

Mathematics for molecular simulation: accomplishments and challenges

Claude Le Bris¹

¹*CERMICS, Ecole Nationale des Ponts et Chaussées, 6-8 avenue Blaise Pascal, Champs sur Marne, 77455 Marne La Vallée Cedex 2, FRANCE, and INRIA Rocquencourt, MICMAC project, Domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, FRANCE*

*CERMICS — ENPC
6 et 8 avenue Blaise Pascal
Cité Descartes - Champs sur Marne
77455 Marne la Vallée Cedex 2*

<http://cermics.enpc.fr>

Mathematical and numerical analysis for molecular simulation: accomplishments and challenges

Claude Le Bris

[To appear in *Proceedings of the International Congress of Mathematicians, Madrid 2006, EMS Ph.*]

Abstract. Molecular simulation is explored from the mathematical viewpoint. The field comprises computational chemistry and molecular dynamics. A variety of mathematical and numerical questions raised is reviewed. Placing the models and the techniques employed for simulation on a firm mathematical ground is a difficult task, which has begun decades ago. The time is right for assessing the field, and the issues and challenges ahead.

Mathematics Subject Classification (2000). Primary 35Bxx, 35Jxx, 35Pxx, 49Kxx, 81Q05, 81Q10, 82Bxx.

Keywords. Computational chemistry, molecular simulation, molecular dynamics, Schrödinger equations, Hartree-Fock theory, Thomas-Fermi theory, Density Functional Theory, nonlinear eigenvalue problems, spectral theory, spectral methods, elliptic partial differential equations, optimization, reduced basis methods, sparse grids, Hamiltonian dynamics, symplectic methods, geometric integration, stochastic differential equations, Markov chains, Monte-Carlo techniques.

1. Introduction

Molecular simulation is an increasingly important field of scientific computing. It comprises *computational chemistry*, focused on the calculations of electronic structures and the related properties, and *molecular dynamics*, devoted to the simulation of molecular evolutions, evaluations of ensemble averages and thermodynamic quantities. Examples of reference treatises are [68, 70] and [1, 39], respectively. We also refer to [14, Chapter 1] for a mathematically-oriented introductory text.

1.1. Ubiquity of molecular simulation. The field has several intimate connections with many other fields. Indeed, *molecular simulation* is above all important because many macroscopic properties of matter originate from phenomena at the microscopic scale. Instances are: electrical conductivities, colors, chemical reactivities, mechanical behaviour, aging. Accurate calculations on representative microscopic systems allow for the evaluation of such properties. Additionally, even the macroscopic phenomena that proceed from bulk effects, and which thus ne-

cessitate the consideration of large size microscopic systems, may now be studied by advanced techniques in molecular simulation. Recent record calculations simulate the dynamics of billions of atoms over a microsecond. Molecular biology, chemistry and physics are thus inseparable today from molecular simulation. An easy observation sheds some light on this. Roughly one publication out of ten in chemistry journals presents some numerical simulations performed on theoretical models. This is an impressive ratio for a field so much experimentally oriented.

Computations are first seen as complements to experiments. For instance, all the information about the electronic properties is contained in the wave function; the latter cannot be measured but it can be computed. Computations are also seen as an alternative to experiment. It is possible to simulate molecular systems that have not been synthesized yet, or phenomena inaccessible to experiments (huge temperature or pressure, time scales smaller than the femtosecond, evolutions on decades or more). Additionally, computations can serve for the laser control of molecular systems ([6]), and other emerging fields of high energy physics.

Other, apparently distant, fields also make an extensive use of molecular simulation. Rheology of complex fluids and more generally materials science were once focused at the macroscopic scale and based on purely macroscopic descriptions. They used to be far from molecular concerns. However, the accuracy needed in the quantitative evaluation of many properties (think e.g. of constitutive laws or slip boundary conditions) requires models to be more and more precise, involving the finest possible scales in the simulation. This eventually includes the molecular scale.

The last application field that we shall mention, besides the fields using the macroscopic impact of molecular simulation, regards the emerging field of nanotechnology. Nanosystems are indeed accessible today to a direct molecular simulation.

Overall, major technological challenges for the years to come may, or more properly stated must, be addressed by molecular simulation techniques. Examples are the detailed simulation of protein folding, and the description of the long time radiation damage of materials in nuclear power plants. To appreciate this ubiquity of molecular simulation, it is sufficient to consider the enormous proportion of computational time devoted to molecular simulation in the largest centers of computational resources worldwide.

1.2. Relation to mathematics. On the other hand, the interface of molecular simulation with mathematics is not yet comparable to the practical importance of the field.

Molecular simulation, and more precisely computational quantum chemistry, were born in the 1950s for molecular systems consisting of a few electrons. Contemporary methods and techniques now allow for the simulation of molecules of hundreds of electrons, modelled by very precise quantum models, up to samples of billions of particles modelled by molecular dynamics. This is an enormous success. The calculations are often surprisingly accurate, but also sometimes desperately inaccurate. Experts in chemistry have constantly improved the models and the

methods. They have turned the field into an allmighty tool. However, in many respects, molecular simulation is still an art. It relies upon a delicate mix of physical intuition, pragmatic cleverness, and practical know-how. Mathematics has already provided with significant contributions to the theoretical understanding. Also, its companion fields, numerical analysis and scientific computing, have definitely improved the efficiency of the techniques. Yet, they all need to irrigate more molecular simulation. To state it otherwise, *there is an enormous gap between the sophistication of the models and the success of the numerical approaches used in practice and, on the other hand, the state of the art of their rigorous understanding.*

We are witnessing an evolution that is due to two different reasons.

First, the mathematical knowledge on the models is rather satisfactory. Efforts were initiated as early as the 1970s by pioneers such as E. H. Lieb, B. Simon, W. Thirring, Ch. Feffermann, focusing on fundamental theoretical issues. Questions were addressed about the well-posedness of the models, and the relation between the various models, in various asymptotic regimes. Researchers such as R. Benguria, J. P. Solovej, V. Bach, G. Friesecke, to only name a few, continued the effort over the years. Those were later joined by contributors following the impulsion given by P-L. Lions: M. J. Esteban, I. Catto, E. Séré, X. Blanc, M. Lewin, and the author. A number of researchers, experts in analysis, spectral theory, partial differential equations, evolution equations, now become involved in the field. The enclosed bibliography cites several of them.

Second, and as a natural follow-up to mathematical analysis, numerical analysis has indeed come into the picture. The numerical analysis of computational chemistry methods was a completely unexplored subject until the mid 1990s. Boosted by the state of the mathematical analysis, it is now a quickly developing topic. The work in this field was pioneered by E. Cancès. Researchers such as Y. Maday, M. Griebel, W. Hackbush, Ch. Lubich, W. E, well known for their contribution in various other fields of the engineering sciences, now get involved, along with their collaborators (G. Turinici, ...) in electronic structure calculations or in molecular dynamics.

It is therefore a good time for assessing the field, and the issues and challenges ahead. Doing so might help to boost the research in the area.

The present contribution rapidly reviews some commonly used models and their mathematical nature, indicating the progress achieved over the last decades in their mathematical understanding. Questions of numerical analysis are also addressed. Important unsolved issues are emphasized. Owing to the evident space limitation, this review is not meant to be exhaustive: see [54, 55] for more comprehensive reviews, and [56] for a recent collection of various contributions. This is rather an invitation for mathematicians to get involved in the endeavour of placing the field on a firm mathematical ground.

2. Mathematical overview of the models of computational chemistry

2.1. The Schrödinger equation. For most applications of molecular simulation, the matter is described by an assembly of nuclei, which are point particles treated classically, equipped with electrons, which are light particles modelled by quantum mechanics. For systems of limited size, called *molecular systems*, there are M nuclei, of charge z_k , located at \bar{x}_k , and N electrons of unit charge. The finest models are called *ab initio* models since they only involve universal constants and no experimentally determined parameters. Assuming the molecular system non-relativistic, placing it at zero temperature, and, for clarity of exposition, omitting the spin variable, the state of the electrons is modelled by the N -body Hamiltonian

$$H_e^{\bar{x}_1, \dots, \bar{x}_M} = - \sum_{i=1}^N \frac{1}{2} \Delta_{x_i} - \sum_{i=1}^N \sum_{k=1}^M \frac{z_k}{|x_i - \bar{x}_k|} + \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|}, \quad (1)$$

where the terms respectively model the kinetic energy, the attraction between nuclei and electrons, the repulsion between electrons. Notice that the positions \bar{x}_k of the nuclei are *parameters* of this operator. The *electronic ground-state* is by definition the minimizer of the energy:

$$W(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M) = \inf \{ \langle \psi, H_e^{\bar{x}_1, \dots, \bar{x}_M} \psi \rangle, \quad \psi \in \mathcal{W}_N \} + \sum_{1 \leq k < l \leq M} \frac{z_k z_l}{|\bar{x}_k - \bar{x}_l|}. \quad (2)$$

The variational space reads

$$\mathcal{W}_N = \left\{ \psi \in \bigwedge_{i=1}^N L^2(\mathbb{R}^3), \quad \int_{\mathbb{R}^{3N}} |\psi|^2 = 1, \quad \int_{\mathbb{R}^{3N}} |\nabla \psi|^2 < +\infty \right\} \quad (3)$$

where the wedge product denotes the antisymetrized tensor product (owing to the Pauli exclusion principle). The Euler-Lagrange equation of (2) is the (time-independent) Schrödinger equation

$$H_e^{\bar{x}_1, \dots, \bar{x}_M} \psi = E \psi \quad (4)$$

where the energy E , lowest possible eigenvalue of $H_e^{\bar{x}_1, \dots, \bar{x}_M}$ on \mathcal{W}_N is called the ground-state energy. The resolution of (2) (or one approximation of it, which we will detail below) is at the core of any computational chemistry calculation, prior to any calculation related to excited states, energies, linear response, etc. We therefore focus on this problem here.

Analogously, a time-dependent version of the problem exists: then the *time-dependent* Schrödinger equation

$$i \frac{\partial}{\partial t} \psi = H_e^{\bar{x}_1, \dots, \bar{x}_M} \psi \quad (5)$$

is to be solved. The treatment of the electronic problem is usually the *inner* loop of the simulation, the *outer* loop consisting of the treatment of the nuclei. In

the static setting, this consists in solving the *molecular mechanics* problem (also termed *geometry optimization*): finding the configuration of nuclei that minimizes the overall energy, i.e. the minimizer of

$$\inf_{(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M) \in \mathbb{R}^{3M}} W(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M). \quad (6)$$

The time-dependent setting requires solving the equations of *molecular dynamics*, i.e. the Newton equations of motion for the nuclei:

$$m_k \frac{d^2}{dt^2} \bar{x}_k = -\nabla_{\bar{x}_k} W(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M). \quad (7)$$

2.2. Standard approximations. Problem (2) is well explored mathematically: [37, 49]. In addition to their own interest, theoretical studies of (2) provide with useful practical informations on the quantities (wavefunction and energy) to be evaluated in practice (see [48] and other works by the same authors). The practical bottleneck of quantum chemistry calculations is however that state-of-the-art numerical techniques only allow for (2) to be solved for ridiculously small numbers of electrons. Indeed, the dimension of the tensor product $\wedge_{i=1}^N L^2(\mathbb{R}^3)$ makes the problem untractable by usual techniques of scientific computing for the practically relevant numbers of electrons, say a few tens to thousands. The practice of computational chemistry is thus to *approximate* (2). The purpose of such approximations is to reduce the computational complexity of the problem, whilst providing the accuracy required by chemistry. The energy of molecular systems must indeed be determined within an incredibly demanding degree of accuracy (often termed the *chemical accuracy*). Energies such as (2) are typically 10^3 to 10^6 as large as the energy of an hydrogen bond. As the interest lies in the *difference* of energy between two systems, in order to determine which is the more stable one, the difficulty is challenging. Surprisingly, clever approximations do succeed in this task. We now review them. For more details on the analysis, implementation and efficiency of all the numerical techniques mentioned below, see [53].

In chemistry, approximations of (2) are schematically sorted into two categories.

Wavefunctions methods are used preferably by chemists, on small systems, when accuracy is the primary goal, and computational time is a secondary issue. The focus is on the *interaction* between electrons. The prototypical example is the Hartree-Fock model. The latter is the best known model in the mathematical community. The bottom line for deriving the Hartree-Fock model is a variational approximation of the set (3) by the subspace of wavefunctions ψ that read as *determinants* (antisymmetrized products) of wavefunctions of *one* electron. More precisely, the Hartree-Fock problem reads

$$\inf \left\{ E^{HF}(\{\phi_i\}), \quad \phi_i \in H^1(\mathbb{R}^3), \quad \int_{\mathbb{R}^3} \phi_i \phi_j^* = \delta_{ij}, \quad 1 \leq i, j \leq N \right\} \quad (8)$$

with

$$E^{HF}(\{\phi_i\}) = \frac{1}{2} \sum_{i=1}^N \int_{\mathbb{R}^3} |\nabla \phi_i(x)|^2 dx + \int_{\mathbb{R}^3} \rho(x) V(x) dx + \frac{1}{2} \iint_{(\mathbb{R}^3)^2} \frac{\rho(x)\rho(y)}{|x-y|} dx dy - \frac{1}{2} \iint_{(\mathbb{R}^3)^2} \frac{|\tau(x,y)|^2}{|x-y|} dx dy, \quad (9)$$

with $V = -\sum_{k=1}^M \frac{z_k}{|\cdot - \bar{x}_k|}$, $\tau(x,y) = \sum_{i=1}^N \phi_i(x) \phi_i(y)^*$, $\rho(x) = \tau(x,x) = \sum_{i=1}^N |\phi_i(x)|^2$,

where the star denotes the complex conjugate. *Post-Hartree-Fock methods* consist in enlarging the variational space by considering *linear combinations* of determinants: *Configuration Interaction* (CI) methods, *Multiconfiguration Self Consistent Field* (MCSCF) methods. Nonvariational correction methods, mostly based on linear perturbation theory, are also employed: *Möller-Plesset*, *Coupled Cluster*.

On the other hand, *Density Functional Theory based methods* are used preferably for larger systems (and beyond for materials science), when computational time matters and wavefunctions methods are too expensive. They consist in rephrasing the problem (2) in terms of the electronic density

$$\rho(x) = N \int_{\mathbb{R}^{3(N-1)}} |\psi(x, x_2, x_3, \dots, x_N)|^2 dx_2 dx_3 \dots dx_N.$$

Formally, a minimization problem of the type

$$\inf \left\{ \mathcal{E}(\rho); \int_{\mathbb{R}^3} \rho(x) dx = N \right\} \quad (10)$$

is obtained. The idea has a rigorous theoretical grounding, but making it tractable in practice requires some approximation procedure. The energy $\mathcal{E}(\rho)$, which is a reformulation of $\langle \psi, H_e^{\bar{x}_1, \dots, \bar{x}_M} \psi \rangle$, is not explicit. Adequately adjusting the parameters (and even the terms) of the approximate energy functional $\mathcal{E}(\rho)$ is an issue, sometimes controversial. Ancestors of DFT-based methods are Thomas-Fermi type theories, very well investigated mathematically (see [60, 63, 74] for reviews). The latter currently see a revival through *orbital-free* methods, which precisely consist in discretizing ρ itself as the primary unknown. They therefore allow for the treatment of larger systems, notably for materials science applications.

The general trend is that DFT-based models are increasingly popular. A commonly used setting is the *Kohn-Sham Local Density Approximation* (KS-LDA) setting that explicitly reads as the minimization problem

$$\inf \left\{ E^{KS-LDA}(\{\phi_i\}), \quad \phi_i \in H^1(\mathbb{R}^3), \quad \int_{\mathbb{R}^3} \phi_i(x) \phi_j^*(x) dx = \delta_{ij}, \quad 1 \leq i, j \leq N \right\} \quad (11)$$

with

$$E^{KS-LDA}(\{\phi_i\}) = \frac{1}{2} \sum_{i=1}^N \int_{\mathbf{R}^3} |\nabla \phi_i(x)|^2 dx + \int_{\mathbf{R}^3} \rho(x) V(x) dx + \frac{1}{2} \iint_{\mathbf{R}^3 \times \mathbf{R}^3} \frac{\rho(x) \rho(y)}{|x-y|} dx dy - \int_{\mathbf{R}^3} F(\rho(x)) dx, \quad (12)$$

where F is a nonlinear function determined on chemical basis.

Reducing the complexity of (2) comes at a price: *nonlinearity*. Whereas the optimality equation (4) is a linear eigenvalue problem (in a high dimensional space), the equation to be solved for most of the approximations of (2) is a *nonlinear eigenvalue problem* (in a space of lower dimension, though). This is easily seen on the expressions (9) and (12). However they are derived and irrespective of their chemical meaning, the wavefunctions methods and DFT-based methods both lead to a *nonlinear eigenvalue problem*:

$$-\Delta \phi_j + \mathcal{W}(\phi_1, \dots, \phi_N) \phi_j = \lambda_j \phi_j, \quad j = 1, \dots, N, \quad (13)$$

where the λ_j are the Lagrange multipliers of the constraints. Equations (13) are often called *Self-Consistent Field* (SCF) equations to emphasize the nonlinear feature, encoded in the operator \mathcal{W} .

There are many questions of mathematical interest. The existence of a minimizer (under appropriate physically relevant conditions) for several models related to HF and DFT-type is now established. Very important contributions in this direction are [62, 61, 8, 65, 66, 73, 40, 59]. For most models of practical interest, the existence of a minimizer is known. In contrast, nothing is known on the uniqueness. A major reason for this is that almost all models of practical interest are *nonconvex*. The relation of these approximated models with the original model (2) has also been investigated, e.g. in [4, 38] for some physically relevant asymptotic regimes.

Mathematically, all problems arising in electronic structure theory are *nonlinear minimization problems* with possible lacks of compactness at infinity (most of them are posed on the whole space \mathbf{R}^N , and are subject to a constraint, see (8) and (11)). The Euler-Lagrange equation is a system of *nonlinear elliptic partial differential equations* such as (13). The ellipticity basically comes from the Laplacian operator, modelling the kinetic energy in (1). At one stage or another, *spectral theory* comes into the picture. More precisely, the spectral theory of *Schrödinger operators* $-\Delta + V$ often plays a key role. All this concerns the search for the ground state in the nonrelativistic setting, at zero temperature. When relativistic effects have to be accounted for, the Laplacian operator is replaced by the Dirac operator (unbounded from below), and the theoretical setting drastically changes. Important mathematical contributions on the relativistic setting are [36, 34, 35]. They have given birth to more efficient computational techniques. On the other hand, temperature effects may also be accounted for, through the introduction of a statistics, see [64] for one of the rare mathematical studies. Like for temperature

effects, the theory of excited states is not in a satisfactory state. Attempts to place the latter notion on a sound ground are [59, 23].

In the numerical practice, the problem is discretized using *Galerkin techniques*, and more precisely *spectral methods*. The basis functions used for discretization are typically gaussian approximations of the eigenfunctions of a hydrogen-like operator ($M = N = 1$ in (1)), or plane waves. The latter is very well adapted to solid state calculations. The former is incredibly efficient for calculations of molecular systems. A remarkable accuracy is reached with a limited number of basis functions. One reason why hydrogen-like basis functions outperform all other basis sets is that they are problem-dependent basis functions, which very well reproduce the exponential decrease at infinity and the cusp of wavefunctions at the point nuclei. More general purpose basis sets, such as finite elements, have difficulties in doing so, unless expensive mesh refinement techniques are employed. Finite-difference methods also exist, termed in this context *real-space* methods, but they are used for very specific applications, related to solid-state calculations.

After discretization, the equations are solved using *nonlinear optimization* techniques. Surprisingly, the problem is not addressed as a minimization problem, but in the form of the optimality equations (13). The latter is the only possible approach, considering the number of local minima, and despite the fact there is no theoretical basis for this. It reveals as an efficient approach, mostly because computations often benefit from prior calculations for adequately preconditioning the solution procedure. The algorithms in use for solving (13) are known as *SCF-algorithms*. Formally, they are elaborate variants of fixed-point iterations such as

$$-\Delta\phi_j^{n+1} + \mathcal{W}(\phi_1^n, \dots, \phi_N^n)\phi_j^{n+1} = \lambda_j^{n+1}\phi_j^{n+1}, \quad j = 1, \dots, N. \quad (14)$$

Their numerical analysis, initiated in [3], was performed only recently, see [18, 19] and [54] for a review. A rigorous mathematical insight into SCF-algorithms has led to definite improvements of their efficiency [52]. Alternative techniques may also be used. An original approach, based on *a posteriori error estimators* and related to Newton-type algorithms, is introduced in [69].

Notice that each inner loop of the nonlinear procedure involves a *linear* eigenvalue problem. This limitates the range of tractable systems (say typically that systems with a few hundreds of electrons can be standardly treated on a workstation). Ad hoc techniques may however be employed to broaden the spectrum of tractable systems. The latter are known as *linear scaling techniques*, for they significantly reduce the complexity of the diagonalization step, which in principle scales cubically with respect to the size of the system, see [12, 41, 42]. The bottom line for such a reduction is that the eigenelements are not explicitly needed: only the projector on the space spanned by the first N eigenvectors is needed for the computation of all quantities of practical interest. The problem is thus rephrased so that an explicit diagonalization is avoided. Correspondingly, advanced techniques such as *Fast Multipole techniques* [51], are used for assembling the huge matrices to be manipulated. Using a combination of such techniques, larger systems, consisting of thousands of electrons, may be treated on a workstation. The approach however still waits for a rigorous mathematical analysis.

The above description of the numerical approach concerns isolated molecular systems. Specific models are employed for the simulation of the liquid phase, and of the solid (crystalline) phase, respectively. In the former case, a commonly used setting is the *continuum model*: the molecule is placed in a cavity, surrounded by a dielectric medium modelling the solvent. Consequently, the Coulomb interaction potential appearing in V (see (9) and (12)) is replaced by the Green function of electrostatics set on the cavity. *Integral equation methods* are utilized for the numerical resolution: [20, 21]. On the other hand, the modelling of the crystal phase corresponds to a periodic setting [2]: loosely speaking, the functions ψ_j are indexed by a vector, i.e. for each k , $\psi_j^k(x)$ is the j -th eigenvector periodic in x up to a phase factor $e^{-ik \cdot x}$. In practice, the set of vectors k is discretized, and the corresponding equations (13), now indexed by k , are solved. For the practical discretization and resolution of the equations, dedicated techniques are employed: see [53, 30]. Several theoretical issues regarding the rigorous derivation of the models for the crystalline phase have already been considered: see [62, 24], and other works by the same authors. Seminal contributions by L. Van Hove, F. Dyson, A. Lenard, D. Ruelle, E. Lieb, J. Lebowitz, B. Simon, Ch. Fefferman predated those. The bottom line is to justify the models of the solid phase proving they are the limits of models for molecular systems, as the system size grows. More generally, this is part of an enormous body of literature in mathematical physics addressing questions related to *thermodynamic limits*.

2.3. Emerging approaches. Wavefunctions methods and DFT-based methods are dominant computational methods. Apart from the main stream, there are three promising tracks followed either by chemists or mathematicians, that need to be advertised. They consist in addressing the problem (2) in its original form, without any approximation, in principle.

The first approach ([28]), actually almost as old as theoretical chemistry itself, is based on a rephrasing of the minimization problem in terms of the marginals

$$\gamma(x_1, x_2, x'_1, x'_2) = \int_{\mathbf{R}^{3(N-2)}} \psi(x_1, x_2, x_3, \dots, x_N) \psi^*(x'_1, x'_2, x_3, \dots, x_N) dx_3 \dots dx_N,$$

called *second-order reduced density matrices*. This is possible because the operator (1) only involves the positions x_i and x_j of two electrons simultaneously.

The second approach (called *diffusion Monte-Carlo* in the specific context of chemistry) consists in determining the minimizer to problem (2) by solving the fictitious evolution equation

$$\frac{\partial \psi}{\partial t} + H_e \psi = 0,$$

using the Feynmann-Kac representation formula. Considering the long time limit provides with a strategy to evaluate (2), see [67, 22].

The third approach, advocated by some mathematicians ([44, 43]), consists in recognizing (4) as a high-dimensional partial differential equation and applying the techniques of *sparse tensor products*. The technique relies upon a theoretical framework set in this context in [78] (see also other works by the same author).

For all these three approaches, enormous theoretical and practical difficulties are still unsolved. For the first approach, the theoretical challenge is to determine the variational space for γ corresponding to the variational space (3) for ψ . This is where *approximations* are again introduced. The story is not closed. Current techniques rely upon *semi-definite programming* or *Augmented Lagrangian methods* to solve the associated discretized problem. Somehow related to this, the difficulty for the last two approaches lies in the fermionic nature of the electrons: the wavefunction is constrained to be antisymmetric. In addition, the problem also requires appropriate techniques such as high-dimensional integration techniques, mainly based upon Monte-Carlo, or Quasi Monte-Carlo, techniques.

These three approaches are not in position today to compete with the other more classical ones, which have benefited over the past years from constant efforts shared by a huge community. They are however instances of approaches that may be turning points and may change the landscape of computational chemistry in the years to come.

All the above describes approaches to determine the electronic structure. As mentioned in the Introduction, this is most often the inner part of a calculation. The outer part concerns the nuclei, parameters of the inner calculation so far. In the static setting, the problem is usually to determine their optimal position, i.e. the most stable conformation. This is the molecular mechanics problem (6). In biology, such a problem is crucial. It is the well-known question of determining the 3-dimensional structure of the molecule (protein,...) under study. Techniques of *discrete optimization*, *combinatorial optimization*, in particular using *stochastics-based algorithms*, are employed. Notice that the mathematical question of the existence of such a most stable configuration is mostly open for all models of interest, in spite of outstanding contributions on academic models [25].

3. Dynamical problems and problems at larger scales

Regarding time-dependent problems, the evolution of the nuclei is again often considered classical. The Newton equations of motion (7) are solved. This is the extremely popular field of molecular dynamics. It is called *ab initio* when W in the right-hand side of (7) is calculated *on-the-fly* from quantum mechanical models for the electronic structure (see [76] for a review), and *classical* when W has a parameterized analytic form, fitted on previous calculations or experiments. Parameterized potentials reportedly work well in biological applications, but experience some difficulties for materials science applications.

For the explicit evaluation of W , a very common assumption in chemistry is *adiabaticity* (see [45, 46, 75] and other works by the same authors for mathematical discussions). When adiabaticity is assumed, W ideally takes the form (2) and is computed using the static models and the techniques of Section 2.

In some cases such as collisions that involve electronic excited states with inter-

secting energy surfaces, the adiabatic approximation is not valid. Then equations (7) are coupled with the explicit time evolution of the electronic structure, simulated by (5), or one of its approximation (analogous to those of Section 2, adapted to the dynamics setting): see [17, 26, 50] for related mathematical studies.

A peculiarity to be borne in mind, which has a huge impact on the mathematical analysis, is that (7) is not only solved in order to determine the precise evolution of the system. Often, based on the *ergodicity* assumption, (7) serves as a tool for sampling the configuration space of the system in the microcanonical ensemble. Averages on this space are indeed related to quantities of macroscopic interest. Examples include the determination of the temperature, or the pressure, of a liquid system, or the determination of some mechanical properties such as the Young modulus of a crystalline solid.

Numerous challenging issues in numerical analysis arise from molecular dynamics. First, system (7) contains several, disproportionate, timescales. Think e.g. of bond lengths or angles oscillating either rapidly (i.e. at the femtosecond (10^{-15} s) scale) or slowly (hundreds of femtoseconds). Adequate techniques must be employed: multi-timestep techniques, homogenization, damping of rapid degrees of freedom, integration of differential algebraic equations. Second, the integration of (7) over long times raises specific questions: geometric integration, backward error analysis, integration of Hamiltonian, symplectic, reversible systems, etc. For related questions, reference treatises or reviews in the numerical analysis literature are e.g. [11, 58, 47, 71, 13]. See also [72, 31] in the molecular dynamics community. Third, the longest timescales that may be reached using an explicit Hamiltonian dynamics are not sufficient to cover the practical needs. Say the limit is, in good cases, the microsecond and, more generally, the nanosecond. A major reason for this is that the evolution of the system basically consists of long period of oscillations around metastable sets (basins of energy), separated by rapid hoppings between these states (*simulation of infrequent events*). Techniques for reaching extremely long simulation times or for efficiently sampling the phase space are mandatory to complement standard molecular dynamics: *stochastic differential equations, Markov chains, path integrals*, etc. In addition, other ensembles than the microcanonical ensemble may be sampled by adequate deterministic modifications of Hamiltonian dynamics (*thermostated* equations of motion) or by stochastic equations (*Langevin dynamics*). See [27, 29, 33, 32, 77] and many other references by these authors and others, for examples of techniques. The above shows that molecular dynamics problems have a twofold *multiscale nature*: even on small time frames, they involve degrees of freedom with drastically different characteristic times, and in addition to this, the integration must be carried over extremely long times. This is a significant difficulty. In spite of this, molecular dynamics simulation, along with acceleration techniques, is an extremely successful field and provides with impressively good quantitative results on some macroscopic quantities. Standard calculations on workstations simulate 10^8 atoms over the nanosecond, record calculations largely outperform this. Here again, some practical and theoretical pitfalls remain and the mathematical understanding of the methods is to be improved. In a nutshell, one could say that it is not thoroughly understood

why molecular dynamics techniques perform so well, i.e. why averages calculated from erroneous or approximate trajectories are so close to the actual values of macroscopic quantities.

The connection between the microscopic scale and the macroscopic scale is a broad subject. Calculations of ensemble averages using molecular dynamics and related techniques is one instance of it. Other questions concern the relation of molecular simulation with continuum mechanics. An example of a theoretical work in this direction is [9]. See also [10] for a review and references on the numerous practical applications, in particular applications related to computational materials science where strategies coupling molecular simulation techniques and continuum mechanics techniques are rapidly developing.

4. Trends

As briefly overviewed above, molecular simulation is an extremely rich application field of mathematics. Only a tiny part of the models and methods used in practice have been explored mathematically to date. There is much room for improvement in the mathematical understanding, the numerical analysis, the design of advanced techniques, to further enhance the field.

Some theoretical challenges concern the uniqueness of the ground state, the definition of excited states, the foundations of models at finite temperature, etc.

On the numerical side, current efforts in the mathematical community are directed towards the development of novel methods: sparse grids techniques, domain decomposition methods [7], stochastic methods for electronic structure calculations [22], methods for the determination of excited states [23], reduced basis methods [15, 16], parallel-in-time methods [5], stochastic methods for the computation of free energies [57], etc.

Acknowledgements

I wish to express my deepest gratitude to a number of colleagues for many years of collaboration on the various topics addressed in the present article. Thanks are due to P.-L. Lions, and X. Blanc, E. Cancès, I. Catto, M. J. Esteban, F. Legoll, T. Lelièvre, Y. Maday, G. Turinici.

References

- [1] Allen, M.P., Tildesley, D.J., *Computer simulation of liquids*, Oxford Science Publications, 1987.
- [2] Ashcroft, N.W., Mermin, N. D., *Solid-State Physics*, Saunders College Publishing, 1976.

- [3] Auchmuty, G., Jia, W., Convergent iterative methods for the Hartree eigenproblem, *Math. Model. and Num. Anal.*, 28 (1994), 575-610.
- [4] Bach, V., Error bound for the Hartree-Fock energy of atoms and molecules, *Comm. Math. Phys.*, 147 (1992), 527-548.
- [5] Baffico, L., Benard, S., Maday, Y., Turinici, G., Zérah, G., Parallel in time molecular dynamics simulations, *Phys. Rev. E.*, 66 (2002), 057701.
- [6] Bandrauk, A., Delfour, M., Le Bris, C., Editors, *Quantum control: mathematical and numerical challenges*, American Mathematical Society, CRM proceedings series, 2004.
- [7] Barrault, M., Cancès, E., Hager, W. W., Le Bris, C., Multilevel domain decomposition for electronic structure calculations, submitted to *J. Comp. Phys.*.
- [8] Benguria, R., Brezis, H., Lieb, E.H., The Thomas-Fermi-von Weizsäcker theory of atoms and molecules, *Comm. Math. Phys.* 79 (1981), 167-180.
- [9] Blanc, X., Le Bris, C., Lions, P.L., From molecular models to continuum mechanics, *Arch. Rat. Mech. and Anal.*, 164 (2002), 341-381.
- [10] Blanc, X., Le Bris, C., Lions, P.L., Atomistic to Continuum limits for computational materials science, *Math. Model. Num. Anal.*, to appear.
- [11] Borneman, F.A. *Homogenization in time of singularly perturbed mechanical systems*, Lectures Notes in Mathematics 1697, Springer, 1998.
- [12] Bowler, D.R., Miyazaki, T., Gillan, M.J., Recent progress in linear ab initio electronic structure techniques, *J. Phys: Condens. Matter*, 14 (2002), 2781-2798.
- [13] Budd, Ch., Piggott, M. D., Geometric Integration and its application, *Handbook of numerical analysis*, XI, North-Holland, 2004.
- [14] Cancès, E., Defranceschi, M., Kutzelnigg, W., Le Bris, C., Maday, Y., *Computational Quantum Chemistry: a Primer*, in *Handbook of Numerical Analysis, Special volume, Computational Chemistry*, volume X, North-Holland, 2003.
- [15] Cancès, E., Le Bris, C., Maday, Y., Turinici, G., Towards reduced basis approaches in ab initio electronic computations, *J. Scientific Computing*, 17 (1-4) (2002), 461-469.
- [16] Cancès, E., Le Bris, C., Maday, Y., Patera, T., Pau, G., Turinici, G., in preparation.
- [17] Cancès, E., Le Bris, C., On the time-dependent electronic Hartree-Fock equations coupled with a classical nuclear dynamics, *Math. Mod. and Meth. in App. Sci.*, 9 (1999), 963-990.
- [18] Cancès, E., Le Bris, C., On the convergence of SCF algorithms for the Hartree-Fock equations, *Math. Model. Num. Anal.*, 34 (2000), 749-774.
- [19] Cancès, E., Le Bris, C., Can we outperform the DIIS approach for electronic structure calculations, *Int. J. Quantum Chem.*, 79 (2000), 82-90.
- [20] Cancès, E., Le Bris, C., Mennucci, B., Tomasi, J., Integral equation methods for molecular scale calculations in the liquid phase, *Math. Mod. and Meth. in App. Sci.*, 9 (1999) 35-44.
- [21] Cancès, E., Mennucci, B., Tomasi, J., A new integral formalism for the polarizable continuum model: theoretical background and applications to isotropic and anisotropic dielectrics, *J. Chem. Phys.*, 101 (1997), 10506-10517.
- [22] Cancès, E., Jourdain, B., Lelièvre, T., Quantum Monte-Carlo simulations of fermions: a mathematical analysis of the fixed-node approximation, to appear in *Math. Mod. and Meth. in App. Sci.*

- [23] Cancès, E., Galicher, H., Lewin, M., Computing electronic structures: a new multi-configuration approach for excited states, *J. Comp. Phys.*, to appear.
- [24] Catto, I., Le Bris, C., Lions, P.L., *Mathematical theory of thermodynamic limits: Thomas-Fermi type models*, Oxford University Press, 1998.
- [25] Catto, I., Lions, P.L., Binding of atoms and stability of molecules in Hartree and Thomas-Fermi type theories, *Comm. Part. Diff. Equ.*, 17-18 (1992-1993), 1051-1110, 305-354, 381-429, 1149-1159.
- [26] Chadam, J.M., Glassey, R.T., Global existence of solutions to the Cauchy problem for time-dependent Hartree equations, *J. Math. Phys.*, 16 (1975), 1122-1230.
- [27] Chandler, D., Barrier crossings: classical theory of rare but important events, in *Classical and quantum dynamics in condensed phase simulations*, Berne et al. Eds., World Scientific, 1998.
- [28] Coleman, A. J., Yukalov, V.I., *Reduced density matrices*, Lecture Notes in chemistry 72, Springer, 2000.
- [29] Darve, E., Wilson, M., Pohorille, A., Calculating Free Energies Using Scaled-force Molecular Dynamics Algorithm, *Molecular Simulation*, 28 (2002), 113.
- [30] Defranceschi, M., Le Bris, C., Eds., *Mathematical models and methods for ab initio quantum chemistry*, Lecture Notes in chemistry 74 Springer, 2001.
- [31] Deuffhard, P. et al., Eds, *Computational molecular dynamics: challenges, methods, ideas*, Lecture Notes in computational science and engineering, Springer, 1999.
- [32] Deuffhard, P., Huisinga, W., Fischer, A., Schuette, Ch., Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains, *Linear Algebra Appl.*, 315, No.1-3 (2000), 39-59.
- [33] E, W., Ren, W., Vanden-Eijnden, E., String method for the study of rare events, *Phys. Rev. B.*, 66 (2002), 052301.
- [34] Dolbeault, J., Esteban, M. J., Séré, E., On the eigenvalues of operators with gaps. Application to Dirac operators. *J. Funct. Anal.*, 174 (2000), 208-226.
- [35] Dolbeault, J., Esteban, M. J., Séré, E., A variational method for relativistic computations in atomic and molecular physics. *Int. J. Quant. Chem.*, 93 (2003), 149-155.
- [36] Esteban, M. J., Séré, E., Solutions of the Dirac-Fock equations for atoms and molecules. *Comm. Math. Phys.*, 203 (1999), 499-530.
- [37] Fefferman, Ch., The N-body problem in quantum mechanics, *Commun. Pure Appl. Math.*, 39 (1986), S67-S109.
- [38] Fefferman, Ch., Seco, L.A., On the energy of a large atom, *Bull. Amer. Math. Soc.*, 23, 2, (1990), 525-530.
- [39] Frenkel, D., Smit, B., *Understanding molecular simulation*, Academic Press, Second edition, 2002.
- [40] Friesecke, G., The multiconfiguration equations for atoms and molecules: charge quantization and existence of solutions, *Arch. Rat. Mech. Anal.*, 169 (2003), 35-71.
- [41] Galli, G., Large scale electronic structure calculations using linear scaling methods, *Phys. Stat. Sol (b)*, 217 (2000), 231-249.
- [42] Goedecker, S., Linear scaling electronic structure methods, *Rev. Mod. Phys.*, 71 (1999), 1085-1123.

- [43] Griebel, M., Hamaekers, J. Sparse grids for the Schrödinger equation, submitted.
- [44] Hackbusch, W., The efficient computation of certain determinants arising in the treatment of Schrodinger's equations, *Computing*, 67 (2001), 35-56.
- [45] Hagedorn, G.A., Crossing the interface between Chemistry and Mathematics, *Notices of the Amer. Math. Soc.*, March 1996.
- [46] Hagedorn, G. A., Joye, A., Molecular propagation through small avoided crossings of electron energy levels, *Rev. Math. Phys.*, 11 (1999), 1, 41-101.
- [47] Hairer, E., Lubich, Ch., Wanner, G., *Geometric numerical integration*, Springer, 2002.
- [48] Hoffmann-Ostenhof, M., Hoffmann-Ostenhof, T., Sorensen, T. Electron wavefunctions and densities for atoms. *Ann. Inst. Henri Poincaré*, 2, no. 1 (2001), 77-100.
- [49] Hunziker, W. and Sigal, I.M., The quantum N -body problem, *J. Math. Phys.*, vol. 41 (2000), 3348-3509.
- [50] Koch, O. and Lubich, Ch., Regularity of the multi-configuration time-dependent hartree approximation in quantum molecular dynamics, submitted to M2AN.
- [51] Kudin, K., Scuseria, G.E., A fast multipole algorithm for the efficient treatment of the Coulomb problem in electronic structure calculations of periodic systems with Gaussian orbitals, *Chem. Phys. Lett.*, 289 (1998), 611-616.
- [52] Kudin, K., Scuseria, G.E., Cancès, E., A black-box self-consistent field convergence algorithm: one step closer, *J. Chem. Phys.*, 116 (2002), 8255-8261.
- [53] Le Bris, C., Guest-Editor, *Handbook of Numerical Analysis, Special volume, Computational Chemistry*, volume X, North-Holland, 2003.
- [54] Le Bris, C., Computational chemistry from the perspective of numerical analysis, *Acta Numerica*, 14 (2005), 363-444.
- [55] Le Bris, C., Lions, P.L., From atoms to crystals: a mathematical journey, *Bull. Amer. Math. Soc.*, 42 (2005), 291-363.
- [56] Le Bris, C., Patera, A.T., Eds. Computational chemistry, a special issue of *Math. Model. Num. Anal.*, to appear.
- [57] Le Bris, C., Lelièvre, T., Vanden-Eijnden, E., Sampling Boltzmann-Gibbs distributions restricted on a manifold with diffusions, submitted.
- [58] Leimkueller, B., Reich, S., *Geometric integrators in Hamiltonian Mechanics*, Cambridge University Press, 2004.
- [59] Lewin, M., Solutions of the Multiconfiguration Equations in Quantum Chemistry, *Arch. Rat. Mech. Anal.*, 171, 1 (2004), 83-114.
- [60] Lieb, E.H., The stability of matter: from atoms to stars, *Bull. A.M.S.*, 22 (1990), 1-49.
- [61] Lieb, E.H., Simon, B., The Hartree-Fock theory for Coulomb systems, *Comm. Math. Phys.*, 53 (1977), 185-194.
- [62] Lieb, E.H., Simon, B., The Thomas-Fermi theory of atoms, molecules and solids, *Advances in mathematics*, 23 (1977), 22-116.
- [63] Lieb, E.H., Thomas-Fermi and related theories of atoms and molecules, *Rev. Mod. Phys.*, 53 (1981), 603-641.

- [64] Lions, P.-L., Hartree-Fock and related equations, Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. IX (Paris, 1985–1986), Pitman Res. Notes Math. Ser., volume 181, 304-333.
- [65] Lions, P.-L., Solutions of Hartree-Fock equations for Coulomb systems, *Comm. Math. Phys.*, 109 (1987), 33-97.
- [66] Lions, P.-L., The concentration-compactness principle in the calculus of variations. The locally compact case. I and II. *Ann. Inst. Henri Poincaré, Anal. Non Linéaire* 1, (1984), 109-145 and 223-283.
- [67] Lions, P.-L., Remarks on mathematical modelling in quantum chemistry, *Computational Methods in Applied Sciences*, Wiley, (1996) 22-23.
- [68] McWeeny, R., *Methods of molecular quantum mechanics*, 2nd edition, Academic Press, 1992.
- [69] Maday, Y., Turinici, G., Error bars and quadratically convergent methods for the numerical simulation of the Hartree-Fock equations, *Numer. Math.*, 94, 4 (2003), 739-770.
- [70] Parr, R. G., Yang, W., *Density functional theory of atoms and molecules*, Oxford University Press, 1989.
- [71] Sanz-Serna, J.M., Calvo, M.P., *Numerical Hamiltonian Problems*, Chapman and Hall, 1994.
- [72] Schlick, T., Ed., Computational Molecular Biophysics, *J. Comp. Phys.*, Special Issue, 151 (1999).
- [73] Solovej, J.P., Universality in the Thomas-Fermi-von Weizsäcker model of atoms and molecules, *Comm. Math. Phys.*, 129 (1990), 561-598.
- [74] Spruch, L., Pedagogic notes on Thomas-Fermi theory (and on some improvements): atoms, stars and the stability of bulk matter, *Rev. Mod. Phys.*, 63, (1991), 151-209.
- [75] Teufel, S., *Adiabatic perturbation theory in quantum dynamics*, Lecture Notes in Mathematics, 1821, Springer, 2003.
- [76] Tuckerman, M. E., Ab initio molecular dynamics: Basic concepts, current trends and novel applications, *J. Phys. Condens. Matter*, 14 (2002), R1297-R1355.
- [77] Voter, A. F., Sorensen, M. R., Accelerating atomistic simulations of defect dynamics: Hyperdynamics, parallel replica dynamics and temperature accelerated dynamics, *Mat. Res. Soc. Symp. Proc.*, 538 (1999), 427-439.
- [78] Yserentant, H., On the regularity of the electronic Schrödinger equation in Hilbert spaces of mixed derivatives, *Numer. Math.*, 98 (2004), 4, 731-759.

CERMICS, École Nationale des Ponts et Chaussées, 6 & 8, avenue Blaise Pascal, 77455 Marne-La-Vallée, FRANCE, and INRIA Rocquencourt, MICMAC project, Domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, FRANCE
E-mail: lebris@cermics.enpc.fr