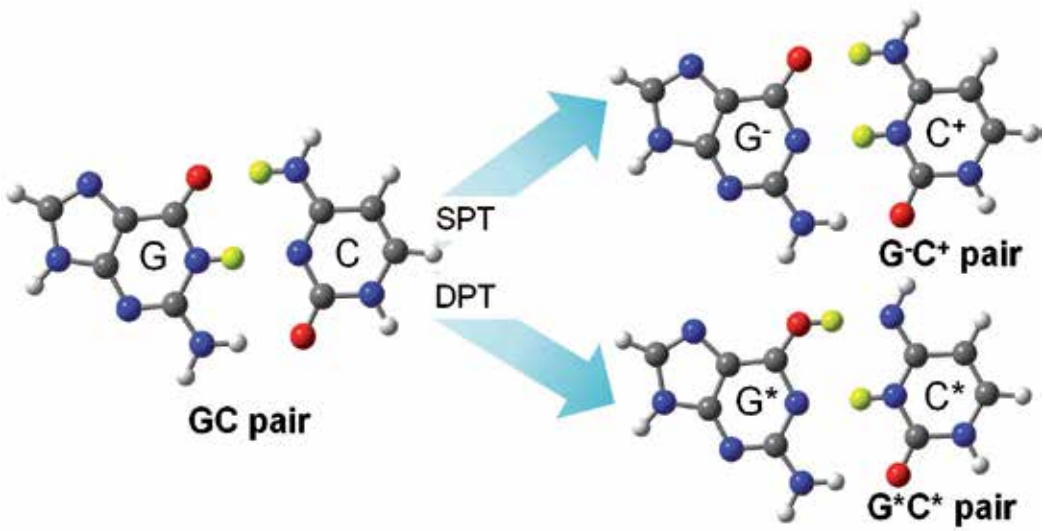


Fig. 2. Proton-transfer reaction between guanine-cytosine (GC) pair.

In order to understand the transition probability to the product or the dynamics simulation of PT reaction, we need the potential energy surface (PES) of the proton between base pair. Since the PES should be drawn within the error of chemical accuracy (1 kcal/mol= 4.2

kJ/mol), we use the quantum chemical approach to obtain the potential energy of the proton. The quantum chemical methods (e.g. Hartree-Fock (HF), post-HF theories and density functioanl theory (DFT)) are derived from the Schrödinger equation or Kohn-Sham equation. Since the introduction of the program package such as GAUSSIAN, the quantum chemical approaches have recently been a powerful tool to investigate the molecular properties which include the PES.

Many theoretical studies have investigated the PT reactions in AT and GC pairs. Florian et al. investigated the energetics of the PT reactions in GC and AT pairs [4, 5]. They reported that the GC pair was more stable than its DPT product（G*C*）by 10 kcal/mol, and the barrier height was estimated to be 14–20 kcal/mol. The DPT reaction would occur rather than the SPT reaction because the SPT reaction causes the charge separation, forming no stable product in the ground state. As an approach to the PES of proton between base pair, Villani has investigated the PES of the protons between AT and GC pair [6] and we also performed the dynamics simulation in semi-classical method [7].

However, these PT reactions can occur easily in the presence of chemical modification of nucleobases such as metal complex binding or ionization of GC pair [8]. We take a cisplatin ($[Pt(NH_3)_2Cl_2]$) bound GC pair as an example of metal complex bound DNA base pair [9, 10]. Since cisplatin was discovered by Rosenberg et al. in 1969 [11], Pt complexes have received much attention for their effects as antitumor drugs. Cisplatin distorts the structure of DNA by making a bridged structure with N_7 of guanine (G) or adenine (A). It causes a cell disorder that leads to apoptosis of the living cell. Because cisplatin contains only 11 atoms, it has been a good target for study by quantum chemistry and has been investigated from both the experimental and theoretical viewpoints. Experimentally, it is known that the bridged structure consists of 65% 1,2-d (GpG) (denoted as cis-G-Pt-G), 25% 1,2-d (ApG) (cis-G-Pt-A), and the rest is other bridged structures [12]. Nevertheless, the existence of 1,2-d (ApA) (cis-A-Pt-A) is difficult to confirm. The distorted DNAs are observed in X-ray analysis at 1.65–2.50 Å resolutions and in NMR experiments. These structures can be freely taken from the protein databank (PDB). We draw the whole structure of DNA and cisplatin bound DNA in Figure 3.

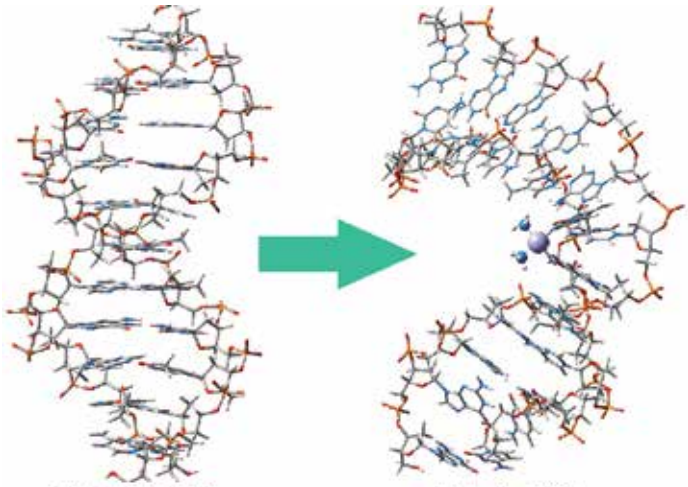


Fig. 3. Observed structure of normal DNA (B-type) and distorted structure by cisplatin.

From theoretical aspects, cisplatin has many interesting topics such as a ligand substitution, a hydration reaction, differences with transplatin and DNA binding. Ligands of cisplatin become NH_3 and/or H_2O by hydration reactions, depending on the pH of the solution. Note that the ligands of cisplatin in the human body are recognized as NH_3. In view of the DNA binding, there have been many studies of the reaction between cisplatin and DNA bases. In particular, Burda's group [13, 14] assumed [cis-Pt$(NH_3)_2$(N_7-G(or A)),(N_7-G(or A))]$^{2+}$ as the bridging structures and estimated binding energies of Pt-G and Pt-A. They found that the binding of base pair to the Pt atomtend to occur in the order of cis-G-Pt-G, cis-G-Pt-A, and cis-A-Pt-A. The angle of N_7-Pt-N_7 was about 90° for all cases. They confirmed that these bridges were stabilized by a hydrogen bond between G (or A) and a ligand of the Pt atom, such as H_2O and/or NH_3. Furthermore, reactions between cisplatin and two DNA purine bases (such as 1,2-d (GpG), 1,2-d (GpA)) were studied. Their calculated results of reaction barriers and the reaction constants of substitution reactions reproduced the experimental data well. There are two possible influences of Pt-DNA formation: (1) global structural changes, such as the distortion of the DNA duplex structure, or (2) local structural changes, such as a DNA mutation because of proton-transfer reactions. The former case is very hard to tackle with full quantum chemistry computation, because such a system is too huge study with currently available computational resource. For the latter case, the DNA mutation can be estimated using small molecular systems such as several DNA base pairs.
In this chapter, we will discuss the possibilities of simultaneous single proton transfers in one or two base pair(s). In Section 2, we will introduce a quantum chemical calculation methods, which we used on the basis of the quantum mechanics. As an application to the DNA-systems, we discuss the possibility of proton-transfer reactions in GC pair in Section 3 & 4. In Section 5, we give the concluding remarks.

2. A brief introduction to the quantum chemical calculations

In this section, we briefly explain the quantum chemical approaches used in this chapter. There are two main methods in the filed of quantum chemistry. The former is *ab initio* calculation, which is mainly used in Section 3. The latter is density functional theory (DFT), which is used in Section 4. Because of the limitation in this chapter, for the reader who wants to know the detail of the theory, we refer some textbooks for the quantum chemical approaches [15, 16].

2.1 *Ab initio* theory

The field of quantum chemistry starts from solving the Schrödinger equation for the electrons in the molecule. Because nuclei are much heavier than electrons, we can approximate that the electrons moves in the field of fixed nuclei. Then the Schrödinger equation for the electrons becomes equation (1).

$$\mathrm{H}_{\mathrm{elec}}\Psi = \left[\sum_{\mu} -\frac{1}{2}\nabla_{\mu}{}^{2} + \sum_{A}\sum_{\mu} -\frac{Z_A}{r_{A\mu}} + \sum_{\mu>\nu}\frac{1}{r_{\mu\nu}}\right]\Psi = E_{\mathrm{elec}}\Psi\,, \tag{1}$$

where A (and B in (2)), μ (and ν) represent the position of nuclei and electrons, respectively. Note here that we chose the "atomic unit" in which we set m=ħ=e=1 for simplicity. This approximation, called Born-Oppenheimer approximation, is central to quantum chemistry. The total energy E_{total} includes the repulsion between fixed nuclei, i.e.,

$$E_{total} = E_{\text{elec}} + \sum_{A>B} \frac{Z_A Z_B}{R_{AB}} . \tag{2}$$

As a wave function Ψ for the electrons, quantum chemists adopt the Slater determinant shown in (3).

$$\Psi = \sum_k c_k \Psi_k(\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_N}), \quad \Psi_k = (2\mathrm{N})!^{-1/2} \begin{vmatrix} \varphi_{k1}(\mathbf{x_1}) & \varphi_{k2}(\mathbf{x_1}) & \cdots & \varphi_{kN}(\mathbf{x_1}) \\ \varphi_{k1}(\mathbf{x}_2) & \varphi_{k2}(\mathbf{x}_2) & \cdots & \varphi_{kN}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{k1}(\mathbf{x}_N) & \varphi_{k2}(\mathbf{x}_N) & \cdots & \varphi_{kN}(\mathbf{x}_N) \end{vmatrix} . \tag{3}$$

The variable $\mathbf{x_i}$ denotes both the coordinate **r** and its spin state (α or β). $(2N)!^{-1/2}$ is a normalization factor. This determinant satisfies the requirement called antisymmetry principle, where a sign of the wave function is changed by interchanging positions of a pair of electrons. In the quantum chemical calculation, each molecular orbital can be described as the linear combination of atomic orbitals (LCAO) $\chi_n(\mathbf{r})$ such as

$$\varphi(\mathbf{r}) = \sum_k C_k \chi_k(\mathbf{r}) . \tag{4}$$

Our goal is to minimize the $E_{\text{elec}}=\langle\Psi|H_{\text{elec}}|\Psi\rangle$ subject to the constraint that the trial wave function must be orthonormalized by using a variational method. This scheme is called as "Hartree-Fock(HF) theory". HF theory is categolized into the mean field theory. The working equation is given by

$$\left(-\frac{\nabla^2}{2} - \sum_A \frac{Z_A}{|\mathbf{r}-\mathbf{R}_A|} + 2\sum_{j\in\text{occ}} \int d\mathbf{r} \frac{\varphi_j^*(\mathbf{r}')\varphi_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \right) \varphi_i(\mathbf{r}) - \sum_{j\in\text{occ}} \int d\mathbf{r} \frac{\varphi_i^*(\mathbf{r}')\varphi_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \varphi_j(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}) , \tag{5}$$

where the fourth term arise from the antisymmetry principle and is referred as exhange contribution. Within the LCAO approach, above nonlinear coupled equations result in a following secular equation,

$$\mathbf{FC} = \varepsilon \mathbf{SC} , \tag{6}$$

where **F** and **S** are so-called Fock and Overlap matrices. The matrix elements are given by

$$\begin{aligned} F_{kl} &= H_{kl} + \sum_{mn} \gamma_{mn} (2V_{klmn} - V_{knml}) \\ S_{kl} &= \int d\mathbf{r} \chi_k^*(\mathbf{r}) \chi_l(\mathbf{r}) \\ H_{kl} &= \int d\mathbf{r} \chi_k^*(\mathbf{r}) \left(-\frac{\nabla^2}{2} - \sum_A \frac{Z_A}{|\mathbf{r}-\mathbf{R}_A|} \right) \chi_l(\mathbf{r}) , \\ V_{klmn} &= \iint \frac{2\chi_k^*(\mathbf{r}) \chi_m^*(\mathbf{r}') \chi_l(\mathbf{r}) \chi_n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ \gamma_{mn} &= \sum_{j\in\text{occ}} C_{mj}^* C_{nj} \end{aligned} \tag{7}$$

where H_{kl} and V_{klmn} are one- and two-electron matrix elements, respectively, and γ_{mn} is the density matrix. Since the two-electron term depends on the input density matrix, the HF equation is solved self-consistently.

HF theory covers 99.5% of the total energy of the whole molecule. However, the remaining 0.5% (equals to the order of 1-10 kcal/mol for small molecules) is quite important in the field of chemistry. To reach this accuracy, we have to consider the "electron correlation" which is the interaction between electrons. There are so many approaches (post-HF theories) to obtain the electron correlation energy. The most frequently used theory for application is the second order of Møller-Plesset (MP2) theory which is the perturbation theory for the dynamical electron correlation. This method is efficient when the HF theory gives the good results. Moreover, the computational cost is the lowest among the post-HF theories. In Section 3, we adopted this theory to obtain the PES of the proton between guanine and cytosine.

2.2 Density functional theory

The DFT is derived from the Hohenberg-Kohn theorem, which consists of two theorems: (1) The ground state electron density $\rho_0(\mathbf{r})$ of a many electron system in the presence of an external potential uniquely determines the external potential. (2) The functional $E[\rho]$ for the ground state energy is minimized by the ground state electron density ρ_0, i.e. $E[\rho_0] \leq E[\rho]$ for every trial electron density ρ. When we determine $\varphi_i(\mathbf{r})$ as a molecular orbital, as we derived the HF equation in previous paragraph, the total electron density can be determined as

$$\rho(\mathbf{r}) = \sum_{i}^{N_{occ}} |\varphi_i(\mathbf{r})|^2 . \tag{8}$$

This summation for variable i runs to the number of occupied molecular orbitals N_{occ}. Kohn-Sham equation can be written as follows:

$$\left(-\frac{\nabla^2}{2} - \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} + V_{\mathrm{H}}(\mathbf{r}) + v_{\mathrm{xc}}(\mathbf{r}) \right) \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}) , \tag{9}$$

where V_H is the Hartree potential which expresses the Coulomb interaction between two electrons. The term $v_{XC}[\rho]$ is "exchange-correlation" potential which replaces the effect from the exchange of electrons used in the HF & post-HF theories. Since the equation (9) is similar to the equation (5), DFT can be used by the same procedure, i.e. LCAO approxach, as the HF theory. Although DFT includes the electron correlation effects through the empirical correlation potential, the computational cost is much lower than post HF methods. However, no one knows the exact formulation for the correlation potential so far. To obtain the exchange-correlation energy, many types of model exchange-correlation functional have been proposed. Among them, Becke proposed the exchange-correlation functional which includes the HF exchange energy. This method is often called "hybrid functional". In Section 4, we used "mPW1PW91", one of the hybrid functionals. The detailed expressions are given in ref 17.

2.3 ONIOM method

The computational cost of *ab inito* or DFT is over $O(N^3)$ where N denotes the number of atoms. Therefore, it is necessary to reduce the computational cost for larger sized molecules which is discussed in this chapter.

ONIOM, which is the abbreviation of "our own *n*-layered integrated molecular orbital and molecular mechanics" proposed by Morokuma et al. [18], is one of the candidates for the approximate ways of this problem. ONIOM is a kind of the hybrid methods for quantum mechanics/molecular mechanics (QM/MM) calculation, which is *de facto standard* of biophysical simulations. This method divides the whole molecule into two (or more) parts in order to enable us to compute different level of theories, i.e.,

$$E_{\mathrm{ONIOM}} = E_{\mathrm{low}}^{\mathrm{large}} - E_{\mathrm{low}}^{\mathrm{small}} + E_{\mathrm{high}}^{\mathrm{small}} . \tag{10}$$

When we choose the QM (*ab initio* or DFT) as the higher level calculation and molecular mechanics (MM) as the lower level, we can perform the QM/MM like calculations with this method. In the ONIOM calculation, we replace the boundary atom with a hydrogen atoms. Figure 4 shows the simple scheme for ONIOM calculations. The potential function of MM E_{MM} are generally written as follows:

$$\begin{aligned} E_{\mathrm{MM}} = & \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\varphi - \gamma)] \\ & + \sum_{i<j} [\frac{A_{ij}}{r_{ij}^{12}} + \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}}] \end{aligned} . \tag{11}$$

These terms are the stretch (bonds), bend (bond angles), torsional (dihedral angles) and non-bonded interactions (van der Waals interaction and Coulomb interaction). The constants r_{eq}, θ_{eq}, γ are obtained by geometry optimization with quantum chemical calculation. We also have to determine the parameter K_{r}, K_θ, V_{n}, A_{ij}, B_{ij}. The variable q_{i} is called as "MM charge" which depends on the computational models. We used this scheme in Section 4.

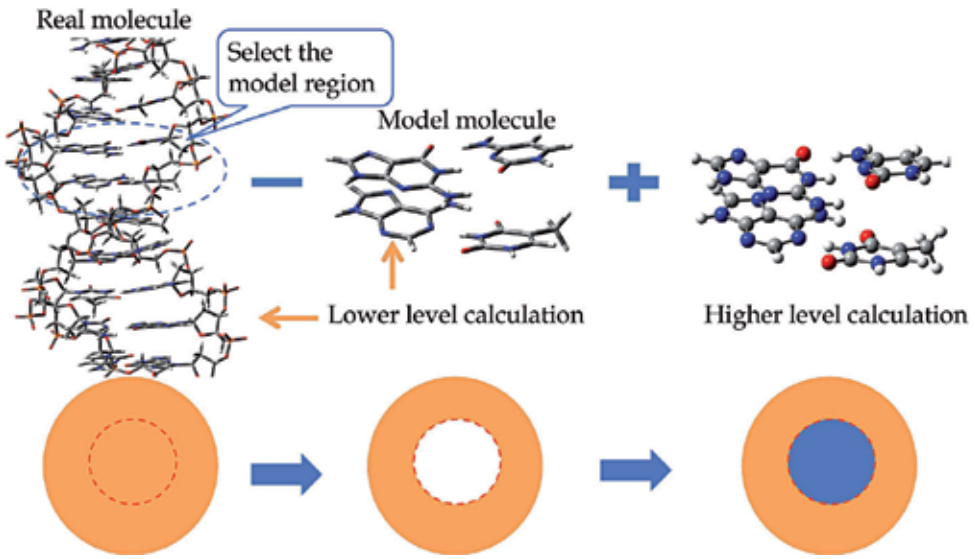


Fig. 4. An image of ONIOM scheme. In general, the region in orange is computed by the lower level theory such as molecular mechanics (MM). The region in blue is for higher level theory such as *ab initio* or DFT.

3. Local change in a cisplatin bound GC pair

3.1 Computational details

Firstly, we focused on the proton-transfer reaction between a GC pair. We simplified the model to the system which consists of one GC pair and cisplatin, i.e. $[Pt(NH_3)_3GC]^{2+}$. Geometry optimizations were first performed by the MP2 method. As the platinum is a heavy element, we have to consider the relativistic effect. In order to reduce the computational cost, we adopted an effective core potential (ECP) which implicitly includes the relativistic effect. As a typical ECP, we used the LanL2DZ basis function (We here denote this set of calculation as "MP2/LanL2DZ"). Later hydrogen bonding energy was calculated at those optimized structures with correction of a basis set superposition by the counterpoise method. All calculations were performed by GAUSSIAN03 [19]. The hydrogen bond lengths, energies, induced charges and all possible proton transfer reactions among the bases were calculated.

3.2 Local changes of properties in platinum complex bound GC pair

In this subsection, we compared neutral GC pair with the platinum complex bound GC pair. Especially, we focused on the hydrogen bonding energy, hydrogen bonding length and IR spectra. First, we summarized the results of the optimization in Table 1.

	Hydrogen bonding energy	Hydrogen bonding length [in Å]		
	[kcal/mol]	O_6(G)-N_4(C)	N_1(G)-N_3(C)	N_2(G)-O_2(C)
GC pair	27.2	2.98	2.93	2.85
Pt + GC	31.4	3.06	2.91	2.90

Table 1. Summary of the change caused by the coordination of platinum complex.

Hydrogen bond length and energy: The major difference between normal GC pair and platinum complex bound GC pair is the length between O_6(G) and N_4(C). This is caused by Pt^{2+} cation which attracts the O_6(G). As a result, the hydrogen bond length O_6(G)-N_4(C) becomes larger. On the other hand, the hydrogen bond length N_1(G)-N_3(C) is not affected by the platinum complex binding. The hydrogen bond energy of the base pairs is given by the energy difference between $[Pt(NH_3)_3G]^{2+}$ + C and $[Pt(NH_3)_3GC]^{2+}$. The results show that the hydrogen bond energy increases by 4 kcal/mol by Pt complex formation.

Changes in IR spectra: Next, the vibrational analyses were performed by MP2/LanL2DZ at the optimized structure obtained by the same method. The GC pair has three peaks of N-H stretching modes related to the hydrogen bonds. We defined three N-H stretching modes O_6…H-N_4, N_1-H…N_3 and N_2-H…O_2 as stretch1, stretch 2 and stretch 3, respectively.

Figure 5 shows the results of peak shifts before and after the Pt binding. In the case of N-H stretch between O_6 (G)-N_4 (C), about 250 cm^{-1} blue shift was observed and the intensity became lower. This means the hydrogen bonding becomes distant and weaker. In the case of N-H stretch between N_2 (G)-O_2 (C), on the other hand, about 300 cm^{-1} red shift was observed and the intensity became higher showing that the hydrogen bonding becomes stronger. These shifts can be explained by the hydrogen bonding length. In the case of N-H stretch between N_1 (G)-N_3 (C), about 400 cm^{-1} red shift was observed. From these results, it was found that the additional hydrogen bond between the H atom of the ligand and O_6 of G affects the stability of the complex.

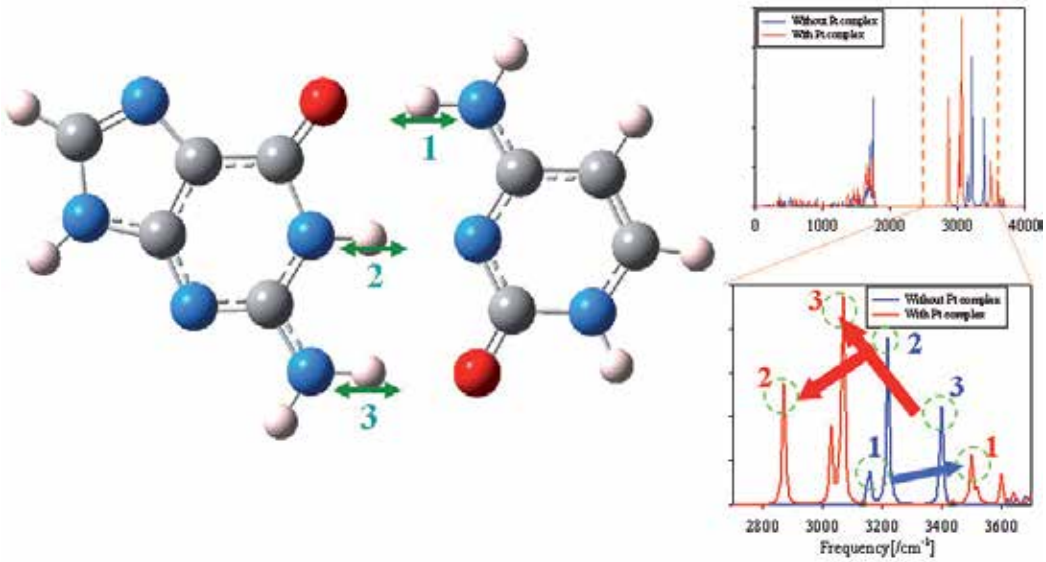


Fig. 5. IR spectra of GC pair and platinum complex bound GC pair. The peak numbers correspond to the N-H stretching modes.

3.3 Potential energy surface of proton

The other hydrogen bonds become closer because of the electrostatic interaction induced by the Pt binding. Especially, N_1 (G) and N_3 (C) become closer and the potential energy curve of the hydrogen atom between them has a double minimum. Next we proceed to investigate a feature of local changes at the hydrogen bonds by observing the PES as a function of positions of H atoms in order to understand the peak shifts. The GC pair has three hydrogen bonds: O_6…H-N_4, N_1-H…N_3 and N_2-H…O_2, respectively. We focused on the two protons: O_6…H-N_4 and N_1-H…N_3, which were concerned with the SPT or DPT reactions as shown in the introduction. We obtained an energy of optimized geometry $V(x,y)$ as a function of bond length O_6-H (x) and N_1-H (y). We fitted the PES as the following equation.

$$V(x,y)= \sum_{\substack{i,j \\ i+j\le 6}} C_{ij} x^i y^j \ , \tag{12}$$

where C_{ij} is the expansion coefficient. We used the PES of normal GC pair computed by Villani [6]. We draw the PESs as shown in Figure 6 and listed the expansion coefficients in Table 2. There are explicit differences in PESs after the Pt binding. In the normal GC pair, two local minima were found, which correspond to the DPT reaction. On the other hand, O_6-H are fixed near the global minimum, N_1-H has two minima in the platinum complex bound GC pair. The former is near N_1-H=1.10 and the latter is near N_1-H=1.70, respectively, which corresponds to SPT reaction. It was confirmed that its energy difference and energy barrier are very small.

3.4 Proton-transfer reaction in Pt+GC pair

We show the existence of the structure that undergoes SPT reaction between N_1 (G) and N_3 (C), which is hereafter denoted as a G-C^+ pair. The barrier of the DPT reaction, which occurs

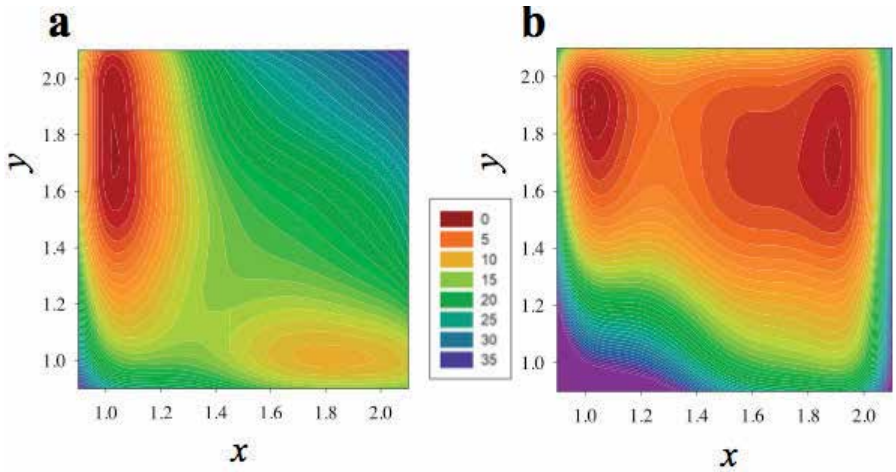


Fig. 6. The potential energy surface of two protons in (a) normal GC pair (b) platinum complex bound GC pair. The unit of the number is kcal/mol.

	j=0	*j*=1	*j*=2	*j*=3	*j*=4	*j*=5	*j*=6
i=0	6935.30	-20913.1	38457.0	-35870.0	18315.3	-4867.65	527.568
i=1	-6004.82	-9088.74	8023.79	-4442.78	1219.89	-131.039	-
i=2	14320.4	4568.72	-1246.22	394.393	-52.4822	-	-
i=3	-14376.9	-2318.65	161.588	-17.1170	-	-	-
i=4	7795.29	692.153	-12.5530	-	-	-	-
i=5	-2195.60	-85.3225	-	-	-	-	-
i=6	251.950	-	-	-	-	-	-

Table 2. Expansion coefficients of C_{ij} in platinum complex bound GC pair.

in normal GC pair, is 15.2 kcal/mol, which gives the similar result to Florian's work [5]. On the other hand, there is no DPT structure of the cisplatin bound GC pair. Figure 7 sketches the optimized structures of the GC and G·C$^+$ pairs, and the transition state (TS). It is confirmed that all of the structures keep their planarity, even after the Pt complex binding and proton transfer. Nevertheless, the distances between the atoms involved in hydrogen bond vary significantly as shown in Table 3. The hydrogen bond of O_6–N_4 increases by 0.15 Å due to the influence of the "additional" hydrogen bond between O_6 and the ligand of cisplatin. The other hydrogen bonds shrink because of the electrostatic interaction induced by the Pt binding, as discussed below. In particular, the hydrogen bond between N_2 (G) and O_2 (C) decreases by 0.19 Å. There is a weak dependence of the ligands in these distances of hydorgen bonds for the GC pair. On the other hand, a stronger dependence is found for the distance between bases of G·C$^+$ pair. O_6–N_4, N_1–N_3, and N_2–O_2 distances in G·C$^+$ pair are shortened by 0.05, 0.08, and 0.02 Å compared with the normal GC pair, respectively. These facts indicate that all of the hydrogen bonds become stronger in comparison with the normal GC pair. In what follows, we examine where these differences arise.

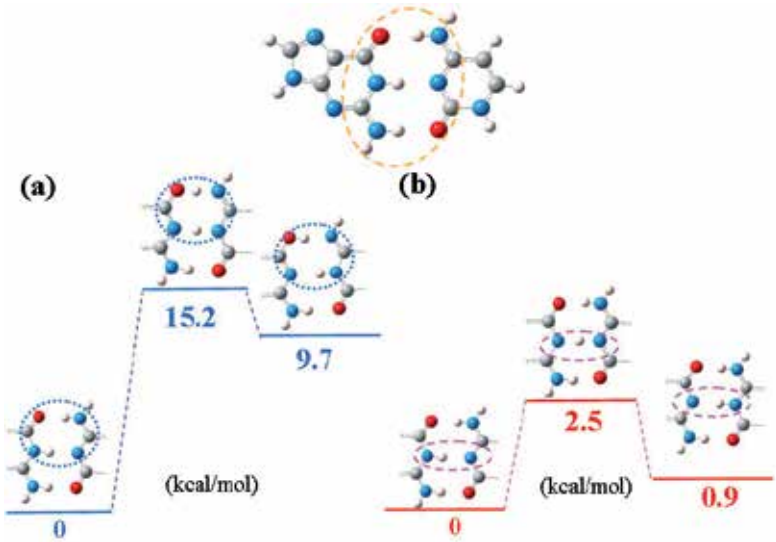


Fig. 7. The reaction diagram for proton-transfer reaction in GC pair. (a) normal GC pair (b) platinum complex bound GC pair. In (a), double proton-transfer reaction occurred, and in (b), single proton-transfer reaction occurred.

	Pt + GC	TS	Pt + G-C$^+$	Experimental
O_6-N_4	3.06	2.80	2.77	2.79
N_1-N_3	2.91	2.64	2.88	2.88
N_2-O_2	2.72	2.77	2.94	2.84

Table 3. Hydrogen bonding length in Pt+GC pairs (in Å)

Charge distributions: The natural bond orbital (NBO) charges of atoms associated with hydrogen bonds and Pt-G coordination, listed in Table 4, shows that in the GC pair the charge on O_6 of G increases by interaction with the ligand of Pt. The hydrogen bond strength between a ligand and O_6 of G influence the charge of O_6, which leads to the difference in the whole hydrogen bonds. The NBO charge on Pt decreases. Therefore, a charge transfer takes place mainly from the guanine and partially from the cytosine to the Pt atom. Such a charge transfer leads to a decrease in the charge on N_1 in guanine and an increase in the charges on all H atoms between the bases. The foregoing discussion gives clues for the reasons why the hydrogen bonds are strengthened: (i) the charge transfer mainly from O_6 of G to the Pt and H atoms, and (ii) additional formation of a hydrogen bond between L_1 and O_6 of G, as pointed out above. Strong dependences of the ligand on the charge of Pt, N_7, and O_6 are found. In the platinum bound GC pair, only a small influence is observed in the Mulliken charges on the atoms in C, whereas the changes are more significant in the G-C$^+$ pair. The charges on all the heavy elements except for O_6 increase compaired to the GC pair. Moreover, strong influences of the ligand on the charge are found for Pt, N_7, O_6, as well as N_1. For further understanding of the interactions among the Pt atom, ligands, and the bases, the sum of the NBO charge on the bases and the ligands are listed in Table 5. As a reference, the total charges in the unbound case is estimated

	Pt-G coordination		Hydrogen bond 1			Hydrogen bond 2			Hydrogen bond 3		
Atom	Pt	N_7	O_6	H_1	N_4	N_1	H_2	N_3	N_2	H_3	O_2
GC	0.72[a]	-0.44	-0.57	0.48	-0.86	-0.81	0.47	-0.64	-0.90	0.46	-0.56
Pt+GC	0.47	-0.63	-0.60	0.46	-0.89	-0.80	0.50	-0.67	-0.87	0.50	-0.57
Pt+G-C$^+$	0.74	-0.53	-0.61	0.47	-0.70	-0.38	0.51	-0.54	-0.69	0.45	-0.32

a: Estimated from $[Pt(NH_3)_4]^{2+}$ as a reference.

Table 4. NBO charges calculated by MP2/LanL2DZ

	G	C	Ligands	Pt
GC pair	-0.026	0.026	1.389	0.611
$Pt(NH_3)_3GC$	0.273	0.125	0.910	0.692
$Pt(NH_3)_3G\text{-}C^+$	-0.444	0.886	0.881	0.686

Table 5. Sum of NBO charge

using $[Pt(NH_3)_4]^{2+}$. In the normal GC pair, both G and C are almost neutral. In contrast, both G and C in the platinum bound GC pair are positive due to the charge transfer from the bases mainly to the Pt atom. As for the G-C$^+$ pair, however, G and C become negative and positive, respectively. Note here that C includes the hydrogen atom that initially belongs to the G in the GC pair as a result of SPT from G to C. Since the charge on the Pt atom increases, charge transfer from G to the ligands of cisplatin accompanies the SPT reaction.

3.5 Summary of section 3

We have numerically elucidated the changes in the hydrogen bonds of the base pairs due to Pt complex formation by means of MP2 theory. For the platinum bound GC pair, the hydrogen bonding energy exceeds by 3–10 kcal/mol than GC pair without platinum complex. The hydrogen bond between O_6–N_4 is lengthened, whereas the other hydrogen bonds are shortened. These observations are explained by a rearrangement of the charge distribution. In this section, we revealed that the binding of platinum complex causes a single proton-transfer (SPT) reaction between N_1 (G) and N_3 (C) in the GC pair. Note here that the SPT reaction does not occur in the GC pair itself, while double proton-transfer (DPT) does. Its reaction barrier decreases from 15-20 kcal/mol of the DPT reaction without the Pt complex to 1.5–3 kcal/mol of the SPT reaction with the Pt complex. The structure that underwent the SPT reaction is as stable as the original structure.

4. Global and local change of GC pairs by coordination of platinum complex

In stacked two GC pairs, althogh DPT reactions can occur independently. However, these reactions hardly take place because of high energy barriers as seen in one GC pair [20]. For further research, it is natural to study what happens when two or more base pairs are stacked. For the simplicity, we hereafter denote a "G-C$^+$ pair" (the product of SPT reaction in GC pair) as a "G*C pair".

4.1 Computational details

Modeling: We first define two types of model systems: the "two base pair" (2bps) model and the "four base pair" (4bps) model. The former consists of two base pairs and cisplatin without the backbone molecules, such as sugar and phosphate groups, while the latter consists of four base pairs and the backbone molecules so that it includes the effects of both the backbone molecules and base stacking.

The 2bps model is small enough to compute with the full quantum chemistry calculation. Density functional theory (DFT) was adopted and the modified Parr–Wang functional (mPW1PW91) was chosen as an exchange–correlation functional, because the functional is modified to better describe the hydrogen bonding. For the calculations described in this chapter, the Stuttgart/Dresden ECPs were used for the Pt atom and the 6-31G (d, p) basis for the other atoms.

ONIOM method: To investigate the effects of DNA stacking, the backbone and counter cations, we proceeded to the 4bps model by ONIOM method introduced in Section 2.3. We here treated two of the four base pairs as the higher layer and the rest as the lower layer. We utilized the method used in the 2bps model for the higher level calculation and the universal force fields (UFF) for the lower level calculation in ONIOM cacluation. We took the initial structure 1,2-d (CpX_1pX_2pT) from PDB (PDBID:1A84), where X_1 and X_2 are purine moieties bound to the Pt complex so that there are three patterns of X_1 and X_2. The ligands of the cisplatin were assumed as NH_3, as in the human body. We also assumed that the Pt atom binds to N_7 of G and A. To keep the whole system neutral, we added sodium atom at every PO_4^- molecule as the counter cation.

4.2 2bps model

Figure 8 shows the all optimized geometries of the three types of model molecules: cis-(CG)-Pt-(GC), cis-(CG)-Pt-(AT) and cis-(TA)-Pt-(AT). The structure of cis-(CG)-Pt-(GC) was distorted in comparison with those in the DNA because of repulsion between the two O_6 atoms. Although one of the GC pairs keeps a planar structure, the other GC pair was greatly distorted. We distinguish them by referring to the former planar pair as the "$(GC)_p$ pair" and the latter distored pair as the "$(GC)_d$ pair". This distortion may lead to the stabiity of whole system in cis-(CG)-Pt-(GC). Hereafter, we will describe the structure of the 2bps model in the form cis-$(GC)_p$-Pt-$(GC)_d$. Unlike the structures of cis-$(GC)_p$-Pt-$(GC)_d$, all base pairs in the cis-(CG)-Pt-(AT) and cis-(TA)-Pt-(AT) almost keep their planarity.

We evaluated the binding energy of cisplatin and the base pair. To estimate the hydrogen bonding energy, we assumed a two-step reaction: (1) $[Pt\ (NH_3)_4]^{2+} + B_p\text{-}B'_p$ -> $[Pt\ (NH_3)_3\ B_p\ B'_p]^{2+} + NH_3$ (reaction energy: ΔE_1) (2) $[Pt\ (NH_3)_3\ B_p\ B'_p]^{2+} + B_d\ B'_d$ -> $[cis\text{-}Pt\ (NH_3)_2\ B_p\ B'_p\ B_d\ B'_d]^{2+} + NH_3$ (reaction energy: ΔE_2), where B_x is A or G, B'_x represents the complementary base of B_x (x = p, d). Results are shown in Table 6. Comparing the ΔE_1 values, the binding energy of Pt-(GC) is much higher than that of Pt-(AT), by 30 kcal/mol. ΔE_1 is larger than ΔE_2 because of the Coulomb repulsion between the original $[Pt\ (NH_3)_3\ B_p\ B'_p]^{2+}$ and the additional base pair. Moreover the bases are likely to bind to the Pt complex in the order cis-(CG)-Pt-(GC), cis-(CG)-Pt-(AT) and cis-(TA)-Pt-(AT). In particular, the binding energy of the cis-(TA)-Pt-(AT) is remarkably low compared with those of cis-(CG)-Pt-(GC) and cis-(CG)-Pt-(AT). These results support both experimental evidence and the tendencies of the models studied by Burda and Leszyczynski [13], as mentioned in the introduction.

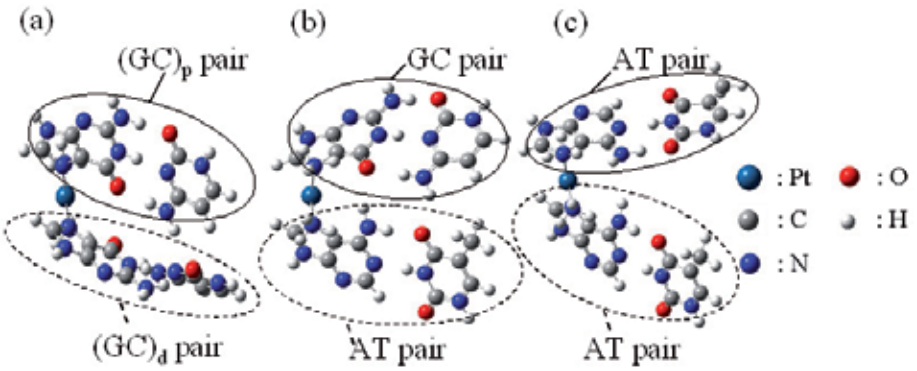


Fig. 8. The optimized structures of the 2bps models: (a) cis-$(CG)_p$-Pt-$(GC)_d$, (b) cis-(CG)-Pt-(AT), and (c) cis-(TA)-Pt-(AT).

	B_1=G, B_2=G	B_1=G, B_2=A	B_1=A, B_2=G	B_1=A, B_2=A
ΔE_1	46.1 (47.9)	46.1 (47.9)	14.7 (16.4)	14.7 (16.4)
ΔE_2	24.5 (25.9)	6.9 (8.2)	38.2 (41.7)	2.4 (4.9)
Total	70.5 (73.8)	52.9 (56.1)	52.9 (56.1)	17.1 (21.3)

a: The "2bps model" was used.
b: The optimized NH_3 and $[Pt(NH_3)_4]^{2+}$ is used for energy calculation.
c: The numbers in parentheses represent the energy including zero-point energy.

Table 6. Energy differences of substitution reaction (in kcal/mol)

Why cisplatin prefers guanine to adenine: Figure 9 shows the reason why cisplatin prefers guanine to adenine. It is known that cisplatin attacks to major groove of DNA base. Guanine has N_7 and O_6 which attract positive charge. On the other hand, adenine has amino group in N_6. One of H atoms in amino group prevent positive charge from attacking to major groove. If cisplatin binds to adenine, this amino group has to be rotated which needs much energy. Thus, cisplatin prefers guanine to adenine. Table 6 shows this tendency exactly. Therefore, we will focus on the models of cis-$(CG)_p$-Pt-$(GC)_d$ and cis-(CG)-Pt-(AT) in further discussion.

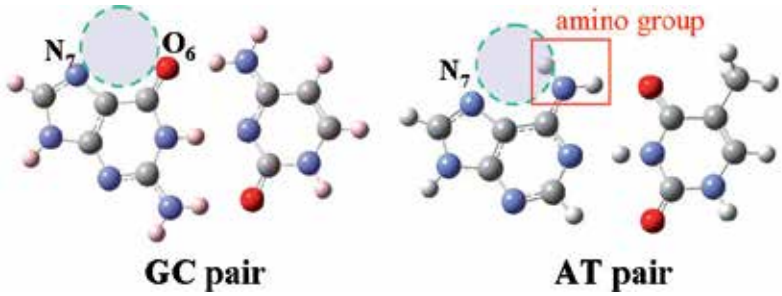


Fig. 9. The major groove of GC and AT pair where cisplatin attacks. In the case of AT pair, amino group of N_6(A) prevents cisplatin from attacking to N_7.

Proton-transfer reactions in 2bps model: Next, we discuss the possibility of multiple proton-transfer reactions in these systems. Here we depict a restricted two-dimensional potential energy surface (PES) of two different hydrogen atoms between N_1 ($G_{p/d}$) and N_3 ($C_{p/d}$) in figure 10, where the geometry except for the two hydrogen atoms is fixed. The rigid PES used here is very rough approximation, because it does not consider the effect of structure relaxation. Nevertheless, it is at least useful to intuitively understand the proton transfer reactions between bases. In this subsection, we depicted this PES in order to confirm whether proton-transferred structures exist or not two stacked base pairs. The origin of the PESs are at the center of both hydrogen bonds, and the PESs were plotted every 0.05 Å. From this figure, it shown that there are three possible minima, at the points marked as X in the figures. No two simultaneous SPT reactions are found because the potential energy of cis-$(CG^*)_p$-Pt-$(G^*C)_d$ is very high as shown in ▲. We here confirmed that one SPT reaction can occur even in two GC pairs. Figure 11 shows the results of geometry optimization of cis-$(CG)_p$-Pt-$(GC)_d$, cis-$(CG^*)_p$-Pt-$(GC)_d$ and cis-$(CG)_p$-Pt-$(G^*C)_d$, where G* again means the guanine donating a proton. This result is similar to the SPT reaction in one GC pair calculation described in previous subsections, where the reaction barrier is about 5–6 kcal/mol. The difference of the energy barrier between cis-$(CG^*)_p$-Pt-$(GC)_d$ and cis-$(CG)_p$-Pt-$(G^*C)_d$ is because of the difference in their planarity. A structure of two simultaneous SPT, cis- $(CG^*)_p$-Pt-$(G^*C)_d$, cannot be found, as expected from the potential surface depicted in figure 10.

Next, the sum of charges obtained by the natural bonding orbital (NBO) is analyzed. Table 7 lists the results of the sum of the NBO charges. Every part of cis-$(CG)_p$-Pt-$(GC)_d$ has a positive charge, but both C_p and C_d are almost neutral. When the SPT reaction occurs, the

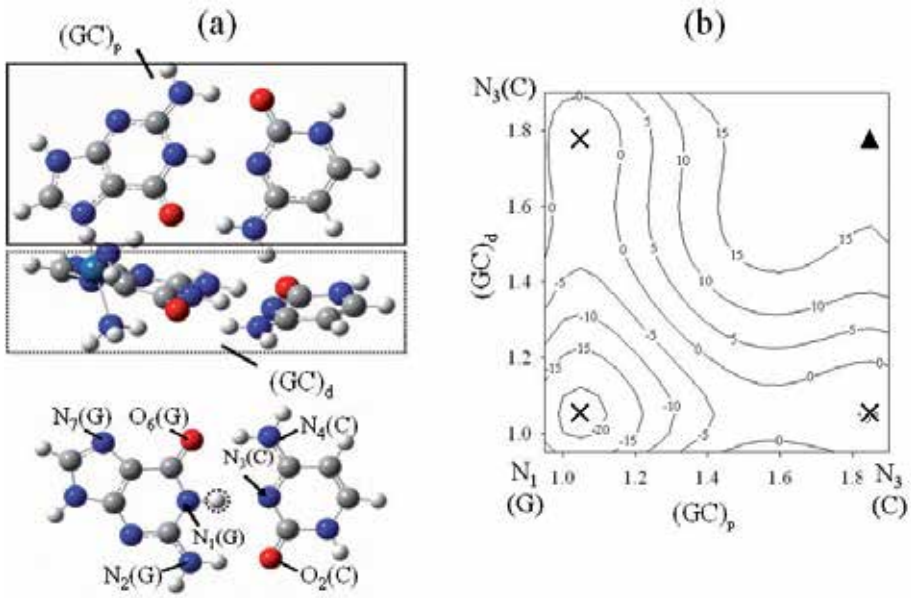


Fig. 10. (a) Definitions of variables in a 2bps model of cis-$(CG)_p$-Pt-$(GC)_d$. Only the hydrogen atoms surrounded by the dotted line in each GC pair were moved. (b) The potential energy surface of hydrogen atoms. The numbers on each axis represent the distance between N_1 (G) and the hydrogen atom (in Å). The origin is set at the point where both hydrogen atoms are at the middle of the hydrogen bonding.

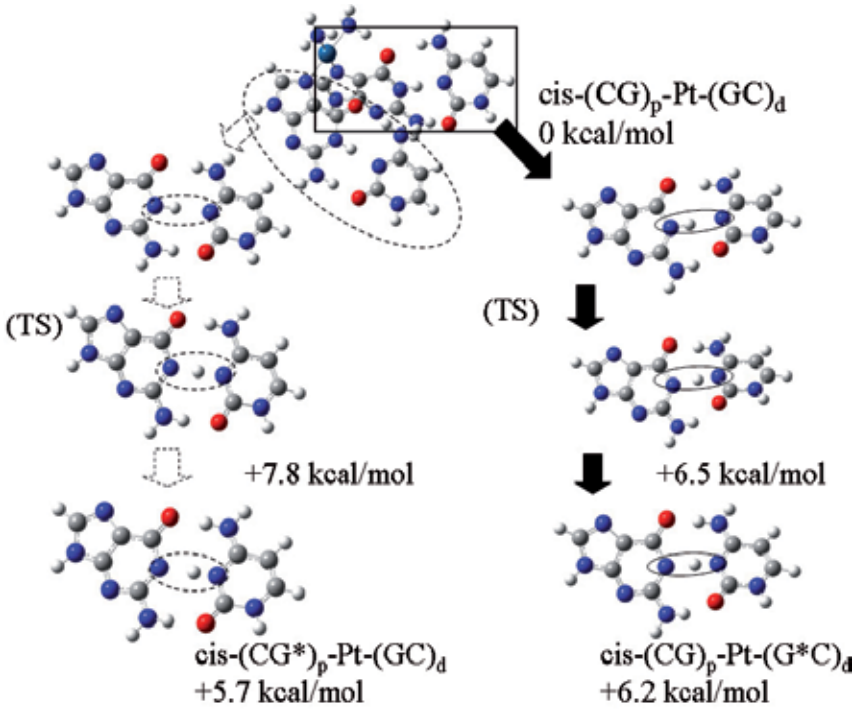


Fig. 11. The reaction diagram for single proton transfer (SPT) between N_1 (G) and N_3 (C) of the 2bps model of cis-(CG)-Pt-(GC). The numbers in kcal/mol in the figure denote relative energies measured from cis-$(CG)_p$-Pt-$(GC)_d$.

	cis-$(CG)_p$-Pt-$(GC)_d$	cis-$(CG^*)_p$-Pt-$(GC)_d$	cis-$(CG)_p$-Pt-$(G^*C)_d$
G_p or G^*_p	0.26	–0.43 [a]	0.25
G_d or G^*_d	0.24	0.25	–0.44 [a]
C_p	0.10	0.58 [b]	0.33
C_d	0.10	0.32	0.58 [b]
Pt	0.73	0.69	0.70
Ligands	0.59	0.58	0.58

a: The sum does not contain the H atom transferred to the N_3 of C.
b: The sum contains the H atom transferred from G.

Table 7. Sum of the charges by natural bond orbital (NBO) analysis

whole charge of a proton donor G* becomes negative. On the other hand, both C_p and C_d become positive after the SPT reaction. The proton donor G* is negative whereas a proton acceptor C is positive. It is expected that these Coulomb repulsions of C_p and C_d and of G_p and G_d prevent further SPT reactions from cis-$(CG)_p$-Pt-$(G^*C)_d$ and $(CG^*)_p$-Pt-$(GC)_d$. In particular, this change of charge distribution can be seen in the case of trans-$(CG)_1$-Pt-$(GC)_2$, in which the two guanines become negative while the two cytosines become positive.
The SPT reaction between N_1 and N_3 of the GC pair also occurs in cis-(CG)-Pt-(AT). The result is similar to the one GC pair shown in Section 3.4. The structure is not distorted in

spite of the SPT reaction between the GC pair. Then, we must know the possibilities of further proton transfers in the AT pair. These pairs have two hydrogen bonds N_6 (A)-O_4 (T) and N_1 (A)-N_3 (T). We show a restricted two-dimensional potential energy surface in Figure 12, where variables r_1 and r_2 denote the distances N_6-H and N_1-H, and the other geometries are fixed. Figure 12 shows two local minima, both cis-(CG)-Pt-(AT) and cis-(CG*)-Pt-(AT). This implies that a multiple SPT reactions can occur in cis-(CG)-Pt-(AT). Nevertheless, the energy difference between local minima is so large that the proton-transfer reaction may not occur at room temperature. The tendency does not change even after the SPT reaction took place in the GC pair. Table 8 summarizes the possibilities for multiple proton-transfer reactions in all the systems.

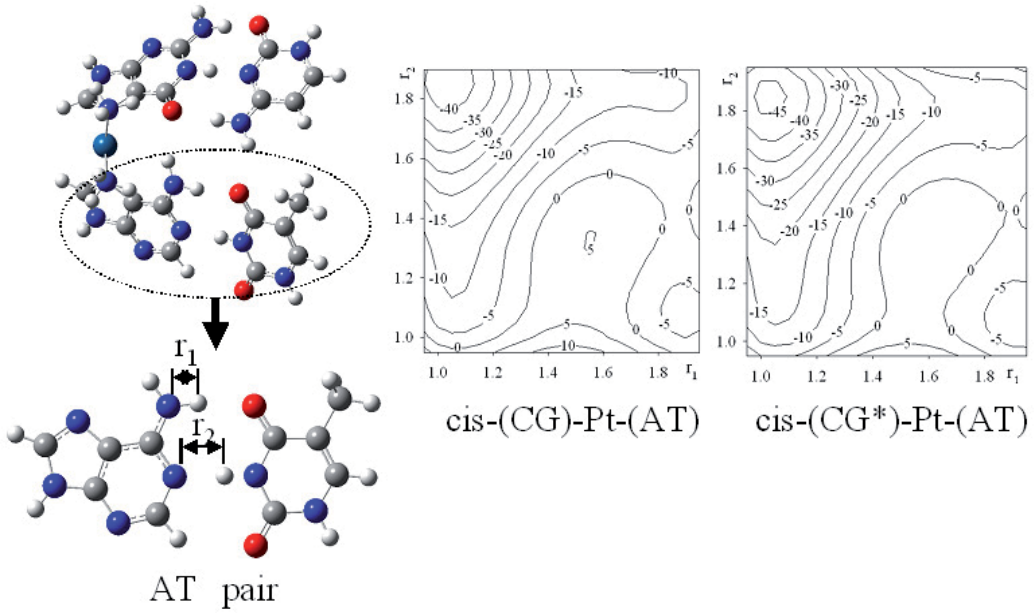


Fig. 12. Definitions of variables in the 2bps model of the cis-(CG)-Pt-(AT). The numbers in the contour plots represent the energy in kcal/mol.

	cis-$(CG)_p$-Pt-$(GC)_d$	
$(CG)_p$ / $(GC)_d$	None	SPT between N_1 and N_3
None	0	6.2
SPT between N_1 and N_3	5.7	N/A (not available)
	cis-(CG)-Pt-(AT)	
GC / AT	None	DPT (N_1 and N_3), (N_6 and O_4)
None	0	17.5
SPT between N_1 and N_3	–1.0	15.2

Table 8. The differences of energy in "2bps. model" molecules (in kcal/mol)

Note that the SPT reaction occurs in platinum-bound GC pair, and the DPT reaction occurs in AT pair. The DPT reaction can also be found even after the SPT reaction between N_1 (G) and N_3 (C) has occurred. The SPT reaction in the AT pair cannot occur, because the Pt complex binding increases the distance between N_1 (A) and N_3 (T). In every case of the DPT reactions, the product of DPT reaction becomes more unstable than the original structure.

4.3 4bps models

Our calculations elucidated that SPT structures of the cis-(CG)-Pt-(GC) type can exist even in 4bps models. Two different optimized structures are shown in Figure 13, which shows one planar GC and one distorted GC pair as well as the 2bps model of cis-(CG)-Pt-(GC). We also found that the backbone and stacking bases do not change their postion much from their original geometry during the SPT reaction.

We extracted the higher layer, i.e., the 2bps and Pt complex, from the 4bps model to investigate the energy differences between cis-$(CG)_p$-Pt-$(GC)_d$ and $(CG^*)_p$-Pt-$(GC)_d$ or cis-$(CG)_p$-Pt-$(G^*C)_d$. The energies of these model molecules are assumed to be approximately those of their higher layer. The energetics of $(CG)_p$-Pt-$(GC)_d$ and $(CG^*)_p$-Pt-$(GC)_d$ are

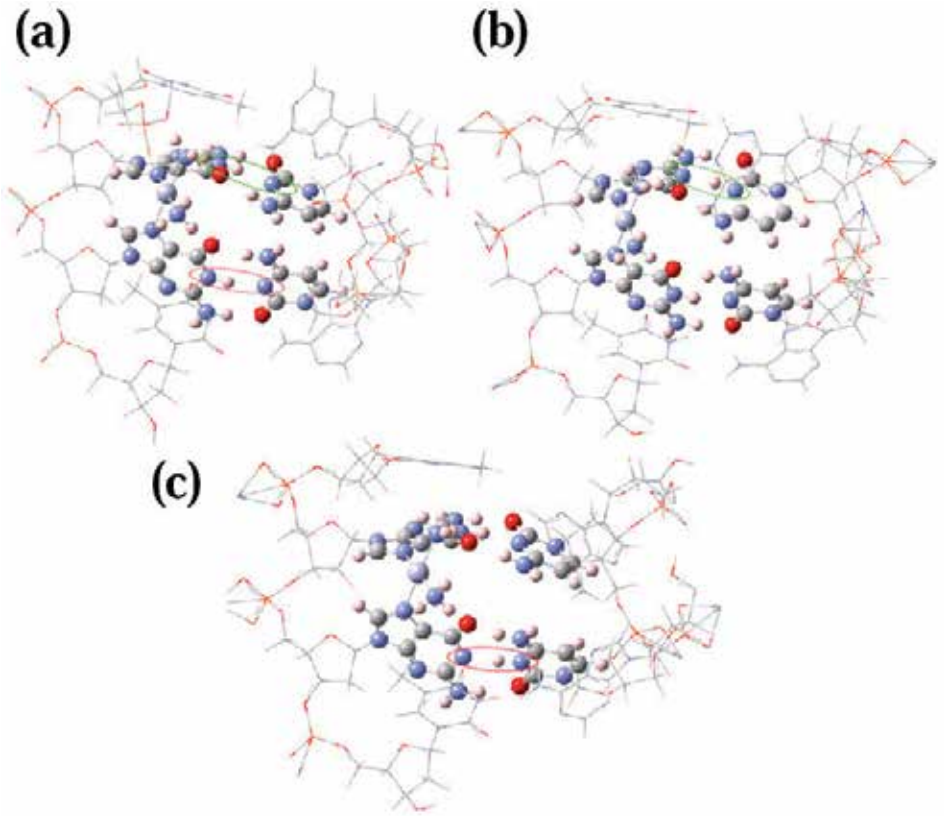


Fig. 13. The optimized structures of the 4bps models of (a) cis-$(CG)_p$-Pt-$(GC)_d$, (b) cis-$(CG^*)_p$-Pt-$(GC)_d$, and (c) cis-$(CG)_p$-Pt-$(G^*C)_d$. Atoms depicted with balls are set to higher layer and with wires are set to lower layer in ONIOM calculations.

shown in Table 9. From this table, the energy differences between the original structure and the proton-transferred structures are estimated as 3-3.5 kcal/mol. From these results, the reaction energy of the SPT reaction is estimated to be lower than the case of 2bps model.

cis-$(CG)_p$-Pt-$(GC)_d$	0
cis-$(CG^*)_p$-Pt-$(GC)_d$	+ 3.2
cis-$(CG)_p$-Pt-$(G^*C)_d$	+ 3.0

a: We set the energy of cis-$(CG)_p$-$(GC)_d$ without sodium atoms as 0.

Table 9. The differences of energy in "4bps model" molecules (in kcal/mol)

We next compared the optimized structure with experimental data from the protein data bank (PDBID: 1A84, 1AU5 and 1KSB) as summarized in Table 10. The hydrogen bond lengths of 1A84 and the optimized structure show poor agreement with each other, where the distance between $(O_6)_d$ and $(N_4)_d$ is about 3.3 Å in the former, which is too large to form hydrogen bonds. Except for this result, the optimized 4bps model of cis-$(CG)_p$-Pt-$(GC)_d$ agrees well with the experimental data in the coordination Pt-G and the error is within 0.05 Å. On the other hand, the error of the hydrogen bond length is significant, e.g., the hydrogen bond length of $(O_2)_p$-$(N_2)_p$ is 2.75 Å for the 4bps model of cis-$(CG)_p$-Pt-$(GC)_d$ and 2.96 Å for 1AU5. This is because the energies of cis-$(CG^*)_p$-Pt-$(GC)_d$ and cis-$(CG)_p$-Pt-$(G^*C)_d$ are as stable as those of cis-$(CG)_p$-Pt-$(GC)_d$ so that the mean of these structures may be observed in the experiment because of the low energy barrier of the SPT reaction. Indeed, the error is improved when the structures of cis-$(CG^*)_p$-Pt-$(GC)_d$ and cis-$(CG)_p$-Pt-$(G^*C)_d$ are taken into account. In this case, it is possible that the dynamic fluctuations may dominate the structure of the system.

	cis-$(CG)_p$-Pt-$(GC)_d$	cis-$(CG^*)_p$-Pt-$(GC)_d$	cis-$(CG)_p$-Pt-$(G^*C)_d$	1A84	1AU5	1KSB
Pt-G coordination						
[a]Pt-G_p	2.04	2.03	2.04	2.05	1.96	2.01
[a]Pt-G_d	2.04	2.04	2.03	2.05	1.98	2.01
[b]Gp-Pt-Gd	89.4	91.4	89.6	90.1	87.4	88.6
G_p-Pt-L_1-G_d	53.7	58.6	63.3	40.8	–1.4	56.9
G_p-Pt-L_2-G_d	69.2	69.5	64.6	72.9	–17.5	55.3
Hydrogen bond length						
$(O_6)_p$-$(N_4)_p$	2.84	2.62	2.85	2.97	2.73	2.76
$(N_1)_p$-$(N_3)_p$	2.85	2.77	2.87	3.01	2.96	2.86
$(O_2)_p$-$(N_2)_p$	2.73	2.92	2.79	2.91	2.87	2.91
$(O_6)_d$-$(N_4)_d$	2.82	2.92	2.65	3.28	2.80	2.83
$(N_1)_d$-$(N_3)_d$	2.81	2.84	2.76	2.99	2.71	2.89
$(O_2)_d$-$(N_2)_d$	2.75	2.74	2.96	2.59	2.96	2.85

a: The distance Pt-G is defined as the distance between Pt and N_7 of guanine.
b: The angle G_p-Pt-G_d is defined as N_7 (G_p)-Pt-N_7 (G_d).
c: The unit of distance is angstrom, the unit of angle is degree.

Table 10. Selected lengths of optimized geometries and the experimental data from the protein databank

For further research, it is also necessary to discuss the possibilities of proton-transfer in larger system. In such systems, it is necessary to consider not only the optimized structure but also structural fluctuations. The mean structure between the original and the SPT structures may be observed at room temperature. The dynamic effects will be investigated in our future work.

4.4 Summary of section 4

The binding affinity of cisplatin to base pairs were in the following order in our calculation, cis-(CG)-Pt-(GC), cis-(CG)-Pt-(AT), and cis-(TA)-Pt-(AT), when complementary base pairs were taken into account. From their energetics, the structure of cis(TA)-Pt-(AT) is expected to be notfound at room temperature. The SPT reaction can occur in systems that consist of two base pairs and cisplatin. The reaction barrier is as low as 6–7 kcal/mol, which is similar to the case of one GC pair with cisplatin, and the SPT structure is as stable as the original structure. From these results, it is possible that the coordination of cisplatin to DNA causes a mispairing of the GC pair that leads to a mutation of the DNA. The SPT reaction causes this in one of the GC pairs. On the other hand, two simultaneous SPT structures like cis-(CG*)-Pt-(G*C) are forbidden. This is explained by the analysis of charge distributions. After one SPT reaction occurs, the proton donor G becomes negative and the proton acceptor C positive. At the same time, the other C also becomes positive through electrostatic effect from virtically stacked base pair that under went SPT reaction. Therefore, the subsequent SPT reaction from the other G is forbidden by Coulomb repulsion. The SPT reaction between G and C can occur with cis-(CG)-Pt-(AT). This result is similar to the case of one GC pair.

By using the ONIOM method, the SPT reaction is also shown to occur in the system consisting of cisplatin and four base pairs containing the backbone molecules (4bps model). Without the effects of the backbone and the stacking base pairs, the structure of cis-$(CG)_p$-Pt-$(GC)_d$ is so distorted that we cannot expect it to describe the actual structure in the DNA. The optimized structure of the 4bps cis-(CG)-Pt-(GC) model agrees with results from NMR experiments in view of the Pt-G coordination, but not of the hydrogen bond length. Because the structures of cis-$(CG^*)_p$-Pt-$(GC)_d$ and cis-$(CG)_p$-Pt-$(G^*C)_d$ are as stable as the original one, their mean structures may be observed in experiments.

5. Conclusion

We investigated the change of proton-transfer reactions in DNA base pairs caused by the coordination of cisplatin by density functional theory (DFT) and ONIOM method. When the cisplatin binds to GC pair, the structure undergoes intermolecular proton transfer from G to C (denoted as G*C pair) resulting in an increase of the bonding energy by 3-10 kcal/mol. This renders the structure to be metastable due to (a) successive processes of charge transfer from G to cisplatin thereby stabilizing the GC and G*C pairs and (b) an additional hydrogen bond between G and the ligand of Pt atom. From the energetics of two base pairs with the cisplatin, it is theoretically confirmed that the Pt complex is likely to bind in the following order: cis-(CG)-Pt-(GC), cis-(CG)-Pt-(AT), cis-(TA)-Pt-(AT).The Pt atom is expected to bind to the N_7 site of G and A. This result supports the experimental evidence, where the structure cis-A-Pt-A is seldom observed at room temperature. The single proton-transfer reaction occurs in one of the two GC pairs. No simultaneous single proton-transfer reaction can occur in both base pairs. Two different single proton-transferred structures (cis-$(CG^*)_d$-

$Pt\text{-}(GC)_p$ and $cis\text{-}(CG)_d\text{-}Pt\text{-}(G^*C)_p$, where * means a proton donor of G) are as stable as the original structures $(CG)_d\text{-}Pt\text{-}(GC)_p$. The same tendency was observed with cis-(CG*)-Pt-(AT). In contrast to cisplatin, multiple single proton-transfer reactions may occur in the system consisted of two base pairs with transplatin. The optimized structure agrees with the experimental data for Pt-G coordination except for the hydrogen bond length.

6. Acknowledgment

This study is supported by a Grant-in-Aid for Young Scientists (A) (No. 22685003) from Japan Society for the Promotion of the Science (JSPS) T.M. is thankful to the research fellowship for young scientists from JSPS. We thank Prof. S. Sakaki in Kyoto University for helpful advice.

7. References

[1] Jurečka, P. & Hobza, P. (2003). True Stabilization Energies for the Optimal Planar Hydrogen-Bonded and Stacked Structures of Guanine·Cytosine, Adenine·Thymine, and Their 9- and 1-Methyl Derivatives: Complete Basis Set Calculations at the MP2 and CCSD(T) Levels and Comparison with Experiment. *Journal of the American Chemical Society* Vol.125, No.50, (December 17, 2003), pp. 15608-15613, ISSN 0002-7863.

[2] Šponer, J. & Jurečka, P.; Hobza, P. (2004). Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs. *Journal of the American Chemical Society*, Vol.126, No.32, (August 18, 2004), pp. 10142-10151, ISSN 0002-7863.

[3] Löwdin, P.-O. (1963). Proton Tunneling in DNA and its Biological Implications. *Review of Modern Physics*, Vol.35, No.3, (July-September), pp. 724-732, ISSN 1539-0756.

[4] Florián, J.; Hrouda, V. & Hobza, P. (1994). Proton Transfer in the Adenine-Thymine Base Pair. *Journal of the American Chemical Society*, Vol.116, No.4, (February, 1994), pp.1457-1460, ISSN 0002-7863.

[5] Florián, J. & Leszczynski, J. (1996). Spontaneous DNA Mutations Induced by Proton Transfer in the Guanine·Cytosine Base Pairs: An Energetic Perspective. *Journal of the American Chemical Society*, Vol.118, No.12, (March 27, 1996), pp.3010-3017, ISSN 0002-7863.

[6] Villani, G. (2006). Theoretical investigation of hydrogen transfer mechanism in the guanine-cytosine base pair. *Chemical Physics*, Vol.324, No.2-3, (May 31, 2006), pp. 438-446, ISSN 0301-0104.

[7] Shigeta, Y.; Miyachi, H.; Matsui, T. & Hirao, K. (2008). Dynamic Quantum Isotope Effects on Multiple Proton-Transfer Reactions. *Bulletin of Chemical Society of Japan*, Vol.81, No.10, (October, 2008), pp. 1230-1240, ISSN 1348-0634.

[8] Yamamura, M.; Ichino, T. & Yoshioka, Y. (2011). A B3LYP Study on Repair of Guanyl and 8-Oxoguanyl Radical by Simultaneous Proton- and Electron-Transfer Reaction. *Bulletin of Chemical Society of Japan*, Vol.84, No.2, (February, 2011), pp. 181-190, ISSN 1348-0634.

[9] Matsui, T.; Shigeta, Y. & Hirao, K. (2006). The influence of Pt complex between guanine-cytosine pair, *Chemical Physics Letters*, Vol.423, No.4-6(June 1, 2006), pp.331-334, ISSN 0006-2614.

[10] Matsui, T.; Shigeta, Y. & Hirao, K. (2007). Multiple proton-transfer reactions in guanine-cytosine pairs by the coordination of Pt complex. *The Journal of Physical Chemistry B*, Vol.111, No. 5(February 15), pp.1176-1181, ISSN 1520-6106.

[11] Rosenberg, B.; van Camp, L.; Trosko, J. L. & Mansour, V.H. (1969), Platinum Compounds: a New Class of Potent Antitumour Agents. *Nature*, Vol.222, No.5191, (April 26, 1969), pp. 385-386, ISSN 0028-0836.

[12] Eastman, A. (1986), Reevaluation of interaction of cis-dichloro(ethylenediamine) platinum(II) with DNA. *Biochemistry* Vol. 25, No.13, (July, 1986), pp. 3912-3915, ISSN 0006-2960.

[13] Burda, J.V. & Leszczynski, J. (2003). How strong can the bend be on a DNA helix from cisplatin? DFT and MP2 quantum chemical calculations of cisplatin-bridged DNA purine bases. *Inorganic Chemistry*, Vol.42, No.22, (November 3, 2003), pp.7162-7172, ISSN 0020-1669.

[14] Zeizinger, M.; Burda, J.V. & Leszczynski, J. (2004). The influence of a sugar-phosphate backbone on the cisplatin-bridged BpB' models of DNA purine bases. Quantum chemical calculations of Pt(II) bonding characteristics. *Physical Chemistry Chemical Physics*, Vol.6, No.13 (July 7, 2004), pp.3585- ISSN 1463-9076.

[15] Szabo, A. & Ostlund, N.S. (1996). *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover publications, ISBN 978-0486691862, New York, USA.

[16] Jensen, F. (2006). *Introduction to Computational Chemistry 2nd edition,* John Wiley & Sons, ISBN 978-0471984252, Chichester, UK.

[17] Adamo, C. & Barone, V.(1998). Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models, *Journal of Chemical Physics*, Vol.108, No.2 (Jan 8, 1998), pp.664-675, ISSN 1089-7690.

[18] Svensson, M.; Humbel, S.; Froese, R.D.J.; Matsubara, T.; Sieber, S. & Morokuma, K. (1996). ONIOM: A multi-layered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and $Pt(P(t\text{-}Bu)_3)_2+H_2$ oxidative addition. *The Journal of Physical Chemistry*, Vol.100, No.50(December 13, 1996), pp.19357-19363, ISSN 1520-6106.

[19] Frisch, M.J. et al. (2004). *Gaussian 03*, *Revision C.02*, Gaussian, Inc., Wallingford CT, USA.

[20] Matsui, T.; Sato, T. & Shigeta, Y. (2009). Sequence dependent proton-transfer reaction in stacked GC pair I: The possibilities of proton-transfer reaction. *International Journal of Quantum Chemistry*, Vol.109, No.10(August 15, 2009), pp.2168-2177, ISSN 0020-7608.

8

Quantum Mechanics on Surfaces

Bjørn Jensen
Departement of Micro and Nano Systems Technology
Vestfold University College
Norway

1. Introduction

Quantum theory in curved spaces has received much attention over the years. It has been applied in the study of black holes and large scales structures in the universe as well as in the study of the Casimir effect, e.g. However, transferring our existing formulations of quantum theory to curved spaces is not straightforward and any approach will be hampered with a number of issues (1). Nanostructures provide an experimental arena which potentially can provide direct evidence for the interplay between geometry and quantum theory. The ability to manufacture micro- and nanosized surfaces has open up new vistas which should be utilized in order to get a firmer grip on the quantum theory of particles living on these curved structures.
In the following a brief account of the most widely accepted formulation (the 'standard' formulation) of Shrödinger theory on surfaces and linear structures in ordinary three-dimensional Euclidean space is given. We then apply this framework to derive a quantum theory on the catenoid in three-dimensional Euclidean space. This will highlight some important features connected with the interplay between quantum theory and geometry. Then follows a partial framework for an alternative formulation of Shrödinger theory on a surface in which we utilize the unique conformal properties of two-dimensional surfaces. Even though most work connected with quantum theory on structures embedded in three-dimensional Euclidean space so far have been concerned with surfaces, wire structures are also of great obvious interest. Next we therefore point to the possible importance of employing ideas from supersymmetric quantum mechanics in order to enhance our understanding of these structures. Workers in the field of quantum mechanics on lower-dimensional structures in flat space have mainly concerned themselves with Shrödinger theory. In the remaining part of this brief account we will concern ourselves with Dirac theory on surfaces in three-dimensional Euclidean space. We look at differences between the first and second order formulations, and device the proper framework for formulating Dirac theory on surfaces and linear structures in a way which makes contact with the standard formulation of Shrödinger theory on these structures. We then explore different issues, including the question of whether it is 'sufficient' to employ an intrinsically defined quantum theory in a surface compared to the standard approach in the context of Dirac theory. This issue should be of particular relevance when formulating effective theories for charge carriers in graphene. No effort has been made to provide an exhaustive list of

references. Those references which are cited, and the references they contain, are those which have been of particular importance for this author.

2. Dimensional reduction

Quantum mechanics has come to age. We can claim that its basic formulation, either in the form of the canonical quantization procedure or in terms of the Dirac quantization program, is well understood in the sense that its formulation is transparent, even though its consequences continue to surprise and baffle us all. However, this claim is only true as long as the theory is formulated in an Euclidean space which is charted with Cartesian coordinates. It was early recognized in the Dirac quantization programe that problems generally arise when non-Euclidean coordinates are used (2). This is a signal that the interplay between quantum theory and geometry is a deep and fundamental one. This interplay took center stage in physics when it was shown that black holes radiate (3). Curved spacetime geometries are not of immediate importance for quantum theory in the laboratory setting, but curved surfaces and linear structures are. The Casimir effect (4) is a well known example which proves this. Aspects of the challenges met in quantum theory in 'exotic' space-times might therefore also appear in more everyday settings. The coordinate challenge in the Dirac quantization program (e.g.) definitely does since curved surfaces in ordinary space can generally not be completely charted with Cartesian coordinates. The most generally accepted adaption of the canonical quantization procedure to curved surfaces and linear structures was developed in (5–7). What follows is a brief account of this adaptation. We will not systematically discuss the Dirac quantization procedure and possible adaptations of it to lower dimensional structures in space in this exposition, but we will briefly comment on an important aspect of the latter in Section 4.

Consider a smooth two dimensional static surface $\mathcal{S}$ in ordinary three dimensional space. We follow the parametrization in (6) and chart the three dimensional embedding space with coordinates X^i. We write the metric as (6; 8)

$$\begin{aligned} ds^2 &= -dt^2 + G_{ij}(X^i)dX^i dX^j + (dX^3)^2 = \\ &= -dt^2 + G_{ab}(x^a)dx^a dx^b + (dx^3)^2\,, \end{aligned} \tag{1}$$

where $G_{ab}(x^a)$ is the metric in the surface $\mathcal{S}$ defined by coordinates x^a. We assume that we can define a normal vector field $\vec{N}$ everywhere on $\mathcal{S}$. The coordinate direction x^3 is assumed to be along $\vec{N}$ in the immediate vicinity of $\mathcal{S}$. Our conventions will be such that indices at the beginning of the alphabet will refer to the coordinates in the surface x^a, while indices in the middle of the alphabet refer to the global coordinates X^i. It follows that (8)

$$G_{ab}(X^i) = g_{ab}(x^a) - 2K_{ab}(x^a)x^3 + K^k{}_a(x^a)g_{km}(x^a)K^m{}_b(x^a)(x^3)^2 \tag{2}$$

$$G(X^i) = \det G_{ab}(X^i) = g(x^a)(1 - 8M(x^a)x^3 + (2K(x^a) + 8M(x^a)^2)(x^3)^2 + ...) \tag{3}$$

$$\sqrt{G(X^i)} \equiv \sqrt{g(x^a)}\xi(X^i)\,,\ \xi(X^i) = 1 - 4M(x^a)x^3 + K(x^a)(x^3)^2 + ... \tag{4}$$

where $g_{ab}(x^a)$ is the induced metric in the surface, $g(x^a) = \det g_{ab}(x^a)$ and $K^i{}_j(x^a)$ is the extrinsic curvature tensor associated with $\mathcal{S}$. $K = \det K^i{}_j$ and $M(x^a) \equiv G^{ij}(X^i)K_{ij}(X^i)$ is the mean curvature in $\mathcal{S}$.

Central to the approach developed in (5–7) is the assumption of the presence of forces which constrain the particle to $\mathcal{S}$. It is assumed that these forces act everywhere normal to $\mathcal{S}$ and that they can be derived from a potential $V_\lambda(X^3)$. λ is a parameter which measures the strength of the potential. The Schrödinger equation describing an electrically neutral particle in the embedding space within this framework is then given by (we use units such that $c \equiv \hbar \equiv 1$)

$$i\partial_t\psi = -\frac{1}{2m}G^{ij}\nabla_i(\nabla_j\psi) + V_\lambda(X^3)\,. \tag{5}$$

m denotes the particle mass. In order to derive a quantum theory in $\mathcal{S}$ we need to *dimensionally reduce* the Schrödinger equation. We therefore decompose the covariant derivative in a coordinate gauge invariant manner as a sum of one part which acts along the surface($||$), and one part which acts normal to the surface ($\perp$)

$$\nabla_i = \nabla_{||i} + \nabla_{\perp i}\,. \tag{6}$$

The purely kinetic term in the Schrödinger equation can then be written

$$G^{ij}\nabla_i\nabla_j\psi \equiv (\nabla^2_{||} + \nabla^2_{\perp})\psi = (\partial^a\partial_a\psi + G^{ab}\,\Gamma^c{}_{ab}\partial_c\psi) + (\partial^3\partial_3\psi + G^{ab}\,\Gamma^3{}_{ab}\partial_3)\psi\,. \tag{7}$$

In the last relation we have used the coordinate gauge Eq.(1). $\Gamma^i{}_{jk}$ represents the Christoffel symbols of the second kind. We will assume that the wave function is normalizable in the three dimensional embedding space, such that the norm is given by

$$N = \int d^3X\sqrt{G}|\psi|^2 = \int d^3x\sqrt{g}|\chi|^2\,. \tag{8}$$

Probability conservation requires that $\psi(X^i) = \xi(x^i)^{-1/2}\chi(x^i)$. We use this relation to compute the kinetic term and rewrite the Schrödinger equation in terms of χ. Clearly,

$$\lim_{x^3\to 0}\nabla^2_{||}\psi = \nabla^2_{||}\chi\,. \tag{9}$$

We also find that

$$\lim_{x^3\to 0}\nabla^2_{\perp}\psi = \lim_{x^3\to 0}\frac{1}{\sqrt{G}}\partial_3(\sqrt{G}\partial^3\psi) = \lim_{x^3\to 0}\xi^{-1}\partial_3(\xi\partial_3(\xi^{-1/2}\chi)) \equiv \partial^2_3\chi - V_0\chi\,. \tag{10}$$

Using these relations we find in the limit $x^3 \to 0$ that the Schrödinger equation becomes

$$i\partial_t\chi = -\frac{1}{2m}\nabla^2_{||}\chi - \frac{1}{2m}\partial^2_3\chi + V_0\chi + V_\lambda\chi\,, \tag{11}$$

where V_0 is given by (6)

$$V_0 = -\frac{1}{2m}(M^2 - K)\,. \tag{12}$$

We see that an effective potential has emerged depending on scalars characterizing the extrinsic curvature of $\mathcal{S}$. V_0 is clearly non-positive on any surface. If χ is separable into one part which is independent of x^3 and one part which only depends on this coordinate we have effectively deduced a quantum theory in the surface $\mathcal{S}$. This program can also be adapted to

linear structures by continuing the dimensional reduction procedure above. The result is (6)

$$-\frac{\hbar^2}{2m}\partial_x^2\psi - \frac{\hbar^2}{8m}\kappa^2(x)\psi = E\psi\,. \tag{13}$$

x is the coordinate along the structure, and $\kappa(x)$ is its local curvature. We will later return to this result in Section 5.

3. The catenoid

Let us apply the dimensional reduction approach above to a concrete surface. We will in particular consider the catenoid surface (9). This is a classical minimal surface. It can conceivably be realized in a bilayer of honeycomb lattices with radially arranged dislocations or in bilayer graphene (10). We choose the following parametrization for the catenoid $x = R\cosh(z/R)\cos\phi$, $y = R\cosh(z/R)\sin\phi$ and $z = z$, with $\phi \in [0, 2\pi]$ where x, y, z represents the canonical Cartesian coordinates in ordinary three-dimensional space (Fig. 1). The local radius $\rho = R\cosh(z/R)$ and the metric is thus given by

$$g_{\rho\rho} = \frac{\rho^2}{\rho^2 - R^2}\ ,\ g_{\phi\phi} = \rho^2. \tag{14}$$

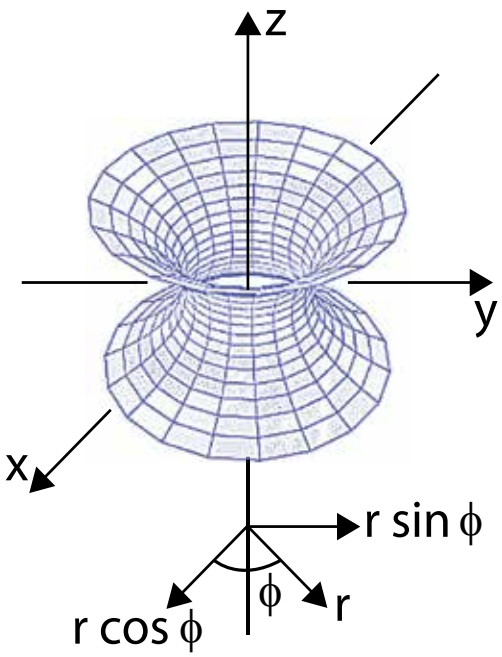


Fig. 1. A two-dimensional section (catenoid) of a three dimensional worm hole geometry with its axis along z and the throat radius R.

It is interesting to note that slices of a wormhole geometry is geometrically equivalent to a catenoid. In cylindrical coordinates (z, r, ϕ) a two-dimensional section of a wormhole is given by (11)

$$z(r) = \pm b_0 \ln\left[\frac{r}{b_0} + \sqrt{\frac{r^2}{b_0^2} - 1}\right], \tag{15}$$

with $l = \pm\sqrt{r^2 - b_0^2}$. The spatial part of the wormhole geometry is given by the following expression (11)

$$ds^2 = dl^2 + (b_0^2 + l^2)(d\theta^2 + \sin^2\theta d\phi^2), \tag{16}$$

where $l \in [-\infty, +\infty]$, $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$. b_0 is the shape function of the wormhole [in general $b = b(l)$ and for $l = 0, b = b(0) = b_0 = const.$ represents the radius of the throat of the wormhole]. Here l is a radial coordinate measuring proper radial distance; θ and ϕ are the spherical polar coordinates. Here we will consider the slice $\theta = \pi/2$ which represents an equatorial section of the wormhole geometry (at constant coordinate time). For this particular slice we thus get the following line element

$$ds^2 = dl^2 + (b_0^2 + l^2)d\phi^2, \tag{17}$$

which is precisely equivalent to the line element on a catenoid (since $l^2 = r^2 - b_0^2$)

$$ds^2 = \frac{r^2}{r^2 - b_0^2}dr^2 + r^2 d\phi^2. \tag{18}$$

Note that if we consider any other section of the three dimensional wormhole, say for $\theta = \theta_0$, the line element will change to

$$ds^2 = \frac{r^2}{r^2 - b_0^2}dr^2 + a^2 r^2 d\phi^2, \tag{19}$$

where $a^2 = \sin^2\theta_0$, and (obviously) $a^2 \in [0, 1]$. For the catenoid this will only mean a rescaling of the radius of the catenoid at the throat (the circle with least local radius) from R to aR. The line element Eq.(19) corresponds to a catenoid with $x = aR\cosh(z/aR)\cos\phi$, $y = aR\cosh(z/aR)\sin\phi$ and $z = z$. Thus all θ-sections of the physical wormhole at constant time coordinates represent a catenoid with radius aR. The catenoid with the biggest radius corresponds to the equatorial section $\theta = \frac{\pi}{2}$ and the one with zero radius to the section $\theta = \pi$. Returning to the catenoid and focusing on the (z, ϕ) coordinates (instead of (ρ, ϕ)), the line element is given by

$$ds^2 = \cosh^2(z/R)dz^2 + R^2\cosh^2(z/R)d\phi^2, \tag{20}$$

with the principal curvatures given by

$$\kappa_1 = \frac{1}{R}\mathrm{sech}^2(z/R), \quad \kappa_2 = -\frac{1}{R}\mathrm{sech}^2(z/R). \tag{21}$$

This implies that the mean curvature $M = (\kappa_1 + \kappa_2)/2 = 0$ (meaning that the surface is a minimal surface) and the Gaussian curvature $K = \kappa_1\kappa_2 = -(1/R^2)\mathrm{sech}^4(z/R)$. The

corresponding curvature induced potential on a catenoid is then given by

$$V(z) = -\frac{\hbar^2}{2m_0}(H^2 - K) = -\frac{\hbar^2}{2m_0 R^2}\text{sech}^4(z/R). \tag{22}$$

Note that for $a^2 \ll 1$ the potential becomes very deep at the origin. The corresponding Schrödinger equation is given by

$$-\frac{\hbar^2}{2m_0 R\cosh^2(z/R)}\left[R\frac{\partial^2\psi}{\partial z^2} + \frac{1}{R}\frac{\partial^2\psi}{\partial\phi^2}\right] - \frac{\hbar^2}{2m_0 R^2}\text{sech}^4(z/R)\psi = E\psi, \tag{23}$$

or simplifying

$$-R\frac{\partial^2\psi}{\partial z^2} - \frac{1}{R}\frac{\partial^2\psi}{\partial\phi^2} - \frac{\text{sech}^2(z/R)}{R}\psi = \frac{2m_0 R}{\hbar^2}\cosh^2(z/R)E\psi. \tag{24}$$

Using the cylindrical symmetry along the z-axis we set $\psi = e^{imz}\Phi$. We thus get the following equation for Φ

$$\Phi_{zz} - \frac{m^2}{R^2}\Phi + \frac{\text{sech}^2(z/R)}{R^2}\Phi + \frac{2m_0 E\cosh^2(z/R)}{\hbar^2}\Phi = 0. \tag{25}$$

Defining a dimensionless length $\zeta = z/R$ and energy $\epsilon = 2m_0 ER^2/\hbar^2$ we get the following effective Schrödinger equation

$$-\Phi_{\zeta\zeta} + V(\zeta)\Phi(\zeta) = 0, \tag{26}$$

where the potential now reads

$$V(\zeta) = [m^2 - \epsilon\cosh^2(\zeta)] - \text{sech}^2(\zeta). \tag{27}$$

This potential for $m \neq 0$ bears some similarity to the corresponding geometric potential for the physical wormhole (12). Note that in the ground state ($\epsilon = 0$, also called a *critically bound state* (13)) $V(\zeta)$ becomes the reflectionless Bargmann's potential (14) and the Schrödinger equation becomes the hypergeometric equation with the ground state wavefunction (the 'Goldstone mode') given by $\Phi(\zeta) = \text{sech}(\zeta)$. This result is remarkable since this implies that the catenoid surface enables complete transmission across it for a quantum particle. This does not seem to be the case for the physical wormhole geometry (12). For non-zero values on ϵ the above potential is an inverted double well potential shown in Fig. 2.
Let us consider Eq.'s(26-27) in some more detail. We see in particular that

$$\lim_{\zeta\to\pm\infty}|V| \to \infty\,. \tag{28}$$

This behavior of the potential at infinity is strange since the geometry on the catenoid in these regions approaches the usual Euclidean one. This feature can be traced to the coordinates used since the proper length per unit in the ζ direction diverges when $\zeta \to \pm\infty$. This can be remedied by introducing another set of coordinates on the catenoid.

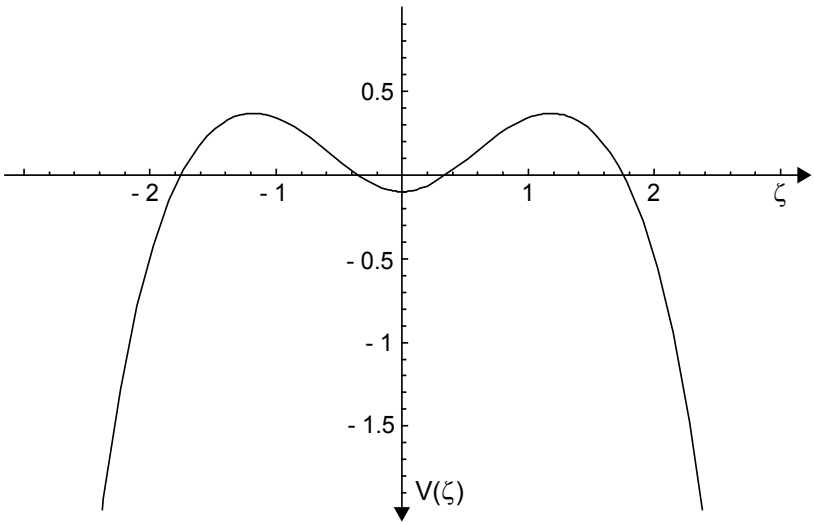


Fig. 2. The inverted double well potential $V(\zeta)$ with $m = \pm 1$ and $\epsilon = 0.1$.

Quantum theory in curved spaces is generally a challenge since the theory is not generally covariant. Classical quantum theory is not even Lorentz invariant. This puts a severe constraint on the coordinate system in which one wishes to describe the physics in order to be able to extract the physical content of the theory. This challenge was even central in the early days of general relativity theory itself in connection with the physical interpretation of the Schwartzshild metric, e.g. Just like as in general relativity theory one is usually safe concerning the physical interpretation, as well as the definitions of physical quantities, when the manifold in question is asymptotically Minkowski (Euclidean). In such asymptotic regions we expect on physical grounds to rederive the usual flat space physical results. The asymptotic properties thus in some sense anchor the curved region and its physics to reality, as we know it. The catenoid is in this sense an asymptotic Euclidean object, thus making this manifold a space anchored to 'reality'.

The catenoid surface can in some sense be perceived as a deformation of the plane. Considering briefly the two-dimensional Schrödinger equation in the plane in polar coordinates we get the Bessel equation. Clearly, the boundary condition at the origin is suspect here. However, in our case we can as a first approximation consider a deformation of the plane in a region around the origin. In the deformed region the Schrödinger equation will generally be very complicated but the solutions of it must nevertheless be matched to the Bessel functions which survive sufficiently far from the deformed region. This reasoning goes *ad verbum* through also on the catenoid even though we here, in addition to curvature corrections, also have a topology change when compared to the plane since not any closed curve on the catenoid surface is null homotopic. Hence, we should seek coordinates on the catenoid such that the Schrödinger equation gives rise to the Bessel equation in the asymptotic region on the catenoid. The coordinates should thus in particular result in a metric which is reminiscent of polar coordinates at infinity. It is possible to find such coordinates if one covers the entire catenoid manifold with two coordinate patches. One patch covers the region $\zeta > 0$,

the other $\zeta < 0$. In the upper part we choose the new radially directed coordinate as

$$\eta^+ = e^{\zeta} - 1\,;\, \zeta > 0\,. \tag{29}$$

In the lower part we correspondingly choose

$$\eta^- = -(e^{-\zeta} - 1)\,;\, \zeta < 0\,. \tag{30}$$

Clearly $\eta^+ = \eta^-$ at $\zeta = 0$. The invariant line-element can then be written as

$$ds^2 = \frac{((\eta^{\pm} \pm 1)^2 + 1)^2}{4(\eta^{\pm} \pm 1)^4}(d\eta^{\pm})^2 + \frac{1}{4}(\frac{(\eta^{\pm} \pm 1)^2 + 1}{\eta^{\pm} \pm 1})^2 d\phi^2\,. \tag{31}$$

In the limit $\eta^{\pm} \to \pm\infty$ the metric can be written as

$$ds^2 = \frac{1}{4}(d\eta^{\pm})^2 + \frac{1}{4}(\eta^{\pm})^2 d\phi^2\,. \tag{32}$$

Hence, the asymptotic form of this metric is very similar to the Euclidean metric expressed in ordinary polar coordinates. Indeed, they are exactly the same something which is easily seen by a simple rescaling of the radial coordinates. These new coordinates should therefore be well suited to explore the physical states of a quantum particle on the catenoid.
Let us now consider the Schrödinger equation (26). In terms of the new coordinates we have in particular that

$$\partial^2_{\zeta}\Phi = (\eta^{\pm} \pm 1)\partial_{\pm}((\eta^{\pm} \pm 1)\partial_{\pm}\Phi)\,, \tag{33}$$

$$\cosh u = \pm\frac{1}{2}(\frac{(\eta^{\pm} \pm 1)^2 + 1}{\eta^{\pm} \pm 1})\,. \tag{34}$$

This gives rise to identical expressions for the Schrödinger equation in the two coordinate patches. In the upper patch the equation is explicitly given by

$$\partial^2_{+}\Phi + \frac{1}{(\eta^+ + 1)}\partial_{+}\Phi + \left[\frac{\epsilon}{4} - \frac{(m^2 - \epsilon/2)}{(\eta^+ + 1)^2} + \frac{1}{4}(\frac{\epsilon}{(\eta^+ + 1)^4} + \frac{16}{((\eta^+ + 1)^2 + 1)^2})\right]\Phi = 0\,. \tag{35}$$

Clearly, letting $\eta^+ \to \infty$ we easily get the Bessel equation, which is well behaved at infinity.
The stationary Schrödinger equation, and assuming a well defined energy E eigenvalue problem, is formally given by

$$(-\nabla^2 + V)\Psi = E\Psi\,. \tag{36}$$

Hence, we have that

$$V - E = -\left[\frac{\epsilon}{4} - \frac{(m^2 - \epsilon/2)}{(\eta^+ + 1)^2} + \frac{1}{4}\left(\frac{\epsilon}{(\eta^+ + 1)^4} + \frac{16}{((\eta^+ + 1)^2 + 1)^2}\right)\right]. \tag{37}$$

In the asymptotic region we find that

$$\lim_{\eta^+ \to \infty} V = E - \frac{1}{4}\epsilon > 0\,. \tag{38}$$

We have plotted the potential for $m = 0$ and $\epsilon = 2$ in Fig. 3. Note that in this s-channel the potential becomes negative sufficiently close to the origin.

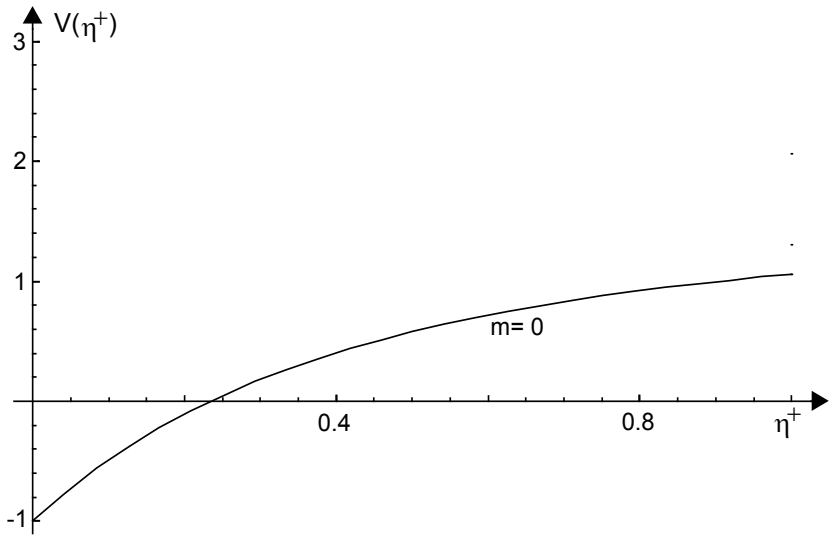


Fig. 3. The potential $V(\eta^+)$ with $m = 0$ and $\epsilon = 2$.

Clearly, the constant part of the potential can be renormalized to zero without any physical consequences. Hence, the renormalized potential V_r can be taken to be

$$V_r - E = -\left[-\frac{(m^2 - \epsilon/2)}{(\eta^+ + 1)^2} + \frac{1}{4}\left(\frac{\epsilon}{(\eta^+ + 1)^4} + \frac{16}{((\eta^+ + 1)^2 + 1)^2} \right) \right]. \tag{39}$$

4. Conformal deformations

The framework derived in (5–7) has become a standard one for analyzing quantum mechanics on surfaces and one-dimensional structures. The resulting theory is a theory framed in curved spaces. As such it is often difficult to deal with both from a purely computational perspective as well as also from an interpretational perspective. On the computational side the kinetic part of the Schrödinger equation will very often be highly complicated function of the metric tensor. This will typically give rise to complicating second order derivative terms which mix the spatial coordinates. The often lack of an 'asymptotically' flat region of a given surface (e.g.) often gives rise to all sorts of interpretational issues which also often appear in quantum theories in curved spaces (1). Our treatment of the catenoid above highlight this issue in particular. Here we sketch a framework which might be of utility in resolving some of these issues. We emphasize that it is a sketch which is presented, and not a complete and polished framework.

We entertain the following idea. It is well known that every open, connected and smooth surface in three-dimensional Euclidean space is conformally flat. Physically we can picture this as it is always possible to deform S in such a way that the local geometry, all the time expressed in the same coordinates, has a conformally flat form all the way to the point when S has become a two-dimensional plane P. We call such a process a *conformal deformation* of S. The

Schrödinger equation is not invariant under conformal transformations. Hence, is it possible in general to re-express the quantum theory on a given surface S on the two-dimensional plane P using conformal techniques in such a way as to bypass (some of) the issues mentioned above which arises in the more physically motivated dimensional reduction framework? We show that it is possible to reformulate the two-dimensional quantum theory on S in the plane such that it can be looked upon as standard free two-dimensional Schrödinger theory coupled to an external potential. The wave-function on P couples to a new potential V_P which exhibits a vector field $\vec{W}$ and a scalar field W, both of a purely geometric origin on S

$$V_P = -\frac{\hbar^2}{2m}(\vec{W}\cdot\nabla_P + W)\,. \tag{40}$$

∇_P denotes the metric compatible connection on P. Hence, certain computational issues is bypassed by this approach since the theory is formulated in a flat space which in Cartesian coordinates will give rise to a 'trivial' kinetic operator. The formulation will also be easier to interpret both because of the non-existing intrinsic curvature in the plane and its vanishing extrinsic curvature, and because V_P can be treated as a completely *external* potential not associated with the space P in which the theory is formulated. This might enhance our understanding of the physical picture on S.

We consider a smooth two-dimensional surface S in ordinary three-dimensional Euclidean space. In the present work we will focus on *open* surfaces. We will briefly deal with compact surfaces at the end of this section. We will repeat some of the mathematical technicalities from section 2 in order to fix the needed notation. We assume that we can define a normal vector field $\vec{N}$ everywhere on S. We will choose coordinates in the embedding space in such a way that one coordinate basis vector is always parallel to $\vec{N}$ in the vicinity of S. We will denote the associated coordinate by x^3 in the following and the coordinates on S by x^1, x^2. We denote the metric tensor induced on S by G_{Sab} such that the indices refer to the coordinates in S. In the dimensional reduction approach it is assumed that the quantum particle is constrained to S by a constraining potential $V_{S\lambda}$ where λ is a measure of the strength of the potential. It is assumed that the constraining forces always act along $\vec{N}$. In order to derive a theory on S we decompose the connection ∇ in the three-dimensional Euclidean space into a component acting solely in the surface ∇_S and a component $\nabla_\perp$ acting along the normal vector

$$\nabla = \nabla_S + \nabla_\perp\,. \tag{41}$$

The potential $V_{S\lambda}$ can be thought of as an infinite well potential such that S is sandwiched between two potential 'walls' (6). In the limit when the width Δx^3 of the potential well goes to zero, when the particle is literally forced to follow the surface S, we get (6)

$$\lim_{\Delta x^3\to 0}\nabla_\perp^2 = \partial_3^2 - V_S(x^1, x^2)\,. \tag{42}$$

In these calculations coordinates are chosen such that $G_{3i} = 0\,, G_{33} = 1\,, ; i \neq 3$. The Schrödinger equation can then be written as

$$i\partial_t\chi_S = -\frac{\hbar^2}{2m}\nabla_S^2\chi_S - \frac{\hbar^2}{2m}\partial_3^2\chi_S + V_S\chi_S + V_{S\lambda}\chi_S\,. \tag{43}$$

If this equation is separable we are left with a theory in the surface S. The theory in S carries with it a 'memory' of the three-dimensional embedding space through the potential V_S.
Any two-dimensional metric can locally be written in a conformally Euclidean form. This means explicitly that there exists a coordinate transformation, *a conformal diffeomorphism* $\mathcal{D}$: $(x^1, x^2) \to (X^1, X^2)$ such that

$$\begin{aligned} ds_S^2 &= G_{Sij} dx^i dx^j \\ &\equiv \omega^2(X^1, X^2)((dX^1)^2 + (dX^2)^2) \equiv \omega^2 ds_P^2 \equiv G_{Pij} dX^i dX^j \,. \end{aligned} \tag{44}$$

Here ds_S^2 and ds_P^2 are the metrics on S and P respectively, and $\omega^2(X^1, X^2)$ is a positive definite scalar function. We will be concerned with the situation when the metric ds_S^2 on S undergoes a *point wise* conformal transformation $\mathcal{T}$ such that

$$\mathcal{T} : ds_S^2 \to \Omega^2(X^1, X^2) ds_S^2 \equiv ds_P^2 \,. \tag{45}$$

$\Omega^2(X^1, X^2)$ is a positive definite function. In the case when none of the coordinates are periodic this represents deformations of S to a plane. We will primely assume this picture in the following. Physically we will picture the transformation $\mathcal{T}$ as either an adiabatic or an instantaneous one in order not to perturb the quantum system out of the quantum state it exhibits on S. For definiteness assume here an instantaneous process such that $\Omega = \omega^{-1}$.
The normal vector field $\vec{N}$ is assumed normal to the surface during the complete deformation process. Hence, the system of coordinates defined by (X^1, X^2, x^3) does therefore represent *comoving isothermal coordinates*. We also assume that the external confining potential V_λ always has a form such that the resulting force acting on the particle is along $\vec{N}$. We will furthermore assume that the particle never escapes the surface; we will assume conservation of probability during the deformation process, i.e. This implies in particular that the integral of the probability density on S must equal the same integral over the plane. If we denote the wave-function on P by χ_P it follows that

$$\int_S dS \sqrt{G_S} |\chi_S|^2 = \int_S dS \sqrt{G_P} |\chi_P|^2 \,, \tag{46}$$

where $G_S = \det(G_{Sij})$, $G_P = \det(G_{Pij})$. Since $\mathcal{T} : G_P = \Omega^4 G_S$, it follows that

$$\chi_S = \Omega \chi_P \,. \tag{47}$$

Hence, the wave-function must transform with conformal weight equal unity under the restricted class of conformal transformations which we deal with in this work. There exist previous studies in the literature of the properties of the Schrödinger equation under conformal transformations. Interestingly, these have to our knowledge only been concerned with space-time or space conformal transformations and not conformal transformations restricted to (hyper-) surfaces.
Two metric compatible connections ∇_S and ∇_P on two conformally related metrics ds_S^2 and ds_P^2 are related by

$$\nabla_{Pi} \omega_b = \nabla_{Si} \omega_j - C^k{}_{ij} \omega_k \tag{48}$$

where ω_k is some one-form, and

$$C^k{}_{ij} = 2\delta^k{}_{(i}\nabla_{Pj)}\ln\Omega - G_{Pij}G_P{}^{kl}\nabla_{Pl}\ln\Omega\,. \tag{49}$$

It follows that

$$\nabla_S^2 = \Omega^2(\nabla_P^2 + G_P{}^{ij}C^c{}_{ij}\nabla_{Pc}) \tag{50}$$

and

$$\nabla_S^2(\chi_S) = \nabla_S^2(\Omega\chi_P) = \Omega^2(\Omega\nabla_P^2\chi_P + V^i\nabla_{Pi}\chi_P + V\chi_P)\,, \tag{51}$$

where

$$V^i = 2G_P{}^{ji}\nabla_{Pj}\Omega + G_P{}^{jk}C^i{}_{jk}\Omega\,, \tag{52}$$

$$V \equiv \nabla_P^2\Omega + G_P{}^{ij}C^k{}_{ij}\nabla_{Pk}\Omega\,. \tag{53}$$

The extrinsic curvature of S is $K^j_{Si} = \nabla_{Si}N^j$. It follows that

$$K_{Sij} = K_{Pij} + C^3{}_{ij}N_3\,. \tag{54}$$

Identify P with the two-dimensional Euclidean plane. We then have that

$$K_{Pij} \equiv 0 \Rightarrow \begin{cases} K_S = \det(K_{Sij}) = \det(C^3{}_{ij}) \\ M_S = \mathrm{tr}(K_{Sij}) = -\mathrm{tr}(C^3{}_{ij}) \end{cases} \tag{55}$$

where

$$C^3{}_{ij} = (\delta^3{}_i\nabla_{Pj} + \delta^3{}_j\nabla_{Pi})\ln\Omega\,. \tag{56}$$

The $(C^3{}_{ij})$-matrix is explicitly given by

$$(C^a{}_{bc}) = \begin{pmatrix} 0 & 0 & C^3{}_{13} \\ 0 & 0 & C^3{}_{23} \\ C^3{}_{31} & C^3{}_{32} & 0 \end{pmatrix}\,. \tag{57}$$

Hence

$$K_S = 0\,,\; M_S = 0\,. \tag{58}$$

We note that this also implies that $G_P^{ij}C^3{}_{ij}\nabla_{P3} = 0$ since G_P^{ij} is diagonal. We then get the highly non-trivial result that the potential V_S transforms to zero under conformal deformations to P. Consequently, the Schrödinger equation on P reads

$$i\partial_t\chi_P = -\frac{\hbar^2}{2m}\Omega^2\nabla_P^2\chi_P - \frac{\hbar^2}{2m}\partial_3^2\chi_P - \frac{\hbar^2}{2m}(\vec{W}\cdot\nabla_P + W)\chi_P + V_{P\lambda}\chi_P\,, \tag{59}$$

where

$$W^i \equiv \Omega^{-1}V^i\;,\; W \equiv \Omega^{-1}V\,. \tag{60}$$

We assume that we can separate the motion orthogonal to the surface from the motion in the surface. We will also assume stationary states such that the time-dependent part of the wave-equation can be written as a simple exponential $\sim e^{-iEt}$ where we identify E with the

total energy of the states. Hence, Eq. (59) can be decomposed into

$$-\frac{\hbar^2}{2m}\nabla_{||}^2\chi_P + \frac{1}{\Omega^2}(E - \kappa^2 + V_P)\chi_P = 0\,, \tag{61}$$

$$-\frac{\hbar^2}{2m}\partial_3^2\chi_P\chi_P + V_{P\lambda}\chi_P = \kappa^2\chi_P \tag{62}$$

where we have defined $V_P \equiv -\frac{\hbar^2}{2m}(\vec{W}\cdot\nabla_P + W)$ and introduced a separation constant κ. This constant can naturally be identified with the momentum in the direction perpendicular to the surface.

Clearly, the set of equations above can be interpreted as describing a particle interacting with two potentials. Equation (61) is in general simpler than Eq. (43). Even though Eq. (61) like Eq. (43) in addition to the second order kinetic terms also exhibits first order differential operator terms contained in the V_P-potential it will not contain differential operator cross terms as Eq. (43) in general will exhibit. Furthermore, note that the potentials W and $\vec{W}$ are functions of Ω. Since Ω does not describe P we can interpret these terms as *external* potentials which are applied to P in very much the same fashion as $V_{P\lambda}$.

In the analysis above we have for simplicity assumed that P is identified with the two-dimensional plane. Clearly, P must not by necessity be identified with the plane in order to have a target manifold P with vanishing intrinsic curvature. One natural example which is also of practical importance is the straight tube. Clearly, when P is a curved surface (with non-vanishing extrinsic curvature, but with vanishing intrinsic curvature) the effective geometric potential on P will become more complicated compared to the exact planar situation. This is easily seen from the transformation property of V_S under a conformal deformation. From the expression for K_{Sij} in terms of K_{Pij} and $C^k{}_{ij}$ it follows that the transformed of the potential V_S will exhibit products between the extrinsic curvature tensor on P and $C^3{}_{ij}$-terms. Hence, changing the *topology* of S away from the planar one (every closed curve is null homotopic) will alter the induced potential on S in a fundamental way.

Let us summarize. In (5–7) a physically motivated framework for dealing with quantum mechanics on surfaces and linear structures in ordinary three dimensional Euclidean space were developed. Here we have attempted to reformulate this framework for quantum mechanics on surfaces into a framework on the two-dimensional plane. On the plane a quantum particle is shown to interact with an external potential V_P in addition to the external potential $V_{P\lambda}$ which constrains the particle to the plane. Clearly, the conformal formulation presented in this work represents *a priori* a simplification computation wise. It also represents an interesting tool in the analysis of quantum mechanics on a given surface. The form of the geometric potential on the plane provides an immediate physical insight. Considering momentum eigen-states it follows that states with opposite sign interacts differently with the geometry since $\vec{W}\cdot\nabla_P \Rightarrow -\vec{W}\cdot\nabla_P$ when $\nabla_P \Rightarrow -\nabla_P$. Hence, a non-trivial geometry on S might lift a degeneracy which is present on the plane provided $\vec{W} \neq 0$. An analogy to the motion of an electrically charged particle moving on the plane with a magnetic field piercing the plane is immediate. Hence, the class of surfaces defined by $\vec{W} \neq 0$ is thus interesting to consider further in order to build a general physical picture of quantum dynamics on surfaces. The tentative framework presented here might furthermore also help shed light on a fundamental problem connected with the understanding of quantum mechanics on surfaces following Dirac's quantization prescription (2). It is well known that the Dirac quantization

scheme does not produce an unique expression for the induced quantum geometric potential V_S (15–17). It is claimed that this result is not related to improper choice of coordinates, but emerges solely due to operator ordering issues (15). Interestingly, considering a plane P it follows that no geometric potential of the kind stemming from the dimensional reduction approach will get induced. The coupling to the external potential V_P should not imply any operator ordering issues of the kind reported in (15). An unique theory should consequently ensue. However, the quantum theory on P is conformally related to a quantum theory on a certain surface S where the very same quantization procedure will not give rise to a unique theory. This apparent 'contradiction' between the '$P-$' and the '$S-$pictures' of the theory might thus hold a key for resolving the disturbing discrepancy in the quantum formulation of classical mechanics on surfaces following from Dirac's prescription.

5. Supersymmetric quantum mechanics

Let us turn to one-dimensional structures living in ordinary three-dimensional Euclidean space. On such strutures the effective theory stemming from the dimensional reduction approach is given by Eq. (13). Ideas from supersymmetry was incorporated into one-dimensional quantum mechanics by E. Witten in (18). This approach leads to a natural notion of isospectral deformations, deformations of the potential in the Schrödinger equation such that that the energy spectrum is identically preserved. It would be interesting to apply this approach to our subject. This should mean that we can set out with a particular linear configuration in space which is described by the local curvature of the structure $\kappa(x)$ (x denotes some coordinate along the structure). From this it should then be possible to generate another potential appearing in the Schrödinger equation which in our context must be related to another curvature configuration $\hat{\kappa}(x)$; to another linear structure in space, i.e. This line of approach, as the one in the previous section, has not been pursued in the previous literature. It seems to represent a promising approach in the work of getting a deeper understanding of the relation between quantum physics and geometry. Let us initiate this study by some relatively straightforward considerations. We will assume that the reader has a basic understanding of supersymmetric quantum mechanics. The recent book (19) represents a nice introduction to the subject. We will follow the notation in that book in the following.

Isospectral deformations in supersymmetric quantum mechanics are connected to the problem of generating another superpotential $\hat{W}(x)$, which gives rise to a new partner potential $\hat{V}_+(x)$, from a given superpotential $W(x)$ and partner potential $V_+(x)$ such that

$$\hat{V}_+(x) = \hat{W}^2 + \hat{W}' = V_+(x) = W^2 + W' \,. \tag{63}$$

$'$ indicates differentiation with respect to the space variable. From the knowledge of $\hat{W}(x)$ a new physical potential $\hat{V}_-(x)$ can be constructed

$$\hat{V}_-(x) = \hat{W}^2 - \hat{W}' \,, \tag{64}$$

which give rise to the same spectrum as the initial one generated by $V_-(x)$. In the literature the only type of deformation that has been studied so far has the form (19)

$$W(x) \to \hat{W}(x) = W(x) + f(x) \,. \tag{65}$$

$f(x)$ is some function to be determined from the condition Eq. (63). Finding the explicit expression for $f(x)$ is straightforward. Inserting Eq. (65) into Eq. (63) we get

$$\frac{d}{dx}f(x) + 2W(x)f(x) + f^2(x) = 0\,. \tag{66}$$

This is the Riccati equation. Further making the substitution $f(x) = 1/y(x)$ we get

$$\frac{d}{dx}y(x) - 2W(x)y(x) = 1\,. \tag{67}$$

This equation is easily integrated and we find

$$f(x;\lambda) = \frac{e^{-\int 2W(x)dx}}{\lambda - \int e^{-\int 2W(x)dx}dx}\,. \tag{68}$$

λ is an integration constant. We note that letting $\lambda \to \infty \Rightarrow f(x;\lambda) \to 0$ results in the identity deformation $\hat{W}(x) = W(x)$. The result Eq. (68) has been taken in the literature as an expression for the most general deformation of $W(x)$ stemming from Eq. (63) when a deformation scheme of the kind in Eq. (65) is employed (19). Let us now identify the physical potential V_- with the potential in Eq. (13)

$$V_-(x) = -\frac{\hbar^2}{8m}\kappa^2(x)\,. \tag{69}$$

This gives the following equation for the superpotential

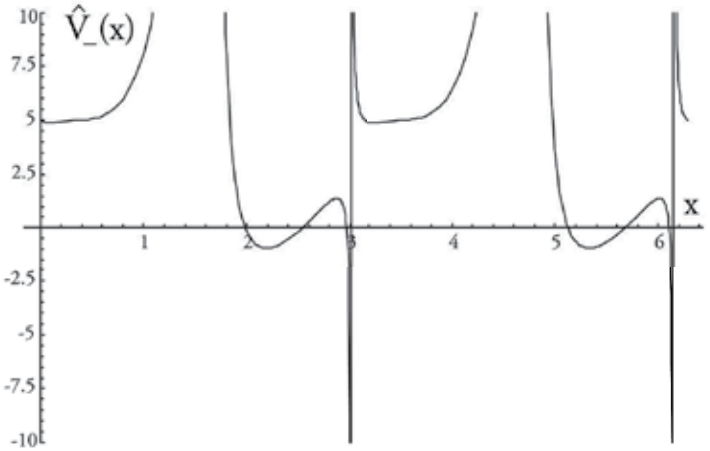


Fig. 4. The effecetive potential $\hat{V}_-(x)$ on the circle with $\lambda = 0.5$.

$$-\frac{dW}{dx} + W^2 + \kappa^2 = 0\,. \tag{70}$$

This is also a Riccati equation. It can be connected to ordinary second order differential equations by the substitution

$$W = \frac{1}{U(x)} \frac{dU(x)}{dx} \,. \tag{71}$$

In terms of $U(x)$ Eq.(70) reads

$$\frac{d^2U}{dx^2} + \kappa^2(x)U = 0 \,. \tag{72}$$

Let us consider the circle which is well understood in quantum theory. We can set $\kappa^2 = 1$ without loss of generality. We then easily get

$$U = A_0 \sin(\omega x) + B_0 \cos(\omega x) \tag{73}$$

where ω is some constant. Let us for simplicity set $A_0 = 0$ and $\omega = 1$. Then we get

$$W = -\tan(x) \,. \tag{74}$$

We are then in position to compute the deformed physical potential $\hat{V}_-(x)$. We find that

$$\begin{aligned} \hat{V}_-(x) &= \hat{W}^2 - \hat{W}' = \\ &= (-\tan(x) + \frac{1}{\lambda \cos^2(x) - 1})^2 - (-\frac{1}{\cos^2(x)} + \frac{2\cos(x)\sin(x)}{(\lambda \cos^2(x) - 1)^2}) \,. \end{aligned} \tag{75}$$

This potential is graphed in Fig. (4) with $\lambda = 0.5$. The potential exhibits singularities. It is not strictly non-positive but can also take positive values. When this happens the corresponding curvature $\hat{\kappa}$ appears to be imaginary. This last feature is clearly unphysical. However, we have only probed a particular solution of a vast solution space and physically acceptable deformations might very well exist. A more systematic study of the circle is left for future research.

Another configuration which is natural to study, and which also is easily tractable by analytical means, is a straight line which has a bent part somewhere along it. A model for such a structure is captured by a curvature function given by

$$\kappa^2 = 1 - \tanh^2(x) \,. \tag{76}$$

κ^2 is graphed in Fig. (5). This gives rise to U-functions given by

$$U(x) = C_1 P_{\frac{1}{2}(\sqrt{5}-1)}(\tanh(x)) + C_2 Q_{\frac{1}{2}(\sqrt{5}-1)}(\tanh(x)) \tag{77}$$

where the P- and Q-functions denote the Legendre functions. The deformed superpotential can formally be written

$$\hat{W} = \frac{1}{U(x)} \frac{dU(x)}{dx} + \frac{1}{\lambda U^2 - 1} \,. \tag{78}$$

From this expression we can deduce an infinite family of new linear structures as we did above. However, the study of this deformation is also left for the future. It is hoped that the relative easy one apparently can derive new physically realizable linear structures from known ones as demonstrated here will inspire more studies along these lines.

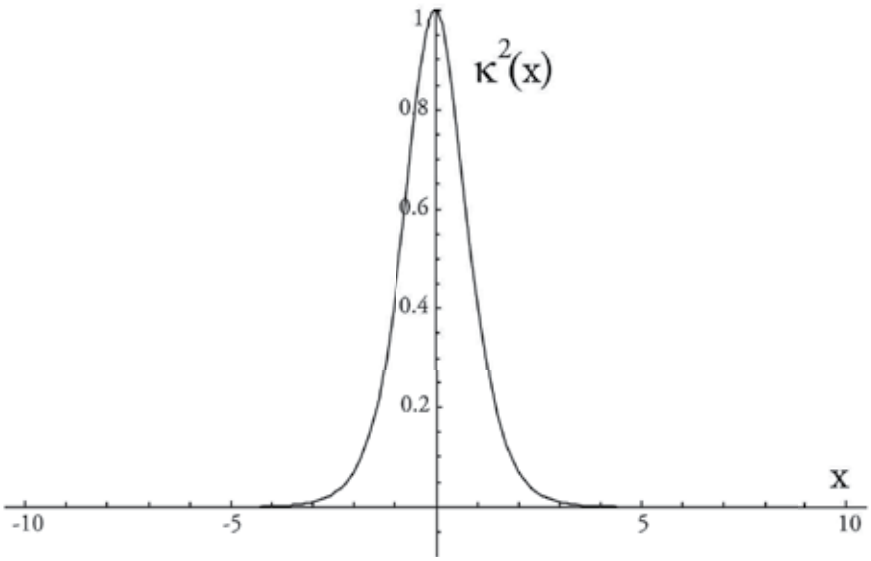


Fig. 5. The curvature $\kappa^2 = 1 - \tanh^2(x)$.

6. The Dirac equation

Let us turn to the Dirac equation. We start our discussion by considering the Dirac equation in a (3+1)-dimensional curved space-time. We apply a curved space-time formalism from the very outset in order to clearly exhibit general computational details which are not readily available elsewhere, but which are of fundamental importance in the discussion of the Dirac equation on general surfaces in three dimensional Euclidean space. Due to the spinor structure the Dirac equation is most easily formulated relative to a local vier-bein field. The massive equation relative to a given viel-bein field ${e_A}^a$ (capital latin letters denote viel-bein components while lower capital latin letters denote coordinate indices) is in general given by (1)

$$(\gamma^A(\partial_A - \Gamma_A) + m)\Psi \equiv (\gamma^A D_A + m)\Psi = 0\,. \tag{79}$$

Ψ represents the Dirac spinor field, and m represents the rest mass of the particle. The γ^A-matrices obey locally the Clifford algebra $\{\gamma^A, \gamma^B\} = 2\eta^{AB}$ with $\eta_{AB} = \mathrm{diag}(-1,1,1,1) = {e_A}^a e_{Ba}$. The matrix valued spin connection Γ_A is explicitly given by

$$\Gamma_i = -\frac{1}{4}\gamma^A\gamma^B {e_A}^j(\partial_i e_{Bj} - {e_B}^k\Gamma_{kji})\,. \tag{80}$$

$\Gamma_{\lambda\nu A}$ represents the Christoffel symbols of the second kind. The current $j^A \equiv \overline{\Psi}\gamma^A\Psi$, where $\overline{\Psi} = \Psi^\dagger\gamma^0$ ($\dagger$ signifies Hermitian conjugation), is covariantly conserved $\nabla_A j^A = 0$. The Dirac field is normalized with respect to the canonical integration measure. When $e^i_A\Gamma_i = 0$ we define $D_a = \nabla_a$. We are primarily interested in the second order form of the Dirac equation in order to make a direct comparison with Schrödinger theory on a two-dimensional surface S, or on a linear structure. However, let us begin with the first order formulation of the Dirac equation.

We may compute the Dirac equation in our coordinate system X relative to an orthonormal tetrade frame so that

$$\underline{\omega}^0 = \underline{d}t \tag{81}$$

$$\underline{\omega}^1 = G_{11}^{1/2}\underline{d}x^1 \tag{82}$$

$$\underline{\omega}^2 = G_{22}^{1/2}\underline{d}x^2 \tag{83}$$

$$\underline{\omega}^3 = \underline{d}x^3\,. \tag{84}$$

The beins may therefore be read off from the knowledge of the metric. Hence, under a change of coordinates from x to X probability conservation requires that the Dirac spinor $\psi(X) \to \psi(X)\xi^{-1/2}$ as for the Schrödinger field. We assume for simplicity that the surface has vanishing intrinsic curvature. We parametrize the surface by Cartesian coordinates. It is then easy to show that

$$\Gamma_3 = 0 \tag{85}$$

$$\Gamma_i = \frac{i}{2}\epsilon_{ij}K_{ii}\sigma_j \cdot I \tag{86}$$

where $i = 1, 2$ is summed and I is the identity for **4** spinor indices. We are now in a position to construct the covariant derivative explicitly. Since $\Gamma_3 = 0$ one sees that this is unnecessary however; there will be no contribution from the spin connection to the geometrical terms. We may therefore proceed to separate the Dirac equation into parallel and perpendicular parts. We have

$$(\gamma^A D_A + m)(\psi(X)\xi^{-1/2}) = 0 \tag{87}$$

where $A = 0, ..., 3$. Pulling ξ through the derivative we have

$$(\gamma^A D_A + \frac{1}{2}\gamma^3 \mathrm{Tr}K + m + O(x^3)^3)\psi(X) = 0\,. \tag{88}$$

We shall assume that the equation is separable and write

$$\psi(x) = \phi_\perp(x^3)\phi_{||}(x^1, x^2, t)\hat{\psi} \tag{89}$$

where $\hat{\psi}$ is a constant, four-component spinor. For later convenience we shall also define $\psi_\perp = \phi_\perp\hat{\psi}$, $\psi_{||} = \phi_{||}\hat{\psi}$, and $A^0 = A^0_\perp(x^3) + A^0_{||}(x^1, x^2, t)$ and $m = m + \Delta m(x^3)$. We then easily see that

$$\phi_{||}^{-1}(\gamma^0 D_0^{||} + \gamma^i D_i + \frac{1}{2}\gamma^3 \mathrm{Tr}K + m)\phi_{||}\hat{\psi} =$$

$$-\phi_\perp^{-1}(\gamma^0(-ieA_0^\perp) + \gamma^3\partial_3 + \Delta m(x^3) + ...O(x^3))\phi_\perp\hat{\psi}\,. \tag{90}$$

The usual argument for the separation of these equations is that the left-hand side is only a function of x^1, x^2, t while the right-hand side is only a function of x^3; thus both sides must be equal to some separation constant k. However, this is not quite correct owing to the terms $...O(x^3)^3$ which contain the curvature K_{ij}, which is a function of x^1, x^2. However, we are

assuming that the excursion into x^3 are small owing to the physics of the problem; thus we shall for the time being adopt the standard procedure and neglect these higher-order terms. To eliminate x^3 from the problem we identify $A_0^{\perp}$ with a squeezing potential $A_0^{\perp} = \frac{\lambda}{e}x^3$, so that nontrivial solutions, given by the vanishing of the determinant of the operator on the right-hand side of Eq.(90), are characterized by

$$\phi_{\perp}(x^3) \sim exp(-(\frac{\lambda}{2}(x^3)^2 + kx^3) - \int dx^3 m(x^3)) \,. \tag{91}$$

The first term in the exponential falls off like a Gaussian function, i.e., a δ function in the limit of infinite λ. The presence of a nonzero separation constant spoils this falloff by introducing a straightforward exponential decay with characteristic length k^{-1}. The mass parameter $m(x^3)$, which is initially at least linear in x^3, will also give rise to a Gaussian falloff. Thus increasing the coupling of this parameter will also have the desired effect of squeezing the system. The residual Dirac equation in the surface S is thus

$$(\gamma^A D_A + \frac{1}{2}\gamma^3 \mathrm{Tr}K + (m-k))\psi_{||} = 0 \,. \tag{92}$$

We have managed to derive a theory in a surface but with a number of assumptions and approximations. Let us next turn to the second order form of the Dirac equation.
We derive the second order form of the Dirac equation by 'squaring' the first order Dirac equation (20)

$$(\gamma^B D_B - m)(\gamma^A D_A + m)\Psi = 0 \tag{93}$$

$$\Rightarrow \; (\gamma^B \gamma^A D_B D_A - m^2)\Psi = 0 \,. \tag{94}$$

Utilizing the algebraic identity

$$\gamma^B \gamma^A = \frac{1}{2}(\{\gamma^B, \gamma^A\} + [\gamma^B, \gamma^A]) \,, \tag{95}$$

and the fundamental Clifford algebra, we can write the second order Dirac equation as

$$((\eta^{BA} + \frac{1}{2}[\gamma^B, \gamma^A])D_B D_A - m^2)\Psi = 0 \,. \tag{96}$$

Note that it is not possible to extract a term corresponding to the antisymmetric part above in a purely bosonic scalar theory like the Schrödinger theory. The antisymmetric part can be written as

$$\frac{1}{2}[\gamma^B, \gamma^A]D_B D_A = \frac{1}{4}[\gamma^B, \gamma^A][D_B, D_A] \,. \tag{97}$$

The commutator between the connection components is per definition proportional to the components of the Riemannian curvature tensor R_{ABCD}

$$[D_B, D_A] = \frac{1}{8}R_{BACD}[\gamma^C, \gamma^D] \,. \tag{98}$$

Relative to the local viel-bein it is furthermore always straightforward to locally decompose the kinetic operator $D^2 = \eta^{AB}D_B D_A$ into a tangential surface component $D_{||}^2$, an normal

component $D_\perp^2$ as well as a time component. Since

$$\frac{1}{8}[\gamma^B,\gamma^A]R_{BACD}[\gamma^C,\gamma^D] = \frac{1}{2}R_{BACD}\gamma^A\gamma^B\gamma^C\gamma^D \tag{99}$$

the squared Dirac equation can thus be written as

$$(-D_t^2 + D_{||}^2 + D_\perp^2 + \frac{1}{8}R_{ABCD}\gamma^A\gamma^B\gamma^C\gamma^D - m^2)\Psi = 0\,. \tag{100}$$

In the dimensional reduction approach (6) one introduces a set of coordinates adapted to the lower dimensional structure S such that the normal field to the surface is always parallel to the tangent vector field to one of the coordinates. $D_\perp^2$ will act in this direction. If one assumes that the dynamic equation in question is separable in parts depending on the coordinates in the surface and the perpendicular coordinate respectively one derives a theory which is living solely in the surface. The corresponding result in Schrödinger theory exhibits a scalar potential which depends on the Gaussian and mean curvatures.
Since we are considering a theory which is living in a space with vanishing intrinsic curvature it follows that Eq. (100) reduces to Schrödinger theory by triviality if we neglect any spin-orbit interactions stemming from the presence of the position dependent spin connection. This conclusion stands out in sharp contrast to the conclusions drawn from our elaborations of the first order form of the Dirac equation. Does this conclusion hold if we follow the intrinsically (2+1)-dimensional approach in which one defines the theory in the surface without taking the embedding space into account? This approach is employed in most current treatments of graphene, e.g. Let us explore this issue.
The standard theory for the charge carriers in graphene in the intrinsically (2+1)-dimensional approach is formulated in a three dimensional Minkowski space with the dynamic equation equal the standard three dimensional Dirac equation. We choose the $\mathbf{4}\times\mathbf{4}$-representation of the γ^A-matrices because the alternative $\mathbf{2}\times\mathbf{2}$-representation, which is available in (2+1)-dimensions, breaks parity invariance (8). In order to adapt this formulation to curved space we simply make the theory generally covariant in the usual fashion. Then the second order formulation is again Eq. (100) but with no $D_\perp^2$-term. The resulting theory will be valid at low energies and in principle on large molecules. In the case of graphene it is well known that a $U(1)$-gauge field which couples to the Dirac spinor is induced on curved surfaces due to the intrinsic curvature (se (21) e.g.). We neglect this field in the following. However, we cannot discard the intrinsic curvature tensor contribution as we could in the dimensional reduction approach. Working in (2+1)-dimensions and taking the symmetries of the curvature tensor into account it follows from some algebraic manipulations that

$$R_{ABCD}\gamma^A\gamma^B\gamma^C\gamma^D = -2R\,, \tag{101}$$

where R is the Ricci curvature scalar in the static surface. In two-dimensional surfaces the Ricci scalar equals twice the Gaussian extrinsic curvature $R = 2K$. Clearly, this Ricci contribution will add to the effective geometrically induced potential in a surface formed from graphene. Inserting the relation Eq. (101) into Eq. (100) we rederive the classical Lichnerowicz formula for the Dirac operator (22).
Let us make a direct comparison with Schrödinger theory by looking at the low energy limit of the intrinsic massive Dirac theory. We neglect all the spin-connection terms. Without these

terms the Dirac equation reduces to

$$(-\partial_t^2 + \nabla_{||}^2 - \frac{1}{2}K - m^2)\psi = 0\,, \tag{102}$$

where ψ is assumed to be a definite spin-state, and $\nabla_{||}$ is the usual covariant derivative acting on a scalar function. We assume energy-eigenstates and denote the total energy of the particle by E. The equation above then takes the following form to first order in a standard $1/m$-expansion (reinstating $\hbar$)

$$E_c\psi \equiv (E-m)\psi = (-\frac{\hbar^2}{2m}\nabla_{||}^2 + \frac{\hbar^2}{2m}K)\psi\,. \tag{103}$$

This is the Schrödinger equation corresponding to the second order Dirac equation, and E_c is the classical energy measure. An effective potential V_D has emerged. It is given by (reinstating $\hbar$)

$$V_D = \frac{\hbar^2}{2m}K\,. \tag{104}$$

Comparing V_D and V_S we see a huge formal difference. Schrödinger theory predicts a negative, or an attractive, potential on any curved surface. However, the sign of K is not restricted. Hence, contrary to Schrödinger theory, intrinsic second order Dirac theory predicts both attractive as well as repulsive geometrically induced quantum potentials depending on the surface in question.

In (23) it was pointed out that on a carbon nanotube the *winding* states in the angular direction will give rise to an effective mass. It is straightforward to see how this will work in our case. Consider first the massless Dirac theory with $m = 0$. This theory does not have a natural low energy limit and the relation between the Dirac and the Schrödinger approaches becomes highly problematic in general. However, assume a perfectly *straight* carbon nanotube with an intrinsic geometry given by $ds^2 = dz^2 + R^2d\theta^2$, where $-\infty \leq z \leq +\infty$ and $\theta \in [0, 2\pi\rangle$. Assuming momentum eigen-states in the angular direction $\psi \sim e^{in\theta}$ (where n represents integers) Eq. (102) (with $m = 0$) becomes

$$(-\partial_t^2 + \partial_z^2 - \frac{1}{2}K - (\frac{n}{R})^2)\psi = 0\ ,\ |n| \in 0, 1, 2, 3, \ldots\,. \tag{105}$$

Clearly, the last term in Eq. (105) can be identified as a mass parameter so that

$$m = \frac{|n|}{R}\,. \tag{106}$$

Since R is on the nanoscale, $1/m$ so defined ($n \neq 0$) is an effective parameter which can be used in the series expansion leading to Eq. (103). Of course, this scheme can also be employed in the dimensional reduction scheme. The effective theory then becomes

$$E_c\psi = (-\frac{1}{2m}\partial_z^2 + \frac{1}{2mR^2}(n^2-1))\psi\,. \tag{107}$$

Here m is the ordinary rest mass of the particle. These results can be utilized as tools to discriminate between the dimensional reduction scheme and the intrinsic Dirac approach. Let us pursue this topic a little bit further.

Assume that the ends of a straight carbon nanotube is welded to two metal slabs (20). Let us for simplicity neglect the physics at the welding areas such that we can assume a free Dirac theory in the metal and the welding areas. Since $K = 0$ on the tube the Dirac theory implies that a current with all the electrons in the lowest winding state $n = 0$ can be transmitted from one of the metal pieces and through the tube without backscattering (reflection). This is not the case according to the dimensional reduction approach. The $n = 0$ states will form bound states implying a net reduction in the net current in the tube. Clearly, when $n = \pm 1$ the dimensional reduction approach also implies a vanishing net potential. However, the dominant transmission channel should correspond to the lowest lying winding mode $n = 0$ due to the relatively large mass scale in the system. Hence, to leading order the two formulations will imply different currents through a straight carbon nanotube connecting two metal slabs. This argument is a very simple one, but the conclusion is robust. In particular, even though one cannot neglect the physics in the welding areas in general (24) the relative difference between the dimensional reduction and the intrinsic Dirac approaches can nevertheless be deduced since the effective potential on the tube in the dimensional reduction picture scales as $1/R^2$. Hence, making the very reasonable assumption that the physics in the welding areas is the same for tubes with differing radii the presence of a geometry induced potential can be extracted. Clearly, for a tube with a large radius $R \to \infty$ the effective potential will vanish while it blows up as $R \to 0$. The intrinsic Dirac theory will not exhibit a similar scaling behavior in any state. Furthermore, note that the effective potential in Eq. (107) changes sign at $n^2 = 1$. Hence, for $|n| \geq 2$ the effective potential becomes repulsive. If real, these differences should be readily observable experimentally.

Of course, the discussion above is not confined to graphene. It should be valid for any surface. As another application let us briefly consider the rolled up nanotubes (RUNTs) discussed in (25). RUNTs are generally made of bi- or multilayer thin films of various materials. A flat thin film might, due to the relaxation of the elastic stresses, curl up and form a RUNT described by an Archimedes spiral. Physically it has the same symmetries as the straight carbon nanotubes considered above, but instead of curling up to form a closed cylinder the film curls to form a structure similar to a book scroll. In (25) the charge carriers were described as an exact two dimensional electron gas within the dimensional reduction framework. Clearly, the Gaussian curvature on the Archimedes spiral vanishes, but not the mean curvature. It was shown in (25) that the non-vanishing mean curvature will result in a number of atomic-like bound states in the spiral surface. The number of bound states equals the number of windings of the spiral. How does this relate to an an intrinsic Dirac theory description? In thin films we are not dealing with massless theory so that $m \neq 0$ in Eq. (102). We are therefore not dependent upon either a doped material, nor a periodic structure in order to deduce a low energy limit as on the fullerenes. We can thus employ Eq. (103) directly. Hence, no geometric potential will appear on the RUNT, and consequently no bound states, since $K = 0$. Just as with the straight carbon nanotube above, measuring for the existence or non-existence of bound states on rolled up nanotubes could prove a veritable tool for discriminating between the different descriptions of quantum mechanics on surfaces.

Even though much work have been done on the quantum mechanics on many different surfaces a complete analytical analysis of quantum mechanics on the simple torus is still missing, even though this surface is an important one. Next we will therefore provide some basic equations as a start for such an analysis.

We will assume the following parametrization of the torus

$$x = (c + a\cos v)\cos u\,, \tag{108}$$

$$y = (c + a\cos v)\sin u\,, \tag{109}$$

$$z = a\sin v\,, \tag{110}$$

where $u, v \in [0, 2\pi\rangle$. a signifies the 'small' radius and c the 'big' radius of the torus such that $c > a$. The coordinates are depicted in Fig. (6). The geometry on the torus can then be expressed as

$$ds^2 = (c + a\cos v)^2 du^2 + a^2 dv^2 \equiv A^2 du^2 + a^2 dv^2\,. \tag{111}$$

The extrinsic curvature is explicitly given by

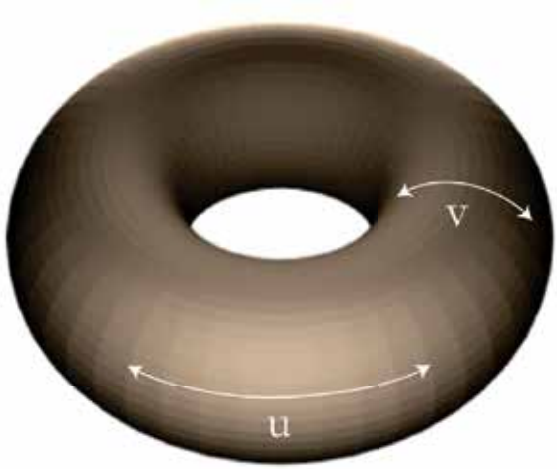


Fig. 6. The torus. The letters denote the angular coordinates.

$$K = \frac{\cos v}{a(c + a\cos v)}\,. \tag{112}$$

Let us consider stationary quantum energy eigenstates on the torus. Due to the non-dependence of the u-coordinate in the metric we also consider momentum eigenstates in the u-direction. We call these states winding states when this momentum is non-zero. Hence, we assume that the wave-function is on the form

$$\Phi \sim e^{-iEt} e^{imu} A^{-1/2} W(v)\ ,\ m \in \{0, \pm 1, \pm 2, ...\} \tag{113}$$

where W is a function to be determined by the wave-equation. With this ansatz the wave-equation reduces, quite remarkably, to the simple form

$$(-\frac{\partial^2}{\partial v^2} + V_e(v) - E^2)W(v) = 0\,, \tag{114}$$

where the effective potential V_e is given by

$$V_e(v) = -(\frac{1}{4}(\partial_v A)^2 + a^2 m^2)A^{-2} + \frac{1}{2}A^{-1}\partial_v^2 A$$
$$= \frac{-a^2}{(c + a\cos v)^2}((\frac{\sin v}{2})^2 - m^2) - \frac{a\cos v}{2(c + a\cos v)} - \frac{\cos v}{2a(c + a\cos v)}\,. \qquad (115)$$

A plot of the effective potential is provided in Fig. (7). If the contribution from the extrinsic curvature is discarded in the expression for the effective potential it still has the same qualitative form as in Fig. (7).

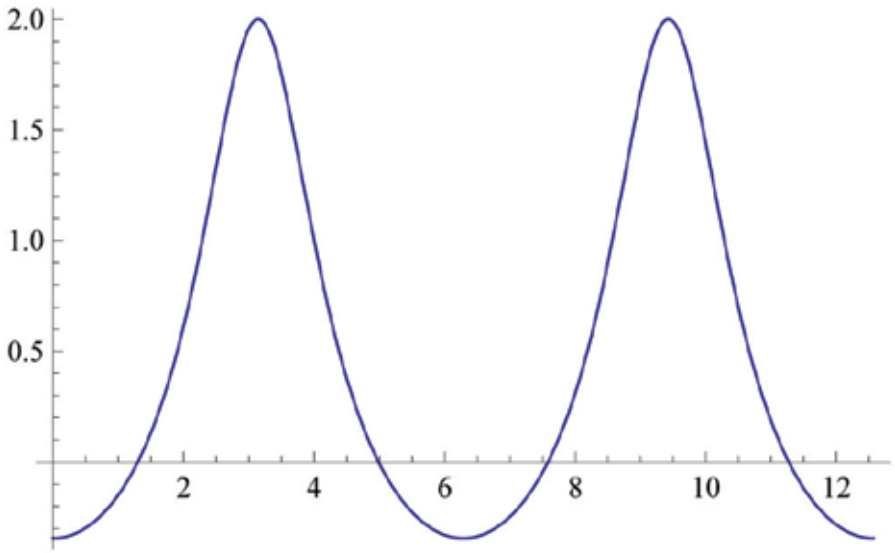


Fig. 7. The effective potential V_{e} with $m = \pm 1$ in the interval $u \in [0, 4\pi]$. It has the same qualitative features irrespective of the parameter values. Negative energy states seemingly exist only for states with $m \in \{0, \pm 1, \pm 2\}$.

Let us finally make a brief comment on Dirac theory on linear structures. Clearly, neglecting the spin-connection the extrinsically defined theory will reproduce Schrödinger theory as in the case of surfaces. Considering an intrinsically defined theory the second order formulation in Eq. (100) is again valid. However, now the Riemann curvature tensor is trivially identically zero when we consider static linear structures since the geometry will be time independent. This means that intrinsically defined fermions on static linear structures will not experience a spin induced potential. However, note that if the structure exhibits a time dependent geometry this is no longer true. This is easily seen already on the algebraic level since now the right hand side of Eq. (101) is non-zero with changed overall sign

$$R_{ABCD}\gamma^A\gamma^B\gamma^C\gamma^D = 2R\,. \qquad (116)$$

This sign flip is induced by the Clifford algebra. Explorations of the implications of this change of sign is left for future work.

7. Conclusion

In this brief account of quantum mechanics on surfaces we introduced the reader to the dimensional reduction approach which is currently the most widely accepted method to construct quantum theories on surfaces and linear structures embedded in ordinary three-dimensional Euclidean space. We generalized this framework to accommodate Dirac theory. We studied quantum theory on explicit geometries; the catenoid in Schrödinger theory and the cylinder in Dirac theory. In addition we presented novel approaches to the study of the interplay between quantum physics and geometry in the form of a conformal approach to defining quantum theory on a surface and by applying ideas from supersymmetric quantum mechanics. Both of these approaches need further elaboration and refinements, but it is the hope of this author that this material might inspire other workers in the field to pursue these new avenues.

This account of quantum mechanics on surfaces also makes an effort to put our subject at the center stage of current physics research. Quantum theory on surfaces might in particular be important in order to get a complete understanding of graphene. We have shown that there is a major discrepancy between the formulation of Dirac theory on surfaces depending on whether one employs an extrinsic or an intrinsic approach. One immediate consequence is that the intrinsic theory implies a new effective scalar potential proportional to the Gaussian curvature in the surface, while the extrinsic approach implies the usual potential stemming from the dimensional reduction approach (6). The new potential emerges due to the Clifford algebra and will thus not be present in any scalar theory defined on surfaces. We also looked at the low energy limit of the intrinsic Dirac theory and derived an effective potential V_D which corresponds to the effective potential stemming from the Schrödinger (or the Dirac) theory in the extrinsic approach (V_S). Clearly, while V_S is strictly non-positive the potential stemming from the low energy intrinsic Dirac theory V_D can carry any sign depending on the surface in question. Graphene is described by the massless Dirac equation near the Dirac points. Hence, near these points the charge carriers might respond very differently to the graphene surface geometry than one would expect from Schrödinger theory. This insight might be of great importance in the modeling of the charge carriers on graphene with consequent experimental and technical implications. We considered one particular technical implication for a system composed of a carbon nanotube bridging two metal slabs. We also considered a particular rolled up nanotube in the same vain with qualitatively the same conclusions as with the carbon tube. We emphasize that we have ignored the effect of the spin-connection. This will induce spin-orbit couplings which will add motion dependent potentials. The effect of these can be very pronounced (26). Clearly, much more work needs to be done in this field, but it is work which apparently has potential to further advance recent discoveries in physics.

8. References

[1] N.D. Birrell, P.C.W. Davies *Quantum fields in curved space* Cambridge University Press (1982).
[2] P.A.M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva University, New York, 1964).
[3] S.W. Hawking, Comm. Math. Phys. 43 199 (1975).
[4] H.B.G. Casimir, Proc. kon. Ned. Akad. wet. 51 793 (1948).
[5] H. Jensen, H. Koppe, Ann. Phys. 63 586 (1971)
[6] R.C.T. da Costa, Phys. Rev. A 23, 1982, (1981).

[7] R.C.T. da Costa, Phys. Rev. A 25, 2893 (1982).
[8] M. Burgess, B. Jensen, Phys. Rev. A 48 1861 (1993).
[9] R. Dandoloff, A. Saxena, B. Jensen, Phys. Rev. A XX, ppp (2010).
[10] Y.N. Joglekar and A. Saxena, Phys. Rev. B 80, 153405 (2009).
[11] M. Morris and K. Thorne, Am. J. Phys. 56, 395 (1988).
[12] R. Dandoloff, Phys. Lett. A 373, 2667 (2009).
[13] J. Lekner, Am.J.Phys., 75, 1151, (2007).
[14] V. Bargmann, Rev. Mod. Phys. 21, 488 (1949); R.E. Crandall, J. Phys. A 16, 3005 (1983).
[15] M. Ikegami, Y. Nagaoka, S. Takagi, T. Tanzawa, Prog. Theor. Phys. 88 229 (1992).
[16] N. Ogawa, Prog. Theor. Phys. 87 513 (1992).
[17] N.M.B. Peres, J.N.B. Rodrigues, T. Stauber, J.M.B. Lopes dos Santos, J.Phys.: Cond. Matter 21 344202 (2009).
[18] E. Witten, Nucl. Phys. B188 513 (1981).
[19] A. Gangopadhyaya, J.V. Mallow, C. Rasinariu, *Supersymmetric quantum mechanics* World Scientific (2011).
[20] B. Jensen, R. Dandoloff, Phys. Lett. A 375 448 (2011).
[21] J.K. Pachos, Contemporary Physics Vol. 00, No.00, 1 (2008), arXiv:0812.1116v1
[22] P. Lichnerowicz *Theorie globale des connexions et des groupes d'holonomie* Dunod, Paris (1955).
[23] A.D. Alhaidari et al. arXiv: 1010.3437.
[24] Ya. M. Blanter, I. Martin, Phys. Rev. B76 155433 (2007).
[25] C. Ortix, J. van den Brink, Phys. Rev. B81 165419 (2010).
[26] B. Jensen, Phys. Rev. A80 02210 (2009).

9

Quantum Statistics and Coherent Access Hypothesis

Norton G. de Almeida
Universidade Federal de Goiás, Campus Samambaia, Goiânia-GO
Brazil

1. Introduction

The properties of quantum systems at equilibrium are mainly displayed by the very nature of the particles involved. As for example, distinguishable particles behave in a quite different manner of indistinguishable bosons, which in turns behave very different from the also indistinguishable fermions. Among the standard methods of statistical mechanics to handle with quantum systems at equilibrium is included the way of counting the accessible states to a given particle (1). Thus, the thermodynamical probability Ω comprising the set of possibilities for a system of N particles to populate the available states will depend on the type of particle being considered, as shown in Table 1. The method for calculating Ω consists in supposing the existence of a number of energy levels, labeled by and index j, each level having energy ε_j and containing a total of n_j particles. Besides, it is assumed that each level contains g_j distinct sublevels, all of which having the same energy. The value of g_j associated with level j is called the "degeneracy" of that energy level. Any number of bosons can occupy the same sublevel, while the Pauli exclusion principle states that only one fermion can occupy any one of the sublevels. Classical behavior emerges when the density is relatively low, which is more common at high temperatures; in such cases, both fermions and bosons behave as if they were classical particles. After identifying what kind of particles we are taking into account, we use counting methods and combinatorial analysis to calculate the expected number n_j of those particles in level j with energy ε_j. This is done by first calculating the number Ω_j of ways of distributing particles among the sublevels g_j of an energy level j and taking the product of the ways that each individual energy level can be populated to find the so called thermodynamical probability Ω. Once we have Ω, it is a standard task to find the statistical distribution for n_j.

In this chapter, the counting method is extended to handle with equilibrium as well as nonequilibrium quantum systems, in order to obtain the new thermodynamical probability emerging from these situations. Also, systems only slightly out of equilibrium are considered, such that a well defined temperature can be assigned to them, and the new results are pointed, which take into account eventual experimentally detectable changes in their statistical properties.

Another method to handle with equilibrium systems is to take advantage of the Boltzmann factor, which establishes that, given two levels a and b whose energies are E_a and E_b,

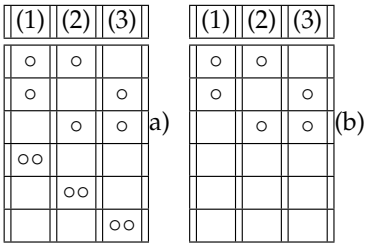


Table 1. Configurations of $M = 3$ accessible states for a) $N = 2$ indistinguishable bosons, $\Omega = 6$ and b) $N = 2$ indistinguishable fermions, $\Omega = 3$. (1) denotes the first available state, (2) and (3) denote the second and third available state, respectively.

respectively, the number n_a of particles in state a with energy E_a is given by

$$n_a = p_a \exp(-\beta E_a), \tag{1}$$

where p_a is a constant indicating the occurrence of probability for state a. The following condition is thus assumed to be satisfied:

$$n_a P_{MB}(E_{a\to b}) = n_b P_{MB}(E_{b\to a}), \tag{2}$$

where $P_{MB}(E_{a\to b})$ stands for the Maxwell-Boltzmann (MB) distribution accounting for the transition rate from state a to another state b. This method will be heuristically used later to obtain a generalized statistics for both equilibrium and nonequilibrium situations concerning quantum systems.

1.1 Quons, anyons, and generalized statistics

The concept of particles with fractional statistics, called anyons, has been studied with increasing interest, now finding applications in, for example, quantum Hall effect (2) and superconductivity (3). Anyons are generally expected to be observed in two space dimensions for a system whose wave function behavior of multiparticle under the exchange of particles obeys $\Psi(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4, ...) = \exp(i\pi\alpha)\Psi(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4, ...)$, where $\{\mathbf{q}\}$ represents generalized coordinates and α is a real number defining the statistics. For bosons and fermions, $\alpha = 0$ and 1, respectively, and α is *any* real number for anyons. Since the concept of anyons was proposed (4; 5), there have been several approaches trying to understand its properties, including a q-deformed bosonic algebra (6) and certain ad hoc conjectures aiming to generalize the Bose-Einstein and Fermi-Dirac distributions (7). Without being exhaustive, it is worthwhile to mention at least a few attempts to achieve a generalized anyonic statistics. As for example, an interesting generalization of the Pauli exclusion principle can be advanced, without explicit reference to spatial dimension (8). This is remarkable, since anyons was generally accepted to exist only in two dimensions. The generalization as proposed in Ref.(8) consists on *defining* a statistical interaction $g_{\alpha\beta}$ given by

$$\Delta d_\alpha = -\sum_\beta g_{\alpha\beta}\Delta N_\beta \tag{3}$$

where d_α is the one-particle Hilbert space dimension, and ΔN_β is an allowed change of the particle number N_β at fixed size and boundary conditions. Thus, for bosons it is required

$g_{\alpha\beta} = 0$, while for fermions the exclusion principle requires $g_{\alpha\beta} = \delta_{\alpha\beta}$. Based on the thermodynamical limit, further requirements must be imposed, such as the independency of the number of particles and the rationality of $g_{\alpha\beta}$. According to the author, at a fixed particle numbers the counting state thus leads to the following size of the Hilbert space of many particle states:

$$\prod_a \frac{(d_a + N_a - 1)!}{N_a!(d_a - 1)!}, \tag{4}$$

with $d_B = d_a$ and $d_F = d_a + N_a - 1$ for the Fock space dimensions of bosons and Fermi, respectively.

We find another interesting proposal for generalizing Bose-Einstein and Fermi-Dirac statistics in (9), where the distribution function for the number of anyons in a given state j is postulated at once as

$$N_j = \frac{\gamma}{\exp\left[\beta\left(\epsilon_j - \mu\right)\right] - f(\alpha)}. \tag{5}$$

The parameter μ characterizes the particle reservoir in the same way that β characterizes the thermal reservoir, $\gamma = 2s + 1$ denotes the multiplicity of states and $f(\alpha)$ generalizes the statistics, since $f(\alpha) = 0, 1, -1$ recovers the Boltzmann, Bose-Einstein and Fermi-Dirac statistics and, for anyons, it is expected that $f(\alpha)$ can assume any real number.

Although the interesting problem of what is exactly the distribution function for anyons in and out of equilibrium remains opened, the above studies provide some clues to this question. It is my purpose, in the remainder of this chapter, to show how to deep our understanding of anyons by looking for some reasons to justify their behavior in a manner that sounds plausible.

2. Generalized thermodynamical probability

It is possible to provide a formula taking into account more general behavior than that presented by fermions and bosons, as I show in the following, although, regarding to equilibrium situations, it is not clear at the present if nature allows for such *any*ons . The

general problem we have to solve is that of obtaining the number Ω of ways that n_j balls can be distributed into g_j boxes, each box having capacity to accommodate p balls. This number is given by (10)

$$\Omega_p = \sum_{k=0}^{[n_j/(p+1)]} (-1)^k C_{g_j}^k \times C_{g_j+n_j-k(p+1)-1}^{n_j-k(p+1)}, \tag{6}$$

where $C_n^m = \frac{n!}{m!(n-m)!}$ and the brackets $[n_j/(p+1)]$ in the upper limit indicate the integer taken from $n_j/(p+1)$. As for example, if $n_j = 2$ and $p = 1$, then $[2/3] = 0$; if $n_j = 5$ and $p = 2$, then $[5/2] = 2$, and so on. This relation can also be written as

$$\Omega_p = \frac{g_j!}{(g_j - 1)!} \sum_{k=0}^{[n_j/(p+1)]} (-1)^k \frac{(g_j + n_j - 1 - k(p+1)!}{k!(g_j - k)![(n_j - k(p+1)]!}. \tag{7}$$

From Eq.(6) it is easy to verify the number of configurations displayed in Tab.I: for the case of bosons, $n_j = p = 2$, $g_j = 3$, and the summation is limited to $k = 0$, thus resulting $\Omega =$

$C_4^2 = 6$; for the case of fermions, $p = 1$, such that in the summation $k = 0, 1$,thus resulting $\Omega = \sum_{k=0}^{1}(-1)^k C_3^k \times C_{4-2k}^{2-2k} = 3$. Proceeding in a general manner, from Eq.(6) we can recover the statistics corresponding to the case $p = 1$, Fermi-Dirac (FD) distribution and $p = n_j$, Bose-Einstein (BE) distribution. For instance, taking $p = n_j$ in Eq.(6), the upper limit $[n_j/(n_j + 1)]$ implies $k = 0$, and it is straightforward to obtain the BE statistics:

$$\Omega_{BE}^{j} = C_{g_j}^{0} \times C_{g_j+n_j-1}^{n_j} = \frac{(g_j + n_j - 1)!}{n_j!(g_j - 1)!}, \tag{8}$$

from which we obtain the thermodynamical probability for all the j levels:

$$\Omega_{BE} = \prod_j \Omega_{BE}^{j} = \prod_j \frac{(n_j + g_j - 1)!}{n_j!(g_j - 1)!}. \tag{9}$$

Following the standard route (11), since the maxima of Ω_{BE} and $\ln \Omega_{BE}$ occur at the same value, we choose to maximize this latter function, which is subject to the constraints $\sum_j n_j = N$ and $\sum_j n_j \epsilon_j = E$. Using Lagrange multipliers λ and β to form the function

$$f(n_j) = \ln \Omega_{BE} + \lambda(\sum_j n_j - N) + \beta(\sum_j n_j \epsilon_j - E), \tag{10}$$

and the the Stirling formula $\ln n! \cong n \ln n - n$ for the factorials, we take the derivative with respect to n_j, setting the result to zero and solving for n_j, to find the Bose–Einstein distribution:

$$n_{jBE} = \frac{g_j}{\exp(\lambda + \beta E_j) - 1}. \tag{11}$$

From thermodynamics, we know that $\beta = 1/kT$, k being the Boltzmann constant, T the temperature, and that λ is related to the chemical potential by $\lambda = -\beta\mu$, thus being zero for a photon gas. Similarly, when we take $p = 1$ and use the closed relation

$$\sum_{k=0}^{[n_j/2]} (-1)^k \frac{(g_j + n_j - 1 - 2k)!}{k!(g_j - k)!(n_j - 2k)!} = \frac{(g_j - 1)!}{n_j!(g_j - n_j)!}, \tag{12}$$

it is straightforward to write

$$\Omega_{FD} = \prod_j \frac{g_j!}{n_j!(g_j - n_j)!}. \tag{13}$$

The same standard procedure now leads to

$$n_{jFD} = \frac{g_j}{\exp\left[\beta(E_j - \mu)\right] + 1}, \tag{14}$$

where the parameter of constraint $\lambda = -\beta\mu$ was written explicitly. Note that Eq.(6) generalizes the concept of indistinguishable particles, and it is possibly a valid starting point to study the statistics of particles whose behavior at equilibrium at temperature T is neither that of bosons nor that of fermions. However, finding a closed relation like that of Eq.(12), for Eq.(7) giving a general capacity p, which would allow one to obtaining a general formula for

the thermodynamical probability Ω_p, remains an opened problem. Of course, once we have such a closed relation, we can apply the standard methods to deduce a generalized occupation number n_p. I hope the readers can consider this problem interesting enough to pursue further. Another way to derive Eq.(11) and Eq.(14) for a gas of photons ($\mu = 0$) is the following. Let E_a and E_b the energy of state a and b, respectively, for a two state system of distinguishable particles. As discussed in the introduction of this Chapter, the maxwell-Boltzmann distribution, Eq.(1), implies that, in equilibrium,

$$\frac{n_a}{n_b} = \frac{P_{MB}(E_a)}{P_{MB}(E_b)} = \exp\left[-\beta(E_a - E_b)\right], \tag{15}$$

where n_b and n_a are the number of distinguishable particles in state b and a, respectively. Now, if we consider indistinguishable particles, as for example bosons, the "Bose-Einstein" rate indicated by $P_{BE}(E_{a\to b})$ for one particle to make a transition from one state (a) to another (b) and *vice-versa* must obey

$$P_{BE}(E_{a\to b}) = (1 + n_b)P_{MB}(E_{a\to b}), \tag{16}$$

$$P_{BE}(E_{b\to a}) = (1 + n_a)P_{MB}(E_{b\to a}), \tag{17}$$

with n_b, $n_a = 0, 1, 2...$ We can assign the following meaning to Eq.(16) and Eq.(17): The transition rate governed by BE statistics, being proportional to the MB distribution, makes the probability to one bosonic particle goes from state a to state b (from state b to state a) to increase with the occupation number of state b (a). As the equilibrium is assumed, we can write

$$n_a P_{BE}(E_{a\to b}) = n_b P_{BE}(E_{b\to a}). \tag{18}$$

From Eqs(16)-(18), we find

$$\frac{n_a}{(1 + n_a)} \exp\left(-\beta E_a\right) = \frac{n_b}{(1 + n_b)} \exp\left(-\beta E_b\right). \tag{19}$$

If we now put $\frac{n_b}{(1+n_b)} \exp\left(-\beta E_b\right) = F(\beta)$ and solve to n_a we find, after changing the subindex a to BE,

$$n_{BE} = \frac{1}{F(\beta)\exp\left(\beta E\right) - 1}, \tag{20}$$

which has the same form as Eq.(11).
Proceeding in the same way, it is straightforward to rederive Eq.(14), for fermions, by this method. We only need to change Eq.(16) and Eq.(17) by these ones:

$$P_{FD}(E_{a\to b}) = (1 - n_b)P_{MB}(E_{a\to b}), \tag{21}$$

$$P_{FD}(E_{b\to a}) = (1 - n_a)P_{MB}(E_{b\to a}), \tag{22}$$

with n_a, $n_b = 0, 1$. We must interpret Eq.(21) and Eq.(22) by saying that the probability for one fermion to populate the state b (a) is either proportional to the Boltzmann factor if in state

b (a) there is no particle, or zero if in state b (a) there exists another fermion. Now, writing $n_a P_{FD}(E_{a\to b}) = n_b P_{FD}(E_{b\to a})$ and using Eqs.(21),(22) we obtain

$$n_{FD} = \frac{1}{F(\beta)\exp(\beta E)+1}, \tag{23}$$

which is the same form as Eq.(14).
A heuristic argument can be developed, based on the proceedings which lead to the BE and FD statistics above, to deduce a generalized distribution function. Bearing in mind Eq.(7) for the generalized thermodynamical probability, we must note that it allows for some fraction p/N of the total number N of particles to populate a given available state. Thus, once the allowed fraction is attained for a certain level, the probability for any anyon making a transition to that level must be zero. With this reminds, consider the following generalization of Eqs.(16)-(17) and Eq.(21)-(22):

$$P_{ANY}(E_{a\to b}) = (\gamma - f n_b) P_{MB}(E_{a\to b}), \tag{24}$$

$$P_{ANY}(E_{b\to a}) = (\gamma - f n_a) P_{MB}(E_{b\to a}), \tag{25}$$

where n_a, $n_b = 0, 1, ..p$, and $P_{ANY}(E_{b\to a})$ stands for the rate of transition from state b to state a governed by the anyonic statistics. Here γ and f are real parameters whose meaning will be cleared in the following. To interpret both Eqs.(24) and (25), we must say that the probability for one anyon to make a transition to state b is proportional to the MB factor, and this probability suddenly becomes zero if the fraction $f n_b$ of particles attains the maximum γ value for a given state b. Eq.(25) can be interpreted in a similar way. If we now require, as it was done previously for the other transition rates, that in equilibrium we must have

$$n_a P_{ANY}(E_{a\to b}) = n_b P_{ANY}(E_{b\to a}), \tag{26}$$

we can combine Eq.(24)-(26) and use Eq.(15) to obtain, after proceeding in the very same way as we did previously:

$$n_{ANY} = \frac{\gamma}{g(\beta)\exp(\beta E)+f}, \tag{27}$$

Eq.(27) generalizes the FD statistics and becomes the FD distribution when, of course, $f = 1$, which means that transitions only occur for states with no particle.
Consider now the case when the transition rates are given by $P_{ANY}(E_{a\to b}) = (\gamma + f n_b) P_{MB}(E_{a\to b})$, $P_{ANY}(E_{b\to a}) = (\gamma + f n_a) P_{MB}(E_{b\to a})$ which generalize Eqs.(16) and (17). It is straightforward to obtain, now

$$n_{ANY} = \frac{\gamma}{g(\beta)\exp(\beta E)-f} \tag{28}$$

Note that the corresponding interpretation of these transition rates must be that the anyons now, being different from those governed by Eq.(27), have increasing probability to populate a given state when the fraction f becomes larger. We see, thus, that is possible to derive, on sound grounds, generalized statistics for anyons at equilibrium.

3. Nonequilibrium systems and coherent access hypothesis

Nonequilibrium situations require a little bit of complexification, since a given particle can now act *as if* it is populating more than one state at once. To convince that this is so it is enough to remind that any initially pure state can be described, using the completeness relation, as a superposition of each state physically accessible to the particle. I call this new situation as *coherent access hypothesis* (12).

3.1 Coherent access for bosons

As we saw from the preceding Section, Tab.1, to define a microstate in statistical mechanics, i.e., a given configuration, it is necessary to take into account the (in)distinguishability of the particles, which gives rise to different configurations. In nonequilibrium situation, to calculate all the possible configurations I now take into account, beside this characteristic, this another one: the possibility to the particle simultaneously access more than one state, or, to avoid eventual difficulties related to interpretations matter inherent to the quantum formalism, the possibility to the particle to coherently access the available states. This situation is shown in Tab. 2 for the case of $p = g_j$ considering two identical particles having two accessible states ($g_j = 2$). Note that if the particles were distinguishable, the corresponding configuration would be different.

(1)	(2)	(12)
○	○	
○		○
	○	○
○○		
	○○	
		○○

Table 2. A system out of equilibrium composed by two particles having two accessible states. (1) denotes the first available state, (2) denotes the second available state, and (12) denotes the coherent access to both states.

Comparing Tab.I and Tab.II, we see that, clearly, the nonequilibrium situation requires a new strategy for counting microstates. We can represent this new strategy to count, as shown in Tab.2, by the following sequences. The number between parentheses indicates the state to be occupied, while the letter following the parenthesis indicates the corresponding occupation by the particle a, which is identical to all the others: $(1)a(2)a(12)$; $(1)a(2)(12)a$; $(1)(2)a(12)a$; $(1)aa(2)(12)$; $(1)(2)aa(12)$; $(1)(2)(12)aa$. The first sequence corresponds to one particle accessing the first available state and the other particle accessing the second available state, with no coherent access by both particles; the second sequence corresponds to one particle accessing the first state and the other one accessing coherently the states (1) and (2), and so on. As the sequence must initiate by a number, and existing three possible number of states, $1, 2$, and 12, there will remain $3-1$ numbers plus two letters a (particles) to be set in whatever order (permutation). Therefore, the number of unrepeated sequences is

$$\Omega^* = \frac{3 \times (3-1+2)!}{2!3!} = 6, \tag{29}$$

where I have put a superscript $(*)$ to remind that we are treating with nonequilibrium situation. Proceeding in a general manner, for g_j sublevels with N_j^* particles, the number Ω_j^* of unrepeated sequences is

$$\Omega_j^* = \frac{G_j(G_j + N_j^* - 1)!}{G_j! N_j^*!} = \frac{(G_j + N_j^* - 1)!}{(G_j - 1)! N_j^*!}, \tag{30}$$

where $G_j = \sum_{k=1}^{g_j} C_{g_j}^k$ is the number of possible sequences formed from g_j, and $C_n^m = n!/(n-m)!m!$. Taking as example the configuration given by Tab.2, where $g_j = 1$, $N_j = 2, G_j = \sum_{k=1}^{2} C_2^k$, thus $G_j = C_2^1 + C_2^2 = 3$; then

$$\Omega_j^* = \frac{(3 + 2 - 1)!}{(3 - 1)!2!} = 6, \tag{31}$$

which is the number of sequences given in $(1)a(2)a(12)$; $(1)a(2)(12)a$; $(1)(2)a(12)a$; $(1)aa(2)(12)$; $(1)(2)aa(12)$; $(1)(2)(12)aa$, corresponding to Tab.2. Therefore, the nonequilibrium thermodynamical probability Ω_j^* for a given macrostate j is

$$\Omega_j^* = \frac{(G_j + N_j^* - 1)!}{(G_j - 1)! N_j^*!}. \tag{32}$$

Also, as $G_j = \sum_{k=1}^{g_j} C_{g_j}^k = C_{g_j}^1 + \sum_{k=2}^{g_j} C_{g_j}^k$, and $C_{g_j}^1 = g_j$, letting $L_{gj} = \sum_{k=2}^{g_j} C_{g_j}^k$, then Eq.(32) can be rewritten and used to compose the thermodynamical probability $\Omega^* = \prod_j \Omega_j^*$:

$$\Omega^* = \prod_j \frac{(g_j + L_{gj} + N_j^* - 1)!}{(g_j + L_{gj} - 1)! N_j^*!}. \tag{33}$$

From Eq.(33) we can see that the only changing in the thermodynamical probability is the appearance of the factor L_{gj} modifying the degeneracy g_j, the number of macrostates Ω_j^*, and also the thermodynamical probability and, consequently, the entropy S of the system, as should be, since $S = k \ln \Omega^*$. Remarkably, note the similarity between Eq.(33) and Eq.(4), provided that we identify $d_j = g_j + L_{gj}$.

We can easily verify that Eq.(32) gives rise to a BE-like statistics, with g_j replaced by G_j. That this is so can be checked proceeding by analogy with the equilibrium situation, as indicated previously (1): First, we take the ln from both sides of Eq.(32). Second, we use the Stirling formula for factorials. Third, we differentiate with respect to N_j^* and use $\partial \ln \Omega_j^* / \partial N_j^* = \epsilon_j^*$, where ϵ_j^* generalizes $\varepsilon_j = \beta E_j$, $\beta = 1/kT$, finding

$$N_j^* / G_j = \frac{1}{\exp(\epsilon_j^*) - 1}. \tag{34}$$

Therefore, we see that the Bose-Einstein statistics is corrected, since the equality $G_j = g_j$ and $\epsilon_j^* = \varepsilon_j$ will be valid only when the complete equilibrium is reestablished.

Since we are treating with nonequilibrium systems, a fundamental point to be addressed is, then, what is the meaning of temperature here, and how temperature enters in these formulas. To apply these formulas, we can think that the system is bring into contact with a reservoir at a well defined temperature T. As will be discussed later, the reservoir, even at idealized zero temperature, will drive the system to a complete mixture of states at the end of the so-called decoherence time τ_D (19). As a consequence, in a time scale shorter than that for occurring the thermalization between the system and the reservoir, the coherence of the system will be lost, implying its incapacity to access, coherently, every possible state, thus recovering the usual BE distribution. Although the temperature of the reservoir is well defined for all times, the temperature for the system is not. However, as I stressed before, we can think of taking the system only slightly out of equilibrium, such that we can correct its temperature using an expansion up to few orders in ϵ_j^* and requiring that $\epsilon_j^* \to \varepsilon_j = \beta E_j$ when the equilibrium is restated.
Thus, by expanding ϵ_j^* in power series of ε_j, which formally can be written as

$$\epsilon_j^* = \epsilon_0^* + \left(\frac{\partial \epsilon_j^*}{\partial \varepsilon_j}\right)_{\varepsilon_0} \varepsilon_j + \frac{1}{2!}\left(\frac{\partial^2 \epsilon_j^*}{\partial \varepsilon_j^2}\right)_{\varepsilon_0} \varepsilon_j^2 + \frac{1}{3!}\left(\frac{\partial^3 \epsilon_j^*}{\partial \varepsilon_j^3}\right)_{\varepsilon_0} \varepsilon_j^3 ..., \tag{35}$$

and requiring that $\epsilon_j^* \to \varepsilon_j = \beta E_j$ when the equilibrium is restated, gives $\epsilon^*(0) = \epsilon_0^* = 0$ and $\left(\frac{\partial \epsilon_j^*}{\partial \varepsilon_j}\right)_{\varepsilon_{0j}} = 1$, such that the first order correction to the Bose-Einstein distribution can be explicitly written as

$$N_j^*/G_j = \frac{1}{\exp\left[\beta E + \alpha_1 (\beta E)^2\right] - 1}, \tag{36}$$

where I have truncated the power series up to second order and put $\frac{1}{2!}\left(\frac{\partial^2 \epsilon_j^*}{\partial \varepsilon_j^2}\right)_{\varepsilon_j=0} = \alpha_1$. Note that, from this approach, the net effect stemming from the nonequilibrium on a given system is the increasing in the degeneracy, which in turn increases the available states given by Ω. Also note that for systems only slightly out of the equilibrium, the energy emitted should be slightly different from that corresponding to the system in equilibrium. I call attention to the fact that some experiments, see Ref.(16), seem to point for the importance of the BE-like statistics given by Eq.(36), which modifies the Boltzmann factor.
I now briefly discuss a point deserving further investigation, which regards to Fermi-Dirac statistics ($p = 1$) for particles out of equilibrium. Take for example Table III, which shows the total of $\Omega = 6$ possibilities of occupation for $N = 2$ fermions having $M = 3$ available states. The situation now is more complicated than that for bosons, since, for example, when the first and second state is being coherently accessed by one fermion, indicated by (12) in Table III, the other fermions cannot populate states labeled using either number 1 or 2; in this case, as shown in Tab. III, when one fermion is in (12) state, the other one only can exist in state (3). It remains a challenge to find the corresponding thermodynamical probability Ω for g_j states and n_j fermions, and thus, the mean occupation number for fermions out of equilibrium, as was done in Eq.(36), for bosons.

(1)	(2)	(3)	(12)	(13)	(23)	(123)
○	○					
○		○				
○					○	
	○	○				
	○			○		
		○	○			

Table 3. Configurations of $M = 3$ accessible states for $N = 2$ indistinguishable fermions. (1) denotes the first available state, (2) and (3) denote the second and third available state, respectively. (12) represents the first and second state being coherently populated, (123) represents the three available states being coherently populated, and so on.

3.2 Coherent access for fermions

Because the plain difficulty in obtaining Ω for fermions out of equilibrium, as pointed above, another route for this purpose must be in order. Thus, let us focus our attention on the partition function, which, by definition, is defined as a sum in all microstates (ms):

$$Z = \sum_{ms} \exp(-\beta E), \tag{37}$$

where $\beta = 1/kT$, and E is the energy of the system, which can also be written using the number of particles n_i in the state of $i-$th energy $\epsilon(i)$ of the system as $E = \sum_i n_i \epsilon(i)$. In this case, the total number of particles is simply $N = \sum_i n_i$.

For an out of equilibrium system consisting of fermions, I introduce the coherent access hypothesis to several states, which consists in maintaining the same form as that of Eq.(37), but replacing $\sum_i n_i \epsilon(i)$ by $\sum_{ij...} n_{ij...} \epsilon(i, j, ...)$, which, for fermions, implies $i \neq j \neq ...$, and n_{kl} is either zero or one and must be interpreted as being the fermion coherently accessing the energy levels $\epsilon(k)$ and $\epsilon(l)$. For example, as discussed in the beginning of this Section and represented in Table III, $\epsilon(k, l)$ represents the coherent access related to the energy levels k and l, and $\epsilon(1, 2)$ represents, for example, the states (1) and (2) being coherently populated by a single fermion. On the other hand, if we are taking into account bosons, there is no constraint on $\sum_{ij...} n_{ij...} \epsilon(i, j, ...)$, and of course $n_{kl} = 0, 1, 2, ...$ is the number of bosons coherently accessing the energy levels $\epsilon(k)$ and $\epsilon(l)$.

For what follows, I am assuming that the partition function preserves its form given by Eq.(37) even at the nonequilibrium situation. To demonstrate this, I make use of this following postulate, which is valid for equilibrium situation: that two systems, in contact with a third one, as for example a reservoir at temperature T, act independently of each other while both the systems exchange energy with the reservoir. Once this postulated is maintained, it is straightforward to proceed the demonstration (1), and, for sake of completeness, let us briefly outline the steps leading to this result. To this end, consider a system composed by two subsystems A and B. The probability for this composed system to be in the energy state E^*_{A+B} is $P_{A+B}(E^*_{A+B})$, where the superscript (*) is to remind us that the system is out of equilibrium. If, as usual, the interaction energy can be neglected, thus the energy of the composed system is $E^*_{A+B} = E^*_A + E^*_B$, and

$$P_{A+B}(E^*_{A+B}) = P_A(E^*_A) + P_B(E^*_B) \tag{38}$$

is the probability for the composed system to be in a particular state such that the subsystem A has an energy E_A^*, and, at the same time, the subsystem B has an energy E_B^*. Now, suppose that these two subsystems is put in contact with a third system, for example, a reservoir at temperature T. While persisting the nonequilibrium situation (and even after that, indeed), the two subsystems A and B act independently of each other, with both subsystems eventually exchanging energy with the reservoir. Besides, the energy exchanged with the reservoir by a given subsystem does not influence the energy that the other subsystem can exchange with this same reservoir. This assumption, valid for two systems in equilibrium with a reservoir, is here assumed to be valid also when the equilibrium was not reached. Therefore, as I am assuming these events as independent, it can be writen

$$P(E_{A+B}^*) = P(E_A^*)P(E_B^*). \tag{39}$$

Differentiating Eq.(39) with respect to E_A^* and E_B^* and equating this result we obtain ($dP/dE^* = P'$)

$$P_A'(E_A^*)P_B(E_B^*) = P_A(E_A^*)P_B'(E_B^*). \tag{40}$$

Next, separating the variables and equating the result to a constant, we have

$$\frac{P_A'(E_A^*)}{P_A(E_A^*)} = \frac{P_B'(E_B^*)}{P_B(E_B^*)} = -\beta^* \tag{41}$$

where β^* is a constant independent from either E_A^* or E_B^*. Of course, in the equilibrium situation we must have $\beta^* \to \beta = 1/kT$. From Eq.(41) follows, therefore, our desired result

$$P(E^*) = \frac{\exp(-\beta^* E^*)}{Z^*}, \tag{42}$$

where the partition function for the nonequilibrium situation is $Z^* = \sum_{ms} \exp(-\beta^* E^*)$ and the index were dropped given the validity of Eq.(42) for the two subsystems.
Thus, according to Eq.(42), if $P(\epsilon_j^* = \beta^* E_j^*)$ is the probability for a given system out of the equilibrium is in a particular microstate whose configuration is described by $\epsilon_j^* = \beta^* E_j^*$, then

$$P(\epsilon_j^*) = \frac{\exp(-\epsilon_j^*)}{Z^*}. \tag{43}$$

Now, using Eq.(35) and requiring that $\epsilon_j^* \to \beta E_j$ when the equilibrium is restated, Eq.(43) can now be written as

$$P(\epsilon_j^*) = \frac{1}{Z^*} \exp\left[-\beta E_j - \alpha_1 \left(\beta E_j\right)^2 - \alpha_2 \left(\beta E_j\right)^3 + \alpha_3 \left(\beta E_j\right)^4 ...\right], \tag{44}$$

where the other constants were renamed for convenience as $\frac{1}{n!}\frac{\partial^n \epsilon_j^*}{\partial \epsilon_j^n} = \alpha_{n-1}$. Such a state of affairs giving origin to an infinite number of free parameters was studied in Refs. (14; 15) in a different context. I will turn to this point in the next Section. Note that for systems only slightly out of the equilibrium this last equation can be written as

$$P(\epsilon^*) = \frac{1}{Z^*} \exp\left[-\beta E - \alpha_1 \left(\beta E\right)^2\right], \tag{45}$$

where I have dropped out the index i. As stressed before, some experiments seem to point for the importance of this last term, which modifies the Boltzmann factor (16). Eq.(45) applies as well as for bosons and fermions just eventually changing the parameter α_1 accounting for the difference in the occupation number for this two particles. It is to be noted that, in order to obtain an equation for the nonequilibrium fermions similar to that of Eq.(36), we must find the nonequilibrium partition function Z^* considering the fermionic nature of the particles, with the occupation number in Table Ib being replaced by the occupation number for out of equilibrium fermions as indicated in Table III. Unfortunately, this is also not an easy task, demanding additional efforts. Notwithstanding, Eq.(45) shows several applications, and exploring its experimental consequences would be a promising road to follow. Besides, following another route yet using Eq.(45), it is possible to find an occupation number for nonequilibrium fermions similar to Eq.(36), as desired. To this end, consider the following reasoning: since the equilibrium is slightly disturbed, it is reasonable to assume that Eq.(45), instead of Eq.(1), must be used to calculate the transition rates. Consider then a two state system of distinguishable particles. Let E_a and E_b the energy of state a and b, respectively. The MB-like distribution, Eq.(45), implies that

$$\frac{n_a^*}{n_b^*} = \frac{P_{MB}(\epsilon_a^*)}{P_{MB}(\epsilon_b^*)} = \exp\left\{-\beta(E_a - E_b) - \alpha_1\left[(\beta E_a)^2 - (\beta E_b)^2\right]\right\}, \tag{46}$$

where n_b^* and n_a^* are the number of distinguishable particles in state b and a, respectively. Eq.(46) thus corrects the usual rate of population of the two states. Now, if the equilibrium is only slightly disturbed, we can assume that the transition rate from a to b is nearly the transition rate from b to a, *i.e.*,

$$n_a^* P_{MB}(\epsilon_{a\to b}^*) \approx n_b^* P_{MB}(\epsilon_{b\to a}^*), \tag{47}$$

where I have indicated the single particle transition rate from state a to b as $P_{MB}(\epsilon_{a\to b}^*)$. By applying the same reasoning for two identical fermions as it was done previously - see Eq.(16)-(23), we can write to the rate $P_{FD}(\epsilon_{a\to b}^*)$ for one particle to make a transition from one state (a) to another (b) and *vice-versa* as

$$P_{FD}(\epsilon_{a\to b}^*) = (1 - n_b^*) P_{MB}(\epsilon_{a\to b}^*), \tag{48}$$

$$P_{FD}(\epsilon_{b\to a}^*) = (1 - n_a^*) P_{MB}(\epsilon_{b\to a}^*), \tag{49}$$

where n_b^* and n_a^* is to be reminded as the nonequilibrium number of fermions which are either zero or one. By the same told, I am assuming that similar to Eq.(47) we can write

$$n_a^* P_{FD}(\epsilon_{a\to b}^*) \approx n_b^* P_{FD}(\epsilon_{b\to a}^*). \tag{50}$$

From Eqs(47)-(50), we find

$$\frac{n_a^*}{(1 - n_a^*)} \exp\left[-\beta E_a - \alpha_1 (\beta E_a)^2\right] \approx \frac{n_b^*}{(1 - n_b^*)} \frac{\exp\left[-\beta E_b - \alpha_1 (\beta E_b)^2\right]}{,}. \tag{51}$$

If we now put $\frac{n_b^*}{(1-n_b^*)}\frac{\exp\left[-\beta E_b-\alpha_1(\beta E_b)^2\right]}{,} = F(\beta)$ and solve to n_a^* we find, after changing the subindex a to FD,

$$n_{FD}^* \approx \frac{1}{F(\beta)\exp\left[\beta E+\alpha_1\left(\beta E\right)^2\right]+1}, \tag{52}$$

which is similar to Eq.(36), for bosons out of equilibrium. In fact, it is straightforward to rederive Eq.(36) by this method, provided we assume $n_a^* P_{BE}(\epsilon_{a\to b}^*) \approx n_b^* P_{BE}(\epsilon_{b\to a}^*)$ and use, now, $P_{FD}(\epsilon_{a\to b}^*) = (1+n_b^*)P_{MB}(\epsilon_{a\to b}^*)$ and $P_{FD}(\epsilon_{b\to a}^*) = (1+n_a^*)P_{MB}(\epsilon_{b\to a}^*)$, instead of Eqs.(48)-(49).

For a simple, yet illuminating application of Eq.(45), consider N two-level systems at equilibrium at temperature T_1 with energies $E_1 = \epsilon$ and $E_2 = -\epsilon$. Then, when this system is bring to a slightly different temperature T_2, we expect that Eq.(45) describes the route to thermalization. In view of Eq.(45) we can write

$$Z^* = \sum_{n=-1}^{1}\exp\left[-\beta E_n-\alpha_1\left(\beta E_n\right)^2\right] \tag{53}$$

$$= 2\exp[-\alpha_1\left(\beta\epsilon\right)^2]\cosh\left(\beta\epsilon\right), \tag{54}$$

such that the N-particle partition function is $\mathcal{Z}^* = Z^{*N}$:

$$\mathcal{Z}^* = 2^N\exp[-N\alpha_1\left(\beta\epsilon\right)^2]\cosh^N\left(\beta\epsilon\right), \tag{55}$$

from which all the relevant experimental quantities can be deduced. As for example, the internal energy is

$$U^* = -\frac{\partial\ln\mathcal{Z}^*}{\partial\beta}$$

$$= -N\epsilon\tanh(\beta\epsilon)+2N\alpha_1\beta\epsilon^2, \tag{56}$$

and the specific heat

$$C = \frac{\partial U^*}{\partial T}$$

$$= Nk(\beta\epsilon)^2\cosh^{-2}(\beta\epsilon)-2Nk\alpha_1\left(\beta\epsilon\right)^2, \tag{57}$$

where it is to remind that $\alpha_1\left(\beta\epsilon\right)^2 \ll 1$, the correction $2Nk\alpha_1\left(\beta\epsilon\right)^2$ being very small.

4. Connection with entropic forms

As discussed in Section II, since the thermodynamical probability Ω was modified to Ω^* to take into account the coherent access, a natural question emerging is what is the best entropic form related to Ω^*. I note that, depending on the choice, we will face with different implications. Once there is a plenty of entropic forms at our disposal, let us take as examples two of them: the Boltzmann-Gibbs (S_{BG}) and the Tsallis (S_q) entropies. As is well known, while the first entropy is extensive, *i.e.* $S_{BG}(A+B) = S_{BG}(A)+S_{BG}(B)$, the second one in general is not, *i.e.*, $S_q(A+B) \neq S_q(A)+S_q(B)$ if $q \neq 1$.

Let us begin by the Boltzmann-Gibbs entropy, assuming for now that the only effect of the nonequilibrium is to increase the degeneracy of the system, as seen in Section II. We shall see that is possible to relate Eq.(44), and thus Eq.(45), either with an extensive or nonextensive entropic form (15). To see how this can be accomplished, I again proceed in analogy with what was done previously, taking advantage of Lagrange multipliers. Thus, given the density operator ρ of the system and the Boltzmann constant k, we want to maximize the Boltzmann-Gibbs entropy

$$S_{BG} = -k\text{tr}\rho\ln\rho, \tag{58}$$

where tr stands for the trace operation, subjected to the constraints given by the moments

$$\left\langle (\Delta E)^n \right\rangle = \text{tr}\rho H^n, \tag{59}$$

where $n = 0, 1, 2, 3, ...$ Note that for $n = 0$ the constraint is just the normalization condition $\text{tr}\rho = 1$, while for $n = 1$ the constraint becomes the mean energy $\left\langle (\Delta E)^n \right\rangle = \text{tr}\rho H$, which are, in general, the two constraints used in the maximization procedures of the entropy. We thus multiply each constraint by the undetermined Lagrange multiplier β_n, adding the result to Eq.(58). After varying ρ, we will obtain

$$\text{tr}\left(\mathbf{1} + \sum_{n=0}^{\infty}\beta_n H^n + \ln\rho\right)\delta\rho = 0. \tag{60}$$

Since all the variations are independent and $\delta\rho$ is arbitrary, the extended (non-Maxwellian) distribution $\ln\rho = -1 - \sum_{n=0}^{\infty}\beta_n H^n$ follows, or, equivalently,

$$\rho = Z^{-1}\exp(-\sum_{n=1}^{\infty}\beta_n H^n), \tag{61}$$

where the partition function is $Z = \text{tr}\exp(-\sum_{n=1}^{\infty}\beta_n H^n)$. In the energy representation where $H\left|E\right\rangle = E\left|E\right\rangle$, Eq.(61) reads

$$\begin{aligned} P(E) &= Z^{-1}\exp(-\sum_{n=1}^{\infty}\beta_n E^n) \\ &= Z^{-1}\exp\left(-\beta_1 E + \beta_2 E^2 + \beta_3 E^3 + \beta_4 E^4 ...\right) \end{aligned} \tag{62}$$

where $Z = \sum_E \exp(-\sum_{n=1}^{\infty}\beta_n E^n)$ and I have used $P(E) = \text{tr}\rho\left|E\right\rangle\left\langle E\right|$. The Lagrange multipliers β_k are formally obtained from $\beta_k = -\frac{\partial \ln Z}{\partial E^k}$, considering $E^k = Y_k$ as independent variables. The equality between Eq.(62) and Eq.(44) is guaranteed, provided that $\beta_1 = \beta$; $\beta_n = \alpha_{n-1}\beta^n$. Therefore, according to the present view, nonequilibrium systems remains extensive, although requiring a *posteriori* knowledge of the variance (second central moment), the coefficient of skewness (third central moment), the kurtosis (fourth central moment), and so on, thus giving rise to a virtually infinite number of free parameters. However, instead of using infinite parameters, we could just use a single one by redefining a new ensemble fully determined by this single parameter. An aesthetically appealing way to do so is to expand Eq.(44) in terms of

the Tsallis entropic index (13). To do this, consider the following expanded form of Eq.(44):

$$P(\epsilon_j^*) = \frac{1}{Z^*} \exp\left[-\beta E_j - \frac{(1-q)}{2}\left(\beta E_j\right)^2 - \frac{(1-q)^2}{3}\left(\beta E_j\right)^3 + ...\right], \tag{63}$$

where in general $\alpha_n = \frac{(q-1)^{n-1}}{n}$. This is equivalent to the statement that the old ensemble which depended of β, $\{\alpha_n\}$ and E_j becomes now a function of only β, q and E_j. Eq.(63) can be rewritten as

$$P(\epsilon_j^*) = \frac{1}{Z^*} \exp\left\{\frac{1}{1-q}\left[(1-q)\,\beta E_j - \frac{(1-q)^2}{2}\left(\beta E_j\right)^2 - \frac{(1-q)^3}{3}\left(\beta E_j\right)^3 - ...\right]\right\}, \tag{64}$$

where it is easily recognized the expanded form of the logarithm function $\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - ...$, $x = (1-q)\beta E_j$, such that Eq.(64) becomes

$$P(\epsilon_j^*) = \frac{1}{Z^*}\left[1 - (1-q)\,\beta E_j\right]^{\frac{1}{(1-q)}}, \tag{65}$$

which is the q-distribution stemming from the extremization of Tsallis entropy,

$$S_q = k\frac{1 - \sum_j p_j^q}{q-1}, \tag{66}$$

when considering a family of constraints determined by the q-expectation value of the energy

$$\langle E\rangle_q = \frac{\sum_j p_j^q E_j}{\sum_j p_j^q}, \tag{67}$$

besides the norm constraint $\sum_j p_j = 1$. Therefore, a complete formal agreement between Tsallis and Boltzmann-Gibbs entropies is possible. This formal equivalence gives rise to an important issue related to a possible pseudononextensivity of the entropy used to describe a given system, since this agreement suggest that nonextensivity can be removed by adding new constraints.

5. Coherent access and decoherence time

In Section II I mentioned that equilibrium destroys the possibility of the system to coherently access all the available states, thus preventing us to experimentally observe results diverging from BE and FD distributions - see Eqs.(36),(52) and (57). In this Section it is briefly discussed possible experiments to be done in order to measure deviation from the equilibrium distributions.

To understand how the coherent access capacity of the system is lost even before the reservoir to compel the system to the equilibrium, it will be instructive to analyze in some detail the decoherence time concept. As I mentioned before, there is a relevant time scale, much less

than that necessary to the system to acquire the thermal equilibrium, given by the decoherence time, which has been extensively studied (17–20). For sake of the clarity, I will briefly present the decoherence induced by the environment interpretation (19). In the so-called Caldeira-Leggett model (21), as proposed initially by Feynman and Vernon (22), a massive quantum system of interest and its environment are represented by a single Hamiltonian. The massive system is represented by a single quantum oscillator, while the environment is modeled as a thermal bath consisting of a set of N oscillators, the coupling occurring via position coordinates. The master equation for the density operator derived by Caldeira and Leggett from this model, in the position representation $\rho(x, x\prime, t)$ and in the high temperature limit, is

$$\dot{\rho} = -\frac{i}{\hbar}\left[H, \rho\right] - \gamma\left(x - x\prime\right)\left(\frac{\partial\rho}{\partial x} - \frac{\partial\rho}{\partial x\prime}\right) - \frac{2m\gamma k_B T}{\hbar^2}\left(x - x\prime\right)^2\rho \tag{68}$$

where H is the Hamiltonian for the particle, γ is the relaxation rate, m is the mass of the system, k_B is the Boltzmann constant, T is the temperature of the heat bath and $\Delta x = x - x'$ is the typical separation from two peaks of the density matrix $\rho(x, x\prime, t)$ in the phase state representation of a quantum superposition; x and x' standing for the diagonal and off-diagonal position of the peak, respectively. In the above equation, the first term corresponds to the unitary von Neumann evolution, while the second term is responsible for dissipation. The third term, having a classical counterpart related to Brownian motion, is the most important for our purposes, since it is responsible for eliminating the off-diagonal terms, thus destroying the quantum coherence, or in other words, the coherent access capacity of the system. For example, for a coherent superposition of two Gaussians, its evolution, according to Eq.(68), will initially present four peaks, two on the diagonal ($x = x\prime$) and two off the diagonal ($x \neq x\prime$). While dissipation is governed in a time scale given by the second term, decoherence is governed by the last term which, being proportional to the square of the difference between diagonal and off diagonal terms, it will be dominant for large Δx, eliminating the off-diagonal terms at the rate $\dot{\rho} \cong -\tau_D\rho$, which implies $\rho \sim \exp(-t/\tau_D)$, where τ_D is the decoherence time

$$\tau_D \equiv \frac{\hbar^2}{2m\gamma k_B T\left(\Delta x\right)^2}. \tag{69}$$

Therefore, this model provides an indicative to the time scale for the coherent access to work and thus for the emergence of the Bose-Einstein and Fermi-Dirac statistics. Besides, it sheds light on the problem of nonexisting superpositions of macroscopic objects: as shown in Ref. (19), according to Eq.(69) it is enough for $1g$ of matter at room temperature $T = 300\ K$ and separation of $1cm$ to the decoherence be destroyed in $\tau_D \sim 10^{-21}s$, even considering relaxations times of the order of the age of the Universe. However, as I are arguing, superposition is the core of the coherent access hypothesis. I shall, therefore, concentrate the attention on systems composed of relatively few particles. For instance, Bose-Einstein condensates (BEC) for N atoms at very low temperature indeed is a good candidate to observe deviations from Bose-Einstein statistics in accordance with Eq.(36). To be specific, considering, for example, a relaxation rate $\gamma \sim 1s^{-1}$taken from the average time of condensates of N atoms of rubidium-87 at $10^{-7}K$ (23), then assuming $\Delta x \sim 1\mu m$, which is much greater than the atom size, the decoherence time will be scaled as $\tau_D \sim N^{-1}s$. For typical N, around $10^2 - 10^6$, we see that the decoherence time is high enough to be experimentally measured,

thus allowing the system to coherently populate the accessible states. Therefore, we can expect that BEC slightly out of equilibrium would emit radiation whose statistics, differing from that of Bose-Einstein, should agree with the predictions given by Eq.(36).

6. Conclusions

Generalized statistics as well as nonequilibrium thermodynamics, being an open subject under current investigation, provides us with several techniques to study the statistical properties of systems. In this Chapter I study possible ways to introduce generalized statistics in both equilibrium and nonequilibrium situations. Also, I developed a new approach to study a nonequilibrium system. This new approach consists in extending the counting methods taken from combinatorial analyses to include the possibility for a system out of equilibrium to act as if it is accessing more that one state at once, a hypothesis I named by coherent access. As a consequence of the coherent access hypothesis, the process of counting how the particles access the available states of a physical system is modified. According to this hypothesis, the statistics resulting from the spectrum of emission of systems out of equilibrium is different from the Bose-Einstein statistics, being this difference in principle experimentally detectable. I then pointed out that coherent access is inhibited at the time scale of the decoherence time, and I explored situations found in Bose-Einstein condensates (BEC) which can be suited to demonstrate the validity of the coherent access hypothesis. Finally, I note that, although BEC systems were mentioned, the above reasoning must be valid for whichever systems having a sufficiently high decoherence time for preserving the coherent access.

7. Acknowledgments

I thanks the CNPq, Brazilian Agency, for the partial support of this work. I also thank Dr. J. N. T. Rabelo, for the careful reading of this manuscript.

8. References

[1] See E. Atlee Jackson, Equilibrium Statistical Mechanics, Dover Publications, Inc., N.Y., 2000.
[2] B. I. Halperin, Phys. Rev. Lett. 52 1583 (1984).
[3] R. B. Laughlin, Phys. Rev. Lett. 60 1583 (1988).
[4] J. M. Leinaas and J. Myrheim Nuevo Cimento B 37 1 (1977)
[5] F. Wilczek, Phys. Rev. Lett. 48, 1144 (1982).
[6] O. W. Greenberg, Phys. Rev. Lett. 64 705 (1990); O. W. Greenberg, Robert C. Hilborn, Phys. Rev. Lett. 83 4460 (1990).
[7] R. Ramanathan, Phys. Rev. D 45 4706 (1992).
[8] F. D. M. Haldane, Phys. Rev. Lett. 67 937 (1991).
[9] R. Acharya and P. N. Swamy, J. Phys. A: Math. Gen. 27 (1994) 7247-7263
[10] J. Elmo and N. G. de Almeida, Unpublished.
[11] F. Reif, Fundamentals of statistical and thermal physics, Mc Graw Hill, 1965.
[12] N. G. de Almeida, Physica. A, 387, 5772 (2008).
[13] C. Tsallis, J. Stat. Phys. 52, 479 (1988).
[14] C. Tsallis, D. Prato, and A. R. Plastino, Astrophysics and Space Science 290, 259 (2004).

[15] N. G. de Almeida, Physica. A, 387, 5772 (2008).
[16] D. D. Clayton, Nature 249, 131 (1974).
[17] E. Joos and H. D. Zeh, Z. Phys. B 59, 223 (1985);
[18] W. G. Unruh and W. H. Zurek, Phys. Rev. D 40, 1071 (1989).
[19] W. H. Zurek, Phys. Today 44, 36 (1991).
[20] B. L Hu. J. P. Paz and Y. Zhang, Phys. Rev. D 45, 2843 (1992).
[21] A. O. Caldeira and A. J. Leggett, Phys. Rev. A 31, 1059 (1985).
[22] R. P. Feynman and F. L Vernon, Annals of Physics 24, 118 (1963).
[23] Eric A. Cornell and Carl E. Wieman, Scientific American 278 (3): 40–45 (1998).

10

Flows of Information and Informational Trajectories in Chemical Processes

Nelson Flores-Gallegos and Carmen Salazar-Hernández.
Unidad Profesional Interdisciplinaria de Ingeniería Campus Guanajuato,
Av. Mineral de Valenciana No. 200 Col. Fracc. Industrial Puerto Interior,
C.P. 36275 Silao de la Victoria, Guanajuato
Instituto Politécnico Nacional
México

1. Introduction

What is the importance of the concept of information measure in quantum mechanics? To answer, it is necessary to try to establish the importance of the concept of information, which is a general concept and perfectly applicable to any case. For example: What do they have in common codes used to send messages from a communications satellite and the bases of a DNA molecule? How does the second law of Thermodynamics and Communication, to the extent that is possible to speak of the entropy of a musical score? Why the intricate problems of probability is related to the way we express ourselves orally or in writing? The answer to all is *information* and the fact that one concept can link different ideas so reveals its great generality and power.

Until the forties, had not been defined information as a scientific term, and this definition was quite new, different from all the common meanings, and precisely because they were described with sufficient accuracy for mathematicians and engineers in telecommunications, the concept became more and more fascinating to the scientific community. The word began to regain some of its meanings that had fallen into disuse. The idea of information as an active agent, i.e. something that informs the material world in a manner similar to the messages of the genes that instruct the cellular machinery to build an organism.

This information emerged as a universal principle operating in the world that shapes, which specifies the special character of living forms, and even help determine, by means of special codes, forms of human thought. Thus, the information covers the disparate fields of computers, Communications Technologies, Physics, Biology, Chemistry, Mathematics, among others.

The information became a scientific concept when it began the era of electronic communication. Scientists were largely what their nineteenth-century predecessors did with the concept of energy. They turned it in theory, gave laws, festooned as usual equations and stripped it of vagueness and mystery. In its pure form, the information theory was discovered by engineer Claude E. Shannon. His most notable achievements were in the transmission of color television, the design of radar systems for warning and recovery of intact messages coming from a communications satellite. Shannon, of *Bell Telephone Laboratories*, presented the world in two papers published in *Bell System Technical Journal* in July and October 1948.

In essence, the Shannon articles contain a set of theorems dealing with the problem of fast delivery, economical and efficient messages from one place to another. But the vast and fascinating implications of Shannon's work focused on the fact that he managed to explain the concept of information with such logic and precision that can be placed within a formal framework of ideas.

By treating information in terms of clearly defined but totally abstract, Shannon was able to generalize and apply laws established not just for a few types of information, but for everyone, anywhere. Although his essays may seem rather abstract and technical in the first reading, provide new ways to analyze any system or process in which exchange messages.

Shannon trials dealing with issues involving shared intellectual concerns: order, disorder, error and control error, and realization of potential possibilities, uncertainty and limitations. Scientists still wonder why the processes of nature manifest such order, where the state would most likely be uncertainty and chaos, a surrender to the forces of disorder that seem so overwhelming universal and so natural. This is considered one of the great paradoxes of science related to the philosophical question of *why exists something instead of nothing?* In his essays, Shannon proved that contrary to what might be expected, *a part* of a message can persist in the middle of the *all* of a random disorder or noise.

The most amazing thing was that Shannon's expression for the amount of information, the first accurate measurement, scientific, satisfying the first definition of the twentieth century was shaped like an equation created many years ago, in the nineteenth century, to a law of physics peculiar and very elusive: *the entropy*, thereby Shannon established an equivalence between entropy and information, fully applicable to any system or process.

This is one of the few theories that have a well defined birth in time, having a birthday and has a parent. The father's name was Claude E. Shannon and the article gives rise to the Information Theory was published in 1948. Entitled *"A Mathematical Theory of Communication"* Shannon (1948) and appeared in Bell System Technical Journal Bell System company. It is curious that this theory has a birth so timely, because usually all important creations are collective creations. In 1998 celebrated fifty years of Information Theory and practice around the world appeared special issues in journals, commemorative symbols that demonstrate all that changed in a very short period of time. Of all the articles published were collected on the occasion of fifty years a series of studies which linked to Telecommunications and, of course, also in Mathematics, Statistics, Economics, Linguistics, Quantum Theory, Astrophysics, Atomic Physics, Genetics, Molecular Biology, Neural Networks and a number of areas that were originally outside the Shannon model.

In the mid 80's, part of the scientific community thought of combining information theory with Quantum Mechanics which led to the emergence of a new branch has been extended to include the until then unexplored territory of the transmission and processing quantum states, as well as the study of quantum information and its relationship with traditional forms of information. We could ask us why this did not happen before?, because some time ago has been accepted quantum principles as the foundation of Modern Physics. Until recently it we was thought the information in terms purely classical and Quantum Mechanics played only a minor role in the design of processing equipment and in setting limits on the rate at which we could send information by certain types of channels.

At a more fundamental level, it has become clear that an Information Theory based on the principles of Quantum Mechanics, expands and complements the Classical Information Theory Vedral (2002). In addition to the quantum generalizations of classical notions such

as sources, channels and codes, this new theory includes two complementary types of quantifiable data: classical information and quantum entanglement.
Finally, the purpose of this chapter is to present the idea that can improve our understanding of Nature by not only analyzing the behavior exclusively in terms of matter and energy, even at the level of elementary particles, but the study by the techniques and methods of modern physics and chemistry, integrate concepts and tools that allow us to comprehensively investigate the behavior of natural systems in order to deepen our understanding of them to incorporate information measures that take into account concepts such as entanglement, known since the early days of Quantum Mechanics, for which, however, there are no measures in classical theories.

2. Informational flows

During the study of quantum systems, is common to use phrases like "suppose the system is in state n", on the understanding that the state $|n\rangle$ is an eigenestate of a set of operators that commute with each other and that this specification is sufficient for our purposes. However a physical system, does not necessarily to be fulfilled this condition, i.e. not necessarily characterized by a wave function as $\psi_n(x) = \langle x \mid n \rangle$, it may be that we are interested in quantum ensemble, consists in turn by various subensembles, each one characterized by a wave function $\psi_1, \psi_2, \psi_3, \ldots$. For that, it is necessary to use a simple and elegant theory called *density matrix*, see for example Fano (1957); McWeeny (1960); Davidson (1976).
Then the description of a physical system in terms of a density matrix is the most general quantum description available; this formalism includes as particular case the description of pure states. To establish what it is a pure state in terms of the density matrix, we notice that if the system is in a state $|p\rangle$, then only in this state is done mixing, so the contribution to it from any other state $|q\rangle$ is null ($q \neq p$). Therefore the pure state is defined by the conditions

$$w_i = \delta_{pi} \tag{1}$$

and the density matrix is reduced to

$$\hat{\rho} = \sum_i w_i \, |i\rangle \langle i| = \sum_i \delta_{pi} \, |i\rangle \langle i| = |p\rangle \langle p| \, , \tag{2}$$

this is consistent with the projector state $|p\rangle$:

$$\hat{\rho} \, |p\rangle \langle p| = \hat{P}_p \tag{3}$$

And now, if we consider the density matrix in coordinate representation about pure state $\psi(x)$ is

$$\rho(x', x) = \langle x' | \, |p\rangle \langle p| \, |x\rangle = \psi_p(x') \psi_p^*(x), \tag{4}$$

from this is straightforward to show that

$$\hat{\rho} = |p\rangle \langle p| \, |p\rangle \langle p| = |p\rangle \langle p| \, , \tag{5}$$

so the density matrix of a pure state is idempotent:

$$\hat{\rho}^2 = \hat{\rho}. \tag{6}$$

This is property that is used to define a pure state.

As an algebraic relation between matrices is also valid for its eigenvalues, the eigenvalues of $\hat{\rho}$ of a density matrix of a pure state satisfy the equation $\hat{\rho}^2 = \hat{\rho}'$ so this values can be only 0 and 1. As $tr\hat{\rho} = 1$, only one of the eigenvalues can be 1, so that its diagonal representation the density matrix of the pure state $|p\rangle$ has a 1 in the diagonal element (pp) and other elements are zero.
In the case of a mixture that is true $tr\hat{\rho}^2 \leq 1$ an the equation (6) implies that equality will be only when the system is a pure state. This permit us define the degree of impurity a mixture as $I = 1 - tr\hat{\rho}^2 = tr(\hat{\rho} - \hat{\rho}^2) = tr\hat{\rho}(1 - \hat{\rho})$, this result represents the expected value of $1 - \hat{\rho}$ and 0 for a pure state and tends to a value of 1 with increasing impurity (when $tr\hat{\rho}^2 \ll 1$).[1]
For a pure state $|p\rangle$, we have

$$\left\langle \hat{A} \right\rangle = tr\hat{\rho}\hat{A} = tr\,|p\rangle\,\langle p|\,\hat{A} = \langle p|\,\hat{A}\,|p\rangle\,,$$

thus recover the expected value definition used in the quantum mechanics of pure states:

$$\left\langle \hat{A} \right\rangle = \langle p|\,\hat{A}\,|p\rangle\,.$$

We use these results to show that the superposition that characterizes a pure state is consistent. To show this, consider the state $|\psi\rangle$, we express the pure state as the superposition

$$|\psi\rangle = \sum_n q_n\,|n\rangle\,; \tag{7}$$

the density matrix is given by

$$\hat{\rho} = |\psi\rangle\,\langle\psi| = \sum_{n,m} q_n q_m^*\,|n\rangle\,\langle m| = \sum_{n=m} |q_n|^2\,|n\rangle\,\langle n| + \sum_{n\neq m} q_n q_m^*\,|n\rangle\,\langle m|\,. \tag{8}$$

The group of diagonal elements can be represented as

$$\sum_n |q_n|^2\,|n\rangle\,\langle n| = \sum_n |q_n|^2\,\hat{\rho}_n,$$

the sum is incoherent (without interference) contributions from each of the pure states $|n\rangle$, with the density matrix $\hat{\rho}_n = |n\rangle\,\langle n|$, the other terms in the expression for $\hat{\rho}$ constitutes the contribution of interference between the states $|n\rangle$ and $|m\rangle$, If the superposition was inconsistent these terms may not appear. In summary, understanding it with due care, we could say that the superposition is a property relative to a pure state (with respect to a base),

[1] The equation (6) can be written as

$$\sum w_n^2\,|n\rangle\,\langle n| = \sum w_n\,|n\rangle\,\langle n|\,,$$

allows us to establish that $w_n^2 = w_n$, i.e. a density matrix weights can only contain 0 or 1. But as the condition on the trace implies that only one of these weights may be different from zero, the condition (6) of idempotency is necessary and sufficient to define a pure state.
At the end opposite the pure case corresponds to a situation where the total lack of information on the possible status of any of the subsystems leads to assign equal weight to all of them, which gives, if the mixture has a total of N components, $w_n \frac{1}{N}$. In this case we obtain

$$tr\hat{\rho}^2 = \sum_{n=1}^{N} w_n^2 = \sum_{n=1}^{N} \frac{1}{N^2} = \frac{1}{N}.$$

For $N \longrightarrow \infty$ this amount can be taken arbitrarily small values.

it is irreducible to a mixture (as well as in the first case is consistent, while the second is incoherent).
Moreover, the uncertainty of a quantum system can be determined from the collection of probabilities $p_i(A)$ of an observable A by von Neumann entropy, von Neumann (1955)

$$S(A) = -\sum_{i=1}^{n} p_i(A) \ln p_i(A), \tag{9}$$

where $\sum_{i=1}^{n} p_i(A) = 1$ y $p_i(A)$ correspond to the elements of a diagonalized density matrix, and will be a maximum when $p_1(A) = p_2(A) = \cdots = p_i(A) = 1/n$, with this condition $S_{\max}(A) = \ln n$, and for a pure estate $S(A) = 0$. Then, within the context of Quantum Chemistry, we consider a pure state as one whose values of the reduced density matrix, fulfilled with the condition $1/p_{11}(A) = 1/p_{22}(A) = \cdots = 1/p_{nm}(A)$. This condition only be satisfied by a monodeterminantal methodologies, such as Hartree-Fock or Density Functional Theory. In this perspective, has been used the Shannon entropy as a measure of correlation in atomic systems Ziesche (1995). Moreover it is important to notice that the Shannon entropy, $S(\vec{r}) = -\int \rho(\vec{r}) \ln \rho(\vec{r}) d\vec{r}$ is not equal or equivalent to the von Neumann entropy, $S(\rho) = -Tr\{\rho \ln \rho\}$. To show this, consider an observable A found in the state ρ, whose expected value is

$$\langle A \rangle = -tr\{\rho, A\}, \tag{10}$$

then if the state is described by a density matrix ρ, the corresponding entropy of that state is

$$S(\rho) = -tr\{\rho \ln \rho\}, \tag{11}$$

while the entropy of the observable is given by

$$S(\rho, A) = -tr\{(\rho, A) \ln(\rho, A)\}, \tag{12}$$

therefore the entropies $S(\rho)$ and $S(\rho, A)$ are not equal or equivalent. The Shannon entropy determines the classical information of the system and is defined in terms of an observable, while von Neumann entropy determines the quantum information of the system and is defined in terms of a density matrix.
Both entropies are subject to

$$S(\rho) \leq S(\rho, A) \tag{13}$$

and will be equal if and only if $[\rho, A] = 0$. That is, only when the values of the observables commute with their density matrix.
From the definition of the uncertainty of a quantum system characterized by its density matrix and using different definitions we can define the following entropies

$$H(A) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(A, B) \ln \sum_{j=1}^{m} P_{i,j}(A, B), \tag{14}$$

$$H(A \mid B) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(A, B) \ln P_{i,j}(A \mid B), \tag{15}$$

$$H(A, B) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(A, B) \ln P_{i,j}(A, B), \tag{16}$$

$$H(A : B) = \sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(A, B) \ln \frac{P_{i,j}(A, B)}{\sum_{i=1}^{n} P_{i,j}(A, B) \sum_{j=1}^{m} P_{i,j}(A, B)}, \tag{17}$$

which, in previous studies have shown the use of von Neumann marginal entropies applied to simple chemical processes, Flores & Esquivel (2008); Flores (2010), where $P_{i,j}(A,B)$ is the joint probability, i.e. the probability from a collection of events $(A_i, B_j) \in \xi$, where ξ is the a set of probabilities obtain form first order density matrix, we can establish which is the probability that a random event occurs with a set of probabilities $\{A\}_i$ when a event $\{B\}_j$ occurs this is given by

$$P\left\{(A_i, B_j) \in \xi\right\} = \sum_{i,j\in\xi}^{n,m} P_{i,j}(A,B) = \sum_{i=1}^{n}\sum_{j=1}^{m} P_{i,j}(A,B) = 1. \tag{18}$$

Also, we can define the marginal probability of the set $\{A\}$ as

$$P(A) = \sum_{j=1}^{m} P_{i,j}(A,B) \tag{19}$$

and the conditional probability $P(A \mid B)$ as

$$P(A \mid B) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} P_{i,j}(A,B)}{\sum_{i=1}^{n} P_{i,j}(A,B)}. \tag{20}$$

In all cases, the marginal probabilities are subject to

$$\sum_{i=1}^{n}\sum_{j=1}^{m} P_{i,j}(A,B) = \sum_{i=1}^{n} P(A) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(A \mid B) = 1. \tag{21}$$

The equations (14-17) are interrelated. For example,

$$H(A,B) = H(A) + H(B \mid A) = H(B) + H(A \mid B) \tag{22}$$

and

$$H(A:B) = H(A) - H(A \mid B) = H(B) - H(B \mid A). \tag{23}$$

this equations have been classified as a *informational balance* Guiaşu (1977). The equations (14-17) are generally represented by the following mnemonic diagram shown in Figure (1).

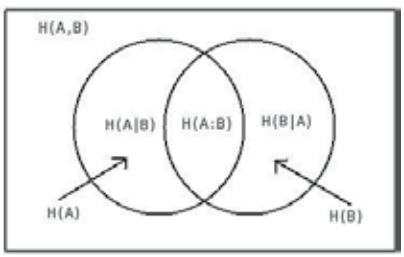


Fig. 1. Relationship between the informational entropies.

The informational entropies have the following properties:

1. The joint entropy and mutual entropy are invariants operationally:

$$H(A,B) \equiv H(B,A)$$

and

$$H(A:B) \equiv H(B:A).$$

2. The conditional entropy always is positive, $H(A \mid B) \geq 0$, consequently

$$H(A : B) \leq H(B)$$

and will be equal if and only if B is a function of A, $B = f(A)$.

3. Subadditivity,

$$H(A, B) \leq H(A) + H(B),$$

will be equal if and only if A and B are independent variables.

4. As $H(A \mid B) \leq H(A)$. Consequently $H(A : B) \geq 0$ and will be equal if and only if A and B are independent variables.

5. Strong subadditivity,

$$H(A, B, C) + H(B) \leq H(A, B) + H(B, C).$$

6. The conditioning reduces the entropy

$$H(A \mid B, C) \leq H(A \mid B).$$

In this case, we can determine the maximum amount of information allows us to quantify each Eqns. (14-17), for which we use the probability distribution that maximizes entropy, which in this case is $\sum_{i=1}^{n} P_{i,j}(A, B) = 1/n + m$, $\sum_{j=1}^{n} P_{i,j}(A, B) = 1/m + n$ and $\sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(A, B) = 1/n + 1/m$, with which we obtain

$$H_{\max}(A) = -\left(\frac{n+m}{nm}\right) \ln\left(\frac{1+nm}{n}\right), \tag{24}$$

$$H_{\max}(A \mid B) = -\left(\frac{n+m}{nm}\right) \ln\left(\frac{n+m}{n(1+nm)}\right), \tag{25}$$

$$H_{\max}(A, B) = -\left(\frac{n+m}{nm}\right) \ln\left(\frac{n+m}{nm}\right), \tag{26}$$

$$H_{\max}(A : B) = \left(\frac{n+m}{nm}\right) \ln\left(\frac{n+m}{(1+nm)^2}\right). \tag{27}$$

To calculate the informational entropies is necessary to construct the density matrix of the system of interest with the purpose of having a collection of probabilities, in this case are obtained from the density matrix of first order, in this context there is ample number of population analysis techniques allow us to obtain these values, for example Reed et al (1985); Löwdin (1955); Bruhn et al. (2006). In previous work Flores & Esquivel (2008); Carrera et al (2010); Flores (2010) have shown that the population analysis of Weinhold Reed et al (1985) and the Lowdin-Davidson Bruhn et al. (2006) this population analysis follows this criteria:

1. Are orthonormal orbitals of maximum occupancy, obtained from sub-blocks of atomic symmetry of the density matrix of first order, located on each atom.
2. Being orthogonal occupations are never negative and sum correctly to the number of electrons.
3. The obtained atomic blocks reflect the most compact description of the density around the isolated atom to which they correspond.

4. The density of natural atomic orbitals is v-representable by obtaining atomic block the first order density matrix and correspond to the potential generated by the Hamiltonian of the isolated atom.
5. The blocks obtained from first-order density matrix reflect intrinsic properties of the wave function and therefore do not have a strong dependence on the choice of the base.
6. Are rotationally invariant, which is appropriate to define the subsystems A and B used to calculate each one of the entropies.
7. Unlike the occupation numbers of Löwdin are not delocalized throughout the molecular orbital and the transform as irreducible representations of the molecular point group and this allow us the study of atomic subsystems and their interaction.

In this chapter, we present a new application of these entropies, that consist in determinate the informational flows and their trajectories of a system that is involved in a chemical process. To do this, start by defining the information amount that a pair of subsystems can be exchanged, i.e.

$$\begin{aligned} H(A) - H(B) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{i,j}(A,B)\ln\sum_{j=1}^{m} P_{i,j}(A,B) + \sum_{i=1}^{n}\sum_{j=1}^{m} P_{i,j}(A,B)\ln\sum_{i=1}^{n} P_{i,j}(A,B), \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{i,j}(A,B)\ln\left(\frac{\sum_{j=1}^{m} P_{i,j}(A,B)}{\sum_{i=1}^{n} P_{i,j}(A,B)}\right). \end{aligned} \tag{28}$$

From this result we notice that the difference $H(A) - H(B)$ represent the amount of information that a pair of subsystems can be exchanged in a process.
Similarly, we can define what level of conditioning that will have the amount of information exchanged, i.e.:

$$H(A \mid B) - H(B \mid A) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{i,j}\ln\frac{\sum_{j=1}^{m} P_{i,j}(A,B)}{\sum_{i=1}^{n} P_{i,j}(A,B)}. \tag{29}$$

Notice that $H(A) - H(B) \equiv H(A \mid B) - H(B \mid A)$, henceforth denote this result as $H(\dot{AB})$.
In applying the above result to a chemical process, we can interpret $H(\dot{AB})$ as the amount of information transferred in the system. This quantity has the following properties

$$H(\dot{AB}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{i,j}(A,B)\ln\frac{\sum_{j=1}^{m} P_{i,j}(A,B)}{\sum_{i=1}^{n} P_{i,j}(A,B)} = \begin{cases} \{A\} > \{B\} \Rightarrow H(\dot{AB}) < 0 \\ \{A\} \equiv \{B\} \Rightarrow H(\dot{AB}) = 0 \\ \{A\} < \{B\} \Rightarrow H(\dot{AB}) > 0 \end{cases}. \tag{30}$$

We can interpret this result as follows, if $H(\dot{AB}) > 0$, informational flow is given in the form $\{A\} \to \{B\}$, if $H(\dot{AB}) < 0$, the informational flow $\{B\} \to \{A\}$, while if $H(\dot{AB}) \equiv 0$ imply that the informational flow is completely equivalent in both directions, i.e. $\{A\} \leftrightarrow \{B\}$, so it is then convenient to define the net flow of information as $H(\dot{AB}) = \|H(A) - H(B)\|$.
As in the case of informational entropy, in this case we can also determine the maximum flow of information, taking into account the probability distributions that maximize this flow to obtain

$$H_{\max}(\dot{AB}) = \left|\frac{n+m}{nm}\ln\frac{n}{m}\right|. \tag{31}$$

3. Informational trajectories

In previous section we defined the amount of information that a system can be exchanged, either between the subsystems that constitute or their environment and other interesting point to study is the possible informational trajectory that follow the amount of information that the subsystems interchanged. To do this, we return the definition informational flow that we defined in the previous section, $H(\dot{AB})$, which allows us to establish or determine the amount of information that a pair of subsystems can exchange with each other, however, it is also important to determine what would be the trajectory that these informational flows.
To do this, is necessary define the informational trajectory that follow the informational flow,

$$H(\vec{AB}) = \int_C H(\dot{AB}) dP(A,B), \tag{32}$$

which has explicit form,

$$H(\vec{AB}) = \int_C H(\dot{AB}) H'(\dot{AB}) d(AB), \tag{33}$$

i.e.

$$H(\vec{AB}) = -\int_C \left[\sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(A,B) \ln \left(\frac{\sum_{j=1}^{m} P_{i,j}(A,B)}{\sum_{i=1}^{n} P_{i,j}(A,B)} \right) \left(\frac{1}{\sum_{i=1}^{n} P_{i,j}(A,B)} - \frac{1}{\sum_{j=1}^{m} P_{i,j}(A,B)} \right) \right] dP(A,B), \tag{34}$$

the last equation can be written explicitly as follow

$$H(\vec{AB}) = -\int_C \left[\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(A,B) \left[\sum_{i=1}^{n} P_{i,j}(A,B) - \sum_{j=1}^{m} P_{i,j}(A,B) \right]}{\sum_{i=1}^{n} P_{i,j}(A,B) \sum_{j=1}^{m} P_{i,j}(A,B)} \ln \left(\frac{\sum_{j=1}^{m} P_{i,j}(A,B)}{\sum_{i=1}^{n} P_{i,j}(A,B)} \right) \right] dP(A,B), \tag{35}$$

notice that $\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(A,B)}{\sum_{i=1}^{n}} = P(A \mid B)$ and $\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} P_{i,j}(A,B)}{\sum_{j=1}^{m}} = P(B \mid A)$, so

$$H(\vec{AB}) = -\int_C (P(B \mid A) - P(A \mid B)) \ln \frac{\sum_{j=1}^{m} P_{i,j}(A,B)}{\sum_{i=1}^{n} P_{i,j}(A,B)} dP(A,B), \tag{36}$$

this integral, is subject to

$$H(\vec{AB}) = \begin{cases} \sum_{i=1}^{n} P_{i,j}(A,B) \equiv \sum_{j=1}^{m} P_{i,j}(A,B);\ H(\vec{AB}) = 0 \\ \sum_{i=1}^{n} P_{i,j}(A,B) > \sum_{j=1}^{m} P_{i,j}(A,B);\ H(\vec{AB}) > 0 \\ \sum_{i=1}^{n} P_{i,j}(A,B) < \sum_{j=1}^{m} P_{i,j}(A,B);\ H(\vec{AB}) > 0 \end{cases}. \tag{37}$$

This would imply that any process, physical, chemical, physicochemical, has a positive informational trajectory which is related to a corresponding flow and informational exchange, so we could postulate the following axiom, *Any system that is subject to a natural process exchanges information*.

As in the flow of information, we can obtain the maximum informational trajectory, using the probability distributions that maximize the flow,

$$H_{\max}(\vec{AB}) = \int_C \left(\frac{n^2 - m^2}{nm}\right) \ln\left(\frac{n}{m}\right) d(n,m), \tag{38}$$

$$= \left(\frac{n^2}{2} + \frac{m^2}{2}\right) \left[\frac{1}{2} \ln\left(\frac{m}{n}\right) + 1\right], \tag{39}$$

and

$$H_{\max}(\vec{BA}) = \left(\frac{n^2}{2} + \frac{m^2}{2}\right) \left[\frac{1}{2} \ln\left(\frac{n}{m}\right) + 1\right], \tag{40}$$

the maximum trajectory $H_{\max}(\vec{AB})$ is subject to the follow conditions

$$H_{\max}(\vec{AB}) = \begin{cases} n > m \Rightarrow H_{\max}(\vec{AB}) > 0 \\ n \equiv m \Rightarrow H_{\max}(\vec{AB}) \equiv 1 \\ n < m \Rightarrow H_{\max}(\vec{AB}) > 0 \end{cases}. \tag{41}$$

4. Informational flows and informational trajectories in chemical processes

Within the context of quantum chemistry, one of the principal interests is sorting systems or chemical processes, in general, this is accomplished by using certain models and approaches in order to obtain and quantify certain parameters of the system or process and to give a physical interpretation of what is possibly happening. Generally, these parameters are: energy, electron density, molecular electrostatic potential, Laplacian of the Density, Molecular or Atomic charges, Chemical Reactivity Parameters (like Hardness, Softness, Chemical Potential, Electrophilicity Parr & Yang (1989), Frequencies, among others.
Reactivity parameters we use in this work are defined as Parr & Yang (1989):

$$\eta = \frac{1}{2}\left(\frac{\partial \mu}{\partial N}\right)_v = \frac{1}{2}\left(\frac{\partial^2 E}{\partial N^2}\right)_v,$$

$$S = \frac{1}{2\eta} = \left(\frac{\partial N}{\partial \mu}\right)_v$$

and

$$\mu = \left(\frac{\partial E}{\partial N}\right)_v,$$

where μ is the chemical potential, N is the number of particles in the system, E is the total energy of the system and v is the external potential. These parameters of reactivity, can be approximated by the Koopmans theorem and finite differences as

$$\eta = \frac{I - A}{2} = \frac{E^{LUMO} - E^{HOMO}}{2},$$

where I is the firts ionization potential and A is the electronic affinity, E^{LUMO} is the energy of Lowest Unoccupied Molecular Orbital and E^{HOMO} is he energy of the Highest Occupied

Molecular Orbital, with this approximation we can calculate the softness and the chemical potential,

$$S = \frac{1}{2\eta}$$

and

$$\mu = \frac{E^{LUMO} + E^{HOMO}}{2}.$$

Under this scheme, in this section show that by applying this parameters and the informational entropies to describe a chemical process its possible obtain trends, which, through its interpretation enable us to establish a link between the parameters commonly obtained with respect to information of the system, we intend to show that changes in the amount of information system are also related to changes in the two principal parameters in the context of the Quantum Chemistry: the electron density and the energy.
In the next section we present some applications concerning the applications of the entropies informational, informational flows and informational trajectories. In each of the examples presented here we used Gaussian 03 program Gaussian (2003), for the electronic occupation probability we used the program NBO 5.0 Glendening et al (2004) and finally the isosurfaces of the molecular electrostatic potential and the Laplacian of the density were obtained with MOLDEN program Schaftenaar & Noordik (2000).

4.1 Dissociation of HCl

In previous work Flores & Esquivel (2008); Esquivel et al (2009) have shown that changes in the informational entropies are associated with physical and chemical changes of a system, such as electron density, atomic charges, molecular electrostatic potential, normal modes of vibration, among others.
In this section we show some applications of informational flows and informational paths Eqns. (28) and (37), presented in such processes. In this case, we present the results for the dissociation HCl, which were calculated with B3LYP/6-311G.
In the Figure 2(a), shows the energy profile of the process of dissociation of the HCl in which we see that the equilibrium internuclear distance is 1.3 Å, and energy is -460.77606 Hartrees. In the Figures 2(b), 3(c) and 3(d) shows the contour surfaces of molecular electrostatic potential at a distance of 1.3 Å, and 4.0 Å, respectively.
From these figures we can notice several important aspects regarding the set of parameters used to characterize this process: i) trends obtained only allow us to get an idea of what is happening in this process based on it, apply or guess that's what happens in the process, and ii) The parameters used, the cases do not necessary represent some of the critical points that may be important to better classify the system and thus seek to identify some important aspect of it.
In the Figures 3(e) - 3(g) present the trends of informational entropy, which were obtained with the Eqn. (14), previously reported Flores & Esquivel (2008), in these trends we define the set $\{A\}$ as the natural atomic probabilities of hydrogen atom and the set $\{B\}$ as the set natural atomic probabilities of chlorine atom. In the Figure 3(e) we present the trends of $H(A)$ and $H(B)$ in which we can see that $H(A)$ has a minimum in $R = 0.45$ Å, and a local maximum in the trend $H(B)$ in the same internuclear distance, to compare trends of the charge density (Figure 2(b)) in this Figure we notice that the change in the curvature trend occurs in the same internuclear distance in both trens of the subsystems, this change of curvature can be attributed to the instability of the system, however, could also say that from this equilibrium

internuclear distance the system starts stabilized due to a redistribution of electron density, which can be viewed through the contour map of the Laplacian Density, see Figures 3(a)-3(b). At a distance of 0.65 Å, the entropy $H(A)$ have a maximum which can be attributed at the beginning in the redistribution of electron density of hydrogen atom which would schedule at the beginning of stabilization of the molecular system to reach equilibrium internuclear distance this is reflected in an abrupt change in the slope of the trend of the entropy. While in the case of the tendency of the entropy $H(B)$, notice how it behaves the chlorine atom in the molecular environment shows a maximum at 1.3 Å, this corresponds to minimum energy, this is agree with a thermodynamic interpretation of the systems. We can also say that in this case, the behavior of the entropy $H(A)$ show a inflexion points in the trend are associated with the physical and chemical changes that occur in this process.

Moreover, in the Figure 3(f), shows the trends of the conditional entropy, which we can interpret them as follows *how the hydrogen atom are conditionated by the presence of chlorine atom* or in a more general form, *how the subsystem* $\{A\}$ *is conditionated by the presence of the subsystem* $\{B\}$, notice the generality of this interpretation, because is it a guideline for board a discussion theme fundamental in the modern chemistry, *What is an atom in a molecule?* Parr et al (2005), Is possible that through the application of the conditional von Neumann informational entropies we can give a opinion to respect.

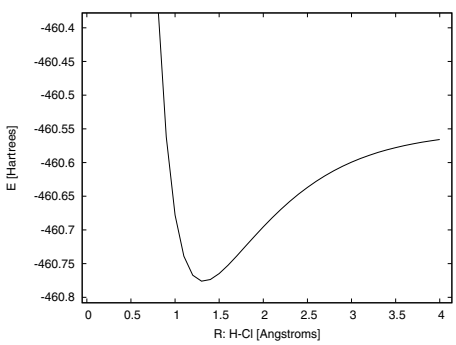
In the trends of the informational entropies, we can notice that $H(A)$ trend, have basically, the same structure of the $H(A \mid B)$ and $H(A, B)$, see Figures 3(g). This leads us to hypothesize that the hydrogen atom, which plays an important role in the process of this nature, for example, in a natural process, this subsystem has primary responsibility for carrying out a transfer of density. With this context, basically recovered several interpretations of the concepts and descriptions commonly used in Chemistry.

Finally, in the Figure 3(h) shows the trends of the flow of information and informational trajectory integral, notice that in this case, both trends have the same structure, this implied that in this process, the trajectory and informational flow depend of the electron density that the subsystems can be transferred to each other, in both cases, the critical points of these trends are the same critical points observed in the trends of informational entropy, Eqns. (14) - (17), which can be attributed to charge transfer process.

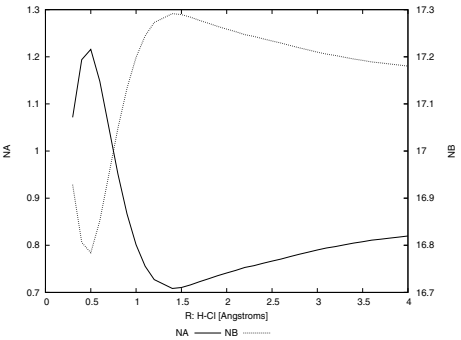
Furthermore, in the trends of $H(A, B)$ and $H(A : B)$, Figure 3(g), find that the mutual entropy, $H(A : B)$, the measure information that subsystems share increase this allows us to establish the nonclassical behavior of this system, because from a classical interpretation, we hope that in the limit when $R \longrightarrow \infty$, classically, the subsystems should not be any interaction between them, however, the entropy $H(A : B)$ shows that the subsystems A y B present a nonclassical types of interaction that is evident when the classical interactions decrease, this is, notice that $H(A : B) \neq 0$, while the total information system $H(A, B) \neq H(A) + H(B)$ and observe that it tends to a constant value. This behavior could be considered as possible evidence of quantum nonlocality Aspect (2007), i.e. evidence shows a strong correlation between pairs of subsystems, which can be separated by an arbitrarily large distance without communication (known) between them. This allows us to incorporate concepts of quantum mechanics such as entanglement, teleportation, among others.

In experimental field, the entanglement of the systems has been determined by ion trap Molmer & Sorensen (1999); Cirac & Zoller (1995); Sorensen & Molmer (1999); Kreuter et al (2004); Schlosser (2001); Monroe et al (1995), by photons Molmer & Sorensen (1999); Tittel et al (1998), phonons Jian-Wei et al (2000), by NMR Ladd et al (2002); Braunstein et al (1999). This property has also been applied to quantum cryptography Beveratos et al (2002); Jenneweinm

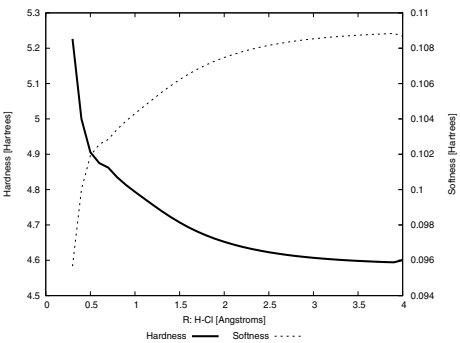
et al (2000); Gisin et al (2002). In such experiments have been used primarily: *Ca*, *Be* and *Mg*. For this circumstance, it is interesting to analyze the processes of dissociation of these systems in order to determine if they have quantum behavior as described above. Notice that these systems are clustered on the block "*s*" of the periodic table, this permit us establish the hypothesis that it is possible to find a quantum behavior in the other elements of the block and try to determine if this property have a periodical pattern that allows us to predict if a chemical system will manifest some type of quantum behavior.
In the following example, we show the characterization of a simple chemical reaction, which is characterized with concepts of modern chemistry, such as; internuclear equilibrium distances, variation of the subsystems charges, normal modes of vibration, as well as concepts of chemical reactivity, such as hardness, softness, chemical potential. In this example, we show, quantitatively, the dependence of the flow of information and the informational trajectory with the charge density of the subsystems, the internuclear distances, and so on.



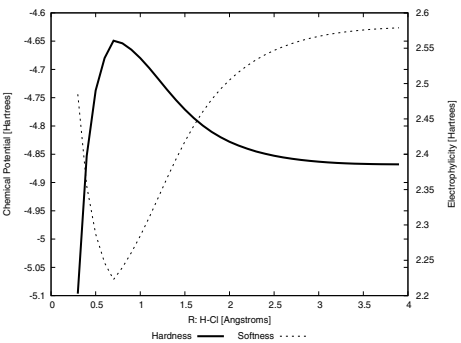
(a) Energy profile of the process of dissociation of *HCl*



(b) Trend charge density of the process dissociation*HCl*

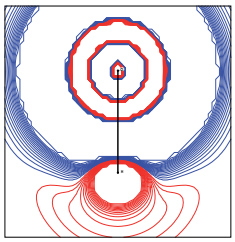


(c) Trends of the hardness and softness of the process of dissociation of *HCl*

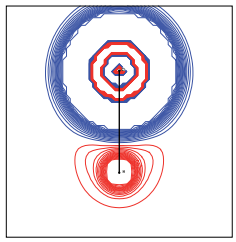


(d) Trend of chemical potential and electrophilicity in the process of dissociation of *HCl*

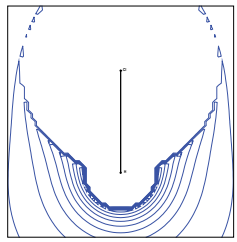
Fig. 2. Parameters that depend of energy and the electron density in the process of dissociation of *HCl*.



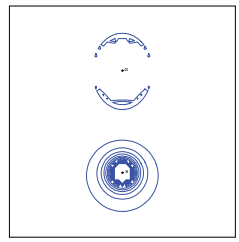
(a) LAP in R = 0.4 Å



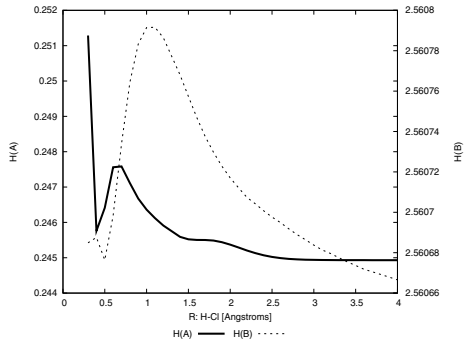
(b) LAP in R = 0.65 Å



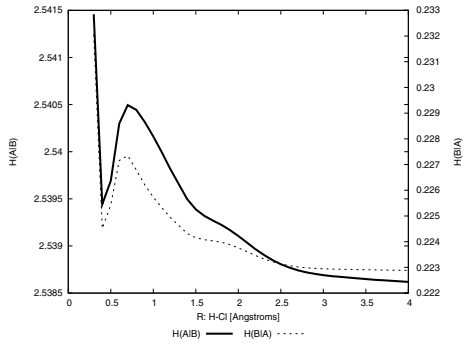
(c) MEP in R = 1.3 Å



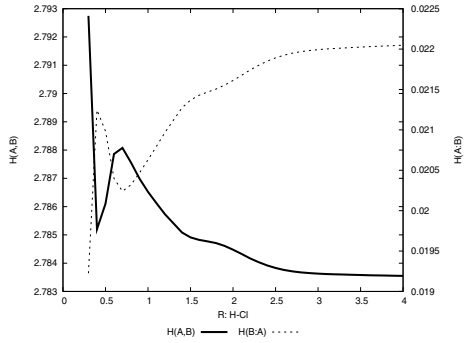
(d) MEP in R = 4 Å



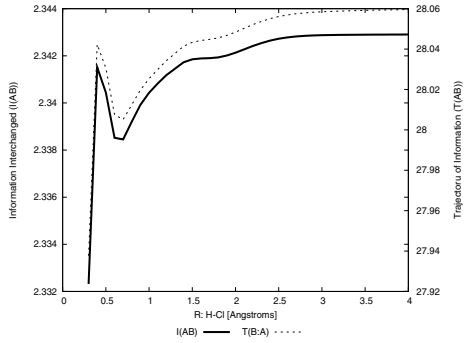
(e) $H(A)$ and $H(B)$



(f) $H(A \mid B)$ and $H(B \mid A)$



(g) $H(A,B)$ y $H(A:B)$



(h) Informational flow and informational trajectory.

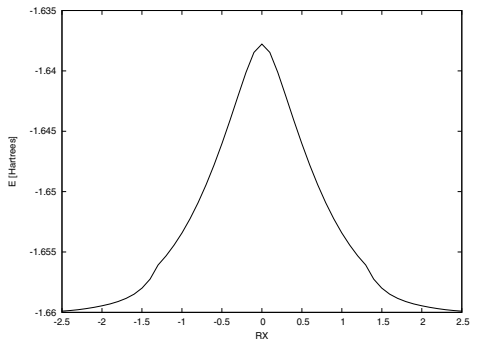
Fig. 3. Trends in informational entropies in the process of dissociation of HCl, LAP (Laplacian of Electron Density), MEP (Molecular Electrostatic Potential).

4.2 Process $H_2 + H^- \longrightarrow H_2 + H^-$

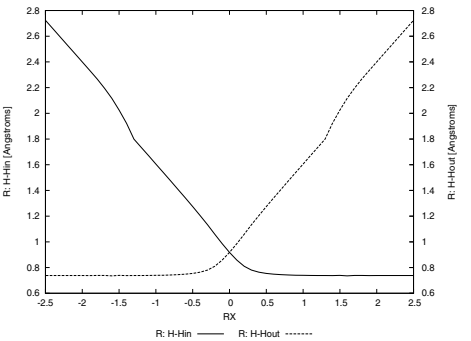
For this case study, we use the method UMP2 with the function basis set 6-311++G** to find the transition state structure. The IRC and each point of the reaction path were also obtained with the same method and function basis set, in each point of the path we verified the wave function stability Seeger & Pople (1977); Bauernschmitt & Ahlrichs (1996). The nomenclature used for this reaction is: $HH_{out} + H_{in} \longrightarrow HH_{in} + H_{out}$. From which, it is possible to generate two analysis. In the first case, selecting the set $\{A\} = H$ and the set $\{B\} = H_{in}H_{out}$, and in

the second analysis is possible take; $\{A\} = HH_{out}$ y $\{B\} = H_{in}$. In this case we just present the firts of them.
In the Figure 4(a) presents the energy profile for this reaction. In which we observe a transition structure that has a maximum, equilibrium internuclear distances at this point are 0.91808 Å, and the energy is -1.63778 Hartrees.
In the Figures 4(b), 4(c) and 4(d) correspond to the trends of the equilibrium internuclear distances, changes in the electronic population of each subsystem, the normal mode of vibration for this process, changes in hardness and softness of the system, which are chemical parameters are related to chemical reactivity.
In the Figures 5(c) - 6(b) we present the trends of the informational entropies, Eqns. (14-17), for this process. In this case two possible combinations were analyzed to generate the sets that allow us to define the joint probability, in the first case was selected set of probabilities of occupation $\{A\}$ how H_2 subsystem and the set $\{B\}$ with the occupation probabilities of H_{in}.
In the Figure 5(c) we shows the $H(A)$ trend, Eqn. (14), which basically shows how is the informational behavior of the H atom, which is directly involved in the process of rupture and formation of the chemical bond. In this figure we notice that from a distance $RX > -1.4$, $H(A)$ exhibits a change of curvature, which is attributed to the beginning of the process of breaking of the chemical bond, and can be visualized by isosurfaces of the molecular electrostatic potential, Figures 6(c) - 6(f), in which there is a decrease of the charge density in the central part of the H_2 molecule. In the coordinate $RX = -0.4$, the $H(A)$ trend shows a change in curvature, this change is associated with the process of breaking and formation of chemical bond, see Figures 6(g) - 6(j), where are the maps of molecular electrostatic potential, from such maps, we can observe a redistribution of electron density in the bond zone of the H_2 molecule, where a decrease in electron density, which is due to a decrease in the internuclear distance of the hydrogen atom which increases the interactions between these systems.
A similar behavior can be observed by comparing the trend of the normal modes of vibration of the system in Figure 4(d), which when compared with the trend of $H(A)$ we notice an excellent agreement with the normal mode of vibration, both trends have a two maximum in $RX = -0.4$ and $RX = 0.4$, these maxima we have attributed to a transition state zone Esquivel et al (2009).
As in the case of the dissociation process of HCl, we notice that $H(A,B) \neq H(A) + H(B)$, Figure 6(a), and $H(A : B) \neq 0$, Figure 6(b), the first result allows us to interpret a new way this kind of results and processes, on the one hand, urges us to leave the classic way of interpreting this type of process, i.e., an even though the subsystems are physically separated and in principle there any known interaction between them, either chemistry or physics, these systems preserve some classical types of interaction. Such results allow us to issues and philosophical implications of Quantum Mechanics, for example, if we accepts the result could be concluded that the interpretation of the results of this process using Quantum Mechanics is inconsistent with local realistic interpretation. This has important consequences. For example must be concluded that quantum systems do not have objective properties, pre measure whose values are simply revealed by the experiment. In the Copenhagen interpretation is accepted contextuality that is inferred here as a fact of Nature, so the value assigned to A, depends if you are determined jointly with B.
There has been a large number of experiments to verify the Bell inequalities Bell (1966), are fulfilled or violated Tittel et al (1998); Aspect (2007); Timothy et al (2007). Generally considered that experiments have shown that the Bell inequalities are violated, and therefore local realism has been refuted in experiments. Notice carefully the nature of this results because imply

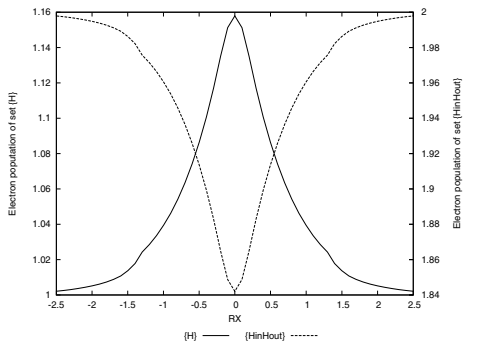
a metaphysical dilemma *Is it valid or not local realism in nature?*, which is resolved through empirical issue that has been classified in experimental metaphysics. In some way, the only precedent in physics of this situation was the EPR theorem Einstein et al (1935).
The reason for the popularity of the Bell inequalities lies in the fact that they can lead the experimental field, which opens the way for long discussions on the fundamental problems of Quantum Mechanics (determinism *vs.* indeterminism, location *vs.* holism, objectivity *vs.* subjectivity, etc.) leave the realm of speculation to go to the empirically verifiable (or refutable as desired). However, the fact is that here the discussions and disagreements persist, and there is good reason for it. Accepting the validity of the quantum description, the proposed scheme is to reject outright the existence of hidden variables, and return with it to indeterminism, or consider the final theory is deterministic, but not necessarily local.



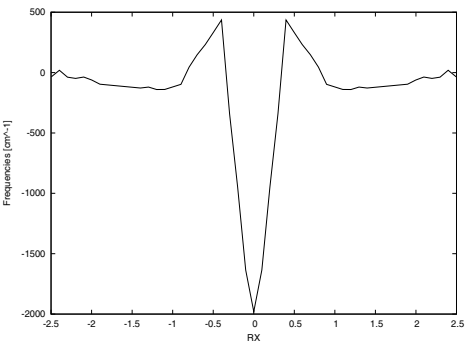
(a) Energy profile



(b) Distances



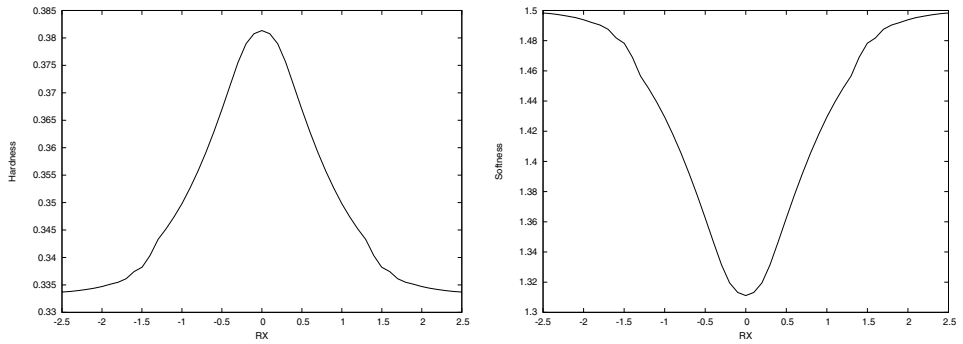
(c) Charges



(d) Frequencies

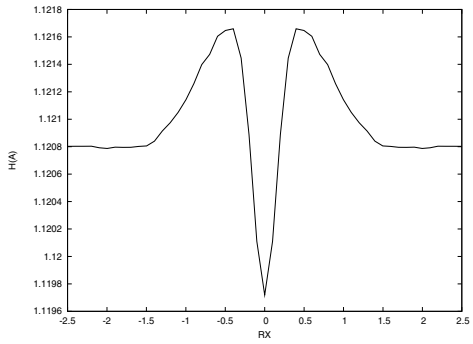
Fig. 4. Parameters characterizing the process $H_2 + H^- \longrightarrow H_2 + H^-$.

One of the most valuable aspects of each of the above figures is that the informational entropies permit us define clearly, there are areas where physical and chemical changes involved in this process. This leads us to establish the following hypothesis; *in any natural process, there is an exchange of matter, energy and information*.
The Figure 7(a) show a contour map which shows as the subsystems exchange information among themselves. In this figure we notice that approximately 1.2,1.2 coordinates correspond to the area where the flow of information is greatest, when comparing this region with the equilibrium internuclear distances Figure 4(b) we can establish a rank in the reaction

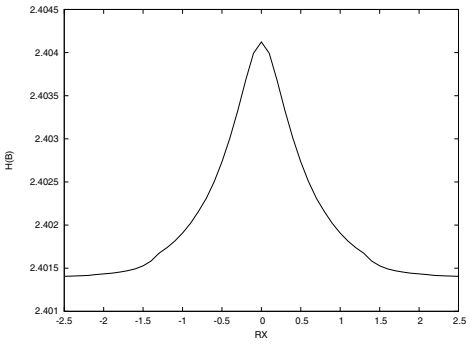


(a) Hardness

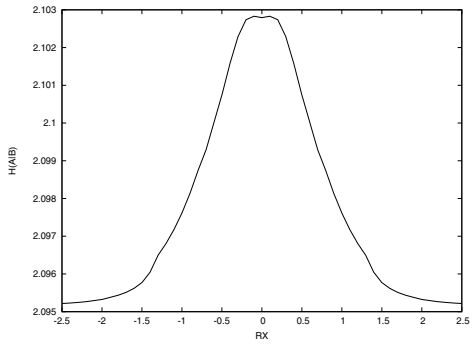
(b) Softness



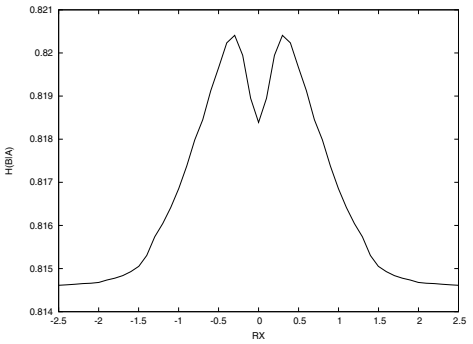
(c) $H(A)$



(d) $H(B)$



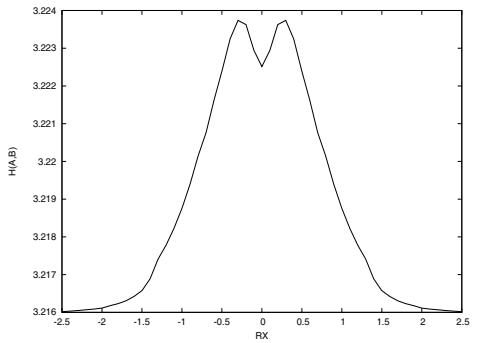
(e) $H(A \mid B)$



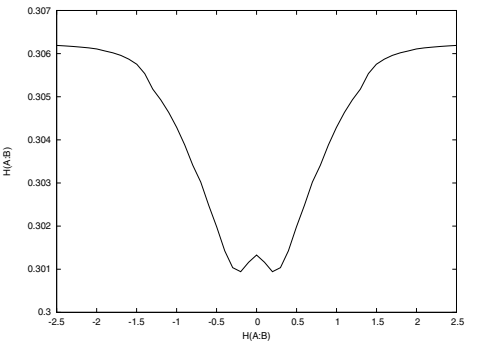
(f) $H(B \mid A)$

Fig. 5. Trends of informational entropies, using $\{A\} = \{H\}$ and $\{B\} = H_{in}H_{out}$ in the process $H_2 + H^- \longrightarrow H_2 + H^-$.

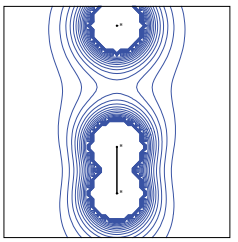
coordinate of $-0.5 < RX < 0.5$, which when compared with the trend of the normal modes of vibration, Figure 4(d), we notice that this is the same range in are the peaks we define the transition zone, also in this range we can say that the charge transfer occurs between the subsystems, see Figure 4(c), this leads us to infer that the area defined by the maximum flow of information is the area where important physical and chemical changes involved in this



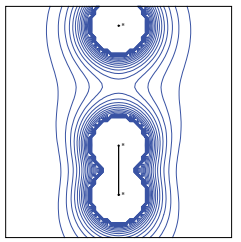
(a) $H(A,B)$



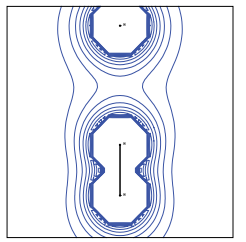
(b) $H(A:B)$



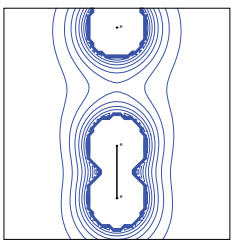
(c) RX = -2.49852



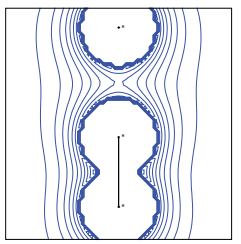
(d) RX = -2.39853



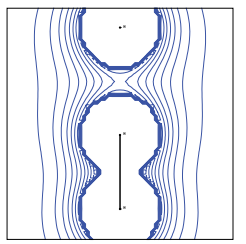
(e) RX = -2.29854



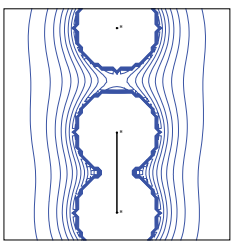
(f) RX = -2.19858



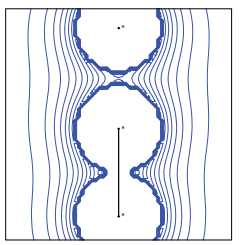
(g) RX = -1.49883



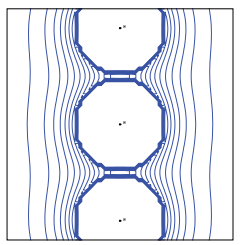
(h) RX = -1.39883



(i) RX = -1.29884



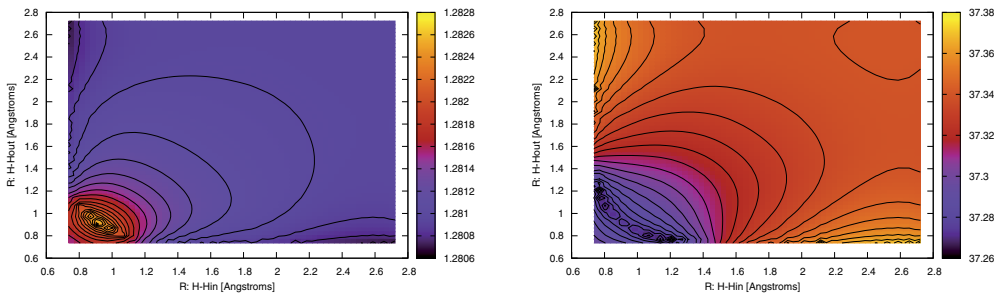
(j) RX = -1.19884



(k) RX = 0.00000 (Transition State)

Fig. 6. Trends of $H(A,B)$, $H(A:B)$ using $\{A\} = \{H\}$ and $\{B\} = H_{in}H_{out}$ and the Molecular Electric Potential in different points of the reaction coordinate in the process $H_2 + H^- \longrightarrow H_2 + H^-$.

process is carried out and while the flow of information in the $-0.5 < RX < 0.5$ is maximum, the trajectory informational, Figure 7(b), decreases when the internuclear distance increases. This same behavior leads to the conjecture that when carrying out a physical or chemical process exist gradients of information, this gradients must be associated with the electronic distribution system, therefore, in Figure 8(a) we present a contour map representing the variation of subsystems charges with respect to information flow. Using this figure we try to relate the informational content of the subsystems that are involved in this process with the charge density from which we notice that when the charge density of the subsystem A, is approximately the values $1.13 < NA < 1.16$ and B of the subsystem at $1.84 < NB < 1.87$, thereby when there are major differences in charges between the subsystems is when we observe the flow of information between the subsystems is maximal.



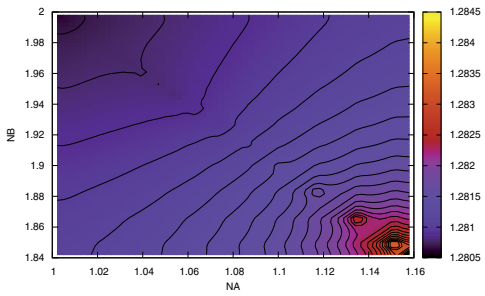
(a) Internuclear distances with informational flow (b) Inter nuclear distances with informational trajectory

Fig. 7. Contour maps of the informational flow and informational trajectories in the process $H_2 + H^- \longrightarrow H_2 + H^-$.

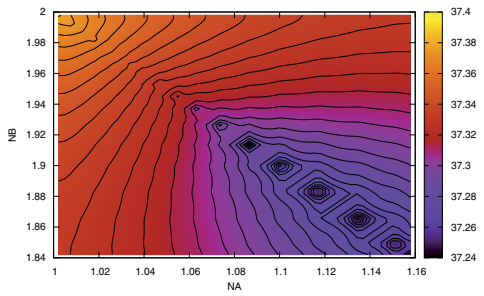
Also, notice that this flow is not constant during the process and is associated with the difference of the charge densities of the various subsystems which leads us to infer that there must be gradients of information which in turn must be related to the charge density of the subsystems, and in Figures 8(a) and 8(b) shows the contour maps showing the change of the charge density of the subsystems with the flow and informational trajectory. This also leads us to establish a conjecture about the existence of gradients of information. This permit us relate these gradients information changes with some parameters of chemical reactivity such hardness and softness.

In Figures 8(c) and 8(d) we present maps of the variation of hardness respect to the flow of information and informational trajectory. Figure 8(c) shows the surface that relates the change in hardness, softness in the information flow, in this case we notice that with the coordinates $0.36 < \eta < 0.385$ and $1.3 < S < 1.33$ we define the area where the information flow is maximum, this area also is an area of $-0.25 < RX < 0.25$, see Figures 5(a) and 5(b) thereby in this range of the intrinsic reaction coordinate important chemical and physical changes occurs, such as charge transfer which is closely related to the process of break and formation of a chemical bond.

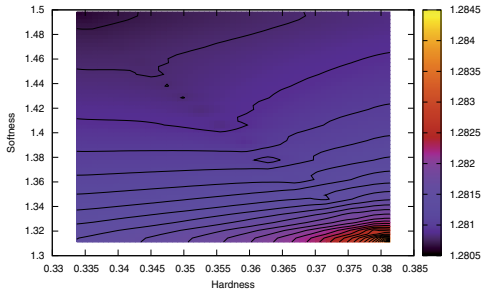
Finally, with this example we have tried to link information from a system that is subjected to a process with the physical and chemical changes. Thus, we have linked the concept of *information*, which is an epistemological concept completely with ontological concepts and



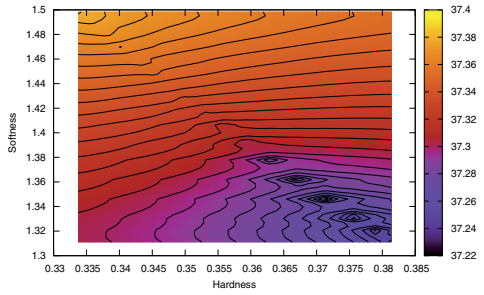
(a) Density charge of subsystems with informational flow



(b) Density charge of subsystems with informational trajectory



(c) Plane of the Hardness-Softness variation, and the informational flow.



(d) Plane of the Hardness-Softness variation and the informational trajectory.

Fig. 8. Contour maps of the informational flow and informational trajectories in the process $H_2 + H^- \longrightarrow H_2 + H^-$.

the solution concepts or interpretation of the results allows us feedback on these concepts in ontological terms, according to the author, abstract and more general.

By other part, is probable that today do not exist a ortodoxical definition of what actually is the *information*, beyond that presented by Shannon and its guidelines, criteria, characterization of it, among other things, the interpretation and the relationship with other concepts such as: energy, electron density, chemical reactivity parameters and many others need be discussed to try of establish a formal relation between concepts.

So, there is no doubt that both knowledge and the praxis and reality as knowledge scientific understanding and also, is it clear that information concept and the model itself is interdisciplinary or transdisciplinary. The concept and moreover, the model itself, promotes a systematic relation with causal analogies and parallelism with scientific knowledge, which transcends the framework of the source domain and extend in various directions, thus making the knowledge acquires an unusual resonance, as this, we believe it is feasible to complement the explanations of natural processes and natural systems.

This model is not intended that the manner of the old school, that using metaphysical substance, the particularities of the processes reveal themselves to us in the end as a progressive manifestation of homogeneous order or a unitary whole and absolute. It is simply to promote and implement a partnership scheme which promises analog route and cover knowledge in a way easier.
Thereby, one of the first goals of the Information Theory that can be invoked to find the link to Modern Chemistry is that through such a link is possible to restore a space for dialogue in this branch of science that in some cases lost or greatly reduced mended because the large number of specialized topics and concepts, thereby, with this scenario is possible consolidate a new vanguard line of research.

5. Conclusion

As the reader will appreciate the above discussion, Information Theory allows us to postulate that in any macroscopic system can identify an attribute inherent in its nature the electron density, energy and entropy. Each of them has a clear and precise physical meaning and its definition is accompanied by an operational rule to quantify it, that is, assigning numerical values. Thus we speak of universal axioms that govern natural processes.
However, we can make the question of what is now known about the informational entropy, which is related to the informational content of a classical or quantum observables and the same. So is there a universal axiom of incontrovertibly establish the relationship between information system with observables?
Thus, the study of physical processes and systems relatively simple physicochemical located within a general outline more or less well understood. But obtaining results that can compare with the experiment is a not a goal yet, in our point of view, it is therefore necessary to understand and apply a theory general enough to try to describe, explain and understand the behavior of natural systems.
For complex systems, like a biological, geophysical, or juts systems constitutes by a large particle number. We think that we are very far from being able to offer a convincing theory.
Finally we can say that now days the interest shifts to the study of systems far from equilibrium or complex processes that interact around through the exchange of matter, energy and information.

6. References

Nelson Flores-Gallegos and Rodolfo O. Esquivel. von Neumann Entropies Analysis in Hilbert Space for the Dissociation Processes of Homonuclear and Heteronuclear Diatomic Molecules. *J. Mex. Chem. Soc.*, Vol. 52, No. 1, (2008) 19-30, ISSN 1870-249X.

Rodolfo O. Esquivel, Nelson Flores-Gallegos, Cristina Iuga, Edmundo M. Carrera, Juan Carlos Angulo, Juan Antolín. Phenomenological description of the transition state, and the bond breaking and bond forming processes of selected elementary chemical reactions: an information-theoretic study. *Theor. Chem. Acc.*, Vol. 240, 445-460 (2009), ISSN 1432-881X.

Edmundo M. Carrera, Nelson Flores-Gallegos, Rodolfo O. Esquivel. Natural atomic probabilities in quantum information theory. *Journal of Computational and Applied Mathematics.*, Vol. 233, 1483-1490 (2010), ISSN 0377-0427.

V. Vedral, The Role of relative entropy in quantum information theory. *Rev. Mod. Phys.*, Vol. 74, 197-234 (2002), ISSN 0034-6861.

C. E. Shannon, A mathematical theory of communication. *The Bell System Technical Journal*. Vol. 27, 379-423, 623-656, July, October (1948), ISSN 0005-8580.

Reed A. E., Weinstock R. B., Weinhold F. Natural Population Analysis. *J. Chem. Phys*. Vol. 83, No. 2, (1985) 735-747 , ISSN 0021-9606.

P. O. Löwdin. Quantum Theory of Many-Particle Systems I. Physical Interpretation by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configuration Interaction. *Phys. Rev.*, Vol. 97 (1955) 1474-1489, ISSN 0031-9007.

G. Bruhn, E. R. Davidson, I. Mayer, and A. E. Clark. Löwdin Population Analysis With and Without Rotational Invariance. *Int. Jour. of Quantum Chem.*, Vol. 106 (2006) 2065-2072, ISSN 0020-7608.

Nelson Flores-Gallegos *Teoría de información cuántica como lenguaje conceptual en Química*. Tesis Doctoral. Universidad Autónoma Metropolitana-Iztapalapa. México D.F. (2010).

M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr. T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewsk, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople 2004, Gaussian Inc. Wallingford CT, Gaussian 03 Revision D.01.

Glendening, E. D., Badenhoop, J. K., Reed, A. E., Carpenter, J. E., Bohmann, J. A., Morales, C. M., Weinhold, F., NBO 5.0, Theoretical Chemistry Institute 2001, University of Wisconsin, Madison.

Schaftenaar, G., Noordik, J. H. MOLDEN: a pre- and post-processing program for molecular and electronic structures., *J. Comput. Aided Mol. Design* Vol. 14 (2000) 123-134, ISSN 0920-654X.

Parr, R. G. & Yang, W. (1989). *Density-Functional Theory of Atoms and Molecules*, Oxford University Press. ISBN 0-19-504279-4, United States of America.

Guiaşu, S. (1977). *Information Theory with Applications*, MacGraw-Hill International Book Company. ISBN 0-07-025109-6, Great Britain.

von Neumann, J. (1955). *Mathematical foundations of quantum mechanics*, Princeton, N. J.: Princeton University Press. L.C.Card 53-10143, London: Geoffrey Cumberlege.

Johnson, B. G., Gonzales, C. A., Gill, P. M. W. Pople, J. A. A density functional study of the simplest hydrogen abstraction reaction. Effect of self-interaction correction. *Chem. Phys. Lett.*, Vol. 221 (1994) 100-108, ISSN 0009-2614.

Ziesche, P. Correlation Strengh and Information Theory., *Int. J. Quantum Chem.* Vol. 56 (1995) 363-369, ISSN 0020-7608.

Parr, R. G., Ayers, P. W. and Nalewajski, R. F., What is an atom in a molecule?. *J. Phys. Chem.*, Vol. 109 (2005) 3957-3959, ISSN 1089-5639.

Fano, U., Descriptions of States in Quantum Mechannics by Density Matrix and Operador Techniques., *Rev. Mod. phys.*, Vol. 29 (1957) 74-93, ISSN 0034-6861.

McWeeny, R., Some Recent Advances in Density Matrix Theory., *Rev. Mod. Phys.*, Vol. 32 (1960) 335-369, ISSN 0034-6861.

Davidson, E. R. (1976). *Reduced Density Matrices in Quantum Chemistry.*, Academic Press Inc. (London), ISBN 0-12-205850-X.

Molmer, K. Sorensen, A. Multiparticle Entanglement of Hot Trapped Ions. *Phys. Rev. Lett.*, Vol. 82 (1999) 1835-1838, ISSN 1079-7114.

Cirac, J. I. and Zoller, P., Quantum Computations with Cold Trapped Ions. *Phys. Rev. Lett.*, Vol. 74 (1995) 4091-4094, ISSN 1079-7114.

Sorensen, A. and Molmer, K., Qantum Computation with Ions in Thermal Motion. *Phys. Rev. Lett.*, Vol. 82 (1999) 1971-1974, ISSN 1079-7114.

Kreuter, A. Becher, C. Lancaster, G. P. T. Mundt, A. B. Russo, C. Haffner, H., Roos, C., Eschner, J., Schmidt-Kaler, F., Blatt, R., Spontaneous Emission Lifetime of a Single Trapped Ca^+ Ion in a High Finesse Cavity., *Phys. Rev. Lett.*, Volume 92 (2004) 203002-203006, ISSN 1079-7114.

Schlosser, N., Reymond, G., Protsenko, I., Grangier, P., Sub-poissonian loading of single atoms in a microscopic dipole trap., *Nature*, Vol. 411 (2001) 1024-1027, ISSN 0028-0836.

Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M., Wineland, D. J., Demostration of a Fundamental Quantum Logic Gate. *Phys. Rev. Lett.*, Vol. 75 (1995) 4714-4717, ISSN 1079-7114.

Tittel, W., Brendel, J. Gisin, B. Herzog, T., Zbinde, H., Gisin, N., Experimental demostration of quantum correlations over more than 10 km., *Phys. Rev. A.*, Vol. 57 (1998) 3229-3232, ISSN 1050-2947.

Jian-Wei, P., Bouwmeester, D., Daniel, M., Experimental test of quantum nonlocality in three-photon., *Nature*, Vol. 403 (2000) 515-519, ISSN 0028-0836.

Ladd, T. D., Goldman, J. R., Yamaguchi, F., Yamamoto, Y., All-Silicon Quantum Computer., *Phys. Rev. Lett.*, Vol. 89 (2002) 017901-017905, ISSN 1079-7114.

Braunstein, S. L., Caves, C. M., Jozsa, R., Linden, N., Popescu, S. Schack, R., Separability of Very Noisy Mixed States and Implications for NMR Quantum Computing., *Phys. Rev. Lett.*, Vol. 83 (1999) 104-107, ISSN 1079-7114.

Beveratos, A., Brouri, R., Gacoin, T., Villing, A., Poizat, J.-P., Grangier, P., Single Photon Quantum Cryptography., *Phys. Rev. Lett.*, Vol. 89 (2002) 187901-187905, ISSN 1079-7114.

Jennewein, T., Simon, C., Weihs, G., Weinfurter, H. Zeilinger, A., Quantum Cryptography with Entangled Photons., *Phys. Rev. Lett.*, Vol. 84 (2000) 4729-4732, ISSN 1079-7114.

Gisin, N., Ribordy, G., Tittel, W. Zbinden, H., Quantum cryptography., *Rev. Mod. Phys.*, Vol. 74 (2002) 145-195, ISSN 0034-6861.

Horodecki, M., Oppenheim, J., Winter, A., Partial quantum information., *Nature*, Vol. 436 (2005) 673-676, ISSN 0028-0836.

Divincenzo, D., Horodecki, M., Leung, D., Smolin J., Terha, B., Locking classical correlation in quantum states., *Phys. Rev. Lett.*, Vol. 92 (2004) 067902-067906, ISSN 1079-7114.

Horodecki, M., Oppenheim, J., Horodecki, R., Are the laws of entanglement theory thermodynamical?., *Phys. Rev. Lett.*, Vol. 89 (2002) 240403-240407, ISSN 1079-7114.

Horodecki, R., Horodecki, P., Horodecki, M., Violating Bell inequality by mixed spin $-\frac{1}{2}$ image states: necessary and sufficient condition., *Phys. Lett. A.*, Vol. 200 (1995) 340-344, ISSN 0375-9601.

Hoderecki, R., Hoderecki, M., Information-theoretic aspects of inseparability of mixed states., *Phys. Rev. A.*, Vol. 54 (1996) 1838-1846, ISSN 0375-9601.

Horodecki, M., Horodecki, P., Horodecki, R., Separability of mixed states: necessary and sufficient conditions., *Phys. Lett. A.*, Vol. 223 (1996) 1-8, ISSN 0375-9601.

Horodecki, P., Separability criterion and inseparable mixed states with positive partial transposition., *Phys. Lett. A.*, Vol. 232 (1997) 333-339, ISSN 0375-9601.

Horodecki, M., Horodecki, P., Horodecki, R., Mixed–State Entanglement and Distillation: Is there a "Bound" Entanglement in Nature?., *Phys. Rev. Lett.*, Vol. 80 (1998) 5239-5242, ISSN 1079-7114.

Horodecki, R., Horodecki, P., Horodecki, M., Horodecki, K., Quantum Entanglement., *Rev. Mod. Phys.*, Vol. 81 (2009) 865-942, ISSN 0034-6861.

Aspect, A., Quantum Mechanics: To be or not to be local., *Nature* Vol. 446 (2007) 866-867, ISSN 0028-0836.

Timothy, Y., Ayman, F., Abouraddy Bahaa E., A., Saleh Malvin C., Teich., Experimental Violation of Bell's Inequality in Spatial-Parity Space., *Phys. Rev. Lett.*, Vol. 99 (2007) 170408-1704012, ISSN 1079-7114.

Einstein, A., Podolsky, B., Rosen, N., Can Quantum–Mechanical Description of Physical Reality Be Considered Complete?, *Phys. Rev.*, Vol. 47 (1935) 777-780, ISSN 1943-2879.

Bell, J. S., On the Problem of Hidden variables in Quantum Mechanics., *Rev. Mod. Phys.*, Vol. 38 (1966) 447-452, ISSN 0034-6861.

Seeger, R. and Pople, J. A., Self-Consistent Molecular Orbital Methods. 28. Constraints and Stability in Hartree-Fock Theory., *J. Chem. Phys.*, Vol. 66 (1977) 3045-3050, ISSN 0021-9606.

Bauernschmitt, R. and Ahlrichs, R., Stability analysis for solutions of the closed shell Kohn-Sham equation., *J. Chem. Phys.*, Vol. 104 (1996) 9047-9052, ISSN 0021-9606.

11

Quantum Mechanics Design of Two Photon Processes Based Solar Cells

Abdennaceur Karoui and Ara Kechiantz
Photovoltaic Nanotechnology and Nanosensors Laboratory,
Shaw University, Department of Natural Sciences and Mathematics
USA

1. Introduction

The photoelectric effect has been utilized in a wide range of optoelectronic sensors and solar energy conversion devices. To respond to today's colossal energy demand there is a need for doing better than using this effect via the utilization of simple semiconductor homojunction based devices, since such structures cannot use all photons available for conversion into electricity. Actually, it is the inefficient response of the material to the various solar photons and the behaviour of photo-generated carriers that bar photovoltaic (PV) cells from fully harvesting the solar energy impinging on its surface. It is well known that there is no semiconductor that can be deployed for making a simple and low cost p-junction based solar cell that converts the entire solar spectrum. Hence, the development of efficient photovoltaic devices has been going through various technology pathways, one of which has led to a variety of multi-junction and cell stacks, i.e., tandem cells; each sub-cell is dedicated to converting a portion of the solar spectrum. Moreover, the conventionally chosen three sub-cell tandem scheme is being extended to stacks with larger number of sub-cells, regardless of the viability of the production technology.
During the last five decades, several theoretical PV energy conversion efficiency thresholds have been established, each being for one type of PV device design. However, none has been reached in spite of the deployed tremendous efforts; the implementation turns out to be challenged by some specific constrains. High efficiencies attained thus far incorporated advances in material refinement, sophisticated cell design, state of the art fabrication technologies, and more importantly a relatively good understanding of the complicated and intertwined physics of charge carrier generation, recombination, and transport. However, for making new generation of high efficiency solar cells, new physics and technology paradigms have to be discovered. Current barriers to high conversion efficiencies are due to current understanding of the physics of photoelectric effect and related phenomena. Breaking those barriers, which would impact photovoltaic technologies, needs refinement of the physics of photoelectric effect with consideration of new materials and new device designs. For instance, development of "ideal" third generation solar cells requires understanding of the early stage of light-matter interaction mechanisms that occur at the nanoscale domains and the related quantum effects. Actually, the improved understanding of these phenomena is repositioning the used linear photoelectric effect occurring in homogenous and continuous planar shallow structure to one that occurs in a volumic

structure with nanoscale features. These induce rather non-linear effects entangled with quantum effects. Additionally, a precise account of the charge carriers within the cell has to be accomplished. Inclusive to that need are (i) the improvement of photon absorption over a wider spectral range by adequately engineering the semiconductor optical properties, (ii) immediate use of photo-generated electron-hole pairs, (iii) the local charge carrier separation and collection, and (iv) transport to remote electric contacts; all of which should be studied at the nanoscale level.

As the quantum processes and the discrete nature of electron transitions are fundamental to photoelectric effect, this chapter focuses on (i) reviewing light-matter interaction quantum processes that are directly or indirectly involved in the photoelectric effect, (ii) discussing those that are useful for third generation PV cells, and (iii) highlighting recent research results by the author. Fundamental quantum processes initiated by photon absorption and their contribution to the solar energy conversion will be reviewed in connection to the electron energy structure of the absorber material. New ideas relevant to absorbers within third generation solar cells will be presented. The nature of light-electron interaction is scrutinized and elementary processes of light absorption, atom excitation, atom ionization, photon excess energy dissipation, emitted charge carrier pairs, carrier transport and/or recombination are discussed in terms of exchanged energy, momentum transfer and photon-electron interactions in nanometer and atomic scales. We will assess the major losses that have quantum bearings and elucidate their effects on third generation solar cells. We will present a novel solution, which employs two-photon absorption . This is a cooperative non-linear process that occurs in two sequential stages. It employs two photons, that have an energy lower than the semiconductor bandgap. The two-photon absorption produces a net optical transition between valence and conduction bands through an Intermediary energy Band (IB) laying within the bandgap of the host material. Photoelectric effect employing two photon processes is a research area of strong scientific interest that drives science and technology research towards making new solar cells among third generation. We will delve into the reasons as to why the two-photon absorption is still not optimally used to make efficient solar cells. In single-photon absorption by indirect semiconductors, phonon cascades are usually involved. The energy and momentum conservation during such photon absorption sequence obligates one to look at the thermodynamics of a solar device and the factors of thermalization, a major effect that reduces the quantum efficiency of the photoelectric effect. The processes and factors controlling the two-photon absorption, which involves virtual or real intermediate states, are presented. Next, quantum phenomenon involved in nanostructured PV device operation is electron tunnelling through energy barriers that spatially separates confined electronic states.

The second part of the chapter will present modelling of new materials that are designed to promote two photon absorption processes in solar cells. The studied system is a circular chain of QDs embedded in nanostructured material, a set of QD chains attached to the inner walls of pores in porous silicon (Karoui & Kechiantz, 2011). Confinement of electron wave-functions breaks the initially continuous electron energy spectrum into discrete energy levels. There are some apparent differences of this effect in QDs and the well understood atom systems, although the underlying physics is the same. We will discuss these differences in relation to their relevance to the studied nanostructured PV material. It is argued that while decaying beyond the potential barrier associated with the QD-host material interface, electron wave-functions penetrate into the host material, which enables overlapping of confined electron wave-functions. We will show how a wave-function that

spills over the interface enables creation of energy bands in the host solid out of the energy levels of the QDs embedded in the nanostructured PV system.
It should be noted that the intention of this chapter is neither to report on a mere quantum mechanics calculation, nor to give an account on photoelectric applications, but to bridge the gap between the physical quantum processes driving the photoelectric effect. This has been mostly over simplified in conventional PV applications. It has become known that modern PV device engineering focuses on increasing efficiency of those devices thereby targeting the 65% limit defined by Luque and Marti. This high goal appears to require the use of nanostructured semiconductor systems. In that context, the optoelectronic quantum processes can no longer be neglected. Therefore, to obtain the most out of the photoelectric effect and overcome the new challenge of 65% conversion efficiency, one needs to deepen the physics knowledge on these processes and put them into the context of real material systems and devices. Detailed experimental and theoretical studies related to those quantum processes occurring in the presented novel PV design will be published by the author of this chapter.

2. Quantum mechanics aspects of solar energy conversion

2.1 Interaction photon-solids

2.1.1 The energy-moment space in semiconductors

The electron energy-momentum *(***E-k***)* relationship is well known to be fundamental for obtaining the physical properties of solids. It governs many optical and electronic related phenomena in solids, in particular photoelectric related processes. This relationship is very sensitive to the periodicity of the crystal potential in which the electron propagates. The *(***E-k***)* relationship is obtained in the one-electron approximation by solving the Schrodinger equation for electron in effective periodic crystal field. The crystal periodic field renormalizes the electron momentum and defines a new complex energy-momentum *(***E-k***)* relationship, given in *Brillouin zone,* that essentially differs from the simple parabolic energy-momentum relationship shown in Figure 1(a) for free electron with m_e the electron mass in the free space. Figure 1(b) and (c) show the very well-known energy-momentum for Si and GaAs, respectively, to remind the complexity of the energy band structure of solids. The *(***E-k***)* relationship is a multifold function of momentum, since the same momentum refers to different electron energies. The *(***E-k***)* is chiefly parabolic near the crystal lattice symmetry points $(\Gamma, L, X, ...)$ only, but the concavity variation is characteristic, in a first order approximation, of crystal symmetry in these points and the band energy variation in the vicinity of the considered point. This variation commensurate with the *electron effective mass*.
Because *Pauli Exclusion principle* limits the occupation of the same energy-momentum-spin electronic state to one electron, some energy bands are completely filled with electrons while others are either empty or partially filled in the vicinity of the symmetry points as shown in Figure 1(b) and (c). Like free electrons, in the vicinity of the symmetry points, the *energy and* to be conserved even though the interaction with other particles, in the solid, can be very strong and can occur in stages. All interactions are bound by the complex *(***E-k)** relation.
Changes in electron energy-momentum *(***E-k***)* relationship in solids are essentially *quantum mechanical effects*, and can be found by solving the Schrödinger Equation.

2.1.2 Absorption and emission, direct and indirect electron transitions

Engineering semiconductors that fully absorb the entire energy of incident photons over the entire solar spectrum constitute an ambitious goal. This can be at best approached by

nanotechnology and used for by third generation PV cells. To better utilize the photoelectric effect, solid state *fundamental processes* must be addressed individually at various time, space, and energy scales. These basically include *photon absorption* and *emission* and subsequent *electron-hole generation, recombination, separation,* and *collection*. Additionally, other processes such as the involvement of phonon cascades, surface plasmons and polarons,… must be taken into account for nanoscale investigations, Single or cascade of such processes move electrons from one energy level to another, which brakes the equilibrium of the electronic system and leads to a cascade of other processes that ultimately bring the system back to its thermodynamic equilibrium.

The blue arrows and red curly-arrows in Figure 1(b) and (c) illustrate photon absorption resulting in the electron transfer from the valence band into the conduction band. The reverse process is the electron transfer from the conduction band into the valence band, which results in photon emission (illustrated by green arrows and red curly-arrows in Figure 1(b) and (c)).

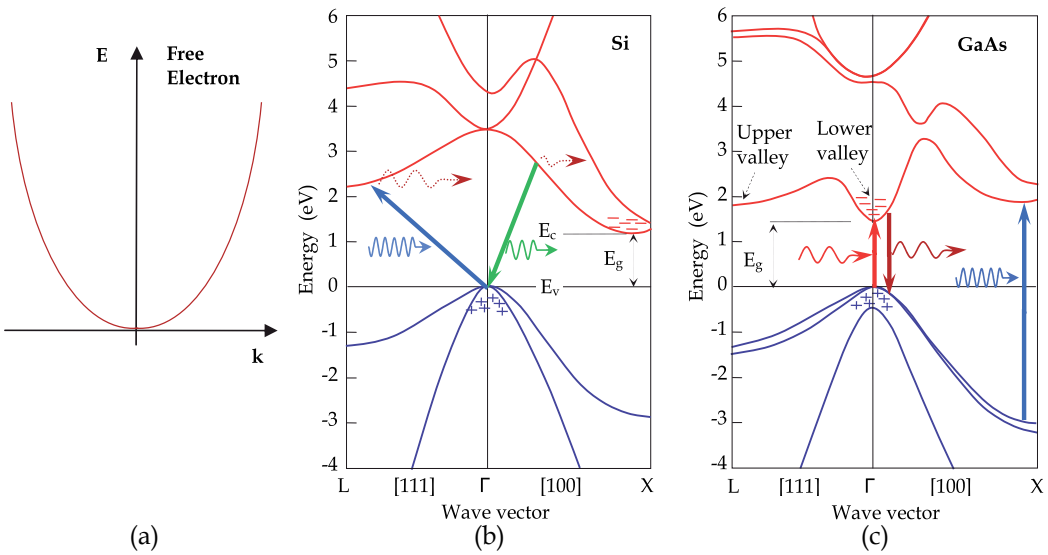


Fig. 1. Electron energy-momentum (E-k) relationship and electron transitions in: a) the free space; b) the indirect bandgap silicon; c) the direct bandgap GaAs.

Since photon momentum is very small compared to that of electron, the momentum conservation allows only electron transitions between energy levels with practically no electron momentum variation (thus the same wave vector) in the Brillouin zone. The vertical blue and green arrows display such *direct electron transitions* in Figure 1(c) for a direct bandgap semiconductor, here GaAs. When there is not an empty state that enables photon induced electron transition with momentum conservation, phonon assistance is required to ensure momentum conservation. *Indirect electron transitions* must involve both phonon and photon, as illustrated in Figure 1(b), where dotted curly-arrows represent phonons.

2.1.3 Radiative and non-radiative recombination

Charge Carrier Recombination is the electron-hole annihilation in semiconductors, where the electron returns from the high-energy electronic state to an empty low-energy electronic

state, for instance from the conduction band to the valence band R*adiative recombination* results in photon emission with conservation of both energy and momentum. However, recombination can also happen without photon assistance, which is known as *non-radiative recombination*. For instance, phonon can assist recombination through impurity related levels within the semiconductor bandgap.

Recombination is very important in photovoltaics and for other optoelectronic devices as well. While photon absorption generates photocurrent, electron-hole recombination suppresses the photocurrent and generates dark current in these devices. Both absorption and recombination processes may become dominant and even complex in low cost photovoltaic materials, where a wide range of defects usually are tolerated, and in semiconductor structures when multi-particle processes like Auger recombination assist electron relaxation.

2.1.4 Shockley–Read–Hall (SRH) recombination

Single and cluster impurities in the crystal either in interstitial or in lattice positions promote electron transition between bands passes through the energy level created within the band gap. During transition of charge carriers from valence band to conduction band, and vice-versa, the carriers take available states and most often those created by impurities. In satisfying the energy and momentum conservation across the entire process, the impurity state can absorb differences in momentum between the carriers. The process of generation and recombination via impurity levels is the dominant in indirect bandgap materials, such as silicon. This process can also be significantly important and dominate the charge carrier dynamics in direct bandgap materials particularly for very low carrier densities. The excess energy is dissipated in the form of lattice vibrations (absorbed phonon by the material give away their energy in the form of thermal energy). Impurities and crystal point defects and clusters create energy levels within the band gap. Figure 2 shows the known energy levels in silicon (Sze & Irvin, 1968) generated by metal impurities and crystal point defects. Light elements and neighbors to the periodic table fourth column mostly induce shallow levels, which are used beneficially as dopants, whereas transition metals for instance induce deep level traps. The latter actively participate in non-radiative carrier recombination and thus are detrimental charge carriers killers.

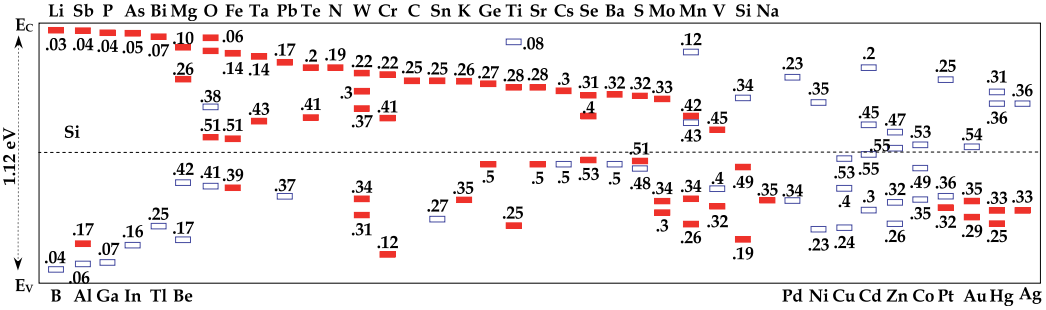


Fig. 2. Discrete energy levels of different impurities in Si energy bandgap. (Sze & Irvin, 1968)

2.1.5 Auger recombination

This type of recombination contributes to the loss of electrons promoted to the conduction band by photon absorption. Under certain conditions this recombination becomes dominant thereby degrading the overall conversion efficiency of PV cells and photo-sensors.

Auger recombination is a non-radiative process that involves three mobile carriers two of which recombine and while vanishing they transfer their energy to a third carrier, the latter will continue to reside in the same band as a bound electron. Figure 4 illustrates the Auger recombination of photo-generated holes in p-type doped Si. Auger recombination process is controlled by majority carriers (here holes) since the recombination rate is proportional to np^2, where n and p are respectively the electron and hole concentrations. The energy transferred to the third carrier increases its energy without ascension to a higher energy band. The stored energy by the third carrier is then lost to thermal vibrations (phonons). This process is fast and involves three-particle interaction, the third is involved at a high-energy unstable state. The process is effective in non-equilibrium conditions, when the carrier density is very high, that is either in heavily doped semiconductor or under high injection levels. The green arrow in Figure 3 (b) represents the electron-hole annihilation while the red dashed arrow shows the energy transferred to another electron in the valence band. The blue arrow displays electron transition to a higher energy state in the valence band, which generates a "hot" hole. The readiness of the third particle for Auger Recombination makes the probability of this process rather low. The probability being proportional to np^2, as stated above in the case of p-type semiconductor, shows that Auger recombination rate swiftly grows with carrier density. This feature makes the Auger recombination a determinant factor in heavily doped solar cells, in particular in the device active zone and the surface as well as under concentrated sunlight irradiation when the density of injected or generated carriers is very large (Sinto & Swanson , 1987). For instance, when the density of photo-carriers is above $10^{18} cm^{-3}$ (Dziewior & Shmid, 1977, Svantesson & Nilsson, 1979). Auger process becomes very effective and dominates the recombination rate in high quality silicon solar cells (Green, 1984), which achieved the highest records of conversion efficiencies , and more as the cell efficiency is increased. Hence, device designs that alleviate Auger recombination at surface and interfaces must be thought out, particularly for concentration solar cells.

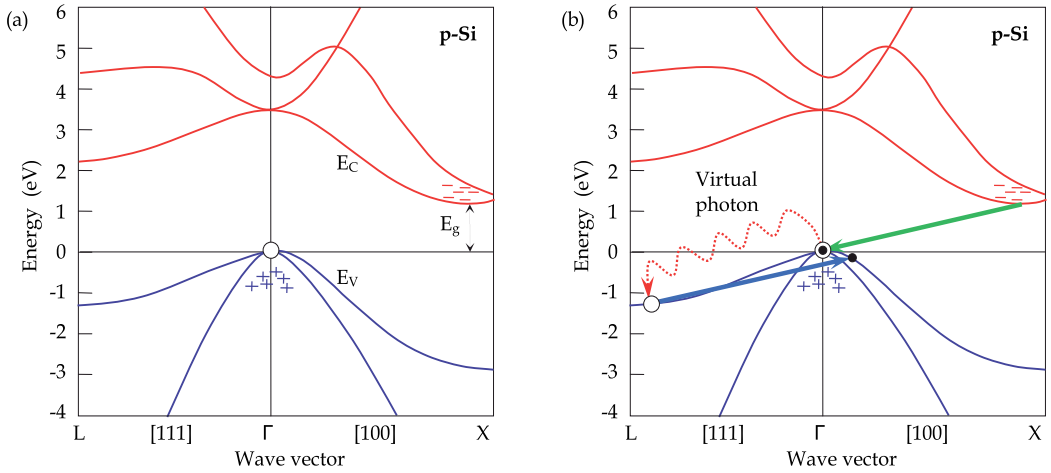


Fig. 3. Photo-generated holes in p-type Si (a), and Auger recombination of that hole (b).

2.2 Major quantum energy losses during carrier photo-generation and transport

For developing high efficiency PV devices one must find new approaches for absorbing the entire energy of incident photons, preventing internal losses, and collecting all produced

electrons. It is well known that there is a slate of energy losses inside and outside the device and the major internal losses have quantum origin. Hence, during the past two decades quantum mechanics has been extensively used for devising new generation of solar cells that minimize those losses. For that purpose, fundamental optoelectronic processes in semiconductors have been revisited aiming at fully exploiting the photon energy available in the solar spectrum as well as the handling of generated charge carriers in the PV devices. As we investigate the photoelectric effect at the nanoscale level, the generation and transport of charge carriers are looked at in details with the associated quantum effects.

2.2.1 Energy excess of photo-generated electrons, and thermalization

The fundamental solid state physics mechanisms that are responsible for energy losses in solar cell are shown in Fig. 4. The internal losses are basically due to the due to the "misfits" of energy-wavevector of incident photons with the electronic band structure characteristic of the material used to make the PV device. The problem is further aggravated as the misfit extends over the entire solar spectrum and thus difficult to correct. Radiative and non-radiative electron transitions between electronic states steadily take place along electron diffusion paths towards the p-n junction. The photoelectron intra-band relaxation processes, which involve thermal photons and phonons, become abundant as the electronic levels within conduction and valence energy bands become dense. Interactions of the material with the photons impinging the PV device surface involves extracted electrons, intermittently produced photons, and phonons, all of which occur via a cascade of radiative and non-radiative intra-band transitions that causes electron and hole thermalization in conduction and valence bands, respectively.

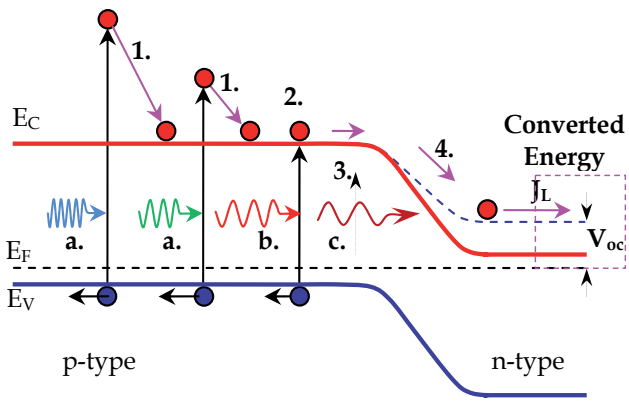


Fig. 4. Photoelectron excess energy leading to energy loss processes in a solar cell made of one p-n-junction; a. High energy photon absorbed, b. photons that have energy exactly equal to the absorber bandgap, c. low energy photons that are usually transmitted, 1. Thermalization via intraband transitions, 2. Absorption without loss, 3. Electronic energy loss during transition through depletion region (i.e., separation of e-h pair).

The three main fundamental processes of energy loss in solar energy conversion devices are schematically shown in Fig. 4. Photon absorption (a. and b. photons) usually produces electrons (electrons of type 1. or 2.) whose energy is larger than or equal to the energy they contribute into the electricity generated by the solar cell. Because the photoelectrons (1.) are in a non-equilibrium thermodynamic state with the electronic system and the crystal lattice,

they swiftly lose the excess energy through thermalization in intra-band inelastic scattering and collision processes with equilibrium electrons and phonons. The arrows attached to electrons labeled (1.) in Fig. 4 illustrate such intra-band electron relaxation processes (thermalization). Also there is a wide range of low-energy solar photons (pin pointed with the arrow labeled (3.) that do not have enough energy for transferring electrons from the valence band into the conduction band; these photons are conventionally transmitted through the material. However, low energy photons (c.) can still promote transitions of type (3.) through levels in the bandgap or participate in two photon absorption mechanism.
Because photoelectron generation uses less than the absorbed energy, intra-band loss mechanisms are unavoidable. After intra-band thermalization the photo-generated electron still retains more energy than what it uses in contributing to the output electricity. The photoelectron loses the difference of these energies during electron-hole separation, shown in Fig. 4 by the arrow labeled (4.), while transiting through the space charge region.
Figure 5 elucidates the impact of the internal losses using the AM1.5 solar irradiance impinging a silicon PV device. The usable energy under Shockley-Queisser limit is 31% of the incident light over the entire spectrum. The convertible energy amount is in regions (1.) and (2.) of Figure 4, which are the result of the effective absorption corresponding to (1.) and (2.). Likewise the converted amount by the state of the art silicon cell is 25% occurs in the same region of the spectrum. High energy photons in the Near-UV-Vis are absorbed at the material surface leading to a high recombination rate at the surface, and a significant loss

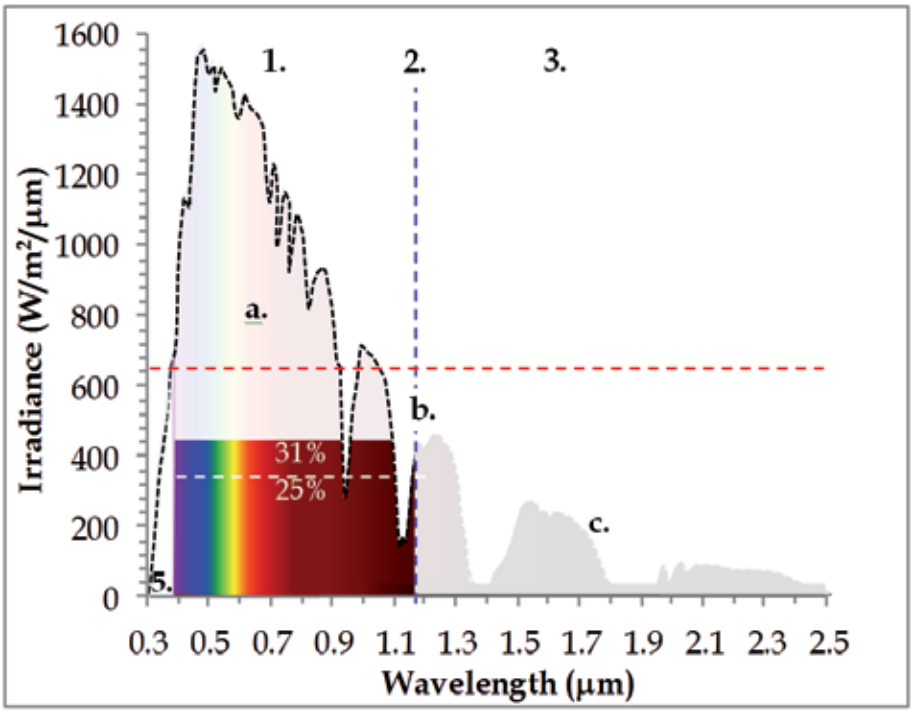


Fig. 5. Solar energy density as a function of the wavelength (AM1.5). Portions that cannot be converted by silicon single p-n homojunction solar cell is represented by greyed areas (1.), for photons that have energy lower than the bandgap, and (2.) for high energy photons, as a result of losses at the level of photon-solid interactions.

labeled (5.) in Fig. 4, this is evaluated to about 8%. Surface passivation has been used to reduce the surface recombination. These loss mechanisms result in a sizable amount of solar energy that cannot be converted in electricity in conventional single p-n-junction solar cells. several energy losses that do not have a quantum origin are also significant and undesirable but difficult to avoid. These are out of the scope of this chapter; nevertheless, it is useful to quickly overview them. Let's start with light reflection at the solar cell surface. Silicon reflectance reaches 31% in the visible range and more than 50% in the UV. Methods have been developed to reduce such losses, which consist in increasing the optical path at the surface by engaging a multiple reflection process. Silicon surface texture, nitrogen annealed and etched silicon (Karoui & Zhang, 2002), and sprayed porous silicon proved to be effective and low cost methods for reducing reflection losses. However, a non-resolvable loss by reflection remains, this is labeled in Fig. 5; and is evaluated to about 8% of the total incident energy.

The second loss stems from the emission of blackbody radiation by solar cells. The equilibrium temperature of a cell in operation leads to blackbody radiation emission. The temperature baseline is that of the surrounding. As the temperature of the cell increases during absorption of sunlight, the cell blackbody radiation increases and ultimately attains thermodynamic equilibrium with the incident radiation, leading to a temperature roughly between 80 and 95°C. The blackbody emitted radiation energy is lost to the ambient and thus cannot be converted to electricity by the cell. The loss by silicon cell via blackbody radiation can reach 70-80% incoming solar radiation energy.

The third category of loss is accidental. It arises when dissipation of electric energy from a PV cell array into low performing cells occurs. In this case the energy is dissipated within such cells in thermal form, which increases the temperature of the cell and thus the blackbody radiation. Henceforth, the associated blackbody loss is increased. Such loss is referred to as hot spot within a cell array. It leads the PV cell to operate at lower conversion efficiency or possibly to a catastrophic breakdown of the cell and the cell array.

2.2.2 Phonon assisted electron transitions and phonon bottleneck effect limiting charge carrier recombination in quantum dots

Phonon quantum energy is very small relative to semiconductor energy bandgap. This makes phonon assistance in electron interband transitions of in direct and ideal semiconductors likely, because of energy conservation. However, in real life semiconductors, decorated extended defects (e.g., dislocations,...) *single and clusters of* point defects (i.e., dissolved impurities, vacancies, self-interstitials,…) spread in the bulk and at the surface generate states and energy levels within energy bandgap. These energy levels are discrete and spatially localized in such non-ideal semiconductors. Both thermal photons and phonons can assist photon initiated electron inter-band transitions between such levels if energy difference between initial and final levels is small and the separation between these defects is low such it prevents any screening. Moreover, cascades of alternating phonon assisting and tunneling electron transitions can result in non-radiative inter-band electron-hole recombination.

Impurities, point defects, crystalline extended defects and their interactions have been assessed during the last three decades. In that respect, silicon is probably the most diagnosed semiconductor. For cost reasons, conventionally used photovoltaic materials are of much lower quality than microelectronic and optoelectronic materials. Acceptable energy level concentrations, and impurity states have been the focus for both academia and photovoltaic industry. Sufficiently efficient silicon solar cells have shown a certain tolerance

margin for impurities and charge carrier lifetime killers in the bulk and at the surface and have defined the baseline for low cost material growth technologies. In such materials non-radiative recombinations are dominant.

Intensity of non-radiative recombination increases with the defect electronic density of states associated with precipitates, decorated dislocations, vacancies, the hardly controllable dissolved impurities (in interstitial or substitution sites) etc., both in the bulk material and at the surface. Radiative inter-band transitions are only possible in solar cells made from ultra-clean materials. Shockley and Queisser have applied this property of clean semiconductors for evaluating limiting conversion efficiency in the best conventional one-junction solar cell (Shockley & Queisser , 1961). More recently, Luque and Marti have used the same property of clean semiconductors for evaluating the utmost conversion efficiency limits of IB solar cells (Luque & Marti, 1997). This is discussed in more details in the following sections.

Consecutive emissions of optical phonons is the main mechanism of photoelectron thermalization. The energy of an optical phonon in silicon is roughly about $50\ \text{meV}$. The characteristic time of such intra-band relaxation is about $10^{-12}s$ for photoelectrons in conventional solar cells. However, in dye or quantum dot (QD) sensitized solar cells, the relaxation process of photoelectrons may become slower. Such cells exploit the sensitizer discrete energy levels to enable additional photoelectron generation in the host material. The reduction of photoelectron relaxation, known as the "*phonon bottleneck*" effect in QDs, results from the lack of optical phonons that match separation of discrete energy levels in QDs (Guyot-Sionnest et al., 1999; Sun et al., 2006). The "phonon bottleneck" could be favorable for photovoltaic devices like solar cells and infrared and terahertz photodetectors operating at room temperature (Zibik et al, 2009) since radiative transitions dominate photoelectron relaxation in such QDs.

2.3 Solar energy conversion limits

2.3.1 Limiting factors for photon conversions and device inherent losses

It is important to contain the photoelectric efficiency under certain conversion limit. This essentially depends on the way the photoelectric effect is exploited and hence the device design and the type of used material. Furthermore, fabrication technologies involve detrimental factors to the exploitation of the photoelectric effect. The various fundamental limits for photon energy conversion will be debated, namely those of thermodynamics origin, detailed balance limit defined by Shockley and Queisser, and the new limit defined by Luke and Marti. The quantum mechanics based discussion will focus on the generation of charge carrier pairs as well as their recombination and how the photon energy can be better used to increase the quantum efficiency of PV devices. Electronic confinement in nanofeatures have become essential for third generation solar cells, thus needs extensive clarification. We will look at how the confinement affects light absorption and the overall conductivity, as these are key matters for high efficiency solar cells.

2.3.2 Photon recycling through optical and electronic transition cascades

The "photon recycling" terminology refers to transformation dynamics of absorbed photon-mode-fluxes through photovoltaic and optical sensor devices. Energetic photon generates electron-hole pair, which in turn produces new photon after the pair recombines radiatively. Because the new photon has energy above or equal to the band gap, it can create another electron-hole pair by re-absorption elsewhere in the cell, and so on and so forth. The "recycled photon" looks like a "virtual photon" in Feynman diagrams. Similar diagrams for

photon-solid interactions relevant to photovoltaics and "photon recycling" processes are presented in Figure 6, where:

- the dashed curled line represents the "photon recycling" process,
- the light blue and green arrows are incoming absorbed photons and the orange and dark blue are emitted photons exiting the PV device,
- CBC and VBC are contacts to conduction and valence bands, respectively, and
- the solid and the dashed brown double-line arrows represent the electron and hole currents, respectively.

At the end of recycling one can obtain one of the following mechanisms: a) incoming photon yields photocurrent; b) electron injected from a contact yields dark current and another photon emission; c) incoming photon causes emission of another photon, with a frequency corresponding to energy adjusted to the bandgap, and d) injected electron or hole through CBC and VBC are contacts to conduction and valence bands, respectively, are absorbed and energy is dissipated in the device. The recycling can be either initiated by photon absorption or charge carrier injection and ends either:

- when the built-in field breaks the generated pair and separates the electron in conduction band from the hole in valence band so that the electrode linked to the conduction band (CBC) in n-doped region absorbs the electron and the electrode linked to the valence band (VBC) in p-doped region absorbs the hole, that is case (a) in Fig. 6. The net result is a separation of electron-hole pairs. The minority carrier currents are shown in Fig. 6 with a brown double-line arrow, solid for the electrons and dashed for the holes.
- or with production of a photon that exits the device, cases (b) and (c) in Fig. 6.

Also electron (or hole) injection from the solar cell electrode (into the device) initiates the "photon recycling" process, without requiring incoming photons. The injected electron contributes into the dark current but only when a photon induced by the "recycling" exits the cell as shown in Figure 6 (c). Otherwise the electrode absorbs the earlier injected electron as it is shown in Figure 6 (d).

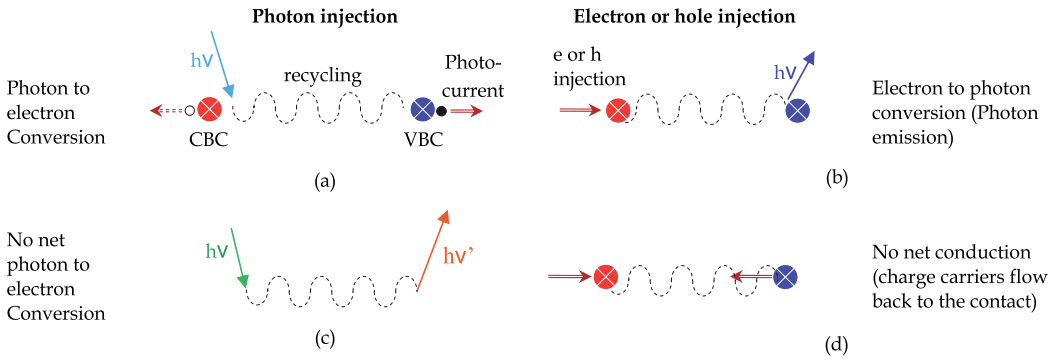


Fig. 6. Feynman-like diagrams for various electron-photon processes in a PV cell. Single and double arrows represent photon and electron, respectively. The periodic path represents the recycling processes of absorption/emission and generation/recombination of photons and electrons, which occurs while the particles propagate in the solid.

2.3.3 Shockley-Queisser limit

For evaluation of the photovoltaic conversion maximum efficiency, Shockley and Queisser have developed a thermodynamic approach based on the detailed balance principle

(Shockley & Queisser , 1961). They have considered the ideal case of solar cells made of perfect materials bestowed with radiative inter-band transitions only. Shockley and Queisser arrived at the paradigm that consisted in equalizing the photo-current and dark-current to the net flux of incoming and emitted photons (Shockley & Queisser , 1961). The most important idealization conditions they have considered are:

- non-radiative interband transitions are prohibited,
- the semiconductor structure is thick enough to ensure full absorption of incoming photons that are able to induce interband transitions in the absorber,
- carrier mobility is infinite so that the quasi-Fermi levels FC and FV exist, respectively, in the conduction band for electrons and in the valence band for holes, and
- contacts are ideal so that one of ohmic contacts absorbs **only electrons** from the conduction band and the other absorbs **only holes** from the valence band.

Dominance of radiative transitions in inter-band transitions maintains electron-hole pairs through "photon recycling". This is schematically illustrated in form of "Feynman-like diagrams" in Figure 6.

In their hallmark model, Shockley and Quiesser have used ideal semiconductor materials where *only radiative electron inter-band transitions are possible*, thereby enabling the development of the detailed balance principle. Fermi's Golden Rule application to electron transitions leads to the same matrix elements for direct and opposite electron transitions in solids so that each photon absorption induced electron transition has its radiative recombination counterpart. The balance of these processes depends on the occupation of the relevant electronic states. The detailed balance principle claims that the net of direct electron transitions between two energy levels equals to the net of opposite electron transitions if the sub-system of the levels is in thermodynamic equilibrium. Shockley and Queisser have shown that, because of intra-band relaxations, the "photon recycling" dynamics brings photon modes traveling long enough within the ideal semiconductor material into the thermodynamic equilibrium distribution of photon modes emitted by the "black body" material; so that the ideal PV cell emits "black body" modes, regardless of material and solar cell structure. Figure 6 shows that the net of photon fluxes emitted from and incoming into the solar cell equals to the net of dark- and photocurrents. One can easily estimate the photocurrent, the dark current, and the conversion efficiency of ideal solar cell. Figure 7 displays conversion efficiencies of *single-junction* based solar cells under normal illumination condition (1 sun) and under full solar light (theoretically possible) as a function of the absorber bandgap (Shockley & Queisser , 1961). The chart shows the Shockley-Quiesser limits for both illumination conditions.

The Shockley-Quiesser detailed balance theory estimates the lowest bound for all recombination processes throughout the device. This lowest bound defines the limit of attainable value of the dark current, which is the recombination rate integrated over the device volume confined within the electron diffusion length. This limit is important because it estimates the potential of the dark current reduction in devices by cleaning materials and improving the device structure.

2.3.4 Luque-Marti limit for quantum dot intermediate band solar cells

The foundation of Intermediary Band (IB) within semiconductor bandgap is very attractive for high efficiency solar cells. In principle, it can be applied for any semiconductor base material, but the benefit differs for one material to another. Formation of IB enables absorption of low energy photons on top of the conventional band to band absorption of

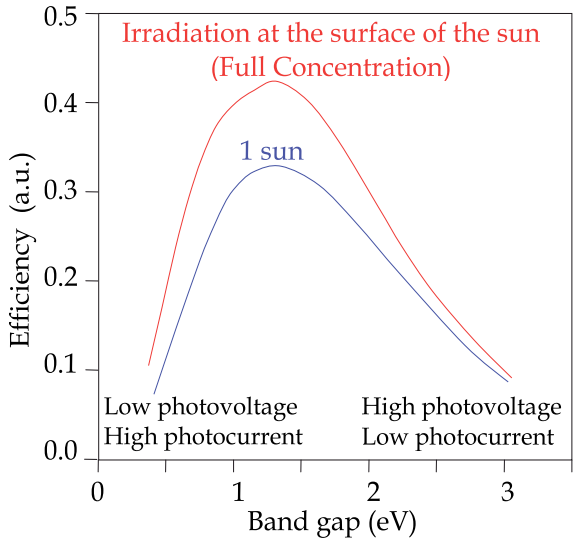


Fig. 7. Limiting efficiency of one-junction conventional solar cells.

photons with energy equal or higher than the bandgap. As shown in Fig. 5, low energy photons that are not converted by conventional silicon solar cells represent a major loss, which justifies the endeavor of modifying silicon to become an IB enabled material.

IB solar cells indeed promise very high conversion efficiency, higher than multi-junction solar cells. New solar cells can use IB states to enable absorption of low energy photons and promote more electronic transitions. This results in producing additional photocurrent and subsequently, a significant increase of solar cell efficiency. However, IB may also induce more recombination of electrons and holes in solar cells (Kechiantz et al., 2007; Kurtz et al., 2005). Because the Shockley-Queisser theory has not considered photogeneration and recombination processes through IB states, Luque and Marti adapted the Shockley-Queisser theory to model IB based solar cell throughput. Guyot-Sionnest et al., (1999) have evaluated the conversion efficiency of ideal IB solar cells in the framework of a new theory.

The IB can be composed by modification of a host semiconductor. A variety of routes can be employed, such as embedding quantum dots (QDs) in a host semiconductor (Marti et al., 1999). An idealized model for the electron energy band diagram in IB solar cell and the associated optoelectronic mechanisms is shown in Figure 8(a). Like in conventional solar cells, absorption of high energy photons brings electrons from the valence band (VB) to the conduction band (CB). Likewise, absorbed low energy photons move electrons from the VB into IB and from IB into the CB , either in synch or differed sequences, depending on the occupation of the states within the IB. The net result is a two-photon absorption with a displacement of an electron from VB to CB.

Luque and Marti have optimized IB energy levels within the bandgap of the base material to match solar spectrum. The matching leads to the utmost net VB-CB electron transitions. The red curve in Fig. 8 (b) displays Luque-Marti limit for conversion efficiency of IB solar cells; the indicated parameter is the higher among energy gaps from VB to IB and from IB to CB. These conversion efficiency curves have been calculated for ideal cells with the following ideal conditions:

- a quasi-Fermi level, F_I , exists for electrons in IB states, and

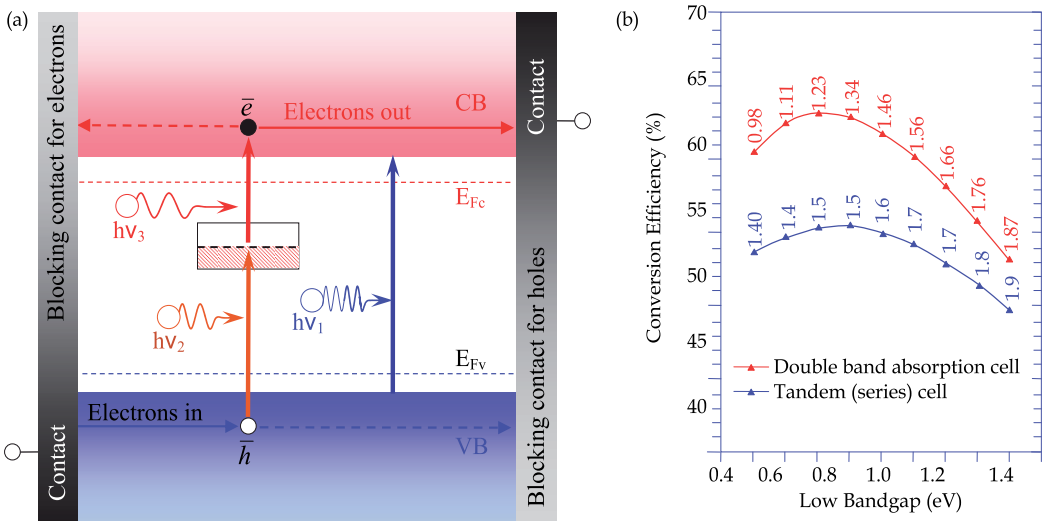


Fig. 8. (a) Energy band diagram of IB solar cell and (b) cell efficiency limits for various IB materials (modified chart in Ref. of Luque & Marti, 2010). Materials with the same bandgap used to make a tandem cell (a completely different cell design) would lead to much lower efficiency limit, as shown with the blue curve.

- no carrier can be extracted from IB into n- and p-doped regions other than by radiative interband transition to continuous bands either VB or CB.

Experiments have confirmed that missing either of the above requisites reduces the efficiency of IB solar cells (Alguno et al., 2006; Luque et al., 2004; Jolley et al., 2010). Also electron transitions through IB must have large matrix elements and solar cells must contain high concentration of IB states for maximizing the generation of additional photocurrent. The red curve in Figure 8(b) exhibits the maximum conversion efficiency for IB solar cells, which is 8 points larger than that for the ideal two-junction tandem solar cells (blue curve). This significant difference is essentially due to a difference in IB solar cell concepts compared to its multi-junction counterpart. While both concepts increment the high energy photon conversion by the low-energy photons in the solar spectrum for achieving higher conversion efficiency, the resulting *photovoltages* are dissimilar. The multi-junction solar cells exploit single-photon absorptions in an array of p-n-junctions connected in series. As the IB solar cells collect current from concurrent processes, the photovoltage can be approached by that of two cells with different bandgaps connected in parallel. In reality, the photovoltage precept is rather more complex. Besides, when an IB cell is shone with *concentrated sunlight, nonlinear effects* driven by the two low-energy photon absorption become significant, which essentially *increase* the generation of *additional photocurrent* in single p-n-junction solar cells (Sun et al., 2006).

2.4 High efficiency solar cells

2.4.1 Down- and up-conversion of solar photon energy

Figure 9 exhibits Si absorption coefficient spectrum and solar energy losses in conventional silicon solar cells. About 10% of incoming solar photons under AM1.5 condition has energy above $3.2eV$. These ultraviolet photons are absorbed in ultra-shallow region, about $10\ nm$

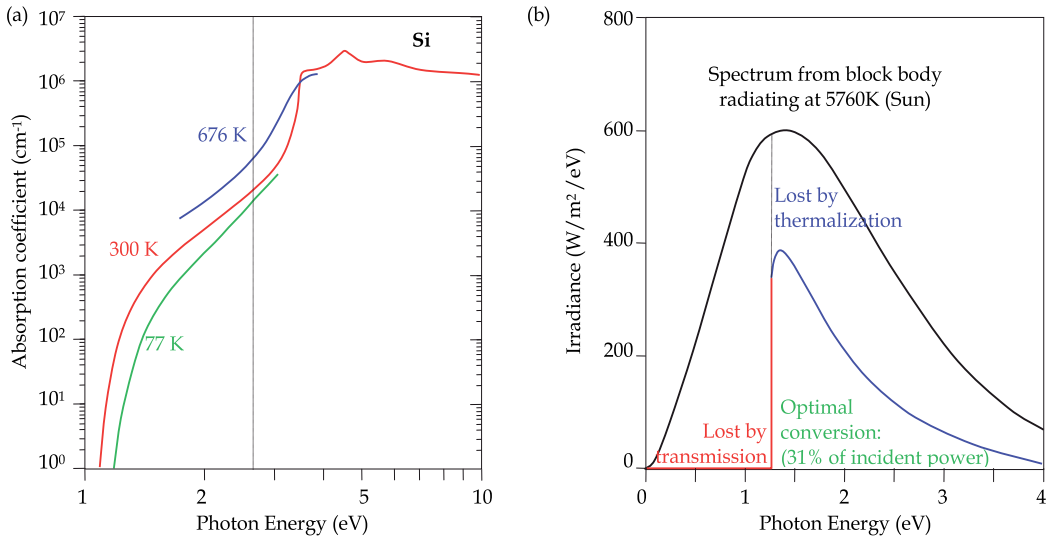


Fig. 9. (a) Absorption coefficient in Si and (b) spectral balance of main losses of solar energy in a one-junction solar cell.

thick in silicon, called the cell dead region because the recombination at the surface eradicates photo-generated carriers.

About 10% of incoming solar photons shown in Fig. 5 have energy above $2.5\,eV$; which is in excess of the energy needed for generating two electron-hole pairs in conventional silicon solar cells (Green, 1998). However, these photons lose most of their energy (in excess of Eg) through intra-band thermalization, as shown in Fig. 9 (b). The *down-conversion* consists in directly splitting the incoming high energy photons into two lower-energy photons or for that matter converting primary optical phonons in low-energy photons. The *down-conversion* has two features, (i) utilization of the excess energy to better match the absorber bandgap leading to more electron hole pairs, and (ii) low energy photons have longer absorption depth and electron-hole pairs can be collected in other areas of the cell. Both features lead to generation of more photoelectrons away from the dead region (Baumgartner et al., 2010).

Also photons with energy lower than the bandgap, represented with small arrow (c.) in Fig. 8, can be used to generate additional photo-carriers in solar cells. Sequenced transitions of two of such photons may, indeed, result in a net ascension of electron to the conduction band. Under some conditions, the most probable path for the electronic system to return to its fundamental state is a radiative transition directly to the valence band. The net effect is the production of a photon of higher energy than the initial ones; this is called *up-conversion*. When up-converted, photon energy matches semiconductor bandgap and hence enables generation of additional photocurrent in solar cells. Noteworthy, the up-conversion is essentially non-linear quantum mechanical effect that gains from sunlight concentration.

2.4.2 Intermediate band for nanostructured silicon materials

Intermediate band (IB) solar cell concept exploits the up-conversion. It uses two low-energy photons for generation of additional photocurrent in solar cells (Luque & Marti, 1997). The detailed balance theory of ideal single p-n junction solar cells allows estimation of conversion efficiency limit of 31% using conventional single-photon absorption mode

(Shockley & Queisser , 1961) and 63% efficiency limit for operation in IB-mode (Luque & Marti, 1997). Such large gain in conversion efficiency will be discussed in more details below.

Like impurity states, QD confined states are easily transformed to fast recombination centers. However, there are experiments that show Type-II QDs made from indirect bandgap materials do not degrade electron–hole recombination lifetime (relative to that of host semiconductor without QDs). The reason stems in the spatial separation of carriers in such QDs. Photo-carriers have much longer lifetime in indirect bandgap materials used in commercial solar cells because phonons must absorb excess momentum, which delays electron–hole recombination. For historical and technical reasons silicon materials have gained incredible advances in material growth and processing. As a consequence, fabricated Si devices have low defect concentration that otherwise degrade carrier lifetime. For instance, electron–hole recombination lifetime can reach 100 μs in CZ Si wafers and epitaxial films and more than 10 ms in float zone silicon. Enhancing photon absorption of indirect band material by hosting specifically designed quantum dots (QDs) opens new opportunities for the foundation of intermediary band (IB) states.

Our choice lies far from standard approaches to PV cell design; we believe that the indirect bandgap silicon, which has poor absorption over the solar spectrum, has a lot of room for improvement through the set-up of an IB. Furthermore, single junction IB cell offers room for significant design optimization.

2.4.2.1 Two approaches to Intermediate Band solar cells

Both IB solar cell concept and the above mentioned requisites (listed in section 2.3.4) do not presume a specific location of IB states in the cell. There are at least two ways for incorporating materials featuring IB(s) in a PV device, which are (i) *within the depletion region* (Luque & Marti, 2010) and (ii) *outside the depletion region* (Sun et al., 2006) of a p-n junction solar cell. Figure 10 (b) displays the energy band diagram of IB solar cell that has type-II QDs imbedded outside the depletion region (Sun et al., 2006) in contrast to the pioneering scheme of IB solar cells, Fig. 10 (a), where QDs are sandwiched within the depletion region between p- and n-doped layers (Luque & Marti, 2010). Recent experiments with InAs QDs embedded in the built-in electric field of depletion layer sandwiched between n- and p-doped GaAs layers have demonstrated that photovoltage of such IB solar cells is lower than in the reference GaAs solar cell (Luque & Marti, 2010). It was proposed that the reason was the overwhelming thermal generation of carriers in InAs QDs. Perhaps InAs/GaAs is not

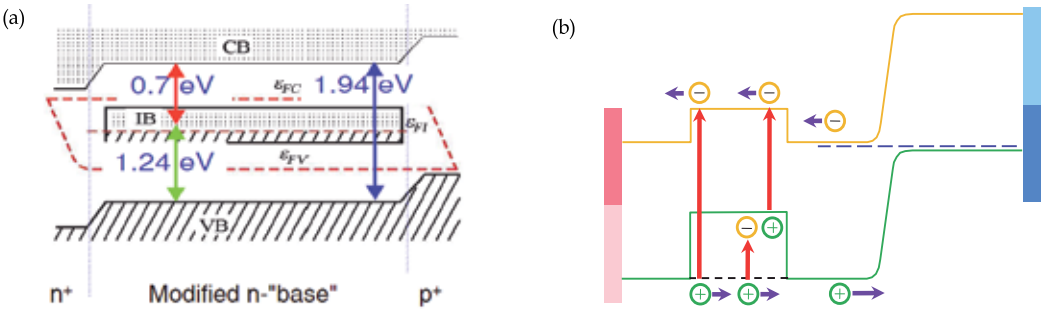


Fig. 10. Two strategies for IB formation in one-junction solar cell: (a) IB within the depletion region (Luque & Marti, 2010); and (b) IB outside the depletion region (Sun et al., 2006).

the best material system for making IB solar cells (Sun et al., 2006). It has then become clear that QDs within the depletion region enhance the dark current.
The recently proposed IB scheme proposed by Sun et al. (2006), see Figure 10 (b), where type-II QDs are placed outside the depletion zone of the cell p-n junction offers the advantage that such QD configuration does not generate additional leakage current. The leading reason is the remoteness of QDs from built-in electric field. This key property isolates confined holes from that field. While holes cannot escape from QDs they still reach the p layer without assistance (for instance from photons, etc….). In the meantime p-n junction separates all mobile photo-carriers generated within a distance less than the carrier diffusion length from the junction and with states either in the VB or in the CB. Such separation includes also photo-carriers transferred trough QD confined states to the continuous bands.
Another important advantage of type-II QDs is the energy band alignment, which results in a potential barrier for majority carriers and a potential well for minority carriers. Such an alignment is shown in Fig. 10 (b) along with carrier transitions and flows. The potential barrier separate carriers photo-generated within the QDs, and hence suppress electron-hole recombination through QDs (Sun et al., 2006). Charge segregation by the QD built-in barrier is an essential feature for irradicating electron-hole recombination in QDs. Energy conversion in IB solar cells strongly depends on the separation of electron and hole quasi-Fermi levels (Kurtz et al., 2005), while recombination through QDs may result in lining-up those quasi-Fermi levels. Hence, it has been argued that QDs confined states can easily turn into fast recombination centers. Likewise, impurities incorporated in GaInNAs solar cells exhibit such conversion into recombination centers (Zibik et al, 2009), whereby the quasi-Fermi level is affixed to the impurity energy level and subsequently the carrier recombination is enhanced. As a result, the photovoltage is degraded to a level lower than expected for GaInNAs solar cells.
Figure 11 shows that sunlight concentration leads to the rise of additional photo-induced potential barrier around type-II QDs in valence band for the new "off-field QDs" strategy (Sun et al., 2006) thereby constituting a new type of IB solar cells. Additionally, such photo-induced barrier suppresses recombination activity of QDs, which results in a more effective two-photon absorption in QDs. Above some value of sunlight concentration the photoelectron generation dominates the recombination in QDs so that the photocurrent generation and the conversion efficiency of such IB solar cell become higher than in the reference solar cell become higher than of a reference cell made of same material and a conventional single p-n junction (Luque et al., 2004).

2.4.2.2 Design of IB solar cell

As the characteristic dimension of solar cell active zones is falling in the nanometer scale, quantum phenomena like the electronic confinement in nanofeatures come to play. These must be adequately used for developing modern electronic and optoelectronic devices and more so the vital third generation solar cells. A detailed understanding of fundamental quantum processes taking place in charge carriers and phonon systems generated within such electronic nanostructures is required. Moreover, those quantum processes must be consolidate with other effects that might otherwise hinder the sought high efficiency devices.
Quantum properties of low-dimensional structured materials are very attractive for designing unique solar cells (Aroutiounian et al., 2001; Nozik, 2002; Luque & Hegedus, 2003;

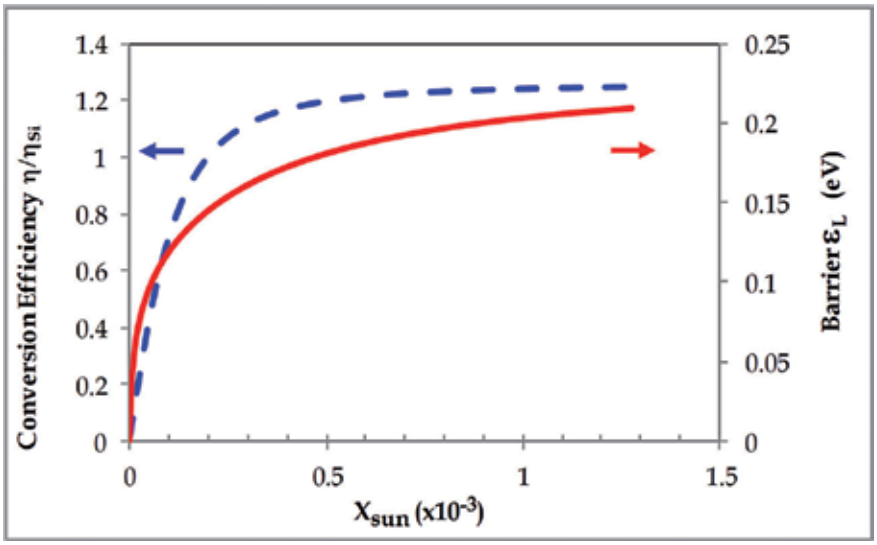


Fig. 11. Irradiation-induced barrier ε_L (dashed line) and conversion efficiency η/η_{Si} (solid line) as a function of sunlight concentration X_{sun}, where η is the conversion efficiency of Ge QD based solar cell and η_{Si} is the efficiency of the reference solar cell (Sun et al., 2006). The reference cell is a Si PN homojunction. The Ge QD induce an IB in the host material outside a similar PN junction.

Tsakalakos, 2008), in particular, IB solar cells with QDs (Marti et al., 2000) QDs imbedded in conventional solar cells introduce resonant energy levels that can be configured to compose a narrow band in the host material band gap (Luque & Marti, 2010). Although silicon is an excellent semiconductor for microelectronics, its indirect bandgap made it inefficient for light absorption, for this matter compare for instance Si absorption coefficient to that of GaAs in Fig. 12. Traditional silicon solar cell design uses at least 150 μm thick wafers to allow capturing enough sunlight. The used substrate is 15 to 20 times thicker than its III-V counter-part. Because of the low absorption and the use of massive material, silicon is in fact not ideal for photovoltaics, and since this expensive material has taken the largest segment in industrial photovoltaic, solar electricity remained a very expensive commodity. The poor absorption makes thick crystalline silicon based PV not apt for further development and cost reduction. However, nanotechnology is being explored to remedy to the silicon fundamental issue of low light absorption. The wide-range of processing knowledge attained during the formidable development of microelectronics, for which silicon still constitutes the base material, can be indeed utilized to compensate the drawbacks of poor light absorption. Also PV silicon suffers from a second problem that is related to the absorption band edge. Wide bandgap semiconductors use more efficiently energy of incident photons as they enable the photocurrent to generate higher photovoltage, whereas the narrower bandgap materials lead to more loss via thermalization of the energy in excess of the bandgap. Silicon bandgap width is considered in the midrange of a host of semiconductors used in photovoltaics. Nevertheless, the loss of high energy excess by thermalization remains considerable in silicon and a solution must be provided.

The large amount of loss practically obstructs one from obtaining high efficiency silicon photovoltaic devices. Industry has been struggling, indeed, for improving silicon solar cell

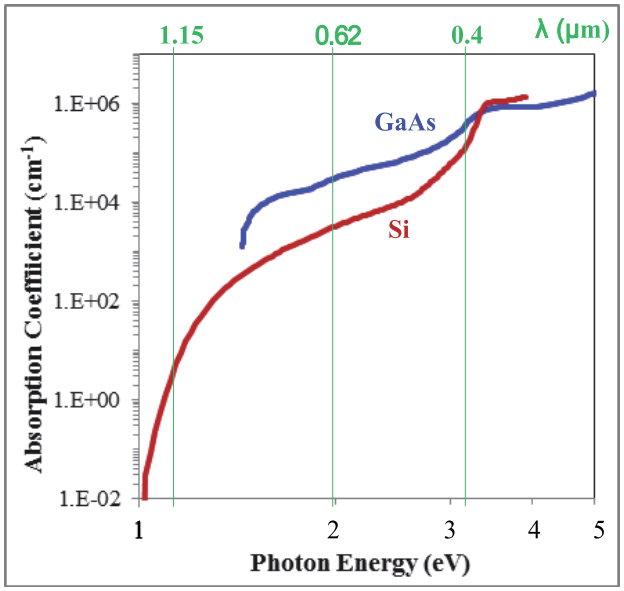


Fig. 12. Direct bandgap (GaAs) versus indirect bandgap material (Si) absorption coefficient.

efficiency at a fraction of percent leaps, while they are getting closer to the theoretical limit. After decades of R&D the fabrication of silicon homojunction solar cells reached its best and has proven to be extremely robust. However, the quantitative assessment of the effect of the bandgap width given in Figs. 4, 5, and 9 shows the shortcomings of the use of silicon in photovoltaics. The main hurdles for PV silicon are the bandgap value, the indirect type and consequently the low light absorption. This explains the efficiency limits over which the PV industry is confronted. Fabrication processes can only be refined for lowering cost and increasing their yield, without adding energy conversion efficiency points that the business is craving of.

However, indirect band gap materials like Ge type-II QD/Si are attractive for IB based solar cells (Sun et al., 2006). This provides an opportunity to overcome the efficiency limitation the current industry is suffering from. Noteworthy is the very slow electron-hole recombination property of IB enabled Si material, which can be exploited for achieving higher efficiency. The large band discontinuities and the real-space indirect fundamental bandgap at the interface in type II Ge QDs embedded in Si are favorable, indeed, for the reduction of carrier recombination. For instance, electron inter-band recombination lifetime is extended to 1 μs in type II Ge QDs in Si (Fukatsu, 1997). Also up to 0.3 eV conduction band offset is possible in this system, which depends on the strain at the interface and the composition of Si spacers (Schaffler, 1997), hence it can be tuned by controlling the strain. Simple tuning can be done by isoelectronic impurities such as Sn, C. The, small C atoms compensate the strain induced by Ge (Schmidt & Eberl, 2000). But the most favorable feature offered by Ge QD, although Ge is an indirect band gap material, is the large cross-sections for photon absorption (Boucaud et al., 1999). Of equal importance is the compatibility of Ge material with Si based technologies, leading to a lower cost integration of Ge in production processes than with other materials.

Luque-Marti concept of IB solar cells does not put limits in the choice of IB material systems and structure. Transition-metal impurities in substitutional positions in III-V compounds

have been under consideration. Narrow energy band predicted in the Brillouin zone of Ga_4As_3Ti compound is shown in Fig. 13, (Wahnon & Tablero , 2002) where an IB.

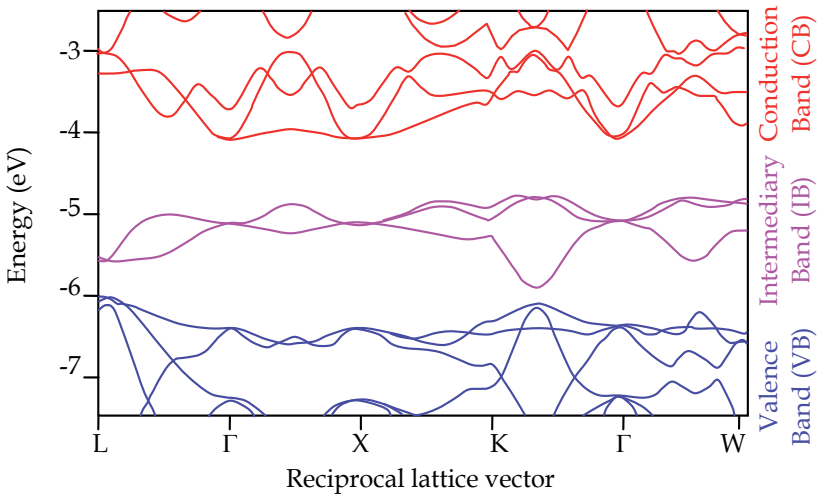


Fig. 13. Transition metal (Ti) induced IB in energy band of III–V compounds, Ga_4As_3Ti (Wahnon & Tablero , 2002).

Layers of alloys with large lattice parameter mismatch is another group suitable for IB (Yu et al.,2003). A relatively small fraction of substitute atoms like nitrogen in III-V compounds and oxygen in II-VI compounds, which are more electronegative than the isovalent host atoms, are used to generate specific features in the electronic structure of such alloys. Figure 14 (a) displays energy band structure of $Zn_{0.88}Mn_{0.12}Te$ with oxygen in substitutional sites, for which one can see the IB right in the middle of the bandgap.
As grown porous silicon (PS) and sensitized PS are the next material systems with unique IB potential for photovoltaic that are proposed by the author and will be studied in the second part of this chapter.

3. Intermediary band enabled silicon for nanostructured PV devices

3.1 Introduction

The currently sought cost reduction of silicon solar cells compels for searching new paradigms for designing solar cells capable of outperforming the widely used silicon solar cells. Hence, modifiers that can make more efficient optoelectronic processes in silicon are desired, while the cost of these substitutes must remain low. Thin film PS is one of such alternatives with a wide range of quasi-classical and quantum properties that can be utilized beneficially for developing the sought new photovoltaic paradigms.

3.2 Quantum properties of nanostructured porous silicon

Electrochemically generated PS constitute a class of complex materials featuring a wide range of possible functionalization. PS chemical composition and surface states vary in a myriad of ways due to fluctuations of etching parameters and exposure of the open surface to chemicals. These appear manageable though. For instance, the size of pores and inner

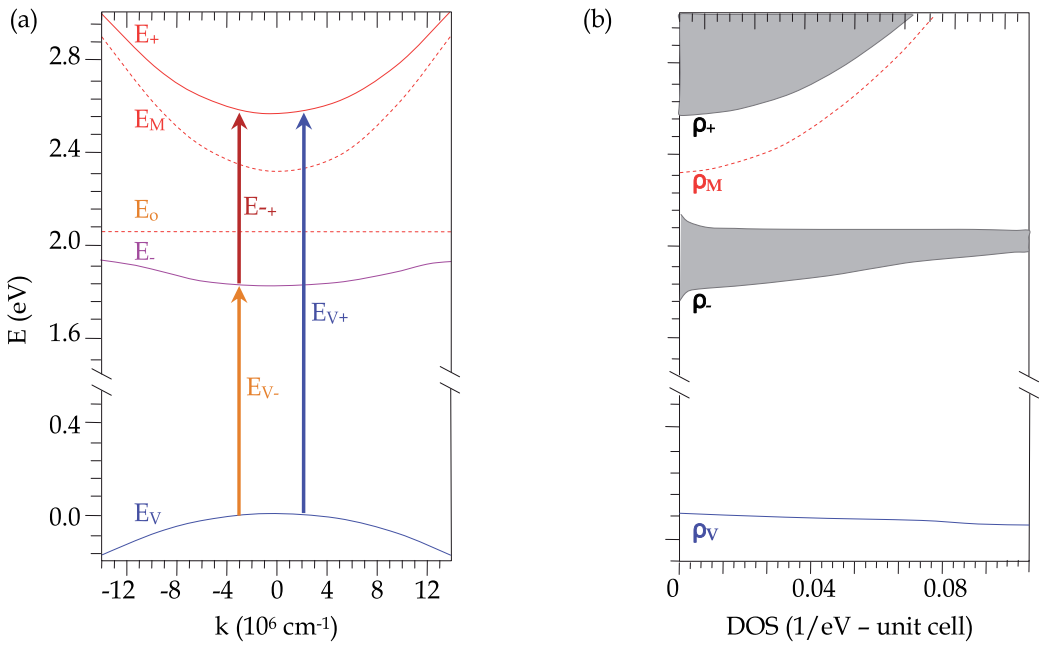


Fig. 14. (a) Band structure of $Zn_{0.88}Mn_{0.12}O_xTe_{1-x}$ with x=0.01 oxygen, exhibiting an IB and (b) the corresponding density of states (Yu et al., 2003).

walls can be deliberately designed to be from a few nm to hundreds; the wall structures can vary from nano- to micro-size and they can contain silicon crystallites and silicon amorphous phases. Porous silicon showed strained silicon phase, that was interpreted based on Raman shift as embedded in oxide (Karoui, 2010). Radiative recombination showed a surface composition that is sensitive to the ambient (Karoui & Zhang, 2010). Potentially, pores may be filled with nanoparticles, or clusters of impurities of other materials attached with physical or chemical bonds to the inner wall surface. Such complexity also provides PS with some useful properties. For instance, QDs can enable PS with electron energy IB. QD sensitized PS is a unique new material system studied by the author for making new solar cells (Karoui & Kechiantz, 2011). For the quantum effects come to play, the electron wavelength must be comparable to the feature size of the considered system, e.g. in QD arrays. Interference of interacting electrons confined in so small volume of a QD may result in quantum effects like Coulomb blockade of electron current through that volume and conductance oscillation. These have been observed in devices like the single electron transistor (SET) (Aleiner et al., 2002). Coulomb blockade enables very strong control of electron transport through SET channel so that the gate switches from insulating state to conducting with one electron charge accuracy (Ai et al., 2010).

Likewise, modified PS that has pore size in the range of several deca-nanometers can behave as a quantum system if the characteristic size of the modifier (here QDs) is comparable with the electron wavelength in that material. Electron energy bands in solids arise from initially discrete energy levels of a very large number of atoms. Each of these N atoms contributes an electronic state so that the initial *s* and *p* atomic levels shown in Fig. 15 are twice N-fold degenerate, considering the two spin states. Reduction of the distance between initially separated atoms brings the orbitals closer, which enables electron

tunneling between neighbor atoms. Furthermore, reduction of inter-atomic distance increases electron interaction in neighboring orbitals, and makes overlapping of electron orbitals possible. In the meantime, interference of interacting electrons enables coherence of their quantum states. This liberates electrons, initially confined around isolated atoms; as a result initial atomic levels are no longer degenerated and are transformed into continuous bands. As the interatomic distances are brought to an optimal value, electron correlation gets stronger and the bands get wider. The well-known diagram given in Fig. 15 illustrates the transformation of energy levels of an atomic assembly into energy bands of the henceforth formed solid. At the atomic spacing R_0 the energy of the assembly is minimal and thus R_0 is the equilibrium interatomic distance at which the assembly becomes stable.

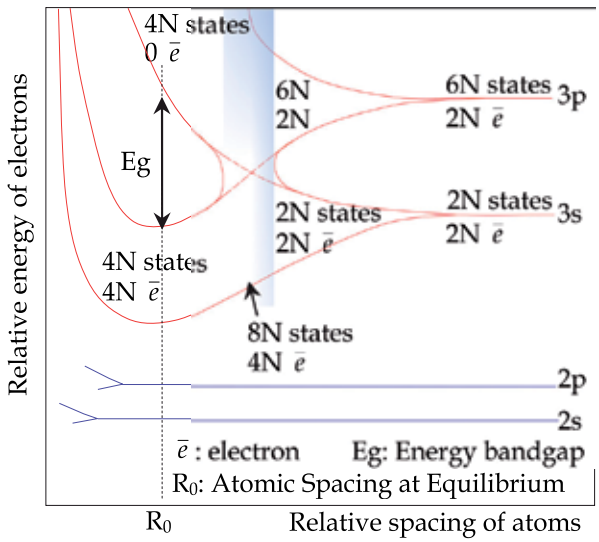


Fig. 15. Merging atomic levels into energy bands in solids.

3.3 Functionalization of porous silicon and size effects

The discrete energy spectrum of Ge QDs buried in crystalline Si modifies the host crystal physical properties. Modification of a host semiconductor with QDs produces material with electronic and optical properties that essentially depend on the size, strain and composition of QDs. However, many other factors like composition of a cap deposited onto a QD layer, separation distance between QDs, strain driven material inter-diffusion at the interface, etc., contribute in the modification of the host material electronic structure. While small QDs are more sensitive to material inter-diffusion, quality interfaces are required for effective charge carrier confinement in small (less than 10 nm) Ge QDs at room temperature (Derivaz et al., 2002). Nanofabricated Ge QDs and the underlying physics have become subjects of thorough investigation for clarifying many new features as these material systems offer humongous potential for using them in making photodetectors, lasers, quantum computers, etc. (Halsall et al., 2004; Meyer et al., 1999). Some of those features have been briefly discussed below. Noteworthy, the behavior of Ge QDs buried in Si system has shown the great potential for building an IB enabled material by embedding Ge QDs in the PS material pores and to control the characteristics of the generated IB.

The understanding of the role of discrete levels in quantum dots and associated electron standing wave-functions, the bunching of energy discrete levels into bands, the origin of splitting of these levels in relation to the coherency of the electron wave tunneling, etc… are crucial for the design of the IB enabled materials. The coherent tunnelling of quasi-confined Bloch wave functions in a finite array of quantum dots (for instance Ge QD attached to inner surfaces of grooves and cavities in porous silicon in the herein discussed case) will emerge as a main mode for charge transfer between QDs and the host material. Carrier tunnelling through critical spacing and its quantum mechanical origin and the coherency of tunnelling waves will be discussed based on calculated electron probability. The quantum levels for QD assemblies and effect of inter-particle transparency, akin to the bunching of atomic levels into energy bands discussed above will be appear, they will depend on the coupling between QDs within the QD array.

3.3.1 Strain induced separation of confined charges

It was reported (Fukatsu, 1997) that diffusion of Ge and the onset of the strain in the Si cap layer may cause spatial separation of confined carriers, i.e., separated holes and electrons. Large mismatch in lattice constants and difference in thermal expansion coefficients enables Si substrate to strain Ge QDs and favors self-assembly of Ge type-II QDs (Lee & Wang , 1994). Most of the band offset occurs in the valence band so that such QDs confine only holes. Square-based pyramidal shape (side ≈100 nm at the base) Ge QDs exhibited intraband absorption at around 300 meV (Boucaud et al., 1999), the authors attributed it to bound-to-continuum transitions in the valence band. The density of these QDs was about $2\times10^{9}cm^{-2}$ and they exhibited an in-plane absorption cross-section of about $2\times10^{-13}cm^{2}$. However, filling the ground state with electrons saturates intraband absorption in those QDs. The use of a virtual substrate, which is a thick buffer layer on graded SiGe, results in a different in-plane lattice constant of the add-on layer due to relaxation. Virtual substrates change the strain configuration in Ge QDs (Fromherz et al., 1994). The $Ge_{1-x}Si_x$ / Si interface layer may become partly strain-free alloy. The new configuration can favor also electron confinement in SiGe buffer layer near Ge QDs so that the structure spatially separates electrons and holes in QDs (Meyer et al., 1999). Inter-subband absorption experiments confirmed such confinement of electron-hole pairs in Ge QDs (Lee & Wang , 1994).

It was reported that free-standing hut clusters of pure Ge are formed at 500°C (Denker et al., 2003) by self-assembling Ge on Si (001) substrate. Photoluminescence characteristic line of these huts is more than 120 meV below the band gap of unstrained bulk Ge, which suggests that a cap Si layer compressively strains huts of pure Ge. The results indicate that the photoluminescence originates by indirect transitions of electrons confined in Si substrate (in the QD vicinity) to holes confined in the strained Ge islands. An activation energy of 40 *meV* ensures that these electrons remain confined around Ge type-II QDs, which is very close to the energy of electron-hole bounding in a neutral exciton (Denker et al., 2003).

3.3.2 Control of QD shape and their density by ultrathin oxide layer

Ultrathin dielectric layer on the inner walls of PS enables the control of QD density. It was reported that 1.2 nm thick SiO_2 covering Si substrate allows production of Ge QDs at a density of 2 to 5×10^{11} cm^{-2} (Derivaz et al., 2002). These dots are hemispherical with 10 nm average size. Based on the RHEED pattern the authors reported that the crystallographic orientations for deposited Ge QDs and the silicon substrate were the same. This points out to the local transfer of the substrate crystallographic order to the add-on material layer. The

transfer can occur through (i) small Si inclusions in the SiO_2 layer or (ii) the local re-order of the oxide layer (largely dependent on the interface stress relief mode). The spatial extent of these processes dictate the mode of self-assembling and the periodicity of the QD array.

3.3.3 Interface QD and host semiconductor

3.3.3.1 Inter-diffusion of Si and Ge at the QD/host material interface

Various techniques are used to fabricate small Ge QDs (smaller than 10 nm) on Si substrates. The most popular is the self-assembled QDs (Konle et al., 2003). Usually capping self-assembled Ge QDs with thin Si or oxide layer follows the formation of quantum dot arrays (Meyer et al., 1999). However, surface segregation and elastic strain relaxation drive QDs above the onset of plasticity to change shape and become flatter. Diffusion of Ge in silicon increases the lateral size of QDs and reduces their height, which may occur along with possible volume loss during the capping. Indeed, the composition of QDs may change during the capping process and even dissolve completely in the surrounding material. More importantly, the capping can induce a strong strain resulting in modification of the QD energy band structure within the QD and in its vicinity. Scanning tunneling spectroscopy study of silicon cap layer showed that deformation of the cap by Ge QDs reduces the band gap of the strained parts of Si cap at the surface (Meyer et al., 1999). The understanding of such reduction is connected to coupling dimer bonds, i.e., Si atoms in the cap and the Ge in the QDs, to stretch in the presence of surface strain. Such stretch reduces overlapping of p-orbitals in the dimers and hence bonding and antibonding splitting of Si p-states may occur. The strain distribution around nanoparticles enables diffusion of Si away from the strained surface regions on top of the Ge nanoparticles. Meanwhile, Ge diffuses towards regions with higher tensile strain, which is one (or the main) driving force for self-assembling of Ge QDs (Tersoff et al., 1996).
Photoluminescence from Ge QDs on Si substrate displays about the same recombination spectral line, which spreads over 0.75–0.9 eV energy range, akin of the high Ge content $Si_{1-x}Ge_x$ QDs and quantum wells on silicon (Miyazaki & Fukatsu, 1999). The broadness of the line indicates significant inter-diffusion between Si and Ge QDs because small particles are very sensitive to the strain. Such inference is also consistent with recently reported theoretical and experimental results of Seok & Kim (2001). The effect of material diffusion must be taken into account in case Ge QD sensitization of PS.

3.3.3.2 Interface transparency

Even in the case of best silicon wafer cleaning via oxide removal, a few silicon oxide monolayers and some amount of chemically or physically adsorbed impurity atoms may still remain. The problem is even more acute for cleaning pore walls in PS (Arenas et al., 2008; Hao et al., 1994). When functionalizing PS with Ge, the remaining oxygen at the surface reacts with Ge during deposition of Ge into the pores of PS. This may result in the encapsulation of Ge QDs incrusted onto the PS walls within a few monolayers of mixture of GeO_2 and SiO_2. Such a layer around a QD constitute an oxide shell (Prabhakaran & Ogino, 1995) that physically separates the dots from the pore wall or other dots. The layer quality determines the transparency of QD-Si or QD-QD interfaces to electrons clinging onto inner walls of pores.

3.3.3.3 Germanium quantum dot-oxide interface

3.3.3.3.1 Decomposition of GeO_2

Electron transport trough Ge QDs incrusted in PS depends on composition of the interface oxide layer. Thus it is essential to precisely know the composition and other properties of

the oxide shell. While SiO_2 and GeO_2 have similar electronic structure, their band gaps are different, $9.0eV$ and $6.1eV$, respectively (Lin et al., 2010). Moreover, unlike SiO_2, GeO_2 lattice poorly matches with Ge lattice, is less stable than SiO_2 on Si, and is soluble in warm water (Prabhakaran & Ogino, 1995; Hovis et al., 1999). Decomposition of isolated GeO_2 into GeO does not take place below 700°C, however, GeO_2 covering Ge dot core has tendency to loose O and transforms to GeO interfacial layer even below 500°C (Prabhakaran et al., 2000). Any thermal processing deteriorates GeO layers. It desorbs with about $2eV$ activation energy into volatile GeO (Wang et al., 2010). This chemical behavior ultimately ends thinning the oxide shell. Ultimately, the thinning stabilizes at a certain thermodynamic equilibrium governed by desorption wells of O atoms and GeO complex out of the GeO_2 layer.

3.3.3.3.2 Dangling bonds at Ge/GeO_2 interface

Tracers ^{73}Ge and ^{18}O isotopes have revealed that interfacial reaction of GeO_2 with Ge generates oxygen vacancies and makes Ge/GeO_2 interfacial region oxygen deficient (Wang et al., 2011). Desorption of GeO occurs due to reaction of GeO_2 with the oxygen vacancies at the GeO_2 surface after diffusion of these vacancies through GeO_2 layer from Ge/GeO_2 interface to GeO_2 surface. Such diffusion is accompanied with the generation of a huge amount of interface states and defects (Kita et al., 2008). Distribution of generated dangling bonds increases from the middle of the bandgap towards conduction and valence band edges like the U-shape distribution of states at Si/SiO_2 interface. However, energy levels of Ge dangling bonds at Ge/GeO_2 interface are below the Ge valence-band maximum E_v, which makes them always negatively charged (Weber et al., 2007). The dangling bonds are acceptor-like and build-up a negative fixed charge at the interface (Afanas'ev et al., 2005). Such negative fixed charge repels electrons from interface. The repulsion is so strong that it prevents layer inversion (Dimoulas et al., 2006).
Both fixed charge and dangling bonds significantly degrade electron transport across the interface Ge QD-Si and between Ge QD-QD clinging onto inner walls of pores in PS.

3.3.3.3.3 Ge/GeO_2 interface passivation Oxygen instead of Hydrogen

The specific feature of Ge/GeO_2 structures is that its straightforward passivation with hydrogen is ineffective because interstitial hydrogen atoms act only as acceptors near Ge/GeO_2 interface (Weber et al., 2007). Repulsion of negatively charged hydrogen from negatively charged interface defects prevents hydrogen passivation of the dangling bonds. As a result, annealing in hydrogen leaves a high density of interface states in Ge/GeO_2 structure though it removes dangling bonds associated with Ge crystal surface atoms (Dimoulas et al., 2006).
It is interesting that high-pressure oxidation at 550°C may stabilize Ge/GeO_2 interface, suppress GeO desorption, and reduce to $2\times10^{11}eV^{-1}cm^{-2}$ the interface density of states (DOS) in the middle of the bandgap (Lee et al., 2009; Matsubara et al., 2008). This is a good example that solves numerous problems like suppression of GeO desorption but passivation of dangling bonds at Ge/GeO2 interface still persists. That is a problem that remains to be solved for obtaining to obtain a working for Ge QD IB PS structures. The problem needs to be solved by further improvement of fabrication technology.

3.3.4 Quantum mechanics model of QD incrusted in porous silicon pores

Identical QDs clinging onto cylindrical pore walls in an ideal PS can be modeled, in a first approximation, as a periodic necklace-like circular chain of QDs, see Fig. 16. Let r_1 be the radius of the pore and r_2 the inner radius of the chain so that the thickness of the chain is

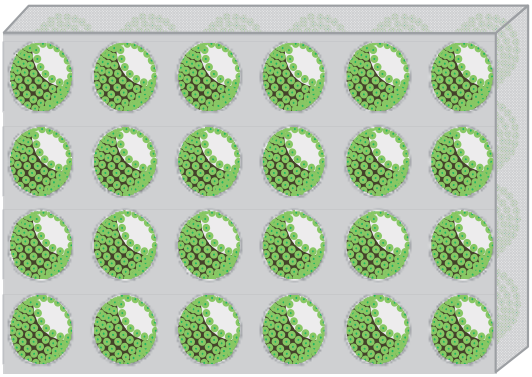


Fig. 16. QD rings inside porous material; in this work SiGe QDs in PS.

$(r_1 - r_2)$. The sketch in Fig. 17 illustrates such model, where the chain and the pore geometry are with perfect surface of inner Si walls.

One must take into account that few monolayer silicon dioxide layers always cover Si surfaces, even in clean silicon samples. Thus , aside from PS surface composition, Si QDs end-up being covered with an oxide shell. This natural oxidation ensures that the PS nanostructure is always covered by an oxide layer as shown in blue in Fig. 17(a). Furthermore, TEM imaging showed that oxide shell in PS separates the Si dots from the Si pore wall and from other dots.

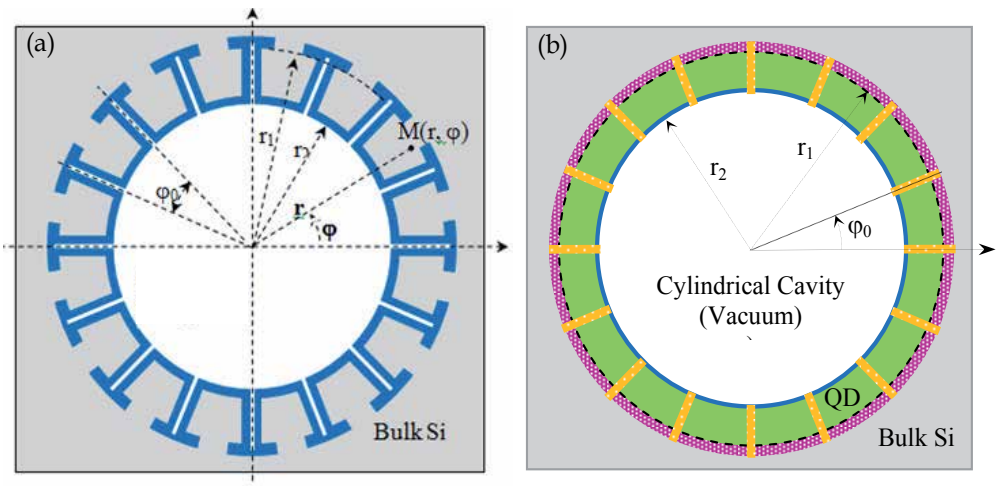


Fig. 17. Model of regular necklace-like circular chains of identical QDs clinging onto cylindrical pore walls in PS, Cross section showing one pore of the PS material and a set of Ge nanodots. The pink lines represent the transparency QD/Host material interfaces to electrons, while the yellow ones represent the transparency of QD/QD separation gaps.

3.3.4.1 Analytical model of QD incrusted in porous silicon pores

We have used polar coordinates (φ, r). QDs are $\varphi_0 r_1$ wide and the chain consists of n_0 QDs so that $\varphi_0 = 2\pi / n_0$.

Since QDs are "artificial atoms", they are expected to have similar energy band formation mechanism for all electronic states confined in an ensemble of closely packed QDs. However,

unlike infinite solids, a small number of QDs can be closely packed within a single pore in PS. Again, interaction and interference of quasi-confined electron waves results in coherence of electron states within such small chain of QDs, which behaves as an "artificial molecule".
Pore walls separate these artificial molecules from each other in PS, leading to a periodic array with symmetry specific to each direction. An artificial crystal is schematically shown in Fig. 17. Such functionalization scheme of PS is new, and has never been studied yet neither experimentally nor theoretically. The work is intended for photovoltaic applications, and will be published elsewhere.

3.3.4.2 Electron and hole wave functions

The simplified model assumes the effective mass approximation and null electron potential in QDs so that $V(r,\varphi)=0$ for $r_2 < r < r_1$. Schrödinger Equation (SE) in cylindrical coordinates, which reflects the symmetry of the physical model, for the electron and hole wave functions in such chain of QDs embedded in the PS can be written as follows:

$$-\frac{\hbar^2}{2m^*}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}+\frac{\partial^2}{\partial z^2}\right]\Psi(r,\varphi,z)=E\Psi(r,\varphi,z) \tag{1}$$

where:
- E and m^* are the electron (hole) energy and effective mass;
- $\Psi(r,\varphi,z)$ is the electron (hole) wave function;

r, φ and z are the radial, angular and axial components of cylindrical coordinates so that $r>0$ and $0<\varphi<2\pi$. Solution for equation (1) allows separation of variables, which reduces that equation to a set of ordinary differential equations for angular $\Phi(\varphi)$, radial $R(r)$, and axial components of the wave functions. The entire wavefunction can be written as:

$$\begin{aligned}\Psi(r,\varphi,z)&=C_{mk}\left[\exp(i\varphi m)+a\times\exp(-i\varphi m)\right]\\&\left[Y_m(kr+b)J_m(kr_2)-J_m(kr+b)Y_m(kr_2)\right]\sin(k_z z)\end{aligned} \tag{2}$$

where:
- m is the angular momentum quantum number;
- k is the radial momentum wave number; it relates to the electron energy E as $k^2=2m^*E/\hbar^2$, $J_m(kr)$, and $Ym\ (kr)$ are the first and second kind Bessel functions of order m,
- k_z is the axial momentum wave number,
- a and b are the integration constants, and
- C_{mk} is the integration constant that represents the amplitude of the wave function in the chain of QDs with respect to that in Si walls of PS.

The boundary condition for electron and hole wave functions is $\Psi(r_2,\varphi,z)=0$ at the inner surface of the chain inside the pore (QD-vacuum), where $r=r_2$. Such condition reduces the electron and hole wave functions within the chain of QDs, $r_2 < r < r_1$, to:

$$\begin{aligned}\Psi(r,\varphi,z)&=C_{mk}\left[\exp(i\varphi m)+a\times\exp(-i\varphi m)\right]\\&\left[Y_m(kr)J_m(kr_2)-J_m(kr)Y_m(kr_2)\right]\sin(k_z z)\end{aligned} \tag{3}$$

The energy band offset at Ge/Si interface is mainly in the valence band. This results in Ge QDs type-II energy band alignment in Si, which ensures that electron and hole wave functions meet different boundary conditions at the Ge/Si interface.

3.3.4.3 Boundary conditions and secular equation

3.3.4.3.1 For holes

We have already illustrated (in section 3, which is about quantum features of nanostructured porous materials) that the close proximity of atoms and the discrete relationship of their energy levels result in the band formations in solids. The same occurs with "artificial molecules" by the Ge QDs ring located inside a pore of PS matrix. If QDs are close enough, the discrete energy levels may form energy bands in the chain of QDs. Transparency of the interface (oxide) layer between QDs controls the phase relationship and induces coherent electron wave functions in QDs. The coherence removes QD discrete energy level degeneracy in the chain and bunches the split sublevels into one energy band.
The electron tunneling probability P is a parameter that characterizes transparency of the barrier located on the ultrathin interface between Ge QDs to electrons, with $0 < P < 1$, where the limit $P = 0$ refers to total opacity of QD-QD interface, i.e., $V = \infty$, and $P = 1$ refers to completely transparent interface, i.e., $V = 0$; V being the barrier height. The transparency parameter (tunneling probability) exponentially decays with the square root of barrier height so that $V \sim (\ln P)^2$. Since tunneling probability is reciprocal to the resistivity, we can attain experimentally the transparency parameter by using nanoscale analysis of electrical resistivity, for instance, by Scanning Kelvin Probe. Semi-transparent interface (oxide) layer between Ge QDs at angle $n_\varphi \varphi_0$, where n_φ is an integer, adds an angular dependent potential barrier $V(\varphi)$ in Schrodinger Equation for the angular component of the wave function,

$$V(\varphi) = \frac{\hbar^2}{m^*}\frac{1}{4\varphi_d}(\ln P)^2 \sum_{n_\varphi}^{2\pi/\varphi_0} \delta(\varphi - n_\varphi \varphi_0) \tag{4}$$

where φ_d is the effective thickness (in angular units) of the barrier layer thickness between Ge QDs; and $\delta(\varphi - n_\varphi \varphi_0)$ is a delta-function. Such semi-transparent potential barrier yields discontinuity at angles $n_\varphi \varphi_0$ at the interface between Ge QDs so that the logarithmic derivative of the wave function jumps by $(\ln P)^2 / 2\varphi_d$,

$$\frac{\partial}{\partial\varphi}\Phi(\varphi_0 + 0) - \frac{\partial}{\partial\varphi}\Phi(\varphi_0 - 0) = \frac{1}{2\varphi_d}(\ln P)^2 \Phi(\varphi_0) \tag{5}$$

In the above equation, the angular term of the wave function, $\Phi(\varphi)$, is invariant with respect to φ_0 angle (rotation about the pore axis shown in Fig.17). Such invariance leads to Bloch theorem, that is mathematically expressed as follows:

$$\Phi(\varphi + \varphi_0) = \exp(i n_\varphi \varphi_0)\, \Phi(\varphi) \tag{6}$$

where, n_φ is limited, however, as compared to Kronig-Penney model.
The continuity of wave function at φ_0, $\Phi(\varphi_0 + 0) = \Phi(\varphi_0 - 0)$, and the jumping of its derivative at the interface combined with the above equations yield a secular equation. The roots of this equation is a set of discrete values for the angular quantum number m:

$$\cos\left(n_\varphi \varphi_0\right) = \cos\left(m\varphi_0\right) + \frac{1}{4\varphi_d m}\left(\ln P\right)^2 \sin\left(m\varphi_0\right) \tag{7}$$

where the integer n_φ is within the interval $0 < n_\varphi \le 2\pi/\varphi_0$.
The hole wave function must swiftly decay for $r_1 < r$ region (from Ge/Si interface into Si bulk material) since the finite potential barrier of the energy band offset at Ge/Si interface confines holes within the chain of Ge QDs,

$$\Psi\left(r,\varphi,z\right) = \left[\exp\left(i\varphi m\right) + a \times \exp\left(-i\varphi m\right)\right] K_m\left(\kappa r\right) \exp\left(\kappa r_1\right) \sin\left(k_z z\right) \tag{8}$$

where $\kappa^2 = \dfrac{2m_{Si}^*(\Delta E_v - E)}{\hbar^2}$,

- m_{Si}^* is the effective mass of holes in Si,
- $K_m\left(\kappa r\right)$ is the order m modified Bessel function of second kind, which is also called MacDonald function, and
- ΔE_v is the energy band offset in the valence band.

The boundary condition requires continuity of the wave functions and their derivatives at Ge/Si interface that is at $r = r_1$. Application of these requirements to equations (3) and (8) yields secular equation that determines a set of discrete energy levels for confined holes

$$\frac{K_{m+1}\left(\kappa r_1\right)}{K_m\left(\kappa r_1\right)} - \frac{k}{\kappa}\frac{Y_{m+1}\left(k r_1\right) J_m\left(k r_2\right) - J_{m+1}\left(k r_1\right) Y_m\left(k r_2\right)}{Y_m\left(k r_1\right) J_m\left(k r_2\right) - J_m\left(k r_1\right) Y_m\left(k r_2\right)} = 0 \tag{9}$$

where the radial momentum (wave number k) is related to the hole energy E as $E = \hbar^2 k^2 / 2m^*$; and the angular quantum number m must be determined from equation (7).

3.3.4.3.2 For electrons

The chain of Ge QDs incrusted in PS allows electron confinement, for instance, if the Si/Ge interface is opaque for electron transport. However, such confinement becomes a virtual "confinement" in case of ultrathin oxide layer that electrons easily pass through so that the electron wave functions expand from the chain into Si. Such interface adds a radial potential barrier in Schrodinger Equation for the radial component of the wave function

$$V\left(r\right) = \frac{\hbar^2}{m^*}\frac{1}{4r_d}\left(\ln P\right)^2 \delta\left(r - r_1\right) \tag{10}$$

r_d is the effective thickness of the barrier layer thickness between Ge QDs and Si wall. Such semi-transparent potential barrier yields discontinuity at radius r_1 at the interface between Ge QDs and Si wall so that the logarithmic derivative of the wave function jumps by $\left(\ln P\right)^2 / 2r_d$.
The continuity of electron wave function requires the continuity for both angular and radial components of the wave functions. Hence for the radial component one obtains the following boundary condition:

$$\begin{aligned}&\left[Y_m\left(k r_1 + b\right) J_m\left(k r_2\right) - J_m\left(k r_1 + b\right) Y_m\left(k r_2\right)\right] \\ &= C_{mk}\left[Y_m\left(k r_1\right) J_m\left(k r_2\right) - J_m\left(k r_1\right) Y_m\left(k r_2\right)\right]\end{aligned} \tag{11}$$

where the angular quantum number m is obtained from equation (7), as indicated above. The discontinuity of the electron wave function derivatives at the interface r_1 adds one more condition for recognition of two unknown parameters, b and C_{mk}, in equation (2) for the electron wave function:

$$k\left[Y_m'(kr_1+b)J_m(kr_2)-J_m'(kr_1+b)Y_m(kr_2)\right]=$$
$$C_{mk}\times\left\{k\left[Y_m'(kr_1)J_m(kr_2)-J_m'(kr_1)Y_m(kr_2)\right]+\frac{(\ln P)^2}{2r_d}\left[Y_m(kr_1)J_m(kr_2)-J_m(kr_1)Y_m(kr_2)\right]\right\} \quad (12)$$

3.3.4.4 Results and discussion

3.3.4.4.1 Virtual "confinement" in conduction band

The continuity of the electron wave functions and the derivative jump at the interface (i..e, at r_1) determine two secular equations. Hence two integration constants, b and C_{mk}, of electron wave function can be determined. Whereas equation (9) yields discrete roots and thus discrete momentum wave numbers k, equations (11) and (12) yield continuous values of b (integration constant) and C_{mk}, as functions of radial momentum wave number k (and hence electron energy E). Therefore, energy solutions for electron localization can exist within the chain of QDs, as equations (11) and (12) do not impose any restriction on the allowed values of electron momentum k.

In the case of total opacity of Ge/Si interface, $P\approx 0$, equation (7) yields integer angular quantum number m, and equation (12) that yields the following simpler equation $Y_m(kr_1)J_m(kr_2)-J_m(kr_1)Y_m(kr_2)=0$, which is very close to $\sin(kr_1-kr_2)=0$. It gives a discrete set of energy states shown in Fig. 18 for electrons confined in the chain of Ge QDs.

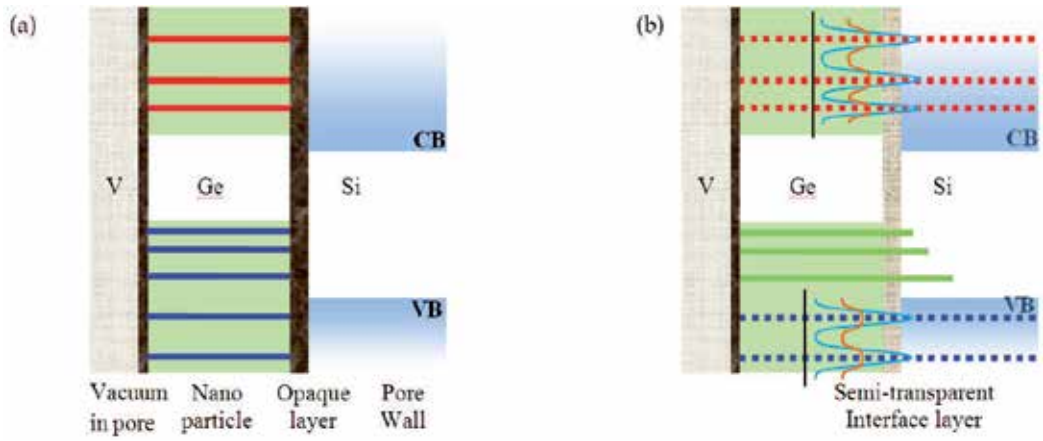


Fig. 18. Schematic diagram of energy bands for the chain of identical QDs clinging onto cylindrical pore walls in porous silicon.

$$E=\frac{\pi^2\hbar^2 n_k^2}{2m^*(r_1-r_2)^2} \quad (13)$$

where n_k =1, 2, 3,... . Here n_k is the principal quantum number for electron confinement along the radius of the QD chain. The principal quantum number is related to the radial momentum *k*.

Schematic diagram of energy bands for the chain of identical QDs clinging onto the cylindrical inner walls of a pore in PS. Green regions represent band edges, separating the Ge bandgap. Red lines are the resonant electronic states in conduction band and the blue lines are the hole confined blue lines are the hole confined states in the valence band. The textured bar is semi-transparent oxide.

Energy levels associated to confined states in Ge dots, with an interface opaque to charge carrier transfer, are visible in Fig. 18 (a), as bleu lines for holes and red for electrons. For comparison the host material energy bands are added. Likewise, Fig.18 (b) shows these levels extending to the silicon (virtual states) due to the semi-transparent interface QD-Si walls. Insets show charge carrier probability related to QD virtual states as a function of the transparency of interface layer barrier (bleu is for higher transparency value). In the case of semi-transparent Ge/Si interface, $0 < P < 1$, solutions of the secular equations (11) and (12) produce a continuous energy spectrum for electrons in the chain of QDs. However, in contrast to electrons in the bulk Si, the amplitude of wave functions is very small in the chain for about all electronic states,

$$\left|C_{mk}\right|_{\min} \approx \frac{\pi\left(2n_k+1\right)r_d}{\left(\ln P\right)^2\left(r_1-r_2\right)} \tag{14}$$

Only for resonant energies that are close to values given by equation (13), the amplitude of electron wave functions is large in the chain

$$\frac{1}{\left|C_{mk}\right|^2} \approx \frac{8r_d^2 m^* E_R}{\hbar^2\left(\ln P\right)^2} + \frac{\left(\ln P\right)^4}{4r_d^2}\left(r_1-r_2\right)^2\left(\sqrt{\frac{E}{E_R}}-1\right)^2 \tag{15}$$

where $E_R = \dfrac{\pi^2\hbar^2 n^2}{2m^*\left(r_1-r_2\right)^2}$ and the width at half-maximum is $\Delta E \approx \dfrac{\sqrt{2}\pi^3\hbar^2 n^3 r_d^2}{m^*\left(\ln P\right)^4\left(r_1-r_2\right)^4}$.

Figure 19 shows that ΔE is small (narrow lineshape) with E_R is very high.

It is worth mentioning that the resonance can occur (i) even between distant Ge QDs in the chain and (ii) even in the case of opaque Ge/Ge interface between QDs due to the electronic coupling through the silicon pore walls. Such "virtually confined" electronic states must also promote electron percolation between QDs. The amplitude of the "virtually confined" electronic waves is so large in Ge QDs that the probability to find electron with energy E_R in QDs is huge compared to that in the same QD-size volume of silicon. Such contrast in amplitudes makes difficult electron escape from "virtually confined" resonant states of QDs into Si and makes easy the opposite transitions.

Large amplitude of electron resonance modes increases probability of two low-energy photon absorption in PS because such absorption is proportional to the electron wave function amplitude in the chain of QDs in PS.

3.3.4.4.2 Confinement in valence band

The roots of the secular equation (9) is a set of discrete energy levels E for holes confined in the chain of QDs, $E = \hbar^2 k^2 / 2m^*$. This transcendental equation is the secular equation of the

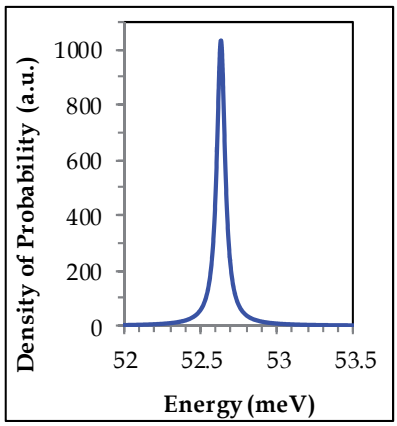


Fig. 19. Square of the amplitude of electron density $|C_{nk}|^2$ as a function of radial momentum kr_1 in the chain of QDs. Interface opacity is $P = 0.5$ and QD size is $r_1 - r_2 = 10\ nm$.

system; it can be solved numerically using a non-linear solver. Approximate roots can be found graphically the zeros of the characteristic function (the term on the left side of the equation).

In the case of opacity interface layer between QDs for hole transport, $P = 0$, each of these levels is n_0 time degenerated; here n_0 is the number of Ge QDs within the chain. However, the degeneracy is removed as the interface transparency increases, $0 < P < 1$. In this case, the discrete energy levels split and give rise to narrow energy bands totaling n_0 electronic states very close in each band.

Because the energy E does not have an explicit dependence on the angular quantum number m, one could conclude that angular confinement of holes would not influence their energy. In fact, k implicitly depends on the angular quantum number m because equation (7) determines allowed values of m for the secular equation (9), which creates implicit dependence of k as well as energy $E = \hbar^2 k^2 / 2m^*$ on the angular quantum number m. Crossing of the blue curve with energy axis in Figure 20 (a) shows the solution of equation (9) for m=400. The red dashed lines point out the solution of the secular equation (9) for m within the range 10 to 100 for which the curve overlap. One can see that the lowest value of energy E for $m = 400$ increases up to 33 meV compare to that for $m = 10$ and $m = 100$. Calculation has showed that this result depends on the ratio $(r_1 - r_2)/(r_1\varphi_0)$ of QD radial to angular sizes.

Solution of Equation (7) yields a series of discrete values for angular quantum numbers m for each integer n_φ from the interval $0 < n_\varphi \le n_0$, where $n_0 = 2\pi/\varphi_0$. Figure 20 (a) displays this series of angular quantum numbers m as a function of the integer n_φ for the chain of $n_0 = 100$ Ge QDs inserted in a cylindrical pore of PS. In the case of opaque interface layer between the chain of QDs, calculation results in the same m for all integer values of n_φ, which means n_0-fold degeneration of these confined electronic states, $n_0 = 100$. The red dashed line in Figure 20 (a) displays such solution. Even a weak transparency of the interface layer between QDs breaks degeneration and splits of the degenerate states into narrow energy bands. The width of the bands depends on the interface transparency P of the interface layer between QDs. The blue curve in Figure 20 (a) displays the removing of n_0-fold degeneracy and the splitting into the band in the case of interface transparency P

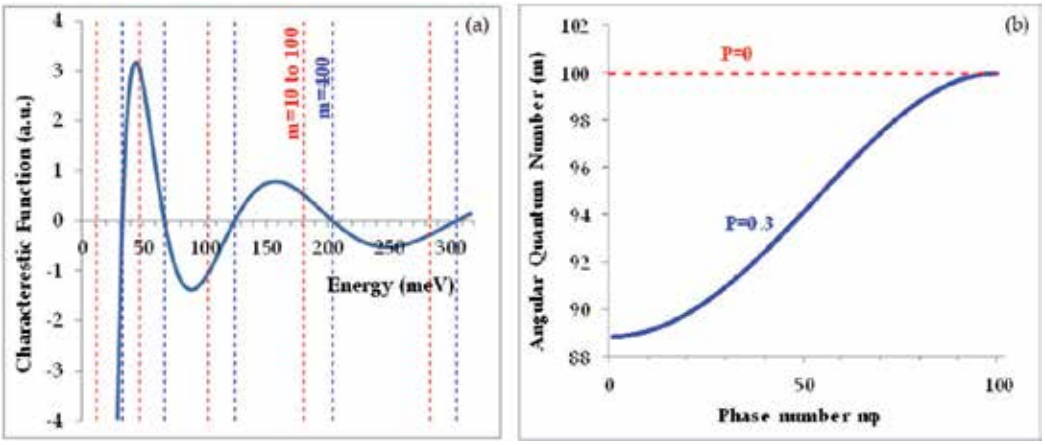


Fig. 20. (a) Eigenvalues obtained by solving the secular equation (9) graphically. Red lines point to the roots for the m within 10 to 100 range, in which case the eigenvalues are about the same. Blue curve is the characteristic function of the secular equation for $m = 400$.
(b) Splitting of degenerated confined states in the chain of Ge QDs clinging onto pore walls in PS. Null intra-chain transparency (between QDs) to electrons shows non-dependence of the angular momentum quantum with the phase number. This results in a full degenerescence of the energy levels.

equal to 0.3. Noteworthy, the angular quantum numbers m of those n_0-split states are not integer, $n_0 = 100$, as all 100 states cluster within narrow interval of close numerical values $88.5 \le m \le 100$.

Calculations show that the reduction of QD angular size φ_0, which increases the number n_0 of QDs in the chain, $n_0 = 2\pi/\varphi_0$, increases the numerical value of the first m in the sets of allowed angular quantum numbers. Because the first m determines the order of the Bessel functions in equation (9) the reduction of φ_0 may also increase the energy of ground confined state in the valence band of Ge QDs. Though the bands are still confined in the valence band of the chain, they also extend into the Si walls because the wave functions decay into Si along a hole tunneling depth. Transparency P of the interface layer between the Si wall and the chain of Ge QDs determines the depth of the band extension into silicon. The salient features of the split bands are (i) within the silicon bandgap, and (ii) extend from the chain of Ge QDs into the Si walls. Both are very favorable features for two low-energy absorption and IB operation of such bands in PS.

3.3.4.4.3 Density of states

The density of states (DOS) in the chain of Ge QDs is the sum of delta-functions over all discrete energy levels, $\rho(E) = \sum_{n_k \ge 1} 2\delta\left(E - E_{n_k}\right)$. The splitting into bands removes degeneration of states and also transforms the DOS $\rho(E)$ in the chain of Ge QDs. The transformation depends on the transparency P of the interface layer between QDs. Figure 21 (a) displays evolution of the density $\rho(E)$ of states per *pore* per *meV* in the confined band split by the interface layer transparency, $P = 0.14$, in the chain of n_0 QDs, ($n_0 = 100$) all QDs are considered identical. The Ge QD thickness, which is along the pore radius,

$\delta r = r_1 - r_2$ is varied from $5nm$ to $3nm$. All QDs have the same angular size $r_1\varphi_0$ .equal to $2\pi r_1/n_0$. The area limited by dark green trapezoids is the same, $n_0 = 100$, for all δr thicknesses because initially all confined states have the same n_0-fold degeneration. The base of green trapezoids is the widths of the split bands. Figure 21 (a) shows that reduction of the QD thickness δr increases the density $\rho(E)$ of states and reduces the width of the split bands. The dashed red and green lines illustrate such dependence on the radial thickness δr of QDs for two ultimate split energies in the band, which refer to angular quantum numbers $m = 88$ and $m = 100$, respectively.

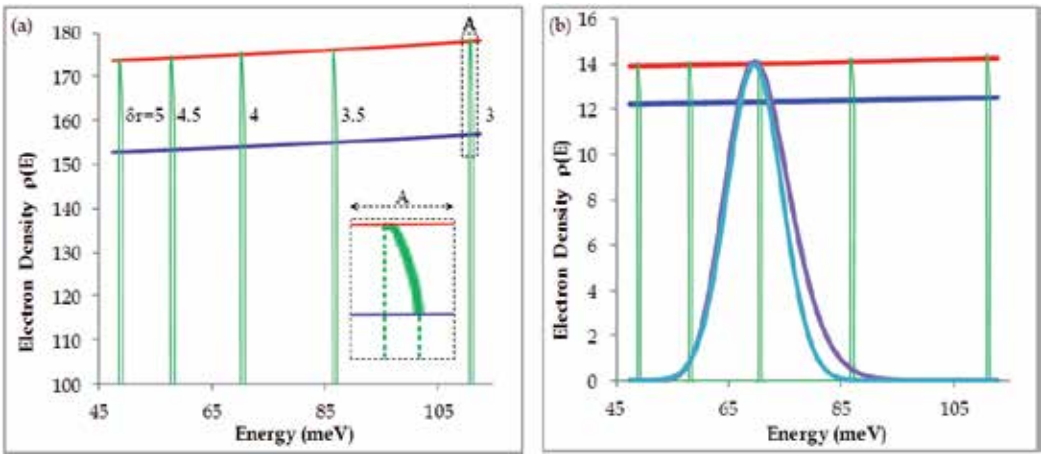


Fig. 21. (a) Density of confined states $\rho(E)$ (per *pore* per *meV*) in the split band. The inset is an enlargement of the electron density function of energy for nanodot thickness of 3 % of the pore radius. (b) The density of confined electronic states (per *pore* per *meV*) in PS for a QD normal size distribution.

The density of the band states in PS could be large as it is related to the large density of pores. The appearance of energy band with large density of confined electronic states in PS is very important for efficient IB solar cell applications. Evidently, properties and potential of such IB in PS are highly sensitive to the ability of controlling the size distribution of Ge QDs in PS (Sun et al., 2005; Abd Rahim et al., 2010).

Fabrication of Ge QDs in PS inevitably imbeds different size Ge QDs into the pores. Depending on the dispersion their contribution in the density of electronic states in PS becomes a weighted function of the dispersion. In case of Gaussian (normal) distribution of QD thicknesses (radial parameter), the weight function is $\left(1/\sqrt{2\pi\sigma^2}\right)\exp\left[-(r-r_{10})^2/(2\sigma^2)\right]$, where r_{10} is the mean thickness and σ is the dispersion of QDs. The dark blue curve in Figure 21 (b) displays the density of confined states in PS for normal distribution of QD thicknesses that $r_{10} = 4nm$ and $\sigma = 0.2nm$. The scale of dark green trapezoids is reduced by 0.08 times for eye guide. One can compare two cases and see that even a small dispersion of QD thicknesses in PS, $\sigma = 0.2nm$, expands the distribution of states into a wider energy band. The dispersion also reduces 0.08 times the maximal

DOS, which occurs at the mean thickness, $r_{10} = 4nm$. Note that the area under both curves are the same as the total number n_0 of Ge QDs in the chain.

The density $\rho(E)$ of states shown in Figure 21 (b) is a skewed Gaussian; compare with the normal distribution (Gaussian) shown in blue (fitted to the actual DOS). The peak of the fitted DOS is at 62.5 *meV* and the dispersion is 4.85 *meV*.

4. Summary

As most of the third generation photovoltaic devices and the used materials are nanostructured, quantum phenomena turn out to be effective. Hence, they must be taken in consideration when designing such devices. Likewise, light-matter interaction quantum processes can be directly or indirectly involved in the photoelectric effect; they are numerous and complex. Therefore, as we move into nanoscale photovoltaic materials targeting higher efficiency solar cells, these processes need extreme attention.

In this chapter we put the emphasis on detailed mechanisms involved in the photoelectric effect in semiconductors. Silicon great relevance for technology calls for more research to overcome the main hurdle related to its indirect bandgap. Also the involvement of a wide range of fundamental processes in connection to the photoelectric effect makes the discussion on Si and Ge indirect semiconductor interesting for this quantum mechanics chapter complexity. We have reviewed fundamental quantum processes taking place in electron and phonon systems in a semiconductor, independent of the device consideration. We then have discussed those systems in nanostructured materials and heterostructures. We have evoked and used (**E**-**k**) energy-momentum dispersion relationship in semiconductors, which is essentially of quantum origin, results in different energy gaps and electron effective masses. These vary with the symmetry points in semiconductor energy-momentum space. Most of the quantum processes involved in the photoelectric effect, which are initiated by photon absorption, have been reviewed in relation to the energy band structure of the absorber. Hence, we described the quantum nature of photon absorption and photon emission processes resulting in electron-hole generation and recombination in solids.

Because photon energy is transferred in a discrete form, generated photoelectron has higher energy than it may contribute in the produced electricity. We assessed the major losses in solar cells and have singled out those that have quantum bearings. We discussed various mechanisms for carrier generation and recombination with consideration of the energy exchange and the net conversion. This led to discussing the thermalization, photon recycling aspect occurring in reality in the course of photoelectric effect processes etc. Ultimately, we went through the net produced charge carrier subsequent to photon absorption or charge carrier injection and recycling.

We went through the three main conceptual limitations of solar energy conversion, namely: the second law of thermodynamics, the Shockley-Queisser, and the Luque-Marti limits. It is necessary to take into account these three physical limitations if one is to exploit every available portion of photon energy in the solar spectrum. Thereafter, we have looked at the physics of multispectral devices known to give the highest efficiency records and compared it to the more recent concept proposed by Luque. For that matter, we analyzed the energy band structure of materials that have an Intermediary Band. We then have compared two published schemes for the generation of IB materials and their use in optoelectronic and photovoltaic devices. For the first, the IB is located within the electronic p-n junction, while

for the second, which is a newer concept, the IB material is outside that region. The first appeared to practically turn QDs intended to enable the IB in recombination centers that degrade the conversion efficiency instead of increasing it. In the newer scheme, the IB material is located outside the electric field within the depletion region, with the intention of preventing charge carrier recombination. In that case the IB material is simply an absorber with an enhanced absorption in the energy range lower than the bandgap. In essence, this is in line with the goal of using IB in photovoltaic devices, that is to generate additional photocurrent while the conventional photocurrent resulting from the VB-CB band to band transitions remains intact.

After clarification of the photoelectric effect fundamental aspects and the routes for high efficiency solar cells, one of the author's approaches to make IB enabled materials was unveiled. It is based on the functionalization of porous silicon with Ge quantum dots. The IB enabled PS solar cells is a new idea that the author (Karoui & Kechiantz, 2011) has been studying experimentally and theoretically through modelling. The author's modeling methodology of the intermediate band enabled porous silicon for solar cells was then summarized and some results have been discussed.

A simple model of QDs necklace-like chain of identical QDs has been put together. The chain is constituted of a regular distribution of QDs, all dots clinging into the pores of PS material. The Schrodinger equation was solved for a single chain in a circular pore, in the effective mass approximation. We realized that this system is described by two essential parameters i.e., angular momentum and radial momentum quantum numbers. We determined the hole and electron wave functions and energy levels in the chain of QDs. The energy spectrum of such system appeared to depend on transparency of the interface layer between the chain of QDs and Si wall, and on transparency of the interface layer between QDs in the chain.

We have discussed the interface transparency factors. For that purpose we looked at published literature about the composition of oxide cap layers, the physical properties such as dangling bond, the layer formation and decomposition, etc. We have analysed the interface layer material properties for Ge-QD-chain/Si-walls material system. The interface layers of such system is a solid solution of SiO_2, GeO_2, and GeO oxides. The composition, structure, and thickness of the interface, which determine its transparency, depend on fabrication technology. To be noted, if the interface is totally impermeable to electrons, the width at half-maximum ΔE of the electron wave function vanishes and the electron energy spectrum becomes discrete. Remarkably, even for extremely small values of interfaces transparency, electrons from the cavity walls (i.e., host material) spill over the QDs, which results in a continuous component in the energy spectrum. However, most wave functions of such electrons vanish in the Ge QDs, while some resonant modes grow large. These resonant electronic states are "virtually confined" in QDs, they easily accommodate electrons from pore walls of PS.

We then showed that Ge QDs introduce a narrow band of evanescent electronic states that being confined in the chain material; these states also expand into the pore wall of the host material due to the quantum mechanics tunnelling. In case of valence band in Ge(QD)/Si(wall) material system, the narrow band of evanescent states is aligned with the Si bandgap, which gives them a potential for being effective within the IB enabled PS material. The density of such IB states is an important parameter for efficient performance of IB PS solar cells. We have shown that IB density in PS may be large if the thickness of pore

walls is comparable with electron tunnelling depth into PS. Our calculation has revealed that the discrete energy spectrum of confined holes is highly sensitive to electron transparency of the interface layer between QDs in the chain. While an opaque interface layer, which completely separates QDs from each other, yields a degenerated discrete spectrum. A semi-transparent interface layer removes such degeneration and widens the discreet spectrum into a narrow IB.

In the valence band of Ge QDs, the confined states induce narrow energy bands lined within the Si bandgap. These Ge valence band states spill over and decay in the PS, whereas conduction band resonant electronic states extend from Si into the Ge QDs, even if the interface between Ge and Si is semi-transparent for charge carriers. The salient features of the split bands are (i) lining up with the silicon band gap, and (ii) extension from the chain of Ge QDs into the Si walls. Both are essential properties for the absorption of two low-energy photons via the IB. Such properties are desirable for Ge QD modified PS based solar cells. The chosen material system put forward an electronic structure that is efficient with respect to electron transitions induced by low energy photons.

Because QDs grown in PS inevitably have a random size distribution, their contribution into the density of electronic states is a weighted function of the size (thickness and breadth) dispersion. We have calculated the density of IB electronic states in PS and studied the size dispersion effect on the DOS in IB. Our study has shown that Gaussian size distribution of QD yields a non-Gaussian distribution of density of states in IB, this result is owing to the quantum confinement in the chain of QDs.

5. Acknowledgment

This material is based upon work supported by the US Department of Defense, Contract Number: W91CRB-10-C-0321. The authors are indebted to Dr Sondes Karoui for discussing the manuscript.

6. References

Arenas, M.C.; Hu, H.; Antonio, del Rıo J.; Salinas, O.H., (2008). Photovoltage & J-V features of porous silicon, *Rev. Mex. Fıs.* Vol.54, (5) pp. 391–396

Abd Rahim, A.F.; Hashim, M.R. & Ali, N.K. (2010). Study of Ge embedded inside porous silicon for potential MSM photodetector, *Microelectronics International, Vol.* 27, No. 3, pp. 154–158

Afanas'ev, V.V.; Fedorenko, Y.G.; & Stesmans, A. (2005). Interface traps & dangling-bond defects in (100)Ge/HfO2, *Appl. Phys. Lett.* , Vol. 87, 032107

Ai, N.; Sul, O.; Begliarbekov, M.; Song, Q.; Kumar, K.; Choi, D. S.; Yang, E.-H. & Strauf, S. (2010). Transconductance & Coulomb Blockade Properties of In-Plane Grown Carbon Nanotube Field Effect Transistors, Nanosci. Nanotechnol. Lett, , Vol. 2, (2) pp. 73-78

Aleiner, I.L.; Brouwer P.W. & Glazman, L.I. (2002). Quantum effects in Coulomb blockade, *Physics Reports,* Vol. 358, pp. 309-440

Alguno, A.; Usami, N.; Ohdaira, K.; Pan, W.; Tayanagi, M. & Nakajima, K. (2006). Influence of stacked Ge isl&s on the dark current–voltage characteristics & the conversion efficiency of the solar cells, *Thin Solid Films,* Vol. 508, No. 1-2, pp. 402 – 405

Aroutiounian, V.; Petrosyan, S.; Khachatryan, A. & Touryan, K. (2001). Quantum dot solar cells, J. *Appl. Phys.,* Vol. 89, pp. 2268-2271

Baumgartner, K.; Holländer, B.; Carius, R.; Ahrens, B.; Angelov, O.; Sendova-Vassileva, M.; Dimova-Malinovska, D. & Schweizer, S. (2010). Photon down-conversion in Terbium(III)-doped thin dielectric films & fluorozirconate glasses for thin film solar cells, Proc. SPIE 7725, 77250T; doi:10.1117/12.853939

Boucaud, P.; Le Thanh, V.; Sauvage, S.; Debarre, D. & Bouchier, D. (1999). Intraband absorption in Ge/Si self-assembled quantum dots, *Appl. Phys. Lett.,* Vol. 74, pp. 401–403

Boucaud, P.; Le Thanh, V.; Sauvage, S.; Debarre, D. & Bouchier, D. (1999). Intraband absorption in Ge/Si self-assembled quantum dots, Appl. Phys. Lett., Vol. 74, pp. 401

Denker, U.; Stoffel, M.; Schmidt, O.G. & Sigg, H. (2003). Ge hut cluster luminescence below bulk Ge band gap, *Appl. Phys. Lett.,* Vol.82, (3), pp. 454-456

Derivaz, M.; Noe, P.; Rouviere, J.L.; Buttard, D.; Sotta, D.; Gentil, P. & Barski, A. (2002). Epitaxial growth of germanium dots on silicon (001) surface covered by a very thin dielectric layer, Materials Science & Engineering B89, pp. 191–195

Dimoulas, A.; Tsipas, P.; Sotiropoulos, A. & Evangelou, E. K. (2006). Fermi-level pinning & charge neutrality level in germanium, *Appl. Phys. Lett.* Vol.89, pp. 252110 -252112

Dziewior, J. & Schmid, W. (1977). Auger coefficients for highly doped & highly excited silicon, *Appl Phys Lett,* Vol. 31(5) pp. 346

Fromherz T.; Koppensteiner E.; Helm M.; Bauer G.; Nutzel J. F. & Abstreiter G. (1994). Hole energy levels & interband absorption in modulation-doped Si/Si1-xGex multiple quantum wells, *Phys Rev B* vol.50, (20), pp. 15073-15085

Fukatsu, S.; Sunamura, H.; Shiraki, Y. & Komiyama, S. (1997). Phononless radiative recombination of indirect excitons in a-Si/Ge type-II quantum dot, Appl. Phys. Lett. Vol. 71, pp. 258

Green, M.A. (1984)Limits on the open-circuit voltage & efficiency of silicon solar cells imposed by intrinsic Auger processes, *IEEE Trans on Electron Devices ED*, Vol. 31, (5) pp. 671-678

Green, M.A. (1998). Solar Cells, *in Modern Semiconductor Device Physics* Ed S M Sze, Wiley, New York

Guyot-Sionnest, P.; Shim, M.; Matranga, C. & Hines, M. (1999). Intraband relaxation in CdSe quantum dots, *Phys Rev,* Vol.60, (4) R2181-R2184

Halsall, M.P.; Dunbar, A.D.F.; Shiraki, Y.; Miura, M. & Wells, J-P R. (2004). Hole confinement & dynamics in d-doped Ge quantum dots, J. Luminescence, Vol.108, pp. 329–332

Handbook of Photovoltaic Science & Engineering, ed Luque, A.; Hegedus, S., Wiley, 2003

Hao, P.H.; Hou, X.Y.; Zhang, F.L. & Wang, X. (1994). Energy band lineup at the porous-silicon/silicon heterointerface measured by electron spectroscopy, *Appl. Phys. Lett.* , Vol. 64, (26) pp. 3602

Hovis, J. S.; Hamers, R. J. & Greenlief, C. M. (1999). Preparation of clean & atomically flat germanium(001) surfaces , *Surface Science*, Vol.440, L815–L819

Jolley, G.; Lu, H.F.; Fu, L.; Tan, H.H. & Jagadish, C. (2010). Electron-hole recombination properties of In0.5Ga0.5As/GaAs quantum dot solar cells & the influence on the open circuit voltage, *Appl Phys Lett*, Vol.97, pp. 123505

Karoui, A.; Kechiantz, A. Sensitization of Porous Silicon with Germanium Quantum Dots for Up-Conversion of Low Energy Photons via Intermediate Band for Third

Generation Solar Cells, (2011) ECS Transactions *Photovoltaics for the 21st Century*, Eds. M. Tao, C. Clayes, L. Deligiani, Vol. 41, Issue (4), pp. 53-60.

Karoui, A. Analysis of Formation Mechanisms of Porous Silicon for Photovoltaics Using Nanoscale Raman Scattering and Photoluminescence, Proc. Conf. Symposium (B) Photovoltaic Materials and Manufacturing Issues, October 4 – 7, 2010, Denver, CO.

Karoui, A.; Zhang, H. Porous Silicon Formation and Photoluminescence Decay Analysis, Proc. (2010) Conf. Symposium E8 - Photovoltaics for the 21st Century 6, The 218th Electrochemical Society Meeting - Las Vegas, Oct. 10 - 15, 2010.

Karoui, A.; Zhang, R., Rozgonyi, G. A.; &, Ciszek, T. (2001) Silicon Crystal Growth and Wafer Processing for High Efficiency Solar Cells and High Mechanical Yield, Proc. Conf. NCPV, NREL /CP-520-31057, Lakewood, CO, 14-17 Oct. 2001.

Kechiantz, A.M.; Kechiyants, H.M. & Kocharyan, L.M. (2007). band Alignment & Conversion Efficiency in Si/Ge Type-II Quantum Dot Intermediate band Solar Cells, *Nanotechnology*, Vol. 18, pp. 5401

Kita, K.; Takahashi, T.; Nomura, H.; Suzuki, S.; Nishimura, T. & Toriumi, A. (2008). Control of high-k/germanium interface properties through selection of high-k materials & suppression of GeO volatilization, Appl. Surf. Sci., Vol. 254, pp. 6100

Konle, J.; Presting, H. & Kibbel, H. (2003). Self-assembled Ge-isl&s for photovoltaic applications Physica E, Vol. 16, pp. 596

Kurtz, S.; Johnston, S., & Branz, H.M. (2005). Capacitance-spectroscopy identification of a key defect in N-degraded GalnNAs solar cells Appl. *Phys. Lett.*, Vol. 86, pp. 113506

Lee, C. H.; Tabata, T.; Nishimura, T.; Nagashio, K.; Kita, K. & Toriumi, A. (2009). Ge/GeO2 Interface Control with High-Pressure Oxidation for Improving Electrical Characteristics, *Appl. Phys. Express* 2, 071404

Lee, C. & Wang, K. L. (1994). Electron intersubband absorption in Ge/Si1−xGexquantum-well structures grown on Si (001) substrate, *Appl. Phys. Lett.* , Vol. 64, pp. 1256

Lin, L.; Xiong, K. & Robertson, J. (2010). Atomic structure, electronic structure, & band offsets at Ge:GeO:GeO2 interfaces, *Appl. Phys. Lett*, Vol. 97, pp. 242902

Luque, A. & Marti, A. (1997). Increasing the efficiency of ideal solar cells by photon induced transitions at intermediate levels, *Phys. Rev. Lett.* , Vol. 78, pp.5014

Luque, A. & Marti, A. (2010). The Intermediate band Solar Cell: Progress Toward the Realization of an Attractive Concept, Adv. Mater, Vol.22, pp.160–174

Luque, A.; Marti, A.; Stanley, C.; Lopez, N.; Cuadra, L.; Zhou, D.; Pearson, J.L. & McKee, A. (2004). General equivalent circuit for intermediate band devices: Potentials, currents & electroluminescence J. Appl. Phys. , Vol. 96, pp. 903-909

Marti, A.; Cuadra, L. & Luque, A. (1999). Quantum dot super solar cell, Proc. Conf. on Sobre Dispositivos Electronicos Madrid pp. 363-366

Marti, A.; Cuadra, L. & Luque, A. (2000) Quantum dot intermediate band solar cell, Proc. 28th IEEE Photovoltaic Specialist Conf. (Fairbanks, AK, 2000) pp.940-943

Matsubara, H.; Sasada, T.; Takenaka, M. & Takagi, S. (2008). Evidence of low interface trap density in GeO2/Ge metal-oxide-semiconductor structures fabricated by thermal oxidation, *Appl. Phys. Lett.* , Vol. 93, 032104

Meyer, T.; Klemenc, M. & Von Kanel, H. (1999). Surface electronic structure modifications due to buried quantum dots, *Phys Rev B* , Vol.60, (12), R pp. 8493-8496

Miyazaki, T. & Fukatsu, S. (1999). Diminished photoluminescence polarization due to exciton ionization in strained Si1−xGex/Si(001)quantum wells, *Appl. Phys. Lett.* , Vol. 75, pp.3962

Nozik, A.J. (2002). Quantum dot solar cells Physica E 14 115

Prabhakaran, K. & Ogino, T. (1995). Oxidation of Ge(100) & Ge(111) surfaces: an UPS & XPS study, *Surface scienc* , Vol. 325, pp. 263

Prabhakaran, K.; Maeda, F.; Watanabe, Y. & Ogino, T. (2000). Thermal decomposition pathway of Ge & Si oxides: observation of a distinct difference, *Thin Solid Films,* Vol. 369, (1-2) pp. 289-292

Schaffler, F. (1997). High-mobility Si & Ge structures, *Semicond. Sci. Technol.* , Vol. 12, pp.1515–1549

Schmidt, O.G. & Eberl, K. (2000). Photoluminescence of monolayer to submonolayer thick Ge1−zCz on Si (111), *Semicond. Sci. Technol.* Vol. 15, pp. 399–402

Seok, J. H. & Kim, J. Y. (2001). Electronic structure & compositional interdiffusion in self-assembled Ge quantum dots on Si(001), *Appl. Phys. Lett* , Vol.78, (20), 3124-3126

Shockley, W. & Queisser, H.J. (1961). Detailed Balance Limit of Efficiency of p-n Junction Solar Cells, *J Appl Phys,* Vol. 32, pp. 510-519

Sinton, R. A. & Swanson, R. M. (1987). Recombination in Highly Injected Silicon, IEEE Trans on Electron Devices ED-34(6) pp.1380-1389

Sun, K. W.; Kechiantz, A.; Lee, B. C. & Lee, C. P. (2006). Ultrafast carrier capture & relaxation in modulation-doped InAs quantum dots, *Appl. Phys. Lett.,* Vol. 88, 163117

Sun, W.; Kherani, N.P.; Hirschman, K.D.; Gadeken L.L. & Fauchet P.M. (2005). A three-Dementional porous silicon p-n-diode for betavoltaics & Photovoltaics, Adv. Mater. 17, 1230-1233

Svantesson, K.G. & Nilsson, N.G. (1979) The temperature dependence of the Auger recombination coefficient of undoped silicon*, J Phys C,* Vol.12, pp.5111

Sze, S. M.; Irvin J. C. (1968). Resistivity, Mobility, and Impurity Levels in GaAs, Ge, and Si at 300 K, Solid-State Electron., Vol. 11, pp. 599–602

Tersoff, J.; Teichert, C. &. Lagally, M. G, (1996). Self-Organization in Growth of Quantum Dot Superlattices, *Phys. Rev. Lett.,* Vol. 76, pp. 1675–1678

Tsakalakos, L. (2008). Nanostructures for photovoltaics, *Materials Science & Engineering R* , Vol. 62 pp. 175–189

Wahnon, P. & Tablero, C. (2002). Ab initio electronic structure calculations for metallic intermediate band formation in photovoltaic materials, *Phys. Rev. B,* Vol.65, pp.165115

Wang, S. K; Kita, K.; Lee, C. H.; Tabata, T.; Nishimura, T.; Nagashio, K. & Toriumi, A. (2010). Desorption kinetics of GeO from GeO2/Ge structure, *J. Appl. Phys.,* Vol. 108, pp. 054104

Wang, S.K.; Kita, K.; Nishimura, T.; Nagashio, K. & Toriumi, A. (2011). Isotope Tracing Study of GeO Desorption Mechanism from GeO2/Ge Stack Using 73Ge & 18O, Jpn. *J. Appl. Phys.,* Vol. 50 04DA01

Weber, J.R.; Janotti, A.; Rinke, P. & van de Walle, C.G. (2007). Dangling-bond defects & hydrogen passivation in germanium, *Appl. Phys. Lett.,* Vol. 91, pp. 142101

Yu, M.; Walukiewicz, W.; Wu, J.; Shan, W.; Beeman, J.W.; Scarpulla, M.A.; Dubon, O.D. & Becla, P. (2003). Diluted II-VI Oxide Semiconductors with Multiple band Gaps, *Phys Rev Lett.*, Vol. 91 pp.246403

Zibik, E.A.; Grange, T.; Carpenter, B.A.; Porter, N.E.; Ferreira, R.; Bastard, G.; Stehr, D.; Winnerl, S.; Helm, M.; Liu, H. Y.; Skolnick, M.S. & Wilson, L.R. (2009)Long lifetimes of quantum-dot intersublevel transitions in the terahertz range, *Nature Materials,* Vol. 8, pp. 803 - 807

12

Quantum Information-Theoretical Analyses of Systems and Processes of Chemical and Nanotechnological Interest

Rodolfo O. Esquivel[1,5] et al.[*]
[1]*Departamento de Química, Universidad Autónoma Metropolitana-Iztapalapa, México D.F.*
[5]*Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, Granada*
[1]*México*
[5]*Spain*

1. Introduction

The application of information-theoretic concepts and techniques to the study of quantum multielectronic systems is presently attracting the attention of numerous researchers in several fields. This area of inquiry is shedding new light on the conceptual foundations of physics and is also at the core of the new field of Quantum Information Theory, which foresees important technological developments through concepts such as "entanglement", "teleportation" and "quantum computation". In line with these developments we present in this Chapter a review of the recent advances performed in our laboratories to study selected molecular processes and mesoscopic systems at the nanoscopic scale by employing information theory concepts to show significant advances of Information Theory applied to chemistry by use of Shannon entropies through the localized/delocalized features of the electron distributions allowing a phenomenological description of the course of elementary chemical reactions by revealing important chemical regions that are not present in the energy profile such as the ones in which bond forming and bond breaking occur. Further, the synchronous reaction mechanism of a S_N2 type chemical reaction and the non-synchronous mechanistic behavior of the simplest hydrogenic abstraction reaction were predicted by use of Shannon entropies analysis. These studies have shown that the information-theoretical measures provide evidence to support the concept of a continuum of transient of Zewail and Polanyi for the transition state rather than a single state, which is also in agreement with other analyses. Although information entropies have been employed in quantum chemistry, applications in large chemical systems are very scarce. For nanostructures, we have been able to show that IT measures can be successfully employed

[*] Edmundo M. Carrera[1], Cristina Iuga[1], Moyocoyani Molina-Espíritu[1], Juan Carlos Angulo[2,5], Jesús S. Dehesa[2,5], Sheila López-Rosa[2,5], Juan Antolín[3,5] and Catalina Soriano-Correa[4]
[1]*Departamento de Química, Universidad Autónoma Metropolitana-Iztapalapa, México D.F., México*
[2]*Departamento de Física Atómica, Molecular y Nuclear, Universidad de Granada, Granada, Spain*
[3]*Departamento de Física Aplicada, EUITIZ, Universidad de Zaragoza, Zaragoza, Spain*
[4]*Laboratorio de Química Computacional-QFB. FES-Zaragoza-UNAM, Iztapalapa, México, DF, México*
[5]*Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, Granada, Spain*

to analyse the growing behaviour of PAMAM dendrimers supporting the dense-core model against the hollow-core one.

2. Essentials of information theory: Classical and quantum

Information theory of quantum many-body systems is at the borderline of the development of physical sciences, in which major areas of research are interconnected, i.e., physics, mathematics, chemistry, and biology. Therefore, there is an inherent interest for applying theoretic-information ideas and methodologies to chemical, mesoscopic and biological systems along with the processes they exert. In this Section we briefly present the theory of the two possible levels, classical (Shannon, Fisher, complexity, etc) and quantum (von Neumann and other entanglement measures). The theoretic-information analyses presented here have been scarcely considered in the literature until recently and reveal important reactivity aspects of elementary chemical reactions which are not accessible by other theoretical methodologies. Therefore, we present important concepts of Information Theory along with the natural atomic probabilities employed for the calculation of the entropies.

2.1 Shannon information theory

The uncertainty in a collection of possible observables A_i with corresponding probability distribution $p_i(A)$ is given by its Shannon entropy $H(A)$ (Shannon and Weaver, 1949):

$$H(A) = -\sum_i p_i(A) \ln p_i(A) \tag{1}$$

This measure is suitable for systems described by classical physics, and is useful to measure uncertainty of observables but it is not suitable for measuring uncertainty of the general state of a quantum system. It is the von Neumann entropy which is appropriate to measure uncertainty of quantum systems since it depends on the density matrix (see below).

Suppose that we have two sets of discrete events A_i and B_j with the corresponding probability distributions, $p_i(A)$ and $p_j(B)$. The relative entropy between these two distributions is defined as

$$H(A\|B) = -\sum_i p_i(A) \ln \frac{p_i(A)}{p_i(B)} \tag{2}$$

This function, also known as Kullback-Leibler entropy (Kullback and Leibler, 1951; Raju et al, 1990) is a measure of the ''distance'' between $p_i(A)$ and $p_j(B)$, even though, strictly speaking, it is not a mathematical metric since it fails to be symmetric:

$$H(A\|B) \neq H(B\|A) \tag{3}$$

Another important concept derived from relative entropy concerns the gathering of information. When one system learns something about another, their states become correlated. How correlated they are, or how much information they have about each other, can be quantified by the mutual information. The Shannon mutual information between two random variables A and B, having a joint probability distribution $p_{ij}(A,B)$ and marginal probability distributions

$$p_i(A) = \sum_j p_{ij}(A,B) \text{ and } p_j(B) = \sum_i p_{ji}(B,A) \tag{4}$$

is defined as

$$\begin{aligned} H(A:B) &= H(A) + H(B) - H(A,B) \\ &= \sum_{ij} p_{ij}(A,B) \ln \frac{p_{ij}(A,B)}{p_i(A) p_j(B)} \end{aligned} \tag{5}$$

where H(A,B) is the joint entropy defined as

$$H(A,B) = -\sum_{ij} p_{ij}(A,B) \ln p_{ij}(A,B) \tag{6}$$

which measures the uncertainty about the whole system *AB*.
The mutual information *H(A:B)* can be written in terms of the Shannon relative entropy. In this sense it represents a distance between the distribution *p(A,B)* and the product of the marginals $p(A) \times p(B)$ As such, it is intuitively clear that this is a good measure of correlations, since it shows how far a joint distribution is from the product one in which all the correlations have been removed, or alternatively, how distinguishable a correlated state is from a completely uncorrelated one. So we have

$$H(A:B) = H\left[p(AB) \| p(A) \times p(B) \right] \tag{7}$$

Suppose that we wish to know the probability of observing *B* if *A* has been observed. This is called a conditional probability and is given by

$$p_{ij}(A|B) = \frac{p_{ij}(A,B)}{p_j(B)} \text{ and } p_{ji}(B|A) = \frac{p_{ji}(B,A)}{p_i(A)} \tag{8}$$

Hence the conditional entropy is,

$$\begin{aligned} H(A|B) &= -\sum_{ij} p_{ij}(A,B) \ln \frac{p_{ij}(A,B)}{p_j(B)} \\ &= -\sum_{ij} p_{ij}(A,B) \ln p_{ij}(A|B) \end{aligned} \tag{9}$$

This quantity, being positive, tells us how uncertain we are about the value of *B* once we have learned about the value of *A*. Now the Shannon mutual information can be rewritten as

$$H(A:B) = H(A) - H(A|B) \tag{10}$$

and the joint entropy as

$$H(A,B) = H(B) + H(A|B) \tag{11}$$

Hence, the Shannon mutual information, measures the quantity of information conveyed about the random variable *A(p(B))* through measurements of the random variable *B(p(A))*.

Note also that, unlike the Shannon relative entropy, the Shannon mutual information is symmetric. Besides, according to the properties of the logarithmic functions (Jensen inequality) it can be established that entropy is a concave function, $-\sum p_i x_i \ln \sum p_i x_i \geq -\sum p_i x_i \ln x_i$ meaning that mixing probability distributions increases entropy, whereas the relative entropy is a convex function, $\sum x_i \ln\left(\sum x_i / \sum a_i\right) \leq \sum x_i \ln\left(x_i / a_i\right)$ i.e., mixing decreases the distance between states (less distinguishable).

The difference between classical and quantum entropies can be seen in the fact that quantum states are described by a density matrix ρ (and not just probability vectors). The density matrix is a positive semidefinite Hermitian matrix, whose trace is unity. An important class of density matrices is the idempotent one, i.e., $\rho = \rho^2$. The states these matrices represent are called pure states. When there is no uncertainty in the knowledge of the system its state is then pure. Another important concept is that of a composite quantum system, which is one that consists of a number of quantum subsystems. When those subsystems are entangled it is impossible to ascribe a definite state vector to any one of them, unless we deal with a bipartite composite system. The most often cited entangled system is the Einstein-Podolsky-Rosen state (EPR) (Einstein et al., 1935: Bell, 1987), which describes a pair of two photons. The composite system is described by $\Psi(1,2) = \frac{\left(|\uparrow(1)\rangle|\downarrow(2)\rangle\right] - \left[|\uparrow(2)\rangle\right]\left[|\downarrow(1)\rangle\right])}{\sqrt{2}}$ which represents the spin directions along the z axis that can either be up or down. We can immediately see that neither of the photons possesses a definite state vector, then if a measurement is made on one photon, let say in the up state, then the other photon will be in the down state. This "assignment" cannot be applied to a general composite system unless its general state is written in a diagonal decomposable form, which not only is mathematically convenient, but also gives a deeper insight into correlations between the two subsystems. According to quantum mechanics the state vector of a composite system, consisting of subsystems *A* and *B*, is represented by a vector belonging to the tensor product of the two Hilbert spaces $H_A \otimes H_B$. The general state of this system can be written as a linear superposition of products of individual states:

$$\Psi^{AB} = \sum_m \sum_n c_{mn} \xi_m(A) \psi_n(B) \tag{12}$$

Where $\{\xi_m(A)$; m=1 to *M*$\}$ and $\{\psi_n(B)$; n=1 to *N*$\}$ are the basis of the subsystems *A* and *B*, respectively. This state can always be decomposed in the Schmidt diagonal form:

$$\Psi^{AB} = \sum_l \lambda_l \chi_l(A) \varphi_l(B) \tag{13}$$

Where $\chi_l(A)$ and $\varphi_l(B)$ are orthonormal bases for *A* and *B*, respectively. Note that in this form the correlations between the two subsystems are completely revealed. If *A* is found in the state $\chi_p(A)$, for example, then the state of B is in the $\varphi_p(B)$ state. This is clearly a multistate generalization of the EPR-state mentioned earlier.

In order to understand the correlation between two subsystems in a joint pure state we point out that the reduced density matrices of both subsystems, written in the Schmidt

decomposed state above, are diagonal and have the same positive spectrum. In particular, the overall density matrix is given by

$$\rho = \sum_{nm} \lambda_n \lambda_m^* \; |\chi_n(A)\rangle\langle\chi_m(A)| \otimes |\phi_n(B)\rangle\langle\phi_m(B)| \tag{14}$$

whereas the reduced ones are

$$\begin{aligned} \rho_A &= \sum_n \langle\phi_n(B)|\rho|\phi_n(B)\rangle \\ &= \sum_m |\lambda_m|^2 \, |\chi_m(A)\rangle\langle\chi_m(A)| \end{aligned} \tag{15}$$

and in analogous way

$$\rho_B = \sum_n |\lambda_n|^2 \, |\phi_n(B)\rangle\langle\phi_n(B)| \tag{16}$$

It is important to note that a *N*-dimensional subsystem can then be entangled with no more than *N* orthogonal states of another one. Schmidt decomposition is, in general, not practical for more than two entangled subsystems since for say *n* entangled systems is uncertain to know at the same time a general state such that by observing the state of one of the subsystems we could instantaneously know the state of the other *n-1*. Clearly, involvement of *n*-subsystems complicates the analysis and produces an even greater mixture and uncertainty. The same reasoning applies to mixed states of two or more subsystems (i.e., states whose density operator is not idempotent), for which we cannot have the Schmidt decomposition in general.

When two subsystems become entangled, the composite state can be expressed as a superposition of the products of the corresponding Schmidt basis vectors. From Eq. (13) it follows that the ith vector of either subsystem has a probability of $|\lambda|^2$ associated with it. We are, therefore, uncertain about the state of each subsystem, the uncertainty being larger if the probabilities are evenly distributed. Since the uncertainty in the probability distribution is naturally described by the Shannon entropy, this classical measure can also be applied in quantum theory. In an entangled system this entropy is related to a single observable. The general state of a quantum system, is described by its density matrix ρ. Let the observables a_i and b_j, pertaining to the subsystems A and B, respectively, have a discrete and non degenerate spectrum, with probabilities $p_i(A)$ and $p_j(B)$ In addition, let the joint probability be $p_{ij}(A,B)$. Then

$$\begin{aligned} H(A) &= -\sum_i p_i(A) \ln p_i(A) \\ &= -\sum_{ij} p_{ij}(A,B) \ln \sum_j p_{ij}(A,B) \end{aligned} \tag{17}$$

and similarly for *H(B).*

An indication of correlation is that the sum of the uncertainties in the individual subsystems is greater than the uncertainty in the total state. Hence, the Shannon mutual information is a good indicator of how much the two given observables are correlated. However, this quantity, as it is inherently classical, describes the correlations between single observables only.

2.2 Quantum information theory

The quantity that is related to the correlations in the overall state as a whole is the von Neumann mutual information which depends on the density matrix. The von Neumann entropy (von Neumann, 1955), may be considered as the proper quantum analog of the Shannon entropy (Wehrl, 1978) for a system described by a density matrix ρ, and is defined as

$$S(\rho) = -Tr(\rho \ln \rho) \tag{18}$$

The Shannon entropy is equal to the von Neumann entropy only when it describes the uncertainties in the values of the observables that commute with the density matrix, i.e., if ρ is a mixed state composed of orthogonal quantum states, otherwise

$$S(\rho) \leq H(A) \tag{19}$$

where A is any observable of a system described by ρ. This means that there is more uncertainty in a single observable than in the whole of the state (Vedral, 2002).

Let ρ_A and ρ_B be the reduced density matrices of subsystems A and B, respectively, and let ρ be the matrix of a composite system, then the entropies of two subsystems are somewhat analogous to its classical counterpart, but instead of referring to observables it is related to the two states which are bounded by the following Araki-Lieb (1970) inequality

$$S(\rho_A) + S(\rho_B) \geq S(\rho) \geq \left| S(\rho_A) - S(\rho_B) \right| \tag{20}$$

Physically, the left-hand side implies that we have more information (less uncertainty) in an entangled state than if the two states are treated separately, hence by treating the subsystems separately the correlations (entanglement) are being neglected. Also, equality in the left-hand side holds when both systems are independent for ρ_A, i.e., if the composite system is in a pure state, then S(ρ)=0, and from the right-hand side it follows that $S(\rho_A)=S(\rho_B)$ (Schmidt decomposition Eq. (13)).

As in the classical case, two important relations can be established (Wehrl, 1978), namely, the entropies of independent systems add up

$$S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B) \tag{21}$$

Further, concavity reflects the fact that mixing states increases uncertainty, i.e.

$$S\left(\sum \lambda_i \rho_i\right) \geq \sum \lambda_i S(\rho_i) \tag{22}$$

According to the definition of the Shannon mutual information which relates only two observables, a quantum analog can be defined which measures the correlation between whole subsystems. The von Neumann mutual information between two subsystems ρ_A and ρ_B of a joint state ρ_{AB} is defined a

$$S(\rho_A : \rho_B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \tag{23}$$

As in the case of the Shannon mutual information this quantity can be interpreted as a distance between two quantum states, the correlated joint state (ρ_{AB}) and the uncorrelated one $\rho_A \otimes \rho_B$, which may be represented through a relative entropy

$$S(\rho_A : \rho_B) = S(\rho_{AB} \| \rho_A \otimes \rho_B) \tag{24}$$

Hence, the relative quantum entropy is an important quantity to classify and quantify quantum correlations (Wehrl, 1978; Vedral, 2002). This measure (Eq. 24) possesses important properties. It is invariant to unitary transformations (the distance between states can not be affected under a change in the basis)

$$S(\rho \| \sigma) = S(U\rho U^+ \| U\sigma U^+) \tag{25}$$

Partial tracing over a part of the system produces a loss of information and hence the subsystems are more difficult to distinguish

$$S(Tr\rho \| Tr\sigma) \leq S(\rho \| \sigma) \tag{26}$$

Therefore, the relative entropy decreases under any combination of these two operations which means that quantum distinguishability never increases.
In order to determine the properties of any good measure of entanglement we have to establish that a bipartite state is "disentangled" if it is in a separable form

$$\rho_{AB} = \sum_i \lambda_i \rho_i^A \otimes \rho_i^B \tag{27}$$

These are the most general states that can be created by local operations and classical communication, Eqs (25) and (26), which contain no quantum correlations as entanglement can only be created through global operations (Wehrl, 1978; Vedral, 2002). Then in order to quantify entanglement is necessary to establish the following: (i) For a disentangled state (separable), the measure of entanglement should be zero, E(ρ)=0, (ii) under any local unitary transformations there is only a change of basis, which is completely reversible for the given entangled state, and then a change of basis should not change the amount of entanglement, i.e.,

$$E(\sigma) = E(U_A \otimes U_B \sigma U_A^+ \otimes U_B^+) \tag{28}$$

Finally, local operations, classical communication and tracing of an ensemble σ which is transformed into subsystems σ_i with probabilities p_i, can not increase the expected entanglement. i.e.,

$$E(\sigma) \geq \sum_i p_i E(\sigma_i) \tag{29}$$

In summary we can conclude that in order to quantify quantum correlations between entangled subsystems, a good measure of quantum correlation has to be non-increasing under local operations (acting separately on A and B), and hence the only way the subsystems become entangled and gain information about each other is by interacting. We will return to this important conclusion in section 2.3.

2.3 Natural atomic probabilities in information theory

We have recently shown (Carrera et al., 2010) that there is an information-theoretic justification for performing Lowdin symmetric transformations (Löwdin, 1970) on the

atomic Hilbert space, to produce orthonormal atomic orbitals of maximal occupancy for the given wavefunction, which are derived in turn from atomic angular symmetry subblocks of the density matrix, localized on a particular atom and transforming to the angular symmetry of the atoms. This alternative information derivation (Carrera et al., 2010) was achieved by minimizing the entropy deficiency between the joint density ρ^{AB} (a reduced first density matrix of a composite fermionic system of the subsystems A and B) with respect to the atomic independent subsystems ρ^A and ρ^B, such that

$$\delta\left\{S\left(\rho^{AB}:\rho^A\otimes\rho^B\right)-\lambda(Tr(\rho^{AB})-1)\right\}=0 \tag{30}$$

according to the constraint

$$Tr(\rho^{AB})=1 \tag{31}$$

The advantages of these kind of atoms-in-molecules (AIM) approaches (Reed & Weinhold, 1983; Davidson, 1967)) are that the resulting natural atomic orbitals are *N*- and *v*-representables (Carrera et al., 2010), positively bounded, and rotationally invariant (Reed et al., 1985; Bruhn et al., 2006). An analogous information-theoretic approach was derived (Nalewajski, 2003) in relation with the Hirshfeld stockholder partitioning of the molecular electron density in Cartesian space (Hirshfeld, 1977).

2.4 Hilbert space partitioning in molecular fragments

We have proposed (Flores-Gallegos & Esquivel, 2008) that a molecule might be considered as a system formed by atomic subsystems, which could be studied through atomic or molecular fragments by means of natural atomic probabilities (Carrera et al., 2010). These probabilities are obtained by diagonalizing the atomic blocks (one center local transformation) of the molecular density matrix which transforms as angular symmetry representations of the isolated atoms, the resulting orthonormal orbitals are thus naturally optimal for the atom in the molecular binding environment. Then, the whole set of diagonalized atomic orbitals is symmetrically orthogonalized as to remove the interatomic overlap, while preserving the atom-like character of the orbitals as nearly as possible (Reed et al., 1985). Thus, the natural atomic probabilities are obtained by local unitary transformations and partial tracing of the molecular density matrix which should decrease the entanglement (Eqs. 26 and 29) by losing information between subsystems. The resulting density matrix is atomic-block diagonal and its spectral decomposition reduces to the atomic angular symmetry instead of the irreducible representation of the symmetry point group of the molecule, hence it can not be reduced to a convex sum of independent subsystems, and therefore its entanglement is not zero. It has been discussed that marginal density matrices with trace-class operators may have their own diagonal representations in terms of orthonormal and complete states in their respective subspaces which do not have marginal (subsystem) probabilities of the composite probability and as a result, the conditional entropies may be negative (Rajagopal et al., 2002); Cerf & Adami, 1997). In this study we have restricted ourselves to the study of a class of entropies (H-type) which possess marginal probabilities of molecular fragments. Thus, we may define atomic density operators through natural atomic probabilities in Cartesian space

$$\rho^A=\sum_{ilm}p_{ilm}(A)\left|\chi_{ilm}(A)\right|^2 \tag{32}$$

And then we may define molecular fragments in an analogous way

$$\rho^{M} = \sum_{A=1}^{M} \rho(A) \tag{33}$$

In Hilbert space we may define a measure of quantum correlations between molecular fragments for a bipartite system through natural atomic probabilities and their joint probability. As we mentioned before, in the study of a bipartite system decomposed through a Schmidt orthogonalization, there are no more than N states that might be entangled (Eqs. 15 and 16). In the natural atomic decomposition scheme we employ there are m states pertaining to molecular fragment A, i.e., $\{p_i(A);$ i= 1 to $m\}$ with n states corresponding to molecular fragment B: $\{ p_j(B);$ j= 1 to $n\}$, thus, we may define the joint entropy through global operations by correlating m x n states as providing that the following constraints are met:

$$\sum_i \sum_j P_{ij}(A,B) = \sum_i P_i(A) = \sum_j P_j(B) = \sum_i P_{ij}(A/B) = 1 \tag{34}$$

And the marginal probabilities are written as

$$P_{ij}(A \mid B) = \frac{P_{ij}(A,B)}{\sum_i P_{ij}(A,B)} \tag{35}$$

We are now in position of using definitions of section 2.1, related to the von Neumann entropies, taking into account that in our natural atomic scheme of probabilities, equality in Eq. (19) holds, and instead of referring to observables we deal with subsystems (molecular fragments), that is why von Neumann entropies are adequate for our study, though we keep the H-terminology to emphasize the orthogonal and commuting properties of the subspaces we are dealing with. It is easy to show that all relations concerning to $H(A)$, $H(B)$, $H(A,B)$, $H(A:B)$ and $H(A \mid B)$ are fulfilled with the definitions above (Eqs 1-11), along with some useful inequalities which follow immediately from definitions in Sec 2.1, i.e.,

$$\begin{aligned}
&H(A) \geq 0 \\
&H(A \mid B) \geq 0 \\
&H(A \mid B) \leq H(A) \\
&H(A,B) \leq H(A) + H(B) \\
&H(A:B) \geq 0 \\
&H(A:B) \equiv H(B:A) \\
&H(A,B) \equiv H(B,A) \\
&H(A,B) \geq H(A:B)
\end{aligned} \tag{36}$$

3. Chemical processes

Theoretic-information measures of the Shannon type have been employed to describe the course of the simplest hydrogen abstraction and the identity S_N2 exchange chemical reactions (Esquivel et al., 2009). For these elementary chemical processes, the transition state

is detected and the bond breaking/forming regions are revealed. A plausibility argument of the former is provided and verified numerically. It is shown that the information entropy profiles posses much more chemically meaningful structure than the profile of the total energy for these chemical reactions. Results support the concept of a continuum of transient of Zewail and Polanyi for the transition state rather than a single state, which is also in agreement with reaction force analyses. Furthermore, the information-theoretic description of the course of these elementary chemical reactions allowed a phenomenological description of their chemical behavior by use of Shannon entropic measures in position and momentum spaces. Interestingly, the analyses also revealed their synchronous/ asynchronous mechanistic behavior (Esquivel et al., 2010).

3.1 Phenomenological description of elementary chemical reactions

The prediction, from first principles, of the structure and energetics of molecules when exerting physical changes such as dissociations or chemical reactions, constitutes a major activity of theoretical/computational chemistry. Such an endeavour has not been an easy one though, involving several research areas which have provided solid grounds and fertile soil for theories and models that have pervaded over the years in ongoing research efforts that have been thoroughly discussed in the literature (Hoffman et al., 2003). We will briefly review here some of the ones that have brought up the need of characterizing chemical processes in terms of physical phenomena such as charge depletion/accumulation, bond breaking/forming, path reaction-following etc.

In an attempt to understand the stereochemical course of chemical reactions, calculations of potential energy surfaces have been carried out extensively at various levels of sophistication (Schlegel, 1988). Within the broad scope of these investigations, particular interest has been focused on extracting information about the stationary points of the energy surface. Considering the Born-Oppenheimer approximation, minima on the N-dimensional potential energy surface for the nuclei can be identified with the classical picture of equilibrium structures of molecules and saddle points can be related to transition states and reaction rates. Since the formulation of transition-state (TS) theory (Eyring, 1935; Wigner, 1939), a great effort has been devoted to developing models for characterizing the TS, which is assumed to govern the height of a chemical reaction barrier, so that any insights into the nature of the TS are likely to provide deeper understanding of the chemical reactivity. Computational quantum chemistry has sidestepped the inherent problems by managing rigorous mathematical definitions of "critical points" on a potential energy hypersurface, and hence assigned to equilibrium complexes or transition states. Within this approach, minima and saddle points have been fully characterized through the first and second derivatives of the energy (gradient and Hessian) over the nuclei positions. That is, if there is more than one minimum on a continuous energy surface, a family of paths can be obtained that connect one minimum to the other and if the highest energy point on each path is considered, the TS can be defined then as the lowest of these maxima along the reaction path, and a minimum for all displacements perpendicular to that path, i.e., a first-order saddle point. The eigenvector corresponding to the single negative eigenvalue of this critical point is the transition vector. It is so that the steepest-descent path from the saddle point to either of the two minima (reactants or products) follows this vector. Since the minima and the transition state are well defined points on the energy surface, it is possible to define a unique reaction path, though it would depend on the particular choice of the coordinate

system. This problem was solved by defining an intrinsic reaction path independently of the coordinate system by appealing to classical mechanics in which the motion equations are the simplest in mass-weighted Cartesian coordinates (Fukui, 1981). Then an intrinsic reaction coordinate (IRC) can be defined as the path traced by a classical particle moving with infinitesimal velocity from a saddle point down to the minima and the IRC is unique in virtue of the classical equations of motion which should be the same in any coordinate system. In mass-weighted coordinates the IRC and the steepest-descent path are the same. Several computational techniques which calculate energy gradients and Hessians have been developed to follow such reaction paths (González & Schlegel, 1990; Peng et al., 1996; Fan & Ziegler, 1992; Safi, 2001; Pople et al., 1978; González-García et al., 2006; Ishida, et al. 1977; Schmidt, et al. , 1985; Baskin et al., 1974).

Notwithstanding that critical points of the energy surface are useful mathematical features for analyzing the reaction-path, their chemical or physical meaning remains uncertain (Shaik, et al., 1994). Chemical concepts such as the reaction rate and the reaction barrier have been thoroughly studied and yet, the pursuit for understanding the TS structure represents a challenge of physical organic chemistry. The many efforts to achieve the latter have produced chemically useful descriptions of the TS such as the one provided by Hammond and Leffler (Hammond, 1955; Leffler, 1953). Hammond postulated that two points on a reaction profile that are of similar energy will also be of similar structure. This allowed predictions regarding the structure of the transition state to be made in highly exothermic and endothermic reactions. Leffler generalized the idea to the entire range of reaction exothermicities by considering the TS as a hybrid of reactants and products whose character is intermediate between these two extremes. These ideas remain as a very practical tool for analyzing chemical reactions now referred as to the Hammond-Leffler postulate which generally states that the properties of the TS are intermediate between reactant and product and are related to the position of the TS along the reaction coordinate.

With the advent of femtosecond time-resolved methods the aforementioned theories have turned out to be more relevant at the present time. Since the seminal studies of Zewail and co-workers (Zewail, 1988, 1990, 2000a, 2000b), femtochemistry techniques have been applied to chemical reactions ranging in complexity from bond-breaking in diatomic molecules to dynamics in larger organic and biological molecules, providing new insights into the understanding of fundamental chemical processes, on Zewail's words: "chemistry on the femtosecond time scale, can be defined as the field of chemical dynamics concerned with the very act of breaking or making a chemical bond. On this time scale the molecular dynamics are "frozen out", and thus one should be able to observe the complete evolution of the chemical event, starting from time zero, passing through transition states, and ultimately forming products". Although most femtochemistry studies deal with excited-state processes, ground-state processes have been studied as well. One of the most promising techniques, the anion photodetachment spectra (Bradforth, 1993), has made possible the direct observation of transition states. In order to explain the experimental results of the femto-techniques it would be necessary to complement the existing chemical reactivity theories with electronic density descriptors of the events taking place in the vicinity of the transition-state region, where the chemical bonds are actually being formed or destroyed.

In connection with the above, there are a number of studies in the literature which have employed a variety of descriptors either to study the TS structure or to follow the course of the chemical reaction path. For instance, Shi and Boyd performed a systematic analysis of

model S_N2 reactions in order to study the TS charge distribution in connection with the Hammond-Leffler postulate (Shi & Boyd, 1991). Bader et al developed a theory of reactivity based solely on the properties of the charge density by employing the properties of the Laplacian of the density so as to align the local charge concentrations with regions of charge depletion of the reactants by mixing in the lowest energy excited state of the combined system to produce a relaxed charge distribution corresponding to the transition density (Bader & MacDougall, 1985). By studying the time evolution of a bimolecular exchange reaction Balakrishnan & Sathyamurthy (1989) showed that information-theoretic entropies in dual or phase space rised to a maximum in a dynamical study. Following the course of two elementary S_N2 reactions, Ho et al showed that theoretic-information measures were able to reveal geometrical changes of the density which were not present in the energy profile, although the transition states were not apparent from the study (Hô et al., 2000). In an attempt to build a density-based theory of chemical reactivity, Knoerr and Eberhart (2001), reported correlations between features of the quantum mechanically determined charge density and the energy-based measures of Shaik and collaborators to describe the charge transfer, stability, and charge localization accompanying an S_N2 reaction (Shaik et al., 1992). Moreover, Tachibana (2001) was able to visualize the formation of a chemical bond of selected model reactions by using the kinetic energy density to identify the intrinsic shape of the reactants, the TS and the reaction products along the course of the IRC, hence realizing Coulson's conjecture (1991) in that the physical meaning of the probability density might be related with the energy density. The reaction force of a system's potential energy along the reaction coordinate has been employed to characterize changes in the structural and/or electronic properties in chemical reactions (Toro-Labbé, et al., 2009; Toro-Labbé et al., 2007; Murray et al., 2009; Jaque et al., 2009) . The Kullback–Leibler information deficiency has been evaluated along molecular internal rotational or vibrational coordinates and along the intrinsic reaction coordinate for several S_N2 reactions (Borgoo et al., 2009).

Notwithstanding that there has been a great interest in the last twenty years in applying Information Theory (IT) measures to the electronic structure of atoms and molecules (Gadre, 2003; Koga & Morita, 1983; Ghosh et al., 1984; Angulo & Dehesa, 1992; Antolín et al., 1993; Angulo, 1994; Massen & Panos, 1998; Ramirez et al., 1998; Nalewajski & Parr, 2001; Nagy, 2003; Romera & Dehesa, 2004; Karafiloglou & Panos, 2004; Sen, 2005; Parr et al. , 2005; Guevara et al., 2005; Shi & Kais, 2005; Chatzisavvas et al., 2005; Sen & Katriel, 2006; Nagy, 2006; Ayers, 2006; Martyusheva & Seleznev, 2006; Liu, 2007), it has not been clearly assessed whether theoretic-information measures are good descriptors for characterizing chemical reaction parameters, i.e., the stationary points of the IRC path (the TS and the equilibrium geometries of the complex species) and the bond breaking/forming regions. The purpose of the study was to perform a phenomenological description of two selected elementary chemical reactions by following their IRC paths with the purpose of analyzing the behaviour of the densities in position and momentum spaces, at the vicinity of the TS, and also at the regions of bond forming/breaking that are not visible in the energy profile, by use of the Shannon entropies in conjugated spaces. In order to witness the density changes exerted by the molecular structures and link them with the theoretic-information quantities during the chemical processes, we will employ several charge density descriptors such as the Molecular Electrostatic Potential (MEP) and some reactivity parameters of Density Functional Theory (DFT), the hardness and the softness. The chemical probes under study are the simplest hydrogen abstraction reaction $H^{\bullet} + H_2 \longrightarrow H_2 + H^{\bullet}$ and the identity S_N2 reaction $H_a^- + CH_4 \longrightarrow CH_4 + H_b^-$.

The central quantities under study are the Shannon entropies in position and momentum spaces (Shannon, 1948), which are analogous to Eq. (1) for the case of the continuous variable distributions, $\rho(\mathbf{r})$ and $\gamma(\mathbf{p})$:

$$S_r = -\int \rho(\mathbf{r}) \ln \rho(\mathbf{r}) d^3\mathbf{r} \tag{37}$$

$$S_p = -\int \gamma(\mathbf{p}) \ln \gamma(\mathbf{p}) d^3\mathbf{p} \tag{38}$$

where $\rho(\mathbf{r})$ and $\gamma(\mathbf{p})$ denote the molecular electron densities in the position and momentum spaces, each normalized to unity. The Shannon entropy in position space S_r behaves as a measure of delocalization or lack of structure of the electronic density in the position space and hence S_r is maximal when knowledge of $\rho(\mathbf{r})$ is minimal and becomes delocalized. The Shannon entropy in momentum space S_p is largest for systems with electrons of higher speed (delocalized $\gamma(\mathbf{p})$) and is smaller for relaxed systems where kinetic energy is low. Entropy in momentum space S_p is closely related to S_r by the uncertainty relation of Bialynicki-Birula and Mycielski (1975), which shows that the entropy sum $S_T=S_r+S_p$, is a balanced measure and cannot decrease arbitrarily. For one-electron atomic systems it may be interpreted as that localization of the electron's position results in an increase of the kinetic energy and a delocalization of the momentum density, and conversely.

In connection with the behaviour of the Shannon entropies above discussed, we expect a simple theoretic-information description of the TS in terms of localized/delocalized distributions. Considering that there is no variational principle for any quantum-mechanical property other than the energy, deriving a direct relation between the TS and the Shannon entropies as functionals of the electron densities seems not practical and surely beyond the scope of the present work. Instead, we present a plausible physical argument to link both quantities: on mathematical grounds the TS represents a first-order saddle point connecting two minima in a topological sense, and physically it represents a maximum in the "potential energy" surface (PES) within the space of all possible nuclear configurations corresponding to the energetically easiest passage from reactants to products constrained to the Born-Oppenheimer approximation. In this sense, TS structure posses (locally) the highest potential energy among all possible chemical structures in that path at the expense of a minimum kinetic energy. Since the transition state theory (Eyring, 1935; Wigner, 1938) is essentially based on the assumption that atomic nuclei behave according to classic mechanics, the presumption for the TS being represented by a chemical structure with a minimum kinetic energy, which corresponds with a highly localized momentum density (locally), seems justified. Simultaneously, the TS will require a highly delocalized density in position space for the uncertainty principle to be satisfied. In this sense the Shannon entropies, as logarithmic descriptors of the electron density distributions in the combined phase space (normalized to unity), would correspond to extrema on the entropy profile at the vicinity of the TS, provided that the densities are adequately represented in the chemical space.

The MEP represents the molecular potential energy of a proton at a particular location near a molecule (Politzer & Truhlar, 1981), say at nucleus *A*. Then the electrostatic potential , V_A , is defined as

$$V_A = \left(\frac{\partial E^{molecule}}{\partial Z_A} \right)_{N, Z_{B \neq A}} = \sum_{B \neq A} \frac{Z_B}{|R_B - R_A|} - \int \frac{\rho(\mathbf{r}) dr}{|\mathbf{r} - R_A|'} \tag{39}$$

where $\rho(\mathbf{r})$ is the molecular electron density and Z_A is the nuclear charge of atom A, located at R_A. Generally speaking, negative electrostatic potential corresponds to an attraction of the proton by the concentrated electron density in the molecules from lone pairs, pi-bonds, etc… (coloured in shades of red in standard contour diagrams). Positive electrostatic potential corresponds to repulsion of the proton by the atomic nuclei in regions where low electron density exists and the nuclear charge is incompletely shielded (coloured in shades of blue in standard contour diagrams).

We have also evaluated some reactivity parameters that may be useful to analyze the chemical reactivity of the processes. Parr and Pearson (1983), proposed a quantitative definition of hardness (η) within conceptual DFT:

$$\eta = \frac{1}{2S} = \frac{1}{2}\left(\frac{\partial \mu}{\partial N}\right)_{v(r)} \tag{40}$$

where $\mu = \left(\frac{\partial E}{\partial N}\right)_{v(r)}$ is the electronic chemical potential of an N electron system in the presence of an external potential ν(**r**), E is the total energy and "S" is called the softness within the context of DFT. Using finite difference approximation, Eq. (39) would be

$$\eta = \frac{1}{2S} \approx \frac{\left(E_{N+1} - 2E_N + E_{N-1}\right)}{2} = \frac{(I - A)}{2} \tag{41}$$

where E_N, E_{N-1} and E_{N+1} are the energies of the neutral, cationic and anionic systems; and I and A, are the ionization potential (IP) and electron affinity (EA), respectively. Applying Koopmans' theorem (Koopmans, 1933; Janak, 1978), Eq. (4) can be written as:

$$\eta = \frac{1}{2S} \approx \frac{\varepsilon_{LUMO} - \varepsilon_{HOMO}}{2} \tag{42}$$

where ε denotes the frontier molecular orbital energies. In general terms, hardness and softness are good descriptors of chemical reactivity, the former measures the global stability of the molecule (larger values of η corresponds to less reactive molecules), whereas the S index quantifies the polarizability of the molecule (Ghanty & Ghosh, 1993; Roy et al., 1994; Hati, & Datta, 1994; Simon-Manso & Fuenteaelba, 1998), thus soft molecules are more polarizable and possess predisposition to acquire additional electronic charge (Chattaraj et al., 2006). The chemical hardness "η" is a central quantity for use in the study of reactivity and stability, through the hard and soft acids and bases principle (Pearson, 1963; 1973; 1997). However, in many cases, the experimental electron affinity is negative rather than positive, an such systems pose a fundamental problem; the anion is unstable with respect to electron loss and cannot be described by a standard DFT ground-state total energy calculation. To circumvent this limitation, Tozer and De Proft have introduced an approximate method to compute this quantity, requiring only the calculation of the neutral and cationic systems which does not explicitly involve the electron affinity (Tozer & De Proft, 2005):

$$\eta = \frac{\varepsilon_{LUMO} - \varepsilon_{HOMO}}{2} + \varepsilon_{HOMO} + I \tag{43}$$

where *I* is obtained from total electronic energy calculations on the *N-1* and *N* electron systems at the neutral geometry $I = E_{N-1} - E_N$ and all energy quantities have to be calculated by continuum approximations such as the local exchange-correlation functionals (GGA) to avoid integer discontinuities. Nevertheless, it has been observed that expression (43) does still work reasonably well with hybrids, such as B3LYP (Tozer & De Proft, 2005). The authors have shown that this approximate method (Eq. 43) provided reasonable estimates for the electron affinities of systems possessing metastable anions, such as the case of CH_4 with large negative experimental electron affinity (-7.8 eV). We have employed Tozer and De Proft approach for computing the hardness of the S_N2 ionic complex in order to test this approximation in a process departing from the ground-state requirement such as the IRP (Intrinsic Reaction Process) of a chemical reaction (De Proft, 2008).
The electronic structure calculations performed in the present study were carried out with the Gaussian 03 suite of programs (Frisch et al., 2004). Reported TS geometrical parameters for the abstraction (Johnson, 1994), and the S_N2 exchange reactions were employed (Shi & Boyd, 1989). Internal reaction coordinate (IRC) calculations (González & Schlegel, 1989) were performed at the MP2 (UMP2 for the abstraction reaction) level of theory with at least 35 points for each one of the reaction directions (forward/reverse) of the IRC path. Then, a high level of theory and a well balanced basis set (diffuse and polarized orbitals) were chosen for determining all the properties for the chemical structures corresponding to the IRC path according to the strategy followed in Esquivel et al., 2009. The hardness and softness chemical parameters were calculated by use of Eqs. (42) and (43) and the standard hybrid B3LYP (UB3LYP for the abstraction reaction) functional (Tozer & De Proft, 2005). Molecular frequencies corresponding to the normal modes of vibration depend on the roots of the eigenvalues of the Hessian (its matrix elements are associated with force constants) at the nuclei positions of the stationary points. We have found illustrative to calculate these values for the normal mode associated with the TS (possessing one imaginary frequency or negative force constant) which were determined analytically for all points of the IRC path at the MP2 (UMP2 for the abstraction reaction) level of theory (Frisch, 2004). The molecular information entropies in position and momentum spaces for the IRC path were obtained by employing software developed in our laboratory along with 3D numerical integration routines (Pérez-Jordá & San-Fabián, 1993; Pérez-Jordá et al., 1994), and the DGRID suite of programs (Kohout, 2007). The bond breaking/forming regions along with electrophilic/nucleophilic atomic regions were calculated through the MEP by use of MOLDEN (Schaftenaar & Noordik, 2000). Atomic units are employed throughout the Chapter unless otherwise is stated.

3.1.1 Radical abstraction reaction

The reaction $H^{\bullet} + H_2 \longrightarrow H_2 + H^{\bullet}$ is the simplest radical abstraction reaction involving a free radical (atomic hydrogen in this case) as a reactive intermediate. This kind of reaction involves at least two steps (S_N1 like): in the first step a new radical is created by homolysis and in the second one the new radical recombines with another radical species. Such homolytic bond cleavage occurs when the bond involved is not polar and there is no electrophile or nucleophile at hand to promote heterolytic patterns. When the bond is made, the product has a lower energy than the reactants and it follows that breaking the bond requires energy. Evidence has been presented (Esquivel et al., 2010) which shows that the two step mechanism observed for this type of reaction is completely characterized by an asynchronous behaviour but yet "concerted".

Our calculations for this reaction were performed at two different levels, the IRC was obtained at the UMP2/6-311G level and all properties at the IRC path were obtained at the QCISD(T)/6-311++G** level of theory. As a result of the IRC, 72 points evenly distributed between the forward and reverse directions of the reaction were obtained. A relative tolerance of $1x10^{-5}$ was set for the numerical integrations (Pérez-Jordá & San-Fabián, 1993; Pérez-Jordá et al., 1994).

In Fig. 1, the energy profile for the process is depicted against the intrinsic reaction coordinate (R_X) which shows the symmetric behavior of the IRC path. Also, in Fig. 1 we have depicted the entropy sum, which shows the exact opposite behavior as that of the energy, i.e., the TS represents a chemical structure with a localized density in the combined space of position and momentum (dual or phase space), which corresponds to a more delocalized position density with the lowest kinetic energy (more localized momentum density) among all the structures at the vicinity of the IRC path (see below). In this way the saddle point might be characterized by IT in the entropy hyper-surface.

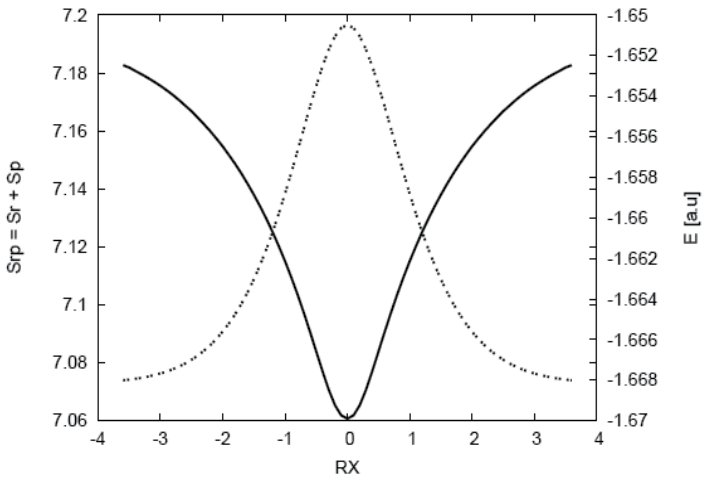


Fig. 1. Total energy values (dashed line) in a.u. and the entropy sum (solid line) for the IRC path of $H_a^{\bullet} + H_2 \longrightarrow H_2 + H_b^{\bullet}$

The Shannon entropies in position and momentum spaces for the abstraction reaction are depicted in Fig. 2 in order to characterize their critical points at the IRC path by use of several density descriptors discussed below. At first glance, we may note from Fig. 2 that the position entropy possesses a local maximum at the TS and two minima at its vicinity, whereas the momentum entropy possesses a minimum at the TS with two maxima at its vicinity, hence we observe that both quantities behave in opposite ways, i.e., the Shannon entropy in position space shows larger values toward the reactant/product complex ($H_a^{\bullet}\cdots H\text{-}H_b$ or $H_a\text{-}H\cdots H_b^{\bullet}$) and tends to decrease toward the TS region. In contrast, the momentum entropy increases as the intermediate radical ($H_a^{\bullet}$) approaches the molecule, reaching maxima at the vicinity of the TS. This behaviour is interpreted as follows: the position entropy values are smaller at the vicinity of the TS region as compared with the ones at the reactive complex region (towards reactants and products) since the densities of the chemical structures are globally more localized at the TS region (see Fig. 1 in connection

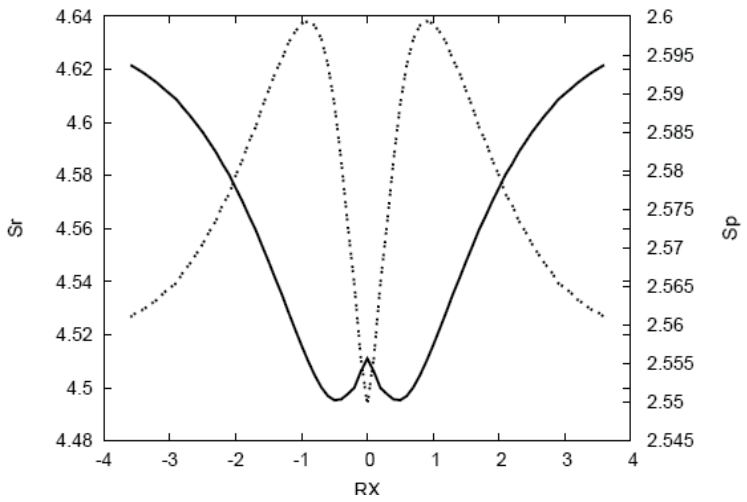


Fig. 2. Shannon entropies in position (solid line) and momentum (dashed line) spaces for the IRC path of $H_a^{\bullet} + H_2 \longrightarrow H_2 + H_b^{\bullet}$

with the entropy sum), which is the zone where the important chemical changes take place. On the perspective of the momentum entropy, we may note that it is minimal at the TS which is linked to a more localized momentum density possessing the lowest kinetic energy value (maximum at the potential energy surface). At the reactive complex regions, momentum entropy values are larger than at the TS and therefore the corresponding kinetic energies are larger too, hence reproducing the typical potential energy surface shown in Fig. 1.

In Fig. 3, the bond distances (in Angstroms) between the entering/leaving hydrogen radicals and the central hydrogen atom are depicted. This clearly shows that in the vicinity of the TS a bond breaking/forming chemical situation is occurring since the R_{in} is elongating at the right side of the TS and the R_{out} is stretching at the left side of the TS. It is worth

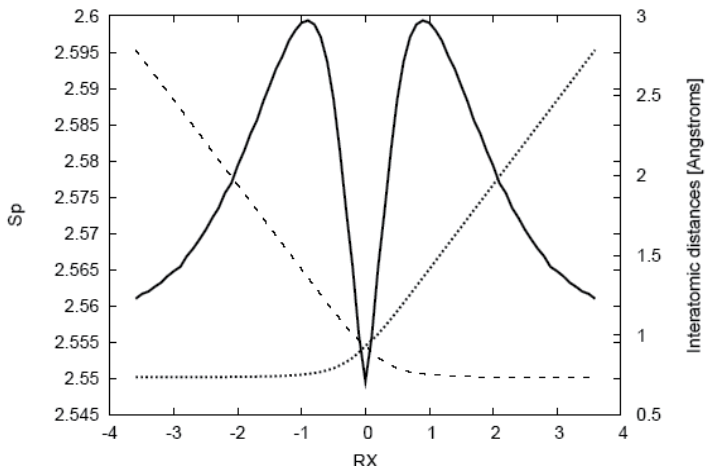


Fig. 3. Shannon entropy in momentum space (solid line) and the bond distances $R(H_0\text{-}H_{in})$ (dashed line for the entering hydrogen) and $R(H_0\text{-}H_{out})$ (dotted line for the leaving hydrogen) in Angstroms for the IRC path of $H_a^{\bullet} + H_2 \longrightarrow H_2 + H_b^{\bullet}$

noting that the chemical process does not happen in a concerted manner, i.e., the homolytic bond breaking occurs first and then the molecule stabilizes by forming the TS structure which is clearly observed in the Fig. 3. As the incoming radical approaches the molecule the bond breaks, at the same location where the position entropy is minimum and the momentum entropy is maximum, then the TS is reached and the new molecule is formed afterwards. This is in agreement with the discussion above with regard to the two step mechanism characterizing this reaction.

The non-polar bond pattern characteristics of homolytic bond-breaking kind of reactions has been studied through the dipole moment of the molecules at the IRC path (Esquivel et al., 2009). This is indeed observed in Fig. 4, where these values along with the ones of the momentum entropy are depicted for comparison purposes. At the TS the dipole moment is zero, and the same is observed as the process tends to the reactants/products in the reaction path, reflecting the non-polar behavior of the molecule in these regions. However, it is also interesting to observe from this property, how the molecular densities get distorted, reaching maximal values at the vicinity of the TS, where the position entropies are minimal, i.e., at the bond breaking/forming regions the complex exerts its largest distortion, molecular geometry gets rigid where the position density is more localized. Once more, we may observed from Fig. 4 that the energy reservoirs for the bond cleavage occur earlier (or later depending on the direction of the reaction) at the IRC path as observed from the maxima of the momentum entropies. We will refer to these chemical regions as to bond cleavage energy reservoirs (BCER) in what follows.

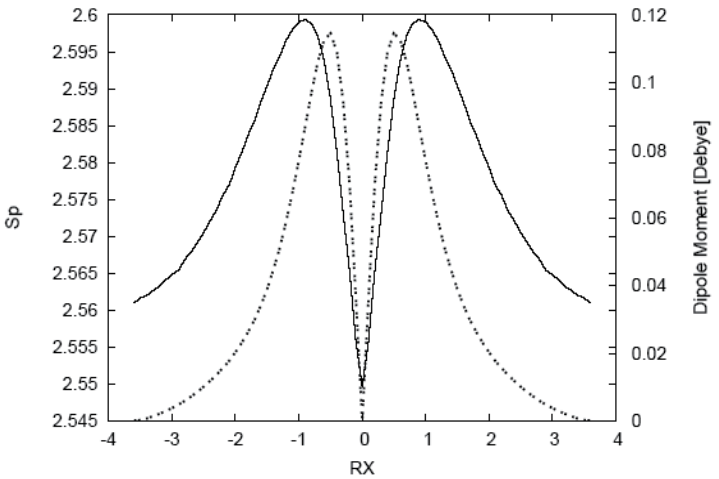


Fig. 4. Shannon entropy in momentum space (solid line) and the dipole moment values in Debye (dashed line) for the IRC path of $H_a^{\bullet} + H_2 \longrightarrow H_2 + H_b^{\bullet}$.

In Fig. 5 the eigenvalues of the Hessian for the normal mode associated with the TS along the path of the reaction are depicted along with the momentum entropy values for comparison purposes. These Hessian values represent the transition vector "frequencies" which show maxima at the vicinity of the TS and a minimal value at the TS. Several features are worth mentioning, the TS corresponds indeed to a saddle point, maxima of the Hessian correspond to high kinetic energy values (largest "frequencies" for the energy cleavage reservoirs) since they fit with maximal values in the momentum entropy profile, and the Hessian is minimal at the TS, where the kinetic energy is the lowest (minimal molecular

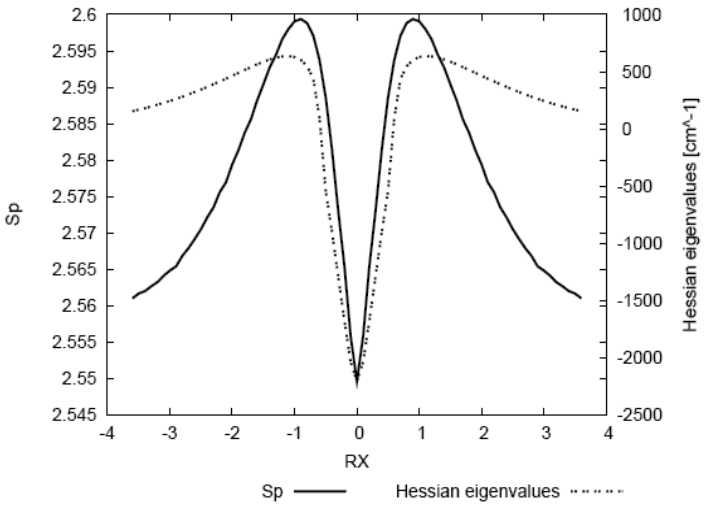


Fig. 5. Shannon entropy in momentum space (solid line) and the eigenvalues of the Hessian (dashed line) for the IRC path of $H_a^{\bullet} + H_2 \longrightarrow H_2 + H_b^{\bullet}$. It should be noted that negative values actually correspond with imaginary numbers (roots of negative force constants) so that the negative sign only represents a flag.

frequency) and it corresponds to a minimal momentum-entropy value. Furthermore, the transition state of a reaction is commonly identified by the presence of a negative force constant for one normal vibrational mode corresponding with an imaginary frequency. The work of Zewail and Polanyi in transition state spectroscopy has led to the concept of a reaction having a continuum of transient, a transition region rather than a single transition state (Zewail, 1988; 1990; 2000a; 2000b; Polanyi & Zewail, 1995). It is worth mentioning that the results of the present study show indeed the existence of such a region between the BCER, before and after the TS. This is in agreement with reaction force, *F(R)*, studies (Toro-Labbé, et al., 2009; Toro-Labbé et al., 2007; Murray et al., 2009; Jaque et al., 2009) where the reaction force constant, *κ(R)*, also reflects this continuum, showing it to be bounded by the minimum and the maximum of *F(R)*, at which *κ(R) = 0*.

The chemical reactivity behavior of the reaction has also been analyzed through density descriptors such as the hardness and softness (Esquivel et al., 2009). From a DFT conceptual point of view, chemical structures with maximal hardness (minimal softness) possess low polarizability and hence are less propense to acquire additional charge (less reactive). These structures are found at the BCER regions, they are maximally distorted, with highly localized position densities (Esquivel et al., 2009).

3.1.2 Hydrogenic identity S_N2 exchange reaction

Continuing with the study of elementary chemical reactions it is of interest to analyze a typical nucleophilic substitution (S_N2) reaction since its chemical process involves only one step in contrast with the two-step S_N1 reaction. In the anionic form, the S_N2 mechanism can be depicted as $Y^- + RX \rightarrow RY + X^-$, which is characterized by being kinetically of second order (first order in each of the reactants; the nucleophile $Y^-$and the substrate RX, where X^- is the nucleofuge or leaving atom). For identity S_N2 reactions X=Y. It was postulated that the observed second order kinetics is the result of passage through the well-known Walden

inversion transition state where the nucleophile displaces the nucleofuge (leaving group) from the backside in a single concerted reaction step. Evidence has been presented (Esquivel et al., 2010) which shows that the one step mechanism observed for this type of reaction is indeed characterized by its synchronous and concerted behaviour.

The $H_a^- + CH_4 \longrightarrow CH_4 + H_b^-$ represents the typical identity S_N2 reaction and we proceed with the calculations as follows: since diffuse functions are important to adequately represent anionic species (Shi & Boyd, 1991), we have performed calculations for the IRC at the MP2/6-311++G** level of theory, which generated 93 points evenly distributed between the forward and reverse directions of the IRC. Then, all entropies and geometrical parameters at the IRC path were calculated at the QCISD(T)/6-311++G** level of theory which has been reported to be adequate for this kind of reactions (Glukhovtsev et al., 1995). A relative tolerance of $1x10^{-5}$ was set for the numerical integrations (Pérez-Jordá & San-Fabián, 1993; Pérez-Jordá et al., 1994).

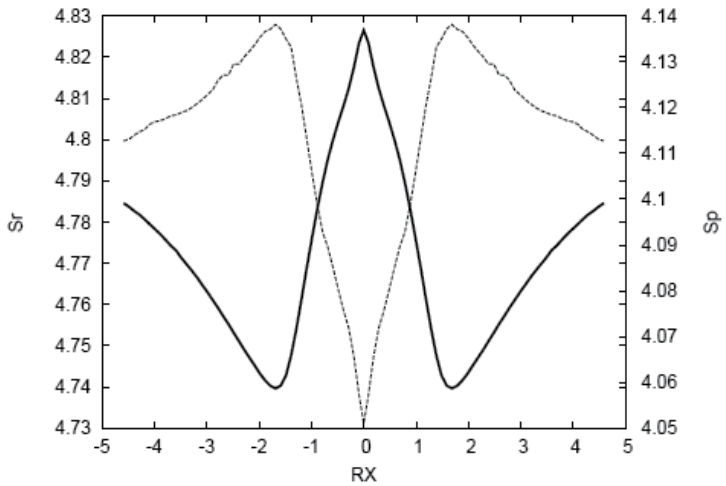


Fig. 6. Shannon entropies in position (solid line) and momentum (dashed line) spaces for the IRC path of the S_N2 reaction at the QCISD(T)/6-311++G** level.

A comparison between the entropy sum (see Fig. 7) and the energy shows that both quantities behave in an opposite manner, although the entropy sum shows much more structure at the vicinity of the TS region as compared to the energy profile (Esquivel et al., 2009). The nature of the richer structure observed for the entropy sum (as compared with the energy) was revealed through the position and momentum entropies depicted in Fig. 6 which show a TS structure characterized by a delocalized position density and a localized momentum density, i.e., corresponding with a structurally relaxed structure with low kinetic energy. In contrast, as compared with the TS; the reactive complex toward reactants/products show more localized position densities with less localized momentum densities, i.e., the chemical structures at these regions are structurally distorted and posses more kinetic energy as compared with the TS. At the vicinity of the TS, at around $|R_X| \approx 1.7$, critical points for both entropies are observed, minima/maxima for the position/momentum entropies, respectively. Thus, ionic complex at these regions characterize position densities which are highly localized and with highly delocalized momentum densities and high kinetic energies. At first glance, it seems likely that these

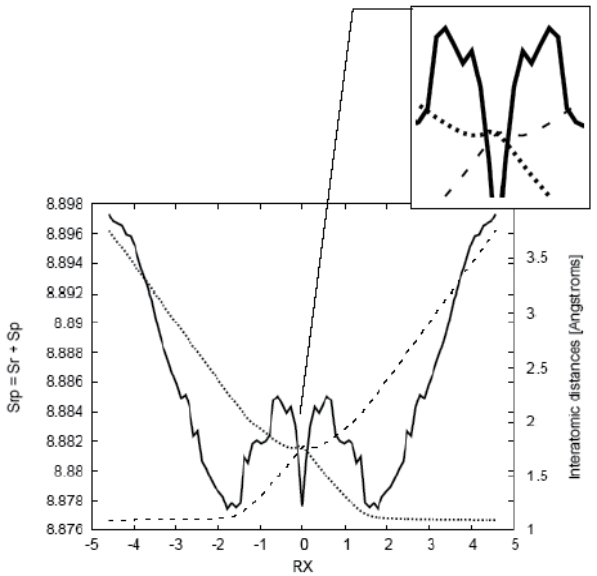


Fig. 7. Shannon entropy in momentum space (solid line) and the bond distance R_a (dotted line), corresponding to the H_a-C distance, and R_b (dashed line) corresponding to the (C-H_b) distance for the IRC path of the S_N2 reaction. In the side frame: detail of the minima observed for the bond distances at $R_X \approx -0.3$. Distances in Angstroms.

regions correspond with BCER where bond breaking may start occurring. Two more features that are worth noting is that both entropies show inflection points at $|R_X| \approx 1.0$ and maxima at $|R_X| \approx 0.5$, regions where the entropy sum shows more defined structure (see Fig. 7), change of curvature and maxima, respectively (Esquivel et al., 2009). We will come back later to these observations in connection with other properties.

In order to support our observations above we find instructive to plot the distances between the incoming hydrogen (H_a) and the leaving hydrogen (H_b) in Fig. 7. Distances show the stretching/elongating features associated with the bond forming/breaking situation that we have anticipated before. In contrast with the previous analyzed abstraction reaction, the S_N2 reaction occurs in a concerted manner, i.e., the bond breaking/forming starts taking place at the same time, in a gradual and more complicated manner as we explain below. An interesting feature which might be observed from Fig. 7 is that whereas the elongation of the carbon-nucleofuge (C-H_b) bond (R_b) changes its curvature significantly at $R_X \approx -1.7$ (forward direction of the reaction) the stretching of the nucleophile-carbon (H_a-C) bond (R_a) does it in a smooth way, posing the argument that bond breaking is occurring first, due to the repulsive forces that the ionic molecule exerts as the nucleophile approaches which provokes the breaking of the carbon-nucleofuge to happen as the molecule starts liberating its kinetic energy (decrease of the momentum entropy). In this sense is that the reaction occurs in a concerted manner, i.e., the bond-breaking/dissipating-energy processes occurring simultaneously. At the near vicinity of the TS, around $R_X \approx -0.3$, we observe small changes for both interatomic distances revealed through minima in the amplified picture, where it is apparent that repulsive forces occur at the TS. Moreover, the analysis of the internal angle between H_a-C-H along with the Shannon entropy in position space for comparison purposes. Thus, the internal angle shows clearly that the molecule starts

exerting the so called "inversion of configuration" at around $R_X \approx -1.7$, where the nucleophile starts displacing the nucleofuge from the backside in a single concerted reaction step. This starts occurring at the BCER regions (see above).
Fig. 8 shows the repulsive effect we mentioned before in connection with the interatomic distances (Fig. 7) by noting that the leaving atom Hb is gaining nucleophilic power (negative MEP). The TS state which is not depicted shows a half and half electrophilic/nucleophilic character among the atoms, where the charge is evenly distributed throughout the molecule.

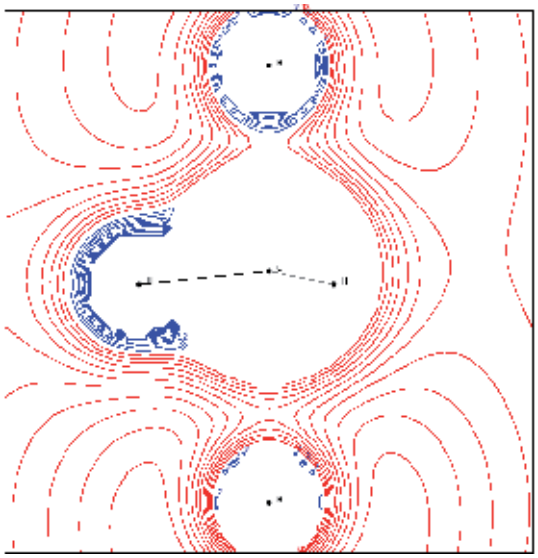


Fig. 8. The MEP contour lines in the plane of H_a-C-H_b (H_a stands for the nucleophilic atom and H_b is the nucleofuge, on bottom and top, respectively) showing positive MEP (nucleophilic regions) and negative MEP (electrophilic regions) at $R_X \approx -0.3$ for the S_N2 reaction.

The S_N2 reaction is an excellent probe to test the polar bond pattern characteristic of heterolytic bond-breaking (with residual ionic attraction because of the ionic nature of the products) which should be reflected through the dipole moment of the molecules at the IRC path (note that the origin of the coordinate system is placed at the molecules's center of nuclear charge). This is indeed observed in Fig. 9, where these values along with the ones of the momentum entropy are depicted for comparison purposes. At the TS the dipole moment is zero showing the non polar character of the TS structure with both nuclephile/nucleofuge atoms repelling each other evenly through its carbon bonding. At this point the momentum/position entropies are minimal/maximal reflecting the low kinetic energy feature of the chemically relaxed TS structure. As the reactive complex approach the reactants/products regions the dipole moment increases monotonically reflecting the polar bonding character of the ionic complex with a significant change of curvature at the TS vicinity at around $|R_X| \approx 1.0$ (a change of curvature was already noted for all entropies at the same region). In going from reactants to products it is evident that the inversion of the dipole moment values reflects clearly the inversion of configuration of the molecule (this reaction starts with a tetrahedral sp^3 carbon in the methyl molecule and ends with a tetrahedral sp3 in the product), which is an inherent feature of S_N2 reactions.

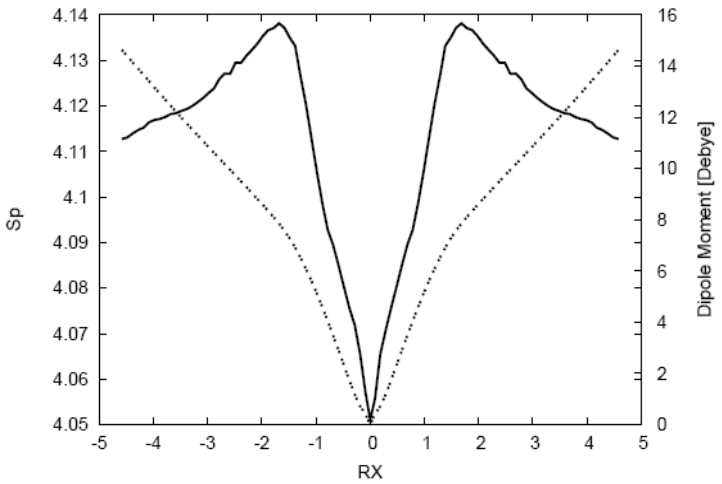


Fig. 9. Shannon entropy in position space (solid line) and the total dipole moment (dotted line) for the IRC path of the S_N2 reaction. Dipole moment in Debye.

It is interesting to note that as in the case of the hydrogenic abstraction reaction, the eigenvalues of the Hessian (Esquivel et al., 2009) for the normal mode associated with the TS along the IRC path show maxima at the BCER and reach their minimal value at the TS. This again validates the concept of a continuum of transient of Zewail and Polanyi, i.e., a transition region rather than a single transition state (see above).

3.2 Reaction mechanisms

A reaction mechanism represents a sequence of elementary steps by which overall chemical change occurs, describing in detail what it takes place at each stage of a chemical transformation, such as the bonds that are being formed or broken, and in what order. The chemical course of a reaction also accounts for the order in which molecules react, either by single- or multi-step conversions, and provides information about the structure of the transition state, reactive complexes, kinetics, catalysis, and stereochemistry. In this section we present a theoretical ab initio study which presents evidence obtained from Shannon theoretic-information concepts in position and momentum spaces that allow a conceptual description of the course of two elementary chemical reactions, revealing all the expected physical transformations predicted for synchronous (one-step) and non-synchronous (two-step) reaction mechanisms.

Dewar (1984) has employed intuitive arguments along with numerical evidence to put forward the notion of "synchronicity being normally prohibited for multibond processes", which is in contrast with the widely accepted Woodward-Hoffman rules (Woodward & Hoffmann, 1969) that establish that multibond "allowed" reactions must be synchronous, i.e., all the bond-forming and bond-breaking processes taking place simultaneously. Furthermore, it has been asserted (Bernasconi, 1992) that a principle for non-perfect synchronization might be derived from the realization that the majority of elementary reactions involve more than one concurrent molecular process such as bond formation/cleavage, delocalization/localization of charge, etc. and that often these processes have made unequal progress at the transition state. For instance, one bond mechanisms are predominant in elementary processes of organic chemistry and most

chemists believe these take place in a synchronous manner. Recently, the simplest prototypical hydrogen exchange reaction (Chandra, 1996) and a variety of other radical exchange reactions have been examined (Chandra, 1999) by use of *ab initio* methods to conclude that despite the fact that these reactions are chemically classified as being concerted (taking place in a single kinetic step) their bond-cleavaging processes are slightly more advanced than the bond-forming ones, proceeding by a two-stages mechanism (Dewar, 1984), and therefore they show asynchronous features.

There has been an increasing interest in the recent years to analyse the electronic structure of atoms and molecules by applying Information Theory (IT) (Gadre, 2003; Koga & Morita, 1983; Ghosh et al., 1984; Angulo & Dehesa, 1992; Antolín et al., 1993; Angulo, 1994; Massen & Panos, 1998; Ramirez et al., 1998; Nalewajski & Parr, 2001; Nagy, 2003; Romera & Dehesa, 2004; Karafiloglou & Panos, 2004; Sen, 2005; Parr et al. , 2005; Guevara et al., 2005; Shi & Kais, 2005; Chatzisavvas et al., 2005; Sen & Katriel, 2006; Nagy, 2006; Ayers, 2006; Martyusheva & Seleznev, 2006; Liu, 2007 These studies have shown that information-theoretic measures are capable of providing simple pictorial chemical descriptions of atoms and molecules and the processes they exert through the localized/delocalized behaviour of the electron densities in position and momentum spaces. In a recent study (Esquivel et al., 2009), we have provided with evidence which supports the utility of the theoretic-information measures in position and momentum spaces to detect the transition state and the stationary points of elementary chemical reactions so as to reveal the bond breaking/forming regions of the simplest hydrogen abstraction and the identity S_N2 exchange chemical processes, thus providing evidence of the concept of a reaction having a continuum of transient of Zewail and Polanyi and also in agreement with reaction force analysis (see above).

The purpose of the present study is to follow the IRC path of the simplest hydrogen abstraction reaction $H_a^{\bullet} + H_2 \longrightarrow H_2 + H_b^{\bullet}$ and the exchange identity S_N2 reaction $H_a^{-} + CH_4 \longrightarrow CH_4 + H_b^{-}$, with the purpose of performing a phenomenological description of two selected elementary chemical reactions with different mechanistic courses by use of Shannon theoretic-information measures in both conjugated spaces, position and momentum.

Abstraction reactions proceed by homolysis and can be characterized by a mechanism being kinetically of first order (S_N1 like). These kind of reactions involve a two-step mechanism, which initiates with the formation of a new radical created by homolysis and continues with the recombination of the new radical with another radical species. Such homolytic bond cleavage occurs when the bond involved is non polar and there is no electrophile or nucleophile at hand to promote heterolytic patterns. When the bond is made, the product has a lower energy than the reactants and it follows that breaking the bond requires energy. In contrast, the hydride-exchange reactions proceed by a S_N2 mechanism which is characterised by being kinetically of second order (first order in each of the reactants: the nucleophile and the nucleofuge atoms). It has been postulated that the observed second order kinetics is the result of passage through the well-known Walden inversion transition state where the nucleophile displaces the nucleofuge (leaving group) from the backside in a single concerted reaction step.

The reaction $H_a^{\bullet} + H_2 \longrightarrow H_2 + H_b^{\bullet}$ is the simplest radical abstraction reaction involving a free radical, $H_a^{\bullet}$ or $H_b^{\bullet}$, as a reactive intermediate (reaction A) whereas the $H_a^{-} + CH_4 \longrightarrow CH_4 + H_b^{-}$ is a typical S_N2 identity exchange reaction (reaction B), where H_a represents the incoming nucleophile and H_b^{-} stands for the leaving nucleofuge. The electronic structure calculations performed in this study were carried out with the Gaussian

03 suite of programs (Frisch et al., 2004). The reported TS geometrical parameters for the abstraction (Johnson, 1994), and the S_N2 exchange reactions (Shi & Boyd, 1989) were employed. Calculations for the structures of the internal reaction path were performed by use of the IRC method at the MP2/6-311++G** (UMP2//6-311G for the abstraction reaction) level of theory. As a result of the latter, 72/93 points (chemical structures) evenly distributed on the forward and reverse directions of the IRC paths were obtained for the A/B reactions, respectively. Finally, a higher level of theory (QCISD(T)) and a properly balanced basis set (6-311++G**) were chosen for both reactions to calculate the Shannon entropies of all chemical structures at the IRC paths. The molecular Shannon information entropies were obtained by employing software developed in our laboratory along with 3D numerical integration routines (Pérez-Jordá & San-Fabián, 1993; Pérez-Jordá et al., 1994), and the DGRID suite of programs (Kohout, 2007).

For reaction A (Fig. 2), both entropies possess richer structure at the vicinity of the TS as compared with the energy profile (which only shows one maximum at this point). By close inspection of Fig. 2, we note that the position entropy possesses a local maximum at the TS and two minima at its vicinity, whereas the momentum entropy decreases abruptly so as to reach a global minimum at the TS with two maxima at its vicinity.

The chemical picture proceeds in this way: as the intermediate radical ($H_a^{\bullet}$) approaches the molecule at the TS region, the molecular density exerts important changes so as to undergo the homolysis. This represents a physical situation where the density in position space gets localized in preparation for the bond rupture, which in turn results in a local increase of the kinetic energy. This provides explanation for well the known fact that bond breaking requires energy. Next, the bond is formed and as a consequence, the TS structure shows lower kinetic energy than the reactant/product complex ($H_a^{\bullet}\cdots H\text{-}H_b$ or $H_a\text{-}H\cdots H_b^{\bullet}$). Interestingly, from an information-theoretic perspective all of the above can be analogously described as observed in Fig. 2: as the radical intermediate approaches the TS region, the position entropies turn minimal when the position densities become localized and the corresponding momentum densities get delocalized (higher kinetic energies). Then at the TS, when the chemical structure relaxes, the position/momentum density gets delocalized/localized where the position/momentum entropy shows a local maximum/ minimum. The process occurs in two steps in the way the reaction dictates, and this also might be observed from the theoretic-information context by closely examining the entropies behaviour at the proximity of the TS. That is, the bond breaking process requires energy which should in turn be dissipated by relaxing the structure at the TS, and we note from Fig. 2 that this is indeed the case in that maxima for the momentum entropy are located before minima of the position entropy (depending on the direction of the process), i.e., reactive complexes gain the necessary energy for bond cleavage at BCER (bond cleavage energy regions and then get localized as we described above. Next, the homolysis provokes energy/density relaxation of the molecules toward the TS which is also observed from the Shannon entropies as explained before.

For reaction B (Fig. 6), again both Shannon entropies show richer structure as compared to the total energy profile (which only possesses one maximum at the TS). By examining Fig. 6 we note that the position entropy profile features a maximum at the TS along with two minima at its vicinity, whereas the profile of the momentum entropy shows a global minimum at the TS with two maxima at its vicinity. It seems likely that these particular regions correspond with BCER where bond breaking is supposed to occur. It is interesting to note that for both entropies the BCER are located at the same IRC coordinate, in contrast

with the two-stages mechanism of reaction A, and this may be indicative of the single step mechanism that characterizes the S_N2 process, which highlights the localized/delocalized combination of the position/momentum densities at this particular position of the IRP. At this point, it is interesting to associate the one step mechanism of this reaction to the chemical events that take place. While the nucleophile approaches the molecule the nucleofuge leaves it at unison, i.e., bond forming and bond breaking must occur in a concerted and synchronous manner. Both of these actions increase the energy of the combination: bond breaking requires energy (momentum density becomes delocalized and its corresponding entropy increases, and so its kinetic energy) as does overcoming the repulsion between the incoming ionic-complex (nucleophile) into close contact with the carbon's bonding shell (position density becomes localized and its corresponding entropy decreases). As the reaction process goes forward, the energy increases until a significant bonding begins to occur between the nucleophile and the molecule (increasing the position entropy and delocalizing its density). This releases enough energy to balance the energy required to break the carbon-nucleofuge bond (low kinetic energy structure with a highly localized momentum density). Then the transition state is reached. It is worth mentioning that the abrupt changes apparently observed for the Shannon entropies in both reactions (Figs. 2 and 6) are largely due to the significant changes exerted by the densities at the vicinity of the TS within the BCER (Esquivel et al., 2009).

4. Nanostructured materials

On the perspective of much larger molecules, Shannon Information Theory (IT) has been employed to analyze the growing behavior of nanostructures (Esquivel, 2009b). Shannon entropies in position and momentum spaces require costly and time-consuming computations as the size of the molecules increases in contrast with information entropies in Hilbert space, which are shown to be highly advantageous for analyzing large molecules. Thus, ab initio electronic structure calculations at the Hartree-Fock level of theory were performed to characterize the initial steps towards growing nanostructured molecules of Polyamidoamine (PAMAM) dendrimers, starting from the monomers, dimers, trimers, tetramers up to generations G0 (with 84 atoms), G1 (228 atoms), G2 (516 atoms), and G3 (1092 atoms). Shannon and Kullback entropies in Hilbert space were employed to provide theoretic-information evidence of the validity of the dense-core model of PAMAM precursors and dendrimers G0 through G3. Furthermore, marginal H-type von Neumann informational entropies (Flores-Gallegos & Esquivel, 2008) have been employed (Esquivel, 2010b) to provide evidence of the validity of the dense-core model of dendrimers.

4.1 Information-theoretical analysis of selected PAMAM dendrimers

Since their introduction in 1985 by Tomalia (Tomalia et al., 1985) dendrimers have attracted much attention because of their fascinating structure and unique properties (Newkome et al., 1996; Fréchet & Tomalia, 2001) Dendrimers are globular, size monodisperse macromolecules in which all bonds emerge radially from a central focal point or core with a regular branching pattern and with units that repeat and contribute to a branch point. They are defined by three components: a central core, an internal dendritic structure (the branches), and an exterrnal surface with functional surface groups. Not all regularly branched molecules are dendrimers because properties of the dendritic state (Fréchet &

Tomalia, 2001), such as core encapsulation (Hecht & Fréchet, 2001). Several applications for dendrimers have been proposed in the literature (Tomalia et al., 1990), with potential applications in biology as mimetic systems of enzymes or redox proteins (Liu & Breslow, 2003), in medicine for drug delivery, gene therapy, and biochemical sensors (Zeng & Zimmerman, 1997)., in optoelectronics for transduction of signals or light-harvesting devices (Vögtle, 2000) and in nanoscience as building units in self-assembled systems or functionalized with groups for molecular recognition and signaling (Fréchet, 2002). Among these, Poly(amido)amine or PAMAM dendrimers are among the most studied families of dendrimers. These organic dendrimers contain tertiary amines as branching points, i.e., the respective branching multiplicity is 2. The core multiplicity varies; the original PAMAM dendrimers were synthesized from an amine core and thus had a 3-fold multiplicity (Newkome et al., 1996; Liu & Breslow, 2003) whereas recently, ethylenediamine-core PAMAMs became more common and these have a 4-fold multiplicity (Jockush et al., 1999) where the typical end-groups are primary amines.

A number of review articles on dendrimers have been reported (Ballauff, 2001; Newkome, 2001). Most of the work was motivated by a dendrimer-model of a hollow core and hence by a dense shell. This model became popular since the work of de Gennes and Hervet (1983) in which they presented the first theoretical treatment of dendrimers where they obtained a density profile with a global minimum at the center and a monotonic density increasing towards the periphery of the dendrimer. On the other hand, Lescanec and Muthukumar (1990) used a kinetic growth algorithm to find, in contrast with precvious studies (de Gennes & Hervet, 1983), density profiles that decreased monotonically towards the edge of the molecule. They were the first to show that a dendritic structure made up from flexible bonds should exhibit its maximum density at the center of the molecule becoming in the so-called dense-core model.

A great deal of progress in the understanding of the conformations of dendrimers has been achieved through computer simulations at various levels of description, from the microscopic (atomistic) to the oversimplified ones with different methodologies: Monte-Carlo (MC), molecular dynamics (MD), and Brownian dynamics (BD) (Allen & Tildesley, 1987). Recently, Maitli et al (2004), performed a systematic series of fully atomistic MC-MD simulations on PAMAM-ethylenediamine (PAMAM-EDA) cored dendrimers from G0 through G11, to characterize the structure and properties of these molecules. On the electronic structure (ES) side, only a few Hartree-Fock (HF) and density functional theory (DFT) studies have addressed to study some aspects of low order generation dendrimers (Tarazona-Vasquez & Balbuena, 2004a; 2004b). However, ab initio studies of the ES type addressing the structural properties of PAMAM dendrimers of higher generations are very scarce. The issue is not a simple one since these molecules possess an enormous number of energetically permissible conformations and a large number of atoms (EDA-PAMAM dendrimers grow from 84 atoms for G0 up to 294852 atoms for G11) which are beyond the present capabilities for the ES packages and present computers.

In recent imvestigations (Esquivel et al., 2009b; 2010b) we have shown that despite the limitations of quantum chemistry methods, it is possible to apply information.theory and chemical concepts to elucidate some of the structural features of dendrimers. The focus was on supporting the validity of the core-dense model for dendrimers from a theoretic information point of view. Besides, the growing behavior of PAMAM precursors and some low generations of its dendrimers was revealed by employing selected properties of soft physics matter along with chemical reactivity parameters of dendrimers.

It is known that flexible-chain dendrimers, although being chemically regular structures, do not assume regular shapes. To quantitatively evaluate the deviation from spherical symmetry, we calculated the shape of the precursors and dendrimers through the principal moments of inertia I_x, I_y and I_z which are calculated through the eigenvalues of the shape tensor **G** describing the mass distribution:

$$G_{pq} = (1/M)[\sum_{i}^{N} m_i(\mathbf{r}_{pi} - \mathbf{R}_p)(\mathbf{r}_{qi} - \mathbf{R}_q) \tag{44}$$

where $p,q = x,y,z$, and $\boldsymbol{r}_{pi}$ is the position of the ith atom relative to the R_p components of the center of mass of the molecule, M is the mass of the molecule and m_i is the mass of the atom. The sum of three eigenvalues (I_x, I_y and I_z) is an invariant of the shape tensor **G**, giving $\left\langle R_g^2 \right\rangle$, which is the mean-square radius of gyration that provides a quantitative characterization of the dendrimer size. The ratio of these three principal moments is a measure of *eccentricity* (minor-major axes ratio) of the shape ellipsoid of the dendrimers and hence the shape of the dendrimer can be assessed from the values of the ratio of the three principal moments of inertia of the molecules I_z/I_y and I_z/I_x. Rudnick and Gaspari (1986) introduced a better definition of *asphericity* frequently used in the literature as

$$\delta = 1 - 3\frac{\left\langle I_2 \right\rangle}{\left\langle I_1^2 \right\rangle} \tag{45}$$

where I_i are the respective invariants of the gyration tensor and are given by

$$I_1 = I_x + I_y + I_z \tag{46}$$

and

$$I_2 = I_x I_y + I_x I_z + I_y I_z \tag{47}$$

Technical details concerning the electronic structure calculations for the present study have been amply discussed (Esquivel et al., 2009b; 2010b). So that, the values of the three principal moments of inertia (Eqs. 46 and 47) were tabulated (see Esquivel, et al, 2010) along with the radius of gyration and the asphericity factor. Figure 10 shows the radius of gyration (Eq. 44), for the different G0-precursors and the G0 dendrimer. From the Figure (and Table 2 in Ref. 96) we may assess the size of the dendrimers through the radius of gyration. As expected, R_g values show a constant increasing trend in going from the monomers up to the G0-dendrimer.

On the side of the chemical reactivity of the precursors and dendrimers we have also evaluated some parameters that might be useful to analyze the chemical properties. In the context of conceptual DFT we have defined several properties in subsection 2.1 (Eqs. (40) to (43)), in particular hardness (η) and softness (S), which are good descriptors of chemical reactivity, the former measures the global stability of the molecule (larger values of η means less reactive molecules), whereas the S index quantifies the polarizability of the molecule (Ghanty & Ghosh, 1993; Roy et al., 1994), thus soft molecules are more polarizable and possess predisposition to acquire additional electronic charge (Chattaraj et al., 2006). It has been noted in Ref. Esquivel et al, 2010 that hardness values show a decreasing tendency as

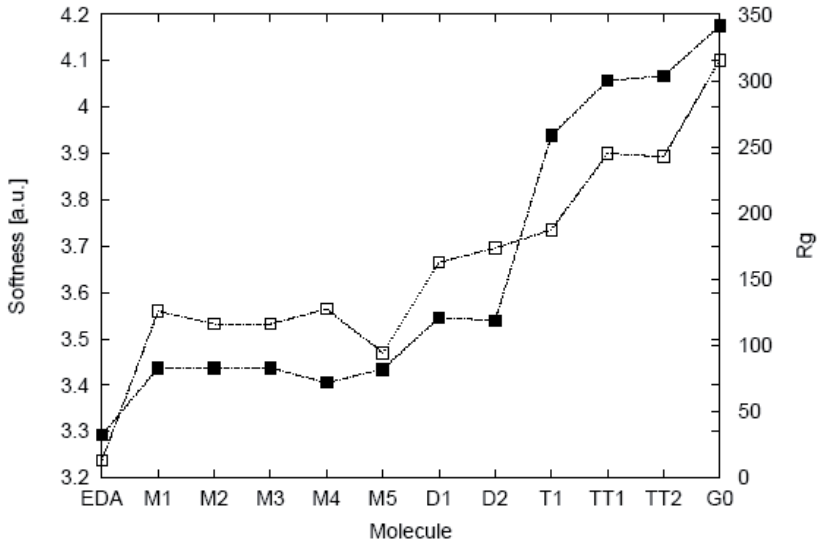


Fig. 10. The softness "S" (empty boxes) and the radius of gyration R_g (solid boxes) values for the G0-precursors and G0-PAMAM generation.

the molecules are sizely bigger, which means that as the size of the precursor increases toward G0 the polarizability of the molecules increases, i.e., their corresponding densities tend to be more delocalized. Further, it is known that less compact molecules are more polarizable, with low hardness values and hence more reactive, and this is indeed the case when comparing the softness values with the size of the molecules through the radius of gyration R_g in Figure 10.

4.2 Information-theoretical analysis of selected PAMAM dendrimers

Next, we examine the Shannon information entropies in real space, Eqs (37) and (38). In Figure 11 the entropy sum (S_r+S_p) and the energy are depicted for the PAMAM G0-precursors and dendrimer G0. It is apparent from Fig. 11 that the total entropy follows the opposite behavior as the energy, i.e., as the size of the molecules increases so does the entropy sum. It is also interesting to note that the total entropy distinguishes the different polymeric structures in that isoelectronic systems possess the same entropy value and so does the energy.

4.3 Information-theoretical entropies in hilbert space

We have recently shown (Carrera, E. M. et al., 2010) that there is an information-theoretic justification for performing Löwdin symmetric transformations (Löwdin, 1970; Reed & Weinhold, 1983; Davidson, 1967). on the atomic Hilbert space, to produce orthonormal atomic orbitals of maximal occupancy for the given wavefunction, which are derived in turn from atomic angular symmetry subblocks of the density matrix, localized on a particular atom and transforming to the angular symmetry of the atoms. The advantages of these kind of atoms-in-molecules (AIM) approaches (Reed et al., 1985; Bruhn et al., 2006) are that the resulting natural atomic orbitals are *N*-representables, positively bounded, and rotationally invariant. We have recently shown (Flores-Gallegos & Esquivel, 2008) that the corresponding

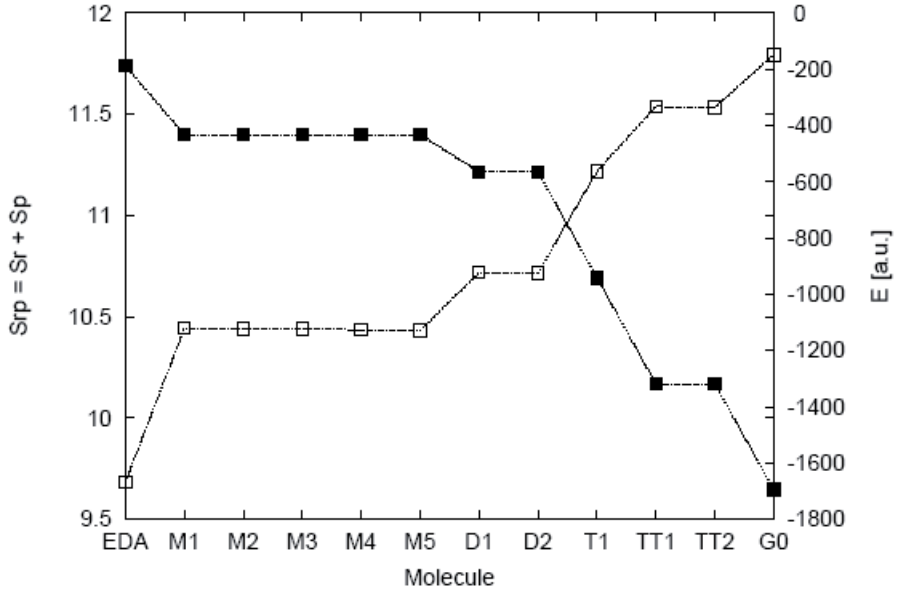


Fig. 11. Shannon entropy sum form Eqs. 1 and 2 (empty boxes) and the Energies in a.u. (solid boxes) for the PAMAM polymeric precursors at the HF/3-21G* level.

"natural atomic probabilities" (NAP) (Carrera, E. M. et al., 2010) are useful to define von Neumann information entropies in Hilbert space which might be able to measure entanglement in the context of Quantum Information Theory (Wehrl, 1978; Vedral, 2002).
The uncertainty of a probability distribution p_i(A) is measured through the Shannon entropy (Shannon, 1948) in Hilbert space Eq. (1). The relative entropy between two probability distributions p_i(A) and p_i(B) is defined through the Kullback-Liebler entropy (Kullback & Leibler, 1951) in Eq. (2) where the p_i(A) (and p_i(B)) in Eqs. (1) and (2) can be determined by use of natural atomic probabilities (Carrera et al., 2010; Flores-Gallegos & Esquivel, 2008). While Eq. (13) represents a distance from a reference probability which is not symmetrized.
Finally, we investigated the information entropies in Hilbert space in connection with the growing behavior of PAMAM dendrimers. It is important to mention that whereas Shannon entropies in real space represent costly and time consuming calculations, Hilbert space entropies do not pose additional computational efforts as to the theoretic-information analyses concerns, of course the obtaining of the wave functions and the NAP values represent a challenge for electronic structure calculations and the present computation capabilities available to us. Thus, we have performed the necessary calculations for analyzing additional structures, the trimer conformations T2 trhough T4, along with G1 through G3 dendrimers (G1 with 228 atoms, G2 with 516 atoms, and G3 with 1092 atoms). In Figure 12 we have depicted the Shannon entropy *H(A)* for all the PAMAM G0-precursors along with the G0 through G3 dendrimers. This measure allows to determine the global information content of the systems and consequently we may observe from Fig. 12 that *H(A)* shows an increasing behavior as the size of the precursors and dendrimers increases. This again supports the above discussed dense-core model of dendrimers as bigger molecules show more uncertainty in Hilbert space, which corresponds to less compact densities in real space and hence to more delocalized electronic distributions.

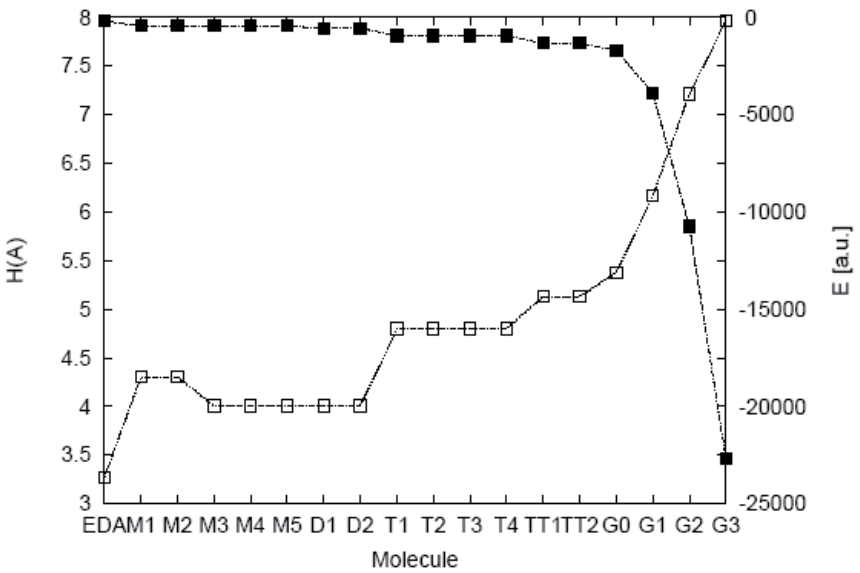


Fig. 12. Shannon entropy in Hilbert space, *H(A)* from Eq. (1) (empty boxes), and the total Energy values in a.u. (solid boxes) for the G0-precursors and generations G0 through G3.

Furthermore, in Figures 13 and 14 we have plotted the *H(A)* values along with the Kullback-Leibler relative entropy in Hilbert space, Eqs. (1) and (2), respectively. The $H(A \mid B)$ measure represents a distance measure from a reference probability distribution, which in this case we have set for the EDA molecule (which is embedded in all the structures). We may observe from the Fig. 14 that as the molecules depart from EDA the Hilbert information distance increases in a monotonically fashion which reflects the core-dense growing behavior

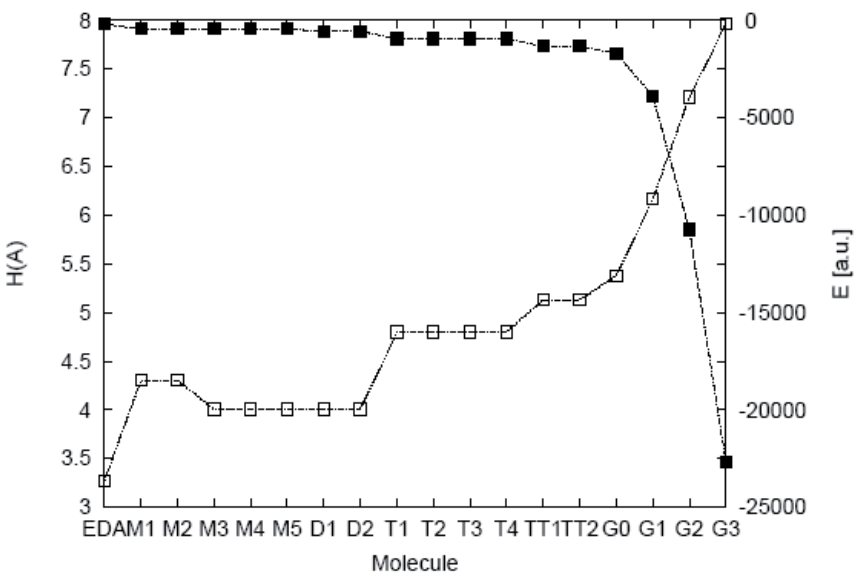


Fig. 13. Shannon entropy in Hilbert space, *H(A)* from Eq. (1) (empty boxes), and the total Energy values in a.u. (solid boxes) for the G0-precursors and generations G0 through G3.

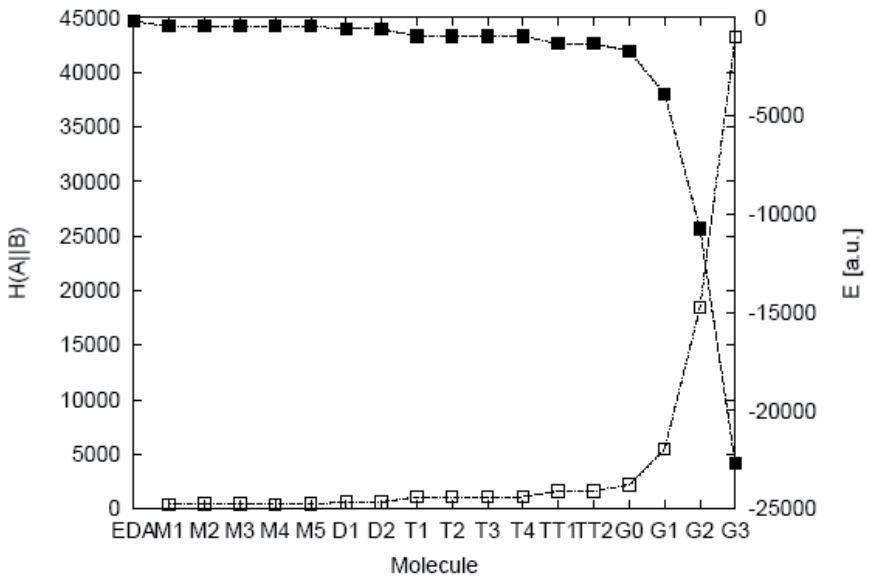


Fig. 14. Kullback-Leibler entropy in Hilbert space, *H(A | B)* from Eq. (2) (empty boxes), and the total Energy values in a.u. (solid boxes) for the G0-precursors and generations G0 through G3.

of dendrimers by simply measuring the information distance between EDA and the increasingly bigger molecules. Assuming that dendrimers follow a hollow-core model of growth and hence a dense shell model, the *H(A | B)* trend would have been just the opposite.

Figures 13 and 14 show the ability of these quantum measures to reveal and corroborate two simple facts already discussed in the literature (Löwdin, 1970): (i) the validity of the core-dense model witnessed by the mutual von Neumann entropy of the marginal type, as the *H(A | B)* Hilbert distance increases monotonically in going from the precursors to higher generation dendrimers, and (ii) the global information content of the systems, *H(A),* which increases monotonically as one would expect from a thermodynamic point of view (entropy). A particular feature that may be noted from Figure 13 is the capacity of *H(A)* of measuring the atomic and electronic content of the systems, regardless of its conformational structure, which is characteristic of the energy. The relevancy of these results might be assessed by considering that the equivalent information in real space implied the calculation of position and momentum space entropies of the Shannon type which, taking into account the size of the systems (1023 atoms for G3 dendrimer), represents indeed a formidable task to compute, even by taking into consideration that the integration quadratures were guided by promolecular grids (Esquivel et al., 2009b). Ongoing research is being undertaken in our laboratories to extend the study to higher generation PAMAM dendrimers mainly on the side of the Hilbert space framework.

5. Concluding remarks

We have shown throughout this Chapter that phenomenological description of chemical phenomena is readily acesible through Information Theory concepts of the Shannon type.

Moreover, it is worth mentioning that other recently published results have extended these ideas to more intrincated processes involving a three-center insertion reaction (Esquivel et al., 2010c), and in a more comprehensive manner in elementary chemical reactions and conformeric analysis through the local information measure of Fisher (López Rosa et al, 2010; Esquivel et al., 2011) and also by use of statistical complexity measures and planes (Esquivel et al., 2010d; 2011b). Furthermore, we have extended the scope of the analyses to quantum information concepts to measure the entanglement for the dissociation of diatomic molecules (Esquivel et al., 2011c).

Throughout these studies in several areas of chemistry and also in nanostructured systems, we believe that information science may conform a new scientific language to explain essential aspects of chemical phenomena (and presumably biological too). These new aspects are not accessible through any other standard methodology in quantum chemistry, allowing to reveal intrincated mechanisms in which chemical phenomena occur. This envisions a new area of research that looks very promising as a standalone and robust science. The purpose of our research is to provide fertile soil to build this nascent scientific area of chemical inquiry through information-theoretical concepts which we have named Quantum Information Chemistry.

The following aspects are to be distinguished from the studies:

i. We have assessed the utility for the theoretic-information measures of the Shannon type to characterize elementary chemical reactions. Through these chemical probes we were capable to observe the basic chemical phenomena of the bond breaking/forming showing that the Shannon measures are highly sensitive in detecting these chemical events not revealed by the energy profile. Furthermore, the transition state of a reaction is commonly identified by the presence of a negative force constant for one normal vibrational mode corresponding with an imaginary frequency. However the work of Zewail and Polanyi in transition state spectroscopy has led to the concept of a reaction having a continuum of transient, a transition region rather than a single transition state (Zewail, 1988; 1990; 2000a; 2000b; Polanyi & Zewail, 1995). It is worth mentioning that the results of the present study show indeed the existence of such a region between the BCER, before and after the TS. This is in agreement with reaction force, $F(R)$, studies (Toro-Labbé, et al., 2009; Toro-Labbé et al., 2007; Murray et al., 2009; Jaque et al., 2009) where the reaction force constant, $\kappa(R)$, also reflects this continuum, showing it to be bounded by the minimum and the maximum of $F(R)$, at which $\kappa(R) = 0$.
ii. We have also shown (Esquivel et al., 2010) that theoretic-information measures of the Shannon type posses the capability of revealing the hidden structure of the chemical reactions through phenomenological concepts (Esquivel et al., 2009) which permit to unveil the asynchronous/synchronous mechanistic behavior which characterize reactions A and B, respectively. Furthermore, it is worth mentioning that the chemical phenomena treated here are largely invisible for most of the standard density descriptors, and certainly not accessible from the energy profile. We believe that this kind of studies may serve to provide with more specific information as to nourish chemical reactivity theories which pursue a full conceptual prediction of the TS structure from basic chemical principles (Shaik et al., 1992). On the other hand, it is our contention that these observations might be amenable of experimental verification through photodetachment techniques in the femtosecond time scale (Zewail, 1988, 1990, 2000a, 2000b).

iii. Throughout this investigation we have employed selected density descriptors which show numerical evidence which supports the dense-core model of dendrimers Besides, from a information-theoretical perspective, it was shown that Shannon entropies defined in real space as well as in Hilbert space are capable of revealing the dense core growing behavior of dendrimers, by showing that bigger molecules possess more delocalized electronic distributions in such a way to span their molecular distributions as the molecular size increases.

6. Acknowledgments

We wish to thank José María Pérez-Jordá and Miroslav Kohout for kindly providing with their numerical codes. We acknowledge financial support through Mexican grants from CONACyT, PIFI, PROMEP-SEP and Spanish grants MICINN projects FIS2011-24540, FQM-4643 and P06-FQM-2445 of Junta de Andalucía. J.A., J.C.A., R.O.E. belong to the Andalusian researchs groups FQM-020 and J.S.D. to FQM-0207. R.O.E. wishes to acknowledge financial support from the Ministerio de Educación of Spain through grant SAB2009-0120 and to Prof. Marcelo Galván for his valuable support. Allocation of supercomputing time from Laboratorio de Supercómputo y Visualización at UAM, Sección de Supercomputacion at CSIRC Universidad de Granada, and Departamento de Supercómputo at DGSCA-UNAM is gratefully acknowledged.

7. References

Allen, M. P. & Tildesley, D. J. (1987). Computer Simulation of Liquids, Clarendon, Oxford, 1987;

Angulo, J. C. & Dehesa, J. S. (1992). J. Chem. Phys. 1992, 97, 6485-6495

Angulo, J. C. (1994). Phys. Rev. A 1994, 50, 311-313

Antolín, J.; Zarzo, A. & Angulo, J. C. (1993). Phys. Rev. A 1993, 48, 4149-4155.

Araki, H. & E. H. Lieb (1970). Commun. Math. Phys. 1970, 18, 160.

Ayers, W. (2006). Theor. Chem. Acc. 2006, 115, 253-256

Bader, R.F.W. & MacDougall P.J. (1985). J. Am. Chem. Soc. 1985, 107, 6788-6795

Balakrishnan N. & Sathyamurthy, N. (1989). Chem. Phys. Lett. 1989, 164, 267-269

Ballauff, M. (2001). Dendrimers III: Design, Dimension, Function:, Top. Curr. Chem. 212, (2001) 177–194

Baskin, C.P.; Bender, C.F.; Bauschlicher, C.W. Jr.; & Schaefer III, H.F. J. Am. Chem. Soc. 1974, 96, 2709-2713

Bell, J (1987).*Speakable and Unspeakable in Quantum Mechanics* 1987 (Cambridge University, Cambridge).

Bernasconi, C.F. (1992). Acc. Chem. Res. 25 (1992) 9.

Bialynicky-Birula, I. & Mycielski, J. (1975). J. Commun. Math. Phys. 1975, 44, 129-132.

Borgoo, A.; Jaque, P. ; Toro-Labbé, A.; Van Alsenoy, C. & Geerlings, P. (2009). Phys. Chem. Chem. Phys., 2009, 11, 76–482

Bradforth, S.E.; Arnold, D.W.; Newmark, D.M. & Manolopoulos, D.E. (1993). J Chem Phys 1993, 99, 6345-6359.

Bruhn, G. ; Davidson, E. R.; Mayer, I. & Clark, A. E. (2006). Löwdin population analysis with and without rotational invariance, Int. J. Quantum Chem. 106 (2006) 2065-2072.

Carrera, E. M. ; Flores-Gallegos, N. & Esquivel , R.O. (2010). Natural Atomic Probabilities in Quantum Information Theory , J Comp. Appl. Math. 233, 1483-1490 (2010).

Chandra, A.K. & Sreedhara-Rao, V. (1996). Int. J. Quantum Chem. 58 (1996) 57.
Chandra, A.K. (1999). Proc. Indian Acad. Sci. (Chem. Sci.) III (1999) 589.
Chattaraj, P.K.; Sarkar, U. & Roy, D. R. (2006). Chem Rev 2006, 106, 2065-2091.
Chatzisavvas, K. Ch.; Moustakidis, Ch. C. & Panos C. P. (2005) J. Chem. Phys. 2005, 123, 174111-174300.
Cerf N. J. & Adami C. (1977). Phys Rev Lett 1997,79, 5194
Coulson, C.A. (1991). "Valence", 2nd. ed., Clarendon, Oxford, 1961
Davidson, E.R. (1967). Electronic Population Analysis of Molecular Wavefunctions, J. Chem. Phys. 46, (1967) 3320-3324
de Gennes, P. G. & Hervet, H. (1983). Statistics of starburst polymers, J. Phys. Lett. 44, (1983) L351- L360.
De Proft, F.; Chattaraj, P. K.; Ayers, P. W.; Torrent-Sucarrat, M. ; Elango, M.; Subramanian, V.; Giri, S. & Geerlings, P. (2008). J. Chem. Theory Comput. 2008, 4, 595–602.
Dewar, M.J.S. (1984). J. Am. Chem. Soc. 106 (1984) 209.
Einstein, A.;. Podolsky B; & Rosen N. (1935). Phys. Rev. 1935, 47, 777.
Esquivel, R O.; Flores-Gallegos, N.; Iuga, C.; Carrera, E. ; Angulo, J. C. & Antolín, J. (2009). Theoretical Chemistry Accounts 124, 445–460 (2009)
Esquivel, R. O. ; Flores-Gallegos, N.; Carrera, E.; Dehesa, J.S. ; Angulo, J. C.; Antolín, J. & Soriano-Correa, C. (2009b). Molecular Simulation, Vol. 35, No. 6, May 2009, 498–511.
Esquivel, R O.; Flores-Gallegos, N.; Iuga, C.; Carrera, E. ; Angulo, J. C. & Antolín, J. (2010). Physics Letters A 374 (2010) 948–951
Esquivel, R. O.; Flores-Gallegos, N.; Carrera, E. & Soriano-Correa, C. (2010b). Journal of Nano Research Vol. 9 (2010) pp 1-15.
Esquivel, R O. ; Flores-Gallegos, N.; Dehesa, J.S.: Angulo, J. C. ; Antolín, J. & Sen, K. (2010c) J. Phys. Chem. A 2010, 114, 1906–1916
Esquivel, R.O.; Angulo, J.C.; Antolin, J.; Dehesa, J. S.; Flores-Gallegos, N. & S. Lopez-Rosa (2010d). Phys. Chem. Chem. Phys., 2010, 12, 7108–7116
Esquivel, R.O. ; Liu, S.; Angulo, J.C. ; Dehesa, J. S. ; Antolin, J. & Molina-Espíritu, M. (2011). J. Phys. Chem. A, 115 (2011) 4406-4415.
Esquivel, R. O. ; Molina-Espíritu, M.; Angulo, J. C. ; Antolín, J. ; Flores-Gallegos, N. & Dehesa, J. S. (2011b). Mol. Phys., 109 (2011), 2353–2365.
Esquivel, R.O.; Flores-Gallegos, N.; Molina-Espíritu, M.; Plastino, A.R.; Angulo, J.C.; Antolín, J. & Dehesa, J.S. (2011c). J. Phys. B: At. Mol. Opt. Phys. 44 175101 (2011)
Eyring, H. (1935). J. Chem. Phys. 1935, 3, 107-115
Fan, L & Ziegler, T. (1992). J. Am. Chem. Soc. 1992, 114, 10890-10897
Fischer, M. & Vögtle, F. (1999). Dendrimers : From Design to Application - A Progress Report, Angew. Chem., Int. Ed., 38, (1999) 884-905
Flores-Gallegos, N. & Esquivel, R. O. (2008). J. Mex. Chem. Soc., Volume 52, Issue 1, 19-30 (2008)
Fréchet, J M J & Tomalia, D A. (2001). Editors, Dendrimers and Other Dendritic Polymers, Chichester, U.K.: Wiley; 2001.
Fréchet, J. M. (2002). Dendrimers and supramolecular chemistry, J. Proc. Natl. U.S.A. 99, (2002) 4782-4787.
Frenkel, D.; Smit, B. , Understanding Molecular Simulation, Academic Press, San Diego, 1996.
Fukui, K. (1981). Acc. Chem. Res. 1981, 14, 363-368.
Frisch MJ; Trucks GW; Schlegel HB; Scuseria GE; Robb MA; Cheeseman JR; Montgomery Jr. JA; Vreven T; Kudin KN; Burant JC; Millam JM; Iyengar SS; Tomasi J; Barone V; Mennucci B; Cossi M; Scalmani G; Rega N; Petersson GA; Nakatsuji H; Hada M; Ehara M; Toyota K; Fukuda R; Hasegawa J; Ishida, M; Nakajima T; Honda Y; Kitao

O; Nakai H; Klene M; Li X; Knox JE; Hratchian HP; Cross JB; Bakken V; Adamo C; Jaramillo J; Gomperts R; Stratmann RE; Yazyev O; Austin AJ; Cammi R; Pomelli C; Ochterski JW; Ayala PY; Morokuma K; Voth GA; Salvador P; Dannenberg JJ; Zakrzewski VG; Dapprich S; Daniels AD; Strain MC; Farkas O; Malick DK; Rabuck AD; Raghavachari K; Foresman JB; Ortiz JV; Cui Q; Baboul AG; Clifford S; Cioslowski J; Stefanov BB; Liu G; Liashenko A; Piskorz P; Komaromi I; Martin RL; Fox DJ; Keith T; Al-Laham MA; Peng CY; Nanayakkara A; Challacombe M; Gill PMW; Johnson B; Chen W; Wong MW; Gonzalez C; Pople JA (2004). Gaussian 03, Revision D.01, Gaussian, Inc., Wallingford CT)

Gadre, S.R. (2003). Reviews of Modern Quantum Chemistry: A Celebration of the Contributions of Robert G. Parr, Vol. 1, p. 108-147, Ed. K. D. Sen, World Scientific, Singapore (2003)

Ghanty, T. K. & Ghosh, S. K. (1993). J. Phys. Chem. 1993, 97, 4951-4953

Ghosh, S. K.; M. Berkowitz & Parr, R. G. (1984). Proc. Natl. Acad. Sci. U.S.A. 1984, 81, 8028-8031

Glukhovtsev, M.N.; Pross, A. & Radom, L. (1995). J. Am. Chem. Soc. 1995, 117, 2024-2032

González, C. & Schlegel, H. B. (1990). J. Phys. Chem. 1990, 94, 5523-5527

González, C. & Schlegel, H. B. (1989). J. Phys. Chem. 1989, 90, 2154-2161

González-García, N.; Pu, J.; González-Lafont, A.; Lluch, J.M. & Truhlar, D.G. (2006). J. Chem. Theory Comput. 2006, 2, 895-904

Guevara, N. L.; Sagar, R. P. & Esquivel, R. O. (2005) J. Chem. Phys. 2005, 122, 084101

Hecht, S. & Fréchet, J M. (2001). Dendritic Encapsulation of Function: Applying Nature's Site Isolation Principle from Biomimetics to Materials Science, J. Angew Chem Int Ed Engl. 2001;40:74–91

Holister, P.; Roman-Vas, C. & Harper, T. (2003). Dendrimers, Technology White Papers", Vol 6, (2003), 1-15.

Ho, M.; Schmider, Weaver, D. F.; Smith, Jr. V. H.; Sagar, R. P. & Esquivel, R.O. (2000) Int. J Quant. Chem. 2000, 77, 376-382

Hammond, G. S. (1955). J. Am. Chem. Soc. 1955, 77, 334-338

Hati, S. & Datta, D. (1994) J. Phys. Chem. 1994, 98, 10451-10454.

Hirshfeld F. L. (1977). Theor. Chim. Acta 44, 129 (1977).

Hoffman, R.; Shaik, S. & Hiberty, P.C. (2003). Acc. Chem. Res. 2003, 36, 750-756

Ishida, K.; Morokuma, K. & Komornicki, A. (1977). J. Chem. Phys 1977, 66, 2153-2156

Janak, J. F. (1978)Phys. Rev B, 1978, 18, 7165-7168.

Jaque P.; Toro-Labbé A.; Geerlings P. & De Proft F. (2009). J. Phys. Chem. A 2009, 113, 332-344.

Jockush, S.; Ramirez, J.; Sanghvi, K. ; Nociti, R.; Turro, N. J. & Tomalia, D. A. (1999). Comparison of Nitrogen Core and Ethylenediamine Core Starburst Dendrimers through Photochemical and Spectroscopic Probes, Macromolecules 32, (1999) 4419-4423.

Johnson, B.A.; Gonzales, C.A.; Gill, P.M.W. & Pople, J.A. (1994). Chem. Phys. Lett 1994, 221, l00-108

Karafiloglou, P. & Panos, C.P. (2004) Chem. Phys. Lett. 2004, 389, 400-404

Knoerr, E.H. & Eberhart M.E. (2001). J. Phys. Chem. A 2001, 105, 880-884

Koga, T. & Morita, M. (1983). J. Chem. Phys. 1983, 79, 1933-1938

Kohout M. (2007). Program DGRID, version 4.2. 2007

Koopmans, T. A. (1933). Physica 1933, 1, 104-113

Kullback, S. & Leibler R. A. (1951). On Information and Sufficiency, Ann. Math. Stat. 1951, 22, 79-86.

Leffler, J. E. (1953). Science 1953, 117, 340-341.

Lescanec, R. L. & Muthukumar, M. (1990). Configurational characteristics and scaling behavior of starburst molecules: a computational study, Macromolecules 23, (1990) 2280 – 2288.

Liu, S. (2007). J. Chem. Phys. 2007, 126, 191107-(1-3).

Liu, L. & Breslow, R. (2003). J. Am. Chem. Soc. 125, (2003)12110.
López-Rosa, S.; Esquivel, R. O. ; Angulo, J. C. ; Antolin, J.; Dehesa, J. S. & Flores-Gallegos, N. (2010). J. Chem. Theory Comput. 2010, 6, 145–154
Löwdin, P.O. (1970). On the orthogonality problem, Adv. Quantum Chem. 5, (1970) 185
Maiti, P.K. ; Cagin, T ; Wang, G. & Goddard III, W.A. (2004). Structure of PAMAM Dendrimers: Generations 1 through 11, Macromolecules 37, (2004) 6236-6254
Martyusheva, L.M. & Seleznev, V.D. (2006) Phys. Rep. 2006, 426, 1 – 45
Massen, S. E. & Panos, C. P. (1998). Phys. Lett. A 1998, 246, 530-533
Murray J.S.; Toro-Labbé A.; Clark T. & Politzer P. (2009). J. Mol. Model. 2009, 15, 701-706
Nagy, A. (2003) J. Chem. Phys. 2003, 119, 9401-9405
Nagy, A. (2006). Chem. Phys. Lett. 2006, 425, 154-156
Nalewajski, R. F. & Parr, R. G. (2001) J. Phys. Chem. A 2001, 105, 7391-7400.
Nalewajski, R.F. (2003). Chem. Phys. Lett. 2003, 372, 28
Newkome, G. R. ; Moorefield, C. N. & Vögtle, F. (1996). Dendritic molecules: Concepts, Synthesis, Perspectives, Wiley-VCH, Weinheim, 1996
Newkome, G. R.; Moorefield, C. & Vogtle, F. (2001). Dendrimers and Dendrons, Wiley-VCH, New York, 2001
Parr, R.G.; Ayers, P.W. & Nalewajski , R. F. (2005). J. Phys. Chem. A 2005, 109, 3957-3959
Parr, R. G. & Pearson, R. G. (1983) J Am Chem Soc 105, (1983) 7512-7516
Parr, R. G. & Yang, W. (1989) "Density-Functional Theory of Atoms and Molecules", Oxford University Press: New York, 1989
Pearson, R. G. (1963). Hard and Soft Acids and Bases. J. Am. Chem. Soc. 1963, 85, 3533
Pearson, R. G. (1973) Hard and Soft Acids and Bases; Dowen, Hutchinson and Ross: Stroudsberg, 1973
Pearson, R. G. (1997). Chemical Hardness; Wiley-VCH; New York, 1997.
Peng, Ch.; Ayala, Ph.Y.; Schlegel, H. B. & Frisch, M.J. (1996). J. Comp. Chem. 1996, 17, 49-56
Pérez-Jordá, J. M. & San-Fabián, E. (1993). Comput Phys Commun 1993, 77, 46–56
Pérez-Jordá, J. M.; Becke, A. D. & San-Fabián, E. J. Chem. Phys. 1994, 100, 6520–6534.
Polanyi J.C. & Zewail A.H. (1995). Acc. Chem. Res. 1995, 28, 119-132
Politzer, P. & Truhlar, D. G. (1981). "Chemical Applications of Atomic and Molecular Electrostatic Potentials"; Academic Press: New York, 1981
Pople, J.; Krishnan, A. R.; Schlegel, H. B. & Binkley, J. S. (1978). Int. J. Quantum Chem. 1978, 14, 545-560
Rajagopal A.K. & Rendell R.W. (2002). Phys. Rev. A 2002, 66, 022104
Raju K.B.K.; Nair P.& V. and Sen K.D. (1990). Chem. Phys. Lett. 1990, 170, 89.
Ramirez, J.C.; Perez, J.M.H.; Sagar, R.P.; Esquivel, R.O.; Ho, M. & Smith Jr., V. H. (1998). Phys. Rev. A 1998, 58, 3507-3515
Reed, A.E. & Weinhold, F. (1983). Natural bond orbital analysis of near-Hartree–Fock water dimmer, J. Chem. Phys. 78, (1983) 4066-4073
Reed, A.E. ; Weinstock, R.B & Weinhold, F. (1985) Natural population analysis J. Chem. Phys. 83, (1985) 735-746
Romera E. & Dehesa, J. S. (2004) J. Chem. Phys. 2004, 120, 8906-8912
Roy, R.; Chandra, A.K. & Pal, S. (1994). J. Phys. Chem., 1994, 98, 10447-10450
Rudnick, G. & Gaspari, G. (1986). The aspherity of random walks, J. Phys. A 4, (1986) L191-194.
Safi, B.; Choho, K. & Geerlings, P. (2001). J. Phys. Chem. A 2001, 105, 591-601
Schaftenaar G. & Noordik J.H. (2000). "MOLDEN: a pre- and post-processing program for molecular and electronic structures" J. Comput.-Aided Mol. Design 2000, 14,123-134.
Schlegel, H. B. (1987). Adv. Chem. Phys. 1987, 67, 249-286.
Schmidt, M. W.; Gordon, M.S. & Dupuis, M. (1985). J. Am. Chem. Soc. 1985, 107, 2585-2589

Sen, K. D. (2005) J. Chem. Phys. 2005, 123, 074110(1-9)
Sen K. D. & Katriel, J. (2006) J. Chem. Phys. 2006, 125, 074117-(1-4)
Shaik, S.; Ioffe, A.; Reddy, A.C. & Pross, A. (1994). J. Am. Chem. Soc. 1994, 116, 262-213
Shaik, S.S.; Schlegel, H.B. & Wolfe, S. (1992). Theoretical Aspects of Physical Organic Chemistry: The SN2 reaction, Wiley, New York, 1992.
Shannon, C.E. (1948). Bell Syst. Tech. J. 1948, 27, 379-423.
Shannon, C. E.& W. Weaver (1949), *The Mathematical Theory of Communication* (University of Illinois, Urbana, IL), (1949).
Shi, Z. & Boyd, R.J. (1991). J. Am. Chem. Soc. 1991, 113, 1072-1076.
Shi, Z. & Boyd, R.J. (1989). J. Am. Chem. Soc. 1989, 111, 1575-1579
Shi, Q. & Kais, S. (2005) J. Chem. Phys. 2005, 309, 127-131
Simon-Manso, Y. & Fuenteaelba, P. (1998) J. Phys. Chem. A 1998, 102, 2029-2032.
Tachibana, A. J. (2001). Chem. Phys. 2001, 115, 3497-3518
Tarazona-Vasquez, F. & Balbuena, P. B. (2004a). Plexation of the Lowest Generation Poly(amidoamine)-NH2 Dendrimers with Metal Ions, Metal Atoms, and Cu(II) Hydrates: An ab Initio Study, J. Phys. Chem. B 108, (2004)15992-16001
Tarazona-Vasquez, F. & Balbuena, P. B. (2004b). Ab Initio Study of the Lowest Energy Conformers and IR Spectra of Poly(amidoamine)-G0 Dendrimers J. Phys. Chem. B 108, (2004) 15982-15991
Tomalia, D. A. ; Baker, H. ; Dewald, J. ; Hall, M.; Kallos, G.; Martin, S. ; Roeck, J. ; Ryder, J. & Smith, P. (1985). A New Class of Polymers: Starburst Dendritic Macromolecules Polym. J. 17, (1985) 117-132
Tomalia, D. A. ; Naylor, A. M. & Goddard III, W. A. (1990). Starburst Dendrimers: Molecular-Level Control of Size, Shape, Surface Chemistry, Topology, and Flexibility from Atoms to Macroscopic Matter Angew. Chem.,Int. Ed. Engl., 29 (1990)138-175
Toro-Labbé, A.; Gutiérrez-Oliva, S.; Murray, J.S. & Politzer, P. (2009). J. Mol. Model., 2009, 15, 707-710
Toro-Labbé, A.; Gutiérrez-Oliva, S.; Murray, J.S. & Politzer, P. (2007). Mol. Phys., 2007, 105, 2619–2625
Tozer, D. J. & De Proft, F. (2005). J. Phys. Chem. A 2005, 109, 8923.
Vedral, V. (2002). The role of relative entropy in quantum information theory, Rev Mod. Phys. 74, (2002)197-234
Vögtle, F. ; Gestermann, S.; Hesse, R.; Schwierz, H. & Windisch, B. (2000). Prog. Polym. Sci., 25, (2000) 987.
von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics*, 1955 translated from the German ed. by R. T. Beyer (Princeton University, Princeton).
Wehrl, A. (1978). General properties of entropy, Rev. Mod. Phys. 50, (1978) 221-260
Wigner, E. (1938). Trans. Faraday SOC. 1938, 34, 29-41.
Woodward, R.B. & Hoffmann, R. H. (1969). Angew. Chem. Int. Ed. Engl. 8 (1969) 781.
Zeng, F. & Zimmerman, S. C. (1997). Chem. Rev., 97, (1997) 1681.
Zewail, A.H. (1988). Science, 1988, 242, 1645-1653
Zewail, A.H. (1990). Sci. Am. 263 (1990) 76
Zewail, A. H. (2000a). J. Phys. Chem. A 2000, 104, 5660-5694
Zewail, A.H. (2000b) Phys. Chem. A 104 (2000) 5660-5694.

13

Quantum Computing and Optimal Control Theory

Kenji Mishima
Research Center for Advanced Science and Technology, The University of Tokyo, Komaba, Meguro-ku, Tokyo Japan

1. Introduction

In recent years, quantum computing and quantum information science have become one of the most important and attractive research areas in a variety of disciplines, e. g., mathematics, information science, physics, chemistry, etc[1]. These new kinds of technologies are predicted to be much more advantageous compared with the classical computers and classical information science and the benefit obtained by these technologies is assumed to be beyond measure in our every-day life. For instance, quantum computers are predicted to be able to solve mathematical problems that today's fastest computers could not solve in years. In particular, entanglement or entangled state plays a key role for quantum computing and quantum information processing. For example, arbitrary quantum states of two-level system can be teleported through classical communication with the help of maximally entangled Bell state from one place to other macroscopic distant places (quantum teleportation)[2], which has no counterpart in classical mechanics. As opposed to the quantum teleportation, classical information can be teleported by using the maximally entangled Bell state (superdense coding)[3]. Needless to say, entanglement is also an essential ingredient in quantum computing[1].

At present, theoretical investigations of the mechanism of quantum computing and quantum information science have become mature although some of the important theoretical problems, e. g., definition of entanglement degree of multipartite systems, have not yet been solved and are still controversial. Yet, one can say that we are now reaching a stage of experimental realizations of quantum computing and quantum information processing proposed and investigated theoretically and numerically. To apply quantum computing and quantum information processing to realistic quantum systems, a number of microscopic quantum systems have been proposed. Just to mention a few, cavity quantum electrodynamics (cavity QED)[4], trapped ions[5 - 7], neutral atoms trapped in optical lattices[8], nuclear magnetic resonance (NMR)[9, 10], superconducting circuits[11], silicon-based nuclear spin[12], diamond-based quantum computer[13, 14] are some of the promising candidates of quantum computing devices.

However, investigation of utilization of molecular internal degrees of freedom for quantum computing and quantum information science, in particular, electronic, vibrational, and rotational degrees of freedom, is still in its infancy. Although molecules are also quantum systems, very few chemists have yet examined how to use molecular internal degrees of

freedom for quantum computing and quantum information science from the chemical viewpoint. The pioneering numerical investigation of usage of molecular vibrational states for constructing elementary quantum gates was reported by de Vivie-Riedle and coworkers at the beginning of this century[15]. Later on, they have stick to pursuing "molecular vibrational" quantum computing in a number of papers[16 - 21]. Soon after their works, some of the other research groups have extended their works and have proposed new ideas of quantum computing and quantum information science[22 - 30]. The purpose of many of these works is to numerically construct elementary gate pulses using optimal control theory (OCT)[31]. Instead of using tailored laser pulses, Teranishi and coworkers have developed a quantum computation scheme to process arbitrary quantum gate operations by using the free propagation of the wavepacket of I_2 molecule[32]. Anyway, OCT that has originally stemmed from the necessity of control of chemical reactions has become one of the main procedures for constructing the quantum gates.

Although the "vibrational" quantum computers are the mainstream for the investigations of molecular quantum computing, two-qubit system consisting of one vibrational and one rotational modes of molecules has also been investigated by several researchers[33, 34]. In[33], single- and two- qubit operations, e. g., NOT and CNOT gates, within rotational and vibrational states of a diatomic molecule using strong-field molecular alignment is proposed. Numerical calculations of IR quantum gate pulses for $^{12}C^{16}O$ molecule using a genetic algorithm instead of employing OCT have been investigated by Momose and coworkers[34].

Another possibility is to use *inter*molecular states instead of the *intra*molecular states mentioned above. In[35], one of the methods of realizing quantum phase gate and generation of entanglement rotational modes of two polar molecules coupled by dipole-dipole interaction has been proposed. Unlike their research, we have numerically constructed several universal gates and applied them to the Deutsch-Jozsa algorithm[36].

On the other hand, attempts of experimental realizations of quantum computers using molecular internal degrees of freedom have also begun to be done in recent years. For example, Vala and coworkers experimentally demonstrated the Deutsch-Jozsa algorithm for three-qubit functions by utilizing pure coherent superposition states of Li_2 rovibrational eigenstates[37]. Rovibrational wave-packet manipulation using phase- and amplitude- modulated midinfrared femtosecond laser pulses for $^{12}C^{16}O$ and $^{14}N^{16}O$ molecules have been investigated experimentally and numerically by Momose and coworkers for the purpose of applying their techniques to quantum computing[38]. Ohmori and coworkers experimentally demonstrated coherent control of wavepacket interference, wavepacket interferometry, using vibraiontal wavepackets of I_2 molecule with the aim of retrieving quantum information such as amplitudes and phases of eigenfunctions involved in the wavepacket[39 - 43].

This present situation mentioned above implies that the research of quantum computing using molecular internal degrees of freedom is gradually attracting many physical chemists and chemical physicists in quite recent years.

Interesting aspects of molecules compared with physical systems such as atoms, photons, electron spins, nuclear spins, etc. are that they possess a variety of quantum mechanical internal degrees of freedom. If we restrict ourselves only to two-qubit systems, several kinds of combinations of modes can be considered. The two-qubit combination studied most frequently is vibrational-vibrational qubit combination as mentioned above. Since the investigation of molecular quantum computers is still immature, we predict that there will

be a number of unsolved problems up to now and recommend chemists to investigate molecular quantum computing in more detail in the future although many of the chemists including us have already contributed to the improvement of the molecular quantum computers. For example, we may expect that scalable quantum computing using many internal degrees of freedom will be realized in the future.

The present chapter is organized as follows. In Section 2, we first introduce some of the basic concepts of quantum computers for the convenience of those who are not familiar with quantum computers. One of the most important quantum algorithms, Deutsch-Jozsa algorithm, is also explained shortly. Then, for this chapter to be self-contained, OCT will briefly be reviewed because molecular quantum computing strongly relies on OCT as mentioned above. In Section 3, our development of free-time and fixed end-point optimal control theories (FRFP-OCTs) without and with dissipation is presented and the theory and the algorithm are applied to entanglement generation and maintenance. One will find that the FRFP-OCT is more convenient and advantageous than the conventional fixed end-point optimal control theory (FIFP-OCT). Finally, Section 4 is devoted to concluding remarks.

2. Quantum algorithms

2.1 Quantum gates

Quantum gates are the counterparts of logic gates of classical computer circuits. The definition of operations of the classical single bit logic gates is given by truth table. For example, the operation of NOT gate is to flip the bits: $0 \to 1$ and $1 \to 0$.

In what follows, we list some of the most important quantum gates that are usually used in quantum circuits:

Hadamard gate: $H_{dm} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ for single-qubit gate,

NOT gate: $NOT = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ for single-qubit gate,

CNOT (controlled-not) gate: $CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ for two-qubit gate,

ID gate: $ID = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ for single-qubit gate,

Z gate: $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ for single-qubit gate,

$\pi/8$ gate: $T = \begin{bmatrix} 1 & 0 \\ 0 & \exp(i\pi/4) \end{bmatrix}$ for single-qubit gate,

phase gate: $S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$ for single-qubit gate,

Toffoli gate: $U_T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 & 0 & 0 & -i & 0 \end{bmatrix}$ for three-qubit gate.

For processing the quantum computation, the two-level unitary gates such as shown above must be universal[48]. Here, the term "universal" means that one can implement an arbitrary two-level unitary transformation on the space of arbitrary numbers of qubits. For example, using the Gray codes, it has been proven that single qubit and CNOT gates are universal[1]. It should be emphasized that the gobal unitary transformations such as CNOT gate cannot be reduced to the direct product of two single-qubit gates. Therefore, if the total Hamiltonian can be reduced to the product of two single-qubit unitary transformations, it is impossible to perform universal quantum computation and quantum information processing.

2.2 Deutsch-Jozsa algorithm

So far, several quantum algorithms have been proposed which outperform the corresponding classical algorithms. These include the Grover's algorithm, Shor's algorithm, the quantum Fourier transform, the Deutsch-Jozsa algorithm, etc.[1]. For example, the Shor's algrorithm is a quantum algorithm for integer factorization[51]. On a quantum computer, to factor an integer N, Shor's algorithm takes polynomial time in log$N$, specifically O((log$N$)^3), demonstrating that integer factorization is in the complexity class BQP. This is exponentially faster than the best-known classical factoring algorithm.

For instance, the flowchart of the two-state Deutsch-Jozsa algorithm is shown in Fig. 1. In short, the story of the Deutsch-Jozsa algorithm is as follows. Let us assume two persons, Alice and Bob. Alice holds the so-called query register while Bob holds the so-called answer register. First, they come close together and they make some promises before they go far apart from each other. When they are close together, Alice promises to send the number 0 or 1 to Bob and he promises to calculate some function f and to send her the answer 0 or 1. At this time, Bob promises to use two kinds of functions f. That is, he sends her the same number for all the numbers that he obtains from her (constant function) or he sends 0 for half of the numbers that he obtains from her and 1 for the remaining half (balanced function). After that, they go far apart from each other. The purpose of this algorithm is that Alice must clarify whether the function f that Bob applies is constant or balanced, which is contained in the oracle denoted by U_f. It is known that classically the algorithm scales as $O(2^n)$, while quantum-mechanically it scales as $O(n)$, where n is the number of qubit registers that Alice holds. This demonstrates the significant speedup of quantum parallelism compared with classical algorithms, in particular, when n is very large. In other words, the advantage of quantum parallelism is obtained when the quantum circuit becomes very large.

In the flowchart of Fig. 1, the initial state of the whole Hilbert space is $|00\rangle$. First, Bob applies the NOT gate and the transition $|00\rangle \rightarrow |01\rangle$ occurs. Bob then applies the Hadamard

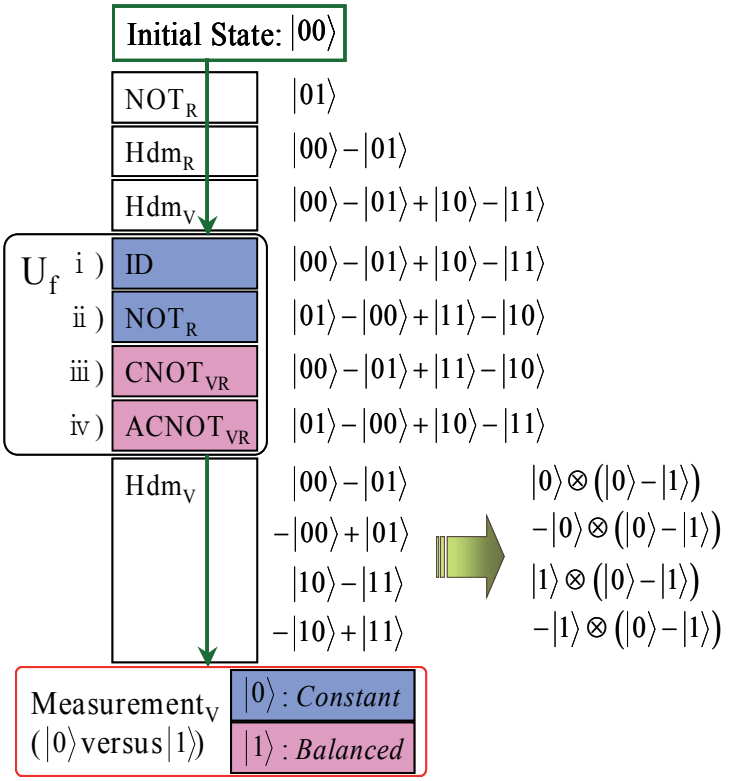


Fig. 1. The two-state Deutsch-Jozsa algorithm

gate Hdm_R and Alice the Hadamard gate Hdm_V. At this moment, the state of the whole system becomes $|00\rangle-|01\rangle+|10\rangle-|11\rangle$. To this quantum superposition state, the unitary transformation, the so-called oracle,

$$U_f : |x,y\rangle \rightarrow |x, y \oplus f(x)\rangle, \tag{1}$$

is applied. Here, $\oplus$ denotes addition modulo 2. The rule of Eq. (1) must be applied for all four possible definitions of f. According to the four definitions, U_f is defined by the four operations (i) ~ (iv) in Fig. 1. Alice then applies the Hadamard transformation Hdm_V. If she recognizes that she obtains the state $\pm|0\rangle$ by her own measurement, f is constant, while f is balanced if she obtains the state $\pm|1\rangle$. These states can be distinguished by measuring her own qubit as shown in Fig. 1.

In Fig. 1, the subscripts of the first and the second entries for elementary quantum gates refer to control bit and target bit, respectively. Here, the abbreviations, V, and R, stand for vibrational and rotational states, respectively.

2.3 Optimal control theory (OCT): General theory and application to molecular quantum computers

As already mentioned in Section 1, to process quantum computing, it is necessary to tailor elementary gate laser pulses appropriately. This particularly holds for molecules. This is because unlike spins molecular modes of internal degrees of freedom are essentially

"qudits", not "qubits." In this section, we will briefly review conventional OCT and multi-target OCT (MTOCT). For more details, we recommend the readers to refer to [52, 53].

If the purpose is just to drive one specific wave function $\psi_i(t)$ to the desired wave function $\Phi(T)$ at the fixed time $t = T$, the objective functional to be maximized is given by[52]

$$J = \left|\langle\psi_i(T)|\Phi(T)\rangle\right|^2 - \alpha_0\int_0^T [E(t)]^2\,dt$$
$$-2\,\mathrm{Re}\left[\langle\psi_i(T)|\Phi(T)\rangle \times \int_0^T \left\langle\psi_f(t)\right|\frac{\partial}{\partial t} + i\left[H_0 + V - \mu E(t)\right]\left|\psi_i(t)\right\rangle dt\right] , \quad (2)$$

where H_0 is the zero-th order Hamiltoinian, V is the potential energy, μ is the transition dipole moment, $E(t)$ is the laser pulse to be optimized, and T is the fixed final time of the laser pulse. The second term restricts the laser intensity, where α_0 is usually called the penalty factor. $|\psi_f(t)\rangle$ is the Lagrange multiplier for $|\psi_i(t)\rangle$.

To incorporate the effect of slow turn-on and turn-off of the laser pulses adequate for practical experimental tailoring, the penalty factor in Eq. (2) is replaced by[53]

$$-\alpha_0\int_0^T \frac{[E(t)]^2}{s(t)}dt\,, \quad (3)$$

where

$$s(t) = \sin^2(\pi t / T)\,. \quad (4)$$

In this case, the optimized external field is expressed as

$$E(t) = -\frac{s(t)}{\alpha_0}\mathrm{Im}\left\{\langle\psi_i(t)|\psi_f(t)\rangle\langle\psi_f(t)|\mu|\psi_i(t)\rangle\right\}. \quad (5)$$

Although the above formalisms may be applicable to tailoring the gate laser pulses for any quantum control problems, they are not appropriate for tailoring *general-purpose* global gate pulses that are required for quantum computing. In other words, the given gate pulse has to process the given quantum gate for any input states and the corresponding output states. In this case, one of the best choices is to resort to multi-target optimal control theory (MTOCT)[54]. For MTOCT, the objective functional to be maximized is given by

$$J_{MTOCT} = \sum_{k=1}^{z}\left\{\left|\langle\psi_{ik}(T)|\Phi_{fk}(T)\rangle\right|^2 - \alpha_0\int_0^T \frac{[E(t)]^2}{s(t)}dt\right.$$
$$\left.-2\,\mathrm{Re}\left\{\langle\psi_{ik}(T)|\Phi_{fk}(T)\rangle \times \int_0^T \left\langle\psi_{fk}(t)\right|\frac{\partial}{\partial t} + i\left[H_0 + V - \mu E(t)\right]\left|\psi_{ik}(t)\right\rangle dt\right\}\right\} , \quad (6)$$

where z is the number of control targets, k denotes the number of targets ranging from 1 to z, $|\Phi_{fk}(T)\rangle$ is the k-th target at time $t = T$, $|\psi_{ik}(t)\rangle$ is the wavefunction of the system of the k-th target, and $|\psi_{fk}(t)\rangle$ is the Lagrange multiplier for $|\psi_{ik}(t)\rangle$. In this case, the optimal external field reads

$$E(t) = -\frac{zs(t)}{\alpha_0}\sum_{k=1}^{z}\mathrm{Im}\left\{\left\langle\psi_{ik}(t)\middle|\psi_{fk}(t)\right\rangle\left\langle\psi_{fk}(t)\middle|\mu\middle|\psi_{ik}(t)\right\rangle\right\}. \tag{7}$$

The number of the control targets z has to be chosen as follows. Recently, de Vivie-Riedle and coworkers[17] proposed a method for phase-correct and basis-set-independent quantum gates in order to perform the correct universal quantum computing. As far as we know, their work is the first one where the phase correction was taken into account adequately. The requirement of the phase-correct quantum gate is that, for example, the NOT operation for the superposition state,

$$|00\rangle + |01\rangle + |10\rangle + |11\rangle \rightarrow \left(|01\rangle + |00\rangle + |11\rangle + |10\rangle\right)e^{i\varphi_5}, \tag{8}$$

must be optimized in addition to the following four conventional pure basis state optimizations,

$$|00\rangle \rightarrow |01\rangle e^{i\varphi_1},$$

$$|01\rangle \rightarrow |00\rangle e^{i\varphi_2},$$

$$|10\rangle \rightarrow |11\rangle e^{i\varphi_3},$$

$$|11\rangle \rightarrow |10\rangle e^{i\varphi_4}. \tag{9}$$

If we do not impose the requirement of Eq. (8), the superposition state will evolve as:

$$|00\rangle + |01\rangle + |10\rangle + |11\rangle \rightarrow |01\rangle e^{i\varphi_1} + |00\rangle e^{i\varphi_2} + |11\rangle e^{i\varphi_3} + |10\rangle e^{i\varphi_4}, \tag{10}$$

which is not the correct NOT operation, because in general $\varphi_1 \neq \varphi_2 \neq \varphi_3 \neq \varphi_4$. Likewise, we must impose additional constraints for the other quantum gates we have in mind. As de Vivie-Riedle and coworkers pointed out[17], the phase correction of quantum gates is one of the key issues for the implementation of quantum algorithms. Therefore, for two-qubit systems, z has to be more than 4.

There are two methods to measure the gate fidelities: the average transition probability given by

$$\overline{P} = \frac{1}{z}\sum_{k=1}^{z}\left|\left\langle\psi_{ik}(T)\middle|\Phi_{fk}(T)\right\rangle\right|^2, \tag{11}$$

and the fidelity expressed as

$$F = \frac{1}{z^2}\left|\sum_{k=1}^{z}\left\langle\psi_{ik}(T)\middle|\Phi_{fk}(T)\right\rangle\right|^2. \tag{12}$$

The average transition probability cannot take into account the phase relation between $\psi_{ik}(T)$ and $\Phi_{fk}(T)$, while the fidelity can. If one uses the average transition probability, the phase correction cannot be determined, while the fidelity is useful for clarifying the phase

correction. Therefore, we must define the laser pulses that have the largest fidelity as the optimal gate pulses.

3. Free-time and fixed end-point optimal control theory (FRFP-OCT)

In a series of our publications[55, 56], we have found that the entanglement generation in general quantum systems crucially rely on the strength of entangling interactions among distinct quantum systems. We have stressed that if the entangling interactions are strong, the maximally entangled state can be created in a short time. This in turn implies that if the strength of the entangling interactions is weak, long laser fields are necessary for creating the maximally entangled states. Approximately, the time duration of the laser pulses by which the maximally entangled states can be created is inversely proportional to the strength of the entangling interaction. Therefore, it can easily be recognized that we need a new OCT that works well even if we do not know the necessary time duration of the laser pulses to create the maximally entangled state efficiently because the actual entangling interactions are usually much more complicated in molecular systems. If this is the case, the necessary OCT will become *free-time* and fixed end-point optimal control theory (FRFP-OCT) since the optimal temporal duration of the laser pulses is not known exactly in advance. Currently, OCT in quantum systems proposed so far has been limited to the *fixed-time* and fixed end-point optimal control theory (FIFP-OCT)[57]. Consequently, we have constructed one of the versions of FRFP-OCTs that can optimize the objective functional and the temporal duration of the laser pulses simultaneously[58, 59]. One of the advantages of our theory is that one does not need to try various final fixed times to achieve the best control of quantum dynamics. To demonstrate the utility of our theory it has been applied to the optimization of laser pulses that can create maximally entangled states efficiently, but it may also be applied to various physical and chemical quantum control problems.

On the other hand, realistic quantum systems that we observe experimentally and calculate theoretically are always interacting with surrounding environment by way of entangling interactions. If the whole quantum system is the sum of the system of our interest and the huge surrounding environment, the quantum state is maintained in pure state (no decoherence). However, the surrounding environment is traced out and our attention is paid only to our small quantum system, our system becomes mixed state (decoherence). This can be easily verified by using, e. g., the von-Neumann entropy used to measure entanglement degree of the pure state of composite systems. In many quantum control problems, the decoherence is unfavorable and should be suppressed.

Quantum computing and quantum information science are also not exceptions. It was pointed out that the decoherence might become one of the crucial obstacles for quantum computers and entanglement generation and manipulation because quantum information processing must be performed in pure states in most cases[60, 61]. Therefore, to achieve accurate quantum computing and quantum information processing in the quantum system in contact with the surrounding environment, it is crucial to maintain the coherence by external active manipulation of the target quantum system.

At present, there are two methods to suppress decoherence that are proposed theoretically. One of these is to utilize quantum error correcting code[62, 63]. The other promising and efficient method of preventing decoherence is the so-called bang-bang control by shining repetitive intense laser pulses on the target quantum system[64].

Although the methods mentioned above are proposed to be applied to simple two-level quantum systems (qubits), most quantum systems are composed of many eigenstates (qudits), e. g., molecular internal degrees of freedom. Therefore, the analytical approaches of the error correcting code and the bang-bang control cannot easily be extended to qudits such as molecular modes. If this is the case, one has to resort to other methods for the purpose of decoherence suppression of the realistic molecular systems. One of the advantageous methods will be to OCT and apply it to concrete calculations of realistic multi-level quantum systems in order to control the dissipative quantum dynamics most efficiently.

In fact, OCT for dissipative quantum dynamics has attracted much attention in recent years. This is because it is possible to construct laser pulses that can manipulate quantum dynamics efficiently in the presence of the surrounding environment and because it is difficult to predict by intuition what kind of laser pulses are the most appropriate for achieving the target dissipative quantum dynamics. OCT for the dissipative quantum dynamics has been developed and improved by many researchers. For example, the OCT for dissipative quantum systems was constructed in a fully systematic and rigorous fashion by Cao and coworkers for the first time[65]. However, their theory can only be applied to the weak response regime. Almost at the same time, the OCT in the strong response regime was developed in terms of the Liouville-space density matrix[66]. Ohtsuki and coworkers developed a monotonically convergent algorithm for dissipative quantum systems[67] and applied their theory to the control of wavepacket dynamics under the influence of dissipation[68]. Recently, there have appeared several numerical applications of OCT in realistic dissipative media for a variety of purposes. For example, simulations of molecular quantum computers using the vibrational modes of molecules including dissipation have been performed by Ndong and coworkers[27]. Seideman and coworkers have applied dissipative OCT to manipulate rotational wavepacket dynamics in a dissipative environment[69, 70]. From the experimental viewpoint, dissipative OCT was used for the quantum control of I_2 in the gas phase and in condensed phase solid Kr matrix[71].

Also for the quantum control in the dissipative environment, only FIFP-OCTs have been developed. Dissipative quantum dynamics can be regarded as on of the most time-sensitive processes. The reason is that the decoherence rate Γ governs the decoherence degree versus time. Therefore, FRFP-OCT also has a significant importance for dynamical control of dissipative quantum dynamics. If this is the case for the quantum system under investigation, the equation of motion should be replaced by, e. g., the Liouville-von Neumann equation in the framework of the density matrix representation. Consequently, one of the main purposes here is to generalize FRFP-OCT suitable only for pure states to *mixed state* FRFP-OCT following the general Master equation in both Markov approximation and without any approximations.

3.1 FRFP-OCT in pure state

We assume that the quantum system of our interest is separated from the surrounding environment so that our system can adequately be described by the Schrödinger equation. The objective functional of our problem to be maximized is just given by

$$J = \left| \left\langle \Psi_i(T) \middle| \Phi_f \right\rangle \right|^2 , \tag{13}$$

where $|\Psi_i(t)\rangle$ is time-dependent wavefunction at time t and $|\Psi_i(T)\rangle$ is the time-dependent wavefunction at the target final time $t = T$. On the other hand, $|\Phi_f\rangle$ is the final target wavefunction at time $t = T$. Our purpose is to maximize the objective function, J, at some time T. Note that we *do not fix T* while J should be maximized. This kind of problem has not yet been investigated in control problems in quantum mechanics so far. It should be noticed that the objective functional given by Eq. (13) is different from that of the optimal control theory investigated so far. In the conventional FIFP-OCT, the objective functional is usually given by

$$J = \left|\langle \Psi_i(T) | \Phi_f \rangle\right|^2 - \alpha \int_0^T E(t)^2 dt \,, \tag{14}$$

where $E(t)$ is the external laser fields and α is usually called penalty factor that is added to minimize the strength of the external laser fields. Defining the objective functional as Eq. (14) and adding the constraints that the system obeys, Rabitz and coworkers proposed, e. g., monotonically convergent OCT[52].

Let us now derive the quantum mechanical FRFP-OCT that is necessary, e. g., for entanglement generation as mentioned above. First, we introduce *real* time t and *fictitious*-time τ, which are related by the following equality:

$$t = T(\tau)\tau \,, \tag{15}$$

where τ is a dimensionless parameter that ranges from zero to unity In addition, we have included the implicit dependence of T on dimensionless parameter τ in Eq. (15). The time-dependent equation for $|\Psi_i(t)\rangle$ is given by the conventional *real-time* Schrödinger equation:

$$i\hbar \frac{\partial |\Psi_i(t)\rangle}{\partial t} = \left\{\hat{H} - \vec{\mu}\cdot\vec{E}(t)\right\} |\Psi_i(t)\rangle \,, \tag{16}$$

where $\hat{H}$ is the zero-th order Hamiltonian and $-\vec{\mu}\cdot\vec{E}(t)$ is the laser-molecule interaction. Using the relationship of Eq. (15) for Eq. (16), we obtain

$$i\hbar \frac{\partial |\Psi_i(\tau)\rangle}{\partial \tau} = \left\{\hat{H} - \vec{\mu}\cdot\vec{E}(\tau)\right\} |\Psi_i(\tau)\rangle T(\tau) \,. \tag{17}$$

We may call Eq. (17) as *fictitious-time* Schrödinger equation.

Usually, the objective functional to be maximized or minimized is constrained by some of the factors, e. g., the equation of dynamics that the problem in mind follows. In this case, we can add such constraints into Eq. (17) using Lagrange multipliers and we obtain the new objective functional,

$$\hat{J} = \left|\langle \Psi_i(\tau=1) | \Phi_f \rangle\right|^2 - 2\,\mathrm{Re}\left[\int_0^1 \langle \Psi_f(\tau)| \left\{\frac{i}{\hbar}\left(\hat{H} - \vec{\mu}\cdot\vec{E}(\tau)\right) + \frac{\partial}{T(\tau)\partial\tau}\right\} |\Psi_i(\tau)\rangle T(\tau) d\tau\right]$$

$$- \int_0^1 \nu_T(\tau) \frac{\partial T(\tau)}{\partial \tau} d\tau \,. \tag{18}$$

Then, we introduce the variational principle for Eq. (18). In order for $\hat{J}$ to be maximized, we can deduce the following equations:

$$i\hbar\frac{\partial\left|\Psi_i(\tau)\right\rangle}{\partial\tau}=\left\{\hat{H}-\vec{\mu}\cdot\vec{E}(\tau)\right\}\left|\Psi_i(\tau)\right\rangle T(\tau)$$

subject to the initial condition

$$\left|\Psi_i(\tau=0)\right\rangle=\left|\Phi_i\right\rangle, \tag{19}$$

where $\left|\Phi_i\right\rangle$ is the initial given state.

$$i\hbar\frac{\partial\left|\Psi_f(\tau)\right\rangle}{\partial\tau}=\left\{\hat{H}-\vec{\mu}\cdot\vec{E}(\tau)\right\}\left|\Psi_f(\tau)\right\rangle T(\tau)$$

subject to the initial condition

$$\left|\Psi_f(\tau=1)\right\rangle=\left|\Phi_f\right\rangle, \tag{20}$$

$$\frac{\partial\nu_T(\tau)}{\partial\tau}=-\frac{2}{\hbar}\mathrm{Im}\left\{\left\langle\Psi_f(\tau)\right|\left[\hat{H}-\vec{\mu}\cdot\vec{E}(\tau)\right]\left|\Psi_i(\tau)\right\rangle\right\}$$

subject to the initial condition

$$\nu_T(\tau=1)=2\,\mathrm{Re}\left\{\left\langle\Psi_i(\tau=1)\middle|\Phi_f\right\rangle\left\langle\Phi_f\middle|\frac{\partial\Psi_i(\tau=1)}{T(\tau=1)\partial\tau}\right\rangle\right\}, \tag{21}$$

When Eqs. (19)~(21) are satisfied, we have

$$\delta\hat{J}=\int_0^1 d\tau g(\tau)\delta E(\tau)+\nu_T(\tau=0)\delta T(\tau=0), \tag{22}$$

where we have defined

$$g(\tau)=-\frac{2}{\hbar}\mathrm{Im}\left\{\left\langle\Psi_f(\tau)\middle|\mu\middle|\Psi_i(\tau)\right\rangle\right\}T(\tau). \tag{23}$$

If the correction of the laser amplitude $E(\tau)$ is represented as $\delta E(\tau)$, we define

$$\delta E(\tau)=\alpha g(\tau). \tag{24}$$

On the other hand, if we defined the correction of $T(\tau)$ as $\delta T(\tau)$, we choose

$$\delta T(\tau)=\beta\nu_T(\tau=0). \tag{25}$$

When Eqs. (24) and (25) are inserted into Eq. (22), we obtain

$$\delta\hat{J}=\int_0^1 d\tau\alpha g(\tau)^2+\beta\nu_T(\tau=0)^2. \tag{26}$$

If both α and β are positive, it is expected that the objective reaches maximum monotonically as is clearly understood from Eq. (26). On the other hand, if both α and β are negative, it is expected that the objective reaches minimum monotonically.
Based on the above equations, we have constructed the following FRFP-OCT in pure state following the Schrödinger equation. In what follows, the superscript (*j*) is used to denote the quantities for the *j*-th iteration.

i. One chooses initial guess external fields $E^{(0)}(\tau)$ and nominal $T^{(0)}$ that is the final time of quantum dynamics. Here and in the following, the superscript (*j*) is used to denote the quantity of the *j*-th iteration. In addition, the trial positive parameters α and β are given because our purpose is to maximize Eq. (13).
ii. The Schrödinger equation, Eq. (19), is propagated forwardly in time from $\tau = 0$ to $\tau = 1$ and the obtained wavefunction $\left|\Psi_i^{(j)}(\tau)\right\rangle$ is stored. At the same time, the objective functional $J^{(j)} = \left|\left\langle \Psi_i^{(j)}(T)\middle|\Phi_f\right\rangle\right|^2$ is calculated.
iii. Equations (20) and (21) are propagated backwardly in time from $\tau = 1$ to $\tau = 0$ and the wavefunction $\left|\Psi_f^{(j)}(\tau)\right\rangle$ is stored. In addition, $\nu_T(\tau = 0)$ is calculated.
iv. Using Eqs. (24) and (25), the laser amplitude $E^{(j)}(\tau)$ and $T^{(j)}$ are updated as follows,

$$E^{(j+1)}(\tau) = E^{(j)}(\tau) + \alpha g(\tau), \tag{27}$$

and

$$T^{(j+1)} = T^{(j)} + \beta \nu_T(\tau = 0). \tag{28}$$

v. One sets the convergence criterion η and if the following criterion

$$\left|J^{(j+1)} - J^{(j)}\right| \le \eta \tag{29}$$

is met, the calculation is terminated.
vi. If the convergence is not sufficient, one updates $E^{(j)}(\tau)$ and $T^{(j)}$ to $E^{(j+1)}(\tau)$ and $T^{(j+1)}$, and loops back to the step (ii).

To show how our theory works concretely by showing calculation results, we have applied the above algorithm to tailoring optimal laser pulses that can create the maximally entangled Bell states between NaCl and NaBr molecules coupled by dipole-dipole interaction. One of the calculation examples is shown in Fig. 2.
In Fig. 2, we show the numerical results for the optimization of the quantum transfer $|0,0\rangle \to \left(|0,0\rangle + |1,1\rangle\right)/\sqrt{2}$ with the nominal $T^{(0)}$ = 300 ps. From panel (a), we can see that the rate of the monotonic convergence of the transition probability is better for FRFP-OCT than that for FIFP-OCT. In addition, the finally obtained transition probability is better for FRFP-OCT. On the other hand, from panel (b), it is seen that the temporal duration of the laser pulse becomes longer with the optimization iteration. This reflects the fact that the longer temporal duration of the laser pulse is more favorable than the shorter one because the nominal $T^{(0)}$ was too short to reach a high transition probability. It is clear from panels (d) and (f), the maximally entangled Bell state cannot be created by both FRFP-OCT and FIFP-

OCT. This is because the tailored laser pulses have a short temporal duration so that it is difficult to reach the maximally entangled state as mentioned above. However, it is clearly seen that FRFP-OCT has attained much higher transition probability than FIFP-OCT has (see panel (f)). The optimal time duration of the laser pulse obtained by FRFP-OCT was 327.95 ps. It is expected that the behaviors shown in these figures are also universal to controls of other physical and chemical phenomena.

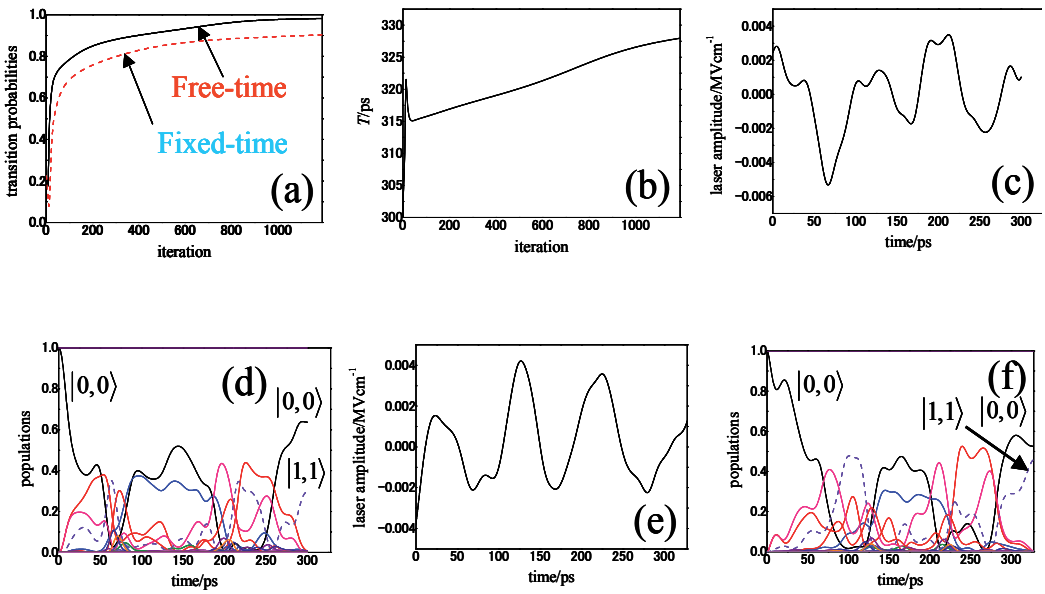


Fig. 2. (a) trasition probability versus iteration number, (b) temporal durationof the optimized laser pulse versus iteration number, (c) optimized laser pulse with α and β being equal to 2×10^{-16} a.u. and 0.0 a.u., respectively, (d) population transfer for panel (c), (e) optimized laser pulse with α and β being equal to $2x10^{-16}$ a.u. and $2x10^{11}$ a.u., respectively, and (f) population transfer for panel (e). The nominal $T^{(0)}$ was set to he 300 ps. The intermolecular distance R is equal to 5.0 nm. In this figure, the target transition $|0,0\rangle \rightarrow (|0,0\rangle + |1,1\rangle)/\sqrt{2}$ was optimized.

From the above numerical results, we can conclude that our FRFP-OCT is much more efficient than the conventional FIFP-OCT because the temporal duration of the laser pulse can also be optimized accurately, which makes OCT more flexible.

3.2 FRFP-OCT in dissipative media

Next, we are interested in the situation where the quantum system of interest is affected by the surrounding environment so that it is necessary to describe the quantum system in the density-matrix representation. In such a case, we start from the assumption that the objective functional to be maximized is simply given by

$$J = \left\langle \left\langle \hat{W} \middle| \hat{\rho}(T) \right\rangle \right\rangle, \tag{30}$$

where $\hat{\rho}(t)$ represents the time-dependent reduced density matrix at time t, $\hat{\rho}(T)$ is the time-dependent reduced density matrix at the target final time $t = T$, and $\hat{W}$ is the objective reduced density matrix. The notation in Eq. (30), $\left\langle\left\langle \hat{B}\middle|\hat{C}\right\rangle\right\rangle$ for arbitrary matrices $\hat{B}$ and $\hat{C}$, is defined by

$$\left\langle\left\langle \hat{B}\middle|\hat{C}\right\rangle\right\rangle = Tr\left(\hat{B}^{\dagger}\hat{C}\right). \tag{31}$$

Equation (31) measures the degree of closeness between the matrices $\hat{B}$ and $\hat{C}$. Then, our purpose is to maximize the objective function, J, at some time T. Note that we *do not fix T* while J should be maximized. It should be noticed that the objective functional given by Eq. (30) is different from that of the conventional FIFP-OCT. In the theory, the objective functional is usually given by[67]

$$J = \left\langle\left\langle \hat{W}\middle|\hat{\rho}(T)\right\rangle\right\rangle - \frac{1}{\hbar A}\int_0^T E(t)^2 dt, \tag{32}$$

where $E(t)$ is the external laser field and the positive constant A is the penalty factor to weigh the significance of the pulse fluence. Because of this difference, our derivation of the OCT in dissipative media is also quite different from theirs.

For the FRFP-OCT, we will again introduced the fictitious time defined by Eq. (15). In real time, the time-dependent equation for the reduced density matrix, $\hat{\rho}(t)$, is expressed as:

$$i\hbar\frac{\partial\hat{\rho}(t)}{\partial t} = \left(\hat{\hat{L}}_0 + \hat{\hat{L}}_{el}(t) - i\hbar\hat{\hat{\Gamma}}\right)\hat{\rho}(t), \tag{33}$$

where

$$\hat{\hat{L}}_0\hat{\rho}(t) = \left[\hat{H}_0, \hat{\rho}(t)\right], \quad \hat{\hat{L}}_{el}(t)\hat{\rho}(t) = \left[\hat{H}_{el}(t), \hat{\rho}(t)\right], \tag{34}$$

and $\hat{\hat{\Gamma}}$ is the damping operator due to the interaction between the system of interest and the surrounding environment. $\hat{H}_0$ is the zeroth-order Hamiltonian and $\hat{H}_{el}(t) = -\vec{\mu}\bullet\vec{E}(t)$ is the laser-molecule interaction with $\vec{\mu}$ being the transition dipole moment. Using the relationship of Eq. (15) for Eq. (33), we obtain the so-called *fictitious time* Master equation,

$$i\hbar\frac{\partial\hat{\rho}(\tau)}{\partial\tau} = \left(\hat{\hat{L}}_0 + \hat{\hat{L}}_{el}(\tau) - i\hbar\hat{\hat{\Gamma}}\right)\hat{\rho}(\tau)T(\tau). \tag{35}$$

When the objective functional to be optimized is constrained by some equations, we should sum up such constraints into Eq. (30) using Lagrange multipliers. Then, we obtain the following new objective function,

$$\begin{aligned}\overline{J} = \left\langle\left\langle \hat{W}\middle|\hat{\rho}(\tau=1)\right\rangle\right\rangle - \int_0^1\left\langle\left\langle \hat{\sigma}(\tau)\middle|\left\{\frac{i}{\hbar}\left(\hat{\hat{L}}_0 + \hat{\hat{L}}_{el}(\tau) - i\hbar\hat{\hat{\Gamma}}\right) + \frac{\partial}{T(\tau)\partial\tau}\right\}T(\tau)\middle|\hat{\rho}(\tau)\right\rangle\right\rangle d\tau \\ -\int_0^1 \nu_T(\tau)\frac{\partial T(\tau)}{\partial\tau}d\tau\end{aligned} \tag{36}$$

For $\overline{J}$ to be maximized, it is possible to deduce the following equations by applying variational principle to Eq. (36):

$$i\hbar\frac{\partial\hat{\rho}(\tau)}{\partial\tau}=\left(\hat{\hat{L}}_0+\hat{\hat{L}}_{el}(\tau)-i\hbar\hat{\hat{\Gamma}}\right)\hat{\rho}(\tau)T(\tau)$$

subject to the initial condition

$$\hat{\rho}(\tau=0)=\hat{W}_0\,, \tag{37}$$

where $\hat{W}_0$ is the initial fixed reduced density matrix,

$$i\hbar\frac{\partial\hat{\sigma}(\tau)}{\partial\tau}=\left(\hat{\hat{L}}_0+\hat{\hat{L}}_{el}(\tau)-i\hbar\hat{\hat{\Gamma}}\right)^{\dagger}\hat{\sigma}(\tau)T(\tau)$$

subject to the initial condition

$$\hat{\sigma}(\tau=1)=\hat{W}\,, \tag{38}$$

where the superscript, †, denotes Hermitian conjugation,

$$\frac{\partial\nu_T(\tau)}{\partial\tau}=\left\langle\left\langle\hat{\sigma}(\tau)\right|\frac{i}{\hbar}\left(\hat{\hat{L}}_0+\hat{\hat{L}}_{el}(\tau)-i\hbar\hat{\hat{\Gamma}}\right)\left|\hat{\rho}(\tau)\right\rangle\right\rangle$$

subject to the initial condition

$$\nu_T(\tau=1)=\frac{1}{T(\tau=1)}\left\langle\left\langle\hat{W}\right|\partial\hat{\rho}(\tau=1)/\partial\tau\right\rangle\right\rangle. \tag{39}$$

When Eqs. (37) - (39) are satisfied, we have

$$\delta\overline{J}=\int_0^1 d\tau g(\tau)\delta E(\tau)+\nu_T(\tau=0)\delta T(\tau=0)\,, \tag{40}$$

where we have defined

$$g(\tau)=\frac{i}{\hbar}\left\langle\left\langle\hat{\sigma}(\tau)\right|\frac{\partial\hat{\hat{L}}_{el}(\tau)}{\partial E(\tau)}\left|\hat{\rho}(\tau)\right\rangle\right\rangle T(\tau)\,. \tag{41}$$

Note that $g(\tau)$ is real. If the correction to the laser amplitude $E(\tau)$ is expressed as $\delta E(\tau)$, we define

$$\delta E(\tau)=\alpha g(\tau)\,. \tag{42}$$

On the other hand, if we define the correction to $T(\tau)$ as $\delta T(\tau)$, we put

$$\delta T(\tau=0)=\beta\nu_T(\tau=0)\,. \tag{43}$$

By inserting Eqs. (42) and (43) into Eq. (40), we obtain

$$\delta\overline{J}=-\int_0^1 d\tau\alpha g(\tau)^2+\beta\nu_T(\tau=0)^2\,. \tag{44}$$

From this equation, it is clear that if α is negative and β is positive, the objective function reaches a maximum monotonically. On the other hand, if α is positive and β is negative, the objective functional reaches minimum monotonically. Here, it should be noted that the units of α and β are Wcm^{-2} and fs^2, respectively.

From the above derivation, we have constructed the following FRFP-OCT in dissipative media following the Master equation. In what follows, the superscript (j) is used to denote the quantity for the j-th iteration.

i. An initial guess is selected for the external field $E^{(0)}(\tau)$ and initial $T^{(0)}$ that is the final time of the quantum dynamics. In addition, the trial negative and positive parameters, α and β, are given because our purpose is to maximize Eq. (30).

ii. The Master equation of Eq. (37) is propagated forward in time from $\tau = 0$ to $\tau = 1$ and the obtained density matrix $\hat{\rho}^{(j)}(\tau)$ is stored. At the same time, the objective function $J^{(j)} = \left\langle\left\langle \hat{W} \middle| \hat{\rho}^{(j)}(\tau = 1) \right\rangle\right\rangle$ is calculated.

iii. Equations (38) and (39) are propagated backward in time from $\tau = 1$ to $\tau = 0$ and the density matrix $\hat{\sigma}^{(j)}(\tau)$ is stored. At the same time, $\nu_T(\tau = 0)$ is calculated.

iv. The laser amplitude $E^{(j)}(\tau)$ and the temporal duration of the external field $T^{(j)}$ are updated as follows,

$$E^{(j+1)}(\tau) = E^{(j)}(\tau) + \alpha g(\tau), \tag{45}$$

and

$$T^{(j+1)} = T^{(j)} + \beta \nu_T(\tau = 0). \tag{46}$$

v. One sets the convergence criterion η and when the following criterion

$$\left| J^{(j+1)} - J^{(j)} \right| \le \eta \tag{47}$$

is met, the calculation is terminated.

vi. If the convergence criterion of Eq. (47) is not satisfied, $E^{(j)}(\tau)$ and $T^{(j)}$ are updated to $E^{(j+1)}(\tau)$ and $T^{(j+1)}$, respectively, and loop back to step (ii).

To apply the theory and the algorithm developed above and demonstrate numerical tests, we will employ the vibrational degrees of freedom of carbon monoxide adsorbed on the copper (100) surface, CO/Cu(100). In this case, the total Hamiltonian $\hat{H}$ in the absence of the laser fields is expressed as

$$\hat{H} = \hat{H}_0 + \hat{V}, \tag{48}$$

where $\hat{H}_0$ is the kinetic energy operator and $\hat{V}$ is the potential energy operator defined in the next section. When we introduce three coordinates r, Z, and X for CO stretch, CO-surface stretch, and frustrated translation modes, respectively, $\hat{H}_0$ is given by

$$\hat{H}_0 = -\frac{\hbar^2}{2\mu_{CO}}\frac{\partial^2}{\partial r^2} - \frac{\hbar^2}{2m_{CO}}\frac{\partial^2}{\partial Z^2} - \frac{\hbar^2}{2m_{CO}}\frac{\partial^2}{\partial X^2}, \tag{49}$$

where the masses are

$$\mu_{CO} = \frac{m_C m_O}{m_C + m_O} = 6.856 \text{ amu}, \; m_{CO} = m_C + m_O = 27.995 \text{ amu}. \tag{50}$$

The eigenstates and eigenenergies of the Hamiltonian, $\hat{H}$, are calculated from

$$\hat{H}\left|n_r, n_Z, n_X\right\rangle = E_n \left|n_r, n_Z, n_X\right\rangle, \tag{51}$$

where we have used the abbreviation $\left|n\right\rangle \equiv \left|n_r, n_Z, n_X\right\rangle$ and E_n is the eigenenergy of the state $\left|n\right\rangle$. Here, n_r, n_Z, and n_X denote the quanta of vibrational modes, CO stretch, CO-surface stretch, and frustrated translation, respectively.
The Liouville-von Neumann equation in the Markov approximation in the energy representation is explicitly expressed as

$$\frac{d\rho_{nn}(t)}{dt} = -\frac{i}{\hbar} E_z(t) \sum_{i=1}^{N} \left\{ \mu_{ni}\rho_{in}(t) - \rho_{ni}(t)\mu_{in} \right\} + \sum_{i=1}^{N} \left\{ \Gamma_{i\to n}\rho_{ii}(t) - \Gamma_{n\to i}\rho_{nn}(t) \right\} \tag{52}$$

for the diagonal elements (populations) of the reduced density matrix and

$$\frac{d\rho_{mn}(t)}{dt} = -i\omega_{mn}\rho_{mn}(t) - \frac{i}{\hbar} E_z(t) \sum_{i=1}^{N} \left\{ \mu_{mi}\rho_{in}(t) - \rho_{mi}(t)\mu_{in} \right\} - \gamma_{m\to n}\rho_{mn}(t) \tag{53}$$

for the off-diagonal elements (coherences). Here, we have defined the energy gap,

$$\omega_{nm} = \left(E_n - E_m\right)/\hbar . \tag{54}$$

The total dephasing rate is given by

$$\gamma_{mn} = \sum_{i=1}^{N} \left(\Gamma_{m\to i} + \Gamma_{n\to i}\right)/2 + \gamma^*_{m\to n}, \tag{55}$$

where $\gamma^*_{m\to n}$ is the pure dephasing rate and $\Gamma_{m\to n}$ is the population transfer rate from the state m to the state n. The values of these parameters were taken from [78]. For the pure dephasing rate, we have taken into account $\gamma^*_{(0,0,0)\to(1,0,0)} \approx \gamma^*_{(1,0,0)\to(2,0,0)} \approx \gamma^*_{(0,0,0)\to(2,0,0)}/4$ with values taken from Table IV of [72]. For the same reason as mentioned in [72], the precise values of the pure dephasing rates are of no concern in the present calculations.
To check the mixedness of the reduced density matrix in the Hilbert space of our interest (CO stretch and CO-surface stretch modes), we explicitly define it by

$$\text{mixedness} = 1 - Tr_{frust}\left\{\rho(t)^2\right\}, \tag{56}$$

where Tr_{frust} denotes the trace over the frustrated translation mode that is of no concern.

Note that we can apply our algorithm to other types of Master equations in addition to the Liouville-von Neumann equation mentioned above.

We have investigated the configuration of the CO/Cu(100) system shown in Fig. 3. We have taken into account two layers of copper atoms and in each layer the nearest nine Cu atoms in the same manner as in [73].

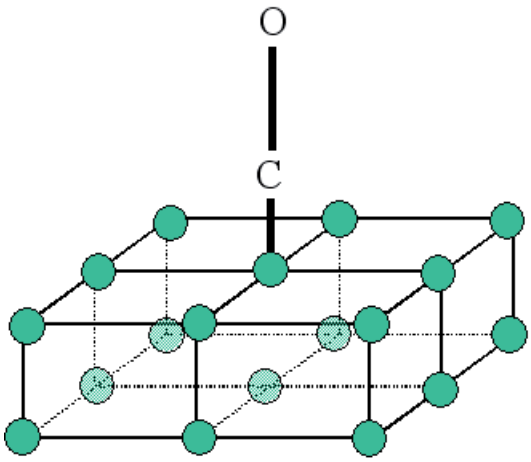


Fig. 3. Schematic of the dissipative CO/Cu(100) system used to apply FRFP-OCT in dissipative media. The Solid circles represent Cu atoms.

The purpose here is two-folds. First, we shall tailor the optimal laser pulses that create maximally entangled Bell state $\left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2}$ from the separable state $|0,0,0\rangle$. Of course, this is of fundamental importance for quantum computing and quantum information science. Second, we assume that the maximally entangled state, $\left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2}$, is prepared at t = 0 fs. We shall examine by what kinds of laser pulses this state is maintained in the presence of dissipation. That is, our target transition is $\left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2}\rightarrow\left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2}$. This problem seems to be important to study in detail because it may be necessary to maintain some specific entangled sates during other processes in large-scale quantum computers composed of many qubits. Because the effect of decoherence generally seems to be negligible in low temperatures, it may be difficult to show the influence of dissipation on the optimal control. Therefore, we shall mainly present numerical results at high temperatures in the following.

In Fig. 4, we show the case where the initial temporal duration of the laser pulse, $T^{(0)}$, is 1000 fs. The maximum transition probability is attained at T = 996.219 fs, as shown in panel (c). In this case, the incident laser pulse has a shape quite different from that of the other cases. As is clear from panel (a), the laser amplitude from the initial time t = 0 fs to around the time t = 800 fs is quite small (~4 MVcm^{-1}). Therefore, we can hardly observe the population transfer due to the laser pulse. Instead, we can see a significant population transfer from the state $|0,0,0\rangle$ to the state $|0,0,1\rangle$ because of the large population transfer rate, $1/\Gamma_{(0,0,0)\rightarrow(0,0,1)}=3.3$ ps. This transition represents the absorption of the single reservoir quantum by the frustrated translation mode. From the time t = 800 fs to the optimal final time T = 996.219 fs, the amplitude of the optimized laser pulse is quite large (~60 MVcm^{-1}) so that a significant population transfer from the state $|0,0,0\rangle$ to the target state $|1,1,0\rangle$ takes

place and coherence between the states, $|0,0,0\rangle$ and $|1,1,0\rangle$, builds up during this period. These trends are reasonable because if the transition to the target state $|1,1,0\rangle$ occurred much earlier as the result of intense laser pulses, the damping of the population of the state $|1,1,0\rangle$ to other states and the decoherence could be quite significant, which would lead to much larger mixedness and a lower transition probability.

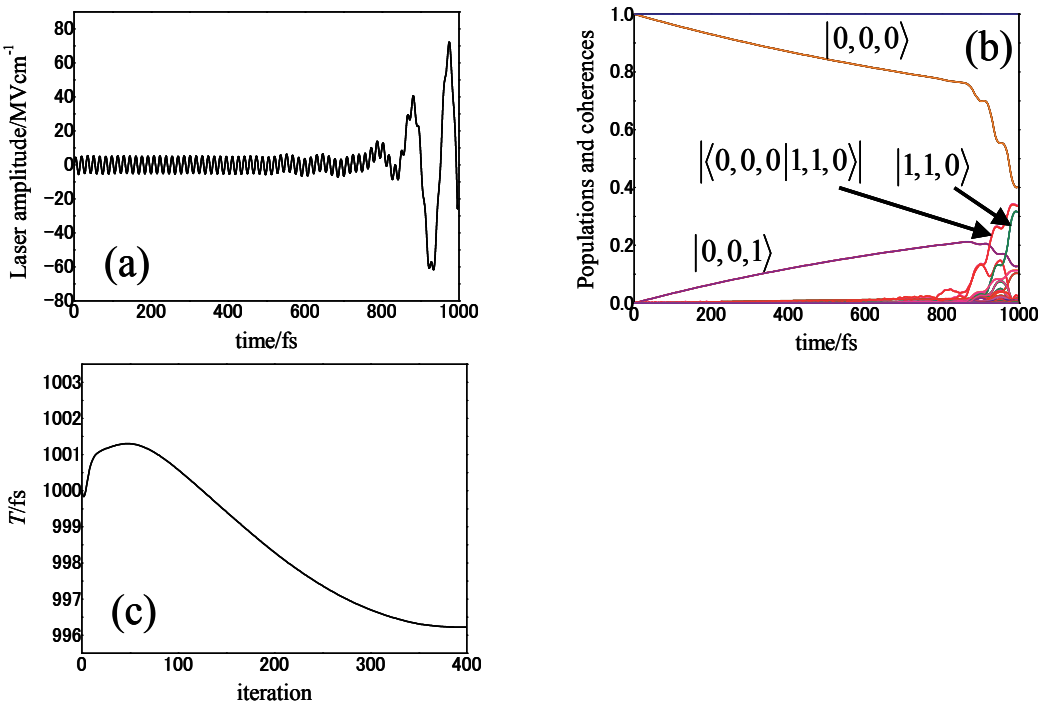


Fig. 4. (a) optimized laser pulse with α and β being equal to -1.755 x10^9 Wcm^{-2} and 5.851 x 10^2 fs^2, respectively, (b) population transfer induced by the optimized laser pulse of panel (a), (c) temporal duration of the optimized laser pulse versus iteration number. The initial $T^{(0)}$ was set to be 1000 fs. The temperature was 300 K. The target transition $|0,0,0\rangle \rightarrow \left(|0,0,0\rangle + |1,1,0\rangle\right)/\sqrt{2}$ was optimized.

When the initial temporal duration, $T^{(0)}$, is 1000 fs and the temperature is 300 K, we observe that the temporal duration becomes a little bit longer, T = 1040.56 fs, as can be seen in Fig. 5. The transition probability and the mixedness at the final time are 66.3430% and 0.50711 for the free-time case and are 65.8890% and 0.50515 for the fixed-time case, respectively. In both the free-time and fixed-time cases, the shape of the optimized laser pulses is interesting (here, we do not show the results for the fixed-time case). For the initial half time of total duration, the amplitude of the laser pulse is strong. In the middle of the temporal duration, it becomes weak. After that, the amplitude of the laser pulse becomes stronger with time. This tendency can be explained as follows. Because it is known that the population of the state $|0,0,0\rangle$ can be excited to the state $|0,0,1\rangle$ during the time evolution because of the

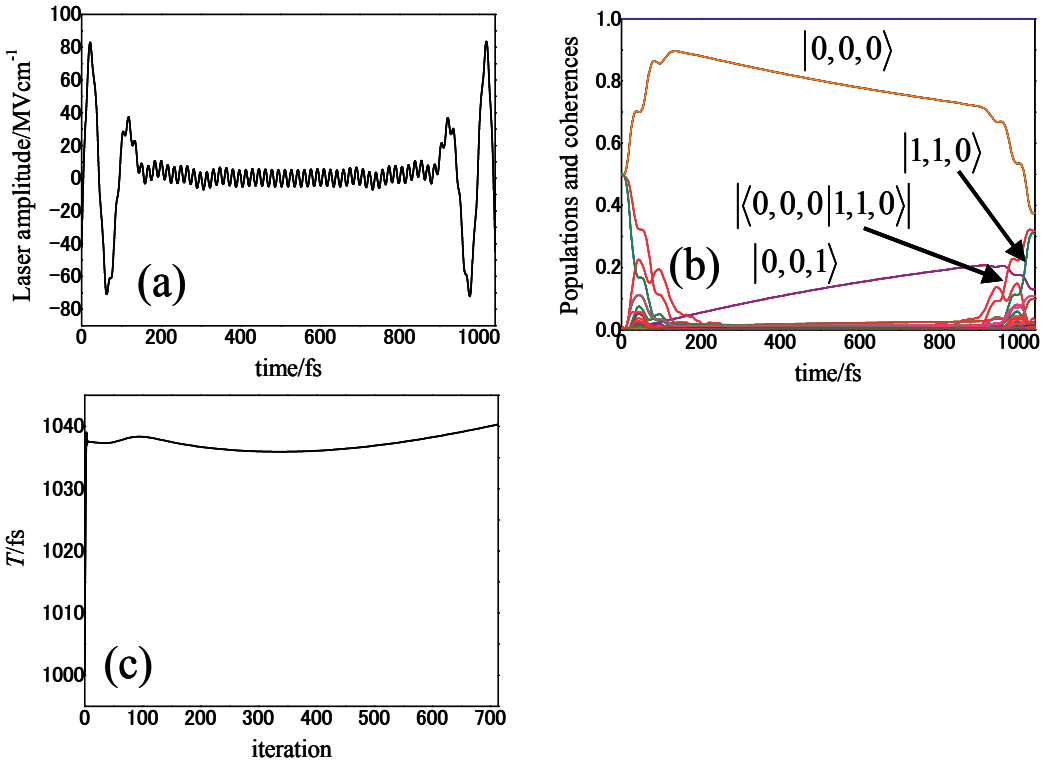


Fig. 5. (a) optimized laser pulse with α and β being equal to -1.755 x10^9 Wcm^{-2} and 5.851 x 10^2 fs^2, respectively, (b) population transfer induced by the optimized laser pulse of panel (a), and (c) temporal duration of the optimized laser pulse versus iteration number. The initial $T^{(0)}$ was set to be 1000 fs. The temperature was 300 K. The target transition $\left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2} \rightarrow \left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2}$ was optimized.

dissipative effect as mentioned above, the population of the state $|0,0,0\rangle$ has to increase for the initial half time of total duration using the large intensity of the laser pulse. During this period, almost all the population of the state $|1,1,0\rangle$ contributes to the population increase of the state $|0,0,0\rangle$. For the last half period of the total duration, because of the large intensity of the laser pulse, almost all the population of the state $|0,0,0\rangle$ is excited to the state $|1,1,0\rangle$, as in the cases shown above, and the optimized laser pulse tries to recover the initial maximally entangled state, $\left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2}$, as much as possible. The reason for the lengthening of the temporal duration compared with the initial guess is that the additional time duration required by the initial recovery of the state $|0,0,0\rangle$ was absent for the target transition $|0,0,0\rangle \rightarrow \left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2}$ shown in Fig. 4.

Figure 6 shows the case where $T^{(0)}$ is 1000 fs and the temperature is 10 K. Comparing panel (a) with panel (a) of Fig. 5, the pulse shapes are rather similar although the temperatures are quite different. However, because of their small difference, the optimized laser pulse in Fig. 6

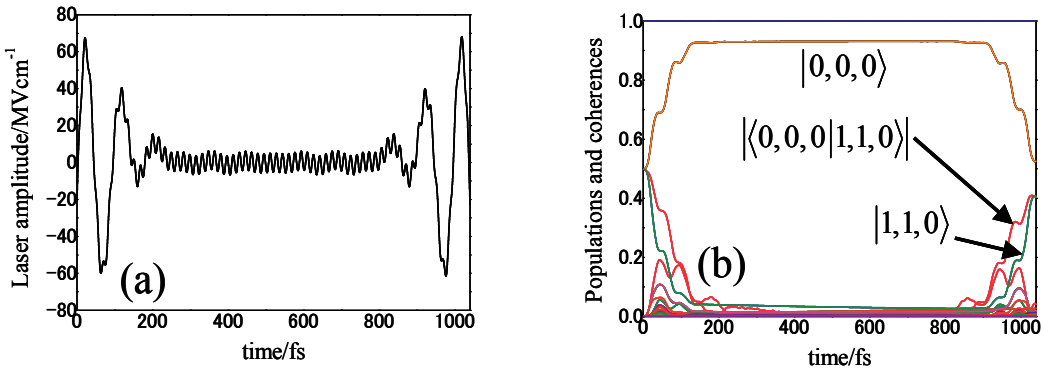


Fig. 6. (a) optimized laser pulse with α and β being equal to -1.755 x10^9 Wcm^{-2} and 2.340 x 10^1 fs^2, respectively, (b) population transfer induced by the optimized laser pulse of panel (a). The initial $T^{(0)}$ was set to be 1000 fs. The temperature was 10 K. The target transition $\left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2} \rightarrow \left(|0,0,0\rangle+|1,1,0\rangle\right)/\sqrt{2}$ was optimized.

creates the population of the state $|0,0,0\rangle$ as much as possible until around t = 200 fs. Unlike panel (b) of Fig. 5, that of Fig. 6 does not show any significant change of population of the state $|0,0,0\rangle$ during the period when the laser pulse is almost off (from around t = 200 fs to around t = 900 fs). This is also due to the small population transfer rate, $1/\Gamma_{(0,0,0)\rightarrow(0,0,1)} = 85300.0$ ps. Therefore, the transition probability is much larger and the mixedness is much smaller than in the case of Fig. 5. That is, the transition probability and the mixedness at the final time are 86.9429% and 0.22626 for the free-time case and are 86.7598% and 0.23058 for the fixed-time case, respectively. In addition, the optimal temporal duration is also longer than the initial guess: T = 1041.12 fs. The reason is the same as that for Fig. 5.

4. Concluding remarks

In the present chapter, we have reviewed our recent main theoretical and numerical contributions to the development of molecular quantum computing and quantum information science. In particular, we have paved a new way for extending the conventional FIFP-OCT to FRFP-OCT.

Now, quantum computing and quantum information science have become an unshakeable important research topics, ranging among a variety of disciplines. However, some basics of the theoretical aspects have not yet been solved and are still debatable. For instance, the definition of multipartite entanglement degree in pure and mixed states is still discussed in the recently published papers. In addition, scalability and decoherence of quantum states in quantum computers have gradually become obvious to be extremely challenging with the rapid development of experiments and theories. At the same time, the experimental realization of quantum computers based on the theories is also very important in order to extremely outperform the present-day classical computers. Although there are a number of experimental data for physical systems, at present there are few experimental evidences for molecules which chemists are interested in. Therefore, we suspect that there may be a

number of rooms for improvement in molecular quantum computers. We chemists hope that molecular quantum computing will be investigated in more detail from the chemical viewpoint in future. In particular, we expect that our and other's theoretical and numerical results will provide important guides to experimental realization of quantum computers and quantum information processing.

Although we have applied our FRFP-OCT to two specific control problems as shown in Section 3, the theory is so general that it may be possible to apply it to a variety of quantum control problems with and without dissipation in future. An experimental application of FRFP-OCTs developed by us for the first time could be expected in the same manner as closed-loop quantum learning control experiments[74 - 77].

Finally, as for the recent advancement of free-time and fixed end-point *multi-target* optimal control theory, the readers are referred to [78].

5. References

[1] M. A. Nielsen, I. Chuang, *Quantum Computation and Quantum Information*, Cambridge Univ. Press, Cambridge, UK, 2000.
[2] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. 70, 1895 (1993).
[3] C. Bennett and S. J. Wiesner, Phys. Rev. Lett. 69, 2881 (1992).
[4] M. Brune, F. Schnidt-Kaler, A. Maali, J. Dreyer, E. Hagley, J. Raimond, and S. Haroche, Phys. Rev. Lett. 76, 1800 (1996).
[5] D. L. Moehring, P. Maunz, S. Olmschenk, K. C. Younge, D. N. Matsukevich, L.-M. Duan, and C. Monroe, Nature, 449, 68 (2007).
[6] J. Benhelm, G. Kirchmair, C. F. Roos, and R. Blatt, Nature Phys. 4, 463 (2008).
[7] Friedenauer, H. Schmitz, J. T. Glueckert, D. Porras, and T. Schaetz, Nature Phys. 4, 757 (2008).
[8] G. K. Brennen, C. M. Caves, P. S. Jessen, and I. H. Deutsch, Phys. Rev. Lett. 82, 1060 (1999).
[9] I. L. Chuang, L. M. K. Vandersypen, X. Zhou, D. W. Leung, and S. Lloyd, Nature, 393, 143 (1998).
[10] J. A. Jones and M. Mosca, J. Chem. Phys. 109, 1648 (1998).
[11] J. Clarke and F. K. Wilhelm, Nature, 453, 1031 (2008).
[12] B. E. Kane, Nature, 393, 133 (1998).
[13] A. P. Nizovtsev, S. Ya. Killin, F. Jelezko, T. Gaebal, I. Popa, A. Gruber, J. Wrachtrup, Optics and Spectroscopy, 99, 248 (2004).
[14] P. Neumann, N. Mizouchi, F. Rempp, P. Hemmer, H. Watanabe, S. Yamasaki, V. Jacques, and T. Gaebel, Science, 320, 1326 (2008).
[15] M. Tesch, L. Kurtz, and R. de Vivie-Riedle, Chem. Phys. Lett. 343, 633 (2001).
[16] M. Tesch and R. de Vivie-Riedle, Phys. Rev. Lett. 89, 157901 (2002).
[17] M. Tesch and R. de Vivie-Riedle, J. Chem. Phys. 121, 12158 (2004).
[18] U. Troppmann, C. M. Tesch, and R. de Vivie-Riedle, Chem. Phys. Lett. 378, 273 (2003).
[19] B. M. R. Korff, U. Troppmann, K. L. Kompa, and R. de Vivie-Riedle. J. Chem. Phys. 123, 244509 (2005).
[20] U. Troppmann, C. Gollub, and R. de Vivie-Riedle, New J. Phys. 8, 100 (2006).
[21] C. Gollub and R. de Vivie-Riedle, J. Chem. Phys. 128, 167101 (2008).
[22] Babikov, J. Chem. Phys. 121, 7577 (2004).

[23] M. Zhao and D. Babikov, J. Chem. Phys. 125, 024105 (2006).
[24] M. Zhao and D. Babikov, J. Chem. Phys. 126, 204102 (2007).
[25] D. Babikov and M. Zhao, J. Chem. Phys. 128, 167102 (2008).
[26] T. Cheng and A. Brown, J. Chem. Phys. 124, 034111 (2006).
[27] M. Ndong, D. Lauvergnat, X. Chapuisat, and M. Desouter-Lecomte, J. Chem. Phys. 126, 244505 (2007).
[28] L. Bomble, D. Lauvergnat, F. Remacle, and M. Desouter-Lecomte, J. Chem. Phys. 128, 064110 (2008).
[29] S. Suzuki, K. Mishima, and K. Yamashita, 2005, Chem. Phys. Lett. 410, 358 (2005).
[30] M. Schröder and A. Brown, J. Chem. Phys. 131, 034101 (2009).
[31] See, for example, S. A. Rice and M. Zhao, *Optical Control of Molecular Dynamics*, Wiley-Interscience Publication, John Wiley & Sons, Inc., 2000.
[32] Y. Teranishi, Y. Ohtsuki, K. Hosaka, H. Chiba, H. Katsuki, and K. Ohmori, J. Chem. Phys. 124, 114110 (2006).
[33] A. Shapiro, I. Khavkine, M. Spanner, and M. Yu. Ivanov, Phys. Rev. A 67, 013406 (2003).
[34] M. Tsubouchi and T. Momose, Phys. Rev. A 77, 052326 (2008).
[35] E. Charron, P. Milman, A. Keller, and O. Atabek, Phys. Rev. A 75, 033414 (2007).
[36] K. Mishima and K. Yamashita, Chem. Phys. 361, 106 (2009).
[37] J. Vala, Z. Amitay, B. Zhang, S. R. Leone, and R. Kosloff, Phys. Rev. A 66, 062316 (2002).
[38] M. Tsubouchi and T. Momose, J. Opt. Soc. Am. B 24, 1886 (2007).
[39] Katsuki, H. Chiba, B. Girard, C. Meier, and K. Ohmori, Science 311, 1589 (2006).
[40] K. Ohmori, H. Katsuki, H. Chiba, M. Honda, Y. Hagihara, K. Fujiwara, Y. Sato, and K. Ueda, Phys. Rev. Lett. 96, 093002 (2006).
[41] Katuski, K. Hosaka, H. Chiba, and K. Ohmori, Phys. Rev. A 76, 013403 (2007).
[42] Katsuki, H. Chiba, C. Meier, B. Girard, and K. Ohmori, Phys. Rev. Lett. 102, 103602 (2009).
[43] Ohmori, Annu. Rev. Phys. Chem. 60, 487 (2009).
[44] Mishima, K. Shioya, and K. Yamashita, Chem. Phys. Lett. 442, 58 (2007).
[45] D. Landau, Phys. Z. Sowj. 2, 46 (1932).
[46] C. Zener, Proc. R. Soc. Lond. A 137, 696 (1932).
[47] V. S. Malinovsky and J. L. Krause, Eur. Phys. J. D 14, 147 (2001).
[48] D. P. DiVincenzo, Phys. Rev. A 51, 1015 (1995).
[49] D. Deutsch, Proc. R. Soc. Lond. A 400, 97 (1985).
[50] D. Deutsch and R. Jozsa, Proc. R. Soc. Lond. A 439, 553 (1992).
[51] P. Shor, *Proceedings 35th Annual symposium on Foundations of Computer Science*, 124. (1994).
[52] W. Zhu, J. Botina, and H. Rabitz, J. Chem. Phys. 108, 1953 (1998).
[53] K. Sundermann and R. de Vivie-Riedle, J. Chem. Phys. 110, 1896 (1999).
[54] Y. Ohtsuki, K. Nakagami, Y. Fujimura, W. Zhu, and H. Rabitz, J. Chem. Phys. 114, 8867 (2001).
[55] K. Mishima and K. Yamashita, Chem. Phys. 342, 141 (2007).
[56] K. Mishima and K. Yamashita, Chem. Phys. 352, 281 (2008).
[57] Shapiro and P. Brumer, *Principles of the Quantum Control of Molecular Processes*, Wiley-Interscience, Hoboken, NJ, 2003.
[58] K. Mishima and K. Yamashita, J. Chem. Phys. 130, 034108 (2009).
[59] K. Mishima and K. Yamashita, J. Chem. Phys. 131, 014109 (2009).

[60] W. G. Unruh, Phys. Rev. A 51, 992 (1995).
[61] G. M. Palma, K. –A. Suominen, and A. K. Ekert, Proc. R. Soc. London, Ser. A 452, 567 (1996).
[62] P. W. Shor, Phys. Rev. A 52, R2493 (1995).
[63] A. R. Calderbank and P. W. Shor, Phys. Rev. A 54, 1098 (1996).
[64] L. Viola and S. Lloyd, Phys. Rev. A 58, 2733 (1998).
[65] J. Cao, M. Messina, and K. R. Wilson, J. Chem. Phys. 106, 5239 (1997).
[66] J. Cheng, Z. Shen, and Y. Yan, J. Chem. Phys. 109, 1654 (1998).
[67] Y. Ohtsuki, W. Zhu, and H. Rabitz, J. Chem. Phys. 110, 9825 (1999).
[68] Y. Ohtsuki, K. Nakagami, W. Zhu, and H. Rabitz, Chem. Phys. 287, 197 (2003).
[69] S. Ramakrishna and T. Seideman, J. Chem. Phys. 124, 034101 (2006).
[70] A. Pelzer, S. Ramakrishna, and T. Seideman, J. Chem. Phys. 129, 134301 (2008).
[71] C. J. Bardeen, J. Che, K. R. Wilson, V. V. Yakovlev, V. A. Apkarian, C. C. Martens, R. Zadoyan, B. Kohler, and M. Messina, J. Chem. Phys. 106, 8486 (1997).
[72] S. Beyvers, Y. Ohtsuki, and P. Saalfrank, J. Chem. Phys. 124, 234706 (2006).
[73] C. Cattarius and H.-D. Meyer, J. Chem. Phys. 121, 9283 (2004).
[74] D. Cardoza, C. Trallero-Herrero, F. Langhojer, H. Rabitz, and T. Weinacht, J. Chem. Phys. 122, 124306 (2005).
[75] P. Gross, D. Neuhauser, and H. Rabitz, J. Chem. Phys. 98, 4557 (1993).
[76] RS. Judson and H. Rabitz, Phys. Rev. Lett. 68, 1500 (1992).
[77] R. J. Levis and H. A. Rabitz, J. Phys. Chem. A 106, 6427 (2002).
[78] K. Mishima and K. Yamashita, Chem. Phys. 379, 13 (2011).

14

Recent Applications of Hybrid *Ab Initio* Quantum Mechanics – Molecular Mechanics Simulations to Biological Macromolecules

Jiyoung Kang and Masaru Tateno
University of Tsukuba, University of Hyogo
Japan

1. Introduction

To understand the functional mechanisms of biological macromolecular systems, investigations of both the three-dimensional geometric and electronic structures are important. Accordingly, quantum mechanical (QM) calculations are essential to obtain the properties relevant to the electronic structures of the macromolecular systems. However, because of the high computational costs, only a few hundred atoms can be included in the usual *ab initio* QM calculations. Therefore, even though biological macromolecular systems include a huge number of atoms, isolated QM model calculations have been performed to date, in which only the reaction centres are considered. However, this may lead to serious misunderstandings of the properties relevant to the electronic structures of the macromolecular systems.

To overcome these problems, the *ab initio* QM calculation is combined with a molecular mechanics (MM) calculation method; this is referred to as a QM/MM calculation. In this strategy, the QM calculation is adopted for the active site (QM regions), and for the remainder of the system, the MM calculation is adopted (MM regions). This QM/MM methodology was originally developed by Warshel and Levitt (Warshel&Levitt, 1976), and since their pioneering work, great progress has been made in the development of QM/MM algorithms and their applications to biological systems (Bruice&Kahn, 2000; Field, 2002; Field et al., 1990; Gao et al., 2006; Lin&Truhlar, 2007; Mulholland, 2005; Ryde, 2003; Senn&Thiel, 2007, 2009).

In this chapter, we discuss some useful QM/MM schemes and their recent applications to analyze the functional mechanisms of biological macromolecular systems. Here, we will utilize three molecular systems as examples of applications of QM/MM calculations to biological macromolecules. We will also introduce the theoretical background to investigate the functional mechanisms in proteins, such as electron transfer.

2. Energy schemes of QM/MM calculations

In the QM/MM calculation, the entire system is separated into the QM and MM regions, as shown in figure 1. Since the QM and MM atoms interact with each other, the total energy can be written as follows:

$$E(\mathrm{ES}) = E(\mathrm{QM}) + E(\mathrm{MM}) + E(\mathrm{QM/MM}). \quad (1)$$

Here, E (ES), E (QM), E (MM), and E (QM/MM) represent the energy of the entire system, the energy of the QM region, the energy of the MM region, and the interaction energy between the QM and MM regions, respectively. In this section, we discuss how to merge the energies that were evaluated by each of the QM and MM calculation programs.

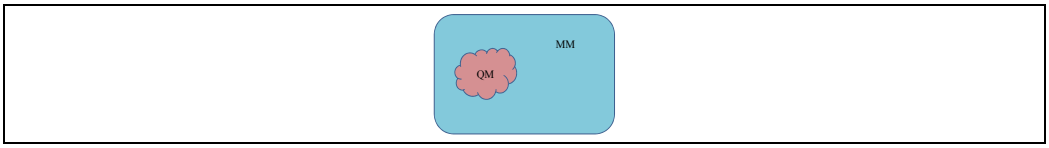


Fig. 1. Division of the entire system (ES) into the QM (pink) and MM (light blue) regions.

2.1 Additive scheme

The additive scheme can integrate the QM and MM interactions, and this is referred to as a hybrid QM/MM calculation. The energy of the entire system in the additive scheme, E_{additive} (ES), is given by the following equation:

$$E_{\mathrm{additive}}(\mathrm{ES}) = E^{\mathrm{QM}}(\mathrm{QM}) + E^{\mathrm{MM}}(\mathrm{MM}) + E^{\mathrm{QM/MM}}(\mathrm{QM, MM}). \quad (2)$$

The superscript of E represents the evaluation method of the system. One can define the QM/MM interaction energy, $E^{\mathrm{QM/MM}}$ (QM, MM), as

$$\begin{aligned} &E^{\mathrm{QM/MM}}(\mathrm{QM, MM}) \\ &= E^{\mathrm{QM/MM}}_{\mathrm{elec}}(\mathrm{QM, MM}) + E^{\mathrm{QM/MM}}_{\mathrm{bond}}(\mathrm{QM, MM}) + E^{\mathrm{QM/MM}}_{\mathrm{vdW}}(\mathrm{QM, MM}) \end{aligned}. \quad (3)$$

Here, the subscript of E represents the type of interaction. For example, elec represents an electronic interaction, bond represents a bond interaction, and vdW represents a van der Waals interaction.

One can calculate the electronic interaction in the QM Hamiltonian with an MM partial charge as follows:

$$\hat{H}^{QM/MM}_{\mathrm{elec}}(\mathrm{QM, MM}) = \sum_{i}^{N}\sum_{\alpha}^{M}\frac{q_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \sum_{\beta}^{L}\sum_{\alpha}^{M}\frac{Q_\beta q_\alpha}{|\mathbf{R}_\beta - \mathbf{R}_\alpha|}. \quad (4)$$

The first term on the right hand side of eq. (4) represents the coulomb interactions between the electron density at the $\mathbf{r}_i$ and the effective partial charges (q_α) of MM atoms at their position $\mathbf{R}_\alpha$, and the second term represents the coulomb interactions between the effective partial charges (q_α) of MM atoms at their positions $\mathbf{R}_\alpha$ and the QM nuclear charges (Q_β) at their positions $\mathbf{R}_\beta$ (figure 2). Therefore, the effects of the partial charges of MM atoms are considered as an external electrostatic field (i.e., one-electron integral term), and this can polarize the electronic structure of the QM region. The consideration of the effects of the environments (MM region) in the QM calculation is a merit of the additive scheme, and is impossible in the subtractive scheme (described in the next subsection). This advantage is clearer in some polarizable systems, such as a transition metal binding protein contained in the system.

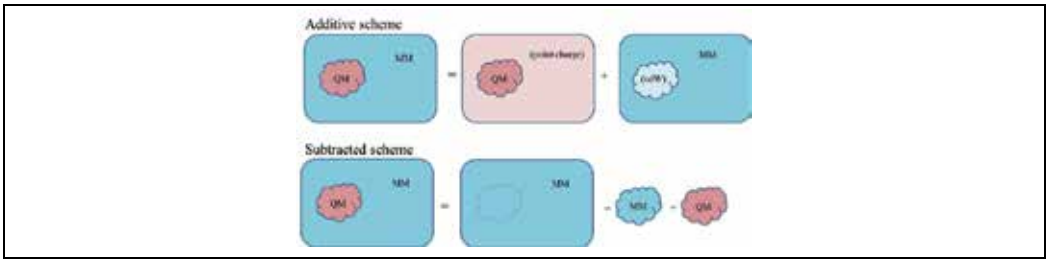


Fig. 2. Energy calculation by additive and subtractive schemes.

2.2 Subtractive scheme

In the subtractive scheme, the energy of the entire system is calculated by three steps. First, the energy of the entire system is calculated by the MM calculation. Second, the energy of the QM region is calculated by the QM calculation. Finally, to correct the double counted energy, the energy of the QM region is calculated by the MM calculation. From these steps, the energy of the entire system, E_{sub} (ES), is given in the following form:

$$E_{\mathrm{subtractive}}(\mathrm{ES}) = E^{MM}(\mathrm{ES}) + E^{\mathrm{QM}}(\mathrm{QM}) - E^{\mathrm{MM}}(\mathrm{QM}). \tag{5}$$

The superscript of E represents the evaluation method (QM or MM) of the system, and the region on which the calculation is performed is described in parentheses. An advantage of the subtractive scheme is the simplicity of the energy calculation (figure 2).
Table 1 shows the advantages and disadvantages of the two calculation schemes. The overpolarization problem is discussed in the next section.

	Additive scheme	Subtractive scheme
advantage	Polarizability in QM Hamiltonian	Simplicity
disadvantage	Overpolarization	Exclusion of polarizability in QM Hamiltonian.

Table 1. Comparison of additive and subtractive schemes

2.3 Treatment of a covalent bond located on the boundary between the QM and MM regions

Note that a problem occurs in the case where a covalent bond crosses the boundary between the QM and MM regions. In most biological systems, the active site contains several amino acid residues, which are connected to other amino acid residues of the protein. Therefore, termination of the QM region should be required (for example, the side chain of an amino acid residue is assigned to the QM region, and terminates at the C_α). Since the QM calculation system should be self-consistent with the QM Hamiltonian, several methods have been developed to control the boundary region between the QM and MM regions. Among these methods, the link atom method is generally used (Lin&Truhlar, 2005). In this method, the boundary atoms between the QM and MM regions are capped by a hydrogen atom, which is called a link atom, along the bond direction between the QM and MM atoms at a distance of ~1 Å. These new atoms are considered in the QM calculations. Although an advantage of this method is its simplicity, overpolarization problems occur, since this

method introduces artificial atoms that are located too close to the boundary MM atoms (Lin&Truhlar, 2005; Senn&Thiel, 2009).
This problem is more serious when one uses a large basis set that can describe an electronic structure more flexibly, such as plane waves (Senn&Thiel, 2009). Therefore, to overcome the overpolarization issue, several schemes have been developed. In one method (Field, 2002; Field et al., 1990), the charges of the MM atoms were not computed in the one-electron integrals. In another (Reuter et al., 2000), the charges of the MM atoms were not included in the one-electron integrals. In yet another strategy (Amara&Field, 2003; Eichinger et al., 1999), the Gaussian functions at the positions of the MM atoms that are located close to the QM region were introduced.
To resolve the disadvantages of the link atom method, the Generalized Hybrid Orbitals (GHO) method was developed. GHO is a major method to overcome the disadvantages of the link atom method, and originated from the Local Self-Consistent Field (LSCF) method (Amara et al., 2000; Assfeld&Rivail, 1996; Ferre et al., 2002; Gao et al., 1998; Monard et al., 1996; Pu et al., 2004; Thery et al., 1994). In this scheme, the "auxiliary" and "active" orbitals are located on the MM boundary atoms, respectively. By incorporating the active orbital, which is directed towards the QM boundary atoms, within the SCF calculation, the valency of the QM subsystem is satisfied, while the auxiliary orbitals are fixed during the SCF calculation.

2.4 Software for QM/MM calculations

In this section, we introduce several QM/MM calculation software packages. Various QM calculation packages, such as Gaussian (Frisch et al., 2003), include the MM module in the QM/MM calculation. In addition, several MM calculation packages, such as AMBER (Case et al., 2005) and CHARMM (Woodcock et al., 2007), can calculate QM/MM schemes by using the included QM modules. The merit of these calculation program packages is simple, and only one calculation job is needed to perform the QM/MM calculation. However, such calculation packages have some restrictions. For examples, in the Gaussian, it is impossible to perform high-level statistical calculation schemes, such as umbrella sampling. On the other hand, AMBER can only perform semi-empirical QM calculations, but not the first principle QM calculations.
To avoid such problems, several interface programs were developed, including ChemShell (Sherwood et al., 2003), QoMMM (Harvey, 2004), and PUPIL (Torras et al., 2008). With these types of methodologies, one can perform QM/MM calculations effectively and couple them to statistical sampling schemes. For example, extended conformational sampling schemes are also available by using molecular dynamics simulations implemented in MM packages, and algorithms to find optimal reaction paths and stationary points on the potential energy surface (PES) are also available, by using the program modules implemented in the QM packages.
Recently, in our laboratory, a new QM/MM interface program was developed by connecting the AMBER (Case et al., 2005) and GAMESS (Schmidt et al., 1993) packages (Kang et al., 2009). Thereby, one can couple various levels of QM calculations, such as Hartree-Fock, density functional theory (DFT), and more advanced schemes, with conformational sampling and free energy estimation techniques, such as a standard molecular dynamics (MD) calculation, replica exchange MD, and potential mean force (PMF).

The roles of the interface program are i) to control the sequence of steps for the QM/MM optimization and the MD simulation, and ii) to exchange information between the two programs. At the first stage of a calculation cycle, a single-point calculation is performed in GAMESS concerning the coordinate set of a system. This yields two types of forces: i) the forces on the QM atoms from the QM–QM and QM–MM interactions, and ii) the forces on the MM atoms derived from the QM–MM interactions. To calculate the second force, the contribution from the QM atoms to the MM atoms is calculated, by integrating the interaction between the partial charges of the MM atoms and the electron density at each grid point of the quantum region, as shown in equation (6).

$$F_j^{\mathrm{QM}} = \sum_i^N \mathbf{r}_{ij} \frac{q_j}{\left|r_{ij}\right|^3} dq_i. \tag{6}$$

Here, q_j is the partial charge of an MM atom, and N is the total number of electron density points. Here, dq_i designates a small volume on the electron density grid, $dq_i = \rho_i dx dy dz$. With respect to the coordinate set, AMBER calculates (1) forces from van der Waals (vdW) interactions between all atoms of the system, and (2) forces on MM atoms from MM–MM interactions. The interface program combines the calculated forces. This step requires the mapping of the relationship between the QM and MM atoms, since the atom labeling in one program package may differ from that in the other one. In order to reduce the overhead cost of this process (in particular, the required computational time would increase when treating solvated biological macromolecules; i.e., the number of atoms included in the calculation system could exceed 100,000 atoms), we employ a combination of a UNIX shell and C programming.

For combining the QM and MM calculations, we should mention another use of (advanced) QM calculations (such as Coupled Cluster); i.e., the development of effective potentials for MM calculations. A recently developed scheme can be applied to a wide range of interactions. For example, the description of cation-π interactions, such as the complex of Na^+ and an aromatic ring, was difficult to achieve using conventional MM potentials. The new scheme mentioned above enabled us to perform accurate MD simulations involving a cation-π interaction that is present in the catalytic site of an enzyme (Hagiwara et al., 2009a; Hagiwara et al., 2011).

3. Application 1: A copper binding protein, azurin

3.1 Azurin

Biological systems possess various types of metalloenzymes, which are involved in biological functions such as electron transfer, metal storage, dioxygen binding, substrate turnover, and protein structure configuration (Holm et al., 1996). An important application of QM/MM methods for metalloenzymes is to investigate the electronic structures of the active sites in blue copper proteins (Paraskevopoulos et al., 2006; Ryde&Olsson, 2001; Ryde et al., 2000; Sinnecker&Neese, 2006; Solomon, 2006), which function in electron transfer (ET) through the bound Cu ion(s) (Gray et al., 2000). X-ray crystallographic analyses have been reported for such enzymes, including azurin, plastocyanin, and stellacyanin. The common features of the active site are one-cysteine coordination and two-histidine coordination to the copper centres, forming an approximate trigonal plane, and their coordination

environments are completed with weak axial ligand(s) (Gray et al., 2000; Paraskevopoulos et al., 2006).
Azurin includes a five-coordinated copper, and possesses methionine (Met) and the backbone carbonyl oxygen of glycine (Gly) as weak axial ligands (figure 3). Stellacyanin and plastocyanin have a four-coordinated copper site coordinated with glutamine (Gln) and Met as weak axial ligands, respectively. These weak coordinative bonds are presumed to contribute critically to the redox potential related to electron transfer, thereby regulating the biological functions. Thus, blue copper proteins have been used as standard metalloenzymes to examine the accuracy of calculations (Ryde&Olsson, 2001; Ryde et al., 2000; Solomon, 2006; Solomon et al., 2004).

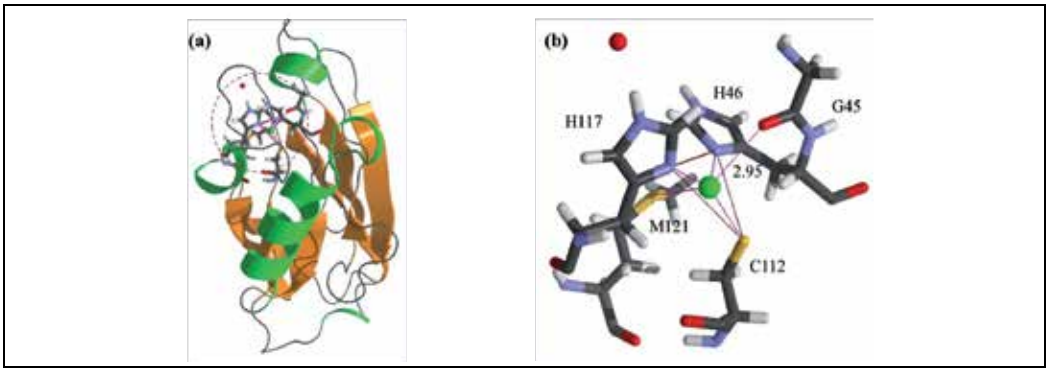


Fig. 3. The three-dimensional structures of azurin and its Cu-binding site.

3.2 Evaluation of QM/MM calculations: Spin polarization

The above-mentioned QM/MM schemes, the additive (termed simulation A here) and subtractive (simulation B) schemes, were adopted by exploiting our interface program that connects QM and MM calculation engines implemented in parallel supercomputers (see section 2.4) (Hagiwara et al., 2009d; Kang et al., 2009). An advantage of this strategy is that the interface program enables us to compare the results of QM/MM calculations using distinct schemes; thus, the long-range effects on the copper active site were evaluated in this theoretical analysis. Here, an *ab inito* DFT calculation was utilized as the QM part in the following QM/MM calculations.
The QM/MM calculations were evaluated by a comparison to the experimental data, as follows. The singly occupied molecular orbital (SOMO) of the copper binding site of blue copper proteins is known as an antibonding feature, as a combination of a Cu 3*d*-orbital and a cysteine sulphur 3*p*-orbital (Solomon, 2006; Solomon et al., 2004; Sugiyama et al., 2005). In simulations A and B, the 128 α- and 127 β-spin orbitals are occupied, and the SOMOs of simulations A and B were found in the 126 α-spin orbitals, as shown in figure 4a and b, respectively. The Cu 3*d*-orbital and the cysteine sulphur 3*p*-orbital contribute to the SOMO in both simulations A and B, while some differences in the orbital patterns are seen at the Gly45 and His117 residues (figure 4).
The electron density of the S atom in Cys112 is significantly delocalized onto the copper *d*-orbital (this was confirmed by the calculation of the Mulliken charge of the S atom) (Kang et al., 2009), resulting in a large spin density on the S atom. In simulation A, the spin polarization is significantly improved (49.9%), as compared to simulation B (56.8%), since

the former calculation result (simulation A) exhibits better agreement with the spectroscopic experiment; i.e., 45% S 3*p*-orbital character was experimentally observed in the SOMO of azurin (George et al., 2003; Solomon et al., 2004). This result indicates that the explicit inclusion of an electrostatic interaction by the QM Hamiltonian is essential for the accurate evaluation of the copper site.

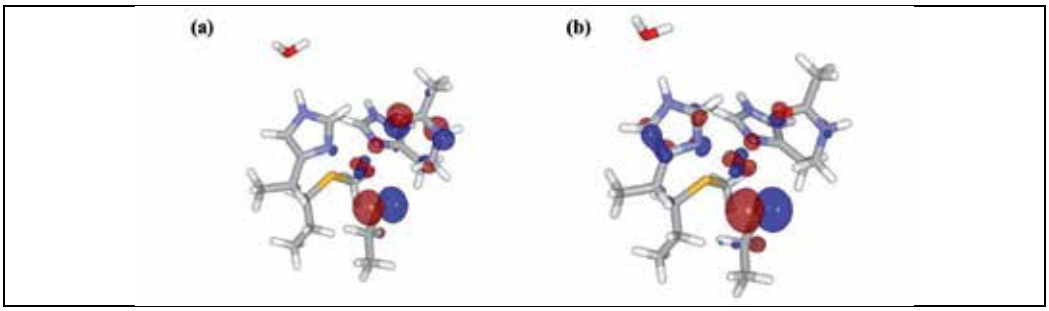


Fig. 4. SOMOs of simulations A (panel a) and B (panel b).

3.3 Evaluation of the QM/MM calculation: Three-dimensional geometrical structure

Next, with respect to the three-dimensional geometrical structure, the atomic distances relevant to the Cu coordination are shown in Table 2. It was pointed out that the intense X-ray beam reduced the blue copper site during data collection, and thus ambiguities still remain in the experimental data. Accordingly, computational investigations are crucial to elucidate the outstanding electronic and geometrical features of the weak axial Cu–S(Met121) and Cu–O(Gly45) ligands and the relatively strong Cu–S(Cys112) ligand observed in azurin.

For the Cu–S(Met121) coordination, a comparison of simulations A and B shows almost the same Cu–S(Met121) distances of ~3.5 Å, indicating that in the QM/MM energy schemes, the weak Cu–S(Met121) coordination is somewhat insensitive to the polarization effect. Note that for the Cu–O(Gly45) distance (2.75–3.16 Å in the crystal structures), Hasnain and co-workers obtained the remarkably short distances of 2.55 and 2.49 Å, using the ONIOM module in the Gaussian package as a QM/MM program, in the presence and absence of the electrostatic energy term in the QM Hamiltonian, respectively (Paraskevopoulos et al., 2006). Thus, the results of the above-mentioned calculations showed much better agreement with the experimental data (Table 2).

Distance	Xtal[a]	Simulation A	Simulation B	ONIOM[b]
Cu-O(Gly45)	2.95	3.01	2.81	2.55 (2.49)
Cu-S(Met121)	3.16	3.49	3.50	3.53 (3.41)
Cu-S(Cys112)	2.27	2.20	2.24	2.17 (2.17)
Cu-N(His117)	1.98	2.00	2.10	2.01 (2.03)
Cu-N(His46)	2.06	2.03	1.93	1.99 (2.01)

a. Experimental data of the crystal structure (Nar et al., 1991)
b. EE (ME) optimized geometries (Paraskevopoulos et al., 2006)

Table 2. Comparison of the three-dimensional geometrical structures of the active site of azurin.

In addition, significantly different Cu–O(Gly45) distances, 3.0 and 2.8 Å, were found between the calculated structures for simulations A and B, respectively, indicating that this weak axial Cu–O(Gly45) coordination is sensitive to the treatment of the electrostatic interaction in the QM Hamiltonian. This is a distinctive feature between the Cu–S(Met121) and Cu–O(Gly45) bonds. Note here that the Cu–O(Gly45) bond is fairly well polarized, whereas the Cu–S(Met121) bond is just slightly polarized (this was confirmed by the calculation of the Mulliken charges) (Kang et al., 2009).

Next, with respect to the electronic structures of the active site of azurin, the molecular orbitals (MOs) that predominantly include the electrons involved in S(Met121)/O(Gly45) atoms were compared between simulations A and B. Here, note that O(Gly45) is included in a peptide group of the protein backbone, and that the electrons of the atom are delocalized onto the peptide group. Therefore, the contributions of the delocalized electrons should be summed up for the peptide groups of the protein backbone.

As a result of the analysis, for the Cu–S(Met121) coordination, the highest occupied molecular orbitals (HOMOs) of the α electrons, which are the 128 MOs in both models, were revealed to include the electrons of the S(Met121) atoms predominantly in both models, and they are equivalent to each other. The energy difference between the two HOMOs is as small as 1.8 kcal/mol, which can be considered to be marginal (figure 5a).

On the other hand, for the Cu–O(Gly45) coordination, the 96 orbital of the α electron in simulation A corresponds to the 97 orbital of the α electron in simulation B, which predominantly includes the electrons of the peptide group involving O(Gly45). The energy difference between the two MOs is 8.6 kcal/mol; this value is definitely larger than that of S(Met121) (figure 5b). Thus, it is likely that the energy level of the electrons of O(Gly45) is shifted through the electrostatic interactions in the QM Hamiltonian, indicating that this polarized coordinative bond is sensitive to the environment surrounding the copper active site.

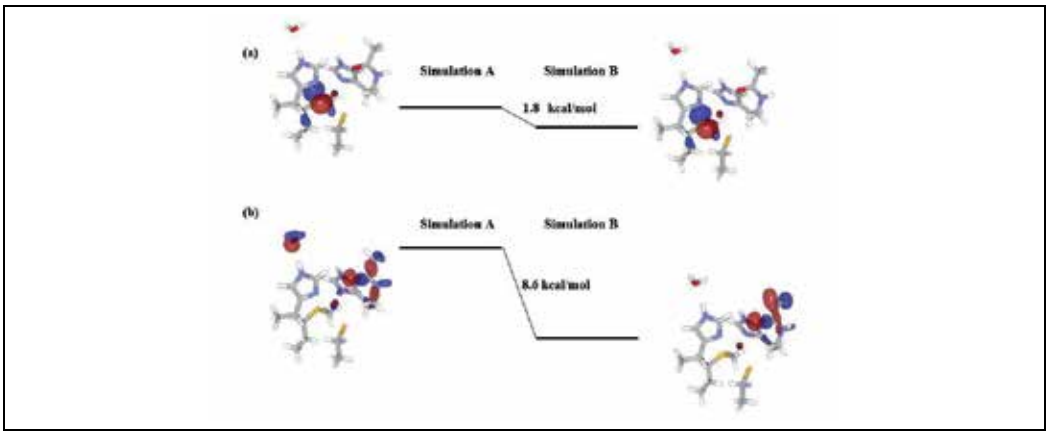


Fig. 5. Examples of the MOs and their energy levels obtained from those of LUMOs, which are common in both calculation results. Panels (a) and (b) represent HOMOs and 96 α MOs, respectively.

3.4 Effect of the long-range interaction

The results indicated that the treatment of the electrostatic interactions of the metal site, the surrounding protein moiety, and the solvent in the QM Hamiltonian is important for an

accurate description of the electronic structure and the geometry of the blue copper site in azurin. Thus, it is suggested that the electronic structures of active sites are actually modulated by the surrounding regions through long-range electrostatic interactions, resulting in contributions to the biological functions of metalloenzymes, such as electron transfer.
In addition, for the other issues, such as the evaluation of solvent effects on the electronic structures of biological macromolecular systems, the QM/MM calculation schemes are also useful. For example, with respect to a DNA-protein complex, the recognition mode between the DNA and PU.1 (a transcriptional factor protein) influences the solvent accessibility (accessible surface area) of the DNA bases, through masking by the protein. The QM/MM calculations revealed that the differences in the solvent accessibilities of the DNA bases are closely relevant to the MO energy levels of the bases, which may contribute to the regulation of the biological functions of the molecular systems, such as the reactivity of each base (Hagiwara et al., 2010b).

4. Application 2: A transition metal binding protein, cytochrome c oxidase

4.1 Cu_A site of cytochrome c oxidase

Cytochrome c oxidase (CcO), which is the terminal enzyme of the electron transport system, reduces an oxygen molecule to water, thereby generating the proton concentration gradient between the matrix and the intermembrane space of mitochondria or the periplasmic space of bacteria (proton pumping). The Cu_A site of CcO (figure 6) receives electrons from cytochrome c, thereby providing the electrons with heme *a*. This triggers various chemical reactions occurring from the reduced state in CcO, which are cooperatively induced, followed by oxygen binding (Kaila et al., 2010; Namslauer&Brzezinski, 2004; Papa et al., 2004). Thus, the Cu_A site acts as the "gate" of the subsequent cooperative reactions in CcO.
The Cu_A site possesses two Cu ions and six coordinated ligands (figure 6). These two Cu ions were found to form a covalent bond characterized by a mixed-valence state (DeBeer

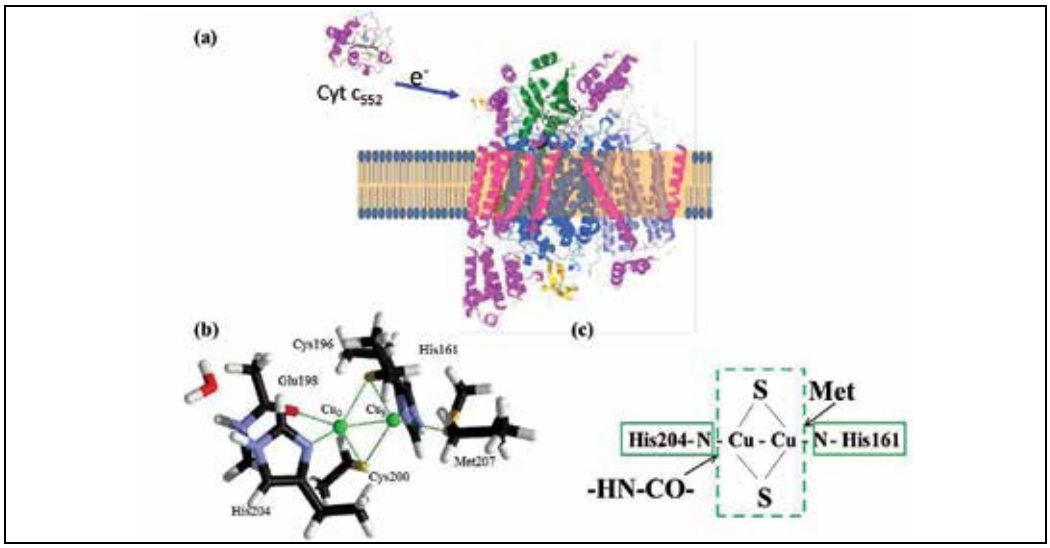


Fig. 6. Geometric structure of bovine CcO (a), and Cu_A site of CcO. Panel (c) is a schematic depiction of the Cu_A site.

George et al., 2001; Gamelin et al., 1998; Kroneck et al., 1988, 1989), which produces a lower reorganization energy and a higher electron transfer rate than those of the blue Cu proteins, such as plastocyanin and azurin (DeBeer George et al., 2001). In bovine CcO, the distances of Cu-S (Met207) and Cu-O (the Glu198 backbone) are slightly longer than the other covalent bonds, and thus the S and O atoms are referred to as the axial-coordinated atoms.

4.2 Controversial experimental data involving discrepancies

Each of the Cu-binding sites of plastocyanin and azurin also possesses an axial Met ligand, and the functional roles of the Cu-binding sites in the blue Cu proteins were considered to be similar to those of the Cu_A site in CcO (Covello&Gray, 1990; DeBeer George et al., 2001). In plastocyanin and azurin, this Met residue is important to control their reduction potentials (Berry et al., 2003; Garner et al., 2006). By analogy to the blue Cu proteins, Met207 of the bovine CcO may also be involved in regulating the electron transfer rate. In fact, this Met is highly conserved in various CcO proteins (Paumann et al., 2004).

However, several distinct experimental results were reported. For example, in the engineered azurin, in which the Cu_A site was incorporated to mimic that of CcO, the roles of the Met residue were analyzed by measuring the redox potentials of several mutants (Hwang et al., 2005; Robinson et al., 1999). This analysis revealed only slight changes in the redox potential of the mutants. In contrast, significant changes in the redox potential were observed in the Met mutants of the bacterial CcO enzymes. In fact, for the CcO from *Rhodobacter sphaeroides*, the replacement of the axial Met with Leu remarkably increased the redox potential, by 118 mV (Wang et al., 2002; Zhen et al., 2002). With respect to the *Thermus thermophilus* ba_3 oxidase, systematic mutagenesis studies of the Cu_A site revealed various ranges of redox potentials in the mutants (Ledesma et al., 2007).

In this manner, the experimental results relevant to the Cu_A site were apparently controversial and also involved some discrepancies. In addition, the intense X-ray beam used for the structural analysis reduces the metal binding site during the data collection, thereby delocalizing the electron density of the transition metal binding site. Thus, the X-ray structure of the Cu_A site may also include some ambiguities, which is a similar case to that of azurin, as discussed below. These led to some difficulties in the detailed and accurate elucidation of the functional roles of the axial ligands in the Cu_A site.

4.3 Electronic structure of Cu_A site

To clarify the functional roles of the amino acid residues in the Cu_A site, intensive theoretical analyses have been performed, and some were coupled to experiments, such as spectroscopic measurements. In a recent QM/MM analysis of the Cu_A site, experimental artifacts were found in its atomic coordinates. The conformation involving the artifacts was improved in the theoretical analysis, and then explicit solvent water molecules were arranged in the calculation model. The accurate descriptions of the electronic structure of the active site were intended to elucidate the functional roles of the axial ligands, where some experimental discrepancies were found, as mentioned above. As a consequence, an interesting feature was revealed in the electronic structure of the Cu_A site of CcO, which may be closely relevant to the roles of the axial residues (Kang et al., 2011). This finding is briefly discussed in this section.

In the Cu_A structure of bovine CcO, a steric clash exists between C_γ of Met207 and C_β of Cys196, which may be derived from the above-mentioned ambiguity of the

crystallographic analysis. Accordingly, for the structural improvement, extended sampling techniques coupled to QM/MM geometry optimization were adopted for the Cu_A site of bovine CcO. Thereby, the steric clash was completely removed without inducing any structural transitions, as compared with the crystallographic electron density distribution map of the Cu_A site. Thus, such computational methodologies are indispensable to minimize the experimental artefacts induced by the X-ray beam in transition metal binding proteins.

The following section includes some descriptions of the electronic structure obtained by the QM/MM calculations, and comparisons with experimental data: the SOMO of the oxidized state of the Cu_A site contains Cu and S (Cys) atomic orbitals (figure 7). In this hybrid QM/MM calculation of bovine CcO, the two Cu and S atoms of the Cu_A site exhibited 47 % and 45 % spin densities, respectively (Kang et al., 2011). On the other hand, to minimize the difficulties in the experimental analysis of the Cu_A site of CcO, Hay et al. incorporated the two Cu ions and the coordinated amino acid residues into azurin, to mimic the original Cu_A structure and the function of CcO (Hay et al., 1996). With respect to this engineered azurin, DeBeer George et al. performed Cu L-edge and S K-edge X-ray absorption spectroscopic (XAS) experiments. This analysis revealed that the *d*-orbitals of the Cu atom (44 %) and the *p*-orbitals of the S atom (46 %) predominantly occupy the SOMO of the Cu_A site (DeBeer George et al., 2001).

In this manner, the results of the hybrid QM/MM calculation of CcO are quite consistent with the experimental data. In previous studies, similar comparisons were performed with respect to the calculated and experimental data, using the engineered azurin for both analyses (Xie et al., 2008). The above-mentioned QM/MM calculation results are better than the previous ones, in comparison with the experimental data.

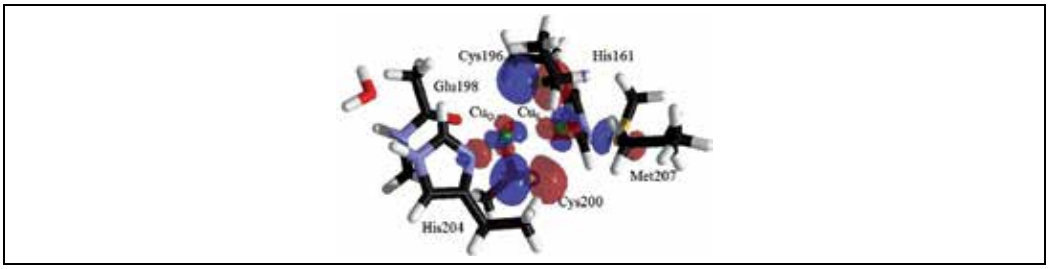


Fig. 7. SOMO of the Cu_A site obtained by hybrid QM/MM calculations of CcO.

Next, to evaluate the effects of the axial Met residue, a simple model was also calculated. This model was composed of only the atoms involved in the Cu_A site where *the Met residue is excluded*. The results were compared with the above-mentioned QM/MM calculations of the CuA site involving CcO in the calculation model. This analysis revealed that the axial Met ligand increases the energy of an MO that predominantly contains the Cu d_{zx} orbital, whereas the other MO energy levels were influenced to a much lower extent (please note that the local coordinate system, such as d_{zx}, depends on the calculation programs used in the analysis, and thus is not substantial) (Kang et al., 2011).

This elevation of the energy level of the Cu d_{zx} orbital can be understood by the hybridization of the atomic orbitals between the Cu ion and the S atom of the axial ligand, as shown in figure 8. Thus, from the view of the electronic structure of the Cu_A site, the role of the axial Met ligand may be the selective modulation of the energy level of a Cu *d* orbital.

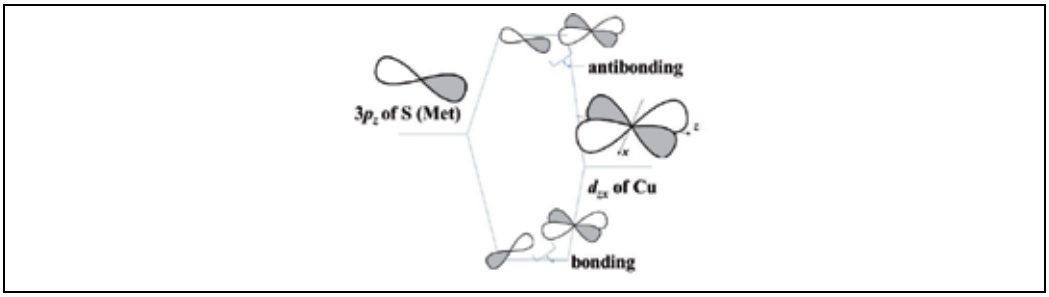


Fig. 8. Hybridization of the $3p_z$ of the S atom of the axial ligand Met residue and d_{xz} of the Cu ion.

4.4 Inner-sphere reorganization energy of Cu_A site

Furthermore, the inner-sphere reorganization energy, which is one of the factors to determine the electron transfer rate in the Marcus theory, was calculated within the framework of the QM/MM calculation, and compared with the experimental data (Kang et al., 2011).

The Marcus theory gives the relationship between the electron transfer rate and the reorganization energy, as follows (Marcus&Sutin, 1985; Moser et al., 1992):

$$k_{ET} = \sqrt{\frac{\pi}{\hbar^2 \lambda k_B T}} |H_{DA}|^2 \exp\left[-\frac{\left(\Delta G^0 + \lambda\right)^2}{4\lambda k_B T} \right]. \tag{7}$$

Here, $|H_{DA}|$ is the electronic coupling between the initial and final states, k_B is the Boltzmann constant, and $\hbar$ is the reduced Planck constant, which is defined as the Plank constant divided by 2π. The ΔG^0 represents the standard Gibbs free energy change, and the λ represents the reorganization energy, which is the sum of the inner-sphere and outer-sphere reorganization energies. In this report, the inner-sphere reorganization energy is evaluated by using the following equation, as mentioned previously (Gorelsky et al., 2006; Kelterer et al., 2001) :

$$\lambda_i = \left(E_{g=oxidized} - E_{g=reduced}\right)_{chg=reduced} + \left(E_{g=reduced} - E_{g=oxidized}\right)_{chg=oxidized}. \tag{8}$$

Here, E represents the energy of the QM region, involving the electrostatic interaction energy between the MM1 and QM regions. The subscript, chg, refers to the charge state of the system. The other subscript, g, refers to the geometry of the redox site, which is obtained by the geometric optimization of the hybrid QM/MM calculations. Thus, each E value should be obtained by QM/MM geometry optimization.

The following two calculation models were used to evaluate the axial Met residue in terms of the inner-sphere reorganization energy: A) QM-Met and B) MM-Met. Model A is the same as the QM/MM model that was described above. In model B, the axial Met residue was removed from the QM region (i.e., the Met residue was incorporated into the MM region), and the atomic partial charges of the Met reside were set to zero.

The computational (model A) and experimental values of the inner-sphere reorganization energy were consistent (i.e., 228 and 250 meV, respectively). Next, the comparison of the two

computational models revealed that the inner-sphere reorganization energy of model A is larger than that of model B, by as little as 44 meV. This indicates that the axial Met residue does not significantly perturb the electron transfer rate of the Cu_A site of bovine CcO. In fact, the mutant in which the Met residue is replaced with Leu exhibited a lower electron transfer rate than that of the wild type, by ten-fold (Wang et al., 2002), which seems to be quantitatively consistent with the above-mentioned calculation result. Therefore, it is concluded that the axial Met residue may be the fine-modulator of the electron transfer rate of the Cu_A site.

4.5 Use of techniques in information science

In this manner, the hybrid QM/MM calculations indicated that the axial Met ligand functions as a fine-modulator of the electronic structure (see 4.3) and the electron transfer rate (see 4.4) of the Cu_A site of bovine CcO. Still, one cannot solve the above-mentioned experimental discrepancy in the differences of the redox potentials of the Cu_A site, depending on the species. The X-ray structure of the axial Met residue of bovine CcO is shown in figure 9. The Met backbone is located in a short turn structure between an α-helix and a β-sheet that form a rigid conformation, and thereby, the residue side chain is also fixed together with the Cu-coordination of the S atom. To solve the experimental discrepancy, the structural changes of the Cu_A sites were examined with respect to the mutants of oxidases from various species, by combining amino acid multiple sequence alignment and computer modeling techniques.

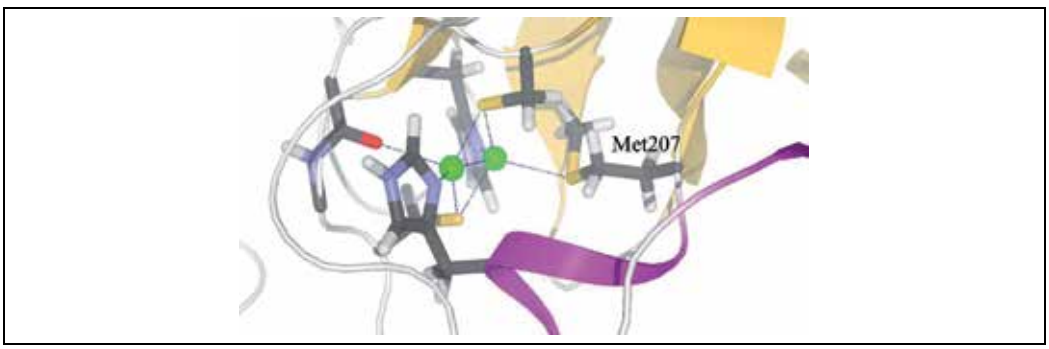


Fig. 9. The axial Met residue (Met207) in the crystal structure of bovine CcO.

The amino acid sequence alignment of the CcOs revealed that the *Rhodobacter* and bovine oxidases (large changes were observed in their redox potential values) possess inserted modular structures close to the Cu_A sites, whereas the *Thermus thermophilus* ba_3 oxidase and the engineered azurin (small changes in the redox potentials) do not. In fact, in the crystal structures of the *Rhodobacter* and bovine oxidases, the axial Met residues contact the inserted modular structural elements. In contrast, in the crystal structures of *Thermus thermophilus* ba_3 oxidase and the engineered azurin, the Met residues can apparently move more flexibly, and so the replacement of the Met residue may cause only small perturbations of the properties relevant to the Cu_A sites. Computer-assisted structural modeling of the various mutants confirmed these findings (Kang et al., 2011).

Thus, the use of information science techniques (bioinformatics) is also crucial to perform thereotical analyses employing *ab inito* QM/MM electronic structure calculations of

biological macromolecular systems. The development of methodologies in both fields is required, and the methodologies should be unified to analyze biological systems.

4.6 Double-stratified architecture of the Cu_A site

On the other hand, with respect to the Cu-coordinated Cys/His residues, their replacement influenced the geometric and electronic structures of the Cu_A site much more drastically (Hwang et al., 2006; Zhen et al., 2002; Zickermann et al., 1997), resulting in structural disruptions and significant increases in the redox potential (Zhen et al., 2002). For instance, the replacement of each cysteine residue (the mixed valence state) with serine in the engineered azurin (C112S and C116S) induced the different electronic states (i.e., the C112S mutation induced two distinct Cu centres, each of which is type II, while the C116S mutation induced one Cu centre, which is type I or II depending on pH) (Hwang et al., 2006). Thus, these Cys/His residues play a critical role in establishing the geometric and electronic structures of the Cu_A site.

In this manner, the functional roles of the amino acid residues of the Cu_A site have been revealed. The two Cys/His residues build the fundamental geometric and electronic structures of the Cu_A site; together, these are considered as the "platform" of the Cu_A structure, since the replacements of those residues exert critical effects on the Cu_A system (figure 10). In contrast, the axial Met ligand functions in the fine modulation of the electronic structure of the "platform"; thereby, this residue may regulate the electron transfer rates from Cyt *c* to heme *a*, thus inducing the various subsequent cooperative reactions occurring in CcO (figure 10) (Kang et al., 2011).

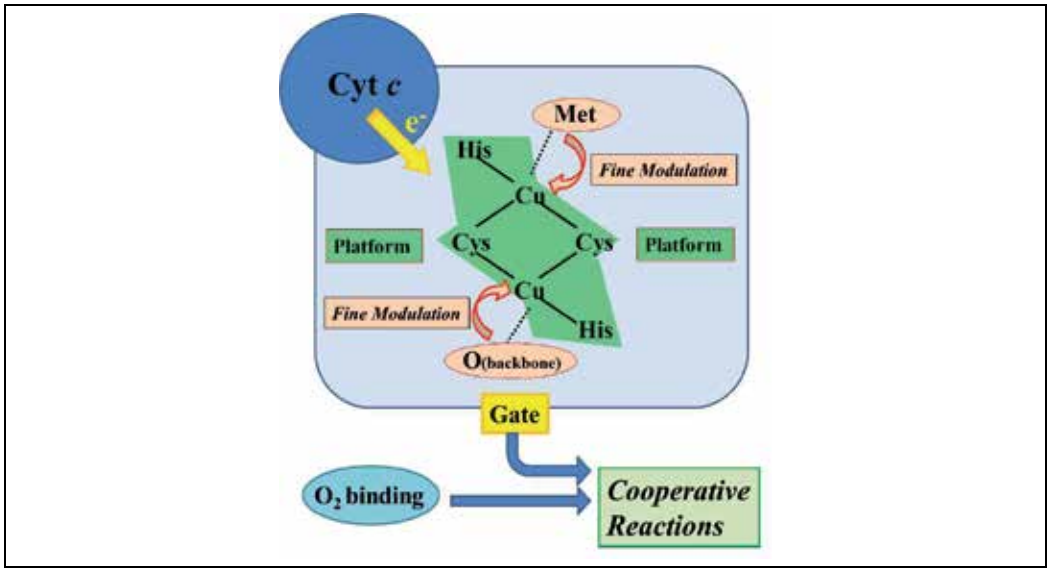


Fig. 10. The double-stratified architecture of the Cu_A site, as proposed in this chapter.

This means that the Cu_A site forms the double-stratified architecture, comprising the platform (i.e., Cu_2, Cys S atoms, His N atoms) and its fine-modulators (Met S and O atoms). This is not trivial. In fact, the weakly coordinated axial Met ligand in the active site of native (wild type) azurin seems to act as a member of the "platform" residues (see 4.1). In addition,

the critical roles of the axial Met residues in the Cu_A sites had been proposed for the oxidases and the engineered azurin, as mentioned in section 4.2; this suggests that the Met ligand is a member of the "platform" residues. However, it has been concluded that the Met ligand is not a "platform" residue, but a "fine-modulator", on the basis of the analysis using *ab initio* QM/MM electronic structure calculation and computer modeling techniques coupled to multiple sequence alignment.

This type of hierarchical structure relevant to the function may be widely utilized in biological systems. For example, some protein structures seem to be established by a similar architecture; i.e., fundamental folds are common in various proteins, and the addition of specific inserted modular elements into such fundamental folds leads to new biological functions of the proteins in molecular evolution. Such multi-stratified architecture and molecular evolution can be found in various biological systems, as observed in the Cu_A sites of various oxidases.

5. Application 3: Novel catalytic reaction mechanism revealed by QM/MM molecular dynamics simulation

5.1 Aminoacyl-tRNA synthetases

Aminoacyl-tRNA synthetases (aaRSs) are responsible for the correct attachment of their cognate amino acid to the 3'-end of the specific tRNA (aminoacylation). Thereby, these enzymes are critical for the conversion of the nucleotide sequences of mRNAs into the amino acid sequences of proteins, in the translational process of every known organism. Translational fidelity is determined by the accuracy with which specific amino acids are attached to their appropriate tRNAs. This reaction proceeds as follows: first, an amino acid is activated to an aminoacyl adenylate by ATP transfer, with the generation of pyrophosphate; second, the amino acid moiety of the aminoacyl adenylate is transferred to the 3'-end of the specific tRNA.

The aaRSs are divided into classes I and II, according to their primary and tertiary structures, and are further subdivided into subclasses, a, b and c, within each class (Sankaranarayanan&Moras, 2001). The class I aaRSs are characterized by the nucleotide binding (Rossmann) fold on which the active site is located, whereas the active sites of the class II aaRSs comprise a characteristic seven-stranded β sheet with flanking α helices (Cusack et al., 1990; Eriani et al., 1990).

As mentioned above, the fidelity of translation is assured by the strict discrimination of cognate from non-cognate amino acids. However, for the leucine, isoleucine, valine, threonine, alanine and phenylalanine systems, which each share structural similarity to some other systems, their cognate enzymes, i.e., the leucyl- (LeuRS), isoleucyl- (IleRS), valyl- (ValRS), threonyl- (ThrRS), alanyl- (AlaRS) and phenylalanyl- (PheRS) tRNA synthetases, have difficulties in the strict discrimination of their specific amino acids, thus producing mis-activated amino acids or mis-aminoacylated tRNAs.

For example, when isoleucine is replaced with valine in the IleRS system, the error rate is approximately 1 in 5 (Pauling, 1957). This is caused by the similarity in the chemical structures of the amino acids, and results in misactivated amino acids or misaminoacylated tRNAs (Fukai et al., 2000; Fukunaga et al., 2004; Fukunaga&Yokoyama, 2006; Lincecum et al., 2003; Mursinna et al., 2001; Nureki et al., 1998; Silvian et al., 1999; Zhai&Martinis, 2005).

To avoid such incorrect products, these enzymes catalyse editing reactions. Two types of editing, pre-transfer editing and post-transfer editing, correct mis-activated amino acids and

mis-aminoacylated tRNAs, respectively. A misactivated amino acid is hydrolysed to the amino acid and AMP by the pre-transfer editing pathway, and a mis-aminoacylated tRNA is hydrolysed to the amino acid and tRNA by the post-transfer editing pathway (Fukai et al., 2000; Fukunaga et al., 2004; Fukunaga&Yokoyama, 2005; Lincecum et al., 2003; Mursinna et al., 2001; Nordin&Schimmel, 2002; Nureki et al., 1998; Sankaranarayanan&Moras, 2001; Silvian et al., 1999; Zhai et al., 2007). In this manner, the overall error occurring in the above-mentioned isoleucine system is thus reduced to only 1 in 40,000 (Freist et al., 1985).

5.2 QM/MM molecular dynamics simulation

Ab initio QM/MM molecular dynamics (MD) simulation is a state-of-the-art theoretical methodology in current computer simulation techniques for huge molecular systems. To investigate the mechanism of the post-transfer editing reaction by the class Ia *Thermus thermophilus* leucyl-tRNA synthetase (LeuRS) complexed with valyl-tRNA^{Leu}, hybrid QM/MM MD simulations were performed coupled to *ab initio* DFT calculations as the QM simulation (Hagiwara et al., 2010a).

Several possible reaction pathways were explored by combining umbrella sampling techniques with hybrid QM/MM MD simulations to obtain the PESs of the reaction pathways, and the calculated activation barrier led us to identify the most preferable reaction pathway. The result obtained was consistent with the biochemical experimental data, and thus a novel enzymatic reaction mechanism was revealed by the theoretical strategy, as discussed below.

In contrast, reaction pathway exploration by experimental techniques is very difficult. Even if such methodologies are available to identify the enzymatic mechanisms, theoretical approaches are also indispensable to elucidate the electronic structure changes of the active site of the enzymatic systems. This is a recent successful case.

figure 11 depicts the entire calculation system of the LeuRS•valyl-tRNA^{Leu} complex, including solvent water molecules (in total ~165,000 atoms), as well as the active site of the editing reaction (Hagiwara et al., 2009b; Hagiwara et al., 2009c). To obtain the PES relevant to the enzymatic reaction, an adiabatic mapping approach was adopted. This method is based upon conformational/configurational sampling by QM/MM MD simulation and QM/MM geometry optimization. The role of the QM/MM MD simulation is the enhancement of such sampling.

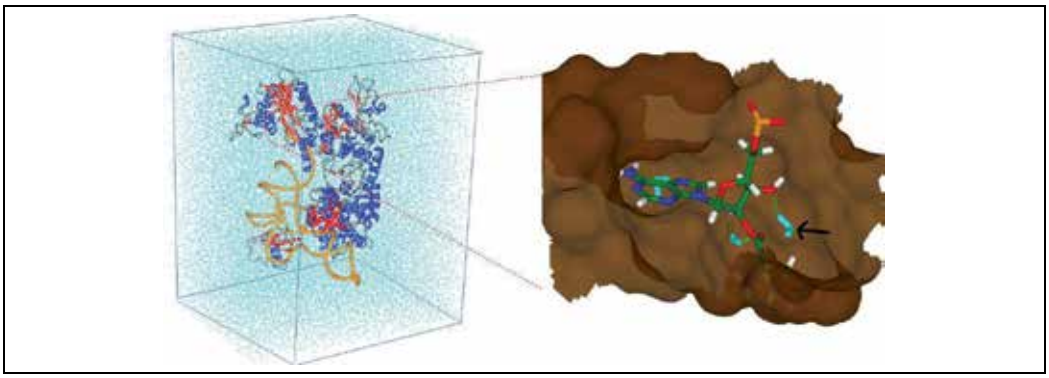


Fig. 11. The entire model system and the active site of the editing reaction in the LeuRS•valyl-tRNA^{Leu} complex. The arrow points to the nucleophilic water.

The reaction space was divided into windows, each defined by a particular set of values of the reaction coordinate variables, and then, for each window, a QM/MM MD simulation was performed, followed by a geometry optimization. Both the MD simulations and geometry optimizations were performed with functions that constrained the reaction coordinate variables, to keep them close to their window reference values. The final PESs were then reconstructed from the energies of the geometrically optimized structures, but without the constraint energy terms.

5.3 Editing mechanism: A novel enzymatic reaction by a hybrid ribozyme/protein catalyst

This computational analysis revealed that the water molecule that acts as the nucleophile in the editing reaction is activated by a 3′-hydroxyl (3′-HO) group at the 3′-end of $tRNA^{Leu}$, and that the O2′ atom of the leaving group of the substrate is capped by one of the water's hydrogen atoms. This was shown to be consistent with experimental data. Thus, it is the tRNA itself, and not the protein, that drives the editing reaction. Unexpectedly and surprisingly, editing was found to be a ribozymal self-cleavage reaction.

The protein does, however, have an important stabilizing effect on certain high-energy intermediates along the reaction path, and thereby plays a critical role in promoting the reaction and providing specificity in ligand recognition. These results led us to call the LeuRS•valyl-$tRNA^{Leu}$ complex a "hybrid ribozyme/protein catalyst", which is more efficient than the ribozyme alone (Hagiwara et al., 2010a).

Furthermore, examinations of previously reported structural and biochemical data revealed that this RNA-driven catalytic mechanism seen in the LeuRS•valyl-$tRNA^{Leu}$ complex appears to be quite widespread, as it occurs in the deacylation and peptidyl-transferase reactions of the ribosome as well as the editing reactions in the other class (class II) of aaRSs. The existence of hybrid ribozyme/protein catalysts, exemplified by the LeuRS editing mechanism, also provides insight into the types of transitional forms that could have been important in the RNA world hypothesis for the origin of life (Cusack et al., 1990; Eriani et al., 1990).

In this manner, by employing hybrid QM/MM MD simulations, we found a new biological catalyst. This system is distinct from conventional RNA•protein complexes such as the group I intron: the proteins (aaRSs) directly contribute to the reaction through hydrogen-bonding with RNA moieties, thus stabilizing the high-energy intermediates. Recently, a similar theoretical analysis of the editing mechanism was conducted by employing our fully-solvated structural model (figure 11), and the author reported almost the same reaction mechanisms as described above (Boero, *J. Phys. Chem.*, 115, 12276, 2011). In this report, an additional possible initial reaction pathway was proposed; however, this process is not feasible, as shown in the previous study (see Simulations 2 and 4 in Hagiwara, et al., 2010a). The difference (artefact) could be due to the insufficient QM regions lacking the crucial atoms, the functional used in the calculations (see Tuckerman, et al., 2006), etc.

5.4 Dynamic electronic structure rearrangements in hybrid ribozyme/protein catalysis

The dynamic changes that occur in the electronic structure of the catalytic site were investigated in the hybrid ribozyme/protein catalysis by the LeuRS•valyl-$tRNA^{Leu}$ complex. As a result of this theoretical analysis, dramatic, functionally important rearrangements of the MOs were observed (Kang et al.). We will briefly describe these dynamic rearrangements of the electronic structure in the biological macromolecular systems.

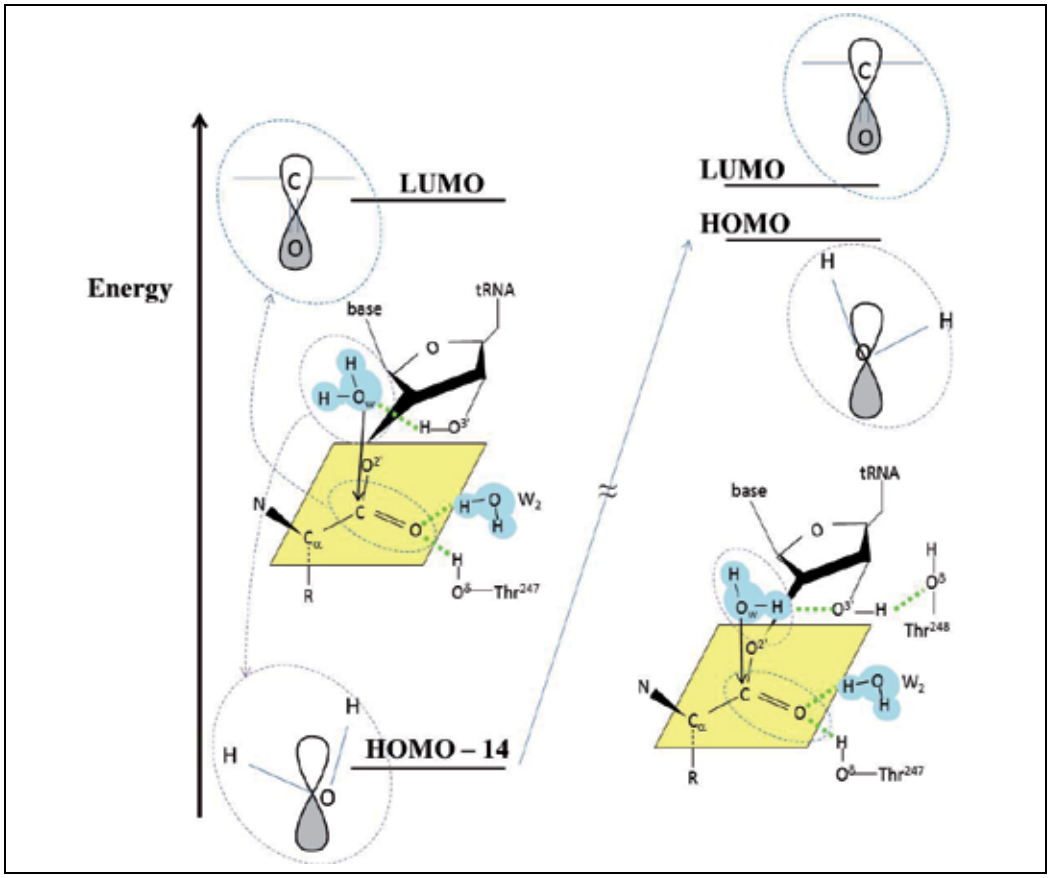


Fig. 12. Increasing the energy level of the MO predominantly contains the *p*-orbital of the nucleophilic water molecule, in the editing reaction by the LeuRS•valyl-tRNA^{Leu} complex. The left and right panels represent the initial and transition states, respectively, in the editing reaction. In the initial state of the editing reaction, the orbital that would be reactive for the bond formation is not HOMO, but HOMO-14. In fact, the energy level of HOMO-14 is elevated as the enzymatic reaction proceeds (i.e., nucleophilic attack occurs). In the transition state, this MO becomes HOMO, which leads to the formation of a new covalent bond between the water and the carbonyl group.

Two changes are particularly noteworthy. One concerns the MO that contains a contribution from the nucleophile (a water molecule). It initially has a much lower energy than the HOMO, but it is activated as the reaction proceeds until it becomes the reactive HOMO (figure 12). The other involves the reactive LUMO with anti-bonding character, which emerges in the bond rupture that leads to the products.

We refer to these processes as the dynamic induction of the reactive HOMO (DIRH) and LUMO (DIRL), respectively. Interestingly, the induction of the reactive HOMO is enhanced by the formation of a low-barrier hydrogen bond (LBHB), which represents a novel role for LBHBs in enzymatic systems.

Previous reports of computational studies of enzymatic reactions have discussed the electronic structures of the catalytic sites in terms of frontier orbital theory (FOT). However,

the above-mentioned DIRH/DIRL are unusual, in that we have highlighted the dynamic rearrangements of the MOs that are relevant to catalysis along the reaction path.
In small compound systems, the reactive HOMO and LUMO are already present in the reactant structures, since the majority of the atoms are directly relevant to the chemical reaction, and so dramatic rearrangements in the MOs are not required. By contrast, in the complex environment of a biological macromolecule, the reactive HOMO and LUMO are much more likely to be hidden, and so these systems must implement the DIRH and DIRL mechanisms that expose them.
Therefore, DIRH and DIRL are probably widespread in the reactions of complex and flexible molecular species, such as enzymes, and thus their analysis should be an essential part of the investigation of many reaction mechanisms. In this manner, state-of-the-art methodologies in computational science are essential to facilitate the elucidation of these complex electronic processes.

6. Conclusion

In this chapter, we introduced two QM/MM schemes, the additive and subtractive schemes, and discussed their advantages and disadvantages. We then showed three types of examples of their applications to biological macromolecular systems. Here, the *ab inito* DFT calculation was utilized as the QM part. In the first application, we compared the calculation results using the two schemes, to evaluate the effects of the long-range electrostatic interactions to the QM Hamiltonian (i.e., the effects of polarizability). As a consequence, the additive scheme was found to be more reliable to reproduce the experimental data. Therefore, it was concluded that the environmental structures of the protein and solvent water molecules crucially affect the geometric and electronic structures of the active site of a protein. Thereby, biological macromolecular systems may regulate the properties of their active sites.
In the second application, a computational technique that extensively explored the conformational space (involving hybrid QM/MM calculations) was important to remove the incorrect structures (e.g. steric clashes) included in the experimental structure. Utilizing the corrected atomic coordinates of the Cu_A site of CcO, the functional roles of the axial Met ligand were investigated, where some properties related to the Met residue showed experimental discrepancy. A comparison of the electronic structure obtained using the hybrid QM/MM calculation of CcO with that of a simple model system lacking the axial Met ligand revealed that the effect of this Met residue is not significant, but "selective".
Thus, the axial Met ligand was theoretically found to be a fine modulator of the electronic structure of the Cu_A site, although the experimental data were controversial concerning the role of this residue. This fine modulation may contribute to a more effective oxygen reduction process, by optimizing the electron transfer reaction to the environment. In this manner, hybrid QM/MM approaches are crucial to resolve the experimental discrepancy related to the electronic properties of biological macromolecules.
In the final application, we showed an example of a hybrid QM/MM MD simulation that was used to elucidate the editing mechanism of the LeuRS•valyl-$tRNA^{Leu}$ complex. The results of this analysis revealed a novel biological catalyst, a hybrid ribozyme/protein enzyme. In this catalysis, the dynamic rearrangement of the electronic structures was discovered. Thus, hybrid QM/MM MD simulation is currently one of the most powerful theoretical methodologies to elucidate the functional mechanisms of enzymatic systems.

7. Acknowledgments

This work was partly supported by Grants-in-Aid under contract Nos. 21340108 and 23654129, and the Global Center of Excellence Program, each provided by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT). Computations were performed using computer facilities under the "Interdisciplinary Computational Science Program" at the Center for Computational Sciences, University of Tsukuba, the Computer Center for Agriculture, Forestry, and Fisheries Research, MAFF, Japan, and the Supercomputer Center, Institute for Solid State Physics, University of Tokyo. J.K. was supported by a Research Fellowship of the Japan Society for the Promotion of Science.

8. References

Amara, P., & Field, M.J. (2003). Evaluation of an ab initio quantum mechanical/molecular mechanical hybrid-potential link-atom method. *Theoretical Chemistry Accounts*, 109, 43-52.

Amara, P., Field, M.J., Alhambra, C., & Gao, J.L. (2000). The generalized hybrid orbital method for combined quantum mechanical/molecular mechanical calculations: formulation and tests of the analytical derivatives. *Theoretical Chemistry Accounts*, 104, 336-343.

Assfeld, X., & Rivail, J.L. (1996). Quantum chemical computations on parts of large molecules: The ab initio local self consistent field method. *Chemical Physics Letters*, 263, 100-106.

Berry, S.M., Ralle, M., Low, D.W., Blackburn, N.J., & Lu, Y. (2003). Probing the role of axial methionine in the blue copper center of azurin with unnatural amino acids. *Journal of the American Chemical Society*, 125, 8760-8768.

Bruice, T.C., & Kahn, K. (2000). Computational enzymology. *Curr Opin Chem Biol*, 4, 540-544.

Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R.J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26, 1668-1688.

Covello, P.S., & Gray, M.W. (1990). RNA sequence and the nature of the Cu_A-binding site in cytochrome c oxidase. *FEBS Letters*, 268, 5-7.

Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N., & Leberman, R. (1990). A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 Å. *Nature*, 347, 249-255.

DeBeer George, S., Metz, M., Szilagyi, R.K., Wang, H., Cramer, S.P., Lu, Y., Tolman, W.B., Hedman, B., Hodgson, K.O., & Solomon, E.I. (2001). A quantitative description of the ground-state wave function of Cu_A by X-ray absorption spectroscopy: comparison to plastocyanin and relevance to electron transfer. *Journal of the American Chemical Society*, 123, 5757-5767.

Eichinger, M., Tavan, P., Hutter, J., & Parrinello, M. (1999). A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *Journal of Chemical Physics*, 110, 10452-10467.

Eriani, G., Delarue, M., Poch, O., Gangloff, J., & Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, 347, 203-206.

Ferre, N., Assfeld, X., & Rivail, J.L. (2002). Specific force field parameters determination for the hybrid ab initio QM/MM LSCF method. *Journal of Computational Chemistry*, 23, 610-624.

Field, M.J. (2002). Simulating enzyme reactions: challenges and perspectives. *Journal of Computational Chemistry*, 23, 48-58.

Field, M.J., Bash, P.A., & Karplus, M. (1990). A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *Journal of Computational Chemistry*, 11, 700-733.

Freist, W., Pardowitz, I., & Cramer, F. (1985). Isoleucyl-tRNA synthetase from bakers' yeast: multistep proofreading in discrimination between isoleucine and valine with modulated accuracy, a scheme for molecular recognition by energy dissipation. *Biochemistry*, 24, 7014-7023.

Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Jr., J.A.M., Vreven, T., Kudin, K.N., Burant, J.C., *et al.* (2003). *Gaussian 03 Gaussian Inc Pittsburgh PA*.

Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D.G., & Yokoyama, S. (2000). Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of $tRNA^{Val}$ and valyl-tRNA synthetase. *Cell*, 103, 793-803.

Fukunaga, R., Fukai, S., Ishitani, R., Nureki, O., & Yokoyama, S. (2004). Crystal structures of the CP1 domain from Thermus thermophilus isoleucyl-tRNA synthetase and its complex with L-valine. *The Journal of biological chemistry*, 279, 8396-8402.

Fukunaga, R., & Yokoyama, S. (2005). Structural basis for non-cognate amino acid discrimination by the valyl-tRNA synthetase editing domain. *Journal of Biological Chemistry*, 280, 29937-29945.

Fukunaga, R., & Yokoyama, S. (2006). Structural basis for substrate recognition by the editing domain of isoleucyl-tRNA synthetase. *Journal of Molecular Biology*, 359, 901-912.

Gamelin, D.R., Randall, D.W., Hay, M.T., Houser, R.P., Mulder, T.C., Canters, G.W., Vries, S.d., Tolman, W.B., Lu, Y., & Solomon, E.I. (1998). Spectroscopy of Mixed-Valence Cu_A-Type Centers: Ligand-Field Control of Ground-State Properties Related to Electron Transfer. *J Am Chem Soc*, 120, 5246-5263.

Gao, J., Ma, S., Major, D.T., Nam, K., Pu, J., & Truhlar, D.G. (2006). Mechanisms and free energies of enzymatic reactions. *Chemical Reviews*, 106, 3188-3209.

Gao, J.L., Amara, P., Alhambra, C., & Field, M.J. (1998). A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *Journal of Physical Chemistry A*, 102, 4714-4721.

Garner, D.K., Vaughan, M.D., Hwang, H.J., Savelieff, M.G., Berry, S.M., Honek, J.F., & Lu, Y. (2006). Reduction potential tuning of the blue copper center in Pseudomonas aeruginosa azurin by the axial methionine as probed by unnatural amino acids. *Journal of the American Chemical Society*, 128, 15608-15617.

George, S.D., Basumallick, L., Szilagyi, R.K., Randall, D.W., Hill, M.G., Nersissian, A.M., Valentine, J.S., Hedman, B., Hodgson, K.O., & Solomon, E.I. (2003). Spectroscopic Investigation of Stellacyanin Mutants: Axial Ligand Interactions at the Blue Copper Site. *Journal of the American Chemical Society*, 125, 11314–11328.

Gorelsky, S.I., Xie, X., Chen, Y., Fee, J.A., & Solomon, E.I. (2006). The two-state issue in the mixed-valence binuclear Cu_A center in cytochrome c oxidase and N_2O reductase. *Journal of the American Chemical Society*, 128, 16452-16453.

Gray, H.B., Malmström, B.G., & Williams, R.J.P. (2000). Copper coordination in blue proteins. *Journal of Biological Inorganic Chemistry*, 5, 551-559.

Hagiwara, Y., Field, M.J., Nureki, O., & Tateno, M. (2010a). Editing mechanism of aminoacyl-tRNA synthetases operates by a hybrid ribozyme/protein catalyst. *Journal of the American Chemical Society*, 132, 2751-2758.

Hagiwara, Y., Kang, J., & Tateno, M. (2011). Structural Instability of the Active Site of T1 Lipase Induced by Replacement of Na+ with Water Complexed with the Phenylalanine Aromatic Ring. *Journal of Chemical Theory and Computation*, 7, 2593-2599.

Hagiwara, Y., Kino, H., & Tateno, M. (2010b). Modulation of electronic structures of bases through DNA recognition of protein. *Journal of Physics: Condensed Matter*, 22, 152101.

Hagiwara, Y., Matsumura, H., & Tateno, M. (2009a). Functional roles of a structural element involving Na^+-π interactions in the catalytic site of T1 lipase revealed by molecular dynamics simulations. *Journal of the American Chemical Society*, 131, 16697-16705.

Hagiwara, Y., Nureki, O., & Tateno, M. (2009b). Identification of the nucleophilic factors and the productive complex for the editing reaction by leucyl-tRNA synthetase. *FEBS Letters*, 583, 1901-1908.

Hagiwara, Y., Nureki, O., & Tateno, M. (2009c). Structural modelling of the complex of leucyl-tRNA synthetase and mis-aminoacylated $tRNA^{Leu}$. *FEBS Letters*, 583, 825-830.

Hagiwara, Y., Ohta, T., & Tateno, M. (2009d). QM/MM hybrid calculation of biological macromolecules using a new interface program connecting QM and MM engines. *Journal of Physics: Condensed Matter*, 21, 064234.

Harvey, J.N. (2004). Spin-forbidden CO ligand recombination in myoglobin. *Faraday Discussions*, 127, 165-177.

Hay, M., Richards, J.H., & Lu, Y. (1996). Construction and characterization of an azurin analog for the purple copper site in cytochrome c oxidase. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 461-464.

Holm, R.H., Kennepohl, P., & Solomon, E.I. (1996). Structural and Functional Aspects of Metal Sites in Biology. *Chemical Reviews*, 96, 2239-2314.

Hwang, H.J., Berry, S.M., Nilges, M.J., & Lu, Y. (2005). Axial methionine has much less influence on reduction potentials in a Cu_A center than in a blue copper center. *Journal of the American Chemical Society*, 127, 7274-7275.

Hwang, H.J., Nagraj, N., & Lu, Y. (2006). Spectroscopic characterizations of bridging cysteine ligand variants of an engineered $Cu_2(Scys)_2$ Cu_A azurin. *Inorganic Chemistry*, 45, 102-107.

Kaila, V.R.I., Verkhovsky, M.I., & Wikström, M. (2010). Proton-Coupled Electron Transfer in Cytochrome Oxidase. *Chemical Reviews*, 110, 7062-7081.

Kang, J., Hagiwara, Y., Kino, H., Field, M.J., & Tateno, M. Electronic Structure Rearrangements in Hybrid Ribozyme/Protein Catalysis. *Journal of Chemical Theory and Computation*, to be published.

Kang, J., Kino, H., & Tateno, M. (2011). A theoretical investigation of the functional role of the axial methionine ligand of the Cu_A site in cytochrome c oxidase. *Biochimica et Biophysica Acta : Bioenergetics*, 1807, 1314-1327.

Kang, J., Ohta, T., Hagiwara, Y., Nishikawa, K., Yamamoto, T., Nagao, H., & Tateno, M. (2009). Electronic and geometric structures of the blue copper site of azurin investigated by QM/MM hybrid calculations. *Journal of Physics: Condensed Matter*, 21, 064235.

Kelterer, A.M., Landgraf, S., & Grampp, G. (2001). Quantum mechanical investigation of the inner-sphere reorganization energy of cyclooctatetraene/cyclooctatetraene radical anion. Part I. *Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy*, 57, 1959-1969.

Kroneck, P.M., Antholine, W.A., Riester, J., & Zumft, W.G. (1988). The cupric site in nitrous oxide reductase contains a mixed-valence [Cu(II),Cu(I)] binuclear center: a multifrequency electron paramagnetic resonance investigation. *FEBS Letters*, 242, 70-74.

Kroneck, P.M., Antholine, W.A., Riester, J., & Zumft, W.G. (1989). The nature of the cupric site in nitrous oxide reductase and of Cu_A in cytochrome c oxidase. *FEBS Letters*, 248, 212-213.

Ledesma, G.N., Murgida, D.H., Ly, H.K., Wackerbarth, H., Ulstrup, J., Costa-Filho, A.J., & Vila, A.J. (2007). The Met Axial Ligand Determines the Redox Potential in Cu_A Sites. *Journal of the American Chemical Society*, 129, 11884-11885.

Lin, H., & Truhlar, D.G. (2005). Redistributed charge and dipole schemes for combined quantum mechanical and molecular mechanical calculations. *Journal of Physical Chemistry A*, 109, 3991-4004.

Lin, H., & Truhlar, D.G. (2007). QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts*, 117, 185-199.

Lincecum, T.L., Jr., Tukalo, M., Yaremchuk, A., Mursinna, R.S., Williams, A.M., Sproat, B.S., Van Den Eynde, W., Link, A., Van Calenbergh, S., Grotli, M., Martinis, S.A., & Cusack, S. (2003). Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase. *Molecular cell*, 11, 951-963.

Marcus, R.A., & Sutin, N. (1985). Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta (BBA) : Reviews on Bioenergetics*, 811, 265-322.

Monard, G., Loos, M., Thery, V., Baka, K., & Rivail, J.L. (1996). Hybrid classical quantum force field for modeling very large molecules. *International Journal of Quantum Chemistry*, 58, 153-159.

Moser, C.C., Keske, J.M., Warncke, K., Farid, R.S., & Dutton, P.L. (1992). Nature of biological electron transfer. *Nature*, 355, 796-802.

Mulholland, A.J. (2005). Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discovery Today*, 10, 1393-1402.

Mursinna, R.S., Lincecum, T.L., Jr., & Martinis, S.A. (2001). A conserved threonine within Escherichia coli leucyl-tRNA synthetase prevents hydrolytic editing of leucyl-$tRNA^{Leu}$. *Biochemistry*, 40, 5376-5381.

Namslauer, A., & Brzezinski, P. (2004). Structural elements involved in electron-coupled proton transfer in cytochrome c oxidase. *FEBS Letters*, 567, 103-110.

Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M., & Canters, G.W. (1991). Crystal structure analysis of oxidized Pseudomonas aeruginosa azurin at pH 5.5 and pH 9.0. A pH-induced conformational transition involves a peptide bond flip. *Journal of Molecular Biology*, 221, 765-772.

Nordin, B.E., & Schimmel, P. (2002). Plasticity of recognition of the 3'-end of mischarged tRNA by class I aminoacyl-tRNA synthetases. *Journal of Biological Chemistry*, 277, 20510-20517.

Nureki, O., Vassylyev, D.G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T.L., Schimmel, P., & Yokoyama, S. (1998). Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science*, 280, 578-582.

Papa, S., Capitanio, N., & Capitanio, G. (2004). A cooperative model for proton pumping in cytochrome c oxidase. *Biochimica et Biophysica Acta : Bioenergetics*, 1655, 353-364.

Paraskevopoulos, K., Sundararajan, M., Surendran, R., Hough, M.A., Eady, R.R., Hillier, I.H., & Hasnain, S.S. (2006). Active site structures and the redox properties of blue copper proteins: atomic resolution structure of azurin II and electronic structure calculations of azurin, plastocyanin and stellacyanin. *Dalton Trans*, 25, 3067-3076.

Pauling, L. (1957). Festschrift Arthur Stroll (Birkhäuser-Verlag, Basel, Switzerland.).

Paumann, M., Lubura, B., Regelsberger, G., Feichtinger, M., Kollensberger, G., Jakopitsch, C., Furtmuller, P.G., Peschek, G.A., & Obinger, C. (2004). Soluble Cu_A domain of cyanobacterial cytochrome c oxidase. *Journal of Biological Chemistry*, 279, 10293-10303.

Pu, J.Z., Gao, J.L., & Truhlar, D.G. (2004). Generalized hybrid orbital (GHO) method for combining ab initio Hartree-Fock wave functions with molecular mechanics. *Journal of Physical Chemistry A*, 108, 632-650.

Reuter, N., Dejaegere, A., Maigret, B., & Karplus, M. (2000). Frontier bonds in QM/MM methods: A comparison of different approaches. *Journal of Physical Chemistry A*, 104, 1720-1735.

Robinson, H., Ang, M.C., Gao, Y.G., Hay, M.T., Lu, Y., & Wang, A.H. (1999). Structural basis of electron transfer modulation in the purple Cu_A center. *Biochemistry*, 38, 5677-5683.

Ryde, U. (2003). Combined quantum and molecular mechanics calculations on metalloproteins. *Current Opinion in Chemical Biology*, 7, 136-142.

Ryde, U., & Olsson, M.H.M. (2001). Structure, strain, and reorganization energy of blue copper models in the protein. *International Journal of Quantum Chemistry*, 81, 335-347.

Ryde, U., Olsson, M.H.M., Roos, B.O., De Kerpel, J.O.A., & Pierloot, K. (2000). On the role of strain in blue copper proteins. *Journal of Biological Inorganic Chemistry*, 5, 565-574.

Sankaranarayanan, R., & Moras, D. (2001). The fidelity of the translation of the genetic code. *Acta Biochimica Polonica*, 48, 323-335.

Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S.J., Windus, T.L., Dupuis, M., & Montgomery, J.A. (1993). General Atomic and Molecular Electronic-Structure System. *Journal of Computational Chemistry*, 14, 1347-1363.

Senn, H.M., & Thiel, W. (2007). QM/MM studies of enzymes. *Current Opinion in Chemical Biology*, 11, 182-187.

Senn, H.M., & Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, 48, 1198-1229.

Sherwood, P., de Vries, A.H., Guest, M.F., Schreckenbach, G., Catlow, C.R.A., French, S.A., Sokol, A.A., Bromley, S.T., Thiel, W., Turner, A.J., Billeter, S., Terstegen, F., Thiel, S., Kendrick, J., Rogers, S.C., Casci, J., Watson, M., King, F., Karlsen, E., Sjøvoll, M., Fahmi, A., Schäfer, A., & Lennartz, C. (2003). QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *Journal of Molecular Structure: THEOCHEM*, 632, 1-28.

Silvian, L.F., Wang, J., & Steitz, T.A. (1999). Insights into editing from an ile-tRNA synthetase structure with $tRNA^{ile}$ and mupirocin. *Science*, 285, 1074-1077.

Sinnecker, S., & Neese, F. (2006). QM/MM calculations with DFT for taking into account protein effects on the EPR and optical spectra of metalloproteins. Plastocyanin as a case study. *Journal of Computational Chemistry*, 27, 1463-1475.

Solomon, E.I. (2006). Spectroscopic Methods in Bioinorganic Chemistry: Blue to Green to Red Copper Sites. *Inorganic Chemistry*, 45, 8012–8025.

Solomon, E.I., Szilagyi, R.K., DeBeer George, S., & Basumallick, L. (2004). Electronic Structures of Metal Sites in Proteins and Models: Contributions to Function in Blue Copper Proteins. *Chemical Reviews*, 104, 419–458.

Sugiyama, A., Sugimori, K., Shuku, T., Nakamura, T., Saito, H., Nagao, H., Kawabe, H., & Nishikawa, K. (2005). Electronic structure of the active site with two configurations of azurin. *International Journal of Quantum Chemistry*, 105, 588–595.

Thery, V., Rinaldi, D., Rivail, J.L., Maigret, B., & Ferenczy, G.G. (1994). Quantum-Mechanical Computations on Very Large Molecular-Systems - the Local Self-Consistent-Field Method. *Journal of Computational Chemistry*, 15, 269-282.

Torras, J., Seabra, G.D.M., Deumens, E., Trickey, S.B., & Roitberg, A.E. (2008). A versatile AMBER-Gaussian QM/MM interface through PUPIL. *Journal of Computational Chemistry*, 29, 1564–1573.

Tuckerman, M. E., Chandra, A., and Marx, D. (2006). Structure and dynamics of OH-(aq). *Acc. Chem. Res.*, 39, 151-8.

Wang, K., Geren, L., Zhen, Y., Ma, L., Ferguson-Miller, S., Durham, B., & Millett, F. (2002). Mutants of the Cu_A site in cytochrome c oxidase of Rhodobacter sphaeroides: II. Rapid kinetic analysis of electron transfer. *Biochemistry*, 41, 2298-2304.

Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103, 227-249.

Woodcock, H.L., Hodoscek, M., Gilbert, A.T.B., Gill, P.M.W., Schaefer, H.F., & Brooks, B.R. (2007). Interfacing Q-chem and CHARMM to perform QM/MM reaction path calculations. *Journal of Computational Chemistry*, 28, 1485-1502.

Xie, X., Gorelsky, S.I., Sarangi, R., Garner, D.K., Hwang, H.J., Hodgson, K.O., Hedman, B., Lu, Y., & Solomon, E.I. (2008). Perturbations to the geometric and electronic structure of the Cu_A site: factors that influence delocalization and their contributions to electron transfer. *Journal of the American Chemical Society*, 130, 5194-5205.

Zhai, Y., & Martinis, S.A. (2005). Two conserved threonines collaborate in the Escherichia coli leucyl-tRNA synthetase amino acid editing mechanism. *Biochemistry*, 44, 15437-15443.

Zhai, Y., Nawaz, M.H., Lee, K.W., Kirkbride, E., Briggs, J.M., & Martinis, S.A. (2007). Modulation of substrate specificity within the amino acid editing site of leucyl-tRNA synthetase. *Biochemistry*, 46, 3331-3337.

Zhen, Y., Schmidt, B., Kang, U.G., Antholine, W., & Ferguson-Miller, S. (2002). Mutants of the Cu_A Site in Cytochrome c Oxidase of Rhodobacter sphaeroides: I. Spectral and Functional Properties. *Biochemistry*, 41, 2288-2297.

Zickermann, V., Wittershagen, A., Kolbesen, B.O., & Ludwig, B. (1997). Transformation of the Cu_A Redox Site in Cytochrome c Oxidase into a Mononuclear Copper Center. *Biochemistry*, 36, 3232-3236.

15

Battle of the Sexes: A Quantum Games Theory Approach

Juan Manuel López R.
Logic and Computation Group, Engineering Physics
EAFIT University
Colombia

1. Introduction

Real life situations where there are strategies to be chosen in order to obtain a profit can be reproduced by games, so game theory is a way to describe the evolution of possible scenarios where players can select a scheme of play. Game theory takes importance in many areas, such as people decisions making, where their choices do affect others benefit (Davis, 1970; Myerson, 1991). It is to remark that the principles of game theory were initiated by trying to understand the behavior of economic strategies, however Von Neumann presented the concept of modern game theory in 1944 (Von Neumann & Morgenstern, 1947).
Quantum mechanics is a tool that creates another point of view for the traditional game theory due to multiple strategies offered for the players, whom possibilities are numerously expanded in contrast of classical ones (Eisert & Wilkens, 2000). There are also games where the player who uses quantum strategies, enhances his payoffs or even always wins against a player who only uses classical moves (Meyer, 1999). It is to remark that there are plenty of applications for quantum game theory such as quantum cryptography and computation, economics and biology (Piotrowski & Sladkowski, 2003; Hanauske et al.,2009).
The Battle of the Sexes game is a largely analyzed problem, based on two players: Alice and Bob and their choice about an activity for a Saturday night with each other. It is pretty important to remark that both want the best possible payoff in the decision, so the game can be developed normally, otherwise it would not be our case. Alice, really loves Opera, but wants to be with Bob; Bob likes Football but he wants to have Alice's company along the activity. This game has a lot of applications in real life scenarios such as the spread of some type of genes from a reproduction between two organisms (Dawkins, 2006); another interesting application is neuroeconomics (Montague & Berns, 2002), where brain studies have been done in order to incite neurons to choose either to "work" for a reward or to "shirk" (Glimcher, 2003).

2. Classical analysis

The global idea of the model used in this section is taken from Richard Dawkins' Battle of the Sexes Model, however the specific method implemented is taken from a worksheet made by Frank Wang (Wang, 2010), due to its facility to be developed in a computer algebra software such as Maple.

2.1 Method

M_A and M_B are payoff matrixes for Alice and Bob, respectively. Where M_{Aii} is the payoff matrix of a female player using strategy i against a male playing strategy j; and M_{Bij} is the payoff matrix for a male playing i against a female who plays j. x_i is the proportion of females playing i; and y_i is the proportion of males playing strategy i. As here we considerate two possible strategies whose proportions satisfy: $x_1 + x_2 = 1$; and the same applies for y_i. In order to reduce the number of variables, we describe x_2 and y_2 in terms of x_1 and y_1, respectively.

$$x_1(t) = x(t) \tag{1}$$

$$x_2(t) = 1 - x_1(t) \tag{2}$$

We substitute (1) in (2):

$$x_2(t) = 1 - x(t) \tag{3}$$

$$y_1(t) = y(t) \tag{4}$$

$$y_2(t) = 1 - y_1(t) \tag{5}$$

(4) in (5) to get a simpler equation for y:

$$y_2(t) = 1 - y(t) \tag{6}$$

The relevant fitness functions are defined as:

$$\begin{gathered} f(x,y) = x_1(t)\,(A_{11}\,y_1(t) + A_{12}\,y_2(t) \\ -\big(A_{11}\,x_1(t)\,y_1(t) + A_{12}\,x_1(t)\,y_2(t) + A_{21}\,x_2(t)\,y_1(t) + A_{22}\,x_2(t)\,y_2(t)\big)\big), \\ g(x,y) = y_1(t)\,(B_{11}\,x_1(t) + B_{12}\,x_2(t) \\ -\big(B_{11}\,x_1(t)\,y_1(t) + B_{12}\,y_1(t)\,x_2(t) + B_{21}\,y_2(t)\,x_1(t) + B_{22}\,y_2(t)\,x_2(t)\big)\big); \end{gathered} \tag{7}$$

The appropriate replicator equations in terms of the previously defined fitness functions are:

$$\begin{gathered} \frac{d\,x(t)}{dt} = f(x,y), \\ \frac{d\,y(t)}{dt} = g(x,y); \end{gathered} \tag{8}$$

In order to find the equilibrium values for x and y, we remove time-dependence by solving (9) when the derivatives are zero, hence $x(t) = x$ and $y(t) = y$. Then we solve the resulting system of two equations and two variables. In order to make the stability analysis, we search for a solution with the form:

$$\begin{gathered} x(t) = x + \varepsilon(t), \\ y(t) = y + \eta(t); \end{gathered} \tag{9}$$

where x and y are the equilibrium values previously found; $\varepsilon(t)$ and $\eta(t)$ are introduced as small perturbations respect to the equilibrium values. Now, we substitute (9) in equations (8). We can neglect the squares of η and ε and their product because their result is very small. So we can define the following simplifications in order to be replaced on the resulting replicator equations.

$$\begin{aligned} \eta(t)\,\varepsilon(t) &= 0, \\ \varepsilon(t)^2 &= 0, \\ \eta(t)^2 &= 0; \end{aligned} \tag{10}$$

The replicator equations are finally linear, so them can be analytically solved. From these solutions we can observe if the equilibrium values are stable.
An alternative method consists on the Jacobian's use and the computation of its eigenvalues, where the stability depends on the resulting sign of the real part of the eigenvalues. However, when the eigenvalues are pure imaginary numbers, the Jacobian method is not appropriate for the stability analysis. In this last case, we use the method of the vector field for the replicator equations; which is graphically implemented.

2.2 Standard case

The previously shown method will be illustrated using three possible examples of the traditional battle of the sexes game:

a. Dawkins example.
b. Wikipedia's example[1].
c. Wikipedia's example

2.2.1 Dawkins model by Frank Wang

This first example is taken from a Maple worksheet designed by Frank Wang, based on Richard Dawkins' Battle of the Sexes Model presented on Chapter 9 of the celebrated book titled "Selfish Gene". The classical payoff matrices for this scenario are:

$$\begin{aligned} A &= \begin{bmatrix} 2 & 0 \\ 5 & -5 \end{bmatrix}, \\ B &= \begin{bmatrix} 2 & 5 \\ 0 & 15 \end{bmatrix}; \end{aligned} \tag{11}$$

According with these payoff matrices, the corresponding fitness functions are:

$$\begin{aligned} f(x,y) &= x_1(t)\ (2\,y_1(t) - 2\,x_1(t)\,y_1(t) - 5\,x_2(t)\,y_1(t) + 5\,x_2(t)\,y_2(t)), \\ g(x,y) &= y_1(t)\ \big(2\,x_1(t) + 5\,x_2(t) - 2\,x_1(t)\,y_1(t) - 5\,x_2(t)\,y_1(t) - 15\,x_2(t)\,y_2(t)\big); \end{aligned} \tag{12}$$

which are reduced to:

$$\begin{aligned} f(x,y) &= x(t)\,(8\,y(t) - 5)\,(-1 + x(t)), \\ g(x,y) &= -2y(t)\ \big(-1 + y(t)\big)\,(6\,x(t) - 5). \end{aligned} \tag{13}$$

[1] Wikipedia – Battle of the Sexes (game theory)

With the previous found fitness, the replicator equations take the form:

$$\frac{d\,x(t)}{dt} = x(t)\,(8\,y(t) - 5)\,\big(-1 + x(t)\big),$$
$$\frac{d\,y(t)}{dt} = -2\,y(t)\,\big(-1 + y(t)\big)\,(6\,x(t) - 5). \tag{14}$$

In order to find the equilibrium values, we remove time dependence for x and y, making $x(t) = x$ and $y(t) = y$; and we obtain:

$$0 = -8\,y\,x + 5\,x + 8\,x^2 y - 5\,x^2$$
$$0 = 12\,y\,x - 12\,x\,y^2 - 10\,y + 10\,y^2 \tag{15}$$

Solving these last equations, we get:

$$\{x = 0, y = 0\}, \{x = 1, y = 0\}, \{x = 0, y = 1\},$$
$$\left\{x = \frac{5}{6}, y = \frac{5}{8}\right\}, \{x = 1, y = 1\}. \tag{16}$$

In order to make the stability analysis, we use the non-trivial equilibrium values and the following small perturbations around these values:

$$x(t) = \frac{5}{6} + \varepsilon(t)$$
$$y(t) = \frac{5}{8} + \eta(t) \tag{17}$$

The resulting replicator equations are:

$$\frac{d}{dt}\,\varepsilon(t) = \left(-\frac{10}{9}\right)\eta(t) + \left(\frac{16}{3}\right)\eta(t)\,\varepsilon(t) + 8\,\eta(t)\,\varepsilon(t)^2,$$
$$\frac{d}{dt}\,\eta(t) = \left(\frac{45}{16}\right)\varepsilon(t) - 3\,\eta(t)\,\varepsilon(t) - 12\,\varepsilon(t)\,\eta(t)^2; \tag{18}$$

and using (10), we obtain the following linear equations:

$$\frac{d}{dt}\,\varepsilon(t) = -\frac{10}{9}\,\eta(t),$$
$$\frac{d}{dt}\,\eta(t) = \frac{45}{16}\,\varepsilon(t); \tag{19}$$

The solution for the last system of equations is:

$$\left\{\varepsilon(t) = C_1 \sin\left(\frac{5}{4}\sqrt{2}\,t\right) + C_2 \cos\left(\frac{5}{4}\sqrt{2}\,t\right), \eta(t)\right.$$
$$\left. = \frac{9}{8}\sqrt{2}\,\left(-C_1 \cos\left(\frac{5}{4}\sqrt{2}\,t\right) + C_2 \sin\left(\frac{5}{4}\sqrt{2}\,t\right)\right)\right\} \tag{20}$$

About the stability analysis, we observe that the small perturbations are oscillating functions without damping. It is to say that the perturbations never end, and therefore the equilibrium values are instable. It is worthwhile to note that in this case the Jacobian has eigenvalues that are pure imaginary numbers, so it is necessary to use the vector field method in order to see the results. The subsequent vector field plot is:

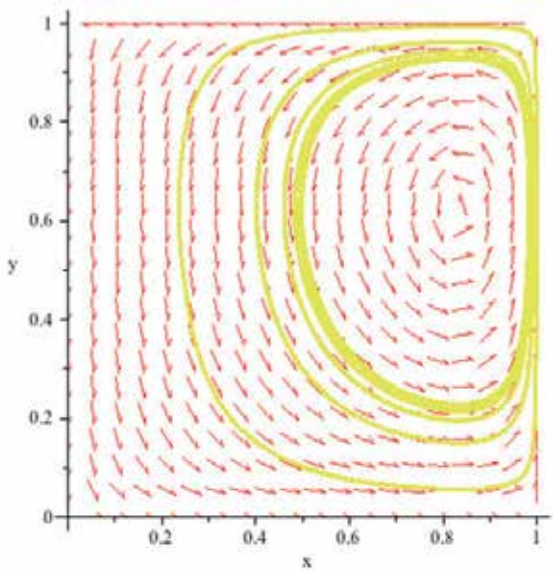


Fig. 1. Vector Field plot of the results.

We can observe that the trajectories in the phase plane are closed curves around the non-trivial equilibrium point, considered as a center. We can also make a numerical solution for the replicator equations, whose results are graphically shown:

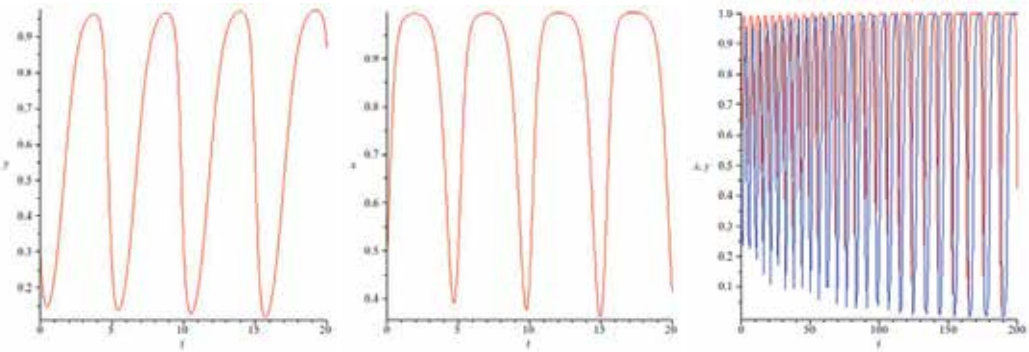


Fig. 2. a) Numerical solution for Alice's, b) Numerical solution for Bob's, c) Alice's and Bob's results plotted together with an extended time.

The three previous graphs are the time series for the player's proportions. The following graph is the phase plane for x and y:

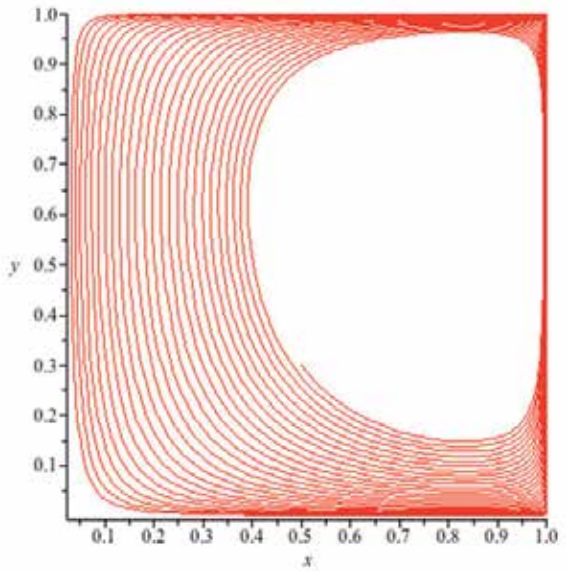


Fig. 3. Phase plane of the solutions.

2.2.2 Wikipedia case 1

The following example is taken from Wikipedia, where is referred as case 1. In this scenario, Alice and Bob receive pays only when both of them meet at the Opera, where Alice gets 3 and Bob 2; or at the football match, where Alice gets 2 and Bob 3. Also, both Alice and Bob receive no pay when they go to the wrong places either Opera or Football. It is to considerate the case where Alice receives a small payoff for going alone to the Opera and Bob for attending to the football match, but this will be analyzed in section 2.2.3. The previous scenario is shown on the following matrices, for Alice and Bob respectively:

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}; \tag{21}$$

Applying matrices at (21) in the method previously shown, the solutions obtained are:

$$\{x = 0, y = 0\}, \{x = 1, y = 0\}, \{x = 0, y = 1\}, \left\{x = \frac{3}{5}, y = \frac{2}{5}\right\}, \{x = 1, y = 1\}. \tag{22}$$

Following the steps presented in the method, the following linear equations are shown, which came from replicators equations:

$$\frac{d}{dt}\,\varepsilon(t) = \frac{6}{5}\,\eta(t), \qquad \frac{d}{dt}\,\eta(t) = \frac{6}{5}\,\varepsilon(t); \tag{23}$$

Finally we can find the solutions for the previous system (23):

$$\left\{\varepsilon(t) = C_1\, e^{-\frac{6}{5}t} + C_2\, e^{\frac{6}{5}t}, \eta(t) = -C_1\, e^{-\frac{6}{5}t} + C_2\, e^{\frac{6}{5}t}\right\} \tag{24}$$

From the previous solution we can deduce that both strategies meets at a specific point. A vector field plot is made:

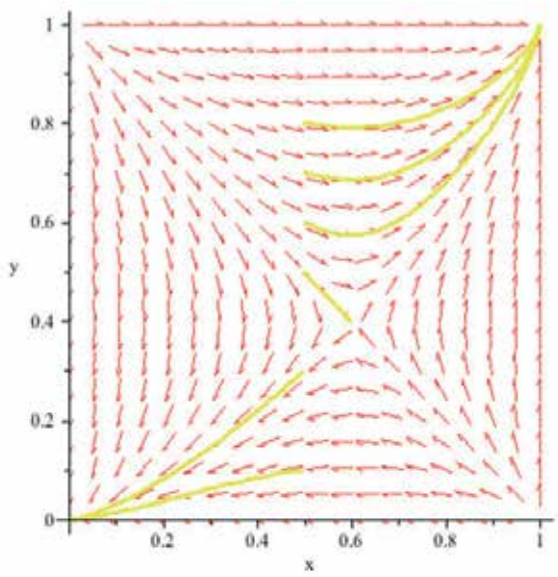


Fig. 4. Vector field plot.

Numerical solutions in the phase plane for both strategies are:

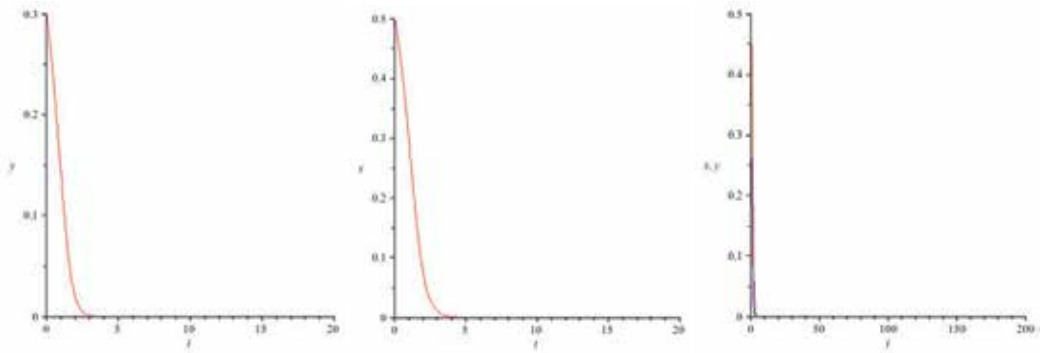


Fig. 5. a), b) Alice's and Bob's trajectories in the phase plane respectively, c) superposed graphs with a longer time period.

Figure 5 shows the evolution in time of the solutions obtained. The following graph is the phase plane for x and y:

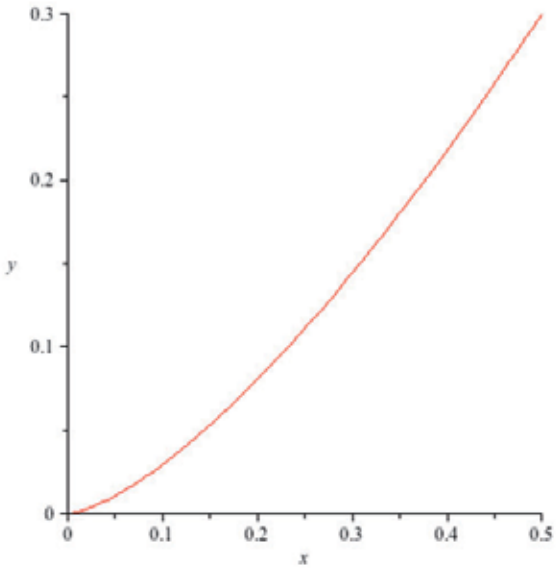


Fig. 6. Phase plane for x and y.

2.2.3 Wikipedia case 2

It's presented a possible modification for the previous case. Here, Alice receives a small pay for attending the Opera alone; likewise, Bob gets a positive payment for going to the football match without Alice. The amount received in both cases will be taken as 1; for the other possible situations, the payoffs keep being the same:

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix},$$
$$B = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}; \tag{25}$$

Using replicator's equation in the method presented, the equilibrium values obtained for the system are:

$$\{x = 0, y = 0\}, \{x = 1, y = 0\}, \{x = 0, y = 1\}, \left\{x = \frac{3}{4}, y = \frac{1}{4}\right\}, \{x = 1, y = 1\}. \tag{26}$$

Using equations (10) and the non-trivial equilibrium values, the following linear equations are found:

$$\frac{d}{dt}\,\varepsilon(t) = \frac{3}{4}\,\eta(t),$$
$$\frac{d}{dt}\,\eta(t) = \frac{3}{4}\,\varepsilon(t); \tag{27}$$

which led to the solutions for $\varepsilon(t)$ and $\eta(t)$:

$$\left\{\varepsilon(t) = C_1 e^{\frac{3}{4}t} + C_2 e^{-\frac{3}{4}t}, \eta(t) = C_1 e^{\frac{3}{4}t} - C_2 e^{-\frac{3}{4}t}\right\} \tag{28}$$

In order to understand the stability analysis, a vector field plot of the solutions is made:

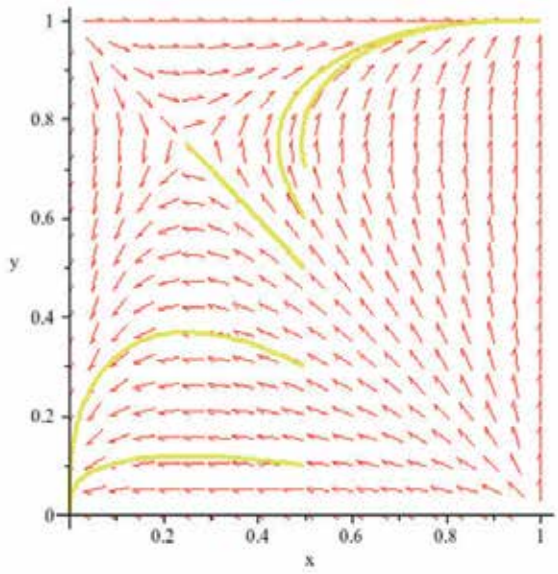


Fig. 7. Vector field plot.

Player's proportions on time's evolution are plotted in the next graphs:

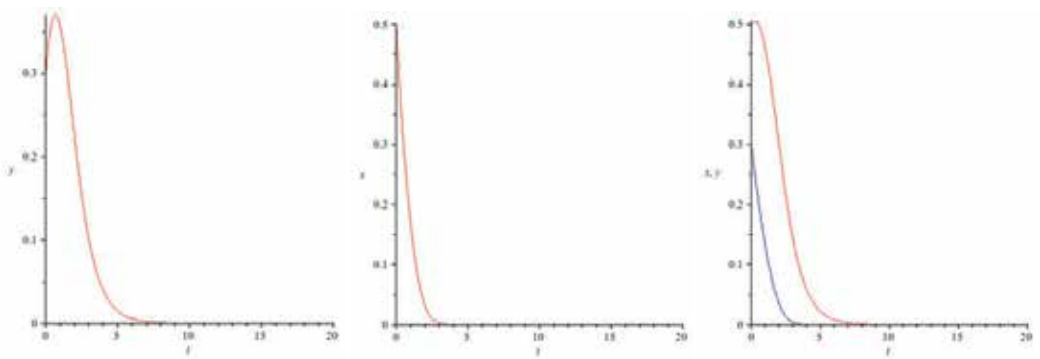


Fig. 8. a) Alice's proportion on time, b) Bob's strategy proportion, c) Alice's and Bob's time proportions.

From the previous three graphs is possible to observe the convergence. It is also presented a phase plane graph of the solutions:

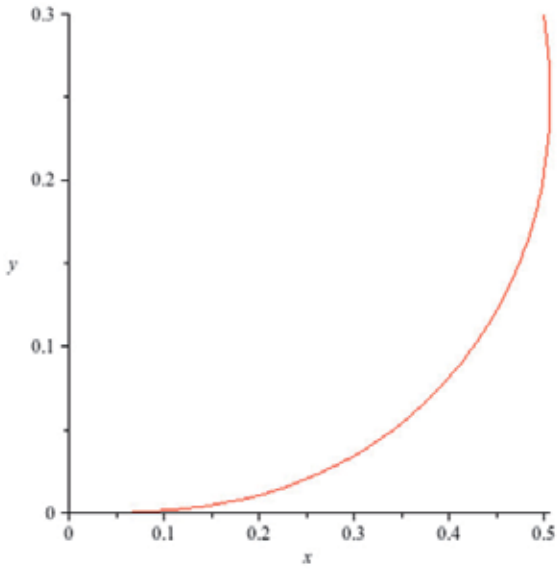


Fig. 9. Phase plane graph of the solution.

2.3 Burning money case

Allowing one of the players to "burn" money is really an interesting idea, remarking that both Alice and Bob are rational and pretend to have the best possible reward. In this work Alice will have the chance of burning money, while Bob plays only with selecting between opera or football match. The idea of burning money is presented as a game that helps understand weakly dominated strategies, modifying the normal game in order to let the players make a rational elimination by knowing that the other player eliminates strategies too. This case is presented in by Herbert Gintis (Gintis, 2009). The case proposed on section 2.2.2 can be taken as an un-burning money case.

2.3.1 Burning money case

The difference for this case is the possibility for one player (Alice) to "burn" money; that is to say, the player can have a negative payment by destroying some of her stuff, hence changing the strategies used by each player. In this case this option will be for Alice, who loses 2 points in each possible payment; while Bob keeps having the same payment:

$$A = \begin{bmatrix} 2 & -2 \\ -2 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}; \tag{29}$$

With the given matrices and the established method, equilibrium values for the system are found:

$$\{x = 0, y = 0\}, \{x = 1, y = 0\}, \{x = 0, y = 1\}, \left\{x = \frac{4}{5}, y = \frac{1}{5}\right\}, \{x = 1, y = 1\}. \quad (30)$$

Simplifying by (10) from the replicator's equation obtained in the method, we get:

$$\frac{d}{dt}\,\varepsilon(t) = \frac{4}{5}\,\eta(t), \qquad \frac{d}{dt}\,\eta(t) = \frac{4}{5}\,\varepsilon(t); \quad (31)$$

The solutions for this system are:

$$\left\{\varepsilon(t) = C_1\, e^{-\frac{4}{5}t} + C_2\, e^{\frac{4}{5}t}, \eta(t) = -C_1\, e^{-\frac{4}{5}t} + C_2\, e^{\frac{4}{5}t}\right\} \quad (32)$$

With the results obtained, it is possible to make a stability analysis and generate a vector field plot, which led us observe the specific point where the solution converges:

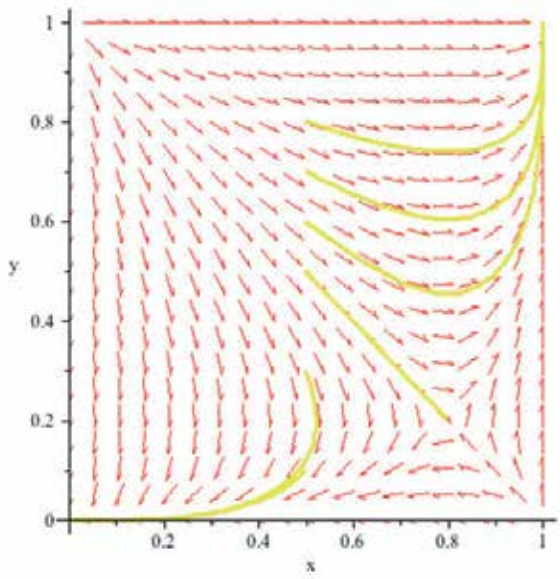


Fig. 10. Vector field plot

Plotting the proportions obtained over time, we can see that both Alice's and Bob's converge.

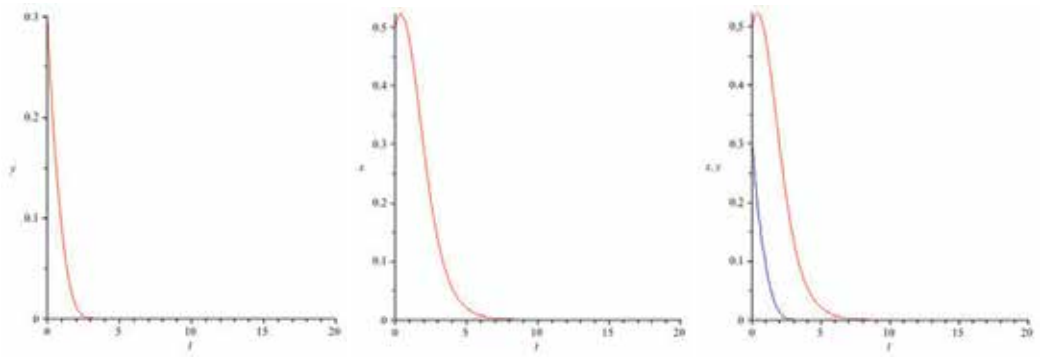


Fig. 11. a) Numerical solution for Alice's, b) Numerical solution for Bob's, c) Alice's and Bob's results plotted together with an extended time.

The three previous graphs are the time series for the player's proportions. The following graph is the phase plane for x and y:

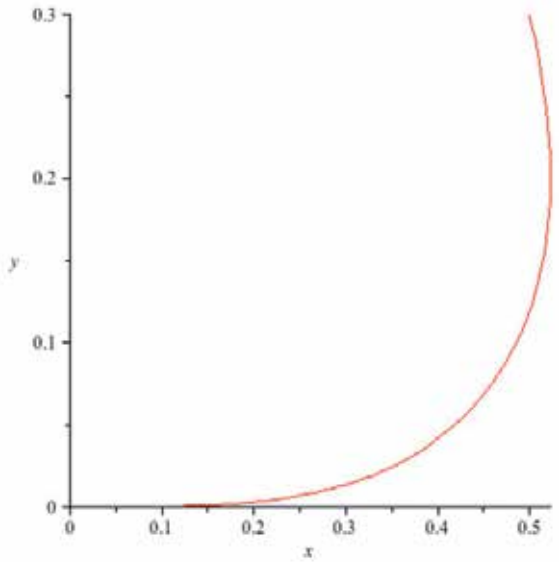


Fig. 12. Phase plane graph

3. Quantum game

Quantum mechanics has been applied to different games such as Prissioner Dilema (Eisert et al., 1999), hawk and dove game (López, 2010) and battle of the sexes game (Frackiewicz, 2009). Several versions for the battle of the sexes game have been proposed from quantum game theory (Neto, 2008; Du et al., 2001; Frackiewicz, 2009), however the conception of this work is to present the wider variety of strategies that a classical analysis. The method implemented is shown using quantum circuits, where operators can be understood as

quantum gates that transform the initial states in order to obtain the payoff matrices for both Alice and Bob. Three possible scenarios are presented, however only the first one is used in this work:

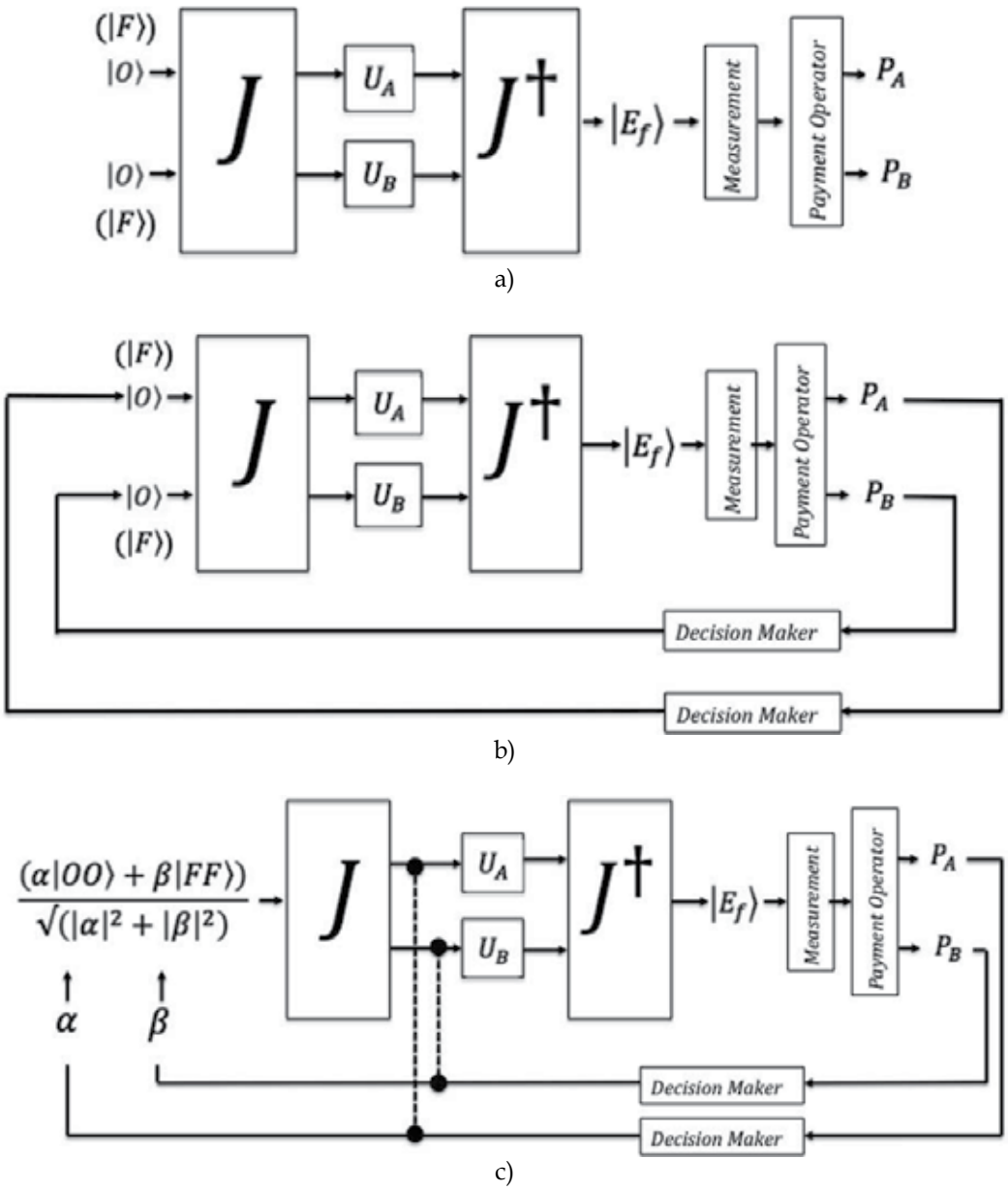


Fig. 13. Quantum circuit for quantum Battle of the Sexes Game: a) open loop case with non-entangled initial states, b) closed loop with non-entangled initial states, c) closed loop with entangled initial states.

3.1 Method

The method that will be used through this section is a formalization of the quantum circuit shown in Figure 13a), which corresponds to the open loop case with non-entangled initial states. The quantum gate named J in Figure 13a), has the form:

$$J = e^{\frac{i\Gamma}{2} C \otimes C}, \tag{33}$$

where

$$C = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, C \otimes C = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \frac{i\Gamma}{2} C \otimes C = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} I\,\Gamma \\ 0 & 0 & -\frac{1}{2} I\,\Gamma & 0 \\ 0 & -\frac{1}{2} I\,\Gamma & 0 & 0 \\ \frac{1}{2} I\,\Gamma & 0 & 0 & 0 \end{bmatrix}. \tag{34}$$

Then, J takes the form:

$$J = \begin{bmatrix} \cos\left(\frac{1}{2}\Gamma\right) & 0 & 0 & I\sin\left(\frac{1}{2}\Gamma\right) \\ 0 & \cos\left(\frac{1}{2}\Gamma\right) & -I\sin\left(\frac{1}{2}\Gamma\right) & 0 \\ 0 & -I\sin\left(\frac{1}{2}\Gamma\right) & \cos\left(\frac{1}{2}\Gamma\right) & 0 \\ I\sin\left(\frac{1}{2}\Gamma\right) & 0 & 0 & \cos\left(\frac{1}{2}\Gamma\right) \end{bmatrix} \tag{35}$$

The possible strategies are represented by the following Kets:

$$|FF\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, |FO\rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, |OF\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, |OO\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{36}$$

where, F refers to Football and O to Opera.
The action of the quantum gate J over the Kets $|FF\rangle$ and $|OO\rangle$ is:

$$J \cdot |FF\rangle = \begin{bmatrix} \cos\left(\frac{1}{2}\Gamma\right) \\ 0 \\ 0 \\ I\sin\left(\frac{1}{2}\Gamma\right) \end{bmatrix}, J \cdot |OO\rangle = \begin{bmatrix} I\sin\left(\frac{1}{2}\Gamma\right) \\ 0 \\ 0 \\ cos\left(\frac{1}{2}\Gamma\right) \end{bmatrix} \tag{37}$$

The quantum gates associated with Alice and Bob are given respectively by:

$$U_A = \begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix} \quad , \quad U_B = \begin{bmatrix} c & d \\ -\bar{d} & \bar{c} \end{bmatrix} \tag{38}$$

and the tensor product between these, is:

$$U_A \otimes U_B = \begin{bmatrix} ac & ad & bc & bd \\ -a\bar{d} & a\bar{c} & -b\bar{d} & b\bar{c} \\ -\bar{b}c & -\bar{b}\bar{d} & \bar{a}c & \bar{a}d \\ \bar{b}\bar{d} & -\bar{b}\bar{c} & -\bar{a}\bar{d} & \bar{a}\bar{c} \end{bmatrix} \tag{39}$$

The action of combined quantum gate (39), over the Kets presented at (37) is:

$$(U_A \otimes U_B)\,(J \cdot |FF\rangle) = \begin{bmatrix} ac\cos\left(\frac{1}{2}\Gamma\right) + I\ bd\ \sin\left(\frac{1}{2}\Gamma\right) \\ -a\bar{d}\cos\left(\frac{1}{2}\Gamma\right) + I\ b\bar{c}\ \sin\left(\frac{1}{2}\Gamma\right) \\ -\bar{b}c\cos\left(\frac{1}{2}\Gamma\right) + I\ \bar{a}d\ \sin\left(\frac{1}{2}\Gamma\right) \\ \bar{b}\bar{d}\cos\left(\frac{1}{2}\Gamma\right) + I\ \bar{a}\bar{c}\ \sin\left(\frac{1}{2}\Gamma\right) \end{bmatrix} \tag{40}$$

$$(U_A \otimes U_B)\,(J \cdot |OO\rangle) = \begin{bmatrix} I\ ac\sin\left(\frac{1}{2}\Gamma\right) + \ bd\ \cos\left(\frac{1}{2}\Gamma\right) \\ -I\ a\bar{d}\sin\left(\frac{1}{2}\Gamma\right) + \ b\bar{c}\ \cos\left(\frac{1}{2}\Gamma\right) \\ -I\ \bar{b}c\sin\left(\frac{1}{2}\Gamma\right) + \ \bar{a}d\ \cos\left(\frac{1}{2}\Gamma\right) \\ I\ \bar{b}\bar{d}\sin\left(\frac{1}{2}\Gamma\right) + \ \bar{a}\bar{c}\ \cos\left(\frac{1}{2}\Gamma\right) \end{bmatrix} \tag{41}$$

According with the Figure 13a), the next step is to take the Hermitian transpose of the quantum gate J, and the result is the following:

$$J^{\dagger} = \begin{bmatrix} \cos\left(\frac{1}{2}\bar{\Gamma}\right) & 0 & 0 & -I\sin\left(\frac{1}{2}\bar{\Gamma}\right) \\ 0 & \cos\left(\frac{1}{2}\bar{\Gamma}\right) & I\sin\left(\frac{1}{2}\bar{\Gamma}\right) & 0 \\ 0 & I\sin\left(\frac{1}{2}\bar{\Gamma}\right) & \cos\left(\frac{1}{2}\bar{\Gamma}\right) & 0 \\ -I\sin\left(\frac{1}{2}\bar{\Gamma}\right) & 0 & 0 & \cos\left(\frac{1}{2}\bar{\Gamma}\right) \end{bmatrix} \tag{42}$$

The last step in the quantum circuit is the action of the quantum gate shown in (42) over the Kets given at (40) and (41):

$$J^{\dagger} \cdot ((U_A \otimes U_B)(J \cdot |FF))$$

$$= \begin{bmatrix} \cos\left(\frac{1}{2}\bar{\Gamma}\right)\left(ac\cos\left(\frac{1}{2}\Gamma\right) + I\ bd\ \sin\left(\frac{1}{2}\Gamma\right)\right) - I\sin\left(\frac{1}{2}\bar{\Gamma}\right)\left(\overline{bd}\cos\left(\frac{1}{2}\Gamma\right) + I\ \overline{ac}\ \sin\left(\frac{1}{2}\Gamma\right)\right) \\ \cos\left(\frac{1}{2}\bar{\Gamma}\right)\left(-a\bar{d}\cos\left(\frac{1}{2}\Gamma\right) + I\ b\bar{c}\ \sin\left(\frac{1}{2}\Gamma\right)\right) + I\sin\left(\frac{1}{2}\bar{\Gamma}\right)\left(-\bar{b}c\cos\left(\frac{1}{2}\Gamma\right) + I\ \bar{a}d\ \sin\left(\frac{1}{2}\Gamma\right)\right) \\ I\sin\left(\frac{1}{2}\bar{\Gamma}\right)\left(-a\bar{d}\cos\left(\frac{1}{2}\Gamma\right) + I\ b\bar{c}\ \sin\left(\frac{1}{2}\Gamma\right)\right) + \cos\left(\frac{1}{2}\bar{\Gamma}\right)\left(-\bar{b}c\cos\left(\frac{1}{2}\Gamma\right) + I\ \bar{a}d\ \sin\left(\frac{1}{2}\Gamma\right)\right) \\ -I\sin\left(\frac{1}{2}\bar{\Gamma}\right)\left(ac\cos\left(\frac{1}{2}\Gamma\right) + I\ bd\ \sin\left(\frac{1}{2}\Gamma\right)\right) \cos\left(\frac{1}{2}\bar{\Gamma}\right)\left(\overline{bd}\cos\left(\frac{1}{2}\Gamma\right) + I\ \overline{ac}\ \sin\left(\frac{1}{2}\Gamma\right)\right) \end{bmatrix} \tag{43}$$

$$J^{\dagger}\cdot((U_A\otimes U_B)\,(J\cdot|OO\rangle))$$
$$=\begin{bmatrix} \cos\left(\frac{1}{2}\bar{\Gamma}\right)\left(I\,ac\sin\left(\frac{1}{2}\Gamma\right)+\,bd\,\cos\left(\frac{1}{2}\Gamma\right)\right) & -I\sin\left(\frac{1}{2}\bar{\Gamma}\right)\left(I\,\bar{b}\bar{d}\sin\left(\frac{1}{2}\Gamma\right)+\,\bar{a}\bar{c}\,\cos\left(\frac{1}{2}\Gamma\right)\right) \\ \cos\left(\frac{1}{2}\bar{\Gamma}\right)\left(-I\,a\bar{d}\sin\left(\frac{1}{2}\Gamma\right)+\,b\bar{c}\,\cos\left(\frac{1}{2}\Gamma\right)\right)+I\sin\left(\frac{1}{2}\bar{\Gamma}\right)\left(-I\,\bar{b}c\sin\left(\frac{1}{2}\Gamma\right)+\,\bar{a}d\,\cos\left(\frac{1}{2}\Gamma\right)\right) \\ I\sin\left(\frac{1}{2}\bar{\Gamma}\right)\left(-I\,a\bar{d}\sin\left(\frac{1}{2}\Gamma\right)+\,b\bar{c}\,\cos\left(\frac{1}{2}\Gamma\right)\right)+\cos\left(\frac{1}{2}\bar{\Gamma}\right)\left(-I\,\bar{b}c\sin\left(\frac{1}{2}\Gamma\right)+\,\bar{a}d\,\cos\left(\frac{1}{2}\Gamma\right)\right) \\ -\sin\left(\frac{1}{2}\bar{\Gamma}\right)\left(I\,ac\sin\left(\frac{1}{2}\Gamma\right)+\,bd\,\cos\left(\frac{1}{2}\Gamma\right)\right) & \cos\left(\frac{1}{2}\bar{\Gamma}\right)\left(I\,\bar{b}\bar{d}\sin\left(\frac{1}{2}\Gamma\right)+\,\bar{a}\bar{c}\,\cos\left(\frac{1}{2}\Gamma\right)\right) \end{bmatrix} \quad (44)$$

Finally a measurement is made over the final states (43) and (44), from which is possible to build the payment operators and the expected payments for Alice and Bob.
In figures 13b) and 13c), the quantum circuits are depicted as closed loops, where the feedback is implemented by special quantum gates that assume the role of "decision makers", it is to say, special quantum gates that transform the expected payments into new possible initial states and new quantum gates: U_A and U_B.

3.2 Simple cases

In this section some models presented by J.J. de Farias Neto in his work titled "Quantum Battle of the Sexes Revisited". These cases are a combination of different values for γ and different initial states, either non-entangled or entangled. In the first case, $\gamma = 0$ and the initial state is non-entangled. It will be initially made for $|OO\rangle$:

$$\left|E_{OO}\right. = (U_A \otimes\ U_B)\cdot\left|OO\right. \quad (45)$$

Replacing values from equation (44), we obtain:

$$\left|E_{OO}\right. = \begin{bmatrix} bd \\ b\bar{c} \\ \bar{a}d \\ \overline{ac} \end{bmatrix} \quad (46)$$

The resulting payoffs for Alice and Bob are:

$$P_A = \left|b^2d^2\right| + 2\left|a^2c^2\right|$$
$$P_B = 2\left|b^2d^2\right| + \left|a^2c^2\right| \quad (47)$$

Likewise for $|FF\rangle$, the final state is given by:

$$\left|E_{FF}\right\rangle = (U_A \otimes\ U_B)\cdot\left|FF\right. \quad (48)$$

Substituting equation (43), we get:

$$|E_{FF}\rangle = \begin{bmatrix} ac \\ -a\bar{d} \\ -\bar{b}c \\ \bar{b}\bar{d} \end{bmatrix} \tag{49}$$

The outcomes in this case are:

$$\begin{aligned} P_A &= 2|b^2d^2| + |a^2c^2| \\ P_B &= |b^2d^2| + 2|a^2c^2| \end{aligned} \tag{50}$$

The second model maintains $\gamma = 0$, but the initial state is now entangled; hence the $|E_i\rangle$ has the form:

$$|E_i\rangle = \frac{1}{2}\sqrt{2}(\,|FF\rangle + |OO\rangle) \tag{51}$$

Replacing values for $|FF\rangle$ and $|OO\rangle$, we obtain the initial state:

$$|E_i\rangle = \begin{bmatrix} \frac{1}{2}\sqrt{2} \\ 0 \\ 0 \\ \frac{1}{2}\sqrt{2} \end{bmatrix} \tag{52}$$

The final state is given by the following equation:

$$\left|E_f\right\rangle = (U_A \otimes U_B)\ \cdot |E_i\rangle \tag{53}$$

Substituting, the final state has the form:

$$\left|E_f\right\rangle = \begin{bmatrix} \frac{1}{2}\sqrt{2}ac + \frac{1}{2}\sqrt{2}bd \\ -\frac{1}{2}\sqrt{2}a\bar{d} + \frac{1}{2}\sqrt{2}b\bar{c} \\ -\frac{1}{2}\sqrt{2}\bar{b}c + \frac{1}{2}\sqrt{2}\bar{a}d \\ \frac{1}{2}\sqrt{2}\bar{b}\bar{d} + \frac{1}{2}\sqrt{2}\bar{a}\bar{c} \end{bmatrix} \tag{54}$$

Alice and Bob's payoffs are, respectively

$$\begin{aligned} P_A &= \frac{3}{2}(ac + bd)\left(\bar{a}\bar{c} + \bar{b}\bar{d}\right) \\ P_B &= \frac{3}{2}(ac + bd)\left(\bar{a}\bar{c} + \bar{b}\bar{d}\right) \end{aligned} \tag{55}$$

Another case is proposed by J.J. de Farias, where he defines a new initial state:

$$|E_i\rangle = \frac{1}{2}\sqrt{2}(\,I|FF\rangle + |OO\rangle) \tag{56}$$

By replacing values for the initial state previously given, we obtain:

$$|E_i\rangle = \frac{1}{2}\sqrt{2}\left(I\begin{bmatrix}1\\0\\0\\0\end{bmatrix} + \begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right) \tag{57}$$

In (53), we replace the initial state, obtaining:

$$|E_f\rangle = \begin{bmatrix} I\frac{1}{2}\sqrt{2}ac + \frac{1}{2}\sqrt{2}bd \\ -I\frac{1}{2}\sqrt{2}a\bar{d} + \frac{1}{2}\sqrt{2}b\bar{c} \\ -I\frac{1}{2}\sqrt{2}\bar{b}c + \frac{1}{2}\sqrt{2}\bar{a}d \\ I\frac{1}{2}\sqrt{2}\bar{b}\bar{d} + \frac{1}{2}\sqrt{2}\bar{a}\bar{c} \end{bmatrix} \tag{58}$$

The outcome for Alice and Bob are:

$$P_A = -\frac{3}{2}(I\bar{b}\bar{d} + \bar{a}\bar{c})(-ac + Ibd)$$

$$P_B = -\frac{3}{2}(I\bar{b}\bar{d} + \bar{a}\bar{c})(-ac + Ibd) \tag{59}$$

The following case is a modification of the previous one. Here, the initial state is modified and has the form:

$$|E_i\rangle = \frac{1}{2}\sqrt{2}(\,|FF\rangle + I|OO\rangle) \tag{60}$$

By replacing the respective Kets, we get:

$$|E_i\rangle = \frac{1}{2}\sqrt{2}\left(\begin{bmatrix}1\\0\\0\\0\end{bmatrix} + I\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right) \tag{61}$$

The final state of the system is given by:

$$|E_f\rangle = \begin{bmatrix} \frac{1}{2}\sqrt{2}ac + I\frac{1}{2}\sqrt{2}bd \\ -\frac{1}{2}\sqrt{2}a\bar{d} + I\frac{1}{2}\sqrt{2}b\bar{c} \\ -\frac{1}{2}\sqrt{2}\bar{b}c + I\frac{1}{2}\sqrt{2}\bar{a}d \\ \frac{1}{2}\sqrt{2}\bar{b}\bar{d} + I\frac{1}{2}\sqrt{2}\bar{a}\bar{c} \end{bmatrix} \tag{62}$$

Alice and Bob's rewards are shown ahead:

$$P_A = -\frac{3}{2}\left(I\bar{b}\bar{d} + \bar{a}\bar{c}\right)(-ac + Ibd)$$
$$P_B = -\frac{3}{2}\left(I\bar{b}\bar{d} + \bar{a}\bar{c}\right)(-ac + Ibd) \tag{63}$$

3.2.1 Dawkins model

The same model for the classical analysis is presented, and so are the describing matrices:

$$M_A = \begin{bmatrix} 2 & 0 \\ 5 & -5 \end{bmatrix} \tag{64}$$

$$M_B = \begin{bmatrix} 2 & 5 \\ 0 & 15 \end{bmatrix} \tag{65}$$

Payments for Alice and Bob are:

$$\begin{aligned} P_A = {} & -5\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a^2 d^2 \cos\left(\tfrac{1}{2}\Gamma\right)^2 + 2\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2 \bar{a}\,\bar{c} - 5\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 \bar{a}^2 d^2 \sin\left(\tfrac{1}{2}\Gamma\right)^2 - 2\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b^2 a^2 \cos\left(\tfrac{1}{2}\Gamma\right)^2 \\ & + 5\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a^2 c^2 \sin\left(\tfrac{1}{2}\Gamma\right)^2 - 5\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2 b\,d + 4\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) b\,d\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) + 2\sin\left(\tfrac{1}{2}\Gamma\right)\bar{a}^2\bar{c}^2\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) - 10\sin\left(\tfrac{1}{2}\Gamma\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\bar{a}\,\bar{c}\cos\left(\tfrac{1}{2}\Gamma\right) - 5\cos\left(\tfrac{1}{2}\Gamma\right)\bar{b}^2\bar{a}^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right) \\ & + 10\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,d\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right) d\,\bar{a}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) + 5\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b\,\bar{c}\sin\left(\tfrac{1}{2}\Gamma\right)^2 c\,b + 5\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) b^2\bar{c}^2\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) \\ & + 5\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) + 5\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,\bar{c} + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,d\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,\bar{c}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b \\ & c\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) a\,d - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b\,c\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right) d\,a - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2 c\,b\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) \\ & + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right) - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,c\sin\left(\tfrac{1}{2}\Gamma\right)^2 b\,d\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b\,d\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right) a\,c + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,d - 8\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) a\,c + 8\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b\,\bar{d}\cos\left(\tfrac{1}{2}\Gamma\right) a\,c\sin\left(\tfrac{1}{2}\Gamma\right) - 8\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) a \\ & \bar{c}\sin\left(\tfrac{1}{2}\Gamma\right)^2\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,d + 8\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,\bar{d}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) - 2\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b^2 d^2\sin\left(\tfrac{1}{2}\Gamma\right)^2 + 5\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)^2 + 2\cos\left(\tfrac{1}{2}\Gamma\right) a^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right) + 2\sin\left(\tfrac{1}{2}\Gamma\right)^2 \bar{a}\,\bar{c}\sin\left(\tfrac{1}{2}\Gamma\right)^2 a\,c - 5\sin\left(\tfrac{1}{2}\Gamma\right) b^2 d^2\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) - 5\cos\left(\tfrac{1}{2}\Gamma\right)^2 \bar{b} \\ & \bar{d}\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,d - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right)\bar{b}\,\bar{d}\cos\left(\tfrac{1}{2}\Gamma\right)^2\bar{a}\,\bar{c}\cos\left(\tfrac{1}{2}\Gamma\right) + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right)^2 a\,c\cos\left(\tfrac{1}{2}\Gamma\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right) - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right)^2\bar{a}\,\bar{c}\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) \\ & \bar{b}\,\bar{d} + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2\sin\left(\tfrac{1}{2}\Gamma\right) a\,c \end{aligned} \tag{66}$$

$$\begin{aligned} P_B = {} & 2\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2 \bar{a}\,\bar{c} - 2\sin\left(\tfrac{1}{2}\Gamma\right)^2 \bar{b}^2\bar{d}^2\cos\left(\tfrac{1}{2}\Gamma\right)^2 - 15\cos\left(\tfrac{1}{2}\Gamma\right)^2 \bar{a}^2\bar{c}^2\sin\left(\tfrac{1}{2}\Gamma\right)^2 + 15\sin\left(\tfrac{1}{2}\Gamma\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2 \bar{b}\,\bar{d} + 4\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,d\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)\bar{b}\,\bar{d}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) + 2\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)\bar{a}^2\bar{c}^2\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) + 30\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,c\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)\bar{a}\,\bar{c}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) + 15\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b^2 a^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,c\sin\left(\tfrac{1}{2}\Gamma\right)^2 a\,d\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2 d\,a\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 \\ & \bar{a}\,d\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) c\,\bar{b} - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a\,\bar{d}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,\bar{c} - 15\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b\,d\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,c\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) + 15\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)\bar{a}\,\bar{c}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 \bar{b} \\ & \bar{d}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) + 15\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right)^2 \bar{b}\,\bar{d}\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\bar{a}\,\bar{c} - 15\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2\cos\left(\tfrac{1}{2}\Gamma\right) b\,d + 12\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,d\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)\bar{a}\,\bar{c} \\ & - 12\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 \bar{b}\,\bar{a}\cos\left(\tfrac{1}{2}\Gamma\right) a\,c\sin\left(\tfrac{1}{2}\Gamma\right) + 12\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,c\sin\left(\tfrac{1}{2}\Gamma\right)^2\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,d - 12\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2 \bar{b}\,\bar{a}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) + 5\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a \\ & \bar{d}\cos\left(\tfrac{1}{2}\Gamma\right)^2 d\,\bar{a} + 5\cos\left(\tfrac{1}{2}\Gamma\right) a^2\bar{d}^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right) + 5\sin\left(\tfrac{1}{2}\Gamma\right)\bar{a}^2 d^2\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) + 5\sin\left(\tfrac{1}{2}\Gamma\right)^2\bar{a}\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2 a\,\bar{a} \\ & - 2\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b^2 d^2\sin\left(\tfrac{1}{2}\Gamma\right)^2 - 15\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)^2 + 2\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) a^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)\sin\left(\tfrac{1}{2}\Gamma\right) + 2\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a\,c\sin\left(\tfrac{1}{2}\Gamma\right)^2 a\,c \\ & + 15\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) b^2 d^2\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)\cos\left(\tfrac{1}{2}\Gamma\right) + 15\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 \bar{b}\,\bar{a}\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,d + 10\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,\bar{c}\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) c\,\bar{b}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) b \\ & d\cos\left(\tfrac{1}{2}\Gamma\right)^2 \bar{a}\,\bar{c}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 a\,c\cos\left(\tfrac{1}{2}\Gamma\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right) - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 \bar{a}\,\bar{c}\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) b\,\bar{a} + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\bar{\Gamma}\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2\sin\left(\tfrac{1}{2}\bar{\Gamma}\right) \\ & a\,c - 5\cos\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 b^2 c^2\sin\left(\tfrac{1}{2}\Gamma\right)^2 - 5\sin\left(\tfrac{1}{2}\bar{\Gamma}\right)^2 \bar{b}^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)^2 \end{aligned} \tag{67}$$

3.2.2 Wikipedia case 1

Like in section 2.2.2, the possible payments for both Alice and Bob has the form:

$$M_A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \tag{68}$$

$$M_B = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \tag{69}$$

From the method established, the payoff matrices for Alice and Bob are presented in an extended form:

$$\begin{aligned} P_A = {} & 3\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2 \overline{a}\,\overline{c} - 3\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b^2 d^2\cos\left(\tfrac{1}{2}\Gamma\right)^2 - 2\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 \overline{a}^2\,\overline{c}^2\sin\left(\tfrac{1}{2}\Gamma\right)^2 + 2\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2 b\,d \\ & + 6\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\overline{b}\,\overline{d}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) + 3\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)\overline{a}^2\,\overline{c}^2\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) + 4\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\overline{a} \\ & \overline{c}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) + 2\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)\overline{b}^2\,\overline{d}^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right) + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)\overline{a}\,\overline{c}\sin\left(\tfrac{1}{2}\Gamma\right)^2\overline{b} \\ & \overline{d}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2\overline{b}\,\overline{d}\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\overline{a}\,\overline{c} - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\,d + 3\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\,c\cos\left(\tfrac{1}{2}\Gamma\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right) \\ & - 3\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right)^2\overline{a}\,\overline{c}\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\overline{b}\,\overline{d} + 3\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2\sin\left(\tfrac{1}{2}\Gamma\right) a\,c - 3\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right)\overline{b}\,\overline{d}\cos\left(\tfrac{1}{2}\Gamma\right)^2\overline{a}\,\overline{c}\cos\left(\tfrac{1}{2}\Gamma\right) \\ & + 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\overline{a}\,\overline{c} - 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right)^2\overline{b}\,\overline{d}\cos\left(\tfrac{1}{2}\Gamma\right) a\,c\sin\left(\tfrac{1}{2}\Gamma\right) + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right)\overline{a}\,\overline{c}\sin\left(\tfrac{1}{2}\Gamma\right)^2\cos\left(\tfrac{1}{2}\Gamma\right) b\,d - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2\overline{b}\,\overline{d}\sin\left(\tfrac{1}{2}\Gamma\right) - 3\cos\left(\tfrac{1}{2}\Gamma\right)^2 b^2 d^2\sin\left(\tfrac{1}{2}\Gamma\right)^2 - 2\sin\left(\tfrac{1}{2}\Gamma\right)^2 a^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)^2 + 3\cos\left(\tfrac{1}{2}\Gamma\right) a^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right) \\ & + 3\sin\left(\tfrac{1}{2}\Gamma\right)^2\overline{a}\,\overline{c}\sin\left(\tfrac{1}{2}\Gamma\right)^2 a\,c + 2\sin\left(\tfrac{1}{2}\Gamma\right) b^2 d^2\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right) + 2\cos\left(\tfrac{1}{2}\Gamma\right)^2\overline{b}\,\overline{d}\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,d \end{aligned} \tag{70}$$

$$\begin{aligned} P_B = {} & 2\cos\left(\tfrac{1}{2}\Gamma\right)^2 a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2\overline{a}\,\overline{c} - 2\sin\left(\tfrac{1}{2}\Gamma\right)^2\overline{b}^2\,\overline{d}^2\cos\left(\tfrac{1}{2}\Gamma\right)^2 - 3\cos\left(\tfrac{1}{2}\Gamma\right)^2\overline{a}^2\,\overline{c}^2\sin\left(\tfrac{1}{2}\Gamma\right)^2 + 3\sin\left(\tfrac{1}{2}\Gamma\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2\overline{b}\,\overline{d} \\ & + 4\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\overline{b}\,\overline{d}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) + 2\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a^2\,c^2\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) + 6\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right) a \\ & \overline{c}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) + 3\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)\overline{b}^2\,\overline{d}^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) + 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\overline{a}\,\overline{c} - 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2\overline{b} \\ & \overline{d}\cos\left(\tfrac{1}{2}\Gamma\right) a\,c\sin\left(\tfrac{1}{2}\Gamma\right) + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right)\overline{a}\,\overline{c}\sin\left(\tfrac{1}{2}\Gamma\right)^2\cos\left(\tfrac{1}{2}\Gamma\right) b\,d - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right)\overline{a}\,\overline{c}\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,d \\ & - 3\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right)^2 b\,d\sin\left(\tfrac{1}{2}\Gamma\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right) + 3\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right)\overline{a}\,\overline{c}\sin\left(\tfrac{1}{2}\Gamma\right)^2\overline{b}\,\overline{d}\sin\left(\tfrac{1}{2}\Gamma\right) + 3\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right)^2\overline{b}\,\overline{d}\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\Gamma\right)\overline{a}\,\overline{c} - 3\,\mathrm{I}\sin\left(\tfrac{1}{2}\right. \\ & \left.\overline{\Gamma}\right) a\,c\cos\left(\tfrac{1}{2}\Gamma\right)^2\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\,d - 2\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b^2 d^2\sin\left(\tfrac{1}{2}\Gamma\right)^2 - 3\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)^2 + 2\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) a^2 c^2\cos\left(\tfrac{1}{2}\Gamma\right)\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)\sin\left(\tfrac{1}{2}\Gamma\right) \\ & + 2\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\,c\sin\left(\tfrac{1}{2}\Gamma\right)^2 a\,c + 3\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) b^2 d^2\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)\cos\left(\tfrac{1}{2}\Gamma\right) + 3\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\,d\cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,d - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) b \\ & \overline{d}\cos\left(\tfrac{1}{2}\Gamma\right)^2 a\,c\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\,c\cos\left(\tfrac{1}{2}\Gamma\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right) - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\,c\sin\left(\tfrac{1}{2}\Gamma\right)\cos\left(\tfrac{1}{2}\Gamma\right)\overline{b}\,\overline{d} + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\right. \\ & \left.\overline{\Gamma}\right) b\,d\sin\left(\tfrac{1}{2}\Gamma\right)^2\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\,c \end{aligned} \tag{71}$$

3.2.3 Wikipedia case 2

The values proposed in section 2.2.3 of the classical analysis are shown again for both Alice and Bob:

$$M_A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \tag{72}$$

$$M_B = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix} \tag{73}$$

The expected payoff for Alice and Bob are:

$$P_A = 3\cos\left(\frac{1}{2}\Gamma\right)^2 a\,c\cos\left(\frac{1}{2}\Gamma\right)^2 \overline{a}\,\overline{c} - 3\sin\left(\frac{1}{2}\Gamma\right)^2 \overline{b}^2\,\overline{d}^2\cos\left(\frac{1}{2}\Gamma\right)^2 - 2\cos\left(\frac{1}{2}\Gamma\right)^2 \overline{a}^2\,\overline{c}^2\sin\left(\frac{1}{2}\Gamma\right)^2 + 2\sin\left(\frac{1}{2}\Gamma\right)^2 b\,d\sin\left(\frac{1}{2}\Gamma\right)^2 \overline{b}\,\overline{d}$$
$$+ 6\cos\left(\frac{1}{2}\overline{\Gamma}\right) b\,d\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right)\overline{b}\,\overline{a}\sin\left(\frac{1}{2}\overline{\Gamma}\right) + 3\sin\left(\frac{1}{2}\overline{\Gamma}\right) a^2\,c^2\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\overline{\Gamma}\right) + 4\sin\left(\frac{1}{2}\overline{\Gamma}\right) a\,c\cos\left(\frac{1}{2}\overline{\Gamma}\right)\sin\left(\frac{1}{2}\Gamma\right) a$$
$$\overline{c}\cos\left(\frac{1}{2}\Gamma\right) + 2\cos\left(\frac{1}{2}\Gamma\right)\overline{b}^2\,\overline{d}^2\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right) - 2\,\mathrm{I}\sin\left(\frac{1}{2}\Gamma\right)^2 b\,d\sin\left(\frac{1}{2}\Gamma\right) a\,c\cos\left(\frac{1}{2}\Gamma\right) + 2\,\mathrm{I}\cos\left(\frac{1}{2}\Gamma\right)\overline{a}\,\overline{c}\sin\left(\frac{1}{2}\Gamma\right)^2\overline{b}$$
$$\overline{d}\sin\left(\frac{1}{2}\Gamma\right) + 2\,\mathrm{I}\cos\left(\frac{1}{2}\Gamma\right)^2\overline{b}\,\overline{d}\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)\overline{a}\,\overline{c} - 2\,\mathrm{I}\sin\left(\frac{1}{2}\Gamma\right) a\,c\cos\left(\frac{1}{2}\Gamma\right)^2\cos\left(\frac{1}{2}\Gamma\right) b\,d - \mathrm{I}\cos\left(\frac{1}{2}\Gamma\right)^2 a\,\overline{a}\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right) b\,\overline{c}$$
$$- \mathrm{I}\cos\left(\frac{1}{2}\overline{\Gamma}\right) b\,\overline{c}\sin\left(\frac{1}{2}\Gamma\right)^2 a\,d\sin\left(\frac{1}{2}\overline{\Gamma}\right) + \mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right) b\,c\cos\left(\frac{1}{2}\Gamma\right)^2 d\,\overline{a}\cos\left(\frac{1}{2}\overline{\Gamma}\right) + \mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right)^2\overline{a}\,d\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right) c\,b + \cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 a$$
$$\overline{d}\cos\left(\frac{1}{2}\Gamma\right)^2 d\,a + \cos\left(\frac{1}{2}\overline{\Gamma}\right) a^2\,\overline{d}^2\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\overline{\Gamma}\right) + \sin\left(\frac{1}{2}\overline{\Gamma}\right) a^2\,d^2\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\overline{\Gamma}\right) + \sin\left(\frac{1}{2}\overline{\Gamma}\right)^2 a\,d\sin\left(\frac{1}{2}\Gamma\right)^2 a\,\overline{d} \quad (74)$$
$$+ 3\,\mathrm{I}\cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 a\,c\cos\left(\frac{1}{2}\Gamma\right) b\,d\sin\left(\frac{1}{2}\Gamma\right) - 3\,\mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right)^2\overline{a}\,\overline{c}\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right)\overline{b}\,\overline{a} + 3\,\mathrm{I}\cos\left(\frac{1}{2}\overline{\Gamma}\right) b\,d\sin\left(\frac{1}{2}\Gamma\right)^2\sin\left(\frac{1}{2}\overline{\Gamma}\right) a\,c - 3\,\mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right)$$
$$\overline{b}\,\overline{d}\cos\left(\frac{1}{2}\Gamma\right)^2\overline{a}\,\overline{c}\cos\left(\frac{1}{2}\Gamma\right) - 3\cos\left(\frac{1}{2}\Gamma\right)^2 b^2\,d^2\sin\left(\frac{1}{2}\Gamma\right)^2 - 2\sin\left(\frac{1}{2}\Gamma\right)^2 a^2\,c^2\cos\left(\frac{1}{2}\Gamma\right)^2 + 3\cos\left(\frac{1}{2}\Gamma\right) a^2\,c^2\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)$$
$$+ 3\sin\left(\frac{1}{2}\overline{\Gamma}\right)^2\overline{a}\,\overline{c}\sin\left(\frac{1}{2}\Gamma\right)^2 a\,c + 2\sin\left(\frac{1}{2}\overline{\Gamma}\right) b^2\,d^2\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\overline{\Gamma}\right)\cos\left(\frac{1}{2}\Gamma\right) + 2\cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,d\cos\left(\frac{1}{2}\Gamma\right)^2 b\,d + 2\cos\left(\frac{1}{2}\overline{\Gamma}\right) b$$
$$c\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right) c\,b\sin\left(\frac{1}{2}\overline{\Gamma}\right) + 4\,\mathrm{I}\cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,d\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right) a\,c - 4\,\mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,d\cos\left(\frac{1}{2}\Gamma\right) a\,c\sin\left(\frac{1}{2}\Gamma\right) + 4\,\mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right) a$$
$$c\sin\left(\frac{1}{2}\Gamma\right)^2\cos\left(\frac{1}{2}\overline{\Gamma}\right) b\,d - 4\,\mathrm{I}\cos\left(\frac{1}{2}\overline{\Gamma}\right) a\,c\cos\left(\frac{1}{2}\Gamma\right)^2\overline{b}\,\overline{a}\sin\left(\frac{1}{2}\overline{\Gamma}\right) - \cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 b^2\,c^2\sin\left(\frac{1}{2}\Gamma\right)^2 - \sin\left(\frac{1}{2}\overline{\Gamma}\right)^2\overline{b}^2\,c^2\cos\left(\frac{1}{2}\Gamma\right)^2$$

$$P_B = -\sin\left(\frac{1}{2}\Gamma\right)^2 a^2\,\overline{d}^2\cos\left(\frac{1}{2}\Gamma\right)^2 + 2\cos\left(\frac{1}{2}\Gamma\right)^2 a\,c\cos\left(\frac{1}{2}\Gamma\right)^2\overline{a}\,\overline{c} - \cos\left(\frac{1}{2}\Gamma\right)^2\overline{a}^2\,d^2\sin\left(\frac{1}{2}\Gamma\right)^2 - 2\sin\left(\frac{1}{2}\Gamma\right)^2\overline{b}^2\,\overline{d}^2\cos\left(\frac{1}{2}\Gamma\right)^2$$
$$- 3\cos\left(\frac{1}{2}\overline{\Gamma}\right)^2\overline{a}^2\,\overline{c}^2\sin\left(\frac{1}{2}\Gamma\right)^2 + 3\sin\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,d\sin\left(\frac{1}{2}\Gamma\right)^2\overline{b}\,\overline{d} + 4\cos\left(\frac{1}{2}\overline{\Gamma}\right) b\,d\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right)\overline{b}\,\overline{a}\sin\left(\frac{1}{2}\overline{\Gamma}\right) + 2\sin\left(\frac{1}{2}\right.$$
$$\left.\Gamma\right)\overline{a}^2\,\overline{c}^2\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right) + 6\sin\left(\frac{1}{2}\Gamma\right) a\,c\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)\overline{a}\,\overline{c}\cos\left(\frac{1}{2}\Gamma\right) + 3\cos\left(\frac{1}{2}\Gamma\right)\overline{b}^2\,\overline{d}^2\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)$$
$$+ 2\sin\left(\frac{1}{2}\overline{\Gamma}\right) a\,\overline{d}\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right) d\,\overline{a}\cos\left(\frac{1}{2}\overline{\Gamma}\right) + \sin\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,\overline{c}\sin\left(\frac{1}{2}\Gamma\right)^2 c\,\overline{b} + \sin\left(\frac{1}{2}\overline{\Gamma}\right) b^2\,\overline{c}^2\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right) + \cos\left(\frac{1}{2}\right.$$
$$\left.\overline{\Gamma}\right) b^2\,c^2\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\overline{\Gamma}\right) + \cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,c\cos\left(\frac{1}{2}\Gamma\right)^2 b\,c - 3\,\mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,d\sin\left(\frac{1}{2}\Gamma\right) a\,c\cos\left(\frac{1}{2}\Gamma\right) + 3\,\mathrm{I}\cos\left(\frac{1}{2}\overline{\Gamma}\right) a\,c\sin\left(\frac{1}{2}\Gamma\right)^2 b$$
$$d\sin\left(\frac{1}{2}\overline{\Gamma}\right) + 3\,\mathrm{I}\cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,d\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right)\overline{a}\,\overline{c} - 3\,\mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right) a\,c\cos\left(\frac{1}{2}\Gamma\right)^2\cos\left(\frac{1}{2}\overline{\Gamma}\right) b\,d + \mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right) a\,d\cos\left(\frac{1}{2}\Gamma\right)^2 b\,\overline{c}\cos\left(\frac{1}{2}\overline{\Gamma}\right) \quad (75)$$
$$+ \mathrm{I}\sin\left(\frac{1}{2}\Gamma\right)^2 b\,\overline{c}\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right) a\,\overline{d} - \mathrm{I}\cos\left(\frac{1}{2}\Gamma\right)^2\overline{b}\,c\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\Gamma\right) d\,\overline{a} - \mathrm{I}\cos\left(\frac{1}{2}\Gamma\right)\overline{a}\,d\sin\left(\frac{1}{2}\Gamma\right)^2 c\,\overline{b}\sin\left(\frac{1}{2}\Gamma\right)$$
$$- 2\cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 b^2\,d^2\sin\left(\frac{1}{2}\Gamma\right)^2 - 3\sin\left(\frac{1}{2}\overline{\Gamma}\right)^2 a^2\,c^2\cos\left(\frac{1}{2}\Gamma\right)^2 + 2\cos\left(\frac{1}{2}\overline{\Gamma}\right) a^2\,c^2\cos\left(\frac{1}{2}\Gamma\right)\sin\left(\frac{1}{2}\overline{\Gamma}\right)\sin\left(\frac{1}{2}\Gamma\right) + 2\sin\left(\frac{1}{2}\overline{\Gamma}\right)^2\overline{a}\,\overline{c}\sin\left(\frac{1}{2}\Gamma\right)^2 a\,c$$
$$+ 3\sin\left(\frac{1}{2}\overline{\Gamma}\right) b^2\,\overline{d}^2\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\overline{\Gamma}\right)\cos\left(\frac{1}{2}\Gamma\right) + 3\cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,d\cos\left(\frac{1}{2}\Gamma\right)^2 b\,d + 4\,\mathrm{I}\cos\left(\frac{1}{2}\overline{\Gamma}\right)^2 b\,d\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right) a\,c - 4\,\mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right)^2 b$$
$$d\cos\left(\frac{1}{2}\Gamma\right) a\,c\sin\left(\frac{1}{2}\Gamma\right) + 4\,\mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right)\overline{a}\,\overline{c}\sin\left(\frac{1}{2}\Gamma\right)^2\cos\left(\frac{1}{2}\overline{\Gamma}\right) b\,d - 4\,\mathrm{I}\cos\left(\frac{1}{2}\overline{\Gamma}\right) a\,c\cos\left(\frac{1}{2}\Gamma\right)^2 b\,d\sin\left(\frac{1}{2}\overline{\Gamma}\right) - 2\,\mathrm{I}\sin\left(\frac{1}{2}\overline{\Gamma}\right) b\,d\cos\left(\frac{1}{2}\Gamma\right)^2\overline{a}$$
$$\overline{c}\cos\left(\frac{1}{2}\Gamma\right) + 2\,\mathrm{I}\cos\left(\frac{1}{2}\Gamma\right)^2 a\,c\cos\left(\frac{1}{2}\Gamma\right) b\,d\sin\left(\frac{1}{2}\Gamma\right) - 2\,\mathrm{I}\sin\left(\frac{1}{2}\Gamma\right)^2\overline{a}\,\overline{c}\sin\left(\frac{1}{2}\Gamma\right)\cos\left(\frac{1}{2}\Gamma\right)\overline{b}\,\overline{a} + 2\,\mathrm{I}\cos\left(\frac{1}{2}\Gamma\right) b\,d\sin\left(\frac{1}{2}\Gamma\right)^2\sin\left(\frac{1}{2}\Gamma\right) a\,c$$

3.3 Burning money case

This section has the same assumptions made for the classical analysis. The burning money case is analyzed in order to find the payment matrices for Alice and Bob. The un-burning money case is presented in section 3.2.2, but with different values as reward.

3.3.1 Burning money

Here, Alice can have a negative reward due to the assumption of burning money as a possible strategy that modifies the outcomes of the game, while Bob plays using only two strategies to choose from. This is shown on the following matrices:

$$M_A = \begin{bmatrix} 2 & -2 \\ -2 & -1 \end{bmatrix}, M_B = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \quad (76,77)$$

The payoffs for Alice and Bob are:

$$
\begin{aligned}
P_A = {} & 2\sin\left(\tfrac{1}{2}\Gamma\right)^2 a^2 \overline{a}^2 \cos\left(\tfrac{1}{2}\Gamma\right)^2 + 2\cos\left(\tfrac{1}{2}\Gamma\right)^2 a\, c \cos\left(\tfrac{1}{2}\Gamma\right)^2 \overline{a}\,\overline{c} + 2\cos\left(\tfrac{1}{2}\Gamma\right)^2 \overline{a}^2 d^2 \sin\left(\tfrac{1}{2}\Gamma\right)^2 - 2\sin\left(\tfrac{1}{2}\Gamma\right)^2 \overline{b}^2 \overline{d}^2 \cos\left(\tfrac{1}{2}\Gamma\right)^2 \\
& + \cos\left(\tfrac{1}{2}\Gamma\right)^2 \overline{a}^2 \overline{c}^2 \sin\left(\tfrac{1}{2}\Gamma\right)^2 - \sin\left(\tfrac{1}{2}\Gamma\right)^2 b\, d \sin\left(\tfrac{1}{2}\Gamma\right)^2 \overline{b}\,\overline{d} + 4\cos\left(\tfrac{1}{2}\Gamma\right) b\, d \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) \overline{b}\,\overline{d} \sin\left(\tfrac{1}{2}\Gamma\right) + 2\sin\left(\tfrac{1}{2}\right. \\
& \left.\overline{\Gamma}\right) a^2 c^2 \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) - 2\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, c \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) a\, c \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) - \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b^2 \overline{a}^2 \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) \\
& + 1\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\, d \sin\left(\tfrac{1}{2}\Gamma\right) a\, c \cos\left(\tfrac{1}{2}\Gamma\right) + 1\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, c \cos\left(\tfrac{1}{2}\Gamma\right)^2 \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\, d - 2\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\, d \cos\left(\tfrac{1}{2}\Gamma\right)^2 d\,\overline{a} - 4\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a \\
& \overline{a} \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) d\,\overline{a} \cos\left(\tfrac{1}{2}\Gamma\right) - 2\sin\left(\tfrac{1}{2}\Gamma\right)^2 b\,\overline{c} \sin\left(\tfrac{1}{2}\Gamma\right)^2 c\,\overline{b} - 2\sin\left(\tfrac{1}{2}\Gamma\right) b^2 \overline{c}^2 \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) - 2\cos\left(\tfrac{1}{2}\right. \\
& \left.\Gamma\right) \overline{b}^2 c^2 \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) - 2\cos\left(\tfrac{1}{2}\Gamma\right)^2 \overline{b}\, c \cos\left(\tfrac{1}{2}\Gamma\right)^2 b\,\overline{c} - 2\cos\left(\tfrac{1}{2}\Gamma\right) a^2 \overline{d}^2 \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) - 2\sin\left(\tfrac{1}{2}\right. \\
& \left.\overline{\Gamma}\right) \overline{a}^2 d^2 \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) - 2\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 \overline{a}\, d \sin\left(\tfrac{1}{2}\Gamma\right)^2 a\,\overline{d} + 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\, d \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) a\,\overline{c} - 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 \overline{b} \\
& \overline{a} \cos\left(\tfrac{1}{2}\Gamma\right) a\, c \sin\left(\tfrac{1}{2}\Gamma\right) + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, c \sin\left(\tfrac{1}{2}\Gamma\right)^2 \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\, d - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, c \cos\left(\tfrac{1}{2}\Gamma\right)^2 \overline{b}\,\overline{d} \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) - 2\cos\left(\tfrac{1}{2}\right. \\
& \left.\Gamma\right)^2 b^2 a^2 \sin\left(\tfrac{1}{2}\Gamma\right)^2 + \sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a^2 c^2 \cos\left(\tfrac{1}{2}\Gamma\right)^2 + 2\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) a^2 c^2 \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) \sin\left(\tfrac{1}{2}\Gamma\right) + 2\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\, c \sin\left(\tfrac{1}{2}\Gamma\right)^2 a\, c - \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) b^2 d^2 \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\right. \\
& \left.\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) - \cos\left(\tfrac{1}{2}\Gamma\right)^2 \overline{b}\,\overline{d} \cos\left(\tfrac{1}{2}\Gamma\right)^2 b\, d - 4\cos\left(\tfrac{1}{2}\Gamma\right) b\,\overline{c} \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) c\,\overline{b} \sin\left(\tfrac{1}{2}\Gamma\right) + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right) b\,\overline{c} \sin\left(\tfrac{1}{2}\Gamma\right)^2 a\,\overline{d} \sin\left(\tfrac{1}{2}\Gamma\right) \\
& - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right) \overline{b}\, c \cos\left(\tfrac{1}{2}\Gamma\right)^2 d\,\overline{a} \cos\left(\tfrac{1}{2}\Gamma\right) - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\Gamma\right)^2 \overline{a}\, d \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) c\,\overline{b} + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\Gamma\right)^2 a\,\overline{d} \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) b\,\overline{c} \\
& + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 \overline{b}\, c \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) d\, a + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, d \sin\left(\tfrac{1}{2}\Gamma\right)^2 c\,\overline{b} \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\,\overline{d} \cos\left(\tfrac{1}{2}\Gamma\right)^2 b\, c \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) \\
& - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\, c \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) a\, d - 1\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, c \sin\left(\tfrac{1}{2}\Gamma\right)^2 b\, d \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) - 1\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\, d \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) a\, c - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) \\
& b\, d \cos\left(\tfrac{1}{2}\Gamma\right)^2 \overline{a}\,\overline{c} \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\, c \cos\left(\tfrac{1}{2}\Gamma\right) b\, d \sin\left(\tfrac{1}{2}\Gamma\right) - 2\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 \overline{a}\,\overline{c} \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) b\, d + 2\,\mathrm{I}\cos\left(\tfrac{1}{2}\right. \\
& \left.\Gamma\right) b\, d \sin\left(\tfrac{1}{2}\Gamma\right)^2 \sin\left(\tfrac{1}{2}\Gamma\right) a\, c + 2\cos\left(\tfrac{1}{2}\Gamma\right)^2 b^2 \overline{c}^2 \sin\left(\tfrac{1}{2}\Gamma\right)^2 + 2\sin\left(\tfrac{1}{2}\Gamma\right)^2 \overline{b}^2 c^2 \cos\left(\tfrac{1}{2}\Gamma\right)^2
\end{aligned}
\tag{78}
$$

$$
\begin{aligned}
P_B = {} & \cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\, c \cos\left(\tfrac{1}{2}\Gamma\right)^2 a\, c - \sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 \overline{b}^2 \overline{d}^2 \cos\left(\tfrac{1}{2}\Gamma\right)^2 - 4\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a^2 c^2 \sin\left(\tfrac{1}{2}\Gamma\right)^2 + 4\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\, d \sin\left(\tfrac{1}{2}\Gamma\right)^2 \overline{b}\,\overline{d} \\
& + 1\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\, c \cos\left(\tfrac{1}{2}\Gamma\right) b\, d \sin\left(\tfrac{1}{2}\Gamma\right) + 2\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\, d \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) \overline{b}\,\overline{d} \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) + 1\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\, d \sin\left(\tfrac{1}{2}\Gamma\right)^2 \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, c - 1\,\mathrm{I}\sin\left(\tfrac{1}{2}\right. \\
& \left.\overline{\Gamma}\right) \overline{b}\,\overline{d} \cos\left(\tfrac{1}{2}\Gamma\right)^2 a\, c \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) + \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a^2 c^2 \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) - 1\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\, c \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) \overline{b}\,\overline{a} + 8\sin\left(\tfrac{1}{2}\right. \\
& \left.\overline{\Gamma}\right) a\, c \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) a\, c \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) + 4\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) \overline{b}^2 \overline{d}^2 \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) + 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\, d \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\Gamma\right) a\, c \\
& - 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 \overline{b}\,\overline{d} \cos\left(\tfrac{1}{2}\Gamma\right) a\, c \sin\left(\tfrac{1}{2}\Gamma\right) + 5\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, c \sin\left(\tfrac{1}{2}\Gamma\right)^2 \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\, d - 5\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, c \cos\left(\tfrac{1}{2}\Gamma\right)^2 \overline{b}\,\overline{d} \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) \\
& - \cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b^2 d^2 \sin\left(\tfrac{1}{2}\Gamma\right)^2 - 4\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a^2 c^2 \cos\left(\tfrac{1}{2}\Gamma\right)^2 - 4\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 b\, d \sin\left(\tfrac{1}{2}\Gamma\right) a\, c \cos\left(\tfrac{1}{2}\Gamma\right) + \cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a^2 c^2 \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\right. \\
& \left.\overline{\Gamma}\right) \sin\left(\tfrac{1}{2}\Gamma\right) + \sin\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 a\, c \sin\left(\tfrac{1}{2}\Gamma\right)^2 a\, c + 4\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) b^2 d^2 \sin\left(\tfrac{1}{2}\Gamma\right) \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) \cos\left(\tfrac{1}{2}\Gamma\right) + 4\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 \overline{b}\,\overline{a} \cos\left(\tfrac{1}{2}\Gamma\right)^2 b\, d + 4\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right) a \\
& c \sin\left(\tfrac{1}{2}\Gamma\right)^2 \overline{b}\,\overline{d} \sin\left(\tfrac{1}{2}\overline{\Gamma}\right) + 4\,\mathrm{I}\cos\left(\tfrac{1}{2}\overline{\Gamma}\right)^2 \overline{b}\,\overline{a} \cos\left(\tfrac{1}{2}\Gamma\right) \sin\left(\tfrac{1}{2}\Gamma\right) a\, c - 4\,\mathrm{I}\sin\left(\tfrac{1}{2}\overline{\Gamma}\right) a\, c \cos\left(\tfrac{1}{2}\Gamma\right)^2 \cos\left(\tfrac{1}{2}\overline{\Gamma}\right) b\, d
\end{aligned}
\tag{79}
$$

4. Conclusion

In this work, we made a classical and quantum analysis of the Battle of the Sexes game. The classical analysis was based on the model proposed by Frank Wang, using ordinary differential equations and the corresponding stability analysis; while the quantum scheme was based on the method introduced by Neto (Neto, 2008) using quantum circuits with quantum gates for two qubits. Both quantum and classical analysis were implemented using computer algebra software, specifically Maple.

From the results obtained along this work, it is possible to observe that the quantum version has more versatility than the classical one, due to the effect of quantum superposition, the

quantum entanglement and the complexity of the payment operators from which resulted expected payments with more engineering possibilities. It is to remark that the burning money case could be analyzed using different reasoning methods, which may lead to specific solutions such as guaranteeing other player's final choice (Ginitis, 2009).
As a future investigation trend we propose the application of the Yang-Baxter operators (Zhang et al., 2005), which will play the role of the quantum gate J, hence generating completely different but more physically implementable quantum operators and expected payoffs.

5. Acknowledgment

The author is very grateful with Prof. Juan Ospina for guiding this project; and also wants to thank Hernando Suarez for his collaboration.

6. References

Davis, M.D. (1970). *Game Theory: A Nontechnical Introduction*. Dover, New York.

Dawkins, R.(2006). *The Selfish Gene*. Oxford University Press, Oxford.

Du, J., Li, H., Xu, X., Shi, M., Zhou, X. & Han, R. (2001). Remark on Quantum Battle of the Sexes Game. Retrieved from http://arxiv.org/abs/quant-ph/0103004

Eisert, J., Wilkens, M. & Lewenstein, M. (1999). Quantum Games and Quantum Strategies. Physical Review Letters 83, 3077.

Eisert, J. & Wilkens, M. (2000). *Quantum Games*. J. Mod. Opt. 47, 2543.

Frackiewicz, P. (2009). *The Ultimate Solution to the Quantum Battle of the Sexes game*. Journal of Physics A: Mathematical and Theoretical. 42 365305

Ginitis, H. (2009). The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences. Princeton University Press, ISBN: 9781400830367, Princeton.

Glimcher, P. (2003). *Decisions, Uncertainty and the Brain*. MIT Press, Cambridge.

Hanauske, M., Kunz, J., Bernius, S. and König, W.(2009). *Doves and hawks in economics revisited. An evolutionary quantum game theory-based analysis of financial crises*. Quantitative Finance Papers 0904.2113, arXiv.org.

López, J.M. (2010). Hawk-Dove-Bully-Retaliator quantum game CAS aided. Quantum Information and Computation - SPIE Proceedings. April 2010, Orlando, FL.

Meyer, D. (1999). *Quantum Strategies*. Physical Review Letters 82, 1052.

Montague,P. R., & Berns, G. (2002). Neural Economics and the Biological Substrates of Valuation. *Neuron*, 36: 265–284.

Myerson, R.B. (1991). *Game Theory: An Analysis of Conflict*. MIT Press, Cambridge.

Neto, J.J.D.F. (2008). *Quantum Battle of the Sexes Revisited*. 12.23, p. 11.

Piotrowski, E. W. & Sladkowski, J. (2003). *An invitation to Quantum Game Theory*. Int.J.Theor.Phys. 42 (2003) 1089.

Von Neumann, J. & Morgenstern, O. (1947) *Theory of Games and Economic Behaviour*. Princeton University Press, Princeton.

Wang, F. (2010). Richard Dawkins' Battle of the Sexes Model, in Maplesoft Application Center. July 2011, available from: http://www.maplesoft.com/applications/view.aspx?SID=95974&view=html

Zhang, Y., Kauffman, L. & Ge, M.L. (2005). Yang-Baxterizations, Universal Quantum Gates and Hamiltonians. *Quantum Information Processing*. Vol. 4, No. 3. August 2005, pp. 159-197. ISSN: 1570-0755

16

Einstein-Bohr Controversy After 75 Years, Its Actual Solution and Consequences

Miloš V. Lokajíček
Institute of Physics of the AS CR, Prague
Czech Republic

1. Introduction

The contemporary fundamental physics has been based fully on phenomenological models starting from the Copenhagen quantum mechanics proposed by Bohr in 1927 (1). However, the Copenhagen model was criticized strongly by Einstein and collaborators (2) from the point of view of ontological interpretation of matter world, i.e., from the grounds on which the whole physical concept was based by Galileo and Newton. Einstein called the attention (with the help of a Gedankenexperiment) to the fact that immediate interaction between different matter objects (in principle between different measuring devices) at great mutual distances was required in Bohr's theory (the effect being unacceptable from the ontological point of view and denoted now usually as entanglement). However, the criticism of Einstein was refused strongly by Bohr (3). And the physical community has accepted fully Bohr's arguments (that the rules in microworld differ from those of macroworld) even if Einstein did not changed his standpoint till the end of his life, having been aware of the correctness of his conclusion (see also (4)).

Later (on the basis of Bohm's paper (5)) two different quantum alternatives were considered, i.e., the Copenhagen alternative and the so called hidden-variable theory. Some differences between these two alternatives started to be discussed when Bell (6) derived his famous inequalities. These inequalities should have been valid for some combinations of four probabilities of the passage of two particles with opposite spins through two polarizers; the original EPR (Einstein-Podolsky-Rosen) Gedankenexperiment having been slightly modified. To obtain the given inequalities Bell had to assume for individual probabilities to be rather strongly mutually correlated; and it has been believed that the given inequalities have been valid in the framework of the hidden-variable theory. It was, therefore, expected that their violation (or not) in an experiment might have helped in deciding between the two given quantum alternatives.

As to the original Gedankenexperiment of Einstein it was based on the coincidence detection measurement of two decay objects emitted in a decay of an unstable particle and going in opposite directions. The basic proposal of this experiment was then modified to be possible actually to perform it. And the coincidence transmission probabilities of two photons having opposite spins and running in opposite directions through two polarizers were being established:

$$\left\| < - - |^{\beta} - - - o - - - |^{\alpha} - - > \right\|$$

where the axes of individual polarizers were deviated by angles α and β from some common zero position. The corresponding measurements were performed in the seventieth years of the past century and ended practically in 1982 (see (7)) with the following results:
- the inequalities of Bell's combinations of coincidence probabilities have been violated;
- the measured values of individual coincidence probabilities have fulfilled approximately the Malus law (see Sec. 4 of this chapter);
and both the results have been interpreted as decisive support for the Copenhagen quantum mechanics.
However, the given conclusion has been based on several mistakes. Both the quantum alternatives (starting from the same Schrödinger equation) have been regarded as two different interpretations of one common theoretical model. In fact each of these interpretations has been based on discrepant sets of assumptions, representing two fully diverse theories (see Sec. 2).
It has been then possible to show further that in the interpretation of the experimental results gained on the basis of the EPR experiments mainly two mistakes have played important roles (8) (see also (9; 10)):

(i) It has been believed all the time that Bell's inequalities have held in the framework of hidden-variable theory. However, they were based on the assumption holding only in the framework of classical physics, which means that no quantum alternative may be excluded on the basis of the results of EPR experiments (see Sec. 3).

(ii) As to the measured Malus law it was argued by Belinfante (11) that a very different result should be obtained in the performed coincidence EPR experiments. However, also this argument has been based on mistaking assumption (see Sec. 4).

It follows, therefore, that contrary to common conviction any of two quantum alternatives cannot be excluded on the basis of the results of EPR experiments. Even if the hidden-variable theory should be at least partially preferred (the lower limit not having been overpassed, see Sec. 3), one must look for other ways how to decide between them.
And it has been possible to demonstrate that the hidden-variable theory being practically equivalent to the Schrödinger equation (Bohr's deforming assumptions not having been added) should be preferred not only to the Copenhagen alternative but also to the classical physics:

(i) The Copenhagen quantum mechanics is to be practically excluded as it involves some important internal contradictions (as it will be shown in Sec. 5) that have been interpreted until now as quantum paradoxes.

(ii) There are also some experimental data that are in clear contradiction to the predictions of Copenhagen quantum mechanics; see Sec. 6.

(iii) And finally, the basic solutions of Schrödinger equation (determined always with the help of one Hamiltonian eigenfunction only) give the same results as classical physics, which prefers this equation (or hidden-variable theory) to be applied to the whole matter reality (microscopic as well as macroscopic); see Sec. 7.

The physical interpretation of Schrödinger equation was strongly influenced by the requirements put by Bohr on the shape of corresponding Hilbert space. If the physical content is to remain undeformed the Hilbert space must consist at least of two individual subspaces being bound by evolution operator as it was shown already by Lax and Phillips (12; 13) in 1967 and as it will be discussed in Sec. 8.

The evolution of physics in the last centuries was strongly influenced by the philosophical ideas that started to influence any thinking of human society from the beginning of the modern age and differed fundamentally from the earlier ontological thinking. The corresponding story will be mentioned in Sec. 9.
Main conclusions and consequences will be then summarized in the last section (Sec. 10).

2. Two quantum alternatives and corresponding assumptions

It is possible to say that the Copenhagen quantum mechanics has been based in principle on four following assumptions:

- first, it is the validity of time-dependent Schrödinger equation (14)

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = H\psi(x,t), \quad H = -\frac{\hbar^2}{2m}\triangle + V(x) \tag{1}$$

  where Hamiltonian H represents the total (kinetic and potential) energy of a given physical system and x represents the coordinates of all matter objects;

- physical quantities are expressed as the mean values of corresponding operators at time t:

$$A(t) = \int \psi^*(x,t)\, A_{op}\, \psi(x,t)dx \tag{2}$$

  where functions $\psi(x,t)$ represent vectors and A_{op} operators in some Hilbert space, the structure of which having been left unspecified (in the framework of this assumption);

- in the Copenhagen alternative the corresponding Hilbert space has been then required to be spanned on one set of Hamiltonian eigenfunctions $\psi_E(x)$:

$$H\psi_E(x) = E\psi_E(x); \tag{3}$$

- any vector of such a Hilbert space has represented a so called "pure" state of a corresponding physical system, which has meant that the mathematical superposition principle (holding in any Hilbert space) has been interpreted in physical sense.

While the Hilbert space has remained unspecified in the second assumption, the first two assumptions have been shared also by the hidden-variable theory. However, to get the whole assumption set belonging to the hidden-variable theory the third and fourth earlier assumptions must be refused and substituted by:

- the Hilbert space must be adapted to a given physical problem, consisting of some mutually orthogonal subspaces and being at least doubled in comparison to the third assumption in the preceding assumption set, if the original physical interpretation of Schrödinger time-dependent solution is to be conserved;

- only the vectors of orthogonal basis of extended Hilbert space (represented in principle by eigenfunctions of the Hamiltonian) may be interpreted as "pure" states; any superposition represents always a "mixed" state.

The given quantum alternatives represent, therefore, two quite different theories. They cannot be interpreted in any case as two mere different interpretations of one common mathematical model. Only the Schrödinger equation is shared commonly, its solutions being interpreted

in quite different ways. In the Copenhagen quantum mechanics the time evolution of any physical system is described as time reversible. In the hidden-variable theory the evolution is irreversible; the former third assumption being refused and the Hilbert space being suitably chosen (i.e., extended) in full agreement with the time-dependent solutions of Schrödinger equation (see Sec. 8).

3. Assumption in Bell's inequalities

It was believed since 1964, even if any more detailed analysis was not performed, that the assumption needed for the derivation of Bell inequalities corresponded to the hidden-variable theory. However, it has been shown later (see, e.g., (9)) that the given assumption has corresponded to the classical physics and it has not been possible to attribute it to any quantum alternative.

We shall describe now shortly the essence of the corresponding mistake. As already mentioned the given experiment has consisted in the measurement of coincidence transmission probabilities of two photons having opposite spins and running in opposite directions through two polarizers:

$$\left\| < - - |^{\beta} - - - o - - - |^{\alpha} - - > \right\|$$

Introducing some rather strong correlation conditions between individual transmission probabilities a_j and b_k Bell (6) (see also (15)) has derived the following inequalities

$$B = a_1 b_1 + a_2 b_1 + a_1 b_2 - a_2 b_2 \leq 2 \tag{4}$$

holding for any two pairs of orientations of both the polarizers (4 different combinations). The inequalities (4) have been then attributed to the hidden-variable alternative (without any deeper analysis). However, the given assumption has been much stronger and has eliminated practically the possibility for photon spins to influence the values of individual transmission probabilities, which corresponds to the original proposal of the given Gedankenexperiment by Einstein.

Any individual probability in the given experiment should be determined in principle by three different parameters: position of the photon source, photon spin direction and polarizer orientation. And the influence of two last parameters has been excluded when the assumption (leading to Bell's inequalities) has been applied to. The given inequalities were derived, of course, in other ways, too; see, e.g., Ref. (15), where instead of one assumption of Bell several weaker assumptions have been made use of. And when some assumptions have corresponded to the hidden-variable alternative one of these assumptions has held always in the classical physics only (see, e.g., (16)).

The given situation may be represented more clearly when the individual probabilities a_j and b_k are substituted by operators representing individual measurement acts and acting in two different subspaces (corresponding to individual polarizers) of the whole Hilbert space

$$\mathcal{H} \;=\; \mathcal{H}_a \otimes \mathcal{H}_b . \tag{5}$$

It holds for the expectation values of these operators (see (17))

$$0 \;\leq\; |\langle a_j \rangle|,\; |\langle b_k \rangle| \;\leq\; 1\,.$$

Eq. (4) represents then the definition of Bell's operator B. And it may be immediately seen that it holds $< B^*B > \leq 16$ and/or $< B > \leq 4$, which represents the highest possible limit. However, Bell's operator B may exhibit different limits according to basic assumptions concerning the relations holding between individual operators.
According to chosen commutation relations between a_j and b_j three different limits may be obtained (see (9)):

$$|B| \leq 2,\ 2\sqrt{2} \text{ or } 2\sqrt{3}.$$

The first limit corresponds to the classical case, when all operators a_j and b_k commute mutually, i.e., if

$$[a_j, b_k] = 0\,, \quad [a_1, a_2] = [b_1, b_2] = 0.$$

The second limit corresponds to the hidden-variable theory, when only the operators belonging to different subspaces commute (no interaction at distance or no entanglement), i.e., if

$$[a_j, b_k] = 0 \quad \text{and} \quad [a_1, a_2] \neq 0,\ [b_1, b_2] \neq 0\,.$$

And finally, the third limit (and actually the highest one (18)) corresponds to the case when the interaction at distance exists and the operators from different Hilbert subspaces do not commute mutually, i.e., if

$$[a_j, b_k] \neq 0\,, \quad [a_1, a_2] \neq 0,\ \ [b_1, b_2] \neq 0\,.$$

Only the classical alternative has been, therefore, excluded by the results of EPR experiment. As to the hidden-variable theory it does not contradict the results of EPR experiments (obtained, e.g., by Aspect et al. (7)); the value $2\sqrt{2}$ not having been overpassed. It is, of course, also the Copenhagen quantum mechanics that has not been excluded at least in principle. And further arguments should be made use of to decide with certainty which of the quantum alternatives represents really the description of microscopic physical reality; see Secs. 5 and 6.

4. Malus law and EPR coincidence experiment

One of the arguments supporting the Copenhagen quantum mechanics was added by Belinfante (11) as already mentioned. He tried to express the coincidence probability in the EPR polarization experiment as a function of $(\alpha - \beta)$. It would be possible easily if the individual dependencies $a(\alpha)$ and $b(\beta)$ for single transmission probabilities were known; α and β being the angles between photon spin (or photon polarization) and polarizer axis.
However, in fact only the transmission of unpolarized light through two polarizers have been measured

$$o - - - |^{\alpha} - - - |^{\beta} - - > \Big\|$$

and the so called generalized Malus law has been approximately obtained

$$M(\alpha - \beta) = (1 - \varepsilon) cos^2(\alpha - \beta) + \varepsilon$$

where ε represents very small correction. And Belinfante came to the given statement for the coincidence measurement when he put quite arbitrarily $a(\alpha) = \cos^2 \alpha$ and $b(\beta) = \cos^2 \beta$, which cannot correspond to real conditions; i.e., the Malus formula holding approximately for a polarizer pair was applied to each polarizer. And there is not any reason to expect for

coincidence measurements the values of the angle dependencies differing significantly from the Malus law, as required by Belinfante.
The given mistake has not represented, of course, the main argument for the then conclusion. This argument followed primarily from the violation of Bell's inequalities. However, Belinfante contributed probably rather strongly to it. In fact the approximate Malus law can be easily obtained in the given EPR experiment if the probabilistic functions $a(\alpha)$ and $b(\beta)$ are suitably chosen in agreement with reality (19). And the second experimental result obtained in 1982 (see Sec. 1) could not represent in principle any contribution to the solution of the given controversy.

5. Copenhagen alternative and internal discrepancies

The Copenhagen quantum mechanics has ascribed to the physical reality some properties that are denoted usually as paradoxes (wave-particle duality, tunnel phenomenon, non-local interaction or entanglement, a.s.o.). However, these paradoxes followed practically from some contradictions contained in the corresponding mathematical model. And it is possible to introduce at least three contradictions which may be denoted as decisive (see also (9)):

- Already the former fourth assumption itself (see Sec. 2) concerning the superposition principle has represented very important discrepancy, which has concerned the existence of discrete states in Schrödinger equation. It is evident that if the given assumption has been added all mathematical superpositions in the corresponding Hilbert space should have represented physical states fully equivalent to basic ones (i.e., to pure states) and no quantized (discrete) states might exist in principle in experimental reality. The given problem has been removed in the hidden-variable theory when only basis vectors corresponding to Hamiltonian eigenstates have represented "pure" states and any superposition of theirs has represented a statistical "mixture" (see also Sec. 7).
- The former third assumption has played then important role in the other discrepancy. Already in 1933 Pauli (20) showed that in such a case it was necessary for the corresponding Hamiltonian to have continuous energy spectrum from $-\infty$ to $+\infty$, which disagreed with the fact that the energy was defined as positive quantity, or at least limited always from below; the existence of discrete states having been excluded, too.
- In 1964 Susskind and Glogover (21) showed then in addition that in disagreement with necessary requirements the exponential phase operator $\mathcal{E} = e^{-i\omega\Phi}$ (where Φ was the phase) was not unitary, as it held $\mathcal{E}^{\dagger}\mathcal{E}\, u_0 \equiv 0$ for any state of zero energy if the third assumption was involved. It indicated that the given Hilbert space was not complete to represent the evolution of a corresponding physical system quite regularly.

Many attempts have been done during the 20th century to remove the last two deficiencies. The reason of having been unsuccessful may be seen in the fact that practically in all these attempts both these shortages were regarded and solved as one common problem. The corresponding solution has been obtained only recently (see Refs. (22; 23)) when it has been shown that it is necessary to remove two mentioned shortages one after the other.
The criticism of Pauli may be removed if the Hilbert space is extended in the accord to the time-dependent Schrödinger equation. E.g., as to the simple system of two free colliding particles it has had to be doubled in comparison to the former third assumption (see Sec. 2) as proposed by Lax and Phillips (12; 13). In this case the corresponding Hilbert space must

consist of two mutually orthogonal subspaces ($\mathcal{H} = \Delta^- \oplus \Delta^+$); being mutually related with the help of evolution operator and each of them being spanned on one basis of Hamiltonian eigenfunctions. The incoming or outgoing states of two particles must belong always to the different orthogonal subspaces.

As to the criticism of Susskind and Glogower the Hilbert space (extended already to solve Pauli's problem) should be further doubled and formed by combining two mutually orthogonal subspaces corresponding to systems with opposite angular momentums. These subspaces must be bound together by adding the action of exponential phase operator to link together the vacuum states (corresponding to minimum energy in individual subspaces), as it was proposed in principle already by Fain (24) also in 1967; see (22; 23). Some other points concerning the Hilbert space corresponding to the Schrödinger equation (or to the hidden-variable theory) will be mentioned in Sec. 8.

6. Experimental data and two quantum theories

The Copenhagen alternative has been often denoted as supported by different experimental data. However, in all such cases only the assumptions corresponding to the Schrödinger equation (i.e., to the hidden-variable theory) have been practically tested; without the last two assumptions (forming Copenhagen alternative) having been taken into account. And therefore, none of two quantum alternatives may be excluded on the basis of the experiments practically available until now.

And it has been shown in the preceding that practically the same has held also in the case of EPR experiment. Also the original statement of Belinfante that the two given quantum alternatives has had to lead to different predictions has been shown as mistaking (see Sec. 4). Having discovered this fact many years ago it was quite natural for us to ask: When in the EPR experiment the same predictions might be obtained for the so different physical concepts, would it not be possible to find an experiment where the predictions would be different? And after a preliminary theoretical analysis the measurement of light transmission through three polarizers seemed to represent suitable way.

The corresponding experiments have been performed and the results (see Refs. (25; 26)) have been obtained being fundamentally different from the predictions of the Copenhagen quantum mechanics as may be seen from Figs. 1 and 2 taken from Ref. (26); see also (9; 27; 28). In Fig. 1 the pairs of angles α and β giving the smallest coincidence probabilities for a chosen α have been shown; angle α having been represented on horizontal axis and corresponding angle β on vertical axis.

The values of light transmission through the polarizer triple for the given angle pairs (shown in Fig. 1) are represented by individual points on dashed line in Fig. 2. The full line represents then the prediction of the standard theory based on the Copenhagen quantum alternative. It may be seen that the existence of two maximums is predicted just where the experimental data show two deep minimums. Predictions similar to the obtained results may follow from the phenomenological polarization theory of Stokes (see (26)) but sufficient agreement with data may be hardly obtained on such a basis.

It is possible to say that the given results have opened also a way to a more realistic theory of polarization phenomena. The polarization process should be interpreted probably as a complex process; consisting practically at least of two main different successive steps: the influence of the enter layer and of the whole medium of a given polarizer. The given problem

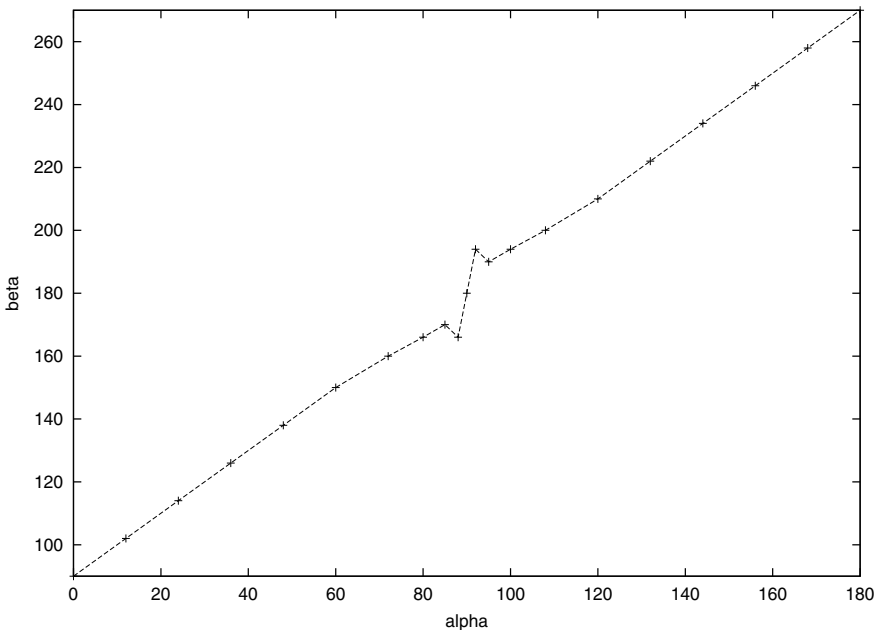


Fig. 1. The pairs of angles α and β (used for the measurement shown in Fig. 2)

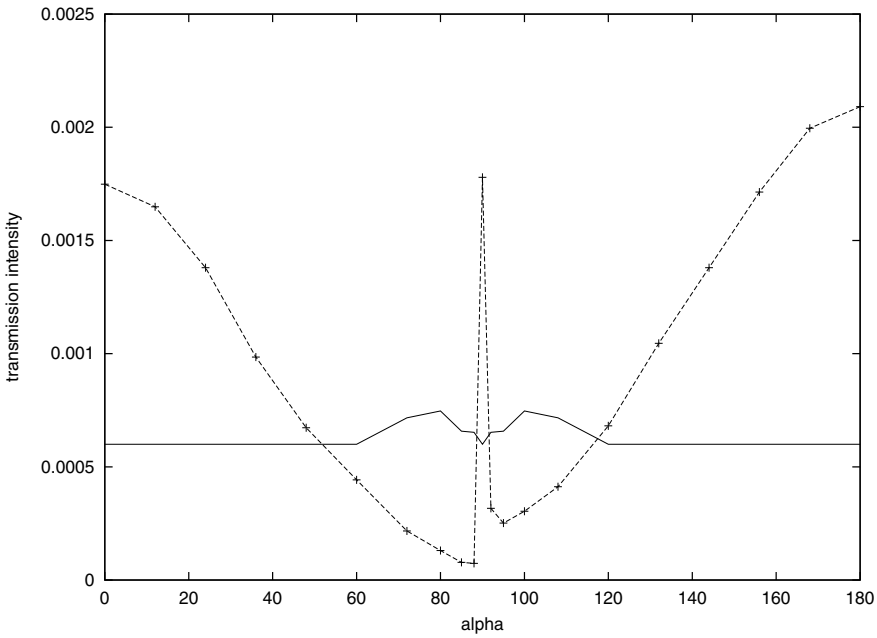


Fig. 2. Light transmission through three polarizers (for $\alpha - \beta$ pairs shown in Fig. 1); experimental data - points on dashed line; quantum-mechanical prediction - full line.

has been studied already theoretically on such a basis by us; however, definite results have not yet been gained.

7. Schrödinger equation and classical physics

As already mentioned the hidden-variable theory has been practically identical with the Schrödinger equation when any other conditions deforming the physical interpretation of original solutions have not been added. And one should ask what is to be understood under the term of hidden variables. Especially it is necessary to answer the question whether some hidden variables exist in the solutions of Schrödinger equation.

The problem concerning the hidden variables in the Schrödinger equation was discussed since the beginning of the Copenhagen quantum mechanics; their existence was, however, refused by von Neumann in 1932 (29). Even if G.Hermann (30) showed already in 1935 that the proof of von Neumann was to be denoted as circle proof the corresponding discussion was renewed only after D. Bohm (5) showed that some parameter existed already in the solution of the simplest Schrödinger equation. However, the actual meaning of such parameters remained unclear.

To answer the question what is the actual essence of "hidden variables" one must ask what is the assumption basis on the grounds of which the Schrödinger equation may be derived, when at least in some cases the predictions identical with classical ones may be obtained (14). And the answer has been given by Hoyer (31) and Ioannidou (32) who have shown that it has been possible to derive it if the set of all basic solutions of Hamilton equations has been considered together with the set of all statistical superpositions of these solutions, limited only by a rather weak condition.

It is then necessary to distinguish between the basic solutions of Eq. (1) and their superpositions, when the basic solutions (or states) are given by

$$\psi_E^b(x,t) = \psi_E(x)e^{-iEt}$$

where $\psi_E(x)$ is eigenfunction of corresponding Hamiltonian (see Eq. (3)). The basic solutions of Schrödinger equation (characterized always by one Hamiltonian eigenfunction only) represent the so called "pure" states and correspond to individual solutions of the Hamilton equations. It does not hold in opposite direction in the case when discrete states are involved (33). In such a case only a limited subset of solutions of Hamilton equations may represent pure states in the quantum alternative.

The Schrödinger equation is linear differential equation and consequently any superposition of basic states represents its solution, too. These superpositions must be interpreted, however, always as "mixed" states; never as "pure" states. And the parameters characterizing a concrete statistical distribution of corresponding basic states have been denoted earlier as hidden parameters. They may be in principle represented by the function $\alpha(E)$ having continuous as well as discrete parts that fulfills the condition $\Sigma_E|\alpha(E)|^2 = 1$ if all solutions are normalized to unit value. The corresponding statistical combination (mixed state) is then represented by the solution of Schrödinger equation:

$$\psi(x,t) = \sum_E \alpha(E)\,\psi_E^b(x,t)$$

where each basic (pure) state conserves the corresponding energy value E. The average energy of the corresponding superposition equals $\sum_E |\alpha(E)|^2 E$ and is also conserved. It follows on the basis of the mentioned derivation of Schrödinger equation from Hamilton equation that the function $|\alpha(E)|^2$ should fulfill some limiting condition (e.g., corresponding to Boltzmann statistics, see (31)) if the mentioned equivalence between Schrödinger equation and classical physics is to hold.

As to the Hilbert space the vectors of which may represent the time dependent solutions of Schrödinger equation it must contain always at least two different subspaces (each spanned on one set of Hamltonian eigenvectors) enabling to represent corresponding time flow. These subspaces must be mutually related with the help of evolution operator $U(t) = e^{-Ht}$ defining the passage from one subspace to the other (see also Sec. 5). In the case of discrete Hamiltonian energy spectrum this pair of subspaces may form the infinite chain of identical pairs related by evolution operator (for more details see (22; 23)). If the system consists of a greater number of interacting objects each of the given subspaces may have a more complicated structure.

All basic (or pure) states are represented by mutually orthogonal vectors forming the basis of individual subspaces and corresponding to states that may be described also by Hamilton equations. Any mixed state (superposition of basic states) corresponds then to a statistical distribution of classical states, and the values $|\alpha(E)|^2$ represent the corresponding probabilities of individual states. It means that the hidden-variable theory (or the Schrödinger equation) can supersede also the classical physics if one admits that macroscopic physical systems exhibit discrete energy values, too; the fact being fully acceptable but experimentally unverifiable.

However, a given basic physical state is not fully determined by total energy value E. The other value that is conserved during the evolution is angular momentum L which cannot be usually uniquely defined in an initial state. It is, of course, possible to choose the coordinate system in which the components $< L_1 > = < L_2 > = 0$ and the given state may by represented by a superposition containing states with different values of $< L_3 >$. In a two-particle system the given value may be substituted by their minimal mutual distance q during the evolution and the given superposition is characterized (in addition to $\alpha(E)$) also by the function $\alpha(q)$; see also the end of Sec. 8.

8. Schrödinger equation and corresponding Hilbert space

All solutions of Schrödinger equation may be represented by the vectors of Hilbert space. It must have, however, the corresponding structure as mentioned in preceding. We shall demonstrate basic characteristics of this space with the help of the physical system consisting of two zero-spin particles which represents the smallest system exhibiting time evolution. Its behavior in the center-of-mass system may be described by Schrödinger equation (1) with the Hamiltonian

$$H = \frac{p_j^2}{2m} + V(q_j) \tag{6}$$

where m is the reduced mass of the given particle pair and q_j and p_j are the operators of relative coordinates and momentum components of one particle (in the center-of-mass system); they are assumed to fulfill the following relations

$$[q_j, p_k] = i\delta_{jk}, \quad [p_j, p_k] = [q_j, q_k] = 0. \tag{7}$$

Introducing two other operators

$$Q = q_j^2, \quad R = \frac{1}{2}\{p_j, q_j\} \tag{8}$$

and assuming $V(q_j) = V(Q)$ (i.e., the mutual potential between particles depends on their distance only) one can write further

$$i[H, q_k] = \frac{p_k}{m}, \quad i[H, p_k] = -2q_k \frac{dV(Q)}{dQ}, \tag{9}$$

$$i[H, Q] = \frac{2}{m}R, \quad i[H, R] = 2H - V(Q) - Q\frac{dV(Q)}{dQ}. \tag{10}$$

It is also possible to introduce angular-momentum operators fulfilling relations

$$L_i = \varepsilon_{ijk}[q_j, p_k], \quad [L_i, H] = 0; \tag{11}$$

and further operator L:

$$L = L_i^2, \quad [L, L_i] = [L, H] = 0. \tag{12}$$

The basic properties of a given physical system are characterized by the expectation values of the operators $< H >$ and $< L >$; and the basic space orientation by expectation values $< L_j >$. Any time-dependent solution of the Schrödinger equation may be then represented uniquely by one trajectory in the corresponding Hilbert space. And it may be suitable to choose the special coordinate system in which $< L_1 >=< L_2 >= 0$ (or $< q_3 >= 0$) and further also $< q_1 >= 0$ and $< p_2 >= 0$ if $< R >= 0$.

As to the structure of corresponding Hilbert space there is a significant difference between the solutions belonging to continuous or discrete spectra of Hamiltonian. We shall start with the continuous spectrum. In this case the instantaneous states with negative (positive) $< R >$ values represent incoming (outgoing) states belonging to two different Hilbert subspaces (both being spanned on the full set of the same Hamiltonian eigenfunctions). And it is necessary to assume that these subspaces are mutually orthogonal as the given states are physically well distinguishable. It is also useful to define the time operator T fulfilling the condition

$$i[H, T] = 1 \tag{13}$$

with the expectation value $t = 0$ corresponding to the state characterized by $< R >= 0$. The whole trajectory may be obtained when the evolution operator

$$U(t) = e^{-iHt} \tag{14}$$

is applied to this initial state; the incoming (outgoing) states corresponding to negative (positive) values of t.

To obtain a reasonable description of an evolving physical system it is necessary always to hold $\lim_{Q\to\infty} V(Q) = 0$. If the system consists of two free particles there is one-to-one correspondence between the operators T and R that may be specified for some concrete shapes of $V(Q)$ (34). And for a given Hamiltonian any time-dependent trajectory is uniquely defined by the function $\psi(x, 0)$; being fully determined by the values of $q_2(0)$ and $p_1(0)$ when the mentioned coordinate system has been made used of.

The corresponding mathematical model describes suitably usual collision experiments. However, in the given experiments (for any energy value E from the continuous spectrum) the collision behavior may be described only as a statistical combination of the states with different values of $q_2(0)$. The frequency of these individual collision events is then proportional to the values of impact parameter (distance of initial momentum direction from system mass center) that must be derived from the values obtained at a time $t_Q < 0$ (expectation value of the operator T) for which $< V(Q) >$ may be fully neglected against $< H >$.

If the influence of potential energy decreases monotony with Q it is possible to assume for sufficiently high values of $|t_Q|$ that the particle moves along a straight line that may be uniquely determined for any state characterized by the values of $q_2(0)$ and $p_1(0)$. And the distance of this straight line from the parallel one going directly through the corresponding coordinate center corresponds then to the original impact parameter for the given pair of colliding particles. And it is possible to assume further that these impact parameters will be equally distributed in the plane perpendicular to the given direction and that the corresponding value of the impact parameter b will represent the weight of the corresponding state in the given experiment.

In the given coordinate system the straight line corresponding to initial momentum transects the q_2-coordinate at the value

$$y = \frac{q_2(t_Q)\, p_1(t_Q) - q_1(t_Q)\, p_2(t_Q)}{p_1(t_Q)} \tag{15}$$

and it holds also

$$\frac{b}{y} = \frac{p_1(t_Q)}{p(t_Q)}.$$

The numerator in Eq. (15) represents one component of angular momentum. And if one takes into account that the angular momentum component commutes with Hamiltonian (and is conserved during the evolution) it holds further

$$b = \frac{|L_3|}{p(t_Q)}$$

or

$$b = |q_2(0)|\, \frac{p(0)}{p(t_Q)}$$

where $p(t_Q)$ may be derived immediately from the corresponding energy E determined with the help of values $p(0) = |p_1(0)|$ and $q_2(0)$. The frequency of a state characterized by the values of $q_2(0)$ and $p_1(0)$ in the given experiment is then proportional to the value of b. It means that the states corresponding to very small values of $|q_2(0)|$ become more frequent for an attractive potential and less frequent for a repulsive one.

The given weight of states characterized by the value $q_2(0)$ in the statistical distribution is important in the case when the interaction between colliding particles is mediated by forces at distance (e.g., Coulomb forces) as the density of states at small values of $|t|$ may be significantly influenced by the corresponding shape of potential dependence at small mutual distances. In the case of contact forces between colliding particles (which may concern strong

interactions) one may regard corresponding tracks as linear in the whole length (before the proper collision process).
The preceding considerations have concerned the case when the Hamiltonian spectrum has been continuous. The situation is more complex in the case of discrete Hamiltonian spectrum as the zero value of $< R >$ repeats during periodical time evolution; four zero values existing already in any period. Each period may be then represented by two pairs of subspaces corresponding to positive and negative values of $< R >$ which continuously repeat and may form an infinite series where individual states are characterized by different expectation values of operator T. More suitable operators than R (represented now by goniometric type functions) may be introduced in the case of special shapes of $V(Q)$; see (34).

9. Physical science and ontological thinking

The human thinking in the preceding century was undoubtedly rather strongly influenced also by the progress of physical research, which was often being connected with the phenomenistic ideas of the Copenhagen quantum mechanics. The corresponding ideas differed fundamentally from the earlier ontological concept proposed by Aristotle and developed further by Thomas Aqu. in the Middle Ages, from which G. Galileo and I. Newton started when the basis of classical physics was built up and the theoretical background was formulated further by W.R. Hamilton (1805-65); and the whole modern civilization has been based in principle fully on it.
However, this train of thought started to change from the beginning of the modern age. The initial impulse was given probably by Occam's razor when the simplicity started to be decisively preferred in explaining the reality that might be often very complex. The decisive step was done, however, by R. Descartes who refused practically any ontology by his "Cogito, ergo sum"; the human thinking started to decide practically about the real existence on the basis of different observed phenomena without any deeper ontological analysis. The given phenomenism started to influence strongly also the physical science when L. Boltzmann declared the rise of uniform distribution in the systems consisting of many particles (i.e., the phenomenistic feature) as basic natural law in 1867. A further important step occurred when N. Bohr based the atom theory on two phenomenological postulates in 1913, but mainly when he proposed the Copenhagen quantum mechanics in 1927. And since Bohr's victory in the controversy with Einstein in 1935 the corresponding phenomenistic approach asserted oneself in the western world practically in all regions of human thinking and knowledge.
As to the microscopic physics region the given research has been divided practically into two different branches. One part of physicists has continued in the research on the earlier ontological basis without taking the conclusions of Copenhagen alternative practically into account when they looked for the solutions of Schrödinger equation based on the same Hamiltonian as the classical physics; any additional deforming assumptions having not been added.
The other (lesser) part of these physicists has looked then mainly for arguments to persuade the human community about the validity of quantum paradoxes. This part of physicists was making use of the successes of modern research based on the solutions of Schrödinger equation often as the decisive support of their conviction, without respecting the fact that any of the assumptions added by Bohr (see Sec. 2) has not been taken into account when only solutions of Schrödinger equation have been looked for and directly interpreted. In

such a case the existence of the correspondingly extended Hilbert space (as described in Sec. 8) has been automatically assumed in the interpretation of physical processes; no deformation according to Bohr's assumptions has been applied to the corresponding solutions of Schrödinger equation.
It follows from the presented results that the phenomenistic approach has brought the human knowledge to misleading conclusions. It should be abandoned and the common human thinking should return again to the earlier ontological approach in studying the matter world. It might open the new way to deeper understanding not only of the fundamental structure of matter, but also of other problems of human kind.

10. Conclusion

Let us introduce yet the short summary of all previous results:
The Copenhagen quantum mechanics may be denoted as falsified on the basis of logical contradictions as well as of experimental data (light transmission through a polarizer triple).
Schrödinger equation itself (or hidden-variable theory) should be preferred not only to the Copenhagen quantum mechanics but also to the classical physics; it may be applied to the whole matter reality (microscopic as well as macroscopic).
The basic solutions of Schrŏdinger equation (characterized by one Hamiltonian eigenfunction only) fulfill always classical properties. It does not hold in opposite direction in the case of discrete Hamiltonian spectrum; some solutions of Hamilton equations having not any quantum counterpart.
There are not in principle any hidden variables in the Schrödinger equation. In the case of superposition solutions only the statistical distribution of potential energy corresponding to individual basic states must be defined with the help of the distribution of some physical quantities in a suitable time instant.
Superposition solutions do not represent any of the so called pure states. They must be always interpreted as mixed states. There is not any gap between microscopic and macroscopic world. Any superposition represents always the set of states that may be interpreted classically (or ontologically).
And finally, one must stress that in the given controversy Einstein was right. The purely phenomenological accession to the physical reality should be refused and the ontological approach to physical problems again strongly preferred. The physics (and all science) is to become again serious problem of human knowledge and not a mere game deciding between different phenomenologically based attitudes (comp., e.g., (4)).

11. References

[1] N.Bohr: The quantum postulate and the development of atomic theory; Nature 121, 580-90 (1928).
[2] A.Einstein, B.Podolsky, N.Rosen: Can quantum-mechanical description of physical reality be considered complete?; Phys. Rev. 47, 777-80 (1935).
[3] N.Bohr: Can quantum-mechanical description of physical reality be considered complete?; Phys. Rev. 48, 696-702 (1935).
[4] R.G.Newton: How physics confronts reality: Einstein was correct but Bohr won the game; World Scientific (2009).

[5] D.Bohm: A suggested interpretation of the quantum theory in terms of "hidden variables"; Phys. Rev. 85, 180-93 (1952).
[6] J.S.Bell: On the Einstein Podolsky Rosen paradox; Physics 1, 195-200 (1964).
[7] A.Aspect, P.Grangier, G.Roger: Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell's inequalities; Phys. Rev. Lett. 49, 91-4 (1982).
[8] M.V.Lokajíček: Einstein-Bohr controversy and theory of hidden variables; NeuroQuantology (section: Basics of Quantum Physics) 8 (2010), issue 4, 638-45; see also /arXiv:1004.3005[quant-ph].
[9] M.V.Lokajíček: Quantum theory without logical paradoxes; Concepts of Physics 6, No.4, 581-604 (2009); see also /arxiv:09050140[quant-ph].
[10] M.V.Lokajíček: Physical theory of the twentieth century and contemporary philosophy; Concepts of Physics 4 (2007), no.2, 317-39; arXiv:quant-ph/0611069.
[11] F.J.Belinfante: A survey of hidden-variable theories; Pergamon, Oxford, p. 283 (1973).
[12] P.D.Lax, R.S.Phillips: Scattering theory; Academic Press (1967).
[13] P.D.Lax, R.S.Phillips: Scattering theory for automorphic functions; Princeton (1976).
[14] E.Schrödinger: Quantisierung als Eigenwertproblem; Ann. Phys. 79, 361-76; 489-527; 80, 437-90; 81, 109-39 (1926).
[15] J.F.Clauser, A.Shimony: Bell's theorem: experimental tests and implications; Rep. Prog. Phys. 41, 1881-91 (1978).
[16] M.V.Lokajíček: Locality problem, Bell's inequalities and EPR experiments; /arXiv:quant-ph/9808005 (1998).
[17] M.Hillery, B.Yurke: Bell's theorem and beyond; Quantum Semiclass. Optics 7, 215-227 (1995).
[18] M.Revzen, M.Lokajíček, A. Mann: Bell's inequality and operators' noncommutavity; Quantum Semiclass. Optics 9, 501-6 (1996).
[19] M.V.Lokajíček: Semiclassical interpretation of microscopic processes; arXiv:quant-ph/0511184.
[20] W.Pauli: Die allgemeinen Prinzipien der Wellenmechanik; Handbuch der Physik XXIV, Springer, Berlin, p. 140 (1933).
[21] L.Susskind, J.Glogover: Quantum mechanical phase and time operator; Physics (Long Island City, N.Y.) 1, 49-61 (1964).
[22] P.Kundrát, M.Lokajíček: Three-dimensional harmonic oscillator and time evolution in quantum mechanics; Phys. Rev. A 67, art. 012104 (2003).
[23] P.Kundrát, M.Lokajíček: Irreversible time flow and Hilbert space structure; New Research in Quantum Physics (eds. Vl.Krasnoholovets, F.Columbus), Nova Science Publishers, Inc., pp. 17-41 (2004).
[24] V.Fajn: Quantum harmonic oscillator in phase representation and the uncertainty relation between the number of quanta and the phase (in Russian); J. Exp. Theor. Phys. 52, 1544-8 (1967).
[25] J.Krása, J.Jiřička, M.Lokajíček: Depolarization of light by an imperfect polarizer; Phys. Rev. E 48, 3184-6 (1993).
[26] J.Krása, M.Lokajíček, J.Jiřička: Transmittance of laser beam through a pair of crossed polarizers; Phys. Letters A 186, 279-82 (1994).
[27] M.V.Lokajíček: Hidden-variable theory versus Copenhagen quantum mechanics; Frontiers of Fundamental and Computational Physics - Proc. of the Ninth International

Symposium, January 2008, Udine and Trieste, edited by Sidharth B.G. et al., Conference Proceedings, No. 1018, American Institute of Physics, pp. 40–5 (2008); see also extended version: /arXiv:0711.3977v3[quant-ph].

[28] M.V.Lokajíček: Hidden-variable theory contra Copenhagen quantum mechanics; Proc. of 16th Conf. of Czech and Slovak Physicists (ed. J.Kříž), MAFY Hradec Kralove, pp.406-10 (2009).

[29] J.von Neumann: Mathematische Grundlagen der Quantenmechanik; Springer (1932).

[30] Grete Herrmann: Die Naturphilosophischen Grundlagen der Quantenmechanik; Abhandlungen der Fries'schen Schule 6 (1935), 75-152.

[31] U.Hoyer: Synthetische Quantentheorie; Georg Olms Verlag, Hildesheim (2002).

[32] H.Ioannidou: A new derivation of Schrödinger equation; Lett. al Nuovo Cim. 34, 453-8 (1982).

[33] M.V.Lokajíček: Schrödinger equation, classical physics and Copenhagen quantum mechanics; New Advances in Physics 1, No. 1, 69-77 (2007); see also /arxiv/quant-ph/0611176.

[34] M.V.Lokajíček: Realistic theory of microscopic phenomena; a new solution of hidden-variable problem; /arxiv:quant-ph/9811030.

Edited by Mohammad Reza Pahlavani

Quantum mechanics, shortly after invention, obtained applications in different area of human knowledge. Perhaps, the most attractive feature of quantum mechanics is its applications in such diverse area as, astrophysics, nuclear physics, atomic and molecular spectroscopy, solid state physics and nanotechnology, crystallography, chemistry, biotechnology, information theory, electronic engineering... This book is the result of an international attempt written by invited authors from over the world to response daily growing needs in this area. We do not believe that this book can cover all area of application of quantum mechanics but wish to be a good reference for graduate students and researchers.

ISBN 978-953-51-4953-8

Photo by LeeYiuTung / iStock

IntechOpen