

Publication Series of the John von Neumann Institute for Computing (NIC)
NIC Series

Volume 10

John von Neumann Institute for Computing (NIC)

Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms

edited by

Johannes Grotendorst
Dominik Marx
Alejandro Muramatsu

Winter School, 25 February - 1 March 2002
Rolduc Conference Centre, Kerkrade, The Netherlands
Lecture Notes

organized by

John von Neumann Institute for Computing
Ruhr-Universität Bochum
Universität Stuttgart

NIC Series

Volume 10

ISBN 3-00-009057-6

Die Deutsche Bibliothek – CIP-Cataloguing-in-Publication-Data
A catalogue record for this publication is available from Die Deutsche
Bibliothek.

Publisher: NIC-Directors

Distributor: NIC-Secretariat
Research Centre Jülich
52425 Jülich
Germany

Internet: www.fz-juelich.de/nic

Printer: Graphische Betriebe, Forschungszentrum Jülich

© 2002 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work
for personal or classroom use is granted provided that the copies
are not made or distributed for profit or commercial advantage and
that copies bear this notice and the full citation on the first page. To
copy otherwise requires prior specific permission by the publisher
mentioned above.

NIC Series Volume 10
ISBN 3-00-009057-6

Preface

This Winter School continues a series of schools and conferences in Computational Science organized by the John von Neumann Institute for Computing (NIC). The topics of the School, Quantum Monte Carlo and Quantum Molecular Dynamics, play an outstanding role in many NIC research projects which use the supercomputing facilities provided by the Central Institute for Applied Mathematics (ZAM) of the Research Centre Jülich. The programme of the Winter School covers modern quantum simulation techniques and their implementation on high-performance computers, in particular on parallel systems. The focus clearly is on numerical methods which are tailored to treat large quantum systems with many coupled degrees of freedom ranging from superfluid Helium to chemical reactions. Among others, the following topics are treated by twenty-five lectures:

- Diffusion and Green's function Monte Carlo
- Path integral Monte Carlo and Molecular Dynamics
- Car-Parrinello / ab initio Molecular Dynamics
- Real-time quantum dynamics for large systems
- Lattice and continuum algorithms
- Exchange statistics for bosons and fermions / sign problem
- Parallel numerical techniques and tools
- Numerical integration and random numbers

This strongly interdisciplinary School aims at bridging three “gaps” in the vast field of large-scale quantum simulations. The first gap is between chemistry and physics, the second one between typical graduate courses in these fields and state-of-the-art research, and finally the one between the Monte Carlo and Molecular Dynamics communities. The participants will benefit from this School by learning about recent methodological advances within and outside their field of specialization. In addition, they get insight into recent software developments and implementation issues involved, in particular in the context of high-performance computing.

The lecturers of this Winter School come from chemistry, physics, mathematics and computer science and this is true for the audience as well. Participants from thirty mainly European countries attend the NIC Winter School, and eighty contributions have been submitted for the poster sessions. This overwhelming international resonance clearly reflects the attractiveness of the programme and demonstrates the willingness of the participants to play an active role in this high-level scientific School.

The scientific programme was worked out by Johannes Grotendorst (Research Centre Jülich), Dominik Marx (Ruhr-Universität Bochum), and Alejandro Muramatsu (Universität Stuttgart). The programme structure consists of overview lectures on various important fields, focus lectures on Quantum Monte Carlo and Quantum Molecular Dynamics methods, and special lectures on numerical and computational techniques.

Many organizations and individuals have contributed significantly to the success of this Winter School. Without the financial support of the European Commission within the framework of the specific research and training programme “Improving Human Research Potential” this one-week School on quantum simulation methods would not have been possible. We are grateful for the generous financial support by the Federal Ministry for Education and Research (BMBF) and by the Research Centre Jülich as well as for the help provided by its Conference Service and its Central Institute for Applied Mathematics.

We are greatly indebted to the local organization committee at Forschungszentrum Jülich who did the bigger part of the preparing work, namely Rüdiger Esser (Finance), Bernd Krahl-Urban (Accommodation and Registration) and Monika Marx (Web Management, Proceedings), and last but not least the conference secretaries Yasmin Abdel-Fattah, Elke Bielitz and Anke Reinartz. Special thanks go to Monika Marx for her tireless commitment concerning the editing and realization of this book. Furthermore, we appreciate the work of Stephan Brück who supported the difficult typesetting with great care. Finally, we would like to thank both the Ruhr-Universität Bochum and the Universität Stuttgart for their support of this activity in the area of high-end scientific education.

The nature of a Winter School requires the notes of the lectures to be available at the meeting. In this way, the participants have the chance to work through the lectures thoroughly during or after the lectures. We are very thankful to all authors who provided written contributions to this book of lecture notes. It is intended to serve as a future standard reference to the rapidly evolving field of quantum simulations of complex many-body systems. The articles give a broad review of modern time-independent and time-dependent methods as well as of the relevant state-of-the-art numerical and parallel computation techniques. In addition to such traditional text-based proceedings, audio-visual proceedings will be produced. All lectures will be recorded on video. After the School these recordings combined with the slides will be made available to the participants on DVD and to the general scientific community in the internet at <http://www.fz-juelich.de/nic-series/volume10>, the same place where the book of lecture notes is published.

Jülich, Bochum, and Stuttgart
February 2002

Johannes Grotendorst
Dominik Marx
Alejandro Muramatsu

Contents

Time-Independent Quantum Simulation Methods

Monte Carlo Methods: Overview and Basics

<i>Marius Lewerenz</i>	1
1 Introduction	1
2 Review of Probability and Statistics	6
3 Sources of Randomness	14
4 Monte Carlo Integration	20

Diffusion and Green's Function Quantum Monte Carlo Methods

<i>James B. Anderson</i>	25
1 Introduction	25
2 History and Overview	26
3 Variational Quantum Monte Carlo	28
4 Diffusion Quantum Monte Carlo	29
5 Green's Function Quantum Monte Carlo	33
6 Node Structure	34
7 Importance Sampling	35
8 Trial Wavefunctions	37
9 Fixed-Node Calculations	39
10 Exact Cancellation Method	40
11 Difference Schemes	43
12 Excited States	45
13 Use of Pseudopotentials	45

Path Integral Monte Carlo

<i>Bernard Bernu, David M. Ceperley</i>	51
1 Introduction	51
2 Mapping of the Quantum to a Classical Problem	52
3 Bose Symmetry	57
4 Applications	58

Exchange Frequencies in 2D Solids: Example of Helium 3 Adsorbed on Graphite and the Wigner Crystal

<i>Bernard Bernu, Ladir Cândido, David M. Ceperley</i>	63
1 Introduction	63
2 PIMC Method	65
3 Reaction Coordinate	67
4 A One Dimensional Toy Model: A Particle in a Symmetrical Double Well	68
5 Results and Magnetic Phase Diagram	70
6 Conclusion	73

Reptation Quantum Monte Carlo	
<i>Stefano Baroni, Saverio Moroni</i>	75
1 Introduction	75
2 From Classical Diffusion to Quantum Mechanics	76
3 From Quantum Mechanics Back to Classical Diffusion	81
4 The Algorithm	86
5 A Case Study of ^4He	88
6 Conclusions	96
Quantum Monte Carlo Methods on Lattices: The Determinantal Approach	
<i>Fakher F. Assaad</i>	99
1 Introduction	99
2 The World Line Approach for the XXZ Model and Relation to the 6-Vertex Model	101
3 Auxiliary Field Quantum Monte Carlo Algorithms	107
4 Application of the Auxiliary Field QMC to Specific Hamiltonians	129
5 The Hirsch-Fye Impurity Algorithm	144
6 Conclusion	147
Effective Hamiltonian Approach for Strongly Correlated Lattice Models	
<i>Sandro Sorella</i>	157
1 Introduction	157
2 The Lanczos Technique	159
3 The Effective Hamiltonian Approach	161
4 The Generalized Lanczos	163
5 Results on the t-J Model	168
6 Conclusions	172
The LDA+DMFT Approach to Materials with Strong Electronic Correlations	
<i>Karsten Held, Igor A. Nekrasov, Georg Keller, Volker Eyert, Nils Blümer, Andrew K. McMahan, Richard T. Scalettar, Thomas Pruschke, Vladimir I. Anisimov, Dieter Vollhardt</i>	175
1 Introduction	175
2 The LDA+DMFT Approach	177
3 Comparison of Different Methods to Solve DMFT: The Model System $\text{La}_{1-x}\text{Sr}_x\text{TiO}_3$	191
4 Mott-Hubbard Metal-Insulator Transition in V_2O_3	194
5 The Cerium Volume Collapse: An Example for a 4f-Electron System	198
6 Conclusion and Outlook	203

Time-Dependent Quantum Simulation Methods

Classical Molecular Dynamics

<i>Godehard Sutmann</i>	211
1 Introduction	211
2 Models for Particle Interactions	215
3 The Integrator	221
4 Simulating in Different Ensembles	229
5 Parallel Molecular Dynamics	235

Static and Time-Dependent Many-Body Effects via Density-Functional Theory

<i>Heiko Appel, Eberhard K. U. Gross</i>	255
1 Introduction	255
2 Basic Concepts of DFT	256
3 Propagation Methods for the TDKS Equations	259
4 Examples for the Solution of the TDKS Equations	263

Path Integration via Molecular Dynamics

<i>Mark E. Tuckerman</i>	269
1 Introduction	269
2 The Density Matrix and Quantum Statistical Mechanics	270
3 Path Integral Formulation of the Canonical Density Matrix and Partition Function	272
4 The Continuous Limit	275
5 Thermodynamics and Expectation Values in Terms of Path Integrals	281
6 Path Integral Molecular Dynamics	283
7 Many-Body Path Integrals	294
8 Summary	297

Ab Initio Molecular Dynamics and Ab Initio Path Integrals

<i>Mark E. Tuckerman</i>	299
1 Introduction	299
2 The Born-Oppenheimer Approximation and <i>Ab Initio</i> Molecular Dynamics	300
3 Plane Wave Basis Sets	305
4 The Path Integral Born-Oppenheimer Approximation and <i>Ab Initio</i> Path Integral Molecular Dynamics	311
5 Illustrative Applications	314

Dynamic Properties via Fixed Centroid Path Integrals

<i>Rafael Ramírez, Telesforo López-Ciudad</i>	325
1 Introduction	325
2 Definition of Auxiliary Quantities	327
3 Definition of Fixed Centroid Path Integrals	328
4 The Schrödinger Formulation of Fixed Centroid Path Integrals	330
5 Constrained Time Evolution of the Operator $\hat{\sigma}(X, P)$	337
6 Numerical Test on Model Systems	344

7	A Review of CMD Applications	351
8	Open Problems	355
9	Conclusions	356

Quantum Molecular Dynamics with Wave Packets

<i>Uwe Manthe</i>	361	
1	Introduction	361
2	Spatial Representation of Wavefunctions	362
3	Propagation of Wave Packets	366
4	Iterative Diagonalization	369
5	Filter Diagonalization	370
6	MCTDH	371

Nonadiabatic Dynamics: Mean-Field and Surface Hopping

<i>Nikos L. Doltsinis</i>	377	
1	Introduction	377
2	Born-Oppenheimer Approximation	378
3	Semiclassical Approach	380
4	Approaches to Nonadiabatic Dynamics	382

Relieving the Fermionic and the Dynamical Sign Problem: Multilevel Blocking

Monte Carlo Simulations

<i>Reinhold Egger, Chi H. Mak</i>	399	
1	Introduction: The Sign Problem	399
2	Multilevel Blocking (MLB) Approach	401
3	Applications	412
4	Concluding Remarks	421

Numerical Methods and Parallel Computing

Statistical Analysis of Simulations: Data Correlations and Error Estimation

<i>Wolfhard Janke</i>	423	
1	Introduction	423
2	Model Systems and Phase Transitions	424
3	Estimators, Autocorrelation Times, Bias and Resampling	430
4	A Simplified Model	435
5	A Realistic Example	439
6	Error Propagation in Multicanonical Simulations	440
7	Summary	443

Pseudo Random Numbers: Generation and Quality Checks

<i>Wolfhard Janke</i>	447	
1	Introduction	447
2	Pseudo Random Number Generators	447
3	Quality Checks	451

4	Non-Uniform Pseudo Random Numbers	454
5	Summary	456

Integrators for Quantum Dynamics: A Numerical Analyst's Brief Review

<i>Christian Lubich</i>	459	
1	Introduction	459
2	The Implicit Midpoint Rule	460
3	The Exponential Midpoint Rule	461
4	Strang Splitting	461
5	Chebyshev Approximation	462
6	Lanczos Approximation	462
7	Magnus Integrators	463
8	Integrators for Almost-Adiabatic Quantum Dynamics	464

Long-Range Interactions in Many-Particle Simulation

<i>Paul Gibbon, Godehard Sutmann</i>	467	
1	Introduction	467
2	Ewald Summation	468
3	Particle-Mesh Techniques	477
4	Multipole Methods	486
5	Performance and Parallelism	498
6	Summary	502

Parallel Programming Models, Tools and Performance Analysis

<i>Bernd Mohr, Michael Gerndt</i>	507	
1	Introduction	507
2	Programming Models	510
3	Parallel Debugging	515
4	Parallel Performance Analysis	515
5	Summary	518

Iteratively Solving Large Sparse Linear Systems on Parallel Computers

<i>Martin Bücker</i>	521	
1	An Algorithmic Shift in Large-Scale Computations	521
2	Difficulties with Direct Methods	522
3	Classical Iterations	526
4	Projection Methods	528
5	Krylov Subspace Methods	529
6	Preconditioning	539
7	Reducing Synchronization	540
8	Matrix-Vector Multiplications and Graph Partitioning	541
9	Concluding Remarks	544
10	Bibliographic Comments	544

Monte Carlo Methods: Overview and Basics

Marius Lewerenz

LADIR/Spectrochimie Moléculaire, UMR 7075
Bât F74, BP 49, Université Pierre et Marie Curie
4, Place Jussieu, F-75252 Paris Cedex 05
E-mail: lewerenz@spmol.jussieu.fr

The first part of this introductory lecture on Monte Carlo methods gives an overview of the essential ideas of Monte Carlo, discusses the relation between the basic sampling concept and statistical mechanics and the probabilistic interpretation of quantum mechanics, and provides a classification of existing Monte Carlo methods. This part is followed by a summary of essential concepts of probability and statistics, the construction of random walks, and the application of random sampling in the estimation of integrals.

1 Introduction

The original title of this lecture was supposed to be “Classical Monte Carlo”, which is somewhat surprising for a winter school about quantum simulations. Since this is the first scientific lecture of this week I decided to interpret the title in a more general sense of (i) providing a sort of classification of the existing Monte Carlo methods, (ii) discussing certain algorithmic elements which were originally invented for classical Monte Carlo simulations but turn out to be useful also for quantum simulations, and last but not least (iii) giving some guidelines of how to handle the stochastic nature of the results. Many of these subjects will be revisited by other speakers in specialized lectures during this week.

The speakers in this workshop come from chemistry, physics, mathematics, and computer science and this is probably true about the audience as well. This fact illustrates the generality of both Monte Carlo and Molecular Dynamics methods and reflects the algorithmic challenges implicit in the simulation of ever larger systems on computers of increasing complexity. The requirements on the stability of trajectories over long simulation times or on the quality of random numbers are continuously increasing.

The section on elementary aspects of random variables and random walks is important because quantum chemists in the broad sense, ranging from electronic structure to reaction dynamics, are mostly not used to error bars in their data. Of course this does not imply the absence of errors in these deterministic calculations, but at least for certain computational techniques we know that we should observe monotonic convergence to a well defined result if a computational parameter (basis set size, grid resolution etc.) is changed. The presence of statistical noise in the result of a computation requires a different spirit in the assessment of the reliability of a calculation and a redefinition of the idea of convergence. The finite error margin on the result from any kind of Monte Carlo calculation can also make it hard to detect subtle errors in a code or deficiencies in the quality of the underlying source of randomness.^{1,2}

In production mode a proper Monte Carlo code will give a different result each time it is run. For test purposes, however, perfect reproducibility has to be possible as an option. A typical Monte Carlo code is far less complex than an electronic structure code but this

apparent simplicity should not lead to an underestimate of the possible troubles. We want to use random numbers to produce results, but we do not want these results to be random!

1.1 What is Monte Carlo

The keyword “Monte Carlo” in the name of a method only indicates that there is some stochastic element in it.^{3–8} Stochastic methods are clearly appropriate for the description of stochastic processes but we will see that it can be useful to replace a deterministic problem by an equivalent stochastic problem. We will restrict ourselves here to stochastic methods of physical or chemical relevance.

The central idea of Monte Carlo methods is to represent the solution of a mathematical (or in our case physical) problem by a parameter of a true or hypothetical distribution and to estimate the value of this parameter by sampling from this distribution. This idea provides a direct tie to the ensemble concept introduced into statistical mechanics by Gibbs and the probabilistic interpretation of wave functions in quantum mechanics. Since we know how to translate a thermodynamic problem into its corresponding ensemble we have, at least in principle, a direct prescription for the computation of an observable, namely to construct a sufficiently large number of microstates compatible with the specified macroscopic variables and to average their properties. The distribution function for the microstates is generally known for this type of problem.

Once we know the wave function of a physical system we can imagine to compute expectation values by drawing samples from the associated probability density. This quantum Monte Carlo approach is very interesting because it provides a very practical interpretation of the abstract concept of a wave function. Unfortunately exact wave functions are known for only a very limited number of systems. However, as you will learn during this workshop, there are several Monte Carlo methods which achieve the difficult goal of simultaneously creating and sampling from the a priori unknown quantum distribution.

Monte Carlo methods play a very important role in statistical physics^{9–14} and have led to a very high technical standard of algorithms and methods to quantify systematic and statistical errors. While classical Monte Carlo has a well established place in chemistry in particular in the (numerical) theory of liquids,^{15,16} the group of methods customarily referred to as quantum Monte Carlo is used by only a relatively small, but rapidly increasing number of people. Unfortunately the transfer of methodological knowledge between statistical physics, electronic structure applications of quantum Monte Carlo, and the rapidly growing group of people interested in dynamical applications is still relatively poor. Specifically in my own field of the application of quantum Monte Carlo techniques to vibrational problems a number of far from optimal algorithms are in widespread use. Hopefully this workshop can contribute to change this situation.

1.2 Monte Carlo vs Molecular Dynamics

Monte Carlo methods, which are usually based on random walks and thus sequences of events, contain an apparent dynamics. It is, however, crucial to recognize that the sequence of events does not correspond to a sequence in real time, as opposed to the role of time as an explicit variable in molecular dynamics methods. The sequence of events is generally independent of the physical phenomena which we wish to describe by a Monte

Carlo method and depends on the chosen algorithm. The creation of smart sequences is in fact an important design goal in the construction of efficient Monte Carlo algorithms.

The absence of time in Monte Carlo algorithms implies the fundamental inability to describe non equilibrium processes. We will, however, encounter the concept of imaginary time, which should rather be regarded as inverse energy, in several quantum Monte Carlo methods in the lectures of my colleagues J. Anderson and D. Ceperley.

Why then do we need Monte Carlo methods if the most common ensembles of statistical mechanics (microcanonical and canonical) are equally well accessible by molecular dynamics methods which give access to time dependent and time independent properties? Besides being easily able to describe situations, which are difficult (e.g. grand canonical ensemble) or impossible (e.g. discrete models, see F. Assaad, S. Sorella) to cast into equations of motion, well designed Monte Carlo methods offer more chances of overcoming one of the most severe problems of simulation methods, namely exhaustive sampling of the relevant configuration space. The local configuration updates imposed by the small time steps in the discretized version of the equations of motion (systematic discretisation error!) allow only slow exploration of configuration space by dynamical methods through a sequence of very many small steps. Our only chance for improvement is a faster computer and maybe a better code design.

Monte Carlo methods do not have the equivalent of a time step error (the target distribution is sampled exactly with only statistical errors) and they allow a great degree of freedom in the design of the sequence of steps, potentially including smart global configuration updates which permit rapid exploration of problems which may suffer from significant ergodicity problems in an equivalent dynamical simulation. Monte Carlo methods allow to replace computer power by creative brain power. There are also relatively straightforward applications of Monte Carlo methods to quantum systems, as you will see in the other lectures of this winter school.

The essential conclusion is that molecular dynamics and Monte Carlo methods do not compete but complement each other.

1.3 Classical vs Quantum Monte Carlo

The existing Monte Carlo methods can be divided into methods which assume that classical mechanics is applicable and that consequently energy is a continuous variable and those which are based on the idea of discrete quantum energy levels. While all of the latter methods are sometimes referred to as quantum Monte Carlo methods, I personally prefer to distinguish methods which assume a known set of energy levels (usually from a model hamiltonian) from those in which the Monte Carlo method is actually used to find these energy levels.

This difference becomes evident if we look at the other physically important distinction, namely the role of temperature. In the classical limit the investigation of the system at T=0 corresponds to finding the global minimum on a multidimensional potential surface, which is conceptually simple (but in practice can be formidable problem). At finite temperature, Monte Carlo methods are used to sample points in configuration space according to a known probability distribution for the available energies (e.g. the Boltzmann distribution). Averages over these points are used to estimate the expectation value of any property of interest which is formally defined as an integral in statistical mechanics. Here

	Classical	Quantum	
$T = 0$	Locating the minimum of a multidimensional surface, e.g. Simulated annealing	Single occupied quantum state of known energy	Single quantum state(s) with unknown properties: variational Monte Carlo (VMC), diffusion Monte Carlo (DMC)
$T > 0$	Classical Monte Carlo (CMC) in various ensembles (Integration over continuous states)	Summation over discrete states (lattice model Hamiltonians, Ising etc.), technically similar to CMC	Direct averaging over many quantum states: path integral Monte Carlo (PIMC)
	known energy levels E_i		unknown energy levels E_i

we have a typical example for the replacement of a deterministic problem (performing a high dimensional integral) by an equivalent stochastic problem.

The same basic techniques can be applied to model quantum systems, where the T=0 situation now corresponds to finding the combination of discrete variables (quantum numbers, spin orientations etc.) which give the lowest energy. This is again a conceptually simple problem, even though in practice the huge size of the available (now discrete) configuration space can still be very troublesome. The computation of thermodynamic averages at $T > 0$ for this type of system does not require any profound modifications of the algorithms of classical Monte Carlo. Instead of performing a random walk in continuous coordinates one attempts to flip discrete spins or to change quantum numbers (sometimes with very elaborate algorithms^{17,18}) before evaluating the quantum energy of the system from a simple formula, which then in turn determines the continuation of the random walk. The Monte Carlo method is used here only to perform the average over available states, but not to actually compute the quantum states or their properties.

The situation becomes much more complicated if we do not even know the states which are available in a given quantum system. Searching for the properties of individual pure quantum states, independent of their occupation, eliminates temperature as a variable, which can formally be set to zero. The simplest of these states, and of course the only one occupied at T=0, is the quantum ground state. The goal of methods like variational Monte Carlo (VMC), Green's function Monte Carlo (GFMC),^{19–21} and its most common variant diffusion Monte Carlo (DMC)^{22–24} is the calculation of wave functions, energies, and other properties of pure states. Whereas the VMC method relies on exploring a proposed quantum distribution and its subsequent optimisation, the distribution from which

the samples should be taken is a priori unknown in GFMC and DMC and is constructed in the course of the Monte Carlo calculation.

Once a sufficiently complete set of eigenstates has been computed by one of these methods, thermodynamic averages at $T>0$ could in principle be computed by direct summation or application of sampling techniques as above. Unfortunately even the computation of a few states is still not easy, as we will see in this workshop.

However, if we are willing to sacrifice detailed knowledge of individual quantum states, direct sampling of the density matrix is possible by the numerical implementation of the path integral approach to quantum mechanics,^{25,26} which again replaces integrals by averages over samples and is known as path integral Monte Carlo (PIMC). This latter method can be mapped onto a classical problem²⁷ with specific modifications to account for quantum statistics.²⁸

This last group of methods merits the designation “quantum Monte Carlo” in the more precise sense of attempts to use Monte Carlo techniques to solve the actual quantum problem. You will hear more about the theory underlying these methods and their implementation and application to actual problems in other lectures of this course. Most of the efforts of advancing quantum Monte Carlo technology in molecular physics (VMC, GFMC, DMC) are oriented towards electronic structure^{29,24,30–33} but the methods are equally applicable to rovibrational problems. While the fermion statistics of electrons causes particular trouble in the construction of stable algorithms for electronic structure problems, the rovibrational ground state of most molecular systems can be computed exactly with these methods.

1.4 Why do we need Quantum Monte Carlo?

Quantum Monte Carlo methods are not here to replace other methods in electronic or vibrational structure, but are an interesting complement to these more conventional methods. The GFMC approach, and DMC as its most common implementation, in principle only needs a potential energy surface. While basis set methods have problems to provide enough flexibility to span large configuration spaces and usually require basis sets specially tailored to a given problem in order to remain technically feasible, DMC is an intrinsically global method.

Due to its general applicability its value is probably even higher in vibrational problems where the different shapes of interaction potentials require a lot of care in basis function design as opposed to electronic structure, where the potential is always just a sum of Coulomb terms and where there is a well developed general technology based on Gaussian basis functions. The construction of the random walks allows the treatment of large amplitude motion and of wave functions which are spread over a large number of potential minima. Of course we have to pay a price for this: DMC will only give a single wave function and not a whole spectrum, and even this single wave function has an unusual representation and the extraction of unbiased expectation values is rather complicated.

Being an exact method except for statistical errors, one can easily follow the evolution of the quantum properties of systems of increasing complexity (e.g. cluster size effects) without biasing the results through the introduction of more and more approximations or more stringent basis limitations. This attractive scaling of the accuracy is accompanied by a relatively modest growth of the computational effort. The only serious, and often

underestimated, problem is the practical achievement of the full exploration of the relevant configuration space, in other words the ergodicity of a random walk of finite length.

Presenting this topic to an audience of young researchers is probably a good occasion to give a brief list of open problems which might be interesting to look into. On the NCSA web site <http://archive.ncsa.uiuc.edu/Apps/CMP/topten/topten.html> you can find the following list of reasons why quantum Monte Carlo methods are still not very popular in electronic structure theory.

Top reasons why quantum Monte Carlo is not generally used in chemistry:

- We need forces, dummy!
- Try getting O₂ to bind at the variational level.
- How many graduate students lives have been lost optimizing wavefunctions?
- It is hard to get 0.01 eV accuracy by throwing dice.
- Most chemical problems have more than 50 electrons.
- Who thought LDA or HF pseudopotentials would be any good?
- How many spectra have you seen computed by QMC?
- QMC is only exact for energies.
- Multiple determinants. We can't live with them, we can't live without them.
- After all, electrons are fermions.
- Electrons move.
- QMC isn't included in Gaussian 90. Who programs anyway?

From my own experience with rovibrational problems I might add:

- How to construct trial wave functions for arbitrary potentials?
- We want a lot of excited states.
- We want rigid body constraints.
- How to handle almost degenerate states?

2 Review of Probability and Statistics

This section introduces several basic notions which we will need to describe the data produced by a Monte Carlo calculation and to specify individual components in the construction of random walks. The presentation follows essentially the discussion given in Ref. 4.

2.1 Probabilities and Random Variables

We consider a reservoir of possible outcomes $\{E\}$ for a random event.

$$\{E\} = \{E_1, E_2, E_3 \dots E_n\} \quad (1)$$

We associate a probability p_k with each E_k :

$$P(E_k) = p_k \quad 1 \geq p_k \geq 0 \quad (2)$$

Properties of p_k :

1. The following relations hold for any pair of E_i, E_j .

$$P(E_i \wedge E_j) \leq p_i + p_j \quad (3)$$

$$P(E_i \vee E_j) \leq p_i + p_j \quad (4)$$

2. If E_i and E_j are mutually exclusive

$(E_i \Rightarrow \neg E_j, E_j \Rightarrow \neg E_i)$:

$$P(E_i \wedge E_j) = 0 \quad (5)$$

$$P(E_i \vee E_j) = p_i + p_j \quad (6)$$

3. For a class of mutually exclusive events, which contains all possible events we have:

$$P(\text{some } E) = 1 = \sum_i p_i \quad (7)$$

2.2 Joint and Marginal Probabilities

Suppose that the events E_i and F_j satisfy the conditions defined above with associated probabilities p_{1i} and p_{2j} ,

$$P(E_i) = p_{1i} \quad P(F_j) = p_{2j} \quad , \quad (8)$$

and we are interested in the probability of the combined event (E_i, F_j) . We define the probability of this event as the *joint probability*

$$P(E_i, F_j) = p_{ij} \quad (9)$$

The events E_i and F_j are called independent if the probability of the combined event can be expressed as

$$p_{ij} = p_{1i}p_{2j} \quad . \quad (10)$$

If the events E_i and F_j are not independent, i. e. $p_{ij} \neq p_{1i}p_{2j}$, it is useful to decompose the *joint probability* as follows:

$$p_{ij} = \left(\sum_k p_{ik} \right) \left[\frac{p_{ij}}{\sum_k p_{ik}} \right] \quad (11)$$

$$p_{ij} = p(i) \left[\frac{p_{ij}}{\sum_k p_{ik}} \right] \quad (12)$$

The quantity $p(i)$ is called the *marginal probability* for the event E_i , the probability of observing E_i combined with any event in the reservoir $\{F\}$. Clearly $p(i) = p_{1i}$ and $\sum_i p(i) = \sum_i \sum_k p_{ik} = 1$.

The second factor $p_{ij}/\sum_k p_{ik} = p(j|i)$ defines the *conditional probability* of observing F_j , provided E_i has occurred. Since we are certain to observe one of the possible F_j in combination with E_i we clearly have

$$\sum_j p(j|i) = \sum_j \frac{p_{ij}}{\sum_k p_{ik}} = \frac{\sum_j p_{ij}}{\sum_k p_{ik}} = 1 \quad (13)$$

The concept of conditional probability is an important element in the construction of random walks, because certain events will be possible only if they are preceded by particular other events.

2.3 Random Variables and Expectation Values

The random events discussed above E, F can be anything of numerical or non numerical character (e.g. a noise amplitude or a logical decision). If we can associate a numerical value x_i with each random event E_i , we call x a random variable.

We define the *expectation value* $E(x)$ of a random variable x as

$$E(x) = \langle x \rangle = \sum_i p_i x_i \quad (14)$$

Assume that g is a function of x , $g(x_i) = g_i$. Then also g_i will be a random variable and we define

$$E(g(x)) = \langle g(x) \rangle = \sum_i p_i g(x_i) \quad (15)$$

Suppose that $g(x_i) = g(x) = \text{const}$:

$$E(g(x)) = \sum_i p_i g(x_i) = g(x_i) \sum_i p_i = g(x) \quad (16)$$

We conclude that the expectation value of a constant is a constant.

In the next step we prove the linearity of the expectation value of two random functions $g_1(x)$ and $g_2(x)$ by substitution of the definition of an expectation value in terms of probabilities:

$$\begin{aligned} E(\lambda_1 g_1(x) + \lambda_2 g_2(x)) &= \langle \lambda_1 g_1(x) + \lambda_2 g_2(x) \rangle \\ &= \sum_i p_i (\lambda_1 g_1(x_i) + \lambda_2 g_2(x_i)) \\ &= \lambda_1 \sum_i p_i g_1(x_i) + \lambda_2 \sum_i p_i g_2(x_i) \\ &= \lambda_1 \langle g_1(x) \rangle + \lambda_2 \langle g_2(x) \rangle \end{aligned}$$

2.4 Moments of a Distribution

We define the n th moment of a distribution as

$$\mu_n = \langle x^n \rangle = \sum_i p_i x_i^n \quad (17)$$

These powers of x are nothing but special cases of the random functions $g(x)$.

Principal moments:

$$\begin{aligned} \mu_1 &= \sum_i p_i x_i && \text{mean of the distribution} \\ \mu_2 &= \sum_i p_i x_i^2 \end{aligned}$$

Central moments:

$$\begin{aligned} \langle m_n(x) \rangle &= \langle (x - \mu_1)^n \rangle \\ &= \sum_i p_i (x_i - \langle x \rangle)^n \end{aligned}$$

The special case of $n = 2$ is called the *variance*:

$$Var\{x\} = \langle m_2(x) \rangle = \langle x^2 \rangle - \langle x \rangle^2 \quad (18)$$

The variance is particularly important because $Var\{x\}$ and μ_1 are sufficient to uniquely specify the important Gaussian distribution which usually results from the superposition of a sufficiently large number of random events from arbitrary underlying distributions.

2.5 Variance of a Random Function

By extension of the definition of the variance of a random variable we can define the variance of a random function $g(x)$:

$$\begin{aligned} Var\{g(x)\} &= \langle (g(x) - \langle g(x) \rangle)^2 \rangle \\ &= \sum_i p_i g^2(x) - 2\langle g(x) \rangle \sum_i p_i g(x_i) \\ &\quad + \langle g(x) \rangle^2 \sum_i p_i \\ &= \langle g^2(x) \rangle - \langle g(x) \rangle^2 \end{aligned}$$

2.6 Variance of a Linear Combination of Random Functions

By insertion of the definition and linearity of expectation values we can find the result for a linear combination of random functions:

$$\begin{aligned}
Var\{\lambda_1 g_1(x) + \lambda_2 g_2(x)\} &= \\
&= \langle (\lambda_1 g_1(x) + \lambda_2 g_2(x) - \langle \lambda_1 g_1(x) + \lambda_2 g_2(x) \rangle)^2 \rangle \\
&= \langle (\lambda_1 g_1(x) + \lambda_2 g_2(x) - \lambda_1 \langle g_1(x) \rangle - \lambda_2 \langle g_2(x) \rangle)^2 \rangle \\
&= \langle (\lambda_1 [g_1(x) - \langle g_1(x) \rangle] + \lambda_2 [g_2(x) - \langle g_2(x) \rangle])^2 \rangle \\
&= \langle (\lambda_1^2 [g_1(x) - \langle g_1(x) \rangle]^2 + \lambda_2^2 [g_2(x) - \langle g_2(x) \rangle]^2 \\
&\quad + 2\lambda_1\lambda_2 [g_1(x) - \langle g_1(x) \rangle][g_2(x) - \langle g_2(x) \rangle]) \rangle \\
&= \lambda_1^2 \langle [g_1(x) - \langle g_1(x) \rangle]^2 \rangle + \lambda_2^2 \langle [g_2(x) - \langle g_2(x) \rangle]^2 \rangle \\
&\quad + 2\lambda_1\lambda_2 \langle [g_1(x)g_2(x) - g_1(x)\langle g_2(x) \rangle \\
&\quad - \langle g_1(x) \rangle g_2(x) + \langle g_1(x) \rangle \langle g_2(x) \rangle] \rangle
\end{aligned} \tag{19}$$

In short this result can be expressed through the variances of the random functions g_1 and g_2 and an extra term:

$$\begin{aligned}
Var\{\lambda_1 g_1(x) + \lambda_2 g_2(x)\} &= \lambda_1^2 Var\{g_1(x)\} \\
&\quad + \lambda_2^2 Var\{g_2(x)\} \\
&\quad + 2\lambda_1\lambda_2 (\langle g_1(x)g_2(x) \rangle \\
&\quad - \langle g_1(x) \rangle \langle g_2(x) \rangle)
\end{aligned}$$

2.7 The Covariance

The mixed term in the preceding equation is called the *covariance* of $g_1(x)$ and $g_2(x)$.

$$Cov\{g_1(x), g_2(x)\} = \langle g_1(x)g_2(x) \rangle - \langle g_1(x) \rangle \langle g_2(x) \rangle \tag{20}$$

This term can be positive or negative and we will show in the next section that this term is related to the mutual dependence between the two random functions g_1 and g_2 .

We have the following special cases for simple random variables x and y :

$$\begin{aligned}
Cov\{x, y\} &= \langle xy \rangle - \langle x \rangle \langle y \rangle \\
Cov\{x, x\} &= \langle xx \rangle - \langle x \rangle \langle x \rangle = Var\{x\}
\end{aligned}$$

Depending on the sign of the *covariance*, the variance of a linear combination of random functions or variables can be larger or smaller than the sum of the individual variances.

$$Var\{g_1 + g_2\} = Var\{g_1\} + Var\{g_2\} + Cov\{g_1, g_2\} \tag{21}$$

The possibility of negative covariance can be exploited in special sampling techniques (correlated sampling, antithetic variates) to achieve *variance reduction*.

$$Var\{g_1 + g_2\} < Var\{g_1\} + Var\{g_2\} \tag{22}$$

2.8 Properties of the Covariance

$$\begin{aligned} \text{Cov}\{x, y\} &= \langle xy \rangle - \langle x \rangle \langle y \rangle \\ \langle xy \rangle &= \sum_{ij} p_{ij} x_i y_j \end{aligned} \quad (23)$$

If the random variables x and y are *independent*, the p_{ij} can be decomposed according to

$$p_{ij} = p_{1i} p_{2j} \quad (24)$$

$$\begin{aligned} \langle xy \rangle &= \sum_{ij} p_{1i} x_i p_{2j} y_j \\ &= \left(\sum_i p_{1i} x_i \right) \left(\sum_j p_{2j} y_j \right) \\ &= \langle x \rangle \langle y \rangle \\ \Rightarrow \text{Cov}\{x, y\} &= 0 \end{aligned} \quad (25)$$

Independence of two random variables x, y is clearly a sufficient but not a necessary condition for $\text{Cov}\{x, y\}$ to be zero!

2.9 Correlation and Autocorrelation

The correlation coefficient $r(x, y)$ is the normalized version of the covariance:

$$r(x, y) = \frac{\text{Cov}\{x, y\}}{\sqrt{\text{Var}\{x\} \text{Var}\{y\}}} \quad (26)$$

$$-1 \leq r(x, y) \leq 1 \quad (27)$$

If one considers the values of y as copies of x with a constant offset δ (in time or some pseudotime establishing an order)

$$y_j = x_i = x_{j-\delta} \quad (28)$$

one can compute a correlation coefficient for each offset δ .

$$r(x, y; \delta) = A(x; \delta) \quad (29)$$

This function $A(x; \delta)$ is called the autocorrelation function and varies between -1 and +1 as a result of the normalisation by the variances of x and y .

The computation of the autocorrelation function is an important tool to assess the statistical independence of events within a sequence of random events. If $A(x; \delta) \neq 0$ we can be sure that there is some serial correlation between events separated by an offset δ . Conversely $A(x; \delta) = 0$ can occur if either the covariance is systematically zero due to statistical independence of the events or accidentally zero in spite of serial correlation.

All random walk methods require careful autocorrelation analysis.

2.10 Continuous Distributions

In the preceding section we have assumed discrete random events to simplify the presentation, but generally random variables can also be continuous.

For a one-dimensional case we have

$$-\infty \leq x \leq \infty \quad (30)$$

We can define a *cumulative distribution function* $F(x)$ as

$$F(x) = P\{\text{randomly selected } y < x\} \quad (31)$$

Assume that $x_2 > x_1$. Then the events $x_2 > y \geq x_1$ and $x_1 > y$ are mutually exclusive and we conclude:

$$\begin{aligned} P\{x_2 > y \geq x_1\} + P\{x_1 > y\} &= P\{x_2 > y\} \\ P\{x_2 > y\} &\geq P\{x_1 > y\} \end{aligned}$$

It follows that $F(x)$ is a monotonically increasing function.

$$F(-\infty) = 0 \quad , \quad F(\infty) = 1 \quad (32)$$

The function $F(x)$ is not necessarily smooth. In differentiable regions one can define the *probability density function* $\rho(x)$:

$$\rho(x) = \frac{dF(x)}{dx} \geq 0 \quad (33)$$

2.11 Moments of Continuous Distributions

The concept of moments can be generalised to continuous distributions by the replacement of summations by integrations and of probabilities p_i by $dF(x)$.

$$E(x) = \langle x \rangle = \int_{-\infty}^{\infty} x dF(x) \quad \left(= \int_{-\infty}^{\infty} x \rho(x) dx \right) \quad (34)$$

$$\int_{-\infty}^{\infty} \rho(x) dx = F(\infty) = 1 \quad (35)$$

$$E(g(x)) = \langle g(x) \rangle = \int_{-\infty}^{\infty} g(x) dF(x) \quad (36)$$

The variance is now given as

$$\begin{aligned} Var\{x\} &= E(x^2) - E(x)^2 \\ &= \int_{-\infty}^{\infty} x^2 \rho(x) dx - \left[\int_{-\infty}^{\infty} x \rho(x) dx \right]^2 \end{aligned}$$

It is important to note that the variance is not a well defined quantity for all probability densities $\rho(x)$. A well known example is the Cauchy-Lorentz-distribution with an arbitrary width parameter a

$$\rho(x) = \frac{1}{\pi} \frac{a}{x^2 + a^2} \quad (37)$$

for which $E(x) = 0$ and $E(x^2) = \infty$.

2.12 Sums of Random Variables

- Suppose we have random variables x_1, x_2, \dots, x_n which are distributed according to some probability density function $\rho(x)$. The variable x_i may represent a multidimensional point.
- We evaluate functions $g_i(x_i)$ for each x_i where the functions g_i may or may not be identical. The $g_i(x_i)$ are then random variables.
- We define a weighted sum G over these functions and its expectation value $E(G)$:

$$G = \sum_i^n \lambda_i g_i(x_i) \quad \lambda_i \in \mathbb{R} \quad (38)$$

$$E(G) = \langle G \rangle = \sum_i^n \lambda_i \langle g_i(x_i) \rangle \quad (39)$$

- A special choice is to use $\lambda_i = 1/n$ for all weights and to consider all g_i to be identical

$$E(G) = E\left(\frac{1}{n} \sum_i^n g(x_i)\right) = \frac{1}{n} \sum_i^n E(g) = E(g) \quad (40)$$

We find that the expectation value $E(G)$ for the sum G is identical with the expectation value $E(g)$ for the function. Consequently G can serve as an *estimator* for $E(g)$. This is in fact the basis of all Monte Carlo methods: Expectation values, which generally correspond to integrals, are approximated by a sum (here G) over values of the integrand (here $g(x)$) sampled at a finite set of points $x_i, i = 1 \dots n$. We now have to establish a measure of the rate of convergence of the estimator G to the true expectation value if we increase n .

2.13 Variance of the Sum of Random Variables

- Assume for simplicity that all x_i are independent. In this case the covariance is zero for all combinations of random variables and the variance of G can be expressed simply as the sum of the variances of its terms:

$$\text{Var}\{G\} = \sum_i^n \lambda_i^2 \text{Var}\{g_i(x)\} \quad (41)$$

- Again assume identical weights and functions $\lambda_i = 1/n, g_i(x) = g(x)$

$$\text{Var}\{G\} = \sum_i^n \frac{1}{n^2} \underbrace{\text{Var}\{g(x)\}}_{\text{some number}} = \frac{1}{n} \text{Var}\{g(x)\} \quad (42)$$

The variance of the estimator G decreases in proportion to $1/n$ with a generally unknown proportionality factor. The determination of this factor would involve the computation of an integral over g , which is of the same complexity as the direct computation of the

expectation value $E(g)$ by integration. If we could compute this integral we would not resort to sampling and summation in the first place!

Statistical convergence:

The deviation of G from $E(g)$ will not decrease monotonically towards zero with increasing n as a result of the random nature of each new term entering the sum. The concept of convergence has to be accordingly redefined: The deviation δ of the estimator from the

true value will exceed a specified limit Δ with a probability which diminishes as $n \rightarrow \infty$.

The non monotonic convergence of results is probably the single most irritating feature of Monte Carlo methods for newcomers, in particular for those who are used to other quantum mechanical methods using basis set expansions or grids.

3 Sources of Randomness

Monte Carlo methods require the creation of random events according to specified probability densities. There are three classes of sources:

- Natural sources of (true ?) randomness.
Historically interesting but inefficient, not reproducible, and of hardly quantifiable quality.
- Deterministic algorithms producing a sequence of numbers with properties which are indistinguishable from a true random sequence as measured by a battery of statistical tests.
- Random walks constructed from primitive random events for all complicated multi-dimensional distributions.

We will look in detail only at **random walk methods** and assume the availability of a good uniform **random number generator**.

3.1 Random Number Generators

We will not attempt an exhaustive discussion of this subject, which is an active subject of research in number theory and cryptography. There is a very rich literature on the subject, including modern developments in the context of parallel computers and the problem of generating independent subsequences on many processors.^{34,35} Many useful methods for the generation of random numbers with a variety of distributions are discussed in Refs. 3, 5, 8, 36. Artefacts and diagnostic tests for the quality of random numbers are discussed in Refs. 1, 37–39. A detailed discussion of random number generators is the subject of other special lectures in this series.

The key features of contemporary random number generators can be summarized as follows:

- No true random sequence due to the underlying deterministic algorithm, therefore more precisely called **pseudo random number generators, PRNG**.
- Construction on the basis of number theory and by extensive statistical testing.
- Usually generation of a sequence of numbers $0 < u \leq 1$ with uniform distribution, $F(u) = u$.
- Several common types: linear congruential generators (LCG), lagged Fibonacci generators (LFG), combined generators.
- Other distributions by transformation algorithms.
- High speed and quality, perfect reproducibility of the sequence for test purposes.
- Deterministic algorithm implies some subtle sequential correlation.
- **There is no publicly known PRNG without at least one reported case of failure.**
- In case of doubt, try a generator of different type and verify statistical consistency.
- New challenges posed by the arrival of parallel computers (Creation of uncorrelated parallel sequences).

3.2 Random Walks

The random walk method has been introduced into statistical physics in the hallmark paper of Metropolis et al.⁴⁰ and is enormously flexible for the creation of random events with any conceivable distribution.

The method allows to generate samples from specified probability density functions $\rho(x)$, in particular if the space x has a high dimensionality. In fact we need not even be able to compute the absolute value of $\rho(x)$ for a given x , which implies a normalisation which in itself involves a multidimensional integration. It is generally sufficient to compute ratios between values of $\rho(x)$ at points x_i and x_j . Random walks are a sequence of events x_1, x_2, x_3, \dots , constructed such that the probability $P\{x_{new}\}$ of finding x_{new} is some function of $f(x_{new}, x_{last})$ of the previous events. The function $f(x_{new}, x_{last})$ describes a strategy to propagate the walk and corresponds in fact to a *conditional probability* $p(x_{new}|x_{last})$. This is the key difference to direct sampling methods where $P\{x_{new}\}$ is independent of the previous event x_{last} . The process has a memory and implies serial correlation.

Random walks are a special example for a Markov process.

The general conditions for random walks which are supposed to generate samples with distribution $\rho(x)$ can be summarized as follows

1. Every point x where $\rho(x) \neq 0$ must be accessible.
2. It must be possible to revisit the same point x any number of times.
3. The walk must not *periodically* pass through the same points x again.

These conditions are equivalent to requiring *ergodicity* of the random walk.

We should bear in mind that even if the fundamental construction principle of a random walk may ensure ergodicity, there is no guarantee that a random walk of finite length has explored all relevant parts of configuration space for a given physical problem. Insufficient length of random walks is the probably most common source of incorrect Monte Carlo results.

3.3 The Stochastic Matrix

Consider for a moment a system with discrete 'states' (position, orientations, quantum numbers etc.) x_1, x_2, \dots, x_n .

$p(x_j|x_i)$ is the conditional probability to observe x_j provided that we had x_i just before and is the transition probability for a Markov process. The ensemble of all $p(x_j|x_i)$ for all combinations of x_i and x_j can be arranged in matrix form, where we use the short notation $p_{ij} = p(x_j|x_i)$ for the transition from x_i to x_j .

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & & & \vdots \\ \vdots & & & & \vdots \\ p_{n1} & \cdots & \cdots & \cdots & p_{nn} \end{pmatrix} \quad (43)$$

The matrix P is a stochastic matrix

All $p_{ij} \geq 0$ because they represent probabilities and $\sum_j p_{ij} = 1$ for all i because each transition from i must lead to one of the available 'states'.

3.4 Properties of the Stochastic Matrix

Consider a row vector

$$\underline{\rho^{(0)}} = \{\rho_1^{(0)}, \rho_2^{(0)}, \dots, \rho_n^{(0)}\} \quad (44)$$

which describes an initial state in which $\rho_i^{(0)}$ is the probability of initially finding the system in state i .

Each step in the Markov chain can be formulated as a multiplication of this row vector with the stochastic matrix P :

$$\begin{aligned} \underline{\rho^{(1)}} &= \underline{\rho^{(0)}} P \\ \underline{\rho^{(2)}} &= \underline{\rho^{(1)}} P \\ &\vdots && \vdots \\ \underline{\rho^{(k)}} &= \underline{\rho^{(0)}} P^k \end{aligned}$$

The asymptotic distribution $\underline{\rho}$ for $k \rightarrow \infty$ is

$$\underline{\rho} = \lim_{k \rightarrow \infty} \underline{\rho^{(0)}} P^k \quad (45)$$

Repeated multiplication of $\rho^{(0)}$ with P converges to a stationary situation if

$$\underline{\rho} = \underline{\rho} P_{\equiv} \quad (46)$$

which implies that $\underline{\rho}$ is an eigenvector of P with eigenvalue 1. Here we already see that $\underline{\rho}$ can be multiplied by any scalar without changing the result. The asymptotic distribution is independent of the initial 'state' $\rho^{(0)}$ and depends exclusively on the matrix P . All initial vectors (except for other eigenvectors of P) converge to the same asymptotic distribution.

Matrices generally have a spectrum of eigenvalues. It is relatively easy to show that repeated multiplication of a vector with a matrix will project out the vector belonging to the eigenvalue with the largest modulus. To complete the proof of the validity of the random walk construction we have to prove that 1 is the largest possible eigenvalue of a stochastic matrix.

3.5 Detailed Balance

In the typical Monte Carlo situation we know which probability density $\underline{\rho}$ we want to sample, while the preceding argument will generate a probability density according to a given stochastic matrix. The key question is how to construct a matrix P_{\equiv} which has an eigenvector corresponding to the desired probability density $\underline{\rho}$.

The eigenvector equation can be written out explicitly as

$$\sum_i^n \rho_i p_{ij} = \rho_j \quad (47)$$

If we now impose detailed balance according to

$$\rho_i p_{ij} = \rho_j p_{ji} \quad (48)$$

we obtain the result

$$\sum_i \rho_i p_{ij} = \sum_i \rho_j p_{ji} = \rho_j \sum_i p_{ji} = \rho_j \quad (49)$$

This is exactly the condition required for an eigenvector of P with eigenvalue 1.

Detailed balance guarantees $\underline{\rho} P_{\equiv} = \underline{\rho}$ and is therefore a sufficient condition to construct a matrix P_{\equiv} with the desired asymptotic distribution but it is not necessarily the only possible way!

3.6 Decomposition of the Transition Process

We can arbitrarily decompose each p_{ij} into a factor describing the *probability of proposing* a particular transition t_{ij} and a factor a_{ij} describing the *probability of accepting* this choice.

$$p_{ij} = t_{ij} a_{ij} \quad (50)$$

This decomposition is valid if the two processes are successive and independent.

Substitution into the detailed balance relationship yields

$$\frac{\rho_j}{\rho_i} = \frac{p_{ij}}{p_{ji}} = \frac{t_{ij}a_{ij}}{t_{ji}a_{ji}} \quad (51)$$

Since we assume that ρ is a known probability density to be generated by the walk, and since we can pick a transition strategy specifying t_{ij} according to our taste, it is useful to convert this expression into a form which defines the required acceptance probability:

$$\frac{a_{ij}}{a_{ji}} = \frac{\rho_j}{\rho_i} \frac{t_{ji}}{t_{ij}} \quad (52)$$

Note that the construction of the random walk requires only that we are able to compute the *ratio* of probability densities. We can consequently work with densities ρ which are not normalized!

3.7 Accepting Proposed Transitions

There are two conventional choices for the acceptance probabilities a_{ij}, a_{ji} which satisfy this relation:

1. Metropolis (1953):

$$a_{ij} = \min \left[1, \frac{\rho_j}{\rho_i} \frac{t_{ji}}{t_{ij}} \right] \quad (53)$$

Proof by verification of the two possibilities:

$$\begin{aligned} (a) \quad \rho_j t_{ji} &\geq \rho_i t_{ij} \Rightarrow a_{ij} = 1 \\ a_{ji} &= \frac{\rho_i}{\rho_j} \frac{t_{ij}}{t_{ji}} \\ \frac{a_{ij}}{a_{ji}} &= \frac{\rho_j}{\rho_i} \frac{t_{ji}}{t_{ij}} \quad q.e.d. \end{aligned}$$

$$\begin{aligned} (b) \quad \rho_j t_{ji} &< \rho_i t_{ij} \Rightarrow a_{ij} = \frac{\rho_j}{\rho_i} \frac{t_{ji}}{t_{ij}} \\ a_{ji} &= 1 \\ \frac{a_{ij}}{a_{ji}} &= \frac{\rho_i}{\rho_j} \frac{t_{ij}}{t_{ji}} \quad q.e.d. \end{aligned}$$

2. Glauber

$$a_{ij} = \frac{1}{1 + \frac{\rho_i}{\rho_j} \frac{t_{ji}}{t_{ij}}} \quad (54)$$

Verification by substitution:

$$\frac{a_{ij}}{a_{ji}} = \frac{1 + \frac{\rho_i}{\rho_j} \frac{t_{ji}}{t_{ij}}}{1 + \frac{\rho_i}{\rho_j} \frac{t_{ij}}{t_{ji}}} \quad (55)$$

$$\frac{a_{ij}}{a_{ji}} = \frac{(\rho_i t_{ij} + \rho_j t_{ji}) \rho_j t_{ji}}{(\rho_j t_{ji} + \rho_i t_{ij}) \rho_i t_{ij}} \quad (56)$$

$$\frac{a_{ij}}{a_{ji}} = \frac{\rho_j t_{ji}}{\rho_i t_{ij}} \quad q.e.d. \quad (57)$$

3.8 Random Walks in Continuous Spaces

The fundamental ideas of the construction of random walks via the stochastic matrix and a detailed balance ansatz can be generalized to an infinite number of states.

$$a(x \rightarrow x') = \min \left[1, \frac{\rho(x') t(x' \rightarrow x)}{\rho(x) t(x \rightarrow x')} \right] \quad (58)$$

3.9 Coordinates of a Random Walk

The coordinates or 'states' of random walks are defined very broadly and can be

- All discrete (e.g. spins on a lattice)
- All continuous (e.g. atomic coordinates in simulations of liquids)
- Any mixture of the two (e.g. particles with spin and continuous space coordinates)

Specifically also the particle number in physical simulations can be a (discrete) coordinate (Grand canonical Monte Carlo).

3.10 The Transition Function t

A common simplification consists in the assumption

$$t(x' \rightarrow x) = t(x \rightarrow x') \quad (59)$$

and specifically the choice $t(x \rightarrow x') = t(|x - x'|)$.

The formula for the Metropolis acceptance probability is then simply

$$a(x \rightarrow x') = \min \left[1, \frac{\rho(x')}{\rho(x)} \right] \quad (60)$$

Examples:

- Proposing a new position x' with uniform probability from a volume surrounding x (e.g. a hypercube of predefined size in the majority of simple random walk methods).
- Picking a new position x' according to a multidimensional Gaussian centered on x :

$$t(x \rightarrow x') \propto \exp \left(-\frac{|x - x'|^2}{2\sigma^2} \right) \quad (61)$$

Explicit inclusion of $t(x \rightarrow x')$ in the formulation allows *guided random walks*. Guided random walks play a major role many smart sampling techniques like Force bias Monte Carlo, J-walking, improved variational quantum Monte Carlo, and diffusion Monte Carlo with importance sampling. The construction of the transition strategy is very often closely tied to the nature of the problem and no general guiding rules can be given here.

A good strategy tends to propose moves which strike a good compromise between large displacement in the underlying coordinate space to minimize the serial correlations and a good acceptance ratio. In diffusion quantum Monte Carlo the transition strategy is given once a trial wave function has been chosen, but in many other cases there is considerable space for human imagination, making this one of the key parts of Monte Carlo algorithms where brain power can beat supercomputer power.

3.11 How to Accept a Proposed Change of State?

The translation of the algebraic relationship for the probability a of accepting a move into a practical algorithm is in fact very simple. The cumulative distribution function $F(x)$ for the probability $P\{u \leq x\}$ that a random number u distributed in the interval $[0, 1]$ with a uniform probability density $\rho(u) = 1$ is less than or equal x ($x \leq 1$) is given by $P\{u \leq x\} = \int_0^x \rho(u)du = \int_0^x du = x$. Consequently $u \leq a$ will be true with probability a and we can accept the proposed move whenever a uniform random number u satisfies $u \leq a$. It is worthwhile to note here that the quality of the uniform random number generator for u has to be very high.

3.12 Summary of Important Random Walk Features

Good features:

- Random sampling from distributions in spaces of high dimension.
- No need to be able to normalize the probability density function (which would involve a multidimensional integration).
- Very general coordinate definition and very broad applicability.

Troublesome Features:

- The desired distribution is reached only asymptotically.
When is a random walk in its asymptotic regime?
- Serial correlation between sampling positions.
Requires careful autocorrelation analysis.

4 Monte Carlo Integration

4.1 The Principle of Monte Carlo Integration

We have seen that the expectation value of the sum of random variables $g(x_i)$ with x_i drawn according to the probability density $\rho(x)$ is identical with the expectation value of $g(x)$ over the underlying distribution:

$$E(G) = E(g(x)) \quad (62)$$

$$G = \frac{1}{n} \sum_i^n g(x_i) \quad , \quad x_i \propto \rho(x) \quad (63)$$

Now recall the original definition of the expectation value as an integral over the distribution:

$$E(g(x)) = \int \rho(x)g(x)dx = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i^n g(x_i) \quad (64)$$

The following conclusions can be drawn from this formula:

- The statistical error for the integral is independent of the dimensionality of the integral.
- The statistical error diminishes proportional to $1/\sqrt{n}$.
- The proportionality constant controlling the absolute size of the error bar depends on the variance of g .
- The integrand need not even be smooth.

The most important observation is the first one, which indicates that Monte Carlo integration can be interesting for multidimensional integrals. While all true random sampling techniques will lead to an error bar on the result which diminishes in proportion to $n^{-\frac{1}{2}}$ for n samples, there are in fact special methods based on socalled quasirandom numbers which achieve a somewhat faster rate of statistical convergence for problems with a moderate number of dimensions.

What we can influence by smart sampling techniques is the prefactor of $n^{-\frac{1}{2}}$ since it depends on the variance of the function being sampled. If parts of the function can be integrated analytically or factored out as a density accessible for direct sampling we may end up with a residual function to be sampled by Monte Carlo which has a smaller variance than the original one. These *variance reduction* techniques allow significant improvements of the statistical errors but do not affect the $n^{-\frac{1}{2}}$ convergence rate.

The last point is also of certain interest in particular in comparison to highly accurate grid based quadrature rules. The latter (e.g. Gaussian quadratures) are very powerful, but require the existence of high order derivatives of the integrand. Functions with discontinuities can be very troublesome to integrate unless one can subdivide the integration domain according to the location of the discontinuities. Monte Carlo is very robust and works for 'spiky' integrands. So even for low dimensional cases where Monte Carlo is not the most efficient method it may be an interesting way to produce a crude estimate due to its simplicity.

We should point out, however, that Monte Carlo integration of oscillatory function is notoriously difficult, as the computation of any integral which is small compared to the variance of the integrand, unless we can devise a trick to absorb the oscillations.

4.2 Efficiency of Monte Carlo Integration

In order to get an idea about the efficiency of Monte Carlo integration and to derive a criterion for the number of dimensions where Monte Carlo becomes interesting we make the following assumptions

- $g(x)$ is a d -dimensional function.
- We have fixed resources which allow the evaluation of $g(x)$ at a fixed number of points n .
- We dispose of a grid based product integration rule with leading error h^q in the grid step h (e.g. Simpson $q = 5$).
- In a balanced treatment of all dimensions the grid step size will then be $h \propto n^{-1/d}$.

Efficiency analysis:

- The product rule quadrature will have overall accuracy

$$\varepsilon_{Grid} \propto n^{-q/d} \quad (65)$$

- The Monte Carlo quadrature will have overall accuracy

$$\varepsilon_{MC} \propto n^{-1/2} \quad (66)$$

- Monte Carlo will be more efficient if $d > 2q$.

References

1. A. M. Ferrenberg, D. P. Landau, and Y. J. Wong. Monte Carlo simulations: Hidden errors from “good” random number generators. *Phys. Rev. Lett.*, 69:3382, 1992.
2. I. Vattulainen and T. Ala-Nissila. Mission impossible: Find a random pseudorandom number generator. *Comput. in Phys.*, 9:500, 1995.
3. R. V. Rubinstein. *Simulation and the Monte Carlo Method*. J. Wiley & Sons, New York, 1981.
4. M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods, Vol. 1*. J. Wiley & Sons, New York, 1986.
5. P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation, 2nd ed.* Springer, Berlin, New York, 1987.
6. C. W. Gardiner. *Handbook of Stochastic Methods, 2nd ed.* Springer, Berlin, New York, 1990.
7. I. M. Sobol. *Die Monte Carlo Methode, 4. Aufl.* Deutscher Verlag der Wissenschaften, Berlin, 1991.
8. G. S. Fishman. *Monte Carlo, Concepts, Algorithms, and Applications*. Springer, New York, 1996.
9. K. Binder, editor. *Monte Carlo Methods in Statistical Physics, 2nd ed. (Topics in Current Physics Vol. 7)*. Springer, Berlin, New York, 1986.
10. K. Binder, editor. *Application of the Monte Carlo Method in Statistical Physics, 2nd ed. (Topics in Current Physics Vol. 36)*. Springer, Berlin, New York, 1987.
11. D. W. Heermann. *Computer Simulation Methods in Theoretical Physics, 2nd ed.* Springer, Berlin, 1990.
12. K. Binder and D. W. Heermann. *Monte Carlo Simulation in Statistical Physics, 2nd ed. (Springer Series in Solid-State Sciences Vol. 80)*. Springer, Berlin, New York, 1992.

13. K. Binder, editor. *The Monte Carlo Method in Condensed Matter Physics, (Topics in Applied Physics Vol. 71)*. Springer, Berlin, New York, 1992.
14. M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, Oxford, 1999.
15. M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, Oxford, 1987.
16. D. Frenkel and B. Smit. *Understanding Molecular Simulation; From Algorithms to Applications*. Academic Press, New York, 1996.
17. R. H. Swendsen and J.-S. Wang. *Phys. Rev. Lett.*, 58:81, 1987.
18. U. Wolff. *Phys. Rev. Lett.*, 62:361, 1989.
19. M. H. Kalos. Monte Carlo calculations of the ground state of three- and four-body nuclei. *Phys. Rev.*, 128:1891, 1962.
20. M. H. Kalos. Energy of a boson fluid with Lennard-Jones potentials. *Phys. Rev. A*, 2:250, 1970.
21. M. A. Lee and K. E. Schmidt. Green's function Monte Carlo. *Comput. in Phys.*, 6:192, 1992.
22. J. B. Anderson. A random walk simulation of the Schrödinger equation: H_3^+ . *J. Chem. Phys.*, 63:1499, 1975.
23. P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr. Fixed-node quantum Monte Carlo for molecules. *J. Chem. Phys.*, 77:5593, 1982.
24. B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds. *Monte Carlo Methods in ab initio Quantum Chemistry*. World Scientific, Singapore, 1994.
25. R. P. Feynman. Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.*, 20:367, 1948.
26. R. P. Feynman and A. R. Hibbs. *Quantum Mechanics and Path Integrals*. McGraw Hill, New York, 1965.
27. D. Chandler and P. G. Wolynes. Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *J. Chem. Phys.*, 74:4078, 1981.
28. D. M. Ceperley. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.*, 67:279, 1995.
29. W. A. Lester, Jr. and B. L. Hammond. *Annu. Rev. Phys. Chem.*, 41:283, 1990.
30. J. B. Anderson. Exact quantum chemistry by Monte Carlo methods. In S. R. Langhoff, editor, *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*, page 1. Kluwer, Dordrecht, 1994.
31. J. B. Anderson. Fixed-node quantum Monte Carlo. *Int. Rev. Phys. Chem.*, 14:85, 1995.
32. D. M. Ceperley and L. Mitas. Quantum Monte Carlo methods in chemistry. *Adv. Chem. Phys.*, 93:1, 1996.
33. J. B. Anderson. Quantum Monte Carlo: Atoms, molecules, clusters, liquids, and solids. *Rev. Comp. Chem.*, 13:133, 1999.
34. M. Mascagni, S. A. Cuccaro, D. V. Pryor, and M. L. Robinson. A fast, high quality, and reproducible parallel lagged-Fibonacci pseudorandom number generator. *J. Comp. Phys.*, 119:211, 1995.
35. M. Mascagni. Parallel linear congruential generators with prime moduli. *Parallel Comput.*, 24:923–936, 1998.

36. R. M. Ziff. Four-tap shift-register-sequence random-number generators. *Comput. in Phys.*, 12:385, 1998.
37. P. Grassberger. On correlations in 'good' random number generators. *Phys. Lett. A*, 181:43, 1993.
38. I. Vattulainen, T. Ala-Nissila, and K. Kankaala. Physical tests for random numbers in simulations. *Phys. Rev. Lett.*, 73:2513, 1994.
39. I. Vattulainen, T. Ala-Nissila, and K. Kankaala. Physical models as tests of randomness. *Phys. Rev. E*, 52:3205, 1995.
40. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087, 1953.

Diffusion and Green's Function Quantum Monte Carlo Methods

James B. Anderson

Department of Chemistry, The Pennsylvania State University
University Park, Pennsylvania 16802, USA
E-mail: jba@psu.edu

Quantum Monte Carlo methods have proved remarkably successful in providing accurate predictions of energies and structures for molecular systems. These methods are 'exact' for systems of a few electrons and highly accurate for systems of as many as a thousand electrons. The scaling of computation effort with molecular size is highly favorable relative to that of other methods. The most commonly used quantum Monte Carlo methods – diffusion and Green's function – are introduced in these notes.

1 Introduction

For systems containing a few electrons – such as the molecular ion H_3^+ , the dimer He-He, the trimer He_3 , the pair He-H, and the molecule H_2 – a quantum Monte Carlo method provides absolute accuracies of better than 0.01 kcal/mole without systematic error. When an 'exact' potential energy surface for the reaction $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$ is needed a quantum Monte Carlo method is the choice ... providing 60,000 points on the surface with accuracies in the range of 0.01 to 0.10 kcal/mole.¹

For systems containing hundreds of electrons – such as the electron gas, metallic lithium, clusters of carbon atoms, crystals of N_2 , large molecules Si_mH_n , and solid silicon – quantum Monte Carlo methods provide the most accurate solutions available. When the stable, lowest-energy structure of C_{20} is desired a quantum Monte Carlo method gives the most reliable result.

Of course, quantum Monte Carlo methods are not so easily packaged as many other methods and they have far fewer practitioners. The program QMagiC is not as user-friendly as Gaussian98. But, there are many problems that demand solutions of very high accuracy – if only to provide benchmarks for calibrating other methods – and these are problems which demand QMC methods. The scaling of QMC methods with the number of electrons is generally favorable compared to that of other methods (see Table 1), and the scaling of QMC methods with increasing accuracy is especially favorable at high accuracies compared to that of other methods.

In these notes we describe the several quantum Monte Carlo methods and discuss their characteristics, their advantages and disadvantages. We present a representative sampling of results of QMC calculations to illustrate the range of systems which have been treated successfully. Our object is to provide an introduction together with an overview of the field.

We call attention to several prior reviews in the QMC area which give different insights and additional details. These include one book³ of general coverage, review articles of a general nature,^{4–11} a review of 'exact' methods,¹² a discussion of fixed-node calculations,¹³ and a review of applications to solids.¹⁴

Theoretical Method	Computational Dependence on Number of Electrons	Maximum Feasible Molecular Size (atoms)
FCI	$N!$	2
CCSD(T)	N^7	8 – 12
CCSD	N^6	10 – 15
MP2	N^5	35 – 50
HF	$N^{3.5} – N^4$	50 – 200
KS-DFT	$N^{3.5} – N^4$	50 – 200
FNQMC	N^3	50 – 200

Table 1. Scaling of computation requirements with number of electrons. Based in part on a table by Head-Gordon.²

2 History and Overview

Among the various ways in which Monte Carlo methods can be utilized in solving the Schrödinger equation there are four methods commonly termed 'quantum Monte Carlo' methods (QMC). These are the variational quantum Monte Carlo method (VQMC), the diffusion quantum Monte Carlo method (DQMC), the Green's function quantum Monte Carlo method (GFQMC), and the path integral quantum Monte Carlo method (PIQMC). These methods are by their nature strongly related and each has its own peculiar advantages and disadvantages relative to the others.

The variational method VQMC is the same as the conventional analytic variational method except that the required integrals are evaluated using special Monte Carlo methods. It has its roots in a numerical method reported by Frost¹⁵ in 1942. In Frost's own words: "A method of approximation to the Schrödinger equation has been developed in which variation functions are used but no integrations are involved. The procedure involves evaluation of the energy for a set of representative points in configuration space. The parameters in the variation function are then chosen by applying the condition that the mean square deviation of the energy from the average should be a minimum." As a part of this calculation Frost estimated the expectation value of the energy $\langle E \rangle$ from the local energies $E_{loc} = H\Psi_0/\Psi_0$ for a trial wavefunction Ψ_0 using Ψ_0^2 as a weighting factor according to

$$\langle E \rangle = \frac{\int \Psi_0^2 \frac{H\Psi_0}{\Psi_0} d\tau}{\int \Psi_0^2 d\tau} \cong \frac{\sum \Psi_0^2 \frac{H\Psi_0}{\Psi_0}}{\sum \Psi_0^2}, \quad (1)$$

where the summations are for points in the configuration space of the electrons, chosen in a manner "to be determined through experience". Frost was successful in investigating preliminary applications to a few simple molecules.

The Monte Carlo aspect of choosing points was introduced by Conroy¹⁶ in 1964. Conroy proposed picking points at random in the configuration space of the electrons with probabilities proportional to Ψ_0^2 and equal weights. Conroy noted "If ... the density function [is] Ψ_0^2 , ... then clearly optimum Monte Carlo sampling has the density of random points proportional to the density of electrons in the actual molecule." The procedure leads to a good approximation of the ratio of the integrals in Eq. (1) for a large number of points

and to the exact value in the limit of a large number of points. Conroy was able to obtain some excellent values for the energies of H_2^+ , H^- , HeH^{++} , He , H_2 , and Li . His calculation for Li was the first application of VQMC to a fermion system with nodes.

Conroy's VQMC calculations were followed very soon by those of McMillan¹⁷ for liquid helium using the Metropolis algorithm to sample the configuration space for points with probabilities proportional to Ψ_0^2 . Only very recently has a much more efficient method for choosing points - anticipated to some extent by Conroy - been devised.¹⁸

The DQMC method is based on the similarity of the Schrödinger equation and the diffusion equation. It has its roots in the Monte Carlo simulation of neutron diffusion and capture by Fermi and others at Los Alamos in the 1940's. Metropolis and Ulam¹⁹ first outlined the method in 1947: "... as suggested by Fermi, the time-independent Schrödinger equation

$$-\frac{1}{2}\nabla^2\Psi(x, y, z) = E\Psi(x, y, z) - V\Psi(x, y, z) \quad (2)$$

could be studied as follows. Re-introduce time by considering

$$\Psi(x, y, z, t) = \Psi(x, y, z) e^{-Et} . \quad (3)$$

and $\Psi(x, y, z, t)$ will obey the equation

$$\frac{\partial\Psi(x, y, z, t)}{\partial t} = \frac{1}{2}\nabla^2\Psi(x, y, z, t) - V\Psi(x, y, z, t) . \quad (4)$$

This last equation can be interpreted however as describing the behavior of a system of particles each of which performs a random walk, i.e., diffuses isotropically and at the same time is subject to multiplication, which is determined by the value of the point function V . If the solution of the latter equation corresponds to a spatial mode multiplying exponentially in time, the examination of the spatial part will give the desired $\Psi(x, y, z)$ – corresponding to the lowest 'eigenvalue' E ." The first applications of DQMC to electronic systems were reported by Anderson²⁰ in 1975 and were followed by a large number of additional developments, along with applications to a wide variety of chemical problems.

The GFQMC method was proposed by Kalos²¹ as an alternative to the DQMC method. As Kalos noted, "It seemed more natural and promising to look for an integral equation formulation of the Schrödinger equation and attempt its solution by Monte Carlo methods." The first applications of GFQMC were in determining the binding energies of three- and four-body nuclei.²¹ For problems having appropriate boundary conditions and potential energy functions the GFQMC method is preferred, but it is not well suited for most electronic systems. However, it provides the basis for 'exact' (i.e., without systematic error) solutions for systems of a few electrons.

The PIQMC method is the result of coupling of Feynmann's path integral formulation of quantum mechanics²² with Monte Carlo sampling techniques to produce a method for finite-temperature quantum systems. In the limit of zero temperature the method is closely related to the GFQMC method. The earliest applications of PIQMC were made to lattice models, but a number of applications to continuum systems of bosons have been made, including some very successful calculations of properties of liquid helium.²³ Applications to fermion systems are more difficult, but a few studies have been carried out.²⁴

3 Variational Quantum Monte Carlo

In the variational quantum Monte Carlo (VQMC) method the expectation value of the energy $\langle E \rangle$ and/or another average property of a system is determined by Monte Carlo integrations. The expectation value of the energy is typically determined for a trial function Ψ_0 using Metropolis sampling²⁵ based on Ψ_0^2 . It is given by

$$\langle E \rangle = \frac{\int \Psi_0^2 \frac{H\Psi_0}{\Psi_0} d\tau}{\int \Psi_0^2 d\tau} = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \frac{H\Psi_0}{\Psi_0}}{\sum_{i=1}^n 1}, \quad (5)$$

where the summations are for samples of equal weights selected with probabilities proportional to Ψ_0^2 . As in analytic variational calculations the expectation value $\langle E \rangle$ is an upper limit to the true value of the energy E ,

$$\langle E \rangle \geq E . \quad (6)$$

The term $\frac{H\Psi_0}{\Psi_0}$ is a local energy E_{loc} . In determining $\langle E \rangle$ it is not necessary to carry out analytic integrations; and, since only differentiation of the trial wavefunction is required to evaluate the local energy, the trial wavefunction may take any desired functional form. It may even include inter-electron distances r_{ij} explicitly. Thus, relatively simple trial functions may incorporate electron correlation effects rather accurately and produce expectation values of the energy well below those of the Hartree-Fock limit. Except in the limit of a large number of terms the VQMC method is not an exact method.

The Metropolis sampling procedure provides a means of sampling points in configuration space with specified probabilities, in this case, with probabilities proportional to the square of the wavefunction. Starting from an arbitrary initial point, one chooses a new point at a fixed distance (or from a distribution of fixed distances) in a random direction. One then calculates the ratio of weights new-to-old, $\Psi_0^2(new)/\Psi_0^2(old)$, and accepts the move to the new point with the probability given by the ratio. If the ratio is greater than unity the move is accepted. If the move is not accepted the old point is treated as a new point. The result of a large number of iterations is a guided random walk which samples points in configuration space with frequencies proportional to Ψ_0^2 . The reader might wish to consider a two-point system of a and b with weights W_a and W_b for which a near-equilibrium distribution is obtained in sampling with just a few steps.

The step sizes for a typical Metropolis walk are usually chosen to give an acceptance ratio of about one-half in order maximize the rate of 'diffusion' and improve the sampling speed. Serial correlation of points is usually high. In many-dimensional (or many-electron) systems the steps may be taken one dimension (or one electron) at a time or all at once. The optimum step sizes and/or combinations of steps depend strongly on the nature of the system treated.

The Metropolis procedure can be made more efficient by using a bias of each step in the direction of higher weight as indicated by the derivative of the weight at the old point. In the limit of small steps this leads to the Fokker-Planck equation, which is applicable to diffusion with drift and is directly related to the 'importance sampling' in diffusion quantum Monte Carlo discussed below. For many systems this type of sampling is more efficient than Metropolis sampling, but care must be taken to eliminate the time-step error²⁶ associated with simulation of the Fokker-Planck equation. The procedure is somewhat more

complicated, offers a greater opportunity for error, and is used less often than Metropolis sampling.

Another alternative, likely to be more efficient than Metropolis sampling, is the use of probability density functions.²⁷ These relatively simple functions which mimic the density of the more complex function Ψ_0^2 can be sampled directly without a Metropolis walk and the associated serial correlation. Sample points of unit weight are obtained with probabilities proportional to the probability density P and their weights are multiplied by the factor Ψ_0^2/P to give overall Ψ_0^2 weighting. The expectation value of the energy $\langle E \rangle$ is then given by

$$\langle E \rangle = \frac{\int \Psi_0^2 \frac{H\Psi_0}{\Psi_0} d\tau}{\int \Psi_0^2 d\tau} = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n (\Psi_0^2/P) \frac{H\Psi_0}{\Psi_0}}{\sum_{i=1}^n (\Psi_0^2/P)} , \quad (7)$$

where the summations are for samples of equal weights selected with probabilities proportional to P .

4 Diffusion Quantum Monte Carlo

The diffusion quantum Monte Carlo method (DQMC) approaches the solution of the Schrödinger equation in a way completely different from that of variational methods. The basic ideas were given above in the succinct description quoted from the original paper by Metropolis and Ulam.¹⁹ Here we give a more complete description.

The DQMC method is basically a simple game of chance involving the random walks of particles through space and their occasional multiplication or disappearance. It may be viewed as based on the similarity between the Schrödinger equation and the diffusion equation (i.e., Fick's second law of diffusion) and the use of the random walk process to simulate the diffusion process. Following the early discussions in the 1940's by Metropolis and Ulam¹⁹ and by King²⁸ a number of related techniques were proposed and discussed in succeeding years, but it was not until fast computers became available that applications to multicenter chemical systems became practical.²⁰

The equation to be solved is the time-independent Schrödinger equation, $H\Psi = E\Psi$, or

$$-\sum_i \frac{\hbar^2}{2m_i} \nabla_i^2 \Psi(\vec{X}) + V(\vec{X}) \Psi(\vec{X}) = E\Psi(\vec{X}) , \quad (8)$$

where the summation is over the electrons or other particles i having masses m_i and the nomenclature is standard. Since we are concerned with the time-independent Schrödinger equation the wavefunction may be treated as a real number rather than a complex number. For simplicity we consider the equation for a single particle of mass m , rearranged to become

$$\frac{\hbar^2}{2m} \nabla^2 \Psi(\vec{X}) - V(\vec{X}) \Psi(\vec{X}) = -E\Psi(\vec{X}) . \quad (9)$$

The equation has as solutions the wavefunctions $\Psi_0(\vec{X})$, $\Psi_1(\vec{X})$, ... which exist only for specific energies E_0, E_1, \dots .

The wavefunction may be treated as a function of an additional variable τ defined according to

$$\Psi(\vec{X}, \tau) = \Psi(\vec{X})e^{-E\tau} . \quad (10)$$

The function then behaves according to

$$\frac{\partial \Psi(\vec{X}, \tau)}{\partial \tau} = -E\Psi(\vec{X}, \tau) \quad (11)$$

and we have

$$\frac{\partial \Psi}{\partial \tau} = \frac{\hbar^2}{2m}\nabla^2\Psi - V\Psi . \quad (12)$$

The function $\Psi(\vec{X}, \tau)$ in Eq. (11) may be considered general, but at large values of τ its solution is given by the $\Psi(\vec{X}, \tau)$ of Eq. (10) corresponding to the lowest-energy or ground-state wavefunction for the system. Since higher-energy states decay faster according to Eq. (10) an arbitrary initial function consisting of a sum of terms containing the wavefunctions for the ground-state and any or all the higher states decays to the ground-state wavefunction. The arbitrary initial function evolves to the ground-state solution of the time-independent Schrödinger equation.

Because of its similarity to the time-dependent Schrödinger equation, Eq. (12) is often referred to as the Schrödinger equation in imaginary time. The analogy is formally correct since solutions of the time-dependent Schrödinger equation have equivalent real and imaginary parts under steady-state conditions.

The Schrödinger equation in imaginary time τ has the same form as the diffusion equation with an added first-order reaction term,

$$\frac{\partial C(\vec{X}, t)}{\partial t} = D\nabla^2C(\vec{X}, t) - kC(\vec{X}, t) . \quad (13)$$

The concentration C corresponds to the wavefunction Ψ , the diffusion coefficient D corresponds to the group $\frac{\hbar^2}{2m}$, and the rate constant k corresponds to the potential energy V .

Differential equations are normally used to model the behavior of physical systems and the diffusion equation above is normally used to model the behavior of a system in which particles undergo diffusion by a random walk process. In quantum Monte Carlo calculations the random walk process is used to simulate the differential equation. Of course, the connection between the random walk process and quantum mechanics may be considered to be direct. In the absence of the Schrödinger equation one might still use the Monte Carlo method to obtain solutions to quantum mechanical problems, but the connection between random walks and quantum mechanics is most easily made with the aid of the Schrödinger equation as above.

The random walk process and the diffusion equation are related through the diffusion coefficient by the Einstein equation,²⁹

$$D = \frac{\overline{(\Delta x)^2}}{2\Delta\tau} , \quad (14)$$

which gives the diffusion coefficient for particles moving a distance Δx at random positive or negative at intervals of time $\Delta\tau$. In the simulation of the Schrödinger equation in imaginary time the time and distance steps are chosen to produce the appropriate value of D (or $\frac{\hbar^2}{2m}$) given by Eq. (13).

The standard quantum mechanical problem of the harmonic oscillator may be used to illustrate the diffusion quantum Monte Carlo method. The potential energy is given by the function $V = \frac{1}{2}kx^2$. The potential energy may be shifted by an arbitrary constant energy to make V negative in the central region near $x = 0$ and positive away from the center.

An initial collection of particles, typically termed 'walkers' but occasionally termed 'psips' and perhaps a dozen other names, is distributed in the region about $x = 0$. Time is advanced one step $\Delta\tau$. To simulate the diffusion term of Eq. (13) each walker is moved right or left at random a distance Δx . To simulate the multiplication term of Eq. (13) each walker then gives birth to a new walker with a probability $P_b = -V\Delta\tau$ if V is negative or disappears with a probability $P_d = V\Delta\tau$ if V is positive. Time is advanced another step and the process is repeated. If the number of walkers falls below an acceptable lower limit or increases beyond an acceptable upper limit, their number may be adjusted by the random multiplication or removal of walkers present. (See comment below on avoiding bias with such adjustments.) For the harmonic oscillator as indicated the walkers diffuse away from the center and disappear at the sides in the regions of high potential energy, but they are replaced by walkers multiplying near the center at negative potential energies. After a large number of iterations the distribution of walkers approaches a fluctuating 'steady-state' distribution – the function $\exp(-ax^2)$ with $a = \frac{1}{2}\sqrt{k}$ – which corresponds to the wavefunction for the ground state of the harmonic oscillator.

The procedure is readily extended to problems having a higher number of dimensions and is clearly most useful for problems in which the number of dimensions is large. A system of n electrons free to move in three dimensions each can be simulated by a collection of walkers moving in $3n$ dimensions each.

For a molecule the procedure is similar. For the case of H_2 the Schrödinger equation in imaginary time for the two-electron system with both nuclei fixed is given, in atomic units, by

$$\frac{\partial\Psi}{\partial\tau} = \frac{1}{2}\nabla_1^2\Psi + \frac{1}{2}\nabla_2^2\Psi - V\Psi . \quad (15)$$

With the electrons labeled 1 and 2 and the two protons labeled A and B the potential energy V , exclusive of the internuclear term, is

$$V = -\frac{1}{r_{1A}} - \frac{1}{r_{1B}} - \frac{1}{r_{2A}} - \frac{1}{r_{2B}} + \frac{1}{r_{12}} , \quad (16)$$

in which r_{1A} is the distance between electron 1 and proton A and so forth. It is convenient to introduce a reference potential V_{ref} so that the operating equation becomes

$$\frac{\partial\Psi}{\partial\tau} = \frac{1}{2}\nabla_1^2\Psi + \frac{1}{2}\nabla_2^2\Psi - (V - V_{ref})\Psi . \quad (17)$$

In terms of the diffusion equation we then have $D = 1/2$ and $k = (V - V_{ref})$.

The random walk in six dimensions is usually executed with non-uniform step sizes in each dimension selected from a Gaussian distribution with probabilities P of step sizes

Δx given by

$$P(\Delta x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\Delta x)^2}{2\sigma^2}\right). \quad (18)$$

The probability of birth is given by $P_b = -(V - V_{ref})\Delta\tau$ for $(V - V_{ref})$ less than zero and the probability of disappearance is given by $P_d = (V - V_{ref})\Delta\tau$ for $(V - V_{ref})$ greater than zero. After each move a random number in the interval $(0,1)$ for each walker is compared with P_b (or P_d) and if smaller than P_b (or P_d) then a birth (or death) is completed.

A calculation is begun with a collection of 1000 or more walkers in positions corresponding to electron configurations in the region of the nuclei and allowed to approach the steady-state distribution. The step size is then fixed at a small value to improve the accuracy of the results in the accumulation of data after steady-state is reached.

In order to maintain the number of walkers approximately constant the arbitrary reference potential V_{ref} may be adjusted occasionally, but to avoid bias a large delay prior to adjustment is advised. At steady-state the energy E corresponding to a wavefunction Ψ may be evaluated using Eq. (11) rearranged as

$$E = -\frac{1}{\Psi} \frac{\partial\Psi}{\partial\tau}. \quad (19)$$

For a given distribution the wavefunction is proportional to the total number of walkers N and one has

$$E = -\frac{1}{N} \frac{\partial N}{\partial\tau}. \quad (20)$$

In the case of the ground state of H_3^+ , which has no boundaries serving as sinks or sources for walkers, the total number of walkers is not directly affected by the diffusion terms of Eq. (17) but changes according to

$$\frac{\partial N}{\partial\tau} = -\sum_N (V - V_{ref}). \quad (21)$$

The energy is thus given by the average potential energy \bar{V} according to

$$E = \bar{V}. \quad (22)$$

After steady-state is reached the energies at each time step are retained for a subsequent determination of the overall average for a large number of samples.

There are five important sources of error in these first diffusion Monte Carlo calculations: (a) Statistical or sampling error associated with the limited number of independent sample energies used in determining the energy from an average of variable potential energies, (b) the use of a finite time-step $\Delta\tau$ rather than an infinitesimal time-step as required for the exact simulation of a differential equation, (c) numerical error associated with truncation and/or round-off in computing, (d) imperfect random number quality, (e) failure of the distributions to reach the steady-state or equilibrium distributions in a finite number of steps. Sources (c), (d), and (e) are common problems in computing. They can be detected relatively easily and eliminated, and they are not found to limit the calculations in any significant way. Sources (a) and (b) seriously limit the accuracy of most DQMC calculations, but twenty years of refinement of methods to reduce time-step error as well as the higher speeds of computers have reduced greatly the magnitude of these errors and uncertainties.

For systems containing two or more electrons of the same spin or other indistinguishable particles, an additional problem appears: the node problem. For these systems it is necessary to restrict the form of the total wavefunction (space and spin parts) such that it is antisymmetric to the exchange of electrons. For any electronic state other than the ground state it is necessary to restrict further the properties of the wavefunction. The effect of these restrictions is the imposition of nodal surfaces, on which $\Psi(\vec{X}) = 0$, in the space part of the wavefunction. For systems of a few electrons the node problem can be overcome by exact cancellation methods (described below) and solutions free of systematic error can be obtained. For systems of more than a few electrons the fixed-node method, which in not an exact method, is usually required.

5 Green's Function Quantum Monte Carlo

For certain boundary conditions the diffusion equation may be solved with the use of standard Green's function methods, and the diffusion equation with an added first-order reaction term may be treated by these methods. The Green's function quantum Monte Carlo method is similar to the DQMC method but takes advantage of the properties of Green's functions in eliminating time-step entirely in treating the steady-state equation. The GFQMC method makes possible very large step sizes, but some of the advantages of large steps are lost for fixed-node calculations. The Green's function quantum Monte Carlo method was proposed by Kalos²¹ for nodeless systems. Procedures for introducing fixed nodes were developed later.

The time-independent Schrödinger equation, Eq. (9), may be written in the form

$$-\nabla^2 \Psi(\vec{X}) + k^2 \Psi(\vec{X}) = k^2 \frac{V(\vec{X})}{E} \Psi(\vec{X}), \quad (23)$$

where

$$k^2 = -\frac{2mE}{\hbar^2}. \quad (24)$$

To keep k^2 positive the energy must be made negative. This can be done by adjusting the reference or zero of the potential energy by an appropriate offset of energy.

The Green's function for Eq. (9) which satisfies the boundary conditions for a problem in electronic structure (i.e., $\Psi \rightarrow 0$ as $X \rightarrow \infty$) is known and is given by

$$G(\vec{X}, \vec{X}') = \frac{1}{(2\pi)^{\frac{3N}{2}}} K_{\frac{3N}{2}-1}(k|\vec{X} - \vec{X}'|)/(k|\vec{X} - \vec{X}'|^{\frac{3N}{2}-1}), \quad (25)$$

where K_ν is the modified Bessel function of the second kind.

The Green's function method is carried out iteratively with steps analogous to time steps. Repetitive sampling is based on the property of the Green's function which reproduces the wavefunction from itself,

$$\Psi(\vec{X}) = \int G_o(\vec{X}, \vec{X}') \frac{V(\vec{X}')}{E} \Psi(\vec{X}') d\vec{X}'. \quad (26)$$

The repeated application of Eq. (26) to an initially arbitrary wavefunction $\Psi(\vec{X}')$ produces a wavefunction $\Psi(\vec{X})$ which is the lowest-energy solution to the Schrödinger equation for

the boundary conditions specified. A walker in the distribution $\Psi(\vec{X}')$ may be transferred to the distribution $\Psi(\vec{X})$ by multiplying its weight by $\frac{V(\vec{X}')}{E}$, sampling the Green's function distribution $G_o(\vec{X}, \vec{X}')$, and moving the walker to its new position \vec{X} . Repetition for an initially arbitrary collection of walkers leads to a set of walkers which is a sample of points from the lowest-energy wavefunction for the boundary conditions and any other constraints imposed. As in DQMC the calculations must be carried out until a 'steady-state' distribution is obtained and sampling is carried out by continuing the calculations.

The imposition of additional boundaries corresponding to nodes for fixed-node calculations has been described by Ceperley,³⁰ by Skinner et al.,³¹ and by Moskowitz and Schmidt.³² The procedures involve conditional sampling together with smaller steps for walkers in the vicinity of the nodes.

6 Node Structure

The structure and properties of the nodal hypersurfaces of the wavefunctions for atomic and molecular systems have received very little attention. In analytic variational calculations the wavefunctions obtained are seldom examined and although electron densities are often examined, these reveal little or nothing about the node structure. Examination of the basis set of a determinantal wavefunction also reveals little or nothing because the many operations of the determinant scramble the properties of the basis functions. Only recently, with a knowledge of node structure required for developing Monte Carlo methods, have the structure and properties of nodal hypersurfaces been examined in detail.

For a system of either bosons or fermions the wavefunction must have the correct properties of symmetry and antisymmetry. Except in the simplest cases the wavefunction for a system of n fermions is positive and negative in different regions of the $3n$ -dimensional space of the fermions. The regions are separated by one or more $(3n - 1)$ -dimensional hypersurfaces which cannot be specified except by solution of the Schrödinger equation.

The procedures described above for DQMC and GFQMC lead to the lowest-energy solutions for boson systems which are nodeless ground-state wavefunctions. They also lead to the ground-state in the case of two electrons (fermions) of opposite spin for which the wavefunction is symmetric to the exchange of the two electrons. For a system of two or more electrons of the same spin the wavefunction must be antisymmetric to the exchange of electrons of the same spin and must contain one or nodal hypersurfaces. The treatment of systems with nodes requires that the solutions be constrained to the appropriate antisymmetry.

Several properties should be noted for a system of two electrons of the same spin. The configuration space of the electrons is divided in half by the nodal surface. The two halves are similar in shape and are nested together face-to-face. The positions of the two electrons are represented by a single point in configuration space and interchange of the two electrons moves the point across the nodal surface to a similar position in the other half of configuration space.

One of the simplest cases is that of the $1s2s\ ^3S$ helium atom for which the wavefunction may be regarded as a function of the electron-nucleus radii r_1, r_2 and the angle θ between them. The nodal surface is the 5-dimensional hypersurface on which the electron-nucleus distances r_1 and r_2 are equal to each other. It is completely specified by the symmetry

of the function. In this case the (r_1, r_2, θ) configuration space is divided by the nodal surface into two equivalent sections, one with the wavefunction positive and one with the wavefunction negative. This is the nodal structure given by the simplest single-determinant wavefunction $\Psi = 1s(1)2s(2) - 1s(2)2s(1)$. For $r_1 = r_2$ the wavefunction is zero and the nodal surface occurs for $r_1 = r_2$ regardless of the functions 1s and 2s provided they are functions of r_1 only and r_2 only.

In the case of $1s2p\ ^3P$ helium the situation is not so simple. For $1s2p\ ^3P$ He the symmetry properties alone are insufficient to specify the node structure. The node structure is not determined by geometric symmetry alone because there are many possible wavefunctions which have the required anti-symmetry on reflection in the $z = 0$ plane and on exchange of electrons. The simplest is given by $\Psi = 1s(1)2p(2) - 1s(2)2p(1)$. But, there is an infinite number of different 1s and 2p functions which may be used and the node structures of the resulting wavefunctions are different. Thus, the symmetry properties alone are insufficient to specify the node structure for this case.

A detailed examination³³ has been made of the node structure of $1s2p\ ^3P$ helium for very accurate wavefunctions. It should be noted that the nodal surface is not a simple plane passing through the origin in the three-dimensional space of one or the other of the electrons. The wavefunction is not the product function $\Psi = 2s(1)2p(2)$ and its node structure is not that of the product function. The node structure is similar to that of the determinantal function and very much different from that of the product function.

DQMC calculations for atoms and molecules such as H₂, H₄, Be, H₂O, and HF using fixed-node structures obtained from optimized single-determinant SCF calculations typically recover more than 90 percent of the correlation energies of these species and yield total electronic energies lower than the lowest-energy analytic variational calculations. These results suggest that optimized single-determinant wavefunctions have node structures which are reasonably correct.

An investigation of the node structure of the wavefunction in the 30-dimensional configuration space of a 10-electron molecule is not an easy task, but it has been carried out for the water molecule. The node structure for the ground state is that suggested by earlier investigations for smaller systems. For a molecule with m spin-up electrons and n spin-down electrons the node structure is approximately that of the product of two functions, one for spin-up and the other for spin-down electrons. The $3m$ -dimensional configuration space of spin-up electrons is divided by a $(3m - 1)$ -dimensional nodal hypersurface into two geometrically similar regions nested together face-to-face. The $(3n - 1)$ -dimensional nodal hypersurface for the spin-down electrons has equivalent characteristics.

7 Importance Sampling

One very important means of improving the accuracies is the technique called “importance sampling” which was introduced by Grimm and Storer³⁴ in 1971. As is clear from their work, one should be able to take advantage of prior knowledge of the properties of wavefunctions to make quantum calculations of many types more efficient. Such prior knowledge is available in the form of wavefunctions from analytic variational calculations at several levels of approximation. It is possible to obtain very high accuracies by extending diffusion quantum Monte Carlo calculations to calculate corrections to trial wavefunctions rather than the complete wavefunction. We discuss those correction methods in a separate

section.

To obtain the importance-sampling version of diffusion quantum Monte Carlo, we first multiply the basic equation, Eq. (4) by a trial wavefunction Ψ_t and define a new term $f = \Psi\Psi_t$ which is the product of the true wavefunction and the trial wavefunction. After several pages of rearrangement we obtain the basic equation for DQMC with importance sampling,

$$\frac{\partial f}{\partial \tau} = \frac{\hbar^2}{2m} \nabla^2 f - \nabla \cdot (f \nabla \ln \Psi_t) - E_{loc} f . \quad (27)$$

The equation has terms on the right side corresponding to diffusion of walkers with a diffusion coefficient of $\frac{\hbar^2}{2m}$, a drift term with a velocity given by $\nabla \ln \Psi_t$, and a first-order rate term for the disappearance of walkers with a rate constant given by the local energy $E_{loc} = \frac{H\Psi_t}{\Psi_t}$ for the trial wavefunction.

In DQMC the simulation of Eq. (27) is carried out in the same way as the simulation of Eq. (9) except that additional walker movement is required by the drift term and walker multiplication depends on the local energy rather than the potential energy. The diffusion and drift terms can be separately simulated. If the trial function is simply a constant the drift term is zero, the local energy is equal to the local potential energy, and the expression reduces to that for diffusion without importance sampling.

The drift term acts to produce a drift of walkers in the direction of higher Ψ_t . The walkers are thus concentrated in the more important regions and their distribution, if Ψ_t is accurate, approximates that of Ψ^2 , the square of the true wavefunction. In the vicinity of a nodal surface the velocity, which may be written as $\frac{\nabla \Psi_t}{\Psi_t}$, is increased and as Ψ_t approaches zero at the nodal surface, the drift velocity approaches infinity in a direction away from the surface. Walkers are thus prevented from crossing the nodes of the trial function.

The computation procedure for diffusion with drift is similar to that of the basic random walk procedure described above. At each time step the values of $E_{loc} = \frac{H\Psi_t}{\Psi_t}$ and the drift velocity $\nabla \ln \Psi_t$ must be determined from the potential energy and from the first and second derivatives of the trial wavefunction. The drift distance is given by the product of the vector drift velocity and the time step. Multiplication is based on the local energy.

A calculation generates a distribution of walkers with a concentration corresponding to the value of the function $f = \Psi\Psi_t$. For the determination of energies an average of local energies is used. Following Grimm and Storer³⁴ one can obtain the expression giving the energy as the average of local energies for the f -particles or walkers. Multiplying the time-independent Schrödinger equation by the trial function we obtain at any point

$$\Psi_t H \Psi = \Psi_t E \Psi . \quad (28)$$

Integrating over all space yields

$$\int \Psi_t H \Psi dX = \int \Psi E \Psi_t dX . \quad (29)$$

The Hermitian properties of wavefunctions, for identical boundary conditions and symmetries, allow a permutation to yield

$$\int \Psi H \Psi_t dX = \int \Psi E \Psi_t dX , \quad (30)$$

which may be rewritten as

$$\int \Psi \Psi_t \frac{H\Psi_t}{\Psi_t} dX = E \int \Psi \Psi_t dX . \quad (31)$$

This may be rearranged to give the energy as

$$E = \frac{\int \Psi H \Psi_t dX}{\int \Psi \Psi_t dX} \quad (32)$$

or

$$E = \frac{\int f E_{loc} dX}{\int f dX} . \quad (33)$$

The equivalent Monte Carlo expression, for equally weighted samples based on f , gives the energy as the average of local energies,

$$E = \frac{\sum \frac{H\Psi_t}{\Psi_t}}{\sum 1} = \frac{\sum E_{loc}}{\sum 1} . \quad (34)$$

The first applications in diffusion Monte Carlo were made for the nodeless ground state of the molecular ion H_3^+ .³⁵ The effect was a substantial improvement in accuracy from an energy of -1.3414 ± 0.0043 hartrees in an earlier calculation to -1.3439 ± 0.0002 hartrees in a similar calculation using importance sampling. The statistical error is reduced by a factor of about 20 and any systematic error is presumed to be similarly reduced.

The nodes of the trial function become the fixed nodes of the wavefunction Ψ which is the exact solution for the Schrödinger equation for boundary conditions corresponding to the fixed nodes. As for simple diffusion with fixed nodes the energy determined is an upper bound to the true energy.

Importance sampling may also be incorporated into GFQMC calculations. Although GFQMC calculations must treat walkers corresponding to the wavefunction Ψ itself rather than the product $\Psi \Psi_0$, one can repeatedly 'split' and/or 'kill' the Ψ -walkers to adjust their weights to be approximately inversely proportional to the local value of Ψ_0 . If this is done the individual weights in the summations of Eq. (34) may be made approximately equal and the calculations made reasonably efficient.

8 Trial Wavefunctions

In VQMC the accuracy of the trial function determines directly the accuracy in the energy obtained in a calculation. With importance sampling in either DQMC or GFQMC for systems without nodes the energy determined is, in principle, independent of the trial function, and only the uncertainty in the energy depends on the trial function. With fixed-node DQMC or GFQMC methods applied to systems with nodes the accuracy in the energy depends only on the accuracy of the node locations, and the uncertainty in the energy depends on the overall accuracy of the entire wavefunction. In any case, a more accurate trial wavefunction will improve a calculation by reducing the statistical uncertainty, by lowering the energy toward the exact value, and/or by reducing the extent of any systematic error such as time-step error.

The ideal trial wavefunction is simple and compact, has simple easily evaluated first and second derivatives, and is accurate everywhere. Since the local energy must be evaluated repeatedly the computation effort required for the derivatives makes up a large part of the overall computation effort for many systems. The typical trial wavefunctions of analytic variational calculations are not often useful, since they are severely restricted in form by the requirement that they be amenable to analytic integrations. The QMC functions are essentially unrestricted in form since no analytic integrations are required. First and second derivatives of trial wavefunctions are needed, but differentiation is in general much easier than integration and most useful trial wavefunctions have reasonably simple analytical derivatives. In most analytic variational calculations to date it has not been possible include the interelectron distances r_{ij} in the trial wavefunction and these wavefunctions are not usually “explicitly correlated”, but for QMC calculations of all types “explicitly correlated” functions containing r_{ij} are the norm.

A simple wavefunction for H₂ in its ground electronic state may be written as

$$\Psi_0 = (e^{-ar_{1A}} + e^{-ar_{1B}})(e^{-ar_{2A}} + e^{-ar_{2B}})e^{\frac{br_{12}}{1+cr_{12}}}. \quad (35)$$

In this the uncorrelated product of the two one-electron terms containing the electron-nucleus distances r_{iN} is multiplied by a Bijl or Jastrow function³⁶ incorporating r_{ij} ,

$$J = e^{\frac{br_{ij}}{1+cr_{ij}}}. \quad (36)$$

For most molecules even the simplest of trial wavefunctions is remarkably accurate. For hydrocarbons a single-determinant SCF function constructed with a minimal basis set and mildly optimized has an expectation value for the energy which corresponds to about 99 percent of the true energy. The nodes of these functions are also remarkably accurate and may be incorporated in functions giving 99.99 percent of the true energy.³⁷ That is not quite good enough in many cases – an error of 0.01 percent for methane corresponds to 2.5 kcal/mole – but it is a very good start.

The typical trial wavefunction for QMC calculations on molecular systems consists of the product of a Slater determinant multiplied by a second function which accounts to some extent for electron correlation with use of interelectron distances. The trial wavefunctions are most often taken from relatively simple analytic variational calculations, in most cases from calculations at the SCF level. Thus, for the 10-electron system methane³⁷ the trial function may be the product of the SCF function, which is a ten-by-ten determinant made up of two five-by-five determinants, and a Jastrow function for each pair of electrons,

$$\Psi_0 = \det^{\text{up}} \det^{\text{down}} \exp \left(\sum_{i < j} \frac{br_{ij}}{(1 + cr_{ij})} \right). \quad (37)$$

The values of b and c may be specified as 1/2 for pairs of electrons with opposite spins and as 1/4 for pairs with identical spins. This avoids infinities in the local energy for two electrons at the same position. The Jastrow functions incorporate the main effects of electron-electron interactions and give a significant improvement over simple SCF trial functions.

More accurate, more flexible expressions are available and these have been used with

considerable success. Schmidt and Moskowitz³⁸ explored functions of the type

$$\Psi_0 = \det^{\text{up}} \det^{\text{down}} \exp \left(\sum_{i < j} \sum_k c_k (q_i^n q_j^l + q_j^n q_i^l) q_{ij}^m \right) \quad (38)$$

in which n, l , and m are integers varying with k , and $q = r/(r+1.0)$. This particular form has been evaluated by Schmidt and Moskowitz³⁸ for a variety of molecular systems, and it has been used recently by Lüchow and Anderson³⁹ for first-row hydrides and by Alexander and Coldwell⁴⁰ for atomic systems. Modifications and extensions of the Schmidt-Moskowitz functional form have been investigated by Umrigar, Nightingale, and Runge²⁶ and by Alexander and Coldwell.⁴⁰

A variety of functional forms has been used for several very small systems. These include the molecules H₂, the ion H₃⁺, and the dimer He-He for which Hylleraas functions, Singer polynomials, and explicitly correlated Gaussian functions of very high accuracies have been used in QMC of all types.

The optimization of these functions has usually been carried out using the technique of minimizing the variance in local energies described by Conroy¹⁶ in the 1960's. In fact, it has only rarely been done in any other way.

9 Fixed-Node Calculations

The problem of node locations – the “sign problem in quantum Monte Carlo” – remains one of the major obstacles to obtaining exact solutions for systems of more than a few electrons. In analytic variational calculations and in VQMC the locations of the nodal surfaces of a trial function may be and usually are optimized along with the rest of the wavefunction in the attempt to reach a minimum in the expectation value of the energy. In DQMC and GFQMC the node locations are not so easily varied. For systems of a few electrons – excited H₂,⁴¹ H-H-H,^{42,43} He-He,^{44,45} H-He⁴⁶ – the node problem can be overcome by exact cancellation methods⁴⁰ (described below) and ‘exact’ solutions (i.e., solutions free of systematic error) can be obtained. But, in general, the method of choice for systems of more than about ten electrons is the fixed-node method. Although the fixed-node method is variational in nature and does not yield exact results, it is the only choice available for quantum Monte Carlo calculations on many larger systems. The fixed-node method is remarkably accurate, and it generally yields energies well below those of the best available analytic variational calculations.

The fixed-node method was first applied in DQMC calculations for the systems H ²P, H₂ ³S_u⁺, H₄ ¹S_g⁻, and Be ¹S.⁴⁷ The results indicated that very good energies could be obtained with node locations of relatively poor quality. Since the nodal surfaces of ground-state systems may be expected to be located in regions of low electron density (i.e., Ψ_0^2), one might expect the calculated energies to be insensitive to small departures in node locations from those of the true wavefunctions.

The fixed-node method is easily demonstrated for the case of the first excited state of a particle in a two-dimensional rectangular box. The true wavefunction has a nodal surface which is a line dividing the region into two rectangles - one in which the wavefunction is positive and the other in which the wavefunction is negative. The wavefunction is zero at the nodal line. A DQMC calculation performed for the positive region or for the negative

region using the true node line as a boundary on which the wave function is zero will produce the true wavefunction and energy for either region. If the true wavefunction is not known in advance, then similar calculations may be made in the same way using the node line of an approximate wavefunction. The energies for the two regions must be the same and for many systems and particularly for electronic systems, this can be assured by choosing a nodal surface which divides the overall space into two regions of the same shape so that one calculation is sufficient to determine the wavefunction and energy for both. The Schrödinger equation is solved exactly within the boundaries.

Unless the assumed nodal surface is exactly correct the overall wavefunction will not be exactly correct and the energy obtained will be an upper bound to the true energy. The fixed-node method is thus variational with respect to node locations. If the nodes are wrong the calculated energy will be higher than the true energy. Approximately correct nodal surfaces are most readily available from approximately correct wavefunctions provided by analytic variational calculations.

Fixed-node calculations may be carried out using the simple diffusion quantum Monte Carlo procedure described above. The nodal surface typically divides the configuration space into identical regions such that a calculation in only one region is required. The boundary condition of $\Psi = 0$ at the nodal surface is enforced by eliminating (killing) any walker which diffuses across a node. Energies may be calculated from the growth rate as described above using Eq. (19), but Eq. (20) is not applicable since walkers may disappear at the boundaries.

The molecule H₂ in its triplet state ${}^3\Sigma_u^+$ was one of the first molecules to be treated using the fixed-node quantum Monte Carlo method and it serves as a simple example. It has two electrons of like spin and a single nodal surface of five dimensions in the six-dimensional configuration space of the electrons, but because of symmetries the nodal surface is easily illustrated. The early variational calculations of James, Coolidge, and Present⁴⁸ give a fairly good energy and a reasonably accurate wavefunction for an internuclear distance of 1.6 bohrs. Their calculations were made with a number of approximate wavefunctions of increasing complexity and flexibility.

Fixed-node calculations⁴⁷ for H₂ ${}^3\Sigma_u^+$ at an internuclear separation of 1.4 bohrs were carried out using a nodal surface given by $\Psi = 0$ for $z_1 = z_2$ as suggested by the analytical wavefunctions. The value obtained for the energy, -0.79 ± 0.01 hartrees, was in good agreement with the value of -0.7831 hartrees obtained in analytic variational calculations by Kolos and Roothaan⁴⁹ and a more accurate value -0.7842 hartrees from more recent calculations.⁴¹

10 Exact Cancellation Method

The exact cancellation method overcomes the node problem for small systems and is thereby able to provide 'exact' solutions, i.e., solutions without systematic error and free of any physical or mathematical assumptions beyond those of the Schrödinger equation itself. The method has been applied successfully to a number of systems such as H-H-H, He-He, He-H, and He-He-He.⁴¹⁻⁴⁶

The method was proposed first by Arnow, Kalos, Lee, and Schmidt⁵⁰ in 1982 and was developed further with several practical improvements⁴¹ in 1991. We describe the improved method here. In its latest form it incorporates some of the best features of fixed-

node, released-node, and other cancellation methods. It takes full advantage of the symmetric and antisymmetric properties of wavefunctions and it offers pairwise cancellations of walkers as well as self-cancellations and multiple collective cancellations.

The basic idea of cancellation is most easily illustrated with the case of the first excited state of the one-dimensional harmonic oscillator. A quantum Monte Carlo calculation for the excited state can be carried out with positive and negative walkers, initially separated left and right of center. In the absence of cancellation the two populations spread throughout the available configuration space, penetrate each other, and independently approach the symmetric distribution for the ground state. If positive and negative walkers in close proximity are occasionally allowed to cancel each other, the two populations tend to cancel each other and produce separated distributions in which the net population on the left of center is positive and that on the right of center is negative. Without any control a fluctuation in populations will eventually lead to the dominance of either positive or negative walkers and a ground-state distribution all positive or all negative. But, if the two populations are controlled to maintain equal numbers of positive and negative walkers and if cancellations are properly executed, the net distribution evolves to that of the first excited state with the node at the center.

There are several ways of cancelling positive and negative particles. Some of these are rigorously correct but not efficient and some are efficient but not rigorously correct. One might cancel positive and negative particles occupying the same position, but the probability of two walkers occupying the same position is vanishingly small. For a one-dimensional system such as the harmonic oscillator one could efficiently cancel walkers passing each other, but that opportunity is not available for systems of higher dimensionality. One might cancel walkers within an arbitrary distance of each other, but that would lead to a bias in the distributions. Fortunately, there is one way which is rigorously correct and reasonably efficient for systems of a few electrons: cancellation on the basis of the overlap of the distributions to which the walkers are moved, specifically on the basis of Green's functions in GFQMC.

The distributions of weights for two walkers with weights W_1 and W_2 and and Green's functions G_1 and G_2 overlap by an amount O_{lap} given by

$$O_{lap} = \int \text{Min}(W_1 G_1, W_2 G_2) dX, \quad (39)$$

where $\text{Min}(W_1 G_1, W_2 G_2)$ is the smaller of $W_1 G_0(X, X'_1)$ and $W_2 G_0(X, X'_2)$. If the distance R separating the two walkers at positions X'_1 and X'_2 is zero the overlap is equal to the lesser of the two weights. For large separations the overlap approaches zero.

The partial cancellation of a pair of walkers may be carried out by a Monte Carlo procedure which may be generalized to multiple collective cancellations if desired.

The move for the first walker of the pair is selected unconditionally from the distribution $G_0(X, X'_1)$ and its weight at the new position X_1 becomes

$$W_1(\text{new}) = \frac{\text{Max}([s_1 W_1 G_1 - s_2 W_2 G_2], 0)}{s_1 G_1}. \quad (40)$$

The move for the second walker is treated similarly and its new weight is given by

$$W_2(\text{new}) = \frac{\text{Max}([s_2 W_2 G_2 - s_1 W_1 G_1], 0)}{s_2 G_2}. \quad (41)$$

Two walkers of the same weight and opposite sign at the same position cancel completely. In the limit of large separation the Green's function for the partner's move falls to zero and each walker keeps its original weight.

For the exact cancellation on the basis of overlapping Green's functions to be useful cancellations must occur often enough to maintain an adequate ratio of positive to negative walkers in regions where the wavefunction is positive and a similar ratio of negative to positive walkers in regions where the wavefunction is negative. Since there are multiple steady-state solutions for the ground state, fluctuations can shift the system from one solution to the other. In the case of the harmonic oscillator one solution is left-positive/right-negative and the other is left-negative/right-positive. To prevent shifts from one to the other and the resulting loss of information there must be an adequate number of walkers as well as an adequate cancellation rate. The required number of walkers and the required cancellation rate depend on the system investigated. Some systems are inherently more stable than others.

The energy for collections of positive and negative walkers may be determined with the aid of an importance sampling trial function having the same symmetry properties imposed on the collection. Equation (4) may be applied directly and, with the use of signs and weights, becomes

$$E = \frac{\sum s_i W_i \Psi_{0i} (\frac{H\Psi_0}{\Psi_0})_i}{\sum s_i W_i \Psi_{0i}}, \quad (42)$$

where the summation is over all walkers.

The most efficient calculations are those for which the sum of positive products $\Psi\Psi_0$ (or $s_i W_i \Psi_{0i}$) is large compared to the sum of negative products $\Psi\Psi_0$ (or $s_i W_i \Psi_{0i}$). This gives the highest signal-to-noise ratio in computing the energy.

The obvious way of increasing walker density to produce a high signal-to-noise ratio is to increase the number of walkers up to the limit of available computer memory. Beyond that one can make full use of symmetry to concentrate walkers in a single region of configuration space. For example, a system with rotational symmetry can be rotated to place a specific electron of a walker configuration in a specific plane. This decreases the distance R between them and increases the overlap O_{lap} . Similarly, electrons of the same spin can be ordered spatially by even numbers of permutations without changing the sign of their walker.

The choice of E affects the ratio of positive to negative walkers in several ways. Since E can be adjusted by arbitrary shifts in the zero of potential energy it can be chosen to optimize the ratio. The multiplication term (V/E) can switch the sign of a walker when the V is positive and E is negative. Since E must be negative the switching of signs can be reduced by shifting the zero of potential energy to make V negative in most regions of configuration space.

With increasing numbers of electrons the cancellation rate falls rapidly and beyond about four electrons, except in special cases, exact cancellation calculations become unstable. The 4-electron system LiH is difficult to treat, but the 8-electron He₄ system with usually well-separated atoms is well within the range of such calculations.

11 Difference Schemes

The difference δ between a true wavefunction Ψ and a trial wavefunction Ψ_0 may be determined directly in quantum Monte Carlo calculations. For an analytic trial function from any source the difference δ may be calculated and used to correct the trial function to obtain a wavefunction of higher accuracy and a more accurate eigenvalue. Successive corrections offer the possibility of unlimited accuracies. Thus far, the number of applications has been very few and the method has not been utilized in treating the problem of node locations, but difference methods offer some very interesting opportunities.

For many atomic and molecular systems approximate wavefunctions are easily obtained from SCF calculations with modest basis sets and the expectation values of the energies for these wavefunctions are typically within a few percent of the exact energies. Unfortunately, this is not good enough for most purposes. Nevertheless, such a wavefunction contains a significant amount of information and can provide a starting point for more accurate calculations.

Importance sampling, difference schemes, and their combinations all have the desirable characteristic of giving small errors for good trial wavefunctions and no errors in the limit of exact trial wavefunctions. Difference calculations have the additional desirable characteristic of correcting good trial wavefunctions to obtain better ones. Rather than calculate a complete wavefunction one may calculate the much smaller correction to a trial wavefunction. The statistical error normally associated with Monte Carlo calculations may then be limited to the correction term and thus reduced in size.

The difference method has been reported in two forms: first for simple diffusion QMC⁵¹ and second for importance sampling diffusion QMC with drift.⁵² In the case of simple diffusion one calculates the difference δ between a true wavefunction Ψ and a trial wavefunction Ψ_0 defined according to

$$\delta(X, \tau) = \Psi(X, \tau) - \Psi_0(X, \tau). \quad (43)$$

Substituting for Ψ in Eq. (4) and specifying Ψ_0 as fixed in time we obtain an equation for the change in δ with time

$$\frac{\partial \delta}{\partial \tau} = \frac{\hbar^2}{2m} \nabla^2 \delta - V \delta + [\frac{\hbar^2}{2m} \nabla^2 \Psi_0 - V \Psi_0]. \quad (44)$$

The equation is similar to Eq. (4). In addition to the diffusion and multiplication terms of Eq. (4) it has the term in brackets which corresponds to a distributed source fixed in time but varying with position.

As in simple diffusion QMC it is convenient to define the potential energy V of Eq. (44) with respect to a reference energy E_{ref} . With this Eq. (44) becomes

$$\frac{\partial \delta}{\partial \tau} = \frac{\hbar^2}{2m} \nabla^2 \delta - (V - E_{ref}) \delta + S. \quad (45)$$

The source term $S(X)$ may also be written in terms of the local energy $E_{loc} = H\Psi_0/\Psi_0$ at X for the trial wavefunction. The source term then becomes

$$S(X) = [-(E_{loc} - E_{ref})\Psi_0]. \quad (46)$$

The source term has the desirable property that as Ψ_0 approaches the true wavefunction and E_{ref} is adjusted to equal the true energy E the term approaches zero everywhere.

The procedure for determining δ is that same as that determining Ψ directly except that additional walkers are fed to the system at each time step as required by the source term. Additional walkers are fed to the system with a probability proportional to $|S|\Delta\tau$ at each point in space. These may be positive- or negative-valued depending on the local sign of S . When the reference energy E_{ref} is adjusted to maintain a fixed net weight (normally zero) of walkers their distribution approaches that of the function δ .

The continued feed of positive and negative δ -walkers leads in time to a large number of walkers in the system and it is necessary to control their number in some way. Cancellation of positive and negative walkers beyond a specified age – i.e., elapsed time since being fed – is perhaps the simplest means. With increasing age walkers fed at any location tend to the same distribution and they may be selected at random for cancellation. The energy E associated with a steady-state distribution may be evaluated from the reference energy E_{ref} required to maintain a fixed net weight of walkers.

The possibilities for successive corrections are apparent. The difference δ_1 determined in a calculation with an input trial wavefunction Ψ_0 may be added to Ψ_0 to obtain an improved trial wavefunction Ψ_1 . This, in turn, may be used as the input for a second calculation yielding a and a second correction δ_2 . The procedure may be extended to produce a series of functions $\Psi_1, \Psi_2, \Psi_3, \dots$ of increasing accuracy.

The simple difference scheme above may be combined with the importance sampling method of Grimm and Storer.³⁴ A new difference function, corresponding to the difference between the products $\Psi\Psi_0$ and $\Psi_0\Psi_0$, is defined as

$$g = (\Psi - \Psi_0)\Psi_0 . \quad (47)$$

When Eq. (47) is introduced to Eq. (27) we obtain upon rearrangement an equation for the feed, diffusion, drift, and multiplication of g -walkers,

$$\begin{aligned} \frac{\partial g}{\partial \tau} &= \frac{\hbar^2}{2m} \nabla^2 f - \nabla \cdot (f \nabla \ln \Psi_0) - \left(\frac{H\Psi_0}{\Psi_0} - E_{ref} \right) g \\ &\quad + \left[-\left(\frac{H\Psi_0}{\Psi_0} - E_{ref} \right) \Psi_0^2 \right] \end{aligned} \quad (48)$$

When Ψ_0 approaches the true wavefunction and E_{ref} approaches E , the feed and multiplication terms both approach zero.

The last term in Eq. (48) is the source term S which may be written as

$$S(X) = [-(E_{loc} - E_{ref})\Psi_0^2] \quad (49)$$

or, in a more convenient form using the expectation value of energy E_{var} ,

$$S(X) = [-(E_{loc} - E_{var})\Psi_0^2] + [-(E_{var} - E_{ref})\Psi_0^2]. \quad (50)$$

The procedure for determining the difference term g is similar to that for determining the difference δ described above. In this case, however, the g -walkers are subject to drift as in a conventional importance sampling calculation to determine f . As in calculating δ it is necessary to control the number of walkers and cancellation of positive and negative walkers beyond a specified age has been found effective. Applications to obtain energies of high accuracy for several systems have recently been described.⁵³

Some of the most interesting prospective applications are those for systems of 10 to 100 or more electrons for which the available trial wavefunctions are SCF wavefunctions.

These are easily generated along with accurate values for many of integrals required in sampling the source terms for difference calculations. The functions are relatively smooth and may allow reasonably large time-steps with minimal time-step error. An even more interesting possibility - that of an extension to correct node locations - remains an elusive but tantalizing target.

12 Excited States

Both DQMC and GFQMC provide the lowest-energy solution to the Schrödinger equation subject to any constraints which may be imposed on the solution. For excited states one must impose the necessary constraints.⁴⁷ In some cases this is relatively easy to do but in others it is difficult or as yet impossible. For these cases alternate methods are available: in particular, a matrix procedure applied to the evolution of several states at once in imaginary time.⁵⁴

The fixed-node method may be used for excited states when the nodes are known in advance as in the case of the 3P helium atom for which the nodal surface occurs at $r_1 = r_2$. For electronic systems of more than two electrons such a specification cannot be made in advance, but for vibrations of diatomic and polyatomic molecules the nodes for many modes of vibration can be specified from geometric considerations. Thus, fixed-node calculations have a place in calculations for excited states – especially for the first few states of small systems.

In GFQMC calculations with exact cancellation the unique symmetry of a desired state may be imposed at each step of a calculation together with importance sampling using a trial function of the same symmetry. This procedure has been used successfully to determine energies in the region of the Jahn-Teller cusp of the H-H-H potential energy surface at which symmetric and antisymmetric potential energy surfaces cross.^{42,43}

One may also impose the restriction of orthogonality to a ground or other lower state if the wavefunction for that state is known. If the wavefunction for the lower state of interest is not known explicitly, it may be possible to generate it in the form of a distribution of walkers in concomitant Monte Carlo calculations and the excited state distribution may then be restricted to a (net) zero overlap with the ground state. Several example systems have been treated in this way.⁵⁵

The matrix procedure applied to the time evolution of states requires only a single distribution of walkers propagated with a guide function as in importance sampling. Using a basis set of N trial wavefunctions one obtains the evolution of N states and their energies from the matrix elements between basis functions. The variance in energies increases exponentially with number of steps as for the released-node method. Nevertheless, excellent results have been obtained for the vibrations of H₂CO with as many as eight levels of vibration each of several modes determined with very high accuracy.⁵⁴

13 Use of Pseudopotentials

Quantum Monte Carlo calculations, like analytic variational calculations, can be considerably simplified – without a great loss in accuracy – by the use of effective potentials to replace core electrons close to the nuclei. In general, it has been found as expected that

as in analytic variational calculations with effective potentials or with frozen core basis sets, the energies of the core electrons and their effect on valence electrons will be almost exactly cancelled in subtracting to obtain relative energies for nearly identical systems. Since the energies of core electrons in heavy atoms are usually very much greater than the energies of valence electrons, including core electrons in QMC calculations is very much more expensive computationally when statistical error in the total energy must be reduced. In terms of local energy the core electrons are very 'noisy' and they contribute a disproportionate share of the variance in local energies. In addition, the sharper gradients in the core region lead to a requirement of much smaller time steps for accuracy in treating core electrons. The acceptable time-step size is much larger for outer electrons. The advantages of eliminating core electrons are large in proportion to the number of core electrons eliminated.

When core electrons are eliminated the Hamiltonian for the valence electrons of an atom becomes

$$\hat{H}_{val} = - \sum_i \frac{-Z_{eff}}{r_i} + \sum_{i < j} \frac{1}{r_{ij}} + \sum_i \hat{W}_i . \quad (51)$$

where the electrons are indexed i and j , Z_{eff} is an effective nuclear charge and \hat{W}_i is a pseudopotential operator for electron i .

The effective potentials normally used in analytic variational calculations are non-local potentials which involve angular projection operators which cannot be simply transferred into QMC calculations. In the earliest QMC calculations to use effective potentials Hurley and Christiansen⁵⁶ and Hammond, Reynolds, and Lester⁵⁷ avoided this difficulty with the use of local potentials defined in terms of trial wavefunctions. The use of effective potentials is, by its nature, not exact and introduces systematic errors which, in most cases thus far, have been found to be small. In later work non-local effective potentials^{58–60} have been used with success as have their more-complex counterparts, effective Hamiltonians. These too introduce systematic errors of finite size, but the errors are not easily analyzed and for that reason it is difficult to make judgements about the relative merits of the several methods.

The results of calculations using effective core potentials of the several types may be compared with experimental measurements, but more useful comparisons can be made with all-electron calculations for the same systems. For example, in studying the use of effective core potentials in QMC calculations Lao and Christiansen⁶¹ calculated the valence correlation energy for Ne and found excellent agreement with previous full-CI benchmark calculations. They recovered 98 to 100 percent of the valence correlation energy and could detect no significant error due to the effective potential approximation.

The advantage of the use of pseudopotentials is very dramatically illustrated by DQMC calculations for the Fe atom carried out by Mitas⁶² for (a) all-electrons, (b) for a neon-core pseudopotential, and (c) for an argon-core pseudopotential. The relative calculation effort for a fixed statistical uncertainty was in the same order (a) 6250, (b) 60, (c) 1. Thus, the appeal of pseudopotentials very strong. Of course, the additional (systematic) uncertainty introduced with the use of pseudopotentials is a disadvantage. Additional work will undoubtedly resolve the relative advantages and disadvantages.

A sampling of studies using effective potentials, model potentials, effective Hamiltonians, and related devices is given in Table 2. The entries range from the three-electron case

Authors	Ref.	Species
Hammond, Reynolds, and Lester (1987)	57	Li/Li ⁺ , Na/Na ⁺
Hurley and Christiansen (1987)	56	Li/Li ⁻ , K/K ⁻
Fahy, Wang, and Louie (1988)	63	Solid C(diamond)
Christiansen and LaJohn (1988)	64	Mg/Mg ⁺
Yoshida, Mizushima, and Iguchi (1988)	65	Cl/Cl ⁻
Carlson, Moskowitz, and Schmidt (1989)	66	Li/LiH, Li ₂ /2 Li
Bachelet, Ceperley, and Chiocchetti (1989)	60	Na ₂ /Na/Na ⁻ /Na ⁺ Mg, Si, Cl dimers and ions
Fahy, Wang, and Louie (1990)	67	Solid C(diamond), solid Si
Li, Ceperley, and Martin (1991)	68	Solid Si/Si
Shirley, Ceperley, and Martin (1991)	69	Be/Be ⁺ , Na/Na ⁺ , Sc/SC ⁺
Flad, Savin, and Preuss (1992)	59	Be/Be ⁺ , also Mg, Ca, Sr, Ba, Li, Na, K mixed dimers
Schrader, Yoshida, and Iguchi (1993)	70	PsF/Ps+F, PsCl/Ps+Cl
Belohec, Rothstein, and Vrbik (1993)	71	CuH (several states)
Tanaka (1993)	72	Solid NiO
Rajagopal, Needs, Kenny, Foulkes, and James (1994)	73	Solid Ge
Mitas (1994)	62	Fe/Fe ⁺ /Fe ⁻
Mitas and Martin (1994)	74	N, N ₂ , solid N, solid N ₂
Grossman, Mitas, and Raghavachari (1995)	75	C ₁₀ , C ₂₀
Greeff and Lester (1997)	76	Si _m H _n
Williamson, Rajagopal, Needs, Fraser, Foulkes, Wang, and Chou (1997)	77	Solid Si (1000 electrons)

Table 2. A sampling of QMC calculations with pseudopotentials.

of the Li atom, one of the earliest to be studied, to Cl atoms using neon-core pseudopotentials, to the atoms Al, Sc, and Fe, to clusters of Si and of silicon hydrides, to the diamond structure of solid C and Si, as well as that of GaAs.

Acknowledgments

Support by the National Science Foundation (Grants No. DGE-9987589 and CHE-9734808) is gratefully acknowledged.

References

1. Y.-S. M. Wu, A. Kuppermann, and J. B. Anderson, *Phys. Chem. Chem. Phys.* **1**, 929-937 (1999).
2. M. Head-Gordon, *J. Phys. Chem.* **100**, 13213 (1996).
3. B. L. Hammond, W. A. Lester, and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry*, World Scientific, Singapore, 1994.
4. K. E. Schmidt and J. W. Moskowitz, *J. Stat. Phys.* **43**, 1027-1041 (1986).
5. M. H. Kalos, in *Monte Carlo Methods in Quantum Problems*. M. H. Kalos, Ed., Reidel, Dordrecht, 1984, pp. 19-31.
6. K. E. Schmidt and M. H. Kalos, in *Monte Carlo Methods in Statistical Physics*. K. Binder, Ed., Springer-Verlag, Berlin, 1987, pp. 125-143.
7. D. Ceperley and B. Alder, *Science* **231**, 555-560 (1986).
8. B. H. Wells, in *Electron Correlation in Atoms and Molecules*. S. Wilson, Ed., Plenum Press, New York, 1987, pp. 311-350.
9. B. L. Hammond, M. M. Soto, R. N. Barnett, and W. A. Lester, *J. Mol. Structure (Theochem)* **234**, 525 (1991).
10. K. Raghavachari and J. B. Anderson, *J. Phys. Chem.* **100**, 12960-12973 (1996).
11. D. M. Ceperley and L. Mitas, in *Advances in Chemical Physics*, Vol. 93. I. Prigogine and S. A. Rice, Eds., Wiley, New York, 1996, pp. 1-38.
12. J. B. Anderson, in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*. S. R. Langhoff, Ed., Kluwer Academic Publishers, Dordrecht, 1995, pp. 1-45.
13. J. B. Anderson, *Int. Rev. Phys. Chem.* **14**, 85-112 (1995).
14. L. Mitas, in *Electronic Properties of Solids Using Cluster Methods*. T. A. Kaplan and S. D. Mahanti, Eds., Plenum Press, New York, 1995, pp. 131-141.
15. A. A. Frost, *J. Chem. Phys.* **10**, 240-245 (1942).
16. H. Conroy, *J. Chem. Phys.* **41**, 1331-1335 (1964). See also H. Conroy, *ibid.* **41**, 1336-1340 (1964), *ibid.* **41**, 1341-1351 (1964), *ibid.* **51**, 3979-3993 (1969).
17. W. L. McMillan, *Phys. Rev. A* **43**, 442-451 (1965).
18. R. Barnett, Z. Sun, and W. A. Lester, *Chem. Phys. Lett.* **273**, 321-328 (1997).
19. N. Metropolis and S. Ulam, *J. Am. Stat. Assoc.* **47**, 335-341 (1949).
20. J. B. Anderson, *J. Chem. Phys.* **63**, 1499-1503 (1975).
21. M. H. Kalos, *Phys. Rev.* **128**, 1791-1795 (1962).
22. R. P. Feynmann, *Statistical Mechanics*, Benjamin, Reading, Massachusetts, 1972.
23. D. M. Ceperley and E. L. Pollock, *Phys. Rev. Lett.* **56**, 351-354 (1986).
24. C. Pierleoni, D. M. Ceperley, B. Bernu, and W. R. Magro, *Phys. Rev. Lett.* **73**, 2145-2149 (1994).
25. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. M. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087-1092 (1953).
26. C. J. Umrigar, M. P. Nightingale, and K. J. Runge, *J. Chem. Phys.* **99**, 2865-2890 (1993).
27. R. Barnett, Z. Sun, and W. A. Lester, *Chem. Phys. Lett.* **273**, 321-328 (1997).
28. G. W. King, in Proceedings, IBM Computation Seminar, Endicott, New York, 1949 (International Business Machines Corp., New York, 1951), pp. 92-95.
29. A. Einstein, *Ann. Phys.* **17**, 549-560 (1905); *Ann. Phys.* **19**, 371-381 (1906).

30. D. M. Ceperley, *J. Comput. Phys.* **51**, 404-422 (1983).
31. D. W. Skinner, J. W. Moskowitz, M. A. Lee, P. A. Whitlock, and K. E. Schmidt, *J. Chem. Phys.* **83**, 4668-72 (1985).
32. J. W. Moskowitz and K. E. Schmidt, *J. Chem. Phys.* **85**, 2868-2874 (1986).
33. J. B. Anderson, *Phys. Rev. A* **35**, 3550 (1987).
34. R. C. Grimm and R. G. Storer, *J. Comput. Phys.* **7**, 134 (1971).
35. F. Menth and J. B. Anderson, *J. Chem. Phys.* **74**, 6307 (1981).
36. D. J. Klein and H. M. Pickett, *J. Chem. Phys.* **64**, 4811 (1976).
37. D. R. Garmer and J. B. Anderson, *J. Chem. Phys.* **86**, 4025 (1987).
38. K. E. Schmidt and J. W. Moskowitz, *J. Chem. Phys.* **93**, 4172 (1986).
39. A. Lüchow and J. B. Anderson, *J. Chem. Phys.* **105**, 7573 (1996).
40. S. A. Alexander and R. L. Coldwell, *Int. J. Quantum Chem.* **63**, 1001 (1997).
41. J. B. Anderson, C. A. Traynor, and B. M. Boghosian, *J. Chem. Phys.* **95**, 7418 (1991).
42. D. L. Diedrich and J. B. Anderson, *Science* **258**, 786 (1992).
43. D. L. Diedrich and J. B. Anderson, *J. Chem. Phys.* **100**, 8089-8095 (1994).
44. J. B. Anderson, C. A. Traynor, and B. M. Boghosian, *J. Chem. Phys.* **99**, 345 (1993).
45. J. B. Anderson, *J. Chem. Phys.* **115**, 4546 (2001).
46. A. Bhattacharya and J. B. Anderson, *J. Chem. Phys.* **100**, 8999 (1994).
47. J. B. Anderson, *J. Chem. Phys.* **65**, 4121 (1976).
48. H. M. James, A. S. Coolidge, and R. D. Present, *J. Chem. Phys.* **4**, 187 (1936).
49. W. Kolos and C. C. J. Roothaan, *Rev. Mod. Phys.* **32**, 219 (1960).
50. D. M. Arnow, M. H. Kalos, M. A. Lee, and K. E. Schmidt, *J. Chem. Phys.* **77**, 5562 (1982).
51. J. B. Anderson and B. H. Freihaut, *J. Comput. Phys.* **31**, 425 (1979).
52. J. B. Anderson, *J. Chem. Phys.* **73**, 3897 (1980).
53. J. B. Anderson, *J. Chem. Phys.* **112**, 9699 (2000).
54. B. Bernu, D. M. Ceperley, and W. A. Lester, *J. Chem. Phys.* **93**, 552-561 (1990).
55. D. F. Coker and R. O. Watts, *Mol. Phys.* **58**, 1113-1123 (1992).
56. M. M. Hurley and P. A. Christiansen, *J. Chem. Phys.* **86**, 1069 (1987).
57. B. L. Hammond, P. J. Reynolds, and W. A. Lester, *J. Chem. Phys.* **87**, 1130 (1987).
58. W. M. C. Foulkes and M. Schluter, *Phys. Rev. B* **42**, 11505 (1990).
59. H.-J. Flad, A. Savin, and H. Preuss, *J. Chem. Phys.* **97**, 459 (1992).
60. G. B. Bachelet, D. M. Ceperley, and M. G. B. Chiocchetti, *Phys. Rev. Lett.* **62**, 2088 (1989).
61. M. Lao and and P. A. Christiansen, *J. Chem. Phys.* **96**, 2162 (1992).
62. L. Mitas, *Phys. Rev. A* **49**, 4411 (1994).
63. S. Fahy, X. W. Wang, and S. G. Louie, *Phys. Rev. Lett.* **61**, 1631 (1988).
64. P. A. Christiansen and L. A. LaJohn, *Chem. Phys. Lett.* **146**, 162 (1988).
65. T. Yoshida, Y. Mizushima, and K. Iguchi, *J. Chem. Phys.* **89**, 5815 (1988).
66. J. Carlson, J. W. Moskowitz, and K. E. Schmidt, *J. Chem. Phys.* **90**, 1003 (1989).
67. S. Fahy, X. W. Wang, and S. G. Louie, *Phys. Rev. B*, **42**, 3503 (1990).
68. X.-P. Li, D. M. Ceperley, and R. M. Martin, *Phys. Rev. B*, **44**, 10929 (1991).
69. E. L. Shirley, L. Mitas, and R. M. Martin, *Phys. Rev. B* **44**, 3395 (1991).
70. D. M. Schrader, T. Yoshida, and K. Iguchi, *J. Chem. Phys.* **98**, 7185 (1993).
71. P. Belohorec, S. M. Rothstein, and J. Vrbik, *J. Chem. Phys.* **98**, 6401 (1993).
72. S. Tanaka, *J. Phys. Soc. Japan* **62**, 2112 (1993).

73. G. Rajagopal, R. J. Needs, S. Kenny, W. M. C. Foulkes, and A. James, Phys. Rev. Lett. **73**, 1959 (1994).
74. L. Mitas and R. M. Martin, Phys. Rev. Lett. **72**, 2438 (1994).
75. J. C. Grossman, L. Mitas, and K. Raghavachari, Phys. Rev. Lett. **75**, 3870 (1995).
76. C. W. Greeff and W. A. Lester, J. Chem. Phys. **106**, 6412 (1997).
77. A. J. Williamson, G. Rajagopal, R. J. Needs, L. M. Fraser, W. M. C. Foulkes, Y. Wang, and M.-Y. Chou, Phys. Rev. B **55**, 4851 (1997).

Path Integral Monte Carlo

Bernard Bernu¹ and David M. Ceperley²

¹ Laboratoire de Physique Théorique des Liquides
UMR 7600 of CNRS, Université Pierre et Marie Curie
boite 121, 4 Place Jussieu, 75252 Paris, France
E-mail: bernu@lptl.jussieu.fr

² Department of Physics and NCSA University of Illinois
Urbana-Champaign, Urbana, IL 61801, USA
E-mail: ceperley@uiuc.edu

In these notes, we present the basis of path integral techniques for indistinguishable and boson particles. The numerical evaluations of physical equilibrium properties needs an accurate action for the many body problem and an efficient algorithm to sample the paths.

1 Introduction

Feynman's path integral formulation of quantum mechanics has been proven to be very well suited to condensed boson systems, such as liquid or solid 4He , and fermion systems, such as liquid or solid 3He or a hydrogen plasma. Here we focus on problems than can be described by the following non-relativistic Hamiltonian:

$$\mathcal{H} = - \sum_{i=1}^N \lambda_i \nabla_i^2 + \sum_{i < j} v(|\mathbf{r}_i - \mathbf{r}_j|) \quad (1)$$

where N is the number of particles and $\lambda_i = \hbar^2 / 2m_i$. The pair potential is exactly known for hydrogen plasma ($v(r) = Z_i Z_j e^2 / r$) and also very well known for the interaction between helium atoms.² The same method applies for both the smoother but long range Coulomb potential and for the short range but strongly repulsive interaction between helium atoms.

These notes are taken from the most part from the review article¹ “*Path Integrals in theory of condensed helium*” published in *Rev. Mod. Phys.* **67** 280 (1995) where the reader can find many more details concerning path integral techniques as well as the physics of condensed helium 4. In these notes we give a minimum needed to start with path integral computations.

The second part of the notes is devoted to the exact mapping of the quantum problem to a classical one and general considerations on how path integrals are done. Because most interesting physical quantum systems deal with indistinguishable particles, the second part is about Bose systems.

2 Mapping of the Quantum to a Classical Problem

All static properties of quantum system in thermal equilibrium are obtained from the thermal density matrix. It can be defined in terms of the exact eigenvalues and eigenfunctions:

$$\rho = e^{-\beta \mathcal{H}} = \sum_i |\phi_i\rangle e^{-\beta E_i} \langle \phi_i|. \quad (2)$$

The equilibrium value of an operator \mathcal{O} is:

$$\langle \mathcal{O} \rangle = Z^{-1} \text{Tr} \rho \mathcal{O} = Z^{-1} \sum_i e^{-\beta E_i} \langle \phi_i | \mathcal{O} | \phi_i \rangle, \quad (3)$$

where Z is the partition function:

$$Z = \text{Tr} \rho = \sum_i e^{-\beta E_i}. \quad (4)$$

In the previous equations, the traces have been expressed in the eigenfunction basis. In the following we shall work exclusively in a position basis where particles are labeled. In such a basis, the density matrix elements are non-negative and can be interpreted as probabilities. The density matrix elements are thus defined as:

$$\rho(R, R'; \beta) \equiv \langle R | e^{-\beta \mathcal{H}} | R' \rangle = \sum_i \phi_i^*(R) e^{-\beta E_i} \phi_i(R'), \quad (5)$$

where $R = (\mathbf{r}^{(i)}, \dots, \mathbf{r}^{(N)})$, and $\mathbf{r}^{(i)}$ is the position of the i th particle. In space dimension 3, $\rho(R, R'; \beta)$ is, in general, a function of $6N + 1$ variables. In position representation, Eqs. (3-4) become:

$$\begin{aligned} \langle \mathcal{O} \rangle &= Z^{-1} \int dR \langle R | \rho \mathcal{O} | R \rangle \\ &= Z^{-1} \int dR dR' \rho(R, R'; \beta) \langle R' | \mathcal{O} | R \rangle, \quad \text{with} \end{aligned} \quad (6)$$

$$Z = \int dR \rho(R, R; \beta), \quad (7)$$

where the identity $\mathbb{I} = \int dR' |R'\rangle \langle R'|$ is introduced. From the definition of $\rho(\beta)$ (Eq. 2), one derives the Bloch Equation:

$$-\frac{\partial \rho}{\partial \beta} = \mathcal{H} \rho \quad (8)$$

with the initial condition $\rho(0) = \mathbb{I}$, which reads $\rho(R, R'; 0) = \delta(R - R')$ in the position representation.

2.1 Basis of Path Integral

The following equation is the basis of the path integral method:

$$\rho(\beta_1 + \beta_2) = e^{-(\beta_1 + \beta_2)\mathcal{H}} = e^{-\beta_1 \mathcal{H}} e^{-\beta_2 \mathcal{H}}, \quad (9)$$

and in position representation:

$$\langle R | \rho(\beta_1 + \beta_2) | R' \rangle = \int dR'' \langle R | e^{-\beta_1 \mathcal{H}} | R'' \rangle \langle R'' | e^{-\beta_2 \mathcal{H}} | R' \rangle, \quad (10)$$

$$\rho(R, R'; \beta_1 + \beta_2) = \int dR'' \rho(R, R''; \beta_1) \rho(R'', R'; \beta_2) \quad (11)$$

By repeating this process M times, one has:

$$e^{-\beta \mathcal{H}} = (e^{-\tau \mathcal{H}})^M \quad (12)$$

$$\begin{aligned} \rho(R_0, R_M; \beta) &= \int \dots \int dR_1 dR_2 \dots dR_{M-1} \rho(R_0, R_1; \tau) \\ &\quad \times \rho(R_1, R_2; \tau) \dots \rho(R_{M-1}, R_M; \tau), \end{aligned} \quad (13)$$

where $\tau = \beta/M$ is called the time step. The succession of the points (R_0, R_1, \dots, R_M) is called a path. Note that Eq. (13) is exact for any $M > 0$. In the limit $M \rightarrow \infty$, the path becomes continuous, but its derivative will be discontinuous at almost all points on the path. Equation (13) is useful because we can find sufficiently accurate analytical approximations of the density matrix at small τ .

In order to go further, we explicitly use the Hamiltonian. Let us suppose it is split in two pieces:

$$\mathcal{H} = T + V, \quad (14)$$

where T and V are the kinetic and potential operators which, in general, do not commute. The primitive approximation neglects all commutators between T and V :

$$e^{-\tau \mathcal{H}} \approx e^{-\tau T} e^{-\tau V}, \quad (15)$$

where the error is proportional to τ^2 . One can think that such approximation will lead to large error when the number of steps M increases ($\tau = \beta/M \rightarrow 0$). But, the Trotter formula shows this is a well controlled process provided the operators are bounded below:³

$$e^{-\beta \mathcal{H}} = \lim_{M \rightarrow \infty} \left[e^{-\frac{\beta}{M} T} e^{-\frac{\beta}{M} V} \right]^M. \quad (16)$$

We now write the primitive approximation Eq. (15) in the position representation:

$$\rho(R, R'; \tau) \approx \int dR'' \langle R | e^{-\tau T} | R'' \rangle \langle R'' | e^{-\tau V} | R' \rangle. \quad (17)$$

Usually the potential is diagonal in position representation:

$$\langle R'' | \exp(-\tau V) | R' \rangle = \exp(-\tau V(R')) \delta(R'' - R'). \quad (18)$$

The kinetic operator is diagonal in the reciprocal space. In a 3-dimensional box of length L , the free particle density matrix elements reads:

$$\begin{aligned} \rho_0(R, R'; \tau) &= \langle R | e^{-\tau T} | R' \rangle \\ &= \sum_{\mathbf{n}} L^{-3N} e^{-\tau \lambda 4\pi^2 n^2 / L^2 - 2i\pi \mathbf{n} \cdot (R - R') / L} \\ &= \frac{1}{(4\pi\lambda\tau)^{3N/2}} \exp\left(-\frac{(R - R')^2}{4\lambda\tau}\right), \end{aligned} \quad (19)$$

where \mathbf{n} is a vector with integer components and the last equality holds when the “size” of a particle (its thermal wavelength) is much smaller than the size of the box:⁴

$$\tau\lambda \ll L^2. \quad (20)$$

Inserting these expressions (Eqs. 18-19) in the primitive approximation (Eq. 17) and using it in Eq. (13) gives:

$$\begin{aligned} \rho(R_0, R_M; \beta) &= \int \dots \int dR_1 dR_2 \dots dR_{M-1} \\ &\frac{1}{(4\pi\lambda\tau)^{3NM/2}} \exp\left(-\sum_{m=1}^M \left[\frac{(R_{m-1} - R_m)^2}{4\lambda\tau} + \tau V(R_m)\right]\right). \end{aligned} \quad (21)$$

Thus the density matrix can be calculated at any temperature from an integral over the paths R_1, \dots, R_{M-1} . Such an integral looks like a partition function of some classical system. In particular, we see here that the integrand is always non-negative and can be interpreted as a probability; an essential property for Monte Carlo simulations.

2.2 Classical Isomorphism

We have now all pieces to understand the classical isomorphism. Let us start with a single particle where V might be some external potential. The path consists of a list of points $\mathbf{r}_0, \dots, \mathbf{r}_M$. In Eq. (21), the first term in the exponential $(\mathbf{r}_{m-1} - \mathbf{r}_m)^2/4\lambda\tau$ can be interpreted as the energy of a spring. Thus, a path is interpreted as a polymer where only first neighbors in the chain are connected with springs. Moving a quantum particle is equivalent to evolve this polymer.⁶ If we have many free particles, using the identity

$$\sum_{m=1}^M (R_{m-1} - R_m)^2 = \sum_{m=1}^M \sum_{i=1}^N (\mathbf{r}_{m-1}^{(i)} - \mathbf{r}_m^{(i)})^2 = \sum_{i=1}^N \sum_{m=1}^M (\mathbf{r}_{m-1}^{(i)} - \mathbf{r}_m^{(i)})^2, \quad (22)$$

we see that each particle can be interpreted as a polymer with strings connecting nearest neighbors in each chain. Thus quantum particles are represented by polymers in Path Integral language. Now adding the potential interaction between particles does not change this picture (see Fig. 1). We only have interacting polymers. But these polymers are simpler than classical polymers as the potential interactions occur only at the same “time” in the path. Indeed, in Eq. (21), the potential term depends only on R_m , which means that only the beads of the paths at the same time (τm) interact. In a real polymer, all beads interact with each other.

Most of the equilibrium quantities, such as energies, are obtained from the partition function (see Eq. (7)). In that case, the starting and ending beads of the polymers are the same: $R_M \equiv R_0$ (see Fig. 2). Quantum particles are represented by ring polymers, and the paths expression is symmetric under a cyclic permutation of (R_0, \dots, R_{M-1}) , thus any time can be chosen as a reference time. Note that this is true only for ring polymers. It is very instructive to keep in mind this picture that quantum particles are mapped on “classical polymers” with specific interactions.

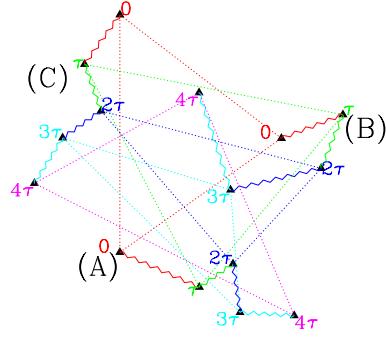


Figure 1. Cartoon of a path for 3 particles (A, B, C) with $M = 4$, $\tau = \beta/4$. The kinetic terms are represented by zig-zag lines and potential by dotted lines. Note that dotted lines connect only beads of the polymers at the same time

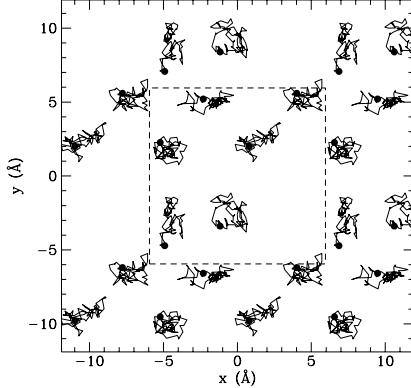


Figure 2. The trace of the close paths of six helium atoms at a temperature of $2K$ with $M = 80$. Straight lines (kinetic term) connect successive beads of the polymers. The filled circles are markers for the (arbitrary) beginning of the path. The paths have been replicated using the periodic conditions. The dashed lines represent the simulation cell.

2.3 Definition of the Action

The action is defined as minus the logarithm of the density matrix. For a given link m (see Eq. 13), one has:

$$S_m \equiv S(R_{m-1}, R_m; \tau) \equiv -\ln[\rho(R_{m-1}, R_m; \tau)]. \quad (23)$$

Then, the exact path integral expression (Eq. 13) becomes:

$$\rho(R_0, R_M; \beta) = \int \dots \int dR_1 dR_2 \dots dR_{M-1} \exp \left(- \sum_{m=1}^M S_m \right). \quad (24)$$

The action has an exact kinetic (free particle) contribution denoted K_m :

$$K_m = \frac{3N}{2} \ln(4\pi\lambda\tau) + \frac{(R_{m-1} - R_m)^2}{4\lambda\tau} \quad (25)$$

All the rest of the action is the inter-action (potential) contribution:

$$U_m \equiv U(R_{m-1}, R_m; \tau) = S_m - K_m \quad (26)$$

The primitive approximation reads:

$$U_m \equiv \frac{\tau}{2} [V(R_{m-1}) + V(R_m)], \quad (27)$$

where Eq. (18) has been symmetrized, a property of the exact action.⁵ Note this symmetrized form of the primitive action leaves unchanged Eq. (21) only for ring polymers where $R_0 \equiv R_M$.

With the primitive action one usually will need a large number of beads to get accurate properties, because this approximation is correct only at very small τ (high temperature). Semi-classical expansions give expressions which are even more singular at short distances (as they involve derivatives of the potential which are usually singular at the origin). But, one can considerably reduce the number of beads by solving “exactly” all partial two body problems.¹ It is then possible to write an N -body action that will be good as long as three body collisions are negligible. For helium such a good action is obtained at $\tau = 1/40K$. See the review for how to get a “good” action.

2.4 Sampling the Path

Once, one has the “best” available action for the many-body system, one has to define the moves. Path Integral methods are usually used within the context of generalized Metropolis, or Markov chain Monte Carlo methods. Other contributions will consider the possibility of using Molecular Dynamics methods to sample path space, but such methods are not applicable when quantum statistics are important.

2.4.1 Displacement

The simplest move consists in a translation of the whole path of a given number of particles. Such a move will not change the relative positions of the beads inside each polymer. Therefore, it does not change the kinetic energy. But they do change the potential energy as they do with classical Monte Carlo, though without relaxation of the internal bead position the potential so computed will be biased. Displacement moves are efficient to evaluate the potential energy, especially in the weak coupling regime.

2.4.2 Multislice Moves

To change the kinetic energy, one has to move the relative positions of the beads inside the polymer. Moving only one bead is very inefficient because of the springs connecting a bead with its neighbors. Thus it is necessary to have moves that will change several adjacent time-slices of the polymer. In practice, one cuts a slice $(r_{m+1}^{(i)} \dots r_{m+l-1}^{(i)})$ of the polymer, and samples a new path starting at $r_m^{(i)}$ and finishing at $r_{m+l}^{(i)}$. Several algorithms may be used to reconstruct the path. One can start at one end and build a path constrained to reach the other end. One can also use the bisection algorithm which first samples the mid point and if it is accepted, it then sample the mid point of each link, and so on. The efficiency of each algorithm is problem dependent. For most simulations of dense systems the bisection algorithm has been successful. The choice of the number of slices l of the polymer which is moved may be fixed or even better may be chosen at random. Multislice moves change the center of mass of the polymer and a change in the potential energy also. One can think such moves are enough, but in practice, it is always better to allow as many types of moves as possible.

2.4.3 Acceptance Ratio

All kind of moves may be tried, but we do not want to waste computer time trying moves that will be always rejected. On the other hand, when all moves are always accepted, we might worry about really moving in the phase space. In classical Monte Carlo, the parameters of the moves are usually chosen to get acceptance ratio of roughly 1/2. For moves with no good *a priori* sampling distribution, an acceptance ratio of 1/2 is correct. But when we have a very good *a priori* sampling distribution, for example at very small τ , the acceptance ratio gets naturally close to 1. It is of great importance to check the acceptance ratio for each type of moves. For example, in the bisection algorithm, we often have a poor sampling distribution at the first levels and the acceptance ratio is small, but when a path is eventually accepted, it has moved far away. Then the rest of the path gets better chance to succeed as the sampling action gets better (because it corresponds to smaller τ). In this algorithm, the time spent to evolve points with a poor action is much smaller than the time spent to evolve points with a good action.

3 Bose Symmetry

The Bose density matrix must be completely symmetric under any permutation of particle labels. It is also the solution of the Bloch equation (Eq. 8) with the symmetrized initial condition:

$$\rho_B(R, R'; 0) = \frac{1}{N!} \sum_{\mathcal{P}} \delta(R - \mathcal{P}R') \quad (28)$$

If ρ is solution for distinguishable particles, then the Bose density matrix is written in position representation:

$$\rho_B(R, R'; \beta) = \frac{1}{N!} \sum_{\mathcal{P}} \rho(R, \mathcal{P}R'; \beta) \quad (29)$$

The bad news is that we have now to evaluate $N!$ terms and there is no efficient algorithm to calculate it explicitly; in mathematics it is called a “permanent.” The good news is that all terms are positive and therefore the sum can be sampled by Monte Carlo. One has an additional discrete variable \mathcal{P} , a variable in the permutation space of $N!$ elements. Thus one has only to add another type of move to the list of Monte Carlo moves.

At high temperature, the identity permutation dominates, while at zero temperature, all permutations have equal probability. In classical polymer language, the action of a cyclic permutation \mathcal{P} with cycle length n , consists in opening the ring polymers involved in this permutation and making a single polymer out of them⁷ with nM beads. Typically only 2, 3 and 4-body ring permutations are tried (see the Appendix). They have the highest probability of success and are enough to walk through the permutation space, because all permutations can always be written as a product of pair permutations. The permutation sampled after p Monte Carlo steps is the product of all the accepted permutation moves since the beginning of the simulation.

Using the classical isomorphism, one can now see how the permutations allow new physics: when the temperature decreases, the particles are less and less localized (the number of beads increases), but two adjacent beads of a polymer are still constrained by the kinetic action which prevents the decrease of the kinetic energy. At low temperature, the ring-polymers try to be as elongated as possible. When two of them collide (which means some of the same-time-beads are close together), only a few beads are concerned. Then when the pair-permutation is tried, it is tried on these few beads: in both polymers, the slices at the same times are cut and new paths connecting polymer 1 to polymer 2 are built. If this move is been accepted, we are left with a ring-polymer with twice as much beads. The resulting polymer is more spread out so the kinetic energy has decreased. At high temperature, the number of beads is small and a polymer looks like a ball and the probability to have exchanging polymers is small. As the temperature decreases, the number of beads increases and the probability of alignment increases. First to appear are pair permutations. Longer cycles appear as the temperature is lowered.

By still decreasing the temperature, more and more polymers join to build longer ones. At some temperature, like in a percolation transition, a polymer of macroscopic size forms, connecting all portions of the systems. When this happens, there is a phase transition, and we have a fraction of the atoms in a superfluid phase. Properties of the transition can be calculated within PIMC. In a periodic system, the order parameter is measured by the winding number: the number of times a polymer wraps around the periodic boundary conditions. With a winding polymer, it is not possible to enclose the polymer completely in the simulation cell by only moving the center of the periodic box. This transition does not change much the local arrangement of particles and thus the pair correlation function changes very little around the transition.

4 Applications

We close the paper with a short list of applications for atomic and molecular problems that have been done using PIMC.

4.1 Boltzmanons

First, there are applications considering only the diffraction effects of quantum mechanics and not the effect of statistics. This is important when the thermal wave length associated with the particles is no longer negligible with respect to the mean spacing between them. Examples are systems of helium atoms or hydrogen molecules for $T > 5K$. These calculations are often quantum corrections to the classical results. Some examples are:

- Quantum effects on the kinetic energy of helium and heavier atoms.⁸
- Quantum effects on solids, melting or liquid-vapor transitions: *i*) the isotopic shift of the helium melting transition,⁹ *ii*) Debye-Waller factor of solid helium,¹⁰ *iii*) quantum effects on the critical point of helium.¹¹
- Atoms and molecules on various surfaces: *i*) rotations of molecules on a surface,¹² *ii*) wetting transition of helium.¹³
- Quantum effects on excess volumes of liquid hydrogen and neon mixtures.¹⁴

4.2 Bosons

There are a number of calculations concerned with the superfluid transition of 4He .¹ Because 4He is one of the simplest bosonic system for experimentalists as well as for theoreticians, it has been studied also in inhomogeneous conditions:

- droplets: superfluidity in doped helium/hydrogen droplets¹⁵
- hydrogen molecules can also Bose condensed and have been studied on surfaces, in mixture with helium and in confined geometry. For a review of bosons in surfaces geometry see ref.¹⁷

The hard sphere model is useful for the new Bose-Einstein condensation,¹⁶ for helium in porous media¹⁸ and vortices in High- T_c map onto bosons.¹⁹

Appendix

Permutation Sampling

In this section we describe the heat-bath algorithm to choose the permutation to be tried. The heat-bath probability for a permutation change between times i and $i + m$ is:

$$T^*(\mathcal{P}) \propto \rho(R_i, \mathcal{P}R_{i+m}) \quad (30)$$

Within the end-point approximation, these density matrix elements have the same potential part (thus can be dropped out) and T^* depends only on the kinetic part:

$$T^*(\mathcal{P}) = \frac{1}{C_I} \exp \left[- \sum_{j=1}^n \left(\mathbf{r}_i^{(j)} - \mathbf{r}_{i+m}^{(\mathcal{P}j)} \right)^2 / (4\lambda^* m\tau) \right], \quad (31)$$

where C_I is a normalization factor, λ^* is an effective mass to account off-diagonal contributions. $T^*(\mathcal{P})$ (Eq. 31) can be computed from the square matrix of the distances between end points:

$$t_{kj} = \exp \left[- \left(\mathbf{r}_i^{(k)} - \mathbf{r}_{i+m}^{(j)} \right)^2 / (4\lambda^* m \tau) \right]. \quad (32)$$

A random walk through this table will try to make a permutation of l atoms: the first atoms k_1 is chosen at random, the second atom k_2 is selected according to the probability $t_{k_1, k_2} / h_{k_1}$ where $h_{k_1} = \sum_k t_{k_1, k}$. After all of the l labels are selected and it is verified that they are unique, the trial permutation is accepted with the probability:

$$A = \min \left[1, \frac{\sum_{i=1}^l h_{k_i} / t_{k_i, k_i}}{\sum_{i=1}^l h_{k_i} / t_{k_i, k_{i+1}}} \right], \quad (33)$$

where $k_{l+1} \equiv k_1$. The sum of terms comes from the various starting point of the cyclic permutation. The probability of acceptance is rare because the last link t_{k_l, k_1} has $1/N$ chance to be not small (i.e. when the cycle closes on itself). But the process of constructing each loop is very rapid.

The physics help to fix the parameter l . On dense liquid, it is much to permute 3 or 4 than 2 atoms. A Monte Carlo simulation starts with the identity permutation. It is thus difficult to get pair exchanges accepted. On the contrary, once three and quadruple exchanges have build long cycles, pair exchanges can efficiently add or subtract from them. Also to get winding number changes, the parameter l must be large enough so that the permutation change can span the whole box.

References

1. David Ceperley, *Rev. Mod. Phys.* **67** 280 (1995).
2. R.A. Aziz, M.J. Slaman, A. Koide, A.R. Allnatt, and W.J. Meath, *Mol. Phys.* **77** 321 (1992).
3. B. Simon, *Functional Integration and Quantum Physics*, (1979) Academic, New-York.
4. The volume of the box is proportional to the number of particles (at fixed density). Thus for large enough number of particles the condition of Eq. (20) is satisfied.
5. The density matrix is symmetric by the exchange of the end points. Thus any “good” approximation should have this property. Also if we expand Eq. (15) in power of τ , the symmetrized form $\exp(-\tau\mathcal{H}) \approx \exp(-\frac{\tau}{2}\mathcal{V}) \exp(-\tau\mathcal{T}) \exp(-\frac{\tau}{2}\mathcal{V})$ is correct to order 2 instead of order 1 for Eq. (15). Note also that this symmetrized form leaves unchanged the Trotter formula.
6. See for example the moves of a single particle in a harmonic well: <http://www.physics.buffalo.edu/phy411-506/lectures.html>
7. Note that only Monte Carlo techniques allow such type of moves, where the polymers often cross each other.
8. Ceperley, D. M., R. O. Simmons and R. C. Blasdell, *Phys. Rev. Lett.* **77** 115 (1996).
9. Boninsegni, M., Pierleoni, C. and Ceperley, D. M., *Phys. Rev. Lett.* **72** 1854 (1994); J.L. Barat, P. Loubeyre, M.L. Klein, *J. Chem. Phys.* **90** 5644 (1989); C. Chakravarty and R. M. Lynden-Bell, *Journal of Chemical Physics* **113** 9239(2000).

10. M. Neumann, M. Zoppi, *Phys. Rev. B* **62** 41 (2000); Draeger, E. W., and D. M. Ceperley, *Phys. Rev. B* **61** 12094 (2000).
11. M. Müser, E. Luijten, cond-mat/0105283.
12. D. Marks, M. Müser, *J. Phys. CM*, **11** R 117 (1999). T Cui, E Cheng, B J Alder, *Phys. Rev. B* **55** 12253 (1997).
13. M. Boninsegni and M.W. Cole, *J. Low Temp. Phys.* **110** 685 (1998).
14. S. R. Challa and J. K. Johnson, *J. Chem. Phys.* **111** 724 (1999).
15. Yongkyung Kwon, K. Birgitta Whaley, *J. Chem. Phys.* **115** 10146 (2001); *J. Chem. Phys.* **114** 3163 (2001).
16. Gruter, P., D. Ceperley and F. Laloe, *Phys. Rev. Lett.* **79** 3549 (1997).
17. D. Ceperley and E. Manousakis, to appear in *J. Chem. Phys.* (2001).
18. M. C. Gordillo and D. M. Ceperley, *Phys. Rev. Lett.* **85** 4735 (2000).
19. Sen, P., N. Trivedi and D. M. Ceperley, *Phys. Rev. Lett.* **86** 4092 (2001).

Exchange Frequencies in 2D Solids: Example of Helium 3 Adsorbed on Graphite and the Wigner Crystal

Bernard Bernu¹, Ladir Cândido², and David M. Ceperley³

¹ Laboratoire de Physique Théorique des Liquides
UMR 7600 of CNRS, Université Pierre et Marie Curie
boite 121, 4 Place Jussieu, 75252 Paris, France
E-mail: bernu@lptl.jussieu.fr

² Instituto de Física de São Carlos, Universidade de São Paulo
13560-970 São Carlos, SP, Brazil
E-mail: ladir@if.sc.usp.br

³ Department of Physics and NCSA University of Illinois
Urbana-Champaign, Urbana, IL 61801, USA
E-mail: ceperley@ncsa.uiuc.edu

In 2d solids of fermion particles, such as helium 3 or electrons, the low temperature physics is governed by spin exchanges, according to the Thouless theory. We present Path Integral Monte Carlo (PIMC) calculation of ring exchange energies on “clean” 2d crystals of both helium 3 and electrons. We see a remarkable similarity of the results in these two “opposite” systems. They are both ferromagnetic in the semi-classical limit (strong coupling) antiferro magnetic near melting transition where the relative exchange energies become equivalent.

1 Introduction

In spin-less 2d solids, the low temperature physics is governed by the low excitations, the phonons. They provide a specific heat in $(T/\theta_D)^2$ in two dimensions, where the Debye temperature θ_D measures typical kinetic energy of a particle in its local potential. When $T/\theta_D \ll 1$, we can consider particles at zero temperature. Helium 3 atoms as well as electrons have a very large zero point motion. They eventually exchange their position resulting in a spin exchange that will modify the thermodynamics. Note here that helium 3 has a spin 1/2 nuclear spin whereas its two electrons are in a total spin 0 state. In the following we will consider helium 3 as atoms with a spin 1/2 interacting through a pair potential, say the Aziz potential. In the clean 2d Wigner crystal, electrons interact through the bare Coulomb potential.

The simplest effective model describing spin exchanges is the Heisenberg model:

$$H_{\text{spin}} = \sum_{\langle i,j \rangle} JP_{ij} = \sum_{\langle i,j \rangle} (2J\mathbf{S}_i \cdot \mathbf{S}_j - \frac{1}{2}). \quad (1)$$

where J is the energy associated to the spin permutation P , and the last equality holds for spin 1/2. Assuming that $J \ll \theta_D$, we look at the leading contribution of H_{spin} to the specific heat at large temperature (meaning $J \ll T \ll \theta_D$) which behaves as $(J/T)^2$. The crossover between a $1/T^2$ and a T^2 law has been well established in specific heat measurements of helium 3 adsorbed on graphite¹ (see Fig. 1). We see here a clear difference in

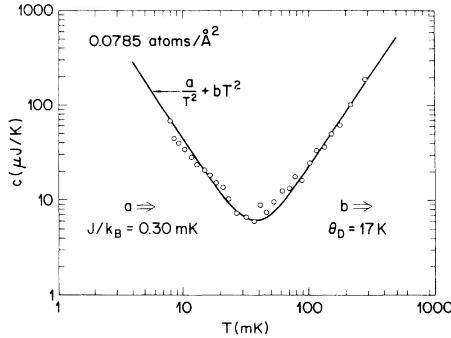


Figure 1. Specific heat measured by D. Greywall on 2d helium 3 adsorbed on graphite. One can see clearly the crossover between the spin contribution in $1/T^2$ and the phonon contribution in T^2 .

energy scale between the degree of freedom associated to the spins and those associated to the spatial coordinates. In the intermediate regime $J \ll T \ll \theta_D$, particles can be considered both at zero temperature for their spatial degree of freedom, moving in their zero point motion, and without spin as they do not contribute anymore to the thermodynamics. In this intermediate regime we use Path Integral Monte Carlo on a spinless 2d solid system to evaluate ring exchange energies using a method first introduced by D. Ceperley for the three dimension solid of helium 3.²

The full understanding of such a fermionic problem assume the Thouless theory can be applied.³ In this approach, one consider the spin-less Hamiltonian in spatial coordinates. At low temperature, each particle oscillate with a zero point motion around a lattice position Z . If the particles are distinguishable, there are $N!$ possibilities of labeling the particles corresponding to $N!$ different points in the $2N$ dimensional phase space. If the barriers between those points would have been infinite the ground state would be $N!$ degenerate. Finite barriers allow tunneling effects between different points Z and PZ which differ in a permutation P of their coordinates (labeling). There is an energy J_P associated to such tunnel effect. The main point is that such an event is very rare ($J_P \ll \theta_D$) so that different permutations never occur simultaneously. Thus one can study each permutation separately. With PIMC, one evaluate the energy of the tunnel effect for various permutations. We find that usually not only the two body exchange is important but also the 3, 4, 5 and 6 exchanges have large contributions, specially near melting. In the semi-classical limit, WKB calculations provide useful informations,^{13,12} specially for the Wigner crystal.¹⁴

The various ring exchange energies are accounted in the Multi spin Exchange (MSE) Heisenberg model:

$$H_{\text{MSE}} = \sum_P J_P P. \quad (2)$$

This Hamiltonian lift the degeneracy of the spin-less Hamiltonian ground state. The eigen-

spectrum of this Hamiltonian has 2^N states that are the fermionic available states among the previous $N!$ states. The fermion problem is thus pushed at the level of the MSE model. Even if solving the MSE Hamiltonian is far from trivial, this is a simpler problem. Thus we solve this fermion problem in two steps. In the first one, PIMC is used on a spin-less solid to evaluate the exchange energies. Then those energies are introduced in the effective MSE Hamiltonian. Finally different techniques may be used to get informations on such Hamiltonian : exact diagonalizations,⁴ high temperature series expansion,⁵ spin wave or Schwinger bosons analysis,⁶ Quantum Monte Carlo,¹⁰

The method to calculate exchange energy using PIMC has been already explained in details.^{8,9} Results on 2d helium 3 on graphite are in Ref.⁹ The helium-graphite potential has a very deep well which leads to up to two solid layers. The helium-helium pair-potential interaction is short range and has a strong repulsive part and each layer solidifies on a triangular lattice. The density of each layer can be tuned by changing the total amount of helium in the system. Helium is the only liquid at very low temperature and the graphite can be made to have very flat surfaces of a few hundred angstrom wide. Therefore, this is a very “clean” system where comparisons between experiments and theories should agree. The semi-classical limit is at high density and the melting occur at density around 6 nm^{-2} . Results on the Wigner crystal can be found in Ref.¹¹ The Coulomb interaction is smooth but long range. Therefore, the semi-classical limit (strong coupling) is at low density. The 2d solid is also a triangular lattice. Such system can be found at the surface of helium¹⁵ or at the interface of semi-conductors.¹⁶ Informations on the phase diagram of the resulting MSE model has been obtained from exact diagonalizations.⁴

In the next section we recall the basic idea of how exchange energies can be calculated by PIMC. In the following section we introduce a reaction coordinate that map the problem on a one-dimensional system. In section 4, we study in more detail a double well problem.

2 PIMC Method

When the temperature is lowered, the crystal of electron attains its ground state and the low energy phonons are frozen. Each electron still has a zero point motion with a substantial kinetic energy. When one continues to decrease the temperature, electrons start to exchange their positions by tunneling, resulting in a spin exchange.

Because exchanges are very rare, each exchange can be studied independently. Suppose we label the particles. There are $N!$ such labeling. Starting with a given numbering, one chooses a given permutation P . In the phase space, we denote by Z the position of the original numbering and PZ the position of the permuted system. We are left here with a two well problem in a multi dimensional space. In this two well system the ground state ψ_0 of energy E_0 is symmetrical and the first excited state ψ_1 of energy E_1 is anti symmetrical. Other states have much higher energies. The diagonal density matrix element $\langle Z | \exp(-\beta H) | Z \rangle$ and the off-diagonal density matrix element $\langle Z | \exp(-\beta H) | PZ \rangle$ can be expanded as :

$$\langle Z | \exp(-\beta H) | Z \rangle = \psi_0^2(Z) e^{-\beta E_0} + \psi_1^2(Z) e^{-\beta E_1} + \dots \quad (3)$$

$$\begin{aligned} \langle Z | \exp(-\beta H) | PZ \rangle &= \psi_0(Z) \psi_0(PZ) e^{-\beta E_0} + \psi_1(Z) \psi_1(PZ) e^{-\beta E_1} + \dots \\ &= \psi_0^2(Z) e^{-\beta E_0} - \psi_1^2(Z) e^{-\beta E_1} + \dots \end{aligned} \quad (4) \quad (5)$$

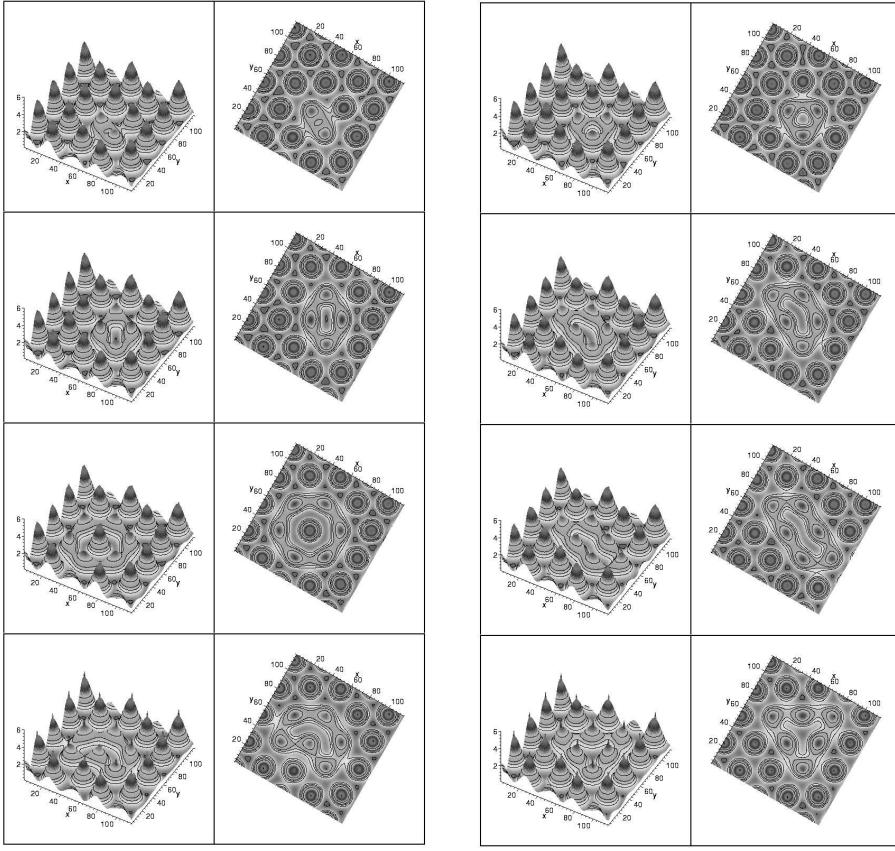


Figure 2. 3d plot of the one body density probability during exchanging paths for the Wigner crystal at $r_s = 40$, each data are represented from the side and from the top. The top 8 figures show 2, 3, 4 and 5 body exchanges, while the bottom 8 figures show the various 6 body exchanges. The “cones” represent the non exchanging particles. Notice that the center of mass of first neighbors of the exchanging particles are displaced from lattice position and their sizes are shrunken, especially in the 2 body exchange case. Note also the low probability density between electrons in the 6 body exchanges of the last row, indicating that those exchanges will be less probable.

where we have used the symmetry properties of the first two states. The ratio of these two density matrix elements is then:

$$F_P(\beta) = \frac{\langle Z | \exp(-\beta H) | P Z \rangle}{\langle Z | \exp(-\beta H) | Z \rangle} = \tanh(J_P(\beta - \beta_0)), \quad (6)$$

where J_P is the exchange frequency and $\beta_0 = \ln[\psi_1(Z)/\psi_0(Z)]$. Because $J_P\beta \ll 1$, one linearizes the previous equation and the slope of $F_P(\beta)$ gives the exchange energy. In terms of path integrals, one interprets $F_P(\beta)$ as the free energy necessary to make an exchange beginning with one arrangement of particles to lattice sites Z and ending on a permuted arrangement PZ . To evaluate $F_P(\beta)$, we use the method introduced by Bennett to calculate free energy differences of 2 chemical species A and B . The idea is to try to transform A in B and to calculate only the probability of success. Bennett⁷ proposed an optimized scheme where one does 2 runs. The first one is an equilibrium run of A where the probability of finding B is calculated. The second run is the reverse. Here run A will be the non permuting system and run B will be the permuted one. In A , we evaluate the probability of successful exchanges P and in B the probability of successful identity. The energy of each type of permutation is evaluated in an independent run. (More clever scheme can be certainly tried. More details can be found in refs.,^{8,2,9,11} specially on the optimized Bennett's method in ref.⁹

Fig. 2 shows the probability of presence of particle in the xy plane during exchanging paths. The first neighbors of exchanging particles must move away in order to free space to exchange paths (in particular, see the 2 body exchange). In this pictures are represented the most important exchanges. Larger cycles as well as exchanges including second neighbor will have significantly smaller contributions.

3 Reaction Coordinate

As seen in Fig. 2, the exchanging particles are mostly localized around there lattice sites. During exchanges, only a small part of the path is involved while the main part of the path stays around Z or PZ . The part of the path that does the exchange uses a “small” amount of imaginary time β_0 . Such process is called an instanton. The instanton occurs at any time between 0 and β . In the WKB calculations, where β goes to infinity, one has to remove first this degeneracy. In PIMC calculations, we must also find this degeneracy. We verify this property by calculating the exchange energy at the various imaginary time t . But, because the beads at times 0 and β are kept fixed in PIMC, there is some effect when the instanton touch the time 0 or β . For $\beta > \beta_0$, we find eventually a plateau. At large β , it can take a huge amount of CPU before obtaining a nice plateau. It is indeed hard to move the instanton at different imaginary times.

In order to get insight in this process, we define a reaction coordinate that help to map this multi dimensional problem onto a one dimensional one. A reaction coordinate allows to determine which part of the path is close to Z or PZ or exchanging :

$$z(t) = \frac{(R(t) - Z) \cdot (PZ - Z)}{|PZ - Z|^2} \quad (7)$$

For $z(t)$ close to 0 (resp. 1), the path is close to Z (resp. PZ). The figure 3-a shows $z(t)$ for the paths of run B . The exchanges take place at all time $0 < t < \beta$, but most of the time $z(t)$ is around 0 or 1. A crossing time t_c is defined by $z(t_c) = 1/2$. In figure 3-b are represented the same paths as functions of $t - t_c$. We see that the exchanges take roughly the same time. Because the exchange is localized in imaginary time, they are called instantons. Fitting these curves with $\tanh(2t/\beta_0)$, one defines the time β_0 needed for the path to go from Z to PZ .

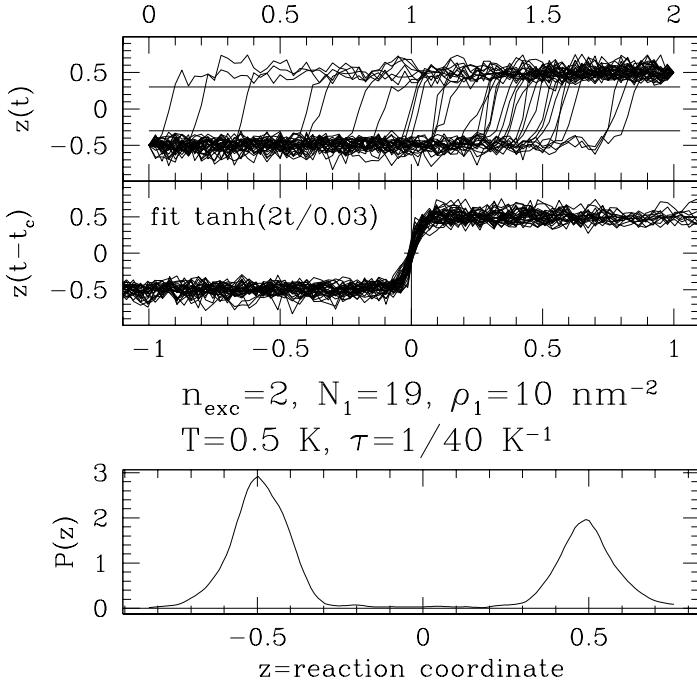


Figure 3. Reaction coordinate with respect to the time obtained in the *B* run. Only a few paths are used. top : row data where we see that exchange time arises at all possible time between 0 and β . middle : each path is now centered at the crossing time t_c . In bold solid line is the fit $\tanh(2t/\beta_0)$. bottom : probability of finding a value of z for these paths. The symmetry is due to the low number of paths used here.

The reaction coordinate can be used to map the $3N$ problem to a one dimensional one. First we build the probability distribution $P(z)$ from the value of $z(t)$, (see fig. 3-*c*). We define $\phi_0(z) = \sqrt{P(z)}$, where ϕ_0 is ground state of the Schrodinger equation $-\frac{1}{\phi_0} \frac{d^2 \phi_0}{dz^2} = E_0 - U(z)$. The pseudo potential $U(z)$ is then fully determined assuming the ground state energy is $E_0 = 0$. The last step is to calculate the anti-symmetric states of this potential. The exchange energy (as well as the potential) is determined to a multiplicative constant λ which represent an effective mass associated with this reaction coordinate. The comparison of this mapping with the direct method fixes this mass λ .

4 A One Dimensional Toy Model: A Particle in a Symmetrical Double Well

As an example suppose, we have a single particle in a symmetrical double well potential. The Hamiltonian reads : $\mathcal{H} = -\lambda \nabla^2 + V$, where V is the potential shown in Fig. 4. The ground state ϕ_0 of energy E_0 is symmetric and the first excited state ϕ_1 of energy E_1 is anti-symmetric. A particle localized in the left well is described by $\phi_L = \frac{1}{\sqrt{2}}(\phi_0 + \phi_1)$ and in the right well by $\phi_R = \frac{1}{\sqrt{2}}(\phi_0 - \phi_1)$. Such a localized particle will oscillate between the left and right wells with a period $h/(E_1 - E_0)$. The time associated with the motion of the particle inside one of the wells is h/K where K is the kinetic energy of the

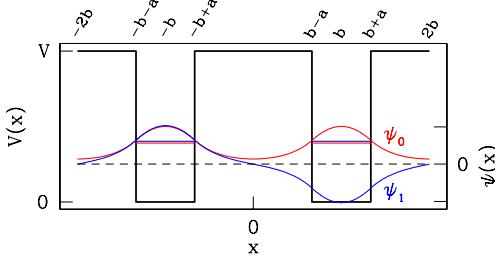


Figure 4. double well potential model. The ground state is a symmetrical function, whereas the first excited state, if it exists is antisymmetrical.

ground state. Let us suppose that $(E_1 - E_0) \ll K$. This means that the particle has a large number of zero-point vibrations in the left well before tunneling into the right well.

More generally, any permutation P may be viewed as a “particle” in two wells. The particle is the N -body system and the two wells represent the system around Z and PZ . The system of particles stay in the well around Z and eventually jump into the well PZ .

In order to understand how the exchange energies varies, we study the effects of the height V , the width $2a$ and separation $2b$ of the wells on the energy difference between the (symmetric) ground state and the first (anti-symmetric) excited state (see fig.4). For simplicity, we assume periodic boundary conditions.

Because of the periodic conditions and the symmetries of the wave functions, it is enough to study them in $[0, b]$ ($\psi(-x) = \pm\psi(x)$ and $\psi(b-x) = \psi(b+x)$). The ground state reads:

$$\psi_0(x) = A \cosh(x\sqrt{(V-E_0)/\lambda}) \quad x \in [0, b-a] \quad (8)$$

$$\psi_0(x) = B \cos((x-b)\sqrt{E_0/\lambda}) \quad x \in [b-a, b] \quad (9)$$

The coefficients A , B and E are determined by the continuity of the wave function and its derivative and the normalization condition. For $x = b-a$ we have:

$$A \cosh((b-a)\sqrt{(V-E_0)/\lambda}) = B \cos(a\sqrt{E_0/\lambda}) \quad (10)$$

$$A \sinh((b-a)\sqrt{(V-E_0)/\lambda})\sqrt{V-E_0} = B \sin(a\sqrt{E_0/\lambda})\sqrt{E_0} \quad (11)$$

The energy E_0 is the solution of the ratio of Eqs.10-11:

$$\tanh((b-a)\sqrt{(V-E_0)/\lambda})\sqrt{V-E_0} = \sqrt{E_0} \tan(a\sqrt{E_0/\lambda}) \quad (12)$$

Note that E_0 is also the ground state energy in a single periodic square well. The normalization condition reads:

$$A^2 \left(\frac{\sinh(2(b-a)\sqrt{(V-E_0)/\lambda})}{\sqrt{(V-E_0)/\lambda}} + 2(b-a) \right) + B^2 \left(\frac{\sin(2a\sqrt{E_0/\lambda})}{\sqrt{E_0/\lambda}} + 2a \right) = 1. \quad (13)$$

The ground state kinetic energy is given by:

$$K_0 = \frac{1}{2} \frac{\sqrt{(V - E_0)/\lambda} E_0 (Va - (V - E_0)bX)}{(\sqrt{(V - E_0)/\lambda}a + \sqrt{(1 - X)V + X\sqrt{(V - E_0)/\lambda}(bE_0 - aV)}}, \quad (14)$$

where $X = 1 - \tanh^2((b - a)\sqrt{(V - E_0)/\lambda})$. For small X , one finds $K_0 = \frac{1}{2}E_0a\sqrt{(V - E_0)/\lambda}/(a\sqrt{(V - E_0)/\lambda} + 1)$.

Similarly, the first excited state is defined by :

$$\psi_1(x) = A \sinh(x\sqrt{(V - E_1)/\lambda}x) \quad x \in [0, b - a] \quad (15)$$

$$\psi_1(x) = B \cos((x - b)\sqrt{E_1/\lambda}) \quad x \in [b - a, b + a] \quad (16)$$

E_1 is solution of :

$$\frac{\sqrt{V - E_1}}{\tanh((b - a)\sqrt{(V - E_1)/\lambda})} = \sqrt{E_1} \tan(a\sqrt{E_1/\lambda}) \quad (17)$$

Let us define the splitting energy by $E_1 = E_0 + \delta$. Inserting this definition in Eq. (17) and using Eq. (12), we get to the leading contribution in δ :

$$\delta = 8 \frac{(V - E_0)E_0 e^{-2(b-a)\sqrt{(V-E_m)/\lambda}}}{V(a\sqrt{(V - E_0)/\lambda} + 1)}. \quad (18)$$

The ratio δ/K_0 reads:

$$\frac{\delta}{K_0} = 16 \frac{V - E_0 e^{-2(b-a)\sqrt{(V-E_0)/\lambda}}}{Va\sqrt{(V - E_0)/\lambda}}. \quad (19)$$

Fig. 5 shows the variations of these quantities with respect to the potential height V and the distance $2(b - a)$ between the wells (V is measured in units of λ/a^2).

When V increases to infinity, E_0 approaches $\lambda(\pi/2a)^2$, K_0 goes to $E_0/2$ and $\delta \sim 8E_0e^{-2(b-a)\sqrt{V/\lambda}-\ln(a\sqrt{V/\lambda})}$. The main dependence in $\log(\delta)$ is $-2(b - a)\sqrt{V/\lambda}$, where $2(b - a)$ represents the width of the barrier. The formula of Eq. (18) is accurate for $(b - a)V/\lambda > 1$. The exchange energy is thus exponentially small with the distance between the wells and with the square root of the potential height.

One can have a large potential barrier, when V is much larger than E_m . But one can also consider a large *kinetic* barrier when V/E_m is of order of unity but the width $b - a$ of the barrier is large.

5 Results and Magnetic Phase Diagram

At strong coupling (high density for helium and low density for electrons), semi-classical (WKB) calculations are accurate. For electrons, the exchange energies are given by:^{12, 14}

$$J_P = A_P b_P^{1/2} r_s^{-5/4} e^{-b_P r_s^{1/2}}. \quad (20)$$

where $b_P r_s^{1/2}$ is the minimum value of the action integral along the exchanging path. This suggests to plot J_p versus $r_s^{1/2}$ as it is shown in Fig. 6. At large r_s , the 3-body exchange is dominant leading to a ferro magnetic ground state.

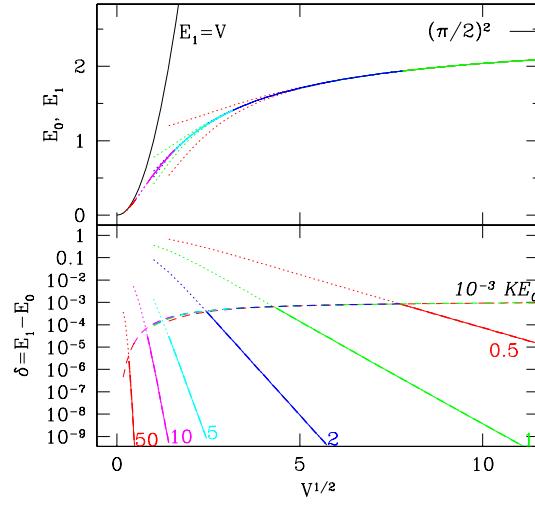


Figure 5. Top : Energies E_0 and E_1 versus V in units of λ/a^2 for $(b-a)/a = 1/2, 125$. Bottom : $\delta = (E_1 - E_0)$ and 10^{-3} times the kinetic energy.

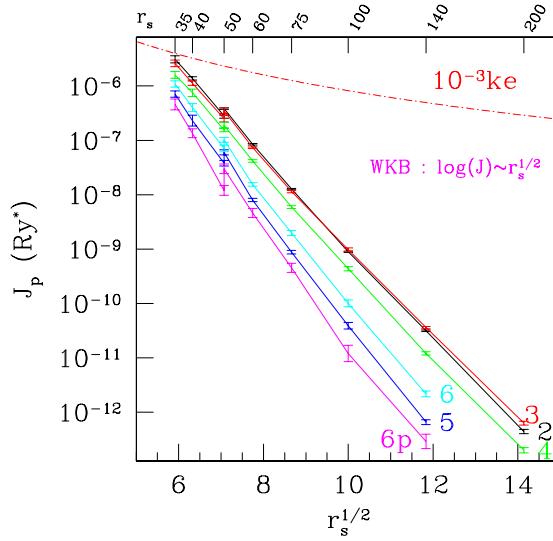


Figure 6. Exchange energies versus $r_s^{1/2}$. For $r_s \geq 50$ non exchanging electrons are distinguishable, and for $r_s \leq 50$, there are polarized (preliminary results). One can see that near melting, exchange energies become comparable with the kinetic energy.

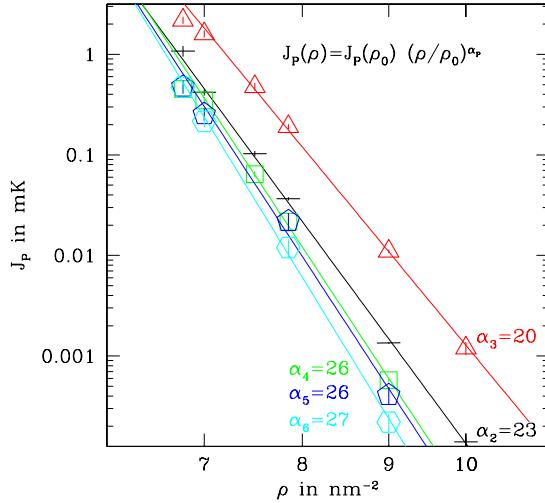


Figure 7. Helium 3 adsorbed on graphite: first layer exchange frequencies versus the density.

For helium 3 adsorbed on graphite, we find also strong variations with the density as shown in Fig. 7. The 3-body exchange is dominant for all densities and its exponent is also the smallest. This imply a ferromagnetic ground state at high density.

For both cases, at intermediate coupling the other 2, 4, 5 and 6 body exchanges are more and more important as the system approaches the melting transition. We put all these exchanges in the Hamiltonian defined in Eq. (2). The ferromagnetic 3 and 5 body exchanges are in competition with the antiferromagnetic 2, 4 and 6 ones. The nature of the ground state is sensitive to there relative values. The phase diagram of this Hamiltonian, obtained from exact diagonalization, is shown Fig. 8.⁴ For spin 1/2, the 3 body permutations can be written in terms of pair permutations defining an effective pair permutation $J_2^{\text{eff}} = J_2 - 2J_3$ which can be positive or negative. Thus we choose J_4 to scale energies and we are left with 3 parameters : J_2^{eff}/J_4 , J_5/J_4 and J_6/J_4 . The straight lines in Fig. 8 separates the ferromagnetic (F) region from the antiferromagnetic (AF) one. The “trajectories” of the Wigner crystal and the second solid layer of helium 3 adsorbed on graphite crosses the F-AF transition line. In the AF region, no long range order has been found but on the contrary they are in a spin liquid state with a gap in all excitations.^{4,6}

A remarkable feature is the similarity of the relative exchanges (the trajectories are closed to each other), in particular when approaching the melting transition. Yet the interactions of these two systems are very different from a short range strongly repulsive potential for helium to a long range smooth potential for electrons. The search of an underlying universal mechanism is thus highly desirable (possibly virtual vacancy-interstitial (VI) mechanism¹³).

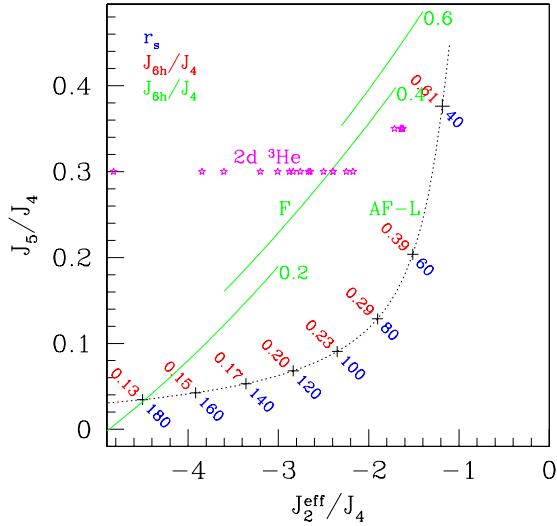


Figure 8. Zero temperature magnetic phase diagram of the MSE hamiltonian.⁴

6 Conclusion

For fermionic solids, PIMC allows to evaluate exchange energies. These energies are then introduced in a MSE hamiltonian which in turn can be studied by different techniques. It is found that various two dimensional systems have a dominant 3 body exchange in the semi-classical limit (strong coupling) leading to a ferromagnetic ground state. As the quantum kinetic contributions increase, all exchanges become comparable with competitive ferro and antiferro interactions. Near melting it is found that the relative exchanges in the Wigner crystal are very similar with those obtained for a solid layer of helium 3 adsorbed on graphite, suggesting a possible universal behavior of the exchange mechanism near the melting transition.

References

1. D. S. Greywall and P. A. Busch, *Phys. Rev. Lett.* **65** 2788 (1990); D. S. Greywall *Phys. Rev. B* **41** 1842 (1990).
2. D. M. Ceperley and G. Jacucci, *Phys. Rev. Lett.* **58**, 1648 (1987).
3. D. J. Thouless, *Proc. Phys. London* **86**, 893 (1965).
4. G. Misguich, B. Bernu, C. Lhuillier and C. Waldtmann. *Phys. Rev. Lett.* **81** 1098 (1998); *Phys. Rev. B* **60** 1064 (1999).
5. M. Roger, C. Bauerle, Yu. M. Bunke, A.-S. Chen and H. Godfrin, *Phys. Rev. Lett.* **80**, 1308 (1998).

6. G. Misguich, B. Bernu and C. Lhuillier. *Journal of Low Temperature Physics* **110**, 327 (1998).
7. C. H. Bennett, *J. Comput. Phys.* **22** 245 (1976).
8. D. M. Ceperley, *Rev. Mod. Phys.* **67**, 279 (1995).
9. B. Bernu and D. Ceperley in *Quantum Monte Carlo Methods in Physics and Chemistry*, eds. M. P. Nightingale and C. J. Umrigar, Kluwer (1999).
10. S. Sorella, *Phys. Rev. B* **64** 024512 (2001), *Phys. Rev. B* **61** 2599 (2000).
11. B. Bernu, L. Candido, D. Ceperley *Phys. Rev. Lett.* **86** 870 (2001).
12. M. Roger, *Phys. Rev. B* **30** 6432 (1984).
13. M. Roger, J. H. Hetherington and J. M. Delrieu, *Rev. Mod. Phys.* **55** 1 (1983).
14. K. Voelker, S. Chakravarty, (cond-mat/0107151).
15. C. C. Grimes and G. Adams, *Phys. Rev. Lett.* **42**, 795 (1979).
16. J. Yoon *et al.*, *Phys. Rev. Letts.* **82**, 1744 (1999).

Reptation Quantum Monte Carlo

Stefano Baroni^{1,2} and Saverio Moroni^{2,3}

¹ SISSA – Scuola Internazionale Superiore di Studi Avanzati

² INFM – Istituto Nazionale per la Fisica della Materia
via Beirut 2-4, 34014 Trieste, Italy

³ Dipartimento di Fisica, Università di Roma “La Sapienza”
Piazzale Aldo Moro 2, 00185 Roma, Italy
E-mail: {baroni, moroni}@sissa.it

<http://xxx.lanl.gov/abs/cond-mat/9808213>

Reprinted from: Quantum Monte Carlo Methods in Physics and Chemistry, edited by M. P. Nightingale and C. J. Umrigar. NATO ASI Series, Series C, Mathematical and Physical Sciences, Vol. 525, (Kluwer Academic Publishers, Boston, 1999), p. 313.

We present an elementary and self-contained account of the analogies existing between classical diffusion and the imaginary-time evolution of quantum systems. These analogies are used to develop a new quantum simulation method which allows the calculation of the ground-state expectation values of local observables without any mixed estimates nor population-control bias, as well as static and dynamic (in imaginary time) response functions. This method, which we name *Reptation Quantum Monte Carlo*, is demonstrated with a few case applications to ${}^4\text{He}$, including the calculation of total and potential energies, static and imaginary-time dependent density response functions, and low-lying excitation energies. Finally, we discuss the relations of our technique with other simulation schemes.

1 Introduction

The theory of stochastic processes plays an important role in modern developments of quantum mechanics both as a deep – and possibly not yet fully exploited – conceptual method^{1,2} and as a powerful practical tool for the computer simulation of interacting quantum systems.³ Quantum Monte Carlo simulations mainly rely on the *static* properties of a random walk of one kind or another that is used to sample the ground-state wave-function or finite-temperature density matrix of a system. Comparatively minor attention has been paid so far to the *dynamical* properties of the random walks used in quantum simulations. The main interest in these properties stems from the study of excitation energies, a notoriously difficult and ill-conditioned problem. These dynamical properties, also determine the magnitude of autocorrelation times which are necessary to estimate statistical errors.

The purpose of these lectures is to provide an elementary and self-contained presentation of the deep relations existing between diffusion in classical systems and the imaginary-time evolution of quantum systems, and to develop some ideas based on these relations⁴ which lead to a new, promising technique for performing quantum simulations. For reasons which will become apparent in the following, we name this technique *Reptation Quantum Monte Carlo (RQMC)*. The main features of RQMC are that it is based on a *purely diffusive* process without branching, that ground-state expectation values of local observables can be evaluated without any mixed estimates, and that static and dynamic (in imaginary time)

response functions are natural by-products of the ground-state simulation. Although some of the ideas presented here are not new,⁴ our method does not suffer from the drawbacks which affected their previous implementations.

2 From Classical Diffusion to Quantum Mechanics

The simplest phenomenological description of classical diffusion is given by the Langevin equation:

$$dx = f(x)d\tau + d\xi, \quad (1)$$

where x is some generalized coordinate describing our system, $f(x) = -\frac{\partial v(x)}{\partial x}$ is the force acting on it – which we suppose to be derivable from a potential – τ is time and $\xi(\tau)$ is a Wiener process: $\langle d\xi \rangle = 0$; $\langle (d\xi)^2 \rangle = 2d\tau$. Although the random walk generated by the Langevin equation, Eq. (1), can be given a mathematically unambiguous meaning in the continuous limit, it is simpler – and more useful in view of applications to quantum simulations – to specialize to a given discretization of time: $\tau_k = k \times \epsilon$. The corresponding random walk is then described by the discrete Markov chain:

$$x_{k+1} = x_k + \epsilon f(x_k) + \xi_k, \quad (2)$$

where the ξ 's are uncorrelated Gaussian random variables of zero mean, $\langle \xi_i \rangle = 0$, and variance 2ϵ , $\langle \xi_i \xi_j \rangle = 2\epsilon \delta_{ij}$.

2.1 The Fokker-Planck Equation

The time evolution of the probability distribution for the variable x is given by the master equation which, in the one-dimensional case, reads (the generalization to many dimensions is straightforward):

$$P(x, \tau + \epsilon) = \int \mathcal{W}_\epsilon(x|y) P(y, \tau) dy, \quad (3)$$

where

$$\mathcal{W}_\epsilon(x|y) = \frac{1}{\sqrt{4\pi\epsilon}} e^{-\frac{(x-y-\epsilon f(y))^2}{4\epsilon}} \quad (4)$$

is the conditional probability that the system is in configuration x at time $\tau + \epsilon$, given that it is found in configuration y at time τ . When the conditional probability is independent of τ (as it is in the present case) the corresponding Markov process is said to be *homogeneous*. In order to convert the master equation, Eq. (3), into a differential equation, we perform a Taylor expansion of its right-hand side in powers of the time step ϵ :

$$P(x, \tau + \epsilon) = \frac{1}{\sqrt{4\pi\epsilon}} \int \delta(x - y - \epsilon f(y) - \xi) e^{-\frac{\xi^2}{4\epsilon}} P(y, \tau) d\xi dy. \quad (5)$$

Taking into account that $\xi \sim \sqrt{\epsilon}$, we now formally Taylor-expand the δ function in powers of $(\epsilon f(y) + \xi)$, and we obtain:

$$P(x, \tau + \epsilon) = \int P(y, \tau) \left(\sum_n \frac{(-1)^n}{n!} \delta^{(n)}(x - y) \times \langle (\epsilon f(y) + \xi)^n \rangle \right) dy, \quad (6)$$

where $\langle \cdot \rangle$ indicates the Gaussian average of the polynomials in ξ , and $\delta^{(n)}$ is the n -th derivative of the δ function:

$$\begin{aligned}\langle (\epsilon f(y) + \xi)^n \rangle &\equiv \frac{1}{\sqrt{4\pi\epsilon}} \int_{-\infty}^{\infty} (\epsilon f(y) + \xi)^n e^{-\frac{\xi^2}{4\epsilon}} d\xi \\ &= (-i\sqrt{\epsilon})^n H_n \left(i \frac{\sqrt{\epsilon}}{2} f(y) \right),\end{aligned}\quad (7)$$

H_n being the Hermite polynomial of order n .⁵ To linear order in ϵ , the first few values of these Gaussian integrals are: $\langle \epsilon f(y) + \xi \rangle = \epsilon f(y)$, $\langle (\epsilon f(y) + \xi)^2 \rangle = 2\epsilon + \mathcal{O}(\epsilon^2)$, $\langle (\epsilon f(y) + \xi)^3 \rangle = \mathcal{O}(\epsilon^2)$. Integrals of the derivatives of the δ function are given by: $\int g(y) \delta^{(n)}(x-y) dy = g^{(n)}(x)$. Using this relation, Eq. (6) can be recast as:

$$P(x, \tau + \epsilon) = P(x, \tau) + \epsilon \left(-\frac{\partial}{\partial x} (f(x)P(x, \tau)) + \frac{\partial^2 P(x, \tau)}{\partial x^2} \right) + \mathcal{O}(\epsilon^2). \quad (8)$$

Eq. (8) is the discrete-time version of the Fokker-Planck equation which, in the continuous limit, reads:⁶

$$\frac{\partial P(x, \tau)}{\partial \tau} = \frac{\partial^2 P(x, \tau)}{\partial x^2} - \frac{\partial}{\partial x} (f(x)P(x, \tau)). \quad (9)$$

The fact that this equation is first order in time is strictly related to the Markovian character of the random walk, Eq. (2). It is easily verified that $P_s(x) \propto e^{-v(x)}$ is a stationary solution of the Fokker-Planck equation, Eq. (9). Using Eq. (8) one sees that the stationary distribution of the *discrete* random walk, Eq. (2), differs from P_s by terms of order ϵ .

The conditional probability, \mathcal{W}_ϵ , defined in Eq. (4), can be seen as the *exact* probability that the system described by the *discrete* Markov chain, Eq. (2), makes a transition from configuration y to configuration x during the time step ϵ ; alternatively, it can be seen as an approximation to the transition probability in the continuous limit, correct to order $\mathcal{O}(\epsilon^2)$. In the continuous limit the exact conditional probability is defined by the relation:

$$P(x, \tau) = \int \mathcal{W}(x, \tau | y, 0) P(y, 0) dy. \quad (10)$$

By inserting this definition into Eq. (9), one sees that $\mathcal{W}(x, \tau | y, 0)$ satisfies Eq. (9) with respect to x , subject to the boundary condition: $\mathcal{W}(x, 0 | y, 0) = \delta(x-y)$, *i.e.* $\mathcal{W}(x, \tau | y, 0)$ is the Green's function of the Fokker-Planck equation, Eq. (9). The Markovian character of random walk, Eq. (2), allows one to define a simple functional-integral representation for $\mathcal{W}(x, \tau | y, 0)$ ⁷ which closely resembles the path-integral representation of the Green's function of the time-dependent Schrödinger equation.⁸ As we will see in the following, this resemblance is by no means accidental nor superficial.

Let $X_N = \{x_0, x_1, \dots, x_N\}$ be a given random walk generated by Eq. (2). Because of the Markovian character of the chain, the corresponding probability density, $\mathcal{P}_\epsilon[X_N]$ satisfies the relation:

$$\begin{aligned}\mathcal{P}_\epsilon[X_N] &\equiv \text{Prob}[x(N\epsilon) = x_N; x((N-1)\epsilon) = x_{N-1}; \dots; x(0) = x_0] \\ &= \mathcal{W}_\epsilon(x_N | x_{N-1}) \text{Prob}[x((N-1)\epsilon) = x_{N-1}; \dots; x(0) = x_0] \\ &= \mathcal{W}_\epsilon(x_N | x_{N-1}) \mathcal{P}_\epsilon[X_{N-1}],\end{aligned}\quad (11)$$

where $x(\tau)$ is the configuration of the system at time τ . By iterating this equation N times, one obtains:

$$\mathcal{P}_\epsilon[X_N] = \mathcal{W}_\epsilon(x_N|x_{N-1})\mathcal{W}_\epsilon(x_{N-1}|x_{N-2}) \cdots \mathcal{W}_\epsilon(x_1|x_0)P(x_0, 0). \quad (12)$$

The probability density that the system is in configuration x_N at time $\tau = N\epsilon$ is obtained from $\mathcal{P}_\epsilon[X_N]$ by integrating out the N variables, $\{x_0, \dots, x_{N-1}\}$:

$$\begin{aligned} P(x_N, N\epsilon) &= \int \mathcal{P}_\epsilon[\{x_0, \dots, x_N\}] dx_0 \cdots dx_{N-1} \\ &= \int \mathcal{W}_\epsilon(x_N|x_{N-1}) \cdots \mathcal{W}_\epsilon(x_1|x_0)P(x_0, 0) dx_0 \cdots dx_{N-1}. \end{aligned} \quad (13)$$

By comparing Eq. (10) with Eq. (13), one obtains the desired functional-integral representation for \mathcal{W} :

$$\begin{aligned} \mathcal{W}(x, \tau|y, 0) &= \\ &\int \mathcal{W}_\epsilon(x|x_{N-1})\mathcal{W}_\epsilon(x_{N-1}|x_{N-2}) \cdots \mathcal{W}_\epsilon(x_1|y) dx_1 \cdots dx_{N-1}, \end{aligned} \quad (14)$$

where $\epsilon = \tau/N$. The above representation is *exact* for the *discrete* Markov chain described by Eq. (2). In the continuous limit ($\epsilon \rightarrow 0$) it is understood that it holds by letting $N = \tau/\epsilon \rightarrow \infty$, while keeping τ fixed.

2.2 The Classical-Quantum Mapping

In order to establish the link between classical diffusion and imaginary-time quantum evolution, we define a wave-function, $\Phi(x, \tau)$, through the relation:

$$P(x, \tau) = \Phi_0(x)\Phi(x, \tau), \quad (15)$$

where $\Phi_0(x) = \sqrt{P_s(x)} \propto e^{-v(x)/2}$. By inserting Eq. (15) into Eq. (8) and dividing by $\Phi_0(x)$, we obtain:

$$\Phi(x, \tau + \epsilon) = (1 - \epsilon\mathcal{H})\Phi(x, \tau) + \mathcal{O}(\epsilon^2), \quad (16)$$

where:

$$\mathcal{H} = -\frac{\partial^2}{\partial x^2} + \mathcal{V}(x), \quad (17)$$

and

$$\begin{aligned} \mathcal{V}(x) &= \frac{1}{4} \left(\frac{\partial v}{\partial x} \right)^2 - \frac{1}{2} \frac{\partial^2 v}{\partial x^2} \\ &= \frac{1}{\Phi_0(x)} \frac{\partial^2 \Phi_0(x)}{\partial x^2}. \end{aligned} \quad (18)$$

The continuous-time limit of Eq. (16) is formally equivalent to an imaginary-time Schrödinger equation,

$$\frac{\partial \Phi(x, \tau)}{\partial \tau} = -\mathcal{H}\Phi(x, \tau), \quad (19)$$

which could as well have been derived directly by inserting Eq. (15) into Eq. (9).

The wave-function Φ_0 is a solution of Eq. (19), which means that it is an eigenfunction of the time-independent Schrödinger equation, corresponding to a zero eigenvalue. A general theorem of quantum mechanics states that the ground-state eigenfunction of a Schrödinger equation with a local potential is non-degenerate and node-less.⁹ Orthogonality with respect to excited-state wave-functions implies that the ground state is the only node-less eigenfunction. We thus arrive at the conclusion that all the excited-state energies are strictly positive and that $\Phi_0(x)$ is the only time-independent solution of Eq. (19) or, equivalently, that $P_s(x) \propto e^{-v(x)}$ is the only stationary solution of the Fokker-Planck equation (9).¹⁰ Furthermore, any solution of Eq. (9) would tend to P_s for large times, irrespective of the initial condition. If the spectrum of \mathcal{H} has a gap, the approach to equilibrium is exponentially fast. This is easily seen by simple inspection. Consider a system whose probability distribution at time $\tau = 0$ is $P(x, 0)$ and its expression in terms of the eigenfunctions and eigenvalues of \mathcal{H} , which we indicate by Φ_n and \mathcal{E}_n :

$$P(x, 0) = \Phi_0(x) \sum_n c_n \Phi_n(x). \quad (20)$$

The time evolution of P is readily derived from the time evolution of the Φ 's:

$$P(x, \tau) = c_0 \Phi_0(x)^2 + \sum_{n \neq 0} c_n e^{-\mathcal{E}_n \tau} \Phi_0(x) \Phi_n(x). \quad (21)$$

The normalization of $P(x, 0)$ and the orthonormality of the Φ 's imply that $c_0 = 1$ and that the norm of $P(x, \tau)$ is conserved. The above equation shows that the thermalization time – *i.e.* the time necessary for the system to reach equilibrium – is $\tau_0 \approx \frac{1}{\mathcal{E}_1}$.

The fact that P_s is the only stationary solution of the Fokker-Planck equation, Eq. (9), implies that classical expectation values over P_s – or, equivalently, quantum expectation values over $\Phi_0(x)$ – can be expressed as time averages over the random walk, $x(\tau)$, generated by the Langevin equation, Eq. (1):

$$\int P_s(x) \mathcal{A}(x) dx \equiv \langle \Phi_0 | \mathcal{A} | \Phi_0 \rangle = \lim_{T \rightarrow \infty} \lim_{t \rightarrow \infty} \frac{1}{t} \int_T^{T+t} \mathcal{A}(x(\tau)) d\tau. \quad (22)$$

A comparison between Eq. (3) and Eq. (16) allows one to establish a relation between the transition probability of the classical random walk, \mathcal{W}_ϵ , and the propagator of the quantum system associated with it:

$$\mathcal{W}_\epsilon(x|y) = \Phi_0(x) \mathcal{G}(x, y; \epsilon) / \Phi_0(y) + \mathcal{O}(\epsilon^2), \quad (23)$$

where:

$$\mathcal{G}(x, y; \tau) \equiv \langle x | e^{-\tau \mathcal{H}} | y \rangle. \quad (24)$$

The fact that the error in Eq. (23) is indeed of second order in ϵ , and not higher, can be proved by pushing the Taylor expansion of Eq. (6) to second order in ϵ and by noting, for instance, that the second-order term would give rise to a non-hermitian contribution to \mathcal{G} . Using Eqs. (12) and (23), the probability density for the random walk X_N can be easily expressed in terms of a product of quantum propagators, \mathcal{G} . Assuming that the system is at equilibrium at $\tau = 0$ – *i.e.* that the probability distribution for x_0 is $P_s(x_0)$ – the

probability distribution for the random walk is:

$$\begin{aligned}\mathcal{P}_\epsilon[X_N] &= \underbrace{\Phi_0(x_N)\mathcal{G}(x_N, x_{N-1}; \epsilon) \cdots \mathcal{G}(x_1, x_0; \epsilon)\Phi_0(x_0)}_{\mathcal{P}[X_N]} + \mathcal{O}(\epsilon) \\ &\equiv \mathcal{P}[X] \times \mathcal{Q}_\epsilon[X_N],\end{aligned}\quad (25)$$

where $\mathcal{Q}_\epsilon[X_N] = 1 + \mathcal{O}(\epsilon)$ is a time-discretization correction factor. The Markovian character of the random walk implies a simple composition law for the \mathcal{P} probability distributions:

$$\begin{aligned}\mathcal{P}[\{x_0, \dots, x_k, \dots, x_N\}] &= \mathcal{P}[\{x_0, \dots, x_k\}] \times \\ &\quad \mathcal{P}[\{x_k, \dots, x_N\}] / P_s(x_k).\end{aligned}\quad (26)$$

Inspection of Eq. (25) shows that the probability distribution $\mathcal{P}[X]$ is invariant under time reversal:

$$\mathcal{P}[\bar{X}] = \mathcal{P}[X], \quad (27)$$

where $\bar{X} \equiv \{x_N, \dots, x_0\}$ is the path obtained from X under time reversal. In the short-time limit, this time-reversal invariance simply expresses the detailed balance condition:

$$\mathcal{W}_\epsilon(x|y)P_s(y) \approx \mathcal{W}_\epsilon(y|x)P_s(x). \quad (28)$$

As it can be seen from Eq. (23), the above relation holds to $\mathcal{O}(\epsilon^2)$, and from a mathematical point of view it is strictly connected with the hermitian character of the quantum propagator, \mathcal{G} .

2.3 Classical vs. Quantum Time Correlation Functions

Eq. (22) expresses the identity between classical thermal expectation values of local observables and quantum expectation values of the same observables calculated over the ground state of a suitably defined auxiliary system. Furthermore, these expectation values can be expressed in terms of time averages over a random walk. In the following we will see how this result can be extended to time correlation functions. It will result that (real) time correlation functions calculated over the random walk coincide with the ground-state imaginary-time correlation functions of the auxiliary quantum system.

The time auto-correlation function of an observable $\mathcal{A}(x)$ is defined as:

$$\langle \mathcal{A}(\tau)\mathcal{A}(0) \rangle = \int \mathcal{A}(x)\mathcal{A}(y) \text{Prob}[x(0) = y; x(\tau) = x] dx dy. \quad (29)$$

If the stochastic process is stationary (*i.e.* if $P(x, \tau) = P_s(x)$), then one has:

$$\begin{aligned}\text{Prob}[x(0) = y; x(\tau) = x] &= \mathcal{W}(x, \tau|y, 0)P_s(y) \\ &= \Phi_0(x)\mathcal{G}(x, y; \tau)\Phi_0(y).\end{aligned}\quad (30)$$

By inserting this relation into Eq. (29), the auto-correlation function reads:

$$\begin{aligned}\langle \mathcal{A}(\tau)\mathcal{A}(0) \rangle &= \int \mathcal{A}(x)\mathcal{A}(y)\Phi_0(x)\mathcal{G}(x, y; \tau)\Phi_0(y) dx dy \\ &= \langle \Phi_0 | e^{\mathcal{H}\tau} \mathcal{A} e^{-\mathcal{H}\tau} \mathcal{A} | \Phi_0 \rangle \\ &\equiv \langle \Phi_0 | \mathcal{A}(-i\tau)\mathcal{A}(0) | \Phi_0 \rangle,\end{aligned}\quad (31)$$

where we have used the fact that $e^{\mathcal{H}\tau}|\Phi_0\rangle = |\Phi_0\rangle$ and the definition of a time-dependent operator in the Heisenberg representation: $\mathcal{A}(t) = e^{i\mathcal{H}t}\mathcal{A}e^{-i\mathcal{H}t}$.

Eq. (31) can be used to express the autocorrelation time of the observable \mathcal{A} in terms of the spectral properties of the auxiliary quantum Hamiltonian, \mathcal{H} . The autocorrelation time is defined as the time integral of the normalized autocorrelation function:

$$\tau_{\mathcal{A}} = \int_0^\infty \underbrace{\frac{\langle \mathcal{A}(\tau)\mathcal{A}(0) \rangle - \langle \mathcal{A} \rangle^2}{\langle \mathcal{A}^2 \rangle - \langle \mathcal{A} \rangle^2} d\tau}_{c_{\mathcal{A}}(\tau)} \quad (32)$$

The integrand in Eq. (32) is defined in such a way that $c_{\mathcal{A}}(0) = 1$ and $c_{\mathcal{A}}(\infty) = 0$. When $c_{\mathcal{A}}$ is exponential, $\tau_{\mathcal{A}}$ is simply its decay constant: $c_{\mathcal{A}}(\tau) = e^{-\tau/\tau_{\mathcal{A}}}$. In practical simulations, $\tau_{\mathcal{A}}$ determines the statistical errors in the measure of \mathcal{A} :

$$\langle \mathcal{A} \rangle \approx \frac{1}{N} \sum_{i=M+1}^{M+N} \mathcal{A}(x_i) \pm \sqrt{\Delta \mathcal{A}^2 \frac{\tau_{\mathcal{A}}}{N\epsilon}}, \quad (33)$$

where M is chosen large enough so that equilibrium is reached (*i.e.* $M > \frac{1}{\epsilon E_1}$), and $\Delta \mathcal{A}^2 \approx \frac{1}{N-1} \sum (\mathcal{A}(x_i) - \langle \mathcal{A} \rangle)^2$. In order to proceed further, we first consider the spectral resolution of the quantum propagator, \mathcal{G} :

$$\mathcal{G}(x, y; \tau) = \sum_n e^{-\mathcal{E}_n \tau} \Phi_n(x) \Phi_n(y). \quad (34)$$

By inserting Eq. (34) into Eq. (31), the auto-correlation time can be easily cast into the form:

$$\tau_{\mathcal{A}} = \frac{1}{\Delta \mathcal{A}^2} \sum_{n \neq 0} \frac{|\langle \Phi_n | \mathcal{A} | \Phi_0 \rangle|^2}{\mathcal{E}_n}. \quad (35)$$

3 From Quantum Mechanics Back to Classical Diffusion

The purpose of many quantum simulation techniques is to study the ground-state properties of a system whose Hamiltonian is

$$H = -\frac{\partial^2}{\partial x^2} + V(x), \quad (36)$$

and whose (unknown) ground-state wave-function and energy we indicate by Ψ_0 and E_0 , respectively. In the variational Monte Carlo method (VMC), an approximate wave-function, Φ_0 , is used to generate a random walk according to the discrete Langevin equation, Eq. (2), with

$$\begin{aligned} f(x) \equiv f_{VMC}(x) &= -\frac{\partial}{\partial x} (-\log \Phi_0(x)^2) \\ &= 2 \frac{1}{\Phi_0(x)} \frac{\partial \Phi_0(x)}{\partial x}, \end{aligned} \quad (37)$$

and the ground-state expectation value of the operator \mathcal{A} is estimated through Eq. (33), where $\mathcal{A}(x) = \frac{1}{\Phi_0(x)} \mathcal{A} \Phi_0(x)$. Systematic errors in Eq. (33) depend on the discretization

of time and are of order ϵ . They can be eliminated in principle by enforcing the detailed-balance condition on the master equation, Eq. (3).¹¹ A variational upper bound, $\bar{\mathcal{E}}$, to the ground-state energy can be estimated from Eq. (33):

$$\bar{\mathcal{E}} \equiv \langle \Phi_0 | H | \Phi_0 \rangle \approx \frac{1}{N} \sum_i \mathcal{E}(x_i), \quad (38)$$

where

$$\mathcal{E}(x) = \frac{1}{\Phi_0(x)} H \Phi_0(x) \quad (39)$$

is the so-called *local energy*. In the following we will show how the *dynamical* properties of the random walk (1) can be used to systematically improve upon this VMC procedure and to estimate, *exactly* within statistical noise, the ground-state properties of quantum systems.

Let us first observe that the trial wave-function, Φ_0 , implicitly defines a reference (unperturbed) Hamiltonian, H_0 , whose *exact* ground state is Φ_0 . The potential function which defines H_0 is simply obtained by inverting the time-independent Schrödinger equation:

$$V_0(x) = \frac{1}{\Phi_0(x)} \frac{\partial^2 \Phi_0(x)}{\partial x^2} + \bar{\mathcal{E}}. \quad (40)$$

A comparison between Eq. (40) and Eqs. (17,18) shows that the potential obtained from the inversion of the Schrödinger equation coincides up to a constant with the effective quantum potential resulting from the classical-quantum mapping discussed in Sec. 2.2 ($V_0(x) = \mathcal{V}(x) + \bar{\mathcal{E}}$), provided that the classical random walk, Eqs. (1,2), is driven by the VMC force defined in Eq. (37).

The original Hamiltonian, Eq. (36), can then be cast into the form:

$$H = \underbrace{-\frac{\partial^2}{\partial x^2}}_{\mathcal{H}} + \mathcal{V}(x) + \bar{\mathcal{E}} + \underbrace{\mathcal{E}(x) - \bar{\mathcal{E}}}_{\Delta \mathcal{H}}. \quad (41)$$

By construction, Φ_0 is the ground state of \mathcal{H} , and $\mathcal{E}(x) = E_0$ if Φ_0 is the ground state of H . When Φ_0 is not the ground state of H , the local energy $\mathcal{E}(x)$ is not a constant, and the ground-state properties of H can in principle be obtained by perturbation theory with respect to $\Delta \mathcal{H}$.

Let us calculate the corrections to the ground-state energy, E_0 , up to second order in $\Delta \mathcal{H}$. The first-order correction vanishes by construction, given that $\langle \Phi_0 | H | \Phi_0 \rangle = \bar{\mathcal{E}}$. The second-order correction is given by:

$$\Delta E_0^{(2)} = - \sum_{n \neq 0} \frac{|\langle \Phi_0 | \mathcal{E} | \Phi_n \rangle|^2}{\mathcal{E}_n}, \quad (42)$$

where \mathcal{E} is the local energy *operator*, and we have taken into account the fact that the unperturbed ground-state energy is zero ($\mathcal{E}_0 = 0$). By comparing this expression with Eq. (35), the second-order correction to the ground-state energy can be expressed in terms of the fluctuations of the local energy and of its auto-correlation time:

$$\Delta E_0^{(2)} = \tau_{\mathcal{E}} \Delta \mathcal{E}^2. \quad (43)$$

In the following, we show how this relation between perturbative corrections to the ground-state energy and (integrals of) auto-correlation functions of the local energy along the random walk can be generalized to arbitrary order. The resulting perturbation series can be effectively summed to infinite order by an appropriate re-sampling of the random walks generated by the VMC Langevin equation, Eqs. (1,2,37).

3.1 Stochastic Perturbation Theory

Given the Hamiltonian H , Eq. (36), and a trial wave-function, Φ_0 , which we suppose to be non-orthogonal to its ground state, Ψ_0 , the ground-state energy of H can be expressed as:

$$E_0 = \lim_{\tau \rightarrow \infty} \frac{\langle \Phi_0 | H e^{-H\tau} | \Phi_0 \rangle}{\langle \Phi_0 | e^{-H\tau} | \Phi_0 \rangle} = - \lim_{\tau \rightarrow \infty} \frac{d}{d\tau} \log \langle \Phi_0 | e^{-H\tau} | \Phi_0 \rangle \quad (44)$$

$$= - \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \langle \Phi_0 | e^{-H\tau} | \Phi_0 \rangle. \quad (45)$$

These equations are easily demonstrated by expressing Φ_0 as a linear combination of the eigenfunctions of H , $\{\Psi_n\}$: $\Phi_0 = \sum_n c_n \Psi_n$, and by inserting this expression into Eqs. (44,45):

$$\log \langle \Phi_0 | e^{-H\tau} | \Phi_0 \rangle = -E_0 \tau + \log(c_0^2) + \mathcal{O}(e^{-(E_1 - E_0)\tau}). \quad (46)$$

The following basic identity holds:

$$\begin{aligned} \mathcal{Z}_0 &\equiv \langle \Phi_0 | e^{-H\tau} | \Phi_0 \rangle = \int e^{-\mathcal{S}[X]} \mathcal{P}[X] \mathcal{D}[X] \\ &\equiv \langle e^{-\mathcal{S}[X]} \rangle, \end{aligned} \quad (47)$$

where $\mathcal{P}[X]$ is the probability distribution for the random walk X , Eq. (25), $\mathcal{D}[X] = dx_0 \cdots dx_N$, and the action $\mathcal{S}[X]$ is a functional of the random walk which in the continuum limit coincides with the time integral of the local energy:

$$\begin{aligned} \mathcal{S}[X] &= \epsilon \sum_{i=1}^N \mathcal{E}(x_i) + \mathcal{O}(\epsilon) \\ &\approx \int_0^\tau \mathcal{E}(x(\tau')) d\tau'. \end{aligned} \quad (48)$$

Eq. (47) is a generalization of the Feynman-Kac formula¹² and can be demonstrated as follows.

$$\mathcal{Z}_0 = \int \Phi_0(x) G(x, y; \tau) \Phi_0(y) dx dy, \quad (49)$$

where

$$G(x, y; \tau) = \langle x | e^{-H\tau} | y \rangle \quad (50)$$

is the imaginary-time propagator of the full Hamiltonian. We now break the propagator in Eq. (49) into the product of N short-time propagators:

$$\mathcal{Z}_0 = \underbrace{\int \Phi_0(x_0) G(x_0, x_1; \epsilon) \cdots G(x_{N-1}, x_N; \epsilon) \Phi_0(x_N)}_{\mathbf{P}[X]} \mathcal{D}[X]. \quad (51)$$

Eqs. (47) and (51) can be seen as a *definition* of the action:

$$\mathbf{P}[X] = \mathcal{P}[X]e^{-\mathcal{S}[X]}. \quad (52)$$

By using the Trotter formula,

$$G(x, y; \epsilon) = \mathcal{G}(x, y; \epsilon)e^{-\epsilon \mathcal{E}(y)} + \mathcal{O}(\epsilon^2), \quad (53)$$

one sees that the action defined by Eq. (52) can indeed be expressed by the time integral given by Eq. (48).

\mathcal{Z}_0 plays the role of a pseudo partition function, in the sense that the energy of the system – as well as other observables, as we will see – can be calculated by differentiating it much in the same way as one would do in classical statistical mechanics. In particular, the two expression (44) and (45) for the ground-state energy correspond to the zero-temperature limits of the internal and free energies, respectively. By inserting Eq. (47) into Eq. (45), one could derive a systematic perturbative expansion of the ground-state energy in powers of $\mathcal{E}(x) - \bar{\mathcal{E}}$. Each term of the series is basically a cumulant of the action, which in turn can be seen as the integral of a suitably defined *connected* time correlation function of the local energy, calculated along the Langevin random walk. This kind of *stochastic* perturbation theory can be effectively carried on to infinite order by inserting Eq. (47) into Eq. (44). The final result reads:

$$\begin{aligned} E_0 &= \lim_{\tau \rightarrow \infty} \frac{\langle \mathcal{E}(x(\tau)) e^{-\mathcal{S}[X]} \rangle}{\langle e^{-\mathcal{S}[X]} \rangle} \\ &\equiv \lim_{\tau \rightarrow \infty} \langle\langle \mathcal{E}(x(\tau)) \rangle\rangle, \end{aligned} \quad (54)$$

where the double bracket $\langle\langle \cdot \rangle\rangle$ indicates that the average over the random walks (*quantum paths*) is re-weighted by the exponential of the action, Eq. (48).

Eq. (54) can be turned into an algorithm for calculating the ground-state energy. The calculation would proceed as in a standard VMC simulation, with the difference that the local energy must be weighted with the exponential of the action, $e^{-\mathcal{S}}$, calculated along a segment of the random walk long τ and ending at the time when the measure is taken. This algorithm, which was first proposed in Ref.,⁴ is bound to fail in all those case where the number of particles is so large or the quality of the trial wave-function is so poor that the fluctuations of the action make the weighting procedure impractical. As a matter of fact, the *pure-diffusion Monte Carlo* (PDMC) of Ref.⁴ has never been applied but to very simple quantum systems. In the next session we will show how the fluctuations of the action can be effectively dealt with through a new algorithm based on importance sampling. Before doing this, we want to show how the ideas exposed so far can be exploited to calculate general ground-state expectation values and response functions.

3.2 Calculation of Observables

Other local observables can be calculated along similar lines starting from the Hellmann-Feynman theorem:¹³

$$\langle \Psi_0 | \mathcal{A} | \Psi_0 \rangle = \left. \frac{dE_\lambda}{d\lambda} \right|_{\lambda=0}, \quad (55)$$

where E_λ is the ground-state energy of a system whose Hamiltonian is

$$H_\lambda = H + \lambda \mathcal{A}. \quad (56)$$

By using Eq. (45), E_λ can be put in the form:

$$E_\lambda = - \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \left\langle e^{-\mathcal{S}_\lambda[X]} \right\rangle, \quad (57)$$

where

$$\mathcal{S}_\lambda[X] = \int_0^\tau (\mathcal{E}(x(\tau')) + \lambda \mathcal{A}(x(\tau'))) d\tau'. \quad (58)$$

The expectation value of \mathcal{A} can then be expressed as:

$$\begin{aligned} \langle \Psi_0 | \mathcal{A} | \Psi_0 \rangle &= \lim_{\tau \rightarrow \infty} \frac{\left\langle \frac{1}{\tau} \int_0^\tau \mathcal{A}(x(\tau')) d\tau' e^{-\int_0^\tau \mathcal{E}(\tau') d\tau'} \right\rangle}{\left\langle e^{-\int_0^\tau \mathcal{E}(\tau') d\tau'} \right\rangle} \\ &\equiv \lim_{\tau \rightarrow \infty} \left\langle \left\langle \frac{1}{\tau} \int_0^\tau \mathcal{A}(x(\tau')) d\tau' \right\rangle \right\rangle \end{aligned} \quad (59)$$

3.3 Response Functions

A simple extension of the ideas used in the previous section leads to a technique for evaluating response functions. Let us suppose that the system is coupled to a set of external variables, $\{\lambda_i\}$, through the local operators $\{\mathcal{A}_i\}$:

$$H_{\{\lambda\}} = H + \sum_i \lambda_i \mathcal{A}_i. \quad (60)$$

In the case of an external potential, for instance, the index i labels the coordinate, x , λ_i is the potential itself, $V_{ext}(x)$, and \mathcal{A}_i is the particle-density operator, $n(x)$. We define a generalized susceptibility, χ , as the derivative of the expectation value of one of the \mathcal{A} operators with respect to one of the λ 's:

$$\begin{aligned} \chi_{ij} &= \frac{\partial \langle \mathcal{A}_i \rangle}{\partial \lambda_j} \\ &= \frac{\partial^2 E_\lambda}{\partial \lambda_i \partial \lambda_j}. \end{aligned} \quad (61)$$

By using Eq. (57), χ can be put in the form:

$$\begin{aligned} \chi_{ij} &= - \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \left[\left\langle \left\langle \left(\int_0^\tau \mathcal{A}_i(\tau') d\tau' \int_0^\tau \mathcal{A}_j(\tau') d\tau' \right) \right\rangle \right\rangle \right. \\ &\quad \left. - \left\langle \left\langle \int_0^\tau \mathcal{A}(\tau') d\tau' \right\rangle \right\rangle^2 \right] \\ &= - \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \left\langle \left\langle \int_0^\tau d\tau_1 \int_0^\tau d\tau_2 (\mathcal{A}_i(\tau_1) \mathcal{A}_j(\tau_2) - \bar{\mathcal{A}}_i \bar{\mathcal{A}}_j) \right\rangle \right\rangle, \end{aligned} \quad (62)$$

where $\bar{\mathcal{A}}$ is the time average of \mathcal{A} over the random walk: $\bar{\mathcal{A}} = \frac{1}{\tau} \int_0^\tau \mathcal{A}(\tau') d\tau'$. We now split the domain of integration $[0 \leq \tau_1 \leq \tau; 0 \leq \tau_2 \leq \tau]$ into two sub-domains $[\tau_1 < \tau_2]$

and $[\tau_2 < \tau_1]$, and change the variables of integration $\{\tau_1, \tau_2\} \rightarrow \{\tau_1, \tau_2 - \tau_1\}$ and $\{\tau_1, \tau_2\} \rightarrow \{\tau_1 - \tau_2, \tau_2\}$ in the two sub-domains respectively. In the large τ limit, the susceptibility can then be cast into the form:

$$\chi_{ij} = - \left\langle \left\langle \int_0^\tau [c_{ij}(\tau') + c_{ji}(\tau')] d\tau' \right\rangle \right\rangle, \quad (63)$$

where the time auto-correlation function is defined by:

$$c_{ij}(\tau) = \frac{1}{\tau} \int_0^\tau \mathcal{A}_i(\tau') \mathcal{A}_j(\tau') d\tau' - \frac{1}{\tau^2} \int_0^\tau \mathcal{A}_i(\tau') d\tau' \int_0^\tau \mathcal{A}_j(\tau') d\tau'. \quad (64)$$

The symmetrized time correlation function $c_{ij} + c_{ji}$ is a functional of the path, whose time integral (*i.e.* whose $\omega = 0$ Fourier component) provides an estimator for the static response function. This argument can be easily generalized to show that other ($\omega \neq 0$) Fourier components of the same time correlation function provide suitable estimators for the dynamic susceptibility.

4 The Algorithm

In the previous section we have shown that the calculation of ground-state expectation values and response functions can be reduced to the weighted average of suitable estimators over the space of random walks. A straightforward evaluation of these averages over a Langevin trajectory would be very inefficient because the weight, $e^{-\mathcal{S}}$, may vary a lot. This problem can be solved by converting the *weighting* of the quantum paths into a *re-sampling* of them, through an appropriate Metropolis algorithm.¹⁴ As a by-product, the use of the Metropolis algorithm allows one to reduce the systematic errors due to the discretization of time in an efficient and convenient way.

Let $G^{(n)}$ be any approximation to the full imaginary-time propagator, Eq. (50), correct to order n in ϵ , and $\mathbf{P}^{(n)}[X]$ the corresponding approximation to $\mathbf{P}[X]$, Eq. (51), which will be affected by errors of order n . Let us also define the corresponding approximations for the unperturbed propagator, $\mathcal{G}^{(n)}$, Eq. (24), path probability distribution, $\mathcal{P}^{(n)}$, and time-discretization correction factor, $\mathcal{Q}_\epsilon^{(n)}$, Eq. (25), and action, $\mathcal{S}^{(n)}[X]$, Eq. (52). For instance, one can take:

$$\begin{aligned} G^{(2)}(x, y; \epsilon) &= \frac{1}{\sqrt{4\pi\epsilon}} e^{-\frac{(x-y)^2}{4\epsilon} - \frac{\epsilon}{2}(V(x) + V(y))} = G(x, y; \epsilon) + \mathcal{O}(\epsilon^3), \\ \mathcal{G}^{(2)}(x, y; \epsilon) &= \frac{1}{\sqrt{4\pi\epsilon}} e^{-\frac{(x-y)^2}{4\epsilon} - \frac{\epsilon}{2}\left(\frac{\Phi_0''(x)}{\Phi_0(x)} + \frac{\Phi_0''(y)}{\Phi_0(y)}\right)} = \mathcal{G}(x, y; \epsilon) + \mathcal{O}(\epsilon^3), \\ \mathcal{P}^{(2)}[X] &= \Phi_0(x_N) \mathcal{G}^{(2)}(x_N, x_{N-1}; \epsilon) \cdots \mathcal{G}^{(2)}(x_1, x_0; \epsilon) \Phi_0(x_0) \\ &= \mathcal{P}[X] + \mathcal{O}(\epsilon^2), \\ \mathcal{S}^{(2)}[X] &= e^{-\frac{\epsilon}{2} \sum_{i=1}^N (\mathcal{E}(x_i) + \mathcal{E}(x_{i-1}))} = \mathcal{S}[X] + \mathcal{O}(\epsilon^2). \end{aligned} \quad (65)$$

Any ground-state expectation value or response function can be put in the form:

$$\langle\langle \mathbf{F}[X] \rangle\rangle = \frac{\int \mathbf{P}^{(n)}[X] \mathbf{F}[X] \mathcal{D}[X]}{\int \mathbf{P}^{(n)}[X] \mathcal{D}[X]} + \mathcal{O}(\epsilon^n). \quad (66)$$

In order to calculate the above expectation value through the Metropolis method, it is necessary to construct a Markov chain of random walks so that the corresponding transition probability, $\mathbf{W}[Y \leftarrow X]$, satisfies detailed balance:

$$\mathbf{W}[Y \leftarrow X] \times \mathbf{P}^{(n)}[X] = \mathbf{W}[X \leftarrow Y] \times \mathbf{P}^{(n)}[Y] \quad (67)$$

In the Metropolis algorithm, the transition probability \mathbf{W} is split into the product of an *a-priori* transition probability, \mathbf{W}^0 , times a probability, \mathbf{A} , that a move *proposed* according to \mathbf{W}^0 is accepted: $\mathbf{W}[Y \leftarrow X] = \mathbf{W}^0[Y \leftarrow X] \times \mathbf{A}[Y \leftarrow X]$. Detailed balance, Eq. (67), can be satisfied by choosing:¹⁵

$$\mathbf{A}[Y \leftarrow X] = \min(1, \mathbf{R}[Y \leftarrow X]), \quad (68)$$

where

$$\mathbf{R}[Y \leftarrow X] = \frac{\mathbf{W}^0[X \leftarrow Y] \times \mathbf{P}^{(n)}[Y]}{\mathbf{W}^0[Y \leftarrow X] \times \mathbf{P}^{(n)}[X]}. \quad (69)$$

Given a quantum path, $X \equiv \{x_0, \dots, x_N\}$, we propose a new path, Y , by propagating the random walk forward by $M < N$ steps, according to the VMC Langevin equation, Eqs. (2,37): $Y \equiv \{x_M, \dots, x_N, \dots, x_{M+N}\}$. The corresponding a-priori transition probability is:

$$\begin{aligned} \mathbf{W}^0[Y \leftarrow X] &= \mathcal{W}_\epsilon(x_{N+M}|x_{N+M-1}) \cdots \mathcal{W}_\epsilon(x_{N+1}|x_N) \\ &= \mathcal{P}^{(n)}[\{x_N, \dots, x_{N+M}\}] \mathcal{Q}_\epsilon^{(n)}[\{x_N, \dots, x_{N+M}\}] / P_s(x_N), \end{aligned} \quad (70)$$

where we have used Eq. (25). By inserting this expression for the a-priori probability into Eq. (69), the latter can be cast into the form:

$$\begin{aligned} \mathbf{R}[Y \leftarrow X] &= \frac{\mathcal{P}^{(n)}[\{x_M, \dots, x_0\}] \times \mathcal{Q}_\epsilon^{(n)}[\{x_M, \dots, x_0\}] / P_s(x_M)}{\mathcal{P}^{(n)}[\{x_N, \dots, x_{N+M}\}] \times \mathcal{Q}_\epsilon^{(n)}[\{x_N, \dots, x_{N+M}\}] / P_s(x_N)} \\ &\quad \times \mathcal{P}^{(n)}[\{x_0, \dots, x_N\}] \times \frac{e^{-\mathcal{S}^{(n)}[X]}}{\mathcal{P}^{(n)}[\{x_M, \dots, x_{N+M}\}] \times e^{-\mathcal{S}^{(n)}[Y]}}. \end{aligned} \quad (71)$$

With the aid of the composition law for the random-walk probability distributions, Eq. (26), and of the time-reversal property, Eq. (27), the above equation finally reads:

$$\mathbf{R}[Y \leftarrow X] = e^{-(\mathcal{S}^{(n)}[Y] - \mathcal{S}^{(n)}[X])} \times \frac{\mathcal{Q}_\epsilon^{(n)}[\{x_M, \dots, x_0\}]}{\mathcal{Q}_\epsilon^{(n)}[\{x_N, \dots, x_{N+M}\}]} \quad (72)$$

Notice that the ratio of the \mathcal{Q} 's goes to one in the continuous-time limit ($\epsilon \rightarrow 0$) and that for finite ϵ it can be explicitly calculated, providing thus a systematic way for reducing to any desired order the time-discretization errors.

The dynamical variables of our simulations are quantum paths, which can be formally associated with polymers, much in the same spirit as this is done in path-integral simulations. The polymer dynamics which would correspond to our algorithm is known in the literature as *reptation*.¹⁶ For this reason, we name our algorithm *Reptation Quantum Monte Carlo* (RQMC), and we refer to each individual quantum path as to a *reptile*.

The practical implementation of RQMC is extremely simple, at the level of a VMC simulation. The algorithm can be summarized as follows:

1. Using Eqs. (2) and (37), generate a reptile long enough that its end point, $x(T)$, is distributed according to the square of the trial wave function, Φ_0^2 .
2. Generate a further segment of the reptile corresponding to the time interval $[T, T + \tau]$. $\tau = N\epsilon$ should be large enough that the limit $\tau \rightarrow \infty$ in Eqs. (54), (59), and (62) is reached to the desired accuracy. Set $X \equiv \{x_0, x_1, \dots, x_N\} \leftarrow \{x(T), x(T + \epsilon), \dots, x(T + \tau)\}$.
3. Select a ‘direction of time’ (*forward* or *backward*) with equal probability. If the choice is *backward*, set $X \leftarrow \bar{X} = \{x_N, x_{N-1}, \dots, x_0\}$. This step is necessary to ensure the micro-reversibility of the algorithm.
4. Generate a segment of the reptile corresponding to the time interval $[T + \tau, T + \tau + \delta]$ and set $Y \equiv \{x_M, x_{M+1}, \dots, x_{N+M}\} \leftarrow \{x(\delta), x(T + \delta + \epsilon), \dots, x(T + \tau + \delta)\}$. The value of δ is sampled from an uniform deviate in the interval $[0, \Delta]$ whose width is chosen so as to minimize the auto-correlation times of the measured quantities. Sampling δ instead of keeping it constant helps to avoid that the reptile remains occasionally stuck at a fixed position for a long time.
5. Evaluate $\mathbf{R}[Y \leftarrow X]$ according to Eq. (72).
6. If $\mathbf{R} > 1$, set $X \leftarrow Y$. If $\mathbf{R} < 1$, set $X \leftarrow Y$ with probability \mathbf{R} .
7. Accumulate the ground-state energy and other observables using appropriate estimators (see the next section for an optimal choice of the estimators).
8. Go to 3.

A preliminary test of this algorithm has been performed for the hydrogen atom using an approximate trial function. Exact results for the average of several moments of electron–nucleus distance have been reproduced within a statistical error pushed down to a small fraction of the difference between the exact value and the extrapolated estimate (i.e. twice the mixed average minus the variational average¹⁷).

5 A Case Study of ${}^4\text{He}$

We now discuss the calculation of several properties of superfluid ${}^4\text{He}$, with the purpose of showing that the method can be successfully applied to systems of actual physical interest. Based on the limited experience gained in this case study, we also present some performance comparisons with related techniques.

5.1 Details of the Simulation

We consider $N_P = 64$ ${}^4\text{He}$ atoms interacting through a pair potential, as obtained from first-principles calculations.¹⁸ The simulation was performed in a cubic box with periodic boundary conditions at the experimental equilibrium density, $\rho = 0.02186 \text{ \AA}^{-3}$. The trial function Φ_0 includes pair and nearly optimal three–body correlations.¹⁹ This is a relatively good trial function. The variational energy is less than 0.3 K above the exact ground-state

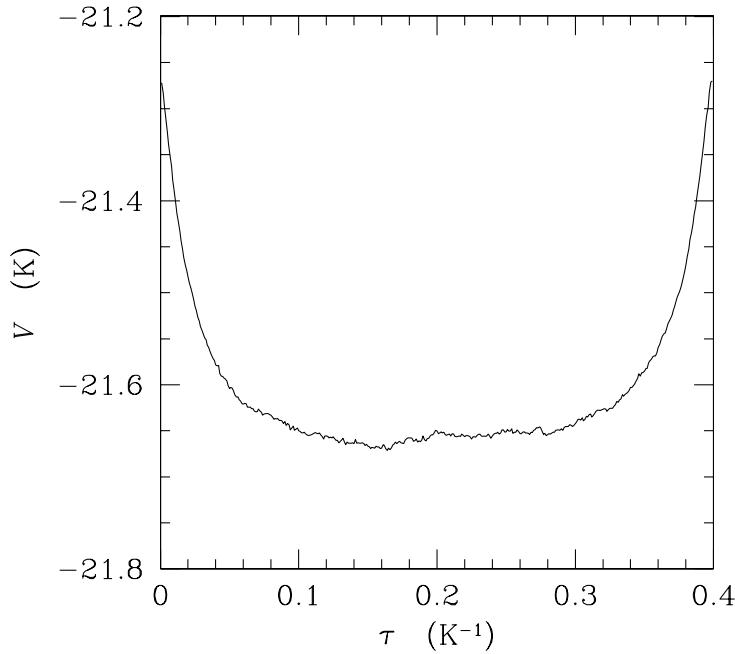


Figure 1. Average of the potential energy in ${}^4\text{He}$, calculated on individual time slices along the path. The statistical error on the central slices is $\simeq 0.03$ K. This result was obtained using a trial function with pair correlations only: note that V converges to the same value given in Table 1, obtained using a trial function with pair and triplet correlations.

energy, whereas the variational bias is larger than 1.1 K using pair correlations only. The quantities we compute are total and potential energies, the imaginary-time correlations of the density fluctuation operators, $\rho_q = \sum_i \exp(-i\mathbf{q} \cdot \mathbf{r}_i)$:

$$F(q, \tau) = \langle \rho_q(\tau) \rho_{-q}^\dagger(0) \rangle / N_P, \quad (73)$$

and the diffusion coefficient of the center of mass motion,

$$D(\tau) = \langle [\mathbf{r}_{CM}(\tau) - \mathbf{r}_{CM}(0)]^2 \rangle N_P / (6\tau). \quad (74)$$

The parameters of our simulations are as follows. The time step is $\epsilon = 0.001 \text{ K}^{-1}$, which gives a systematic bias of the order of 10^{-2} K on the total energy. For the calculation of total and potential energy the length of the path was $\tau = 0.4 \text{ K}^{-1}$, corresponding to $N = 400$ time slices. The calculation of imaginary-time correlations over a significant range required the use of longer paths, up to $N = 700$. The value of the energy resulting from the simulation with such longer paths confirmed that the finite- τ bias in the results obtained with $N = 400$ was smaller than statistical errors. Note that the length of the path adversely affects the efficiency, because the relaxation time of the polymer in the reptation algorithm is proportional to N^2 .²⁰ The number of time slices of each reptation move is uniformly sampled between 0 and 20, yielding an acceptance ratio of $\approx 80\%$.

Table 1. Ground-state energy, E_0 , and potential energy, V , in ${}^4\text{He}$, as computed from RQMC and traditional diffusion Monte Carlo runs of 3×10^6 Monte Carlo steps. Units are K. The length of the path in the RQMC calculation is $\tau = 0.4 \text{ K}^{-1}$, and the length of the forward walk for V in the diffusion Monte Carlo calculation is 0.2 K^{-1} .

	E_0	V
RQMC	-7.4066(27)	-21.644(15)
BDMC	-7.3902(15)	-21.674(21)

Fluctuations in the average of the total energy are reduced using a symmetrized form of Eq. (54), *i.e.* accumulating $[\mathcal{E}(x(\tau)) + \mathcal{E}(x(0))] / 2$. Expectation values of local observables are computed averaging time integrals along the path, Eq. (59) and Eq. (62). Although these expressions are exact in the $\tau \rightarrow \infty$ limit, the contributions coming from the extrema of the path are clearly biased. For instance, the average of \mathcal{A} in the initial or final time slices ($\langle \mathcal{A}(x(0)) \rangle$ and $\langle \mathcal{A}(x(\tau)) \rangle$) yields the mixed estimate:

$$\lim_{\tau \rightarrow \infty} \frac{\left\langle \mathcal{A}(x(0)) e^{- \int_0^\tau \mathcal{E}(\tau') d\tau'} \right\rangle}{\left\langle e^{- \int_0^\tau \mathcal{E}(\tau') d\tau'} \right\rangle} = \lim_{\tau \rightarrow \infty} \frac{\langle \Phi_0 | \mathcal{A} e^{-H\tau} | \Phi_0 \rangle}{\langle \Phi_0 | e^{-H\tau} | \Phi_0 \rangle} = \frac{\langle \Phi_0 | \mathcal{A} | \Psi_0 \rangle}{\langle \Phi_0 | \mathcal{A} | \Psi_0 \rangle}. \quad (75)$$

On the other hand any time slice a distance $\bar{\tau}$ apart from the extrema, such that $\exp(-H\bar{\tau})|\Phi_0\rangle \simeq |\Psi_0\rangle$, gives an unbiased contribution to the time integral of \mathcal{A} . Therefore it is convenient to restrict the time integral of \mathcal{A} in Eqs. (59) and (62), to the inner section of the path, where the bias is reduced. For the potential energy and the imaginary-time correlations we exclude the contributions from 150 time slices on each side of the path. This is a rough estimate of the time it takes for the average potential energy to converge within a few hundredths K from its value at slice 0 or N_τ (which is by Eq. (75) the mixed estimate) to the unbiased estimate. This is demonstrated by the average of the potential energy on individual time slices, shown in Fig. 1. We have not studied the corresponding convergence times for the density–density correlations.

5.2 Results

Our results for the total and potential energies are listed in Table 1. The inter-particle potential adopted¹⁸ overestimates the experimental binding energy of -7.17 K because of the neglect of three-body forces (mostly triple-dipole repulsion). Also reported in Table 1 are the corresponding data obtained from a standard Branching Diffusion Monte Carlo (BDMC) calculation using the same time step and trial function. The small differences between the results of the two algorithms have to be attributed to different time-step biases.

In Fig. 2 we show the density-density correlation function, $F(q, \tau)$, for a few values of q , as obtained from a run of 10^7 MC steps, which required about one week CPU time on a workstation. Note that at virtually no additional cost many more q vectors, belonging to the reciprocal lattice of the simulation box, could have been included in the calculation.

$F(q, \tau)$ is related to several quantities of physical interest,²¹ including the static structure factor,

$$S(q) = F(q, 0) = \int_0^\infty d\omega S(q, \omega), \quad (76)$$

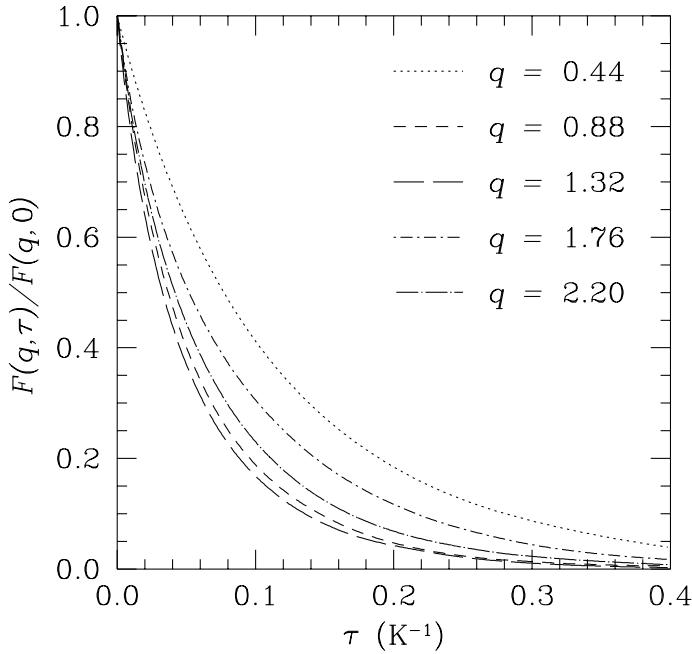


Figure 2. Imaginary-time correlations of the density fluctuation operator in ${}^4\text{He}$. Averages are taken on the inner part of a path of length 0.7 K^{-1} . The statistical error ranges from approximately 0.5% at $\tau = 0$ up to 5% at $\tau = 0.4 \text{ K}^{-1}$.

the static linear response function,

$$\chi(q) = -2 \int_0^\infty d\tau F(q, \tau) = -2 \int_0^\infty d\omega S(q, \omega)/\omega, \quad (77)$$

and the dynamical structure factor,

$$F(q, \tau) = \int_0^\infty d\omega e^{-\omega\tau} S(q, \omega). \quad (78)$$

Low moments of $S(q, \omega)$ can be accurately extracted from the simulation data for $F(q, \tau)$. The static structure factor, Fig. 3, and the static response, Fig. 4, compare very favorably with the experimental results, the discrepancy visible in $S(q)$ at the smallest value of q being due to the finite temperature at which the measurements are performed. The f -sum rule, $\partial F(q, \tau)/\partial\tau|_{\tau=0} = \int d\omega S(q, \omega)\omega = q^2$, is also verified with high precision.

Inferring dynamical properties from imaginary-time correlations, on the other hand, is much harder. Since the inverse Laplace transform (78) with incomplete and noisy data for $F(q, \tau)$ is an ill-conditioned problem,²⁴ a least χ^2 approach to pin down a parametrized form for $S(q, \omega)$ is doomed to failure. Additional constraints can be set on the solution $S(q, \omega)$ by resorting to Maximum Entropy (ME) methods.²⁴ We follow the implementation used in Ref.²⁵ to process data for $F(q, \tau)$ obtained at finite temperature with a PIMC

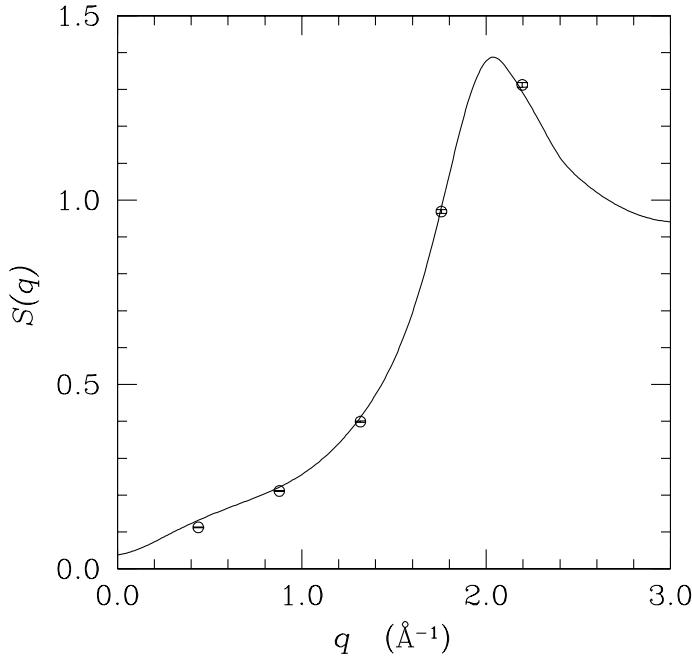


Figure 3. Static structure factor in ${}^4\text{He}$ at the five wave vectors listed in Fig. 2 (open circles). The solid line is the experimental $S(q)$ measured by neutron scattering.²²

simulation. The results are qualitatively similar to those obtained in Ref.²⁵ The ME reconstruction of $S(q, \omega)$, shown in Fig. 5, is too smooth and does not reproduce the sharp features exhibited by the experimental structure factor. Furthermore, the relatively poor quality of the available Monte Carlo data does not allow for a reliable estimate of the statistical uncertainty on the results.²⁴ Nevertheless we do recover some useful information on dynamical properties: the presence of a gap in the excitation spectrum is unambiguously revealed, and the position of the peak of the reconstructed dynamical response closely follows the experimental dispersion of the elementary excitations (see the inset of Fig. 5).

We now outline the calculation of the superfluid density ρ_s . Although the value of ρ_s/ρ is trivially one for pure bulk ${}^4\text{He}$ in the ground state, interacting Bose systems in the presence of an external disordered potential undergo a zero temperature superfluid–insulator transition as the strength of the potential increases.²⁷ We thus consider a model system of static impurities in ${}^4\text{He}$. The external potential V_{ext} is represented by attractive Gaussians, $V_{ext}(\mathbf{r}) = \sum_j A \exp[-\alpha(\mathbf{r} - \mathbf{R}_j)^2]$, where the positions \mathbf{R}_j of the impurities are placed randomly in the simulation box, $A = -50$ K and $\alpha = 0.5 \text{ \AA}^{-2}$. The trial function is multiplied by a one-body factor $\exp[-\sum_{i,j} f(|\mathbf{r}_i - \mathbf{R}_j|)]$ which tends to localize the ${}^4\text{He}$ atoms around the impurities. No average over different realizations of disorder was performed.

In a finite-temperature calculation with periodic boundary conditions, ρ_s can be estimated^{28,29} as $\rho_s/\rho = \langle w^2 \rangle / (6\tau N_P)$, where the winding number

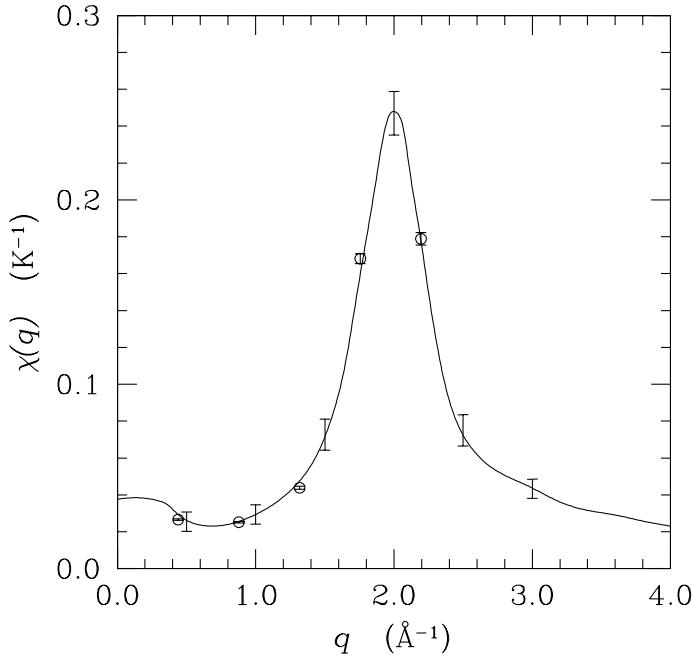


Figure 4. Static response function of ${}^4\text{He}$ at the five wave vectors listed in Fig. 2 (open circles). The solid line is the experimental result of Ref.²³

$\mathbf{w} = \sum_{i=1}^{N_P} \int_0^\tau d\tau' \left[\frac{d\mathbf{r}_i(\tau')}{d\tau'} \right]$ is the displacement of the center of mass of the system, and τ is the inverse temperature. By taking the limit $\tau \rightarrow \infty$, appropriate for a ground-state calculation, the superfluid density at $T = 0$ can be expressed in terms of the diffusion coefficient of the center of mass motion, $\rho_s/\rho = \lim_{\tau \rightarrow \infty} D(\tau)$, where D is defined in Eq. (74).

The results reported in Fig. 6 show that the superfluid fraction, which is correctly one for the pure system, is indeed reduced in the presence of the impurities. Longer simulations would be needed to relate the depletion of the superfluid fraction induced by the disordered potential to changes of the excitation spectrum.

5.3 Comparison With Other Methods

Reptation quantum Monte Carlo utilizes the Metropolis algorithm to sample an explicitly known probability distribution, namely a discretized path integral expansion of $\exp(-\tau H)\Phi_0$ which becomes exact in the limit $\tau \rightarrow \infty$ and $\epsilon \rightarrow 0$. Sampling a distribution, as opposed to carrying weights, avoids the difficulties associated with fluctuating weights which plague applications of ‘pure diffusion’⁴ or ‘single thread’³⁰ Monte Carlo. For instance, we were unable to get converged results for our 64 particle system with using pure-diffusion Monte Carlo.⁴

The idea of sampling a path-integral representation of the imaginary-time evolution

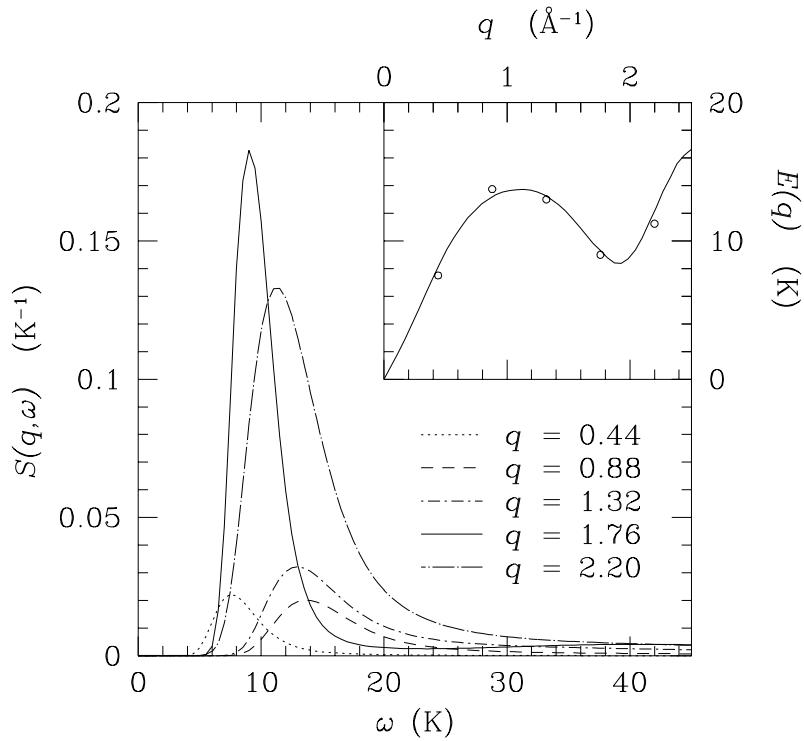


Figure 5. Maximum Entropy reconstruction of the dynamical structure factor of ^4He . In the inset the position of the maxima of the calculated $S(q, \omega)$ is compared with the experimental excitation spectrum.²⁶

to compute ground-state properties is not new.³¹ For instance, Variational Path Integral (VPI)²⁸ uses the pair product approximation to expand the many-body density matrix $\exp(-\tau H)$ and the bisection method to sample the path, just like in the usual Path Integral Monte Carlo method.²⁹

Reptation quantum Monte Carlo features instead an expansion of the imaginary-time propagator based on the Langevin dynamics generated by the trial function, and a reptation algorithm to sample the exponential of the resulting action. From the computational point of view, the advantage of this particular choice can be understood in the limit of perfect importance sampling: if the trial function is exact the local energy is a constant, and reptation moves consisting of an arbitrary number of time slices will be accepted with probability 1. Eventually, a very poor wave function (or equivalently a very large number of particles) will force us to take extremely small reptation moves, and moves of the kind used in VPI will become more efficient.²⁹ Variational path integral has been implemented³² for the simulation of superfluid ^4He at $T = 0$. According to the author of the VPI calculation, for the system size considered here RQMC is considerably more efficient.

Sampling an explicitly known distribution is to be contrasted to the standard branching diffusion Monte Carlo, which samples an unknown distribution (*i.e.* the mixed distribution

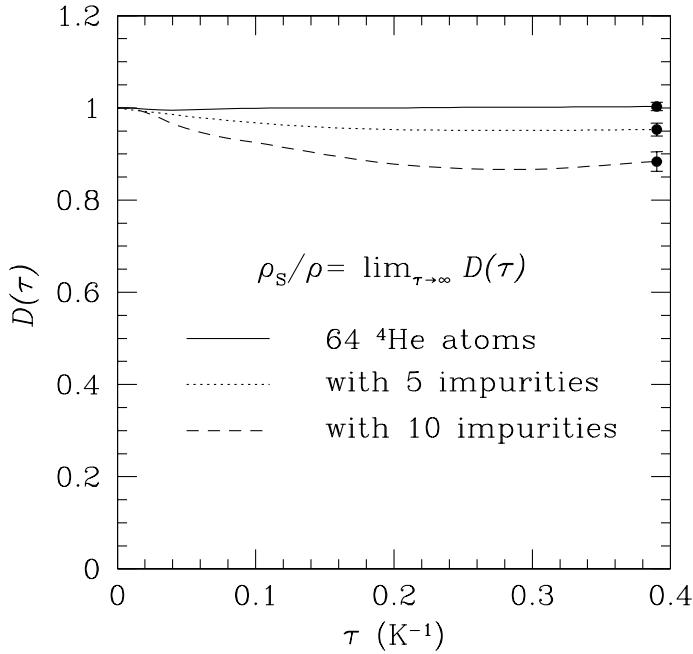


Figure 6. Diffusion coefficient of the center of mass motion of ${}^4\text{He}$ in the presence of quenched disorder.

$\Psi_0 \Phi_0$) obtained from the asymptotic solution of a differential equation.^{33,34} BDMC is designed to compute efficiently the total energy; however it introduces a population-control bias, yields a mixed estimate for operators not commuting with H , and does not retain direct information on the imaginary-time correlations. This information can be retrieved through the ‘forward walking’ technique^{34–36} and used to correct both the population control and the mixed estimate biases, but at the price of additional statistical fluctuations. We believe that reptation quantum Monte Carlo will turn out to be advantageous over branching diffusion Monte Carlo in those cases where recovering information from fluctuating weights becomes too noisy.

The results for the total energy shown in Table 1 are obtained from reptation quantum Monte Carlo and branching diffusion Monte Carlo using the same time step, number of particles, and trial function. From the estimated statistical error we infer that BDMC is roughly 3 times faster than RQMC for the calculation of the total energy. Also listed in Table 1 are the unbiased estimates for the potential energy. In this case branching diffusion Monte Carlo with forward walking (implemented in the “backward storing mode” described in Ref.³⁴) turns out to be roughly two times slower than reptation quantum Monte Carlo. Obviously, factors 2 or 3 for a couple of observables in a particular physical system are not a conclusive assessment of the relative performance of two algorithms. However the resulting factor 6 in the relative improvement of RQMC when BDMC has to be complemented with forward walking suggests that, whenever explicit information on imaginary-time correlations is used, reptation quantum Monte Carlo is likely to be com-

petitive or better.

6 Conclusions

The most attractive feature of the reptation quantum Monte Carlo is that it uses the dynamical properties of the random walk in a way that is directly related to the imaginary-time properties of the physical system. A previous implementation of similar ideas⁴ was based on re-weighting and the scope of its applications was severely restricted by the fluctuations of the weights. We have shown that, by simply complementing the method of Ref.⁴ with a re-sampling of the paths based on the value of the action, the resulting RQMC algorithm can afford system sizes typical of current branching diffusion Monte Carlo simulations of continuous strongly interacting systems. Unbiased estimates, static responses and some insight into dynamical properties can be readily obtained. The dependence of the computational effort on the number of particles and on the quality of the trial function remains to be investigated. Such an analysis will determine whether this method can be useful in more general situations than ^4He bulk liquid. Clusters, films and superfluids in restricted geometries are natural candidates for further applications.

For Fermion problems one has either to resort to the fixed-node approximation,³³ or to cope with the sign problem.³⁷ In the former case, the dynamical information contained in the path is incorrect,⁴ but still the explicit knowledge of the weights along the path makes the algorithm free from the mixed distribution and the population control biases; furthermore it gives easily access to interesting quantities, such as a low-variance estimator of Born–Oppenheimer forces in electronic systems.³⁸ In the latter case the dynamics is correct, and can be used for example to get information on the ground and excited states from the imaginary-time evolution in the transient regime;³⁹ in similar cases however the real bottleneck will remain the sign problem.

Reptation quantum Monte Carlo, variational path integral,²⁸ and the technique discussed in Ref.³¹ sample explicit expressions of the imaginary-time propagator with the Metropolis algorithm to calculate zero temperature properties. We believe that these methods are very promising and deserve more attention than they have received so far. All these methods are based on the Metropolis algorithm: they enjoy therefore of a large freedom in the choice of the transition probability,⁴⁰ which is probably not yet fully exploited.

Acknowledgments

We are indebted with Kevin Schmidt for useful discussions and for communicating to us unpublished details of his VPI calculations on ^4He . We are grateful to Matteo Calandra for lending himself to be the first reader of these lecture notes and to Peter Nightingale for his inspired and careful proof-reading. Thanks to their comments and suggestions, these notes are now considerably better than they were before.

References

1. E. Nelson, Phys. Rev. **150**, 1079 (1966); F. Guerra and P. Ruggiero, Phys. Rev. Lett. **31**, 1022 (1973).
2. G. Parisi and Wu Yongshi, Sci. Sin. **24**, 483 (1981). See also: *Stochastic Quantization*, edited by P.H. Damgaard and H. Hüffel (World Scientific, Singapore, 1988).
3. See, e.g.: *Monte Carlo Methods in Quantum Problems*, edited by M.H. Kalos, NATO ASI series C v. 125 (Reidel, Dordrecht 1984); *Proceedings of the Conference on Frontiers of Quantum Monte Carlo, LANL, September 1985*, edited by J.E. Gubernatis, J. Stat. Phys. **43**, 729–1244 (1986); *Monte Carlo Methods in Theoretical Physics* edited by S. Caracciolo and A. Fabrocini, (ETS, Pisa 1991); *Quantum Monte Carlo Methods in Condensed Matter Physics*, edited by M. Suzuki, (World Scientific, Singapore 1993).
4. M. Caffarel and P. Claverie, J. Chem. Phys. **88**, 1088 (1988); *ibid.* p. 1100.
5. I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. (Academic Press, San Diego, 1994), p. 382 # 3.462–4.
6. For a general textbook on the theory of stochastic processes and its application to the physical sciences, see e.g.: C.W. Gardiner, *Handbook of Stochastic Methods*, 2nd ed. (Springer-Verlag, Berlin, 1985). A nice collection of historically important papers is reprinted in: *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover Publications, New York, 1954).
7. H. Haken, Z. Physik B **24**, 321 (1976).
8. R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw Hill, New York 1965).
9. See e.g. L.D. Landau and E.M. Lifshitz, *Quantum Mechanics* 3rd ed. (Pergamon Press, Oxford, 1977), chapt. III; R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, New York, 1953), vol. I, chapt. VI.
10. This statement is only valid if $v(x) < +\infty$ everywhere inside the domain of x .
11. P.J. Rossky, J.D. Doll, and H.L. Friedman, J. Chem. Phys. **69**, 4628 (1978); the reduction of time-discretization errors in the context of diffusion quantum Monte Carlo is discussed e.g. in: P.J. Reynolds, D. Ceperley, B.J. Alder, and W.A. Lester, J. Chem. Phys. **77** 5593 (1982) and in: C.J. Umrigar, M.P. Nightingale, and K.J. Runge, J. Chem. Phys. **99**, 2865 (1993).
12. M.D. Donsker and M. Kac J. Res. Natl. Bur. Stand. **44**, 581 (1950); R.P. Feynman, *Statistical Mechanics* (Benjamin, Reading, MA, 1972).
13. H. Hellmann, *Einführung in die Quantenchemie* (Deuticke, Leipzig, 1937); R.P. Feynman, Phys. Rev. **56**, 340 (1939).
14. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
15. In the original Metropolis algorithm, the a-priori probability was actually assumed to be symmetric. A generalization to non-symmetric a-priori probabilities was apparently first discussed in: W.K. Hastings, Biometrika **57**, 97 (1970). See also: D. Ceperley, G.V. Chester, and M.H. Kalos, Phys. Rev. B **16**, 3081 (1977) and C.J. Umrigar, Phys. Rev. Lett. **71**, 408 (1993).
16. *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, edited by K. Binder, (Oxford University Press, 1995).

17. D.M. Ceperley and M.H. Kalos, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer-Verlag, 1979).
18. T. Korona, H.L. Williams, R. Bukowski, B. Jeziorski, and K. Szalewicz, *J. Chem. Phys.* **106**, 5109 (1997).
19. S. Moroni, S. Fantoni, and G. Senatore, *Phys. Rev. B* **52**, 13547 (1995).
20. I. Webman, J.L. Lebowitz, and M.H. Kalos, *J. Physique* **41**, 579 (1980).
21. D. Pines and P. Nozières, *Theory of Quantum Liquids* (Benjamin, 1966).
22. E.C. Svensson, V.F. Sears, A.D.B. Woods, and P. Martel, *Phys. Rev. B* **21**, 3638 (1980).
23. A.D.B. Woods and R.A. Cowley, *Rep. Prog. Phys.* **36**, 1135 (1973).
24. J.E. Gubernatis and M. Jarrell, *Phys. Rep.* **269**, 135 (1996).
25. M. Boninsegni and D.M. Ceperley, *J. Low Temp. Phys.* **104**, 339 (1996).
26. R.J. Donnelly, J.A. Donnelly, and R.N. Hills, *J. Low Temp. Phys.* **44**, 471 (1981).
27. M.P.A. Fisher, P.B. Weichman, G. Grinstein, and D.S. Fisher, *Phys. Rev. B* **40**, 546 (1989).
28. D.M. Ceperley, *Rev. Mod. Phys.* **67**, 279 (1995).
29. D.M. Ceperley, in *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by P. Nightingale and C.J. Umrigar. NATO ASI Series, Series C, Mathematical and Physical Sciences, Vol. 525, (Kluwer Academic Publishers, Boston, 1998).
30. P. Nightingale, in *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by P. Nightingale and C.J. Umrigar. NATO ASI Series, Series C, Math. and Physical Sciences, Vol. 525, (Kluwer Academic Publishers, Boston, 1998), sec. III.
31. Ref.,³⁰ sec. IV.
32. K.E. Schmidt, private communication.
33. L. Mitas, in *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by P. Nightingale and C.J. Umrigar. NATO ASI Series, Series C, Mathematical and Physical Sciences, Vol. 525, (Kluwer Academic Publishers, Boston, 1998).
34. Ref.,³⁰ sec. V.
35. K.S. Liu, M.H. Kalos, and G.V. Chester, *Phys. Rev. A* **10**, 303 (1974).
36. P.J. Reynolds, in *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by P. Nightingale and C.J. Umrigar. NATO ASI Series, Series C, Mathematical and Physical Sciences, Vol. 525, (Kluwer Academic Publishers, Boston, 1998).
37. M.H. Kalos, in *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by P. Nightingale and C.J. Umrigar. NATO ASI Series, Series C, Mathematical and Physical Sciences, Vol. 525, (Kluwer Academic Publishers, Boston, 1998).
38. F. Zong and D. M. Ceperley, submitted to *Phys. Rev. E* May 1998.
39. D.M. Ceperley and B. Bernu, *J. Chem. Phys.* **89**, 6316 (1988); B. Bernu, D.M. Ceperley, and W.A. Lester, *J. Chem. Phys.* **93**, 552 (1990). erratum **95**, 7782 (1991); M. Caffarel and D.M. Ceperley, *J. Chem. Phys.* **97**, 8415 (1992).
40. C.J. Umrigar, in *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by P. Nightingale and C.J. Umrigar. NATO ASI Series, Series C, Mathematical and Physical Sciences, Vol. 525, (Kluwer Academic Publishers, Boston, 1998).

Quantum Monte Carlo Methods on Lattices: The Determinantal Approach

Fakher F. Assaad

¹ Institut für Theoretische Physik III, Universität Stuttgart
Pfaffenwaldring 57, 70550 Stuttgart, Germany

² Max Planck Institute for Solid State Research
Heisenbergstr. 1, 70569, Stuttgart, Germany
E-mail: assaad@physik.theo3.uni-stuttgart.de

We present a review of the auxiliary field (i.e. determinantal) Quantum Monte Carlo method applied to various problems of correlated electron systems. The ground state projector method, the finite temperature approach as well as the Hirsch-Fye impurity algorithm are described in details. It is shown how to apply those methods to a variety of models: Hubbard Hamiltonians, periodic Anderson model, Kondo lattice and impurity problems, as well as hard core bosons and the Heisenberg model. An introduction to the world-line method with loop upgrades as well as an appendix on the Monte Carlo method is provided.

1 Introduction

The correlated electron problem remains one of the central challenges in solid state physics. Given the complexity of the problem numerical simulations provide an essential source of information to test ideas and develop intuition. In particular for a given model describing a particular material we would ultimately like to be able to carry out efficient numerical simulations so as to provide *exact* results on thermodynamic, dynamical, transport and ground state properties. If the model shows a continuous quantum phase transition we would like to characterize it by computing the critical exponents. Without restriction on the type of model, this is an extremely challenging goal.

There are however a set of problems for which numerical techniques have and will provide invaluable insight. Here we list a few which are *exact*, capable of reaching large system sizes (the computational effort scales as a power of the volume), and provide ground state, dynamical as well as thermodynamic quantities. i) Density matrix renormalization group applied to general one-dimensional systems¹ ii) world-line loop Quantum Monte Carlo (QMC) applied to non-frustrated spin systems in arbitrary dimensions² and iii) auxiliary field QMC methods.³ The latter method is the only algorithm capable of handling a class of models with spin and charge degrees of freedom in dimensions larger than unity. This class contains fermionic lattice models with an attractive interactions (e.g. attractive Hubbard model), models invariant under a particle-hole transformation, as well as impurity problems modeled by Kondo or Anderson Hamiltonians.

Here we will concentrate primarily on the auxiliary field QMC method and introduce briefly the world line method with loop updates. Both algorithms are based on a path integral formulation of the imaginary time propagator which maps a d -dimensional quantum system on a $d + 1$ -dimensional classical system. The additional dimension is nothing but the imaginary time. For example, within the World Line QMC algorithm,⁴ this mapping

relates the one-dimensional XYZ quantum spin chain to an eight vertex model⁵ or the one-dimensional t - J model to the 15-vertex model.⁶ The classical models may then be solved exactly as in the case of the eight vertex model⁷ or simulated very efficiently by means of cluster Monte Carlo methods.² The latter approach has proved to be extremely efficient for the investigation of non-frustrated quantum spin systems⁸ in arbitrary dimensions. The efficiency lies in the fact that i) the computational time scales as the volume of the $d + 1$ dimensional classical system so that very large system sizes may be achieved and ii) the autocorrelation times are small. In the next section we will briefly, by way of introduction, review the World Line approach and thereby show how the XXZ chain maps onto the 6-vertex model. The attractive feature of the World Line approach is its simplicity. It will also allow us to acquire some insight into the so called sign problem. This is a major, open, issue in QMC methods applied to correlated systems. When it occurs the computational effort scales exponentially with system size and inverse temperature.

In spacial dimensions larger than unity, the World Line approach often fails (i.e. the occurrence of a sign problem) already at the *mean-field* level. That is: consider the paramagnetic mean-field solution of the two dimensional Hubbard model which boils down to solving a free electron problem in a tight binding approximation. This simple model, already leads to a severe sign problem when formulated within the World Line approach. The auxiliary field QMC method³ relies on a different formulation which solves the mean-field problem exactly. With the use of a Hubbard Stratonovich transformation the partition function of a Hamiltonian H at temperature $T = 1/\beta$ and chemical potential μ is written as:

$$Z = \text{Tr} \left[e^{-\beta(H-\mu N)} \right] = \int D\Phi e^{-S(\Phi)}. \quad (1)$$

$S(\Phi)$ is the action of a one-body problem in a imaginary time and space dependent field Φ . As we will see for a given field Φ the computational cost required to compute the action scales as the product of the volume to the cubed and inverse temperature.^a The functional integral is carried by means of Monte Carlo sampling. In this approach the mean-field solution is given by the saddle point approximation: the functional integral over Φ is replaced by a single field Φ^* for which $\frac{\partial S(\Phi)}{\partial \Phi} \Big|_{\Phi=\Phi^*} = 0$. The nature of the mean-field solution depends on the choice of the Hubbard Stratonovich decoupling. Thus, in the auxiliary field QMC mean-field Hamiltonians are solved exactly, the price being the above mentioned scaling of the computational effort. In the above framework, the Monte Carlo integration over the field Φ may be seen as a means of taking into account all fluctuations around the mean-field solution. This will in many cases introduce a sign problem. Nevertheless the method has the advantage that symmetries of the model, such as particle-hole symmetry, may be put to use to avoid the sign problem in many non-trivial cases. Other classes of models where the sign problem does not occur include models with attractive interactions which couple independently to an internal symmetry with an even number of states. The attractive Hubbard model is a member of this class. It is also worth mentioning that when the sign problem occurs in the auxiliary field QMC it is often less severe than in World Line approach so that at least high temperature properties may be investigated.

^aFor the Hirsch-Fye impurity algorithm computational effort scales as the cubed of the inverse temperature.

The auxiliary field quantum Monte-Carlo method is the central topic of this article. In section 3 we will review in all details both the finite temperature^{9–11} and ground state^{12–14} formulation of the method. The application of the method to various models (Hubbard model, periodic Anderson model, Kondo lattice model, hard core boson systems and Heisenberg models) will be discussed in section 4. Since the computational effort scales as the cube of the volume, it is important to control size effects. A simple method to reduce size effects by an order of magnitude in temperature will also be discussed in section 4. In section 5 we review a very much related algorithm, the Hirsch-Fye impurity algorithm, which has been used extensively in the context of dynamical mean-field theories.^{15,16} We will apply this algorithm to the single impurity Kondo problem.

Finally, we provide an Appendix for the Monte Carlo method and error analysis.

2 The World Line Approach for the XXZ Model and Relation to the 6-Vertex Model

To illustrate with a simple example the World Line quantum Monte Carlo method, we consider the XXZ quantum spin chain defined as:

$$H = J_X \sum_i (S_i^x S_{i+1}^x + S_i^y S_{i+1}^y) + J_Z \sum_i S_i^z S_{i+1}^z \quad (2)$$

where \vec{S}_i are the spin 1/2 operators on site i satisfying the commutation relations:

$$[S_i^\eta, S_j^\nu] = \delta_{i,j} i \epsilon^{\eta,\nu,\gamma} S_i^\gamma. \quad (3)$$

In the above, $\epsilon^{\eta,\nu,\gamma}$ is the antisymmetric tensor and the sum over repeated indices is understood. Our aim is to compute observables:

$$\langle O \rangle = \frac{\text{Tr} [e^{-\beta H} O]}{\text{Tr} [e^{-\beta H}]} \quad (4)$$

The basic idea of the World Line algorithm is to split the above Hamiltonian into a set of independent - in this case - two site problems. The way to achieve this decoupling is with the use of a path integral and the Trotter decomposition. First we write

$$H = \underbrace{\sum_n H^{(2n+1)}}_{H_1} + \underbrace{\sum_n H^{(2n+2)}}_{H_2} \quad (5)$$

with

$$H^{(i)} = J_X (S_i^x S_{i+1}^x + S_i^y S_{i+1}^y) + J_Z S_i^z S_{i+1}^z$$

One may verify that H_1 and H_2 are sums of commuting (i.e. independent) two site problems. Hence, on their own H_1 and H_2 are trivially solvable problems. However, H is not.

To put to use this fact, we wish split the imaginary propagation $e^{-\beta H}$ into successive infinitesimal propagations of H_1 and H_2 . This is achieved with the Trotter decomposition:¹⁷

$$\begin{aligned} (e^{-\Delta\tau H_1} e^{-\Delta\tau H_2})^m &= \left(e^{-\Delta\tau H} + \frac{\Delta\tau^2}{2} [H_1, H_2] + \mathcal{O}(\Delta\tau^3) \right)^m \\ &= \left(e^{-\Delta\tau(H - \Delta\tau[H_1, H_2]/2)} + \mathcal{O}(\Delta\tau^3) \right)^m = e^{-\beta(H - \Delta\tau[H_1, H_2]/2)} + \mathcal{O}(\Delta\tau^2) \\ &= e^{-\beta H} + \frac{\Delta\tau}{2} \int_0^\beta d\tau e^{-(\beta-\tau)H} [H_1, H_2] e^{-\tau H} + \mathcal{O}(\Delta\tau^2) \end{aligned} \quad (6)$$

where $m\Delta\tau = \beta$ and $\mathcal{O}(\Delta\tau^n)$ means that for fixed values of β the error scales as $\Delta\tau^n$. In many cases, we will not take the limit $\Delta\tau \rightarrow 0$ and is important to understand the order of the systematic error produced by the above decomposition.^b A priori, it is of the order $\Delta\tau$. However, in many non-trivial cases, the prefactor of the error of order $\Delta\tau$ vanishes. In the World line approach we compute:

$$\frac{\text{Tr} [(e^{-\Delta\tau H_1} e^{-\Delta\tau H_2})^m O]}{\text{Tr} [(e^{-\Delta\tau H_1} e^{-\Delta\tau H_2})^m]} = \frac{\text{Tr} [e^{-\beta H} O] + \frac{\Delta\tau}{2} \text{Tr} [AO]}{\text{Tr} [e^{-\beta H}] + \frac{\Delta\tau}{2} \text{Tr} [A]} + \mathcal{O}(\Delta\tau^2). \quad (7)$$

Here $A = \int_0^\beta d\tau e^{-(\beta-\tau)H} [H_1, H_2] e^{-\tau H}$ and $m\Delta\tau = \beta$. Since A is an antihermitian operator, $A^\dagger = -A$, it follows that $\overline{\text{Tr}[A]} = \text{Tr}[A^\dagger] = -\text{Tr}[A]$ as well as $\overline{\text{Tr}[AO]} = -\text{Tr}[AO]$. Note that the observable O is a hermitian operator. Thus, if O , H_1 and H_2 are simultaneously real representable in a given basis, the systematic error proportional to $\Delta\tau$ vanishes since in this case the trace is real. Hence the systematic error is of order $\Delta\tau^2$.

With the above, the estimation of the partition function reads:

$$\begin{aligned} \text{Tr} [e^{-\beta H}] &= \text{Tr} [(e^{-\Delta\tau H})^m] = \text{Tr} [(e^{-\Delta\tau H_1} e^{-\Delta\tau H_2})^m] + \mathcal{O}(\Delta\tau^2) = \\ &\sum_{n_1 \dots n_{2m}} \langle n_1 | e^{-\Delta\tau H_1} | n_{2m} \rangle \dots \langle n_3 | e^{-\Delta\tau H_1} | n_2 \rangle \langle n_2 | e^{-\Delta\tau H_2} | n_1 \rangle + \mathcal{O}(\Delta\tau^2) \end{aligned} \quad (8)$$

where $m\Delta\tau = \beta$ and the states $|n_\tau\rangle$ span the Hilbert space. We choose the states $|n_\tau\rangle$ to be eigenstates of S_i^z . For each set of states $|n_1\rangle \dots |n_{2m}\rangle$ with non-vanishing contribution to the partition function we have a simple graphical representation in terms of world lines as shown in Fig. 1. Observables are now given by:

$$\langle O \rangle = \frac{\sum_w \Omega(w) O(w)}{\sum_w \Omega(w)} \quad (9)$$

where $\Omega(w)$ corresponds to the weight of a given world line configuration as obtained through multiplication of the weights of the individual plaquettes listed in Fig. 1. Note that although spin-flip processes have negative weight, no sign problem occurs. One can for example carry out the transformation: $S_i^x \rightarrow (-1)^i S_i^x, S_i^y \rightarrow (-1)^i S_i^y, S_i^z \rightarrow S_i^z$ which leaves the commutation relation unaltered (i.e. is canonical) but changes the sign of J_x . Observables O which locally conserve the z-component of the spin are easy to compute. If we decide to measure on time slice τ then $O|n_\tau\rangle = O(w)|n_\tau\rangle$.^c

^bWithin the loop algorithm a continuous time formulation may be achieved¹⁸

^cIn practice, one will measure on all time slices so as to reduce statistical fluctuations

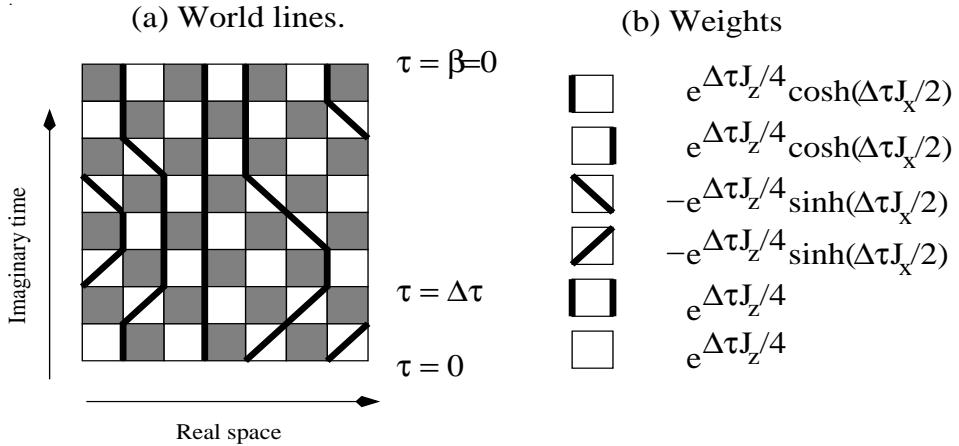


Figure 1. (a) World line configuration for the XXZ model of Eq. (2). Here, $m = 4$ and the system size is $L = 8$. The bold lines follow the time evolution of the up spins and empty sites, with respect to the world lines, correspond to the down spins. A full time step $\Delta\tau$ corresponds to the propagation with H_1 followed by H_2 . Periodic boundary conditions are chosen in the spacial direction. In the time direction, periodic boundary conditions follow from the fact that we are evaluating a trace. (b) The weights for a given world line configuration is the product of the weights of plaquettes listed in the figure. Note that, although the spin-flip processes come with a minus sign the overall weight for the world line configuration is positive since each world line configuration contains an even number of spin flips.

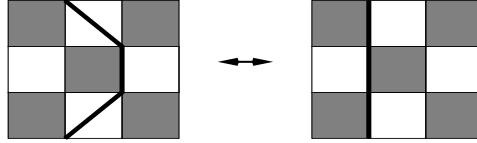


Figure 2. Local updates. A shaded plaquette is chosen randomly and a Word Line is shifted from left to right or vice versa across the shaded plaquette.

The problem is now cast into one which may be solved with classical Monte Carlo methods (see Appendix). To generate a Markov chain through the space of World Lines we need to devise an updating mechanism. Local updates where one locally deforms a World Line configuration have been used successfully (see Fig. 2). The local updates conserve the z-component of the total spin (i.e. canonical in the hard core boson notation introduced in Section 4.4) and are ergodic only in the case of open boundary conditions. Choosing periodic boundary conditions and starting with a configuration with zero winding one will remain in this sector. That is: the configuration of Fig. 4e will not be generated starting from the configuration of Fig. 4a with local updates. Note however, that this is a boundary problem so that when the thermodynamic limit is taken with the above local updates the correct thermodynamic result is obtained.⁶

That the XXZ quantum spin chain is equivalent to the classical two-dimensional 6-vertex model follows from a one to one mapping of a World Line configuration to one of the 6-vertex model. The identification of single plaquettes is shown in Fig. 3(a). The

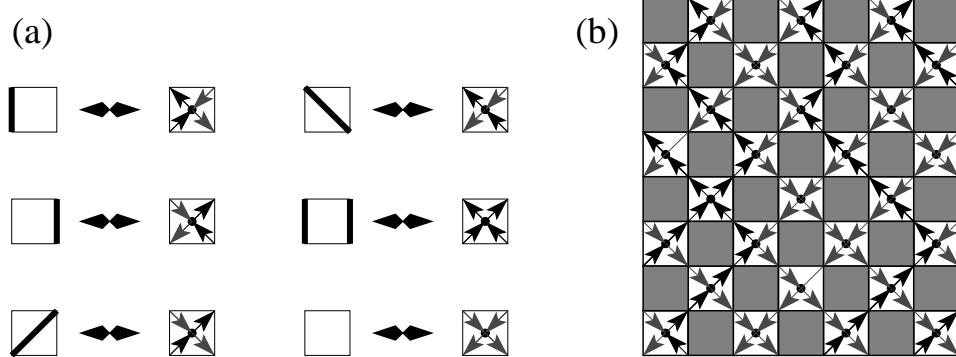


Figure 3. (a) Identification of world lines configurations on plaquettes with the vertices of the 6-vertex model. (b) The World Line configuration of Fig. 1 in the language of the 6-vertex model.

world line configuration of Fig. 1 is plotted in the language of the 6-vertex mode in Fig. 3(b). The vertex model lies on a 45 degrees rotated lattice denoted by bullets in Fig. 1(b). At each vertex (bullets in Fig. 1(b)) the number of incoming arrows equals the number of outgoing arrows. In the case of the XYZ chain, source and drain terms have to be added to yield the 8-vertex model.

The identification of the XXZ model to the 6-vertex model gives us an intuitive picture of loop upgrades.² Consider the World Line configuration in Fig. 4a and it's corresponding vertex formulation (Fig. 4b). One can pick a palquette at random and follow the arrows of the vertex configuration. At each plaquette there are two possible arrow paths to follow. One is chosen, appropriately, and the arrows are followed to arrive to the next plaquette. The procedure is then repeated until one returns to the starting point. Such a loop is shown in Fig. 4c. Along the loop, changing the direction of the arrows generates another valid vertex configuration (see Fig. 4d). The corresponding World Line configuration (after *flipping* the loop) is shown in Fig. 4e. As apparent, this is a global update which changes the winding number and is not achievable with local moves. To gain further insight into the loop algorithm the reader is referred to the review article of H.G. Evertz¹⁹ and references therein. Let us however mention that the loop algorithm has been applied with great success to non-frustrated spin systems. Critical exponents of the order-disorder transition for a two dimensional depleted Heisenberg model were pinned down to show that the transition belongs to the universality class of the three dimensional classical $O(3)$ model.⁸ Furthermore this algorithm has been used to study single particle dynamics in non-frustrated quantum antiferromagnets on various topologies.^{20,21}

2.1 The Sign Problem in the World Line Approach

The Quantum Monte Carlo approach is often plagued by the so-called sign problem. Since the origin of this problem is easily understood in the framework of the World Line algorithm we will briefly discuss it in this section on a specific model. Consider spinless

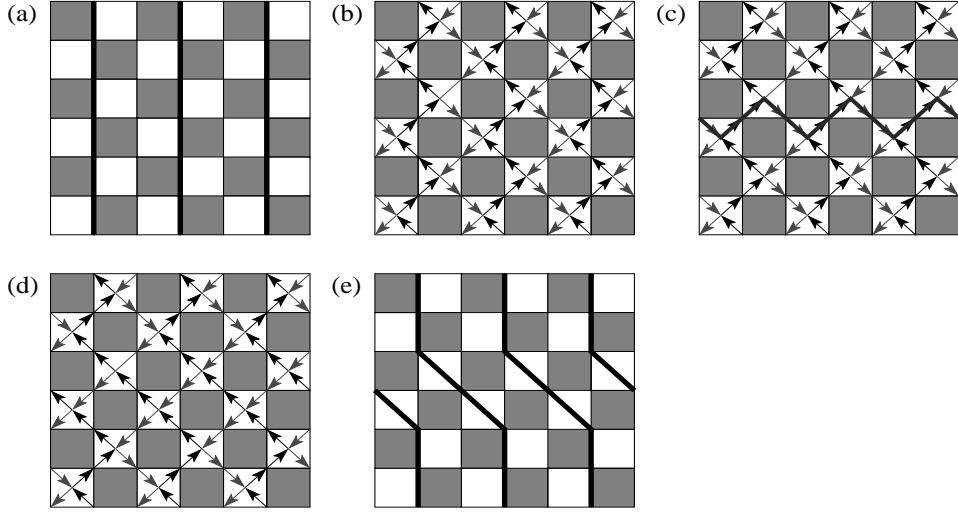


Figure 4. Example of a loop update.

electrons on an L -site linear chain

$$H = -t \sum_i c_i^\dagger (c_{i+1} + c_{i+2}) + \text{H.c.} \quad \text{with} \quad \{c_i^\dagger, c_j^\dagger\} = \{c_i, c_j\} = 0, \{c_i^\dagger, c_j\} = \delta_{i,j}. \quad (10)$$

Here, we consider periodic boundary conditions, $c_{i+L} = c_i$ and $t > 0$. To apply the World Line algorithm to the above Hamiltonian we split it into a set of independent four site problems:

$$H = \underbrace{\sum_{n=0}^{L/4-1} H^{(4n+1)}}_{H_1} + \underbrace{\sum_{n=0}^{L/4-1} H^{(4n+3)}}_{H_2} \quad (11)$$

with

$$H^{(i)} = -tc_i^\dagger \left(\frac{1}{2}c_{i+1} + c_{i+2} \right) - tc_{i+1}^\dagger (c_{i+2} + c_{i+3}) - \frac{t}{2}c_{i+2}^\dagger c_{i+3} + \text{H.c.}$$

With this decomposition one obtains the graphical representation of Fig. 5.²²

The sign problem occurs from the fact that the weights $\Omega(w)$ are not necessarily positive. An example is shown in Fig. 10. In this case the origin of negative signs lies in Fermi statistics. To solve the problem, one decides to carry out the sampling with an auxiliary probability distribution:

$$\overline{Pr}(\omega) = \frac{|\Omega(\omega)|}{\sum_w |\Omega(\omega)|} \quad (12)$$

which in the limit of small values of $\Delta\tau$ corresponds to the partition function of the Hamiltonian of Eq. (10) but with fermions replaced by hard-core bosons. Thus, we can now

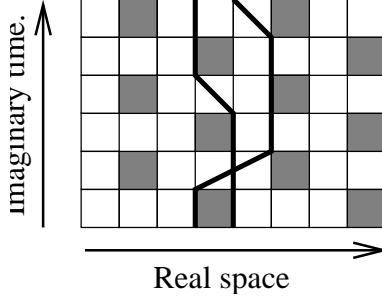


Figure 5. World line configuration for the model of Eq. (10). Here, $m = 3$. Since the two electrons exchange their positions during the imaginary time propagation, this world line configuration has a negative weight.

evaluate Eq. 9 with:

$$\langle O \rangle = \frac{\sum_w \overline{Pr}(\omega) \text{sign}(w) O(w)}{\sum_w \overline{Pr}(\omega) \text{sign}(w)} \quad (13)$$

where both the numerator and denominator are evaluated with MC methods. Let us first consider the denominator:

$$\langle \text{sign} \rangle = \sum_w \overline{Pr}(\omega) \text{sign}(w) = \frac{\sum_w \Omega(w)}{\sum_w |\Omega(w)|} = \frac{\text{Tr} [e^{-\beta H}]}{\text{Tr} [e^{-\beta H_B}]} \quad (14)$$

Here, H_B corresponds to the Hamiltonian of Eq. (10) but with fermions replaced by hard-core bosons. In the limit of large inverse temperatures, β , the partition functions is dominated by the ground state. Thus in this limit

$$\langle \text{sign} \rangle \sim e^{-\beta(E_0 - E_0^B)} = e^{-\beta L \Delta} \quad (15)$$

where $\Delta = (E_0 - E_0^B)/L$ is an intensive, in general positive, quantity. The above equation corresponds to the sign problem. When the temperature is small or system size large, the average sign becomes exponentially small. Hence, the observable $\langle O \rangle$ is given by the quotient of two exponentially small values which are determined stochastically. When the error-bars become comparable to the average sign, uncontrolled fluctuations in the evaluation of $\langle O \rangle$ will occur. Two comments are in order. (i) In this simple example the sign problem occurs due to Fermi statistics. However, sign problems occur equally in frustrated spin-1/2 systems which are nothing but hard core boson models. Note that replacing the fermions by hard core bosons in Eq. (10) and considering hopping matrix elements of different signs between nearest and next nearest neighbors will generate a sign problem in the above formulation. (ii) The sign problem is formulation dependent. In the World Line algorithm, we decide to work in real space. Had we chosen Fourier space, the Hamiltonian would have been diagonal and hence no sign problem occurs. In the auxiliary field approach discussed in the next section the sign problem would not occur for this non-interacting problem since as mentioned in the introduction one body operators are treated exactly. That is, the sum over all World Lines is carried out exactly.

3 Auxiliary Field Quantum Monte Carlo Algorithms

In the remaining, we will concentrate on auxiliary field Quantum Monte Carlo algorithms which we will describe in detail. Consider the Hamiltonian

$$H = H_t + H_I \quad (16)$$

where H_t is the kinetic energy and H_I a two-body interaction term. The ground state expectation value of an observable O is at best obtained by projecting a trial wave function $|\Psi_T\rangle$ along the imaginary time axis:

$$\frac{\langle \Psi_0 | O | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle} = \lim_{\Theta \rightarrow \infty} \frac{\langle \Psi_T | e^{-\Theta H} O e^{-\Theta H} | \Psi_T \rangle}{\langle \Psi_T | e^{-2\Theta H} | \Psi_T \rangle}. \quad (17)$$

The above equation is readily verified by writing $|\Psi_T\rangle = \sum_n |\Psi_n\rangle \langle \Psi_n| \Psi_0\rangle$ with $H|\Psi_n\rangle = E_n |\Psi_n\rangle$. The assumptions that $\langle \Psi_T | \Psi_0 \rangle \neq 0$ and that the ground state is non-degenerate are however required. The algorithm based on Eq. (17) is known as the projector quantum Monte Carlo (PQMC) algorithm.¹²⁻¹⁴

Finite temperature properties in the grand canonical ensemble are obtained by evaluating

$$\langle O \rangle = \frac{\text{Tr} [e^{-\beta(H-\mu N)} O]}{\text{Tr} [e^{-\beta(H-\mu N)}]} \quad (18)$$

where the trace runs over the Fock space, $\beta = 1/k_B T$ and μ is the chemical potential. The algorithm based on Eq. (18) will be referred to as finite temperature QMC (FTQMC) method.^{9,10} Comparison of both algorithms are shown in Fig. (6) for the Hubbard model in standard notation.

$$H_U = -t \sum_{\langle \vec{i}, \vec{j} \rangle, \sigma} e^{\frac{2\pi i}{\Phi_0} \int_{\vec{i}}^{\vec{j}} \vec{A} \cdot d\vec{l}} c_{\vec{i}, \sigma}^\dagger c_{\vec{j}, \sigma} + U \sum_{\vec{i}} n_{\vec{i}, \uparrow} n_{\vec{i}, \downarrow}. \quad (19)$$

Here, and for future use, we have included a magnetic field $\vec{B} = \nabla \times \vec{A}$. At half-filling, the ground state is insulating so that charge fluctuations are absent in the low temperature limit on finite lattices. Hence, in this limit both grand canonical and canonical approaches yield identical results. It is however clear that if one is interested solely in ground state properties the PQMC is more efficient. As we will see, this lies in the choice of the trial wave function which is chosen to be a spin singlet.

3.1 Trotter Decomposition and Hubbard Stratonovich Transformation

As in the World Line approach we wish to evaluate the imaginary time propagation which we will split into infinitesimal successive propagations with H_t followed by H_I . This is again achieved with the Trotter decomposition

$$(e^{-\Delta\tau H_I} e^{-\Delta\tau H_t})^m = e^{-\Theta H} + \frac{\Delta\tau}{2} \int_0^\Theta d\tau e^{-(\Theta-\tau)H} [H_I, H_t] e^{-\tau H} + \mathcal{O}(\Delta\tau^2) \quad (20)$$

or a symmetric variant

$$\left(e^{-\Delta\tau H_t/2} e^{-\Delta\tau H_I} e^{-\Delta\tau H_t/2} \right)^m = e^{-\Theta H} + \mathcal{O}(\Delta\tau^2) \quad (21)$$

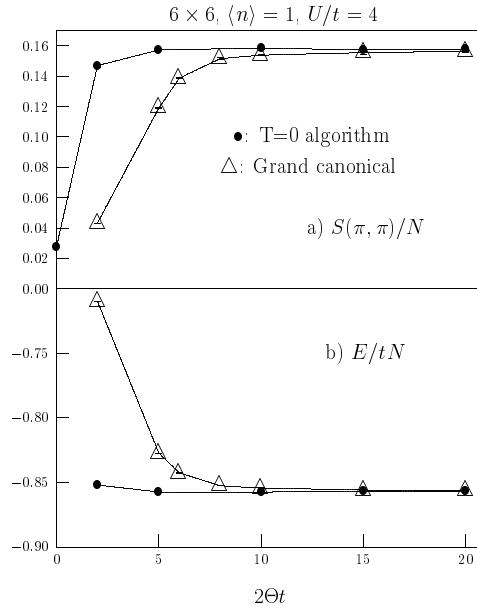


Figure 6. Fourier transform of the spin-spin correlation functions at $\vec{Q} = (\pi, \pi)$ (a) and energy (b) for the half-filled Hubbard model in the absence of magnetic field (19). ●: PQMC algorithm. △: FTQMC algorithm at $\beta = 2\Theta$.

where $m\Delta\tau = \Theta$. As mentioned previously and for the FTQMC if O , H_I and H_t are simultaneously real representable in a given basis, the systematic error proportional to $\Delta\tau$ in Eq. (20) vanishes. To achieve this in the PQMC some care has to be taken. To reduce fluctuations one wishes to measure an observable over several time slices. In the QMC evaluation of O based on the Trotter decomposition of Eq. (20) we hence compute:

$$\frac{1}{2N+1} \sum_{n=-N}^N \frac{\langle \Psi_T | (e^{-\Delta\tau H_I} e^{-\Delta\tau H_t})^{m-n} O (e^{-\Delta\tau H_I} e^{-\Delta\tau H_t})^{m+n} | \Psi_T \rangle}{\langle \Psi_T | (e^{-\Delta\tau H_I} e^{-\Delta\tau H_t})^{2m} | \Psi_T \rangle}. \quad (22)$$

If $[O, H] = 0$, then one can set $N = m$ and the effective projection parameter is 2Θ . On the other hand, if $[O, H] \neq 0$ one has to choose $N < m$ since the effective projection parameter is $(m-N)\Delta\tau$. It is however crucial to measure symmetrically around the central time slice, since only then can show that the systematic error proportional to $\Delta\tau$ vanishes. This is of course valid provided that O , H_I , H_t as well as $|\Psi_T\rangle$ are simultaneously real representable.

Fig. 7 compares both choices of Trotter decompositions (Eqn. (20) and (21)) for the Hubbard model of Eq. (19). Here we use the PQMC algorithm. As apparent the symmetric decomposition of Eq. (21) is much more accurate and due to the variational principle provides an upper bound to the exact energy. It is however often cumbersome to implement.

Having isolated the interaction term H_I with the Trotter decomposition, we may now proceed with the Hubbard-Stratonovich (HS) transformation. The choice of the HS is

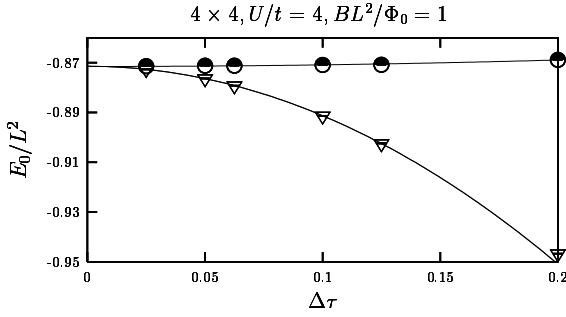


Figure 7. Ground state energy of the Half-filled Hubbard model on a 4×4 lattice, $U/t = 4$, $\langle n \rangle = 1$ and $BL^2/\Phi_0 = 1$ as a function of $\Delta\tau$ as obtained with the PQMC. The Trotter decompositions of Eqn. (20) (∇) and (21) (\circ) are considered. Note that due to the variational principle the Trotter decomposition of Eq. (21) yields an upper bound to the energy. The solid lines correspond to least square fits to the form $a + b\Delta\tau^2$.

important. For Hubbard interactions,

$$H_I = U \sum_{\vec{i}} \left(n_{\vec{i},\uparrow} - 1/2 \right) \left(n_{\vec{i},\downarrow} - 1/2 \right), \quad (23)$$

with $U > 0$ one usually chooses Hirsch's discrete transformation²³

$$\begin{aligned} & \exp \left(-\Delta\tau U \sum_{\vec{i}} \left(n_{\vec{i},\uparrow} - 1/2 \right) \left(n_{\vec{i},\downarrow} - 1/2 \right) \right) \\ &= \tilde{C} \sum_{s_1, \dots, s_N = \pm 1} \exp \left(\tilde{\alpha} \sum_{\vec{i}} s_{\vec{i}} \left(n_{\vec{i},\uparrow} - n_{\vec{i},\downarrow} \right) \right). \end{aligned} \quad (24)$$

where $\cosh(\tilde{\alpha}) = \exp(\Delta\tau U/2)$. On an N -site lattice, the constant $\tilde{C} = \exp(\Delta\tau UN/4)/2^N$. As apparent from the above equation, for a fixed set of HS fields, $s_1 \dots s_N$, $SU(2)$ -spin symmetry is broken since the field couples to the z -component of the magnetization. This symmetry is of course restored after summation over the HS fields with the Monte Carlo method. Alternatively, one may consider²³

$$\begin{aligned} & \exp \left(-\Delta\tau U \sum_{\vec{i}} \left(n_{\vec{i},\uparrow} - 1/2 \right) \left(n_{\vec{i},\downarrow} - 1/2 \right) \right) \\ &= C \sum_{s_1, \dots, s_N = \pm 1} \exp \left(i\alpha \sum_{\vec{i}} s_{\vec{i}} \left(n_{\vec{i},\uparrow} + n_{\vec{i},\downarrow} - 1 \right) \right). \end{aligned} \quad (25)$$

where $\cos(\alpha) = \exp(-\Delta\tau U/2)$ and $C = \exp(\Delta\tau UN/4)/2^N$. With this choice of the HS transformation $SU(2)$ spin invariance is retained for any given HS configuration since the field couples to the density. Even taking into account the overhead of working with complex numbers, it is more convenient to work with this transformation²⁴ since it is often hard to restore the full $SU(2)$ spin symmetry via Monte Carlo sampling of the HS field. When $U < 0$ the transformation (25) may readily be used and involves only real numbers.

We now consider interaction terms of the form:

$$H_I = -W \sum_{\vec{i}} \left(O^{(\vec{i})} \right)^2 \quad (26)$$

where $O^{(\vec{i})}$ is a one-body operator. In general, $[O^{(\vec{i})}, O^{(\vec{j})}] \neq 0$ so that the sum in the above equation has to be split into sums of commuting terms: $H_I = \sum_r H_I^r$, $H_I^r = -W \sum_{\vec{i} \in S_r} \left(O^{(\vec{i})} \right)^2$. For \vec{i} and \vec{j} in the set S_r one requires $[O^{(\vec{i})}, O^{(\vec{j})}] = 0$. The imaginary time evolution may be written as $e^{-\Delta\tau H_t} \approx \prod_r e^{-\Delta\tau H_I^r}$. Thus we are left with the problem of decoupling $e^{\Delta\tau W O^2}$ where we have omitted the index \vec{i} . In principle, one can decouple a perfect square with the canonical HS transformation:

$$e^{\Delta\tau W O^2} = \frac{1}{\sqrt{2\pi}} \int d\Phi e^{-\frac{\Phi^2}{2} + \sqrt{2\Delta\tau W}\Phi O} \quad (27)$$

However, this involves a continuous field which renders the sampling hard. An alternative formulation is given by:²⁵

$$e^{\Delta\tau W O^2} = \sum_{l=\pm 1, \pm 2} \gamma(l) e^{\sqrt{\Delta\tau W} \eta(l) O} + \mathcal{O}(\Delta\tau^4) \quad (28)$$

where the fields η and γ take the values:

$$\begin{aligned} \gamma(\pm 1) &= 1 + \sqrt{6}/3, \quad \gamma(\pm 2) = 1 - \sqrt{6}/3 \\ \eta(\pm 1) &= \pm \sqrt{2(3 - \sqrt{6})}, \quad \eta(\pm 2) = \pm \sqrt{2(3 + \sqrt{6})}. \end{aligned}$$

This transformation is not exact and produces an overall systematic error proportional to $\Delta\tau^3$. However, since we already have a systematic error proportional to $\Delta\tau^2$ from the Trotter decomposition, the transformation is as good as exact. It also has the great advantage of being discrete thus allowing efficient sampling.

Thus, the HS transformation has enabled us to split the two-body interaction term into a one-body operator interacting with an external field. We may now write the imaginary time propagator as:

$$\begin{aligned} \prod_{n=1}^m [e^{-\Delta\tau H_I} e^{-\Delta\tau H_t}] &= \prod_{n=1}^m \left[\sum_{\vec{s}} C(\vec{s}) e^{H_I(\vec{s})} e^{-\Delta\tau H_t} \right] \\ &\quad \sum_{\vec{s}_1 \cdots \vec{s}_m} C(\vec{s}_1 \cdots \vec{s}_m) \prod_{n=1}^m \left[e^{H_I(\vec{s}_n)} e^{-\Delta\tau H_t} \right] \end{aligned} \quad (29)$$

Here, $H_I(\vec{s}_n)$ is a single body operator. The HS fields have a lattice site, \vec{i} and imaginary time, n , index. We will adopt the notation: $\vec{s} = \{s_{\vec{i}, n}\}$ $\vec{i} = 1 \cdots N$ and $n = 1 \cdots m$, and \vec{s}_n denotes the HS fields on time slice n . In the special case of the Hubbard model with the HS transformation of Eq. (25),

$$H_I(\vec{s}_n) = i\alpha \sum_{\vec{j}, \sigma} s_{\vec{j}, n} c_{\vec{j}, \sigma}^\dagger c_{\vec{j}, \sigma}, \quad C(\vec{s}_1 \cdots \vec{s}_m) = \exp(\Theta U N / 4) \exp \left(-i\alpha \sum_{\vec{j}, n} s_{\vec{j}, n} \right). \quad (30)$$

For an interaction term of the form in Eq. (26) $e^{H_I(\vec{s}_n)}$ has to be replaced by

$$\prod_r \exp \left(\sqrt{\Delta\tau W} \sum_{\vec{i} \in S_r} \eta(l_{\vec{i},n}) O^{(\vec{i})} \right) \quad (31)$$

and $C(\{s_{\vec{i},n}\})$ by $C(\{l_{\vec{i},n}\}) = \prod_{\vec{i},n} \gamma(l_{\vec{i},n})$.

At this stage, the fermionic degrees of freedom may be integrated out. In the framework of the PQMC, we will require the trial wave function to be a Slater determinant:

$$|\Psi_T\rangle = \prod_{y=1}^{N_p} \left(\sum_x c_x^\dagger P_{x,y} \right) |0\rangle = \prod_{y=1}^{N_p} (\vec{c}^\dagger P)_y |0\rangle \quad (32)$$

where we have introduced the indices $\vec{x} = (\vec{i}, \sigma)$, which run from $1 \cdots 2N$, N_p is the number of particles and $\vec{c}^\dagger = (c_1^\dagger, \dots, c_{2N}^\dagger)$. For a Hermitian or anti-hermitian matrix T , one may show that:

$$e^{\vec{c}^\dagger T \vec{c}} \prod_{y=1}^{N_p} (\vec{c}^\dagger P)_y |0\rangle = \prod_{y=1}^{N_p} (\vec{c}^\dagger e^T P)_y |0\rangle. \quad (33)$$

To derive the above equation, it is useful to go into a basis where T is diagonal: $U^\dagger T U = D$. U is a unitary matrix and D a real (purely imaginary) diagonal matrix provided that T is hermetian (anti-hermetian). Thus we can define the fermionic operators $\vec{\gamma}^\dagger = \vec{c}^\dagger U$ to obtain:

$$\begin{aligned} e^{\vec{c}^\dagger T \vec{c}} \prod_{y=1}^{N_p} (\vec{c}^\dagger P)_y |0\rangle &= e^{\vec{\gamma}^\dagger D \vec{\gamma}} \prod_{y=1}^{N_p} (\vec{\gamma}^\dagger U P)_y |0\rangle = \\ &\sum_{y_1, \dots, y_{N_p}} e^{\sum_x D_{x,x} \vec{\gamma}_x^\dagger \vec{\gamma}_x} \gamma_{y_1}^\dagger \cdots \gamma_{y_{N_p}}^\dagger |0\rangle (UP)_{y_1,1} \cdots (UP)_{y_{N_p},N_p} = \\ &\sum_{y_1, \dots, y_{N_p}} e^{D_{y_1,y_1}} \gamma_{y_1}^\dagger \cdots e^{D_{y_{N_p},y_{N_p}}} \gamma_{y_{N_p}}^\dagger |0\rangle (UP)_{y_1,1} \cdots (UP)_{y_{N_p},N_p} = \\ &\prod_{y=1}^{N_p} (\vec{\gamma}^\dagger e^D U P)_y |0\rangle = \prod_{y=1}^{N_p} (\vec{c}^\dagger U^\dagger e^D U P)_y |0\rangle = \prod_{y=1}^{N_p} (\vec{c}^\dagger e^T P)_y |0\rangle. \end{aligned}$$

We can now evaluate the imaginary time propagator. It is convenient to define the notation:

$$\begin{aligned} H_I &= \vec{c}^\dagger h_I(\vec{s}_\tau) \vec{c}, \quad H_t = \vec{c}^\dagger h_t \vec{c} \\ U_{\vec{s}}(\tau_2, \tau_1) &= \prod_{n=n_1+1}^{n_2} e^{H_I(\vec{s}_n)} e^{-\Delta\tau H_t} \end{aligned} \quad (34)$$

$$\text{and } B_{\vec{s}}(\tau_2, \tau_1) = \prod_{n=n_1+1}^{n_2} e^{h_I(\vec{s}_n)} e^{-\Delta\tau h_t} \quad (35)$$

where, $n_1 \Delta\tau = \tau_1$ and $n_2 \Delta\tau = \tau_2$, h_t is a $2N \times 2N$ hermetian matrix and $h_I(\vec{s}_n)$ a $2N \times 2N$ hermetian or anti-hermetian matrix depending upon the choice of the HS

transformation. One can then derive:

$$\langle \Psi_T | U_{\vec{s}}(2\Theta, 0) | \Psi_T \rangle = \det(P^\dagger B(2\Theta, 0) P). \quad (36)$$

The above follows from:

$$\begin{aligned} \langle \Psi_T | U_{\vec{s}}(2\Theta, 0) | \Psi_T \rangle &= \langle \Psi_T | \prod_{y=1}^{N_p} (\vec{c}^\dagger B_{\vec{s}}(2\Theta, 0) P)_y | 0 \rangle = \\ &\sum_{\substack{x_1 \cdots x_{N_p} \\ y_1 \cdots y_{N_p}}} P_{1,y_{N_p}}^\dagger \cdots P_{1,y_1}^\dagger \langle 0 | c_{y_1} \cdots c_{y_{N_p}} c_{x_{N_p}}^\dagger \cdots c_{x_1}^\dagger | 0 \rangle \times \\ &(B_{\vec{s}}(2\Theta, 0) P)_{x_1,1} \cdots (B_{\vec{s}}(2\Theta, 0) P)_{x_{N_p},N_p}. \end{aligned}$$

The matrix element $\langle 0 | c_{y_1} \cdots c_{y_{N_p}} c_{x_{N_p}}^\dagger \cdots c_{x_1}^\dagger | 0 \rangle$ does not vanish provided that

- i) $x_1 \neq x_2 \neq \cdots \neq x_{N_p}$ and
- ii) the indices $y_1 \cdots y_{N_p}$ are a permutation (π) of the indices $x_1 \cdots x_{N_p}$:

$$y_1 = x_{\pi(1)} \cdots y_{N_p} = x_{\pi(N_p)}.$$

The matrix element is then equal to the sign of permutation: $(-1)^\pi$. Hence,

$$\begin{aligned} \langle \Psi_T | U_{\vec{s}}(2\Theta, 0) | \Psi_T \rangle &= \\ &\sum_{\substack{x_1 \cdots x_{N_p} \\ \pi \in \mathcal{S}_{N_p}}} (-1)^\pi P_{1,x_{\pi(1)}}^\dagger (B_{\vec{s}}(2\Theta, 0) P)_{x_1,1} \cdots P_{N_p,x_{\pi(N_p)}}^\dagger (B_{\vec{s}}(2\Theta, 0) P)_{x_{N_p},N_p} = \\ &\sum_{\pi \in \mathcal{S}_{N_p}} (-1)^\pi (P^\dagger B_{\vec{s}}(2\Theta, 0) P)_{\pi(1),1} \cdots (P^\dagger B_{\vec{s}}(2\Theta, 0) P)_{\pi(N_p),N_p} = \\ &\det(P^\dagger B_{\vec{s}}(2\Theta, 0) P) \end{aligned}$$

where the last sum runs over the space \mathcal{S}_{N_p} consisting of the $N_p!$ permutations of the integers $[1 \cdots N_p]$.

For the FTQMC, we have to evaluate evaluate the trace over the Fock space.

$$\text{Tr}(U_{\vec{s}}(\beta, 0)) = \det(1 + B_{\vec{s}}(\beta, 0)) \quad (37)$$

where $m\Delta\tau = \beta$. The above equation is readily verified:

$$\begin{aligned}
& \det(1 + B_{\vec{s}}(\beta, 0)) \\
&= \sum_{\pi \in \mathcal{S}_{2N}} (-1)^\pi (1 + B_{\vec{s}}(\beta, 0))_{\pi(1), 1} \cdots (1 + B_{\vec{s}}(\beta, 0))_{\pi(2N), 2N} \\
&= \sum_{\pi \in \mathcal{S}_{2N}} (-1)^\pi \delta_{1, \pi(1)} \cdots \delta_{2N, \pi(2N)} + \\
&\quad \sum_x \sum_{\pi \in \mathcal{S}_{2N}} (-1)^\pi B_{\vec{s}}(\beta, 0)_{\pi(x), x} \delta_{1, \pi(1)} \cdots \widehat{\delta_{x, \pi(x)}} \cdots \delta_{2N, \pi(2N)} + \\
&\quad \sum_{y > x} \sum_{\pi \in \mathcal{S}_{2N}} (-1)^\pi B_{\vec{s}}(\beta, 0)_{\pi(x), x} B_{\vec{s}}(\beta, 0)_{\pi(y), y} \times \\
&\quad \quad \delta_{1, \pi(1)} \cdots \widehat{\delta_{x, \pi(x)}} \cdots \widehat{\delta_{y, \pi(y)}} \cdots \delta_{2N, \pi(2N)} + \\
&\quad \sum_{y > x > z} \sum_{\pi \in \mathcal{S}_{2N}} (-1)^\pi B_{\vec{s}}(\beta, 0)_{\pi(x), x} B_{\vec{s}}(\beta, 0)_{\pi(y), y} B_{\vec{s}}(\beta, 0)_{\pi(z), z} \times \\
&\quad \quad \delta_{1, \pi(1)} \cdots \widehat{\delta_{x, \pi(x)}} \cdots \widehat{\delta_{y, \pi(y)}} \cdots \widehat{\delta_{z, \pi(z)}} \cdots \delta_{2N, \pi(2N)} + \cdots \\
&= 1 + \sum_x \langle 0 | c_x U_{\vec{s}}(\beta, 0) c_x^\dagger | 0 \rangle + \sum_{y > x} \langle 0 | c_x c_y U_{\vec{s}}(\beta, 0) c_y^\dagger c_x^\dagger | 0 \rangle + \\
&\quad \sum_{y > x > z} \langle 0 | c_x c_y c_z U_{\vec{s}}(\beta, 0) c_z^\dagger c_y^\dagger c_x^\dagger | 0 \rangle + \cdots \\
&= \text{Tr}(U_{\vec{s}}(\beta, 0)).
\end{aligned}$$

Here, $\widehat{\delta_{y, \pi(y)}}$ means that this term is omitted in the product: $\prod_{x=1}^{2N} \delta_{x, \pi(x)}$. We have used Eq. (36) to derive the third equality.

3.2 Observables and Wick's Theorem

In the last section, we have shown how to carry out the HS transformation and integrate out the fermionic degrees of freedom. Here, we show both for the PQMC and FTQMC how to compute observables as well as the validity of Wick's theorem for a fixed configuration of HS fields.

3.2.1 PQMC

In the PQMC algorithm we compute:

$$\frac{\langle \Psi_T | e^{-\Theta H} O e^{-\Theta H} | \Psi_T \rangle}{\langle \Psi_T | e^{-2\Theta H} | \Psi_T \rangle} = \sum_{\vec{s}} \text{Pr}_{\vec{s}} \langle O \rangle_{\vec{s}} + O(\Delta\tau^2). \quad (38)$$

For each lattice site, \vec{i} , time slice, n , we have introduced an independent HS field, $\vec{s} = \{s_{\vec{i},n}\}$ and

$$\Pr_{\vec{s}} = \frac{C_{\vec{s}} \det(P^\dagger B_{\vec{s}}(2\Theta, 0)P)}{\sum_{\vec{s}} C_{\vec{s}} \det(P^\dagger B_{\vec{s}}(2\Theta, 0)P)}$$

$$\langle O \rangle_{\vec{s}} = \frac{\langle \Psi_T | U_{\vec{s}}(2\Theta, \Theta) O U_{\vec{s}}(\Theta, 0) | \Psi_T \rangle}{\langle \Psi_T | U_{\vec{s}}(2\Theta, 0) | \Psi_T \rangle}$$

We start by computing the equal time Green function: $O = c_x c_y^\dagger = \delta_{x,y} - \vec{c}^\dagger A^{(y,x)} \vec{c}$ with $A_{x_1, x_2}^{(y,x)} = \delta_{x_1, y} \delta_{x_2, y}$. Inserting a source term, we obtain:

$$\begin{aligned} \langle c_x c_y^\dagger \rangle_{\vec{s}} &= \delta_{x,y} - \frac{\partial}{\partial \eta} \ln \langle \Psi_T | U_{\vec{s}}(2\Theta, \Theta) e^{\eta \vec{c}^\dagger A^{(y,x)}} \vec{c} U_{\vec{s}}(\Theta, 0) | \Psi_T \rangle |_{\eta=0} = \\ &= \delta_{x,y} - \frac{\partial}{\partial \eta} \ln \det \left(P^\dagger B_{\vec{s}}(2\Theta, \Theta) e^{\eta A^{(y,x)}} B_{\vec{s}}(\Theta, 0) P \right) |_{\eta=0} = \\ &= \delta_{x,y} - \frac{\partial}{\partial \eta} \text{Tr} \ln \left(P^\dagger B_{\vec{s}}(2\Theta, \Theta) e^{\eta A^{(y,x)}} B_{\vec{s}}(\Theta, 0) P \right) |_{\eta=0} = \\ &= \delta_{x,y} - \text{Tr} \left[(P^\dagger B_{\vec{s}}(2\Theta, 0) P)^{-1} P^\dagger B_{\vec{s}}(2\Theta, \Theta) A^{(y,x)} B_{\vec{s}}(\Theta, 0) P \right] \\ &\quad \left(1 - B_{\vec{s}}(\Theta, 0) P (P^\dagger B_{\vec{s}}(2\Theta, 0) P)^{-1} P^\dagger B_{\vec{s}}(2\Theta, \Theta) \right)_{x,y} \equiv (G_{\vec{s}}(\Theta))_{x,y} \end{aligned} \quad (39)$$

We have used Eq. (36) to go from the second to third equality. The attentive reader will have noticed that Eq. (36) was shown to be valid only in the case of hermitian or anti-hermitian matrices which is certainly not the case of $A^{(y,x)}$. However, since only terms of order η are relevant in the calculation, we may replace $e^{\eta A}$ by $e^{\eta(A+A^\dagger)/2} e^{\eta(A-A^\dagger)/2}$ which is exact up to order η^2 . For the latter form, one may use Eq. (36). To obtain the fourth equality we have used the relation: $\det A = \exp \text{Tr} \ln A$.

We now show that any multi-point correlation function decouples into a sum of products of the above defined Green functions. First, we define the cumulants:

$$\langle\langle O_n \cdots O_1 \rangle\rangle_{\vec{s}} = \frac{\partial^n \ln \langle \Psi_T | U_{\vec{s}}(2\Theta, \Theta) e^{\eta_n O_n} \cdots e^{\eta_1 O_1} U_{\vec{s}}(\Theta, 0) | \Psi_T \rangle}{\partial \eta_n \cdots \partial \eta_1} \Big|_{\eta_1 \cdots \eta_n = 0}$$

with $O_i = \vec{c}^\dagger A^{(i)} \vec{c}$. (40)

Differentiating the above definition we obtain:

$$\begin{aligned} \langle\langle O_1 \rangle\rangle_{\vec{s}} &= \langle O_1 \rangle_{\vec{s}} \\ \langle\langle O_2 O_1 \rangle\rangle_{\vec{s}} &= \langle O_2 O_1 \rangle_{\vec{s}} - \langle O_2 \rangle_{\vec{s}} \langle O_1 \rangle_{\vec{s}} \\ \langle\langle O_3 O_2 O_1 \rangle\rangle_{\vec{s}} &= \langle O_3 O_2 O_1 \rangle_{\vec{s}} - \\ &\quad \langle O_3 \rangle_{\vec{s}} \langle\langle O_2 O_1 \rangle\rangle_{\vec{s}} - \langle O_2 \rangle_{\vec{s}} \langle\langle O_3 O_1 \rangle\rangle_{\vec{s}} - \langle O_1 \rangle_{\vec{s}} \langle\langle O_3 O_2 \rangle\rangle_{\vec{s}} - \\ &\quad \langle O_1 \rangle_{\vec{s}} \langle O_2 \rangle_{\vec{s}} \langle O_3 \rangle_{\vec{s}}. \end{aligned} \quad (41)$$

The following rule, which may be proven by induction, emerges:

$$\begin{aligned} \langle\langle O_n \cdots O_1 \rangle\rangle_{\vec{s}} &= \langle\langle O_n \cdots O_1 \rangle\rangle_{\vec{s}} + \sum_{j=1}^n \langle\langle O_n \cdots \widehat{O_j} \cdots O_1 \rangle\rangle_{\vec{s}} \langle\langle O_j \rangle\rangle_{\vec{s}} + \\ &\quad \sum_{j>i} \langle\langle O_n \cdots \widehat{O_j} \cdots \widehat{O_i} \cdots O_1 \rangle\rangle_{\vec{s}} \langle\langle O_j O_i \rangle\rangle_{\vec{s}} + \cdots + \\ &\quad \langle\langle O_n \rangle\rangle_{\vec{s}} \cdots \langle\langle O_1 \rangle\rangle_{\vec{s}} \end{aligned} \quad (42)$$

where $\widehat{O_j}$ means that the operator O_j has been omitted from the product.

The cumulants may now be computed order by order. We concentrate on the form $\langle\langle c_{x_n}^\dagger c_{y_n} \cdots c_{x_1}^\dagger c_{y_1} \rangle\rangle$ so that $A_{x,y}^{(i)} = \delta_{x,x_i} \delta_{y,y_i}$. To simplify notation in the next calculation we introduce:

$$B^\rangle = B_{\vec{s}}(\Theta, 0)P, \quad \text{and} \quad B^\langle = P^\dagger B_{\vec{s}}(2\Theta, \Theta) \quad (43)$$

We have already computed $\langle\langle O_1 \rangle\rangle_{\vec{s}}$ (see Eq. (39)):

$$\langle\langle O_1 \rangle\rangle_{\vec{s}} = \langle\langle c_{x_1}^\dagger c_{y_1} \rangle\rangle = \text{Tr} \left((1 - G_{\vec{s}}(\Theta)) A^{(1)} \right) = (1 - G_{\vec{s}}(\Theta))_{y_1, x_1} \quad (44)$$

For $n = 2$ we have:

$$\begin{aligned} \langle\langle O_2 O_1 \rangle\rangle_{\vec{s}} &= \langle\langle c_{x_2}^\dagger c_{y_2} c_{x_1}^\dagger c_{y_1} \rangle\rangle_{\vec{s}} \\ &= \frac{\partial^2 \text{Tr} \ln \left(P^\dagger B_{\vec{s}}(2\Theta, \Theta) e^{\eta_2 A^{(2)}} e^{\eta_1 A^{(1)}} B_{\vec{s}}(\Theta, 0) P \right)}{\partial \eta_2 \partial \eta_1} \Big|_{\eta_2, \eta_1=0} \\ &= \frac{\partial}{\partial \eta_2} \text{Tr} \left[\left(B^\langle e^{\eta_2 A^{(2)}} B^\rangle \right)^{-1} B^\langle e^{\eta_2 A^{(2)}} A^{(1)} B^\rangle \right] \Big|_{\eta_2=0} \\ &= -\text{Tr} \left[\left(B^\langle B^\rangle \right)^{-1} B^\langle A^{(2)} B^\rangle \left(B^\langle B^\rangle \right)^{-1} B^\langle A^{(1)} B^\rangle \right] \\ &\quad + \text{Tr} \left[\left(B^\langle B^\rangle \right)^{-1} B^\langle A^{(2)} A^{(1)} B^\rangle \right] \\ &= \text{Tr} \left(\overline{G_{\vec{s}}(\Theta)} A^{(2)} G_{\vec{s}}(\Theta) A^{(1)} \right) \\ &= \langle c_{x_2}^\dagger c_{y_1} \rangle_{\vec{s}} \langle c_{y_2} c_{x_1}^\dagger \rangle_{\vec{s}}, \quad \text{with } \overline{G} = 1 - G \end{aligned} \quad (45)$$

To derive the above, we have used the cyclic properties of the trace as well as the relation $G = 1 - B^\rangle (B^\langle B^\rangle)^{-1} B^\langle$. Note that for a matrix $A(\eta)$, $\frac{\partial}{\partial \eta} A^{-1}(\eta) = -A^{-1}(\eta) \left(\frac{\partial}{\partial \eta} A(\eta) \right) A^{-1}(\eta)$. There is a simple rule to obtain the third cumulant given the second. In the above expression for the second cumulant, one replaces B^\rangle with $B^\langle e^{\eta_3 A^{(3)}} \rangle$. This amounts in redefining the Green function as

$$G(\eta_3) = 1 - B^\langle \left(B^{\langle} e^{\eta_3 A^{(3)}} B^\rangle \right)^{-1} B^{\langle} e^{\eta_3 A^{(3)}} \rangle. \text{ Thus,}$$

$$\begin{aligned} \langle\langle O_3 O_2 O_1 \rangle\rangle_{\vec{s}} &= \langle\langle c_{x_3}^\dagger c_{y_3} c_{x_2}^\dagger c_{y_2} c_{x_1}^\dagger c_{y_1} \rangle\rangle_{\vec{s}} \\ &= \frac{\partial}{\partial \eta_3} \text{Tr} \left(\overline{G_{\vec{s}}(\Theta, \eta_3)} A^{(2)} G_{\vec{s}}(\Theta, \eta_3) A^{(1)} \right) \Big|_{\eta_3=0} \\ &= \text{Tr} \left(\overline{G_{\vec{s}}(\Theta)} A^{(3)} G_{\vec{s}}(\Theta) A^{(2)} G_{\vec{s}}(\Theta) A^{(1)} \right) - \\ &\quad \text{Tr} \left(\overline{G_{\vec{s}}(\Theta)} A^{(3)} G_{\vec{s}}(\Theta) A^{(1)} \overline{G_{\vec{s}}(\Theta)} A^{(2)} \right) \\ &= \langle c_{x_3}^\dagger c_{y_1} \rangle_{\vec{s}} \langle c_{y_3} c_{x_2}^\dagger \rangle_{\vec{s}} \langle c_{y_2} c_{x_1}^\dagger \rangle_{\vec{s}} - \\ &\quad \langle c_{x_3}^\dagger c_{y_2} \rangle_{\vec{s}} \langle c_{y_3} c_{x_1}^\dagger \rangle_{\vec{s}} \langle c_{x_2}^\dagger c_{y_1} \rangle_{\vec{s}} \end{aligned} \quad (46)$$

since

$$\frac{\partial}{\partial \eta_3} G_{\vec{s}}(\Theta, \eta_3) \Big|_{\eta_3=0} = -\overline{G_{\vec{s}}(\Theta)} A^{(3)} G_{\vec{s}}(\Theta) = -\frac{\partial}{\partial \eta_3} \overline{G_{\vec{s}}(\Theta, \eta_3)} \Big|_{\eta_3=0}.$$

Clearly the same procedure may be applied to obtain the $n^{th}+1$ cumulant given the n^{th} one. It is also clear that the n^{th} cumulant is a sum of products of Green functions. Thus with equation (42) we have shown that any multi-point correlation function may be reduced into a sum of products of Green functions: Wicks theorem. Useful relations include:

$$\langle c_{x_2}^\dagger c_{y_2} c_{x_1}^\dagger c_{y_1} \rangle_{\vec{s}} = \langle c_{x_2}^\dagger c_{y_1} \rangle_{\vec{s}} \langle c_{y_2} c_{x_1}^\dagger \rangle_{\vec{s}} + \langle c_{x_2}^\dagger c_{y_2} \rangle_{\vec{s}} \langle c_{x_1}^\dagger c_{y_1} \rangle_{\vec{s}}. \quad (47)$$

3.2.2 FTQMC

For the FTQMC we wish to evaluate:

$$\frac{\text{Tr} [e^{-\beta H} O]}{\text{Tr} [e^{-\beta H} O]} = \sum_{\vec{s}} \text{Pr}_{\vec{s}} \langle O \rangle_{\vec{s}} + O(\Delta \tau^2). \quad (48)$$

where

$$\text{Pr}_{\vec{s}} = \frac{C_{\vec{s}} \det (1 + B_{\vec{s}}(\beta, 0))}{\sum_{\vec{s}} C_{\vec{s}} \det (1 + B_{\vec{s}}(\beta, 0))}, \quad \langle O \rangle_{\vec{s}} = \frac{\text{Tr} [U_{\vec{s}}(\beta, \tau) O U_{\vec{s}}(\tau, 0)]}{\text{Tr} [U_{\vec{s}}(\beta, 0)]}.$$

Here, we measure the observable on time slice τ . Single body observables, $O = \vec{c}^\dagger A \vec{c}$ are evaluated as:

$$\begin{aligned} \langle O \rangle_{\vec{s}} &= \frac{\partial \ln \text{Tr} [U_{\vec{s}}(\beta, \tau) e^{\eta O} U_{\vec{s}}(\tau, 0)]}{\partial \eta} \Big|_{\eta=0} \\ &= \frac{\partial \ln \det [1 + B_{\vec{s}}(\beta, \tau) e^{\eta A} B_{\vec{s}}(\tau, 0)]}{\partial \eta} \Big|_{\eta=0} \\ &= \frac{\partial \ln \text{Tr} [1 + B_{\vec{s}}(\beta, \tau) e^{\eta A} B_{\vec{s}}(\tau, 0)]}{\partial \eta} \Big|_{\eta=0} \\ &= \text{Tr} [B_{\vec{s}}(\tau, 0) (1 + B_{\vec{s}}(\beta, 0))^{-1} B_{\vec{s}}(\beta, \tau) A] \\ &= \text{Tr} \left[\left(1 - (1 + B_{\vec{s}}(\tau, 0) B_{\vec{s}}(\beta, \tau))^{-1} \right) A \right] \end{aligned} \quad (49)$$

In particular the Green function is given by:

$$\langle c_x c_y^\dagger \rangle_{\vec{s}} = (1 + B_{\vec{s}}(\tau, 0) B_{\vec{s}}(\beta, \tau))_{x,y}^{-1} \quad (50)$$

Defining the cumulants as

$$\langle\langle O_n \cdots O_1 \rangle\rangle_{\vec{s}} = \left. \frac{\partial^n \ln \text{Tr} [U_{\vec{s}}(\beta, \tau) e^{\eta_n O_n} \cdots e^{\eta_1 O_1} U_{\vec{s}}(\tau, 0)]}{\partial \eta_n \cdots \partial \eta_1} \right|_{\eta_1 \cdots \eta_n = 0} \quad (51)$$

with $O_i = \vec{c}^\dagger A^{(i)} \vec{c}$

one can derive Wicks theorem in precisely the same manner as shown for the PQMC. Thus both in the PQMC and FTQMC, it suffices to compute the equal time Green functions to evaluate any equal time observable.

3.3 The Monte Carlo Sampling

We now have to sum over the HS fields. The lowest temperature results in Fig. 6 require summing over 2^{5760} Ising field configurations. An exact summation is thus clearly not possible. We will thus use the Monte Carlo method (see Appendix) and use a single site upgrading method which requires the calculation of the ratio

$$R = \frac{\text{Pr}_{\vec{s}''}}{\text{Pr}_{\vec{s}}} \quad (52)$$

where \vec{s} and \vec{s}'' differ only at one point in space and imaginary time, \vec{i}, n . For the Ising field required to decouple the Hubbard interaction (Eqn. (24) and (25)):

$$s'_{\vec{i}',n'} = \begin{cases} s_{\vec{i}',n'} & \text{if } \vec{i}' \neq \vec{i} \text{ and } n' \neq n \\ -s_{\vec{i},n} & \text{if } \vec{i}' = \vec{i} \text{ and } n' = n \end{cases} \quad (53)$$

For HS fields \vec{l} required to decouple perfect square terms (Eq. (28)):

$$l'_{\vec{i}',n'} = \begin{cases} l_{\vec{i}',n'} & \text{if } \vec{i}' \neq \vec{i} \text{ and } n' \neq n \\ \mathcal{F}(l_{\vec{i},n}) & \text{if } \vec{i}' = \vec{i} \text{ and } n' = n \end{cases} \quad (54)$$

where $\mathcal{F}(l_{\vec{i},n})$ flips the HS field $l_{\vec{i},n}$ into one of the three other choices with probability 1/3. The calculation of R boils down to computing the ration of two determinants:

$$\frac{\det[1 + B_{\vec{s}''}(\beta, 0)]}{\det[1 + B_{\vec{s}}(\beta, 0)]} \quad \text{for the FTQMC} \quad (55)$$

$$\frac{\det[P^\dagger B_{\vec{s}''}(2\Theta, 0) P]}{\det[P^\dagger B_{\vec{s}}(2\Theta, 0) P]} \quad \text{for the PQMC.}$$

For the Hubbard interaction with HS transformation of Eq. (25) only $h_I(\vec{s}_n)$ will be affected by the move.

$$\begin{aligned} e^{h_I(\vec{s}'_n)} &= e^{i\alpha \sum_{\vec{j}} s'_{\vec{j},n} A^{(\vec{j})}} = e^{i\alpha s'_{\vec{i},n} A^{(\vec{i})}} e^{i\alpha \sum_{\vec{j} \neq \vec{i}} s'_{\vec{j},n} A^{(\vec{j})}} \\ &= e^{-i\alpha s_{\vec{i},n} A^{(\vec{i})}} e^{i\alpha \sum_{\vec{j} \neq \vec{i}} s_{\vec{j},n} A^{(\vec{j})}} = e^{-2i\alpha s_{\vec{i},n} A^{(\vec{i})}} e^{i\alpha \sum_{\vec{j}} s_{\vec{j},n} A^{(\vec{j})}} \\ &= (1 + \underbrace{(e^{-2i\alpha s_{\vec{i},n} A^{(\vec{i})}} - 1)}_{=\Delta(\vec{i})}) e^{h_I(\vec{s}_n)}. \end{aligned} \quad (56)$$

Here $A_{x,y}^{(\vec{i})} = \delta_{x,y} \delta_{x,(\vec{i},\sigma)}$ so that $[A^{(\vec{i})}, A^{(\vec{j})}] = 0 \forall \vec{i}, \vec{j}$. The matrix $\Delta^{(\vec{i})}$ is diagonal and has only two non-zero entries at $x = (\vec{i}, \sigma)$. Thus, for the Hubbard model:

$$B_{\vec{s}'}(\bullet, 0) = B_{\vec{s}}(\bullet, \tau) \left(1 + \Delta^{(\vec{i})} \right) B_{\vec{s}}(\tau, 0) \quad (57)$$

where the \bullet stands for 2Θ or β and $\tau = n\Delta\tau$. For interaction term in the form of perfect squares isolating the HS filed $l_{\vec{i},n}$ is a bit more cumbersome. First let us assume that \vec{i} is in the set S_r (see paragraph after Eq. (26)) and that $O^{(\vec{i})} = \vec{c}^\dagger A^{(\vec{i})} \vec{c}$. We will work in a basis where $A^{(\vec{i})}$ is diagonal: $U^{(\vec{i}),\dagger} A^{(\vec{i})} U^{(\vec{i})} = D^{(\vec{i})}$.

$$\begin{aligned} e^{h_I^r(\vec{l}_n)} &= e^{\sqrt{\Delta\tau W} \sum_{j \in S_r} \eta(l'_{j,n}) A^{(\vec{j})}} = e^{\sqrt{\Delta\tau W} (\eta(l'_{i,n}) - \eta(l_{i,n})) A^{(\vec{i})}} e^{h_I^r(\vec{l}_n)} \\ &= U^{(\vec{i})} \left(1 + \underbrace{\left(e^{\sqrt{\Delta\tau W} (\eta(l'_{i,n}) - \eta(l_{i,n})) D^{(\vec{i})}} - 1 \right)}_{=\Delta^{(\vec{i})}} \right) U^{(\vec{i}),\dagger} e^{h_I^r(\vec{l}_n)}. \end{aligned} \quad (58)$$

Thus, $B_{\vec{l}}(\bullet, 0)$ takes the form:

$$\begin{aligned} B_{\vec{l}}(\bullet, 0) &= \overbrace{B_{\vec{l}}(\bullet, \tau) \left[\prod_{r'=1}^{r-1} e^{h_I^{r'}(\vec{l}_n)} \right] U^{(\vec{i})} \left(1 + \Delta^{(\vec{i})} \right)}^{\tilde{B}_{\vec{l}}(\bullet, \tau)} \\ &\quad \underbrace{U^{(\vec{i}),\dagger} \left[\prod_{r' \geq r} e^{h_I^{r'}(\vec{l}_n)} \right] B_{\vec{l}}(\tau, 0)}_{\tilde{B}_{\vec{l}}(\tau, 0)} \end{aligned} \quad (59)$$

After a redefinition of the B matrices as shown in the above equation, a similar form to that for the Hubbard model is obtained. Note that \tilde{B} matrices satisfy: $\tilde{B}_{\vec{l}}(\bullet, \tau) \tilde{B}_{\vec{l}}(\tau, 0) = B_{\vec{l}}(\bullet, 0)$. It is often important to work in a basis where the $A^{(\vec{i})}$ matrices are diagonal since computer time is saved if some of the eigenvalues vanish. In the the so-called t - U - W model^{26,25} a factor 6 in CPU time in the upgrading is saved.

Concentrating first on the PQMC, and again introducing the notation $B_{\vec{s}}^{\langle} = P^\dagger B_{\vec{s}}(2\Theta, \tau)$ and $B_{\vec{s}}^{\rangle} = B_{\vec{s}}(\tau, 0)P$ we have to evaluate:

$$\begin{aligned} \frac{\det [B_{\vec{s}}^{\langle} (1 + \Delta^{(\vec{i})}) B_{\vec{s}}^{\rangle}]}{\det [B_{\vec{s}}^{\langle} B_{\vec{s}}^{\rangle}]} &= \det [B_{\vec{s}}^{\langle} (1 + \Delta^{(\vec{i})}) B_{\vec{s}}^{\rangle} (B_{\vec{s}}^{\langle} B_{\vec{s}}^{\rangle})^{-1}] \\ &= \det [1 + B_{\vec{s}}^{\langle} \Delta^{(\vec{i})} B_{\vec{s}}^{\rangle} (B_{\vec{s}}^{\langle} B_{\vec{s}}^{\rangle})^{-1}] = \det [1 + \Delta^{(\vec{i})} B_{\vec{s}}^{\rangle} (B_{\vec{s}}^{\langle} B_{\vec{s}}^{\rangle})^{-1} B_{\vec{s}}^{\langle}] \end{aligned} \quad (60)$$

where the last equation follows from the identity $\det [1 + AB] = \det [1 + BA]$ for arbitrary rectangular matrices.^d We can recognize the Green function

^dThis identity may be formally shown by using the relation $\det(1 + AB) = \exp \text{Tr} \log(1 + AB)$, expanding the logarithm and using the cyclic properties of the trace.

$B_{\vec{s}'}^\langle \left(B_{\vec{s}}^\langle B_{\vec{s}}^\rangle \right)^{-1} B_{\vec{s}}^\rangle = 1 - G_{\vec{s}}(\tau) \equiv \bar{G}_{\vec{s}}(\tau)$. By construction, $\Delta^{(\vec{i})}$ is diagonal. For a form with Q non vanishing entries,

$$(\Delta^{(\vec{i})})_{z,z_1} = \delta_{z,z_1} \sum_{q=1}^Q \Delta_{x_q,z}^{(\vec{i})} \delta_{x_q,z} \quad (61)$$

one has,

$$\begin{aligned} \det \left[1 + \Delta^{(\vec{i})} \bar{G}_{\vec{s}}(\tau) \right] &= \\ \det \left(\begin{array}{ccc} (1 + \Delta^{(\vec{i})} \bar{G}_{\vec{s}}(\tau))_{x_1,x_1} & \dots & (1 + \Delta^{(\vec{i})} \bar{G}_{\vec{s}}(\tau))_{x_1,x_Q} \\ \vdots & & \vdots \\ (1 + \Delta^{(\vec{i})} \bar{G}_{\vec{s}}(\tau))_{x_Q,x_1} & \dots & (1 + \Delta^{(\vec{i})} \bar{G}_{\vec{s}}(\tau))_{x_Q,x_Q} \end{array} \right). \end{aligned} \quad (62)$$

For the FTQMC, we have to evaluate:

$$\begin{aligned} \frac{\det \left[1 + B_{\vec{s}}(\beta, \tau) (1 + \Delta^{(\vec{i})}) B_{\vec{s}}(\tau, 0) \right]}{\det [1 + B_{\vec{s}}(\beta, 0)]} &= \\ \det \left[1 + \Delta^{(\vec{i})} B_{\vec{s}}(\tau, 0) (1 + B_{\vec{s}}(\beta, 0))^{-1} B_{\vec{s}}(\beta, \tau) \right] &= \\ \det \left[1 + \Delta^{(\vec{i})} \left(1 - (1 + B_{\vec{s}}(\tau, 0) B_{\vec{s}}(\beta, \tau))^{-1} \right) \right]. \end{aligned} \quad (63)$$

Since the finite temperature equal time Green function is given by: $G_{\vec{s}}(\tau) = (1 + B_{\vec{s}}(\tau, 0) B_{\vec{s}}(\beta, \tau))^{-1}$ so that precisely same form as in the PQMC is obtained.

Having calculated the ratio R for a single spin-flip one may now decide stochastically if the move is accepted or not. In case of acceptance, the Green function is to be upgraded since this quantity will be required for the next spin-flip. The upgrading of the Green function is based on the Sherman-Morrison formula:²⁷

$$(A + \vec{u} \otimes \vec{v})^{-1} = A^{-1} - \frac{(A^{-1} \vec{u}) \otimes (\vec{v} A^{-1})}{1 + \vec{v} \bullet A^{-1} \vec{u}} \quad (64)$$

where the outer product is defined by $(\vec{u} \otimes \vec{v})_{x,y} = \vec{u}_x \vec{v}_y$. In the case of the FTQMC and $\Delta^{(\vec{i})}$ given by Eq. (61),

$$\begin{aligned} G_{\vec{s}'}(\tau) &= \left(1 + B_{\vec{s}}(\tau, 0) (1 + \Delta^{(\vec{i})}) B_{\vec{s}}(\beta, \tau) \right)^{-1} \\ &= \left(1 + B_{\vec{s}}(\tau, 0) B_{\vec{s}}(\beta, \tau) + \sum_q \vec{u}^{(q)} \otimes \vec{v}^{(q)} \right)^{-1} \end{aligned} \quad (65)$$

where

$$(\vec{u}^{(q)})_x = (B_{\vec{s}}(\tau, 0))_{x,x_q} \Delta_{x_q}^{(\vec{i})} \quad \text{and} \quad (\vec{v}^{(q)})_x = (B_{\vec{s}}(\beta, \tau))_{x_q,x}.$$

Identifying A to $(G_{\vec{s}}(\tau))^{-1} = 1 + B_{\vec{s}}(\tau, 0) B_{\vec{s}}(\beta, \tau)$, Eq. (64) may now be applied Q times to upgrade the Green function.

For the PQMC, the upgrading of the Green function is equivalent to the upgrading of $(B_{\vec{s}'}^{\langle \rangle} B_{\vec{s}''}^{\rangle})^{-1}$, which is achieved with the Sherman-Morrison formula:

$$(B_{\vec{s}'}^{\langle \rangle} B_{\vec{s}''}^{\rangle})^{-1} = \left(B_{\vec{s}}^{\langle \rangle} (1 + \Delta^{(\vec{i})}) B_{\vec{s}}^{\rangle} \right)^{-1} = \left(B_{\vec{s}'}^{\langle \rangle} B_{\vec{s}''}^{\rangle} + \sum_q \vec{u}^{(q)} \otimes \vec{v}^{(q)} \right)^{-1} \quad (66)$$

with $(\vec{u}^{(q)})_x = (B_{\vec{s}}^{\langle \rangle})_{x,x_q} \Delta_{x_q}^{(\vec{i})}$ and $(\vec{v}^{(q)})_x = (B_{\vec{s}}^{\rangle})_{x_q,x}$. Here x runs from $1 \cdots N_p$ where N_p corresponds to the number of particles contained in the trial wave function.

In principle, we now have all elements required to carry out the simulations. The equal time Green function is the central quantity. On one hand it is used to compute any observables. On the other hand, it determines the Monte Carlo dynamics. It is convenient to adopt a sequential upgrading scheme. Given the Green function at imaginary time $\tau = \Delta\tau$, one upgrades the HS fields on this time slice deterministically or randomly. In case of acceptance, the Green function is upgraded after each single spin flip. To reach the next time slice, the relation:

$$G_{\vec{s}}(\tau + 1) = B_{\vec{s}}(\tau + 1, \tau) G_{\vec{s}}(\tau) (B_{\vec{s}}(\tau + 1, \tau))^{-1} \quad (67)$$

is used and the procedure is repeated till $\tau = \beta$ (FTQMC) or $\tau = 2\Theta$ (PQMC). Having reached $\tau = \beta$ or $\tau = 2\Theta$ we propagate the Green function to back $\tau = 1$ and on the way upgrade the HS fields. The whole procedure may then be repeated. We note that for interactions of the form (26) the propagation of the Green function from time slice τ to time slice $\tau + 1$ is split into intermediate steps according to (59) so as to upgrade the HS fields in the sets S_r successively.

The above corresponds precisely to the procedure adopted in the case of the FTQMC. For the PQMC, it is more efficient to keep track of $(P^\dagger B_{\vec{s}}(2\Theta, 0) P)^{-1}$ since (i) it is of dimension $N_p \times N_p$ in contrast to the Green function which is a $N \times N$ matrix and (ii) it is τ independent. When Green functions are required - to compute R , or observables - they are computed from scratch.

3.4 Numerical Stabilization

In the previous section, we have assumed that we are able to compute the Green functions. On finite precision machines this is unfortunately not the case. To understand the sources of numerical instabilities, it is convenient to consider the PQMC. The rectangular matrix P is just a set of column orthonormal vectors. Typically for a Hubbard model, at weak couplings, the extremal scales in the matrix $B_{\vec{s}}(2\Theta, 0)$ are determined by the kinetic energy and range from $e^{8t\Theta}$ to $e^{-8t\Theta}$ in the two-dimensional case. When the set of orthonormal vectors in P are propagated, the large scales will wash out the small scales yielding an numerically ill defined inversion of the matrix $P^\dagger B_{\vec{s}}(2\Theta, 0) P$. To be more precise consider a two electron problem. The matrix P then consists of two column orthonormal vectors, $\vec{v}(0)_1$ and $\vec{v}(0)_2$. After propagation along the imaginary time axis, will be dominated by the largest scales in $B_{\vec{s}}(2\Theta, 0)$ so that $\vec{v}(2\Theta)_1 = \vec{v}(2\Theta)_2 + \vec{e}$, where $\vec{v}(2\Theta)_1 = B_{\vec{s}}(2\Theta, 0) \vec{v}_1$. It is the information contained in \vec{e} which renders the matrix $P^\dagger B_{\vec{s}}(2\Theta, 0) P$ non-singular. For large values of Θ this information is lost in round-off errors.

To circumvent this problem a set of matrix decomposition techniques were developed.^{13, 14, 10} Those matrix decomposition techniques are best introduced with the Gram-Schmidt orthonormalization method of N_p linearly independent vectors. At imaginary

time τ , $B_{\vec{s}}(\tau, 0)P \equiv B^{\rangle}$ is given by the N_p vectors $\vec{v}_1 \cdots \vec{v}_{N_p}$. Orthogonalizing those vectors yields:

$$\begin{aligned} \vec{v}'_1 &= \vec{v}_1 \\ \vec{v}'_2 &= \vec{v}_2 - \frac{\vec{v}_2 \cdot \vec{v}'_1}{\vec{v}'_1 \cdot \vec{v}'_1} \vec{v}'_1 \\ &\vdots \\ \vec{v}'_{N_p} &= \vec{v}_{N_p} - \sum_{i=1}^{N_p-1} \frac{\vec{v}_{N_p} \cdot \vec{v}'_i}{\vec{v}'_i \cdot \vec{v}'_i} \vec{v}'_i. \end{aligned} \quad (68)$$

Since \vec{v}'_n depends only on the vectors $\vec{v}_n \cdots \vec{v}_1$ we can write,

$$(\vec{v}'_1, \dots, \vec{v}'_{N_p}) = (\vec{v}_1, \dots, \vec{v}_{N_p}) V_R^{-1} \quad (69)$$

where V_R is an upper unit triangular $N_p \times N_p$ matrix, that is the diagonal matrix elements are equal to unity. One can furthermore normalize the vectors $\vec{v}'_1, \dots, \vec{v}'_{N_p}$ to obtain:

$$B^{\rangle} \equiv (\vec{v}_1, \dots, \vec{v}_{N_p}) = \underbrace{\left(\frac{\vec{v}'_1}{|\vec{v}'_1|}, \dots, \frac{\vec{v}'_{N_p}}{|\vec{v}'_{N_p}|} \right)}_{\equiv U^{\rangle}} D_R V_R \quad (70)$$

where D is a diagonal matrix containing the scales. One can repeat the procedure to obtain: $B^{\langle} \equiv P^\dagger B_{\vec{s}}(2\Theta, \tau) = V_L D_L U^{\langle}$. The Green function for the PQMC is now particularly easy to compute:

$$\begin{aligned} 1 - G_{\vec{s}}(\tau) &= B^{\rangle} \left(B^{\langle} B^{\rangle} \right)^{-1} B^{\langle} \\ &= U^{\rangle} D_R V_R \left(V_L D_L U^{\langle} U^{\rangle} D_R V_R \right)^{-1} V_L D_L U^{\langle} \\ &= U^{\rangle} D_R V_R (D_R V_R)^{-1} \left(U^{\langle} U^{\rangle} \right)^{-1} (V_L D_L)^{-1} V_L D_L U^{\langle} \\ &= U^{\rangle} \left(U^{\langle} U^{\rangle} \right)^{-1} U^{\langle} \end{aligned} \quad (71)$$

Thus, in the PQMC, all scales which are at the origin of the numerical instabilities disappear from the problem when computing Green functions. Since the entire algorithm relies solely on the knowledge of the Green function, the above stabilization procedure leaves the physical results invariant. Note that although appealing, the Gram-Schmidt orthonormalization is itself unstable, and hence is more appropriate to use singular value decompositions based on Householder's method to obtain the above UDV form for the B matrices.²⁷ In practice the frequency at which the stabilization is carried out is problem dependent. Typically, for the Hubbard model with $\Delta\tau t = 0.125$ stabilization at every 10th time slice produces excellent accuracy.

The stabilization procedure for the finite temperature algorithm is more subtle since scales may not be doped in the calculation of the Green function. Below, we provide two ways of computing the Green function.

The first approach relies on the identity:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & (C - DB^{-1}A)^{-1} \\ (B - AC^{-1}D)^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix} \quad (72)$$

where A, B, C and D are matrices. Using the above, we obtain:

$$\begin{pmatrix} 1 & B_{\vec{s}}(\tau, 0) \\ -B_{\vec{s}}(\beta, \tau) & 1 \end{pmatrix}^{-1} = \begin{pmatrix} G_{\vec{s}}(0) & -(1 - G_{\vec{s}}(0))B_{\vec{s}}^{-1}(\tau, 0) \\ B_{\vec{s}}(\tau, 0)G_{\vec{s}}(0) & G_{\vec{s}}(\tau) \end{pmatrix} \quad (73)$$

The diagonal terms on the right hand side of the above equation correspond to the desired equal time Green functions. The off-diagonal terms are nothing but the time displaced Green functions which will be discussed in the next section. To evaluate the left hand side of the above equation, we first have to bring $B_{\vec{s}}(\tau, 0)$ and $B_{\vec{s}}(\beta, \tau)$ in UDV forms. This has to be done step by step so as to avoid mixing large and small scales. Consider the propagation $B_{\vec{s}}(\tau, 0)$, and a time interval τ_1 , with $n\tau_1 = \tau$, for which the different scales in $B_{\vec{s}}(n\tau_1, (n-1)\tau_1)$ do not exceed machine precision. Since $B_{\vec{s}}(\tau, 0) = B_{\vec{s}}(n\tau_1, (n-1)\tau_1) \cdots B_{\vec{s}}(\tau_1, 0)$ we can evaluate $B_{\vec{s}}(\tau, 0)$ for $n = 2$ with:

$$B_{\vec{s}}(2\tau_1, \tau_1) B_{\vec{s}}(\tau_1, 0) = \underbrace{(B_{\vec{s}}(2\tau_1, \tau_1)U_1)D_1}_{U_1 D_1 V_1} V_1 = U_2 D_2 V_2 \quad (74)$$

where $V_2 = VV_1$. The parenthesis determine the order in which the matrix multiplication are to be done. In all operations, mixing of scales are avoided. After the multiplication with diagonal matrix D_1 scales are again separated with the used of the singular value decomposition.

Thus, for $B_{\vec{s}}(\tau, 0) = U_R D_R V_R$ and $B_{\vec{s}}(\beta, \tau) = V_L D_L U_L$ we have to invert:

$$\begin{aligned} & \begin{pmatrix} I & V_L D_L U_L \\ -U_R D_R V_R & I \end{pmatrix}^{-1} = \\ & \left[\begin{pmatrix} V_L & 0 \\ 0 & U_R \end{pmatrix} \underbrace{\begin{pmatrix} (V_R V_L)^{-1} & D_L \\ -D_R & (U_L U_R)^{-1} \end{pmatrix}}_{UDV} \begin{pmatrix} V_R & 0 \\ 0 & U_L \end{pmatrix} \right]^{-1} = \\ & \left[\begin{pmatrix} (V_R)^{-1} & 0 \\ 0 & (U_L)^{-1} \end{pmatrix} V^{-1} \right] D^{-1} \left[U^{-1} \begin{pmatrix} (V_L)^{-1} & 0 \\ 0 & (U_R)^{-1} \end{pmatrix} \right] \end{aligned} \quad (75)$$

In the above, all matrix multiplications are well defined. In particular, the matrix D contains only large scales since the matrices $(V_R V_L)^{-1}$ and $(U_L U_R)^{-1}$ act as a cutoff to the exponentially small scales in D_L and D_R . This method to compute Green functions is very stable and has the advantage of producing time displaced Green functions. However, it is numerically expensive since the matrices involved are twice as big as the B matrices.

Alternative methods to compute $G_{\vec{s}}(\tau)$ which involve matrix manipulations only of the size of B include:

$$\begin{aligned} (1 + B_{\vec{s}}(\tau, 0)B_{\vec{s}}(\beta, \tau))^{-1} &= (1 + U_R D_R V_R V_L D_L U_L)^{-1} \\ &= (U_L)^{-1} \underbrace{((U_L U_R)^{-1} + D_R (V_R V_L) D_L)^{-1}}_{UDV} U_R^{-1} \\ &= (V U_L)^{-1} D^{-1} (U_R U^{-1}). \end{aligned} \quad (76)$$

Again, $(U_L U_R)^{-1}$ acts as a cutoff to the small scales in $D_R(V_R V_L)D_L$ so that D contains only large scales.

The accuracy of both presented methods may be tested by in the following way. Given the Green function at time τ we can can upgrade and wrap (see Eq. (67)) this Green function to time slice $\tau + \tau_1$. Of course, for the time interval τ_1 the involved scales should lie within the accuracy of the the computer, $\sim 1^{-12}$ for double precision numbers. The thus obtained Green function at time $\tau + \tau_1$ may be compared to the one computed from scratch using Eq. (75) or Eq. (76). For a 4×4 half-filled Hubbard model at $U/t = 4$, $\beta t = 20$, $\Delta\tau t = 0.1$ and $\tau_1 = 10\Delta\tau$ we obtain an average (maximal) difference between the matrix elements of both Green functions of 1^{-10} (1^{-6}) which is orders of magnitude smaller than the statistical uncertainty. Had we chosen $\tau_1 = 50\Delta\tau$ the accuracy drops to 0.01 and 100.0 for the average and maximal differences.

3.5 Imaginary Time Displaced Green Functions

Imaginary time displaced correlation yield important information. On one hand they may be used to obtain spin and charge gaps,^{28,29} as well quasiparticle weights.²⁰ On the other hand, with the use of the Maximum Entropy method^{30,31} dynamical properties such as spin and charge dynamical structure factors, optical conductivity, and single particle spectral functions may be computed. Those quantities offer the possibility of direct comparison with experiments, such as photoemission, neutron scattering and optical measurements.

Since there is again a Wick's theorem for time displaced correlation functions. it suffices to compute the single particle Green function for a given HS configuration. We will first start with the FTQMC and then concentrate on the PQMC.

3.5.1 FTQMC

For a given HS field, we wish to evaluate:

$$G_{\vec{s}}(\tau_1, \tau_2)_{x,y} = \langle T c_x(\tau_1) c_y^\dagger(\tau_2) \rangle_{\vec{s}} \quad (77)$$

where T corresponds to the time ordering. Thus for $\tau_1 > \tau_2$ $G_{\vec{s}}(\tau_1, \tau_2)_{x,y}$ reduces to

$$\begin{aligned} \langle c_x(\tau_1) c_y^\dagger(\tau_2) \rangle_{\vec{s}} &= \frac{\text{Tr} [U_{\vec{s}}(\beta, \tau_1) c_x U_{\vec{s}}(\tau_1, \tau_2) c_y^\dagger U_{\vec{s}}(\tau_2, 0)]}{\text{Tr} [U_{\vec{s}}(\beta, 0)]} \\ &= \frac{\text{Tr} [U_{\vec{s}}(\beta, \tau_2) U_{\vec{s}}^{-1}(\tau_1, \tau_2) c_x U_{\vec{s}}(\tau_1, \tau_2) c_y^\dagger U_{\vec{s}}(\tau_2, 0)]}{\text{Tr} [U_{\vec{s}}(\beta, 0)]} \end{aligned} \quad (78)$$

Evaluating $U^{-1}(\tau_1, \tau_2) c_x U_{\vec{s}}(\tau_1, \tau_2)$ boils down to the calculation of

$$c_x(\tau) = e^{\tau \vec{c}^\dagger A \vec{c}} c_x e^{-\tau \vec{c}^\dagger A \vec{c}}$$

where A is an arbitrary matrix. Differentiating the above with respect to τ yields

$$\frac{\partial c_x(\tau)}{\partial \tau} = e^{\tau \vec{c}^\dagger A \vec{c}} [\vec{c}^\dagger A \vec{c}, c_x] e^{-\tau \vec{c}^\dagger A \vec{c}} = - \sum_z A_{x,z} c_z(\tau).$$

Thus,

$$c_x(\tau) = (e^{-A} \vec{c})_x \quad \text{and similarly} \quad c_x^\dagger(\tau) = (\vec{c}^\dagger e^A)_x. \quad (79)$$

We can use the above equation successively to obtain:

$$\begin{aligned} U_{\vec{s}}^{-1}(\tau_1, \tau_2) c_x U_{\vec{s}}(\tau_1, \tau_2) &= (B_{\vec{s}}(\tau_1, \tau_2) \vec{c})_x \\ U_{\vec{s}}^{-1}(\tau_1, \tau_2) c_x^\dagger U_{\vec{s}}(\tau_1, \tau_2) &= (\vec{c}^\dagger B_{\vec{s}}^{-1}(\tau_1, \tau_2))_x \end{aligned} \quad (80)$$

Since B is a matrix and not an operator, we can pull it out of the trace in Eq. (78). Note that the above equation automatically leads to a Wick's theorem for time displaced Green function since the theorem holds for equal time quantities. Thus, we obtain the result:

$$G_{\vec{s}}(\tau_1, \tau_2)_{x,y} = \langle c_x(\tau_1) c_y^\dagger(\tau_2) \rangle_{\vec{s}} = B_{\vec{s}}(\tau_1, \tau_2) G_{\vec{s}}(\tau_2) \quad \tau_1 > \tau_2 \quad (81)$$

where $G_{\vec{s}}(\tau_1)$ is the equal time Green function computed previously. A similar calculation will yield for $\tau_2 > \tau_1$

$$G_{\vec{s}}(\tau_1, \tau_2)_{x,y} = -\langle c_y^\dagger(\tau_2) c_x(\tau_1) \rangle_{\vec{s}} = -(1 - G_{\vec{s}}(\tau_1)) B_{\vec{s}}^{-1}(\tau_2, \tau_1). \quad (82)$$

Returning to Eq. (73) we see that we have already computed the time displaced Green functions $G_{\vec{s}}(0, \tau)$ and $G_{\vec{s}}(\tau, 0)$ when discussing a stabilization scheme for the equal time Green functions. However, this calculation is expensive and is done only at times $\tau = n\tau_1$ where τ_1 is time scale on which all energy scales fit well on finite precision machines. To obtain the Green functions for arbitrary values of τ one uses the relations:

$$\begin{aligned} G_{\vec{s}}(0, \tau + \tau_2) &= G_{\vec{s}}(0, \tau) B_{\vec{s}}^{-1}(\tau_2, \tau) \\ G_{\vec{s}}(\tau + \tau_2, 0) &= B_{\vec{s}}(\tau_2, \tau) G_{\vec{s}}(\tau, 0) \end{aligned} \quad (83)$$

where $\tau_2 < \tau_1$.

With the above method, we have access to all time displaced Green functions $G_{\vec{s}}(0, \tau)$ and $G_{\vec{s}}(\tau, 0)$. However, we do not use translation invariance in imaginary time. Clearly, using this symmetry in the calculation of time displaced quantities will reduce the fluctuations which may sometimes be desirable. A numerically expensive but elegant way of producing all time displaced Green functions has been proposed by Hirsch.³² Let β be a multiple of τ_1 , $l\tau_1 = \beta$ and τ_1 small enough so that the scales involved in $B_s(\tau + \tau_1, \tau)$ fit on finite precision machines. Using the definition $\tau_i = i\tau_1$ with $i = 1 \dots l$ one can show:

$$\left(\begin{array}{ccccc} 1 & 0 & \cdot & 0 & B_{\vec{s}}(\tau_1, 0) \\ -B_{\vec{s}}(\tau_2, \tau_1) & 1 & 0 & \cdot & 0 \\ 0 & -B_{\vec{s}}(\tau_3, \tau_2) & 1 & \cdot & 0 \\ \cdot & 0 & -B_{\vec{s}}(\tau_4, \tau_3) & \cdot & \cdot \\ \cdot & \cdot & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & -B_{\vec{s}}(\tau_l, \tau_{l-1}) & 1 \end{array} \right)^{-1} = \left(\begin{array}{c} G_{\vec{s}}(\tau_1, \tau_1) \ G_{\vec{s}}(\tau_1, \tau_2) \dots G_{\vec{s}}(\tau_1, \tau_l) \\ G_{\vec{s}}(\tau_2, \tau_1) \ G_{\vec{s}}(\tau_2, \tau_2) \dots G_{\vec{s}}(\tau_2, \tau_l) \\ \cdot \ \cdot \ \cdot \ \cdot \\ G_{\vec{s}}(\tau_l, \tau_1) \ G_{\vec{s}}(\tau_l, \tau_2) \dots G_{\vec{s}}(\tau_l, \tau_l) \end{array} \right) \quad (84)$$

The matrix to inverse is l times the size of the B matrices, and hence expensive to compute. It is worth noting that on vector machines the performance grows with growing vector size so that the above method can become attractive. Having computed the Green functions $G_{\vec{s}}(\tau_i, \tau_j)$ we can obtain Green functions on any two time slices by using equations of the type (83).

3.5.2 PQMC

Zero temperature time displaced Green functions are given by:

$$\begin{aligned} & G_s \left(\Theta + \frac{\tau}{2}, \Theta - \frac{\tau}{2} \right)_{x,y} \\ &= \frac{\langle \Psi_T | U_{\vec{s}}(2\Theta, \Theta + \frac{\tau}{2}) c_x U_{\vec{s}}^{\dagger}(\Theta + \frac{\tau}{2}, \Theta - \frac{\tau}{2}) c_y^{\dagger} U_{\vec{s}}(\Theta - \frac{\tau}{2}, 0) | \Psi_T \rangle}{\langle \Psi_T | U_{\vec{s}}(2\Theta, 0) | \Psi_T \rangle} \\ &= \left[B_{\vec{s}} \left(\Theta + \frac{\tau}{2}, \Theta - \frac{\tau}{2} \right) G_{\vec{s}} \left(\Theta - \frac{\tau}{2} \right) \right]_{x,y} \end{aligned} \quad (85)$$

and

$$\begin{aligned} & G_s \left(\Theta - \frac{\tau}{2}, \Theta + \frac{\tau}{2} \right)_{x,y} \\ &= - \frac{\langle \Psi_T | U_{\vec{s}}(2\Theta, \Theta + \frac{\tau}{2}) c_y^{\dagger} U_{\vec{s}}^{\dagger}(\Theta + \frac{\tau}{2}, \Theta - \frac{\tau}{2}) c_x U_{\vec{s}}(\Theta - \frac{\tau}{2}, 0) | \Psi_T \rangle}{\langle \Psi_T | U_{\vec{s}}(2\Theta, 0) | \Psi_T \rangle} \\ &= - \left[\left(1 - G_{\vec{s}} \left(\Theta - \frac{\tau}{2} \right) \right) B_{\vec{s}}^{-1} \left(\Theta + \frac{\tau}{2}, \Theta - \frac{\tau}{2} \right) \right]_{x,y}. \end{aligned} \quad (86)$$

Here $\tau > 0$ and we have used Eq. (80), as well as the equal time Green function (Eq. (39)). Two comments are in order. (i) For a given value of τ the effective projection parameter is $\Theta - \tau$. Thus, before starting a simulation, one has to set the maximal value of τ which will be considered, τ_M and the effective projection parameter $\Theta - \tau_M$ should be large enough to yield the ground state within the desired precision. (ii) In a canonical ensemble, the chemical potential is meaningless. However, when single particle Green functions are computed it is required to set the reference energy with regards to which a particle will be added or removed. In other words, it is the chemical potential which delimits photoemission from inverse photoemission. Thus, it is useful to have an estimate of this quantity if single particle or pairing correlations are under investigation. For observable such as spin-spin or charge- charge time displaced correlations this complication does not come into play since they are in the particle-hole channel.

We are now left with the problem of computing the Green functions. A direct multiplication of the equal time Green function with B matrices is unstable for larger values of τ . This can be understood in the framework of free electrons on a two-dimensional square lattice:

$$H = -t \sum_{\langle \vec{i}, \vec{j} \rangle} c_{\vec{i}}^{\dagger} c_{\vec{j}}, \quad (87)$$

where the sum runs over nearest-neighbors. For this Hamiltonian one has:

$$\langle \Psi_0 | c_{\vec{k}}^{\dagger}(\tau) c_{\vec{k}} | \Psi_0 \rangle = \exp(\tau(\epsilon_{\vec{k}} - \mu)) \langle \Psi_0 | c_{\vec{k}}^{\dagger} c_{\vec{k}} | \Psi_0 \rangle, \quad (88)$$

where $\epsilon_{\vec{k}} = -2t(\cos(\vec{k}\vec{a}_x) + \cos(\vec{k}\vec{a}_y))$, \vec{a}_x, \vec{a}_y being the lattice constants. We will assume $|\Psi_0\rangle$ to be non-degenerate. In a numerical calculation the eigenvalues and eigenvectors of the above Hamiltonian will be known up to machine precision, ϵ . In the case $\epsilon_{\vec{k}} - \mu > 0$, $\langle \Psi_0 | c_{\vec{k}}^{\dagger} c_{\vec{k}} | \Psi_0 \rangle \equiv 0$. However, on a finite precision machine the later quantity will take a value of the order of ϵ . When calculating $\langle \Psi_0 | c_{\vec{k}}^{\dagger}(\tau) c_{\vec{k}} | \Psi_0 \rangle$ this roundoff error will be

blown up exponentially and the result for large values of τ will be unreliable. In Eq. (86) the B matrices play the role of the exponential factor $\exp(\tau(\epsilon_{\vec{k}} - \mu))$ and the equal time Green functions correspond to $\langle \Psi_0 | c_{\vec{k}}^\dagger c_{\vec{k}} | \Psi_0 \rangle$. In the PQMC, the stability problem is much more severe than for free electrons since the presence of the time dependent HS field mixes different scales.

We present two methods to circumvent this problem. The first method is numerically expensive but has been used extensively and hence tested for many models.²⁸ The second method has recently been developed, is elegant, simple to implement and very cheap in CPU time.³³

The first method lies on the observation that introducing a large but finite temperature $T = 1/\beta$ stabilizes the calculation. For the free electrons:

$$\langle \Psi_0 | c_{\vec{k}}^\dagger(\tau) c_{\vec{k}} | \Psi_0 \rangle = \lim_{\beta \rightarrow \infty} \frac{\exp(\tau(\epsilon_{\vec{k}} - \mu))}{1 + \exp(\beta(\epsilon_{\vec{k}} - \mu))}. \quad (89)$$

Even if the eigenvalues are known only up to machine precision, the right hand side of the above equation for large but finite values of β is a numerically stable operation. To implement this idea in the PQMC, we use the fact that the trial wave function is a single Slater determinant. Thus, we can find a single particle Hamiltonian, $H_0 = \sum_{x,y} c_x^\dagger(h_0)_{x,y} c_y$, which has $|\Psi_T\rangle$ as a non-degenerate ground state. Hence the equation

$$\begin{aligned} G_{\vec{s}}\left(\Theta - \frac{\tau}{2}\right) &\equiv \frac{\langle \Psi_T | U_{\vec{s}}(2\Theta, \Theta - \frac{\tau}{2}) c_x c_y^\dagger U_{\vec{s}}(\Theta - \frac{\tau}{2}, 0) | \Psi_T \rangle}{\langle \Psi_T | U_{\vec{s}}(2\Theta, 0) | \Psi_T \rangle} \\ &= \lim_{\beta \rightarrow \infty} \frac{\text{Tr}[e^{-\beta H_0} U_{\vec{s}}(2\Theta, \Theta - \frac{\tau}{2}) c_x c_y^\dagger U_{\vec{s}}(\Theta - \frac{\tau}{2}, 0)]}{\text{Tr}[e^{-\beta H_0} U_{\vec{s}}(2\Theta, 0)]} \\ &= \lim_{\beta \rightarrow \infty} \left[1 + B_{\vec{s}}\left(\Theta - \frac{\tau}{2}, 0\right) e^{-\beta h_0} B_{\vec{s}}\left(2\Theta, \Theta - \frac{\tau}{2}\right) \right]_{x,y} \end{aligned} \quad (90)$$

is valid and yields a link between the FTQMC and PQMC, so that at finite but large values of β we can use the methods introduced for the FTQMC to compute time displaced Green functions. In particular, we can use the relation

$$\begin{aligned} \lim_{\beta \rightarrow \infty} \begin{pmatrix} I & B_{\vec{s}}(\Theta - \frac{\tau}{2}, 0) e^{-\beta h_0} B_{\vec{s}}(2\Theta, \Theta + \frac{\tau}{2}) \\ -B_{\vec{s}}(\Theta + \frac{\tau}{2}, \Theta - \frac{\tau}{2}) & I \end{pmatrix}^{-1} = \\ \begin{pmatrix} G_{\vec{s}}(\Theta - \frac{\tau}{2}) & G_{\vec{s}}(\Theta - \frac{\tau}{2}, \Theta + \frac{\tau}{2}) \\ G_{\vec{s}}(\Theta + \frac{\tau}{2}, \Theta - \frac{\tau}{2}) & G_{\vec{s}}(\Theta + \frac{\tau}{2}) \end{pmatrix} \end{aligned} \quad (91)$$

where the Green functions refer to the PQMC Green functions of Eq. (86). To compute in a numerically stable way, similar methods to those applied for the FTQMC are used (see Eq. (75)). A convenient choice of H_0 is obtained in a basis where the trial wave function may be written as:

$$|\Psi_T\rangle = \prod_{n=1}^{N_p} \gamma_n^\dagger |0\rangle. \quad (92)$$

In this basis, we define H_0 through

$$H_0 \gamma_n^\dagger |0\rangle = \begin{cases} -\gamma_n^\dagger |0\rangle & \text{if } \gamma_n^\dagger \gamma_n |\Psi_T\rangle = |\Psi_T\rangle \\ +\gamma_n^\dagger |0\rangle & \text{if } \gamma_n^\dagger \gamma_n |\Psi_T\rangle = 0 \end{cases} \quad (93)$$

For the specific case of the Hubbard model and with the above choice of H_0 values of $\beta t \sim 40$ were well sufficient to satisfy Eq. (90) up to an average precision of 10^{-11} . Clearly this method is ad-hoc since it requires adding a *fictitious* temperature. Furthermore, since the upgrading is done with the PQMC, every time one wishes to compute time displaced correlation function all required quantities have to be computed from scratch.

A more elegant and efficient method rests on the observation that in the PQMC the Green function is a projector. Consider again the free electron case. For a non-degenerate Ground state, $\langle \Psi_0 | c_{\vec{k}}^\dagger c_{\vec{k}} | \Psi_0 \rangle = 0, 1$ so that

$$\langle \Psi_0 | c_{\vec{k}}^\dagger(\tau) c_{\vec{k}} | \Psi_0 \rangle = \left(\langle \Psi_0 | c_{\vec{k}}^\dagger c_{\vec{k}} | \Psi_0 \rangle \exp((\epsilon_{\vec{k}} - \mu)) \right)^\tau. \quad (94)$$

The above involves only well defined numerical manipulations even in the large τ limit provided that all scales fit onto finite precision machines for a unit time interval.

The implementation of this idea in the QMC algorithm is as follows. First, one has to notice that the Green function $G_{\vec{s}}(\tau)$ is a projector:

$$G_{\vec{s}}(\tau)^2 = G_{\vec{s}}(\tau). \quad (95)$$

We have already seen that for $P^\dagger B_{\vec{s}}(2\Theta, \tau) = V_L D_L U^\langle$ and $B_{\vec{s}}(\tau, 0) = U^\rangle D_R U_R$, $G_{\vec{s}}(\tau) = 1 - U^\rangle (U^\langle U^\rangle)^{-1} U^\langle$. Since

$$\left[U^\rangle (U^\langle U^\rangle)^{-1} U^\langle \right]^2 = U^\rangle (U^\langle U^\rangle)^{-1} U^\langle$$

we have:

$$G_{\vec{s}}^2(\tau) = G_{\vec{s}}(\tau) \quad \text{and} \quad (1 - G_{\vec{s}}(\tau))^2 = 1 - G_{\vec{s}}(\tau). \quad (97)$$

This property implies that $G_{\vec{s}}(\tau_1, \tau_3)$ obeys a simple composition identity

$$G_{\vec{s}}(\tau_1, \tau_3) = G_{\vec{s}}(\tau_1, \tau_2) G_{\vec{s}}(\tau_2, \tau_3). \quad (98)$$

In particular for $\tau_1 > \tau_2 > \tau_3$

$$\begin{aligned} G_{\vec{s}}(\tau_1, \tau_3) &= B_{\vec{s}}(\tau_1, \tau_3) G_{\vec{s}}^2(\tau_3) = G_{\vec{s}}(\tau_1, \tau_3) G_{\vec{s}}(\tau_3) \\ &= \underbrace{G_{\vec{s}}(\tau_1, \tau_3) B_{\vec{s}}^{-1}(\tau_2, \tau_3)}_{G_{\vec{s}}(\tau_1, \tau_2)} \underbrace{B_{\vec{s}}(\tau_2, \tau_3) G_{\vec{s}}(\tau_3)}_{G_{\vec{s}}(\tau_2, \tau_3)} \end{aligned}$$

A similar proof is valid for $\tau_3 > \tau_2 > \tau_1$

Using this composition property (98) we can break up a large τ interval into a set of smaller intervals of length $\tau = N\tau_1$ so that

$$G_{\vec{s}}\left(\Theta + \frac{\tau}{2}, \Theta - \frac{\tau}{2}\right) = \prod_{n=0}^{N-1} G_{\vec{s}}\left(\Theta - \frac{\tau}{2} + [n+1]\tau_1, \Theta - \frac{\tau}{2} + n\tau_1\right) \quad (99)$$

The above equation is the generalization of Eq. (94). If τ_1 is *small* enough each Green function in the above product is accurate and has matrix elements bounded by order unity. The matrix multiplication is then numerically well defined.

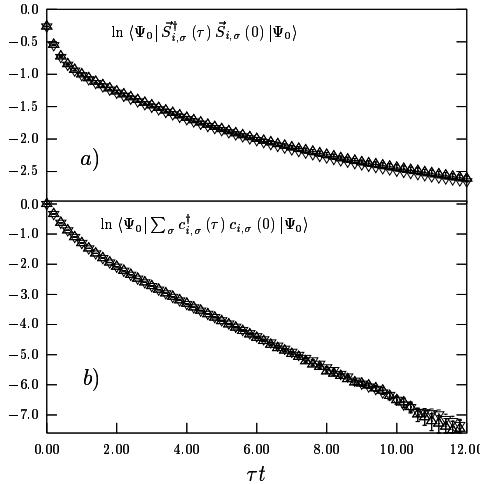


Figure 8. Imaginary time displaced on-site spin-spin (a) and Green function (b) correlation function. We consider a 6×6 lattice at half-filling and $J/t = 1.2$. In both (a) and (b) results obtained from Eq. (99) (\triangle) and (91) (∇) are plotted.

We conclude this section by comparing both presented approaches for the calculation of time displaced correlation functions in the PQMC. We consider the special case of the Kondo lattice model (see Fig. 8). As apparent the results are identical within error-bars. The important point however, is that the method based on Eq. (99) is for the considered case an order of magnitude quicker in CPU time than the method based on Eq. (91).

3.6 The Sign Problem

The sign problem remains the central challenge in QMC simulations of correlated electron systems. We have shown how it appears in the framework of the World Line algorithm for the case of free fermions with nearest and next nearest neighbor hopping. Since in the PQMC and FTQMC the one body terms are treated exactly the sign problem does not occur in this simple example. However, interactions will in many cases lead to a sign problem. Insight into the origin of the sign problem in the PQMC and FTQMC is obtained through the work of Fahy and Hamann^{34,35} which we will briefly review.

The starting point is to rewrite the imaginary time propagation of the trial wave function as diffusion equation in the space of Slater determinants:

$$\sum_{\vec{s}} U_{\vec{s}}(2\Theta, 0) |\Psi_T\rangle = \int d\Psi f(\Psi, 2\Theta) |\Psi\rangle \quad (100)$$

Since $U_{\vec{s}}(2\Theta, 0)$ is a one-body propagator in a time dependent external field, $U_{\vec{s}}(2\Theta, 0) |\Psi_T\rangle$ is a single Slater determinant (See Eq. (33)) so that the sum over the HS fields may be replaced by a weighted average over normalized Slater determinants $|\Psi\rangle$. Fahy and Hamann,^{34,35} have shown that $f(\Psi, \tau)$ satisfies a diffusion type equation in the space of Slater determinants which is invariant under parity $|\Psi\rangle \rightarrow -|\Psi\rangle$. Thus, the solutions of the diffusion equations may be classified according to this symmetry:

$f^+(\Psi, \tau) = f^+(-\Psi, \tau)$ and $f^-(\Psi, \tau) = -f^-(-\Psi, \tau)$. The reader will convince himself that only f^- solutions contribute to the integral over Slater determinants. However, the integrand, $f(\Psi, \tau)$, is exponentially dominated by the even parity solutions $f^+(\Psi, \tau)$ of the diffusion equation.^{34,35} Hence, the relevant odd parity solutions are exponentially damped in comparison to the even parity solutions. This leads to exponential increase of the signal to noise ratio.

Based on the above analysis, Fahy and Hamann have proposed a solution to the sign problem, which is referred to as positive projection.³⁴ Let us start the diffusion process with the trial wave function $|\Psi_T\rangle$ which we require - without loss of generality - to have positive overlap with the ground state: $\langle\Psi_0|\Psi_T\rangle > 0$. At infinitesimal time $\tau = \epsilon$ a population of Slater determinants $\{f(\Psi, \epsilon)|\Psi\rangle\}$ with $f(\Psi, 0) = \delta(\Psi - \Psi_T)$ is obtained. If one of those Slater determinants, $|\Psi_1\rangle$ at time $\tau = \epsilon$, is orthogonal to the ground state, $\langle\Psi_0|\Psi_1\rangle = 0$, it may be discarded since it will provide no information on the ground state at all times $\tau > \epsilon$. This may easily be seen since starting from $|\Psi_1\rangle$ the population of Slater determinants generated by the diffusion equation has zero overlap with the ground state:

$$\langle\Psi_0| \int d\Psi f(\Psi, \tau)|\Psi\rangle = \langle\Psi_0|e^{-(\tau-\epsilon)H}|\Psi_1\rangle = 0, \quad (101)$$

since $f(\Psi, \tau = \epsilon) = \delta(\Psi - \Psi_1)$ and $\langle\Psi_0|\Psi_1\rangle = 0$. The important point is that computed analytically the ensemble of Slater determinants originating from Ψ_1 cancel. It is the stochastic cancellation of those Slater determinants which lies at the origin of the sign problem. The above procedure is repeated iteratively and at each infinitesimal time step Slater determinants with $\langle\Psi_0|\Psi\rangle = 0$ are discarded. This procedure eliminates the exponential growth of the signal to noise ratio.

The above procedure is exact provided we know a-priori the surface \mathcal{N} defined by: $\langle\Psi_0|\Psi\rangle = 0$. This information is in general not available. Thus, the positive projection scheme has to be implemented based on an approximate knowledge of \mathcal{N} . An implementation of this scheme has been carried out by^{36,37} in an algorithm which is referred to as the constrained path QMC (CPQMC). However, a major drawback is that the surface \mathcal{N} in the CPQMC is approximated by a single Slater determinant corresponding to a mean-field solution of the model under consideration. In contrast, the Green function method has the ability to incorporate optimized correlated wave functions as starting point for the fixed-node approximation.

There are nevertheless a set of problems where the sign problem may be avoided in FTQMC and PQMC. As we will see in the subsequent section, particle-hole symmetry allows us to avoid the sign problem. Furthermore, models with attractive interactions which couple independently to an internal symmetry with an even number of states lead to weights, for a given HS configuration, which are an even power of a single determinant. If the determinant itself is real (i.e. absence of magnetic fields), the overall weight will be positive. An example is the attractive Hubbard model.

4 Application of the Auxiliary Field QMC to Specific Hamiltonians

In this section, we will concentrate on applications of the FTQMC and PQMC. Both approaches involve a computational effort which scales as the cubed of the volume. This makes it hard to achieve very large lattice sizes. Hence the importance of reducing finite

size effects. Below we will describe a method which turns out to be extremely efficient for the case of free electrons where size effects are known to be severe.³⁸ We then show how to apply the FTQMC and PQMC to the attractive and repulsive Hubbard models, the Kondo-lattice model, hard core bosons and the Heisenberg model. It is beyond the scope of this article to review in detail the physics related to the models. We will concentrate primarily on the technical aspects. In most cases we have the two-dimensional case in mind, the generalization to higher or lower dimensions being straightforward.

4.1 Size Effects

Size effects become particularly severe when the ground state turns out to be a metallic state with large coherence temperature. On the other hand, insulators are characterized by the localization of the wave function and are hence rather insensitive to boundary conditions on finite sized systems. It thus becomes apparent, that the worst case scenario for severe size effects are just free electrons in a tight binding approximation:

$$H = -t \sum_{\langle \vec{i}, \vec{j} \rangle} c_{\vec{i}}^\dagger c_{\vec{j}} + \text{H.c.} \quad (102)$$

In many cases before turning on the interaction which will automatically restrict the size of the lattice under consideration it is important to control size effects for this simple case. We will concentrate on the two dimensional case on a torus geometry which for the above model reduces to imposing periodic boundary conditions: $c_{\vec{i}+L\vec{e}_x}^\dagger = c_{\vec{i}}^\dagger$, $c_{\vec{i}+L\vec{e}_y}^\dagger = c_{\vec{i}}^\dagger$ where L is the linear length of the lattice lying in the \vec{e}_x, \vec{e}_y plane.

To reduce size effects on thermodynamic quantities one may in principle consider the Hamiltonian:

$$H(L) = \sum_{\langle \vec{i}, \vec{j} \rangle} t_{\vec{i}, \vec{j}}(L) c_{\vec{i}}^\dagger c_{\vec{j}} + \text{H.c.} \quad (103)$$

where $t_{\vec{i}, \vec{j}}(L)$ are arbitrary hopping parameters which have to satisfy

$$\lim_{L \rightarrow \infty} t_{\vec{i}, \vec{j}}(L) = -t. \quad (104)$$

Clearly this choice of hopping matrix elements on finite lattices will break the lattice symmetry. This is a price we are willing to pay provided that the convergence as a function of system size of thermodynamics quantities is greatly improved. Eq. (104) nevertheless guarantees that in the thermodynamic limit this symmetry is restored. To determine the hopping matrix elements $t_{\vec{i}, \vec{j}}(L)$ so as to reduce size effects on say the specific heat, $C_v(L, T) = \frac{\partial E(L)}{\partial T}$, one may minimize

$$\chi^2 = \sum_T [C_v(L, T) - C_v(L = \infty, T)]^2 \quad (105)$$

where the sum extends over a given range of temperatures. Taking into account only amplitude modulations of the hopping matrix elements leads already to a cumbersome minimization problem which does not provide satisfactory results.

Instead of carrying out a complicated minimization problem we can try to guess which matrix elements $t_{\vec{i}, \vec{j}}(L)$ will minimize size effects. It turns out that introducing a magnetic

field produces remarkable results. The magnetic field is introduced via the Peirls phase factors:

$$H(L) = -t \sum_{\langle \vec{i}, \vec{j} \rangle} e^{\frac{2\pi i}{\Phi_0} \int_{\vec{i}}^{\vec{j}} \vec{A}_L(\vec{l}) \cdot d\vec{l}} c_{\vec{i}}^\dagger c_{\vec{j}} + \text{H.c.} \quad (106)$$

with $\vec{B}_L(\vec{x}) = \vec{\nabla} \times \vec{A}_L(\vec{x})$ and Φ_0 the flux quanta. The torus geometry imposes restrictions on the \vec{B}_L field. Since, a translation in the argument of the vector potential may be absorbed in a gauge transformation:

$$\begin{aligned} \vec{A}_L(\vec{x} + L\vec{e}_x) &= \vec{A}_L(\vec{x}) + \vec{\nabla}\chi_x(\vec{x}), \\ \vec{A}_L(\vec{x} + L\vec{e}_y) &= \vec{A}_L(\vec{x}) + \vec{\nabla}\chi_y(\vec{x}), \end{aligned} \quad (107)$$

we chose, the boundary condition

$$c_{\vec{i}+L\vec{e}_x}^\dagger = e^{\frac{2\pi i}{\Phi_0} \chi_x(\vec{i})} c_{\vec{i}}^\dagger, \quad c_{\vec{i}+L\vec{e}_y}^\dagger = e^{\frac{2\pi i}{\Phi_0} \chi_y(\vec{i})} c_{\vec{i}}^\dagger \quad (108)$$

to satisfy the requirement:

$$[H(L), T_{L\vec{e}_x}] = [H(L), T_{L\vec{e}_y}] = 0. \quad (109)$$

Here, $T_{\vec{x}}$ corresponds to a translation by \vec{x} . However, magnetic translation operators belong to the magnetic algebra:³⁹

$$T_{L\vec{e}_x} T_{L\vec{e}_y} = e^{-i2\pi \frac{(L\vec{e}_x \times L\vec{e}_y) \cdot \vec{B}}{\Phi_0}} T_{L\vec{e}_y} T_{L\vec{e}_x}. \quad (110)$$

Thus, to obtain a single valued wave function the condition of flux quantization has to be satisfied: $\frac{(L\vec{e}_x \times L\vec{e}_y) \cdot \vec{B}}{\Phi_0} = n$ where n is an integer. Here, we consider a static magnetic field running along the z -axis perpendicular to the x, y plane in which the lattice lies. Hence, the smallest magnetic field which we can consider on a given lattice size satisfies:

$$\frac{BL^2}{\Phi_0} = 1. \quad (111)$$

With this choice of magnetic field and associated vector potential Eq. (104) holds.

To illustrate the reduction of size effects caused by the inclusion of the magnetic field, we first consider the single particle density of states. In a basis where $H(L)$ is diagonal, $H(L) = \sum_{n=1}^N \epsilon_n \gamma_n^\dagger \gamma_n$ with $c_n^\dagger = \sum_m \gamma_m^\dagger U_{m,n}^\dagger$ and $U^\dagger U = I$, the local density of states reads:

$$N(r, \omega) = \text{Im} \sum_n^N \frac{|U_{n,r}|^2}{\epsilon_n - \omega - i\delta} \quad (112)$$

where δ is a positive infinitesimal and N the total number of sites. Since the magnetic field breaks translation invariance (it is the site dependent vector potential which enters the Hamiltonian) $N(r, \omega)$ is site dependent. Averaging over sites yields the density of states $N(\omega)$ plotted in Fig. 9. As apparent, without the magnetic field and up to $L = 32$, $N(\omega)$ is dominated by size effects for the considered value of $\delta = 0.01t$. In contrast, the presence of the magnetic field provides remarkable improvements. In particular the van-Hove singularity is well reproduced already on $L = 16$ lattices and at $L = 32$ the result is next to exact for the considered value of δ . It is instructive to look at the $L = 8$ case

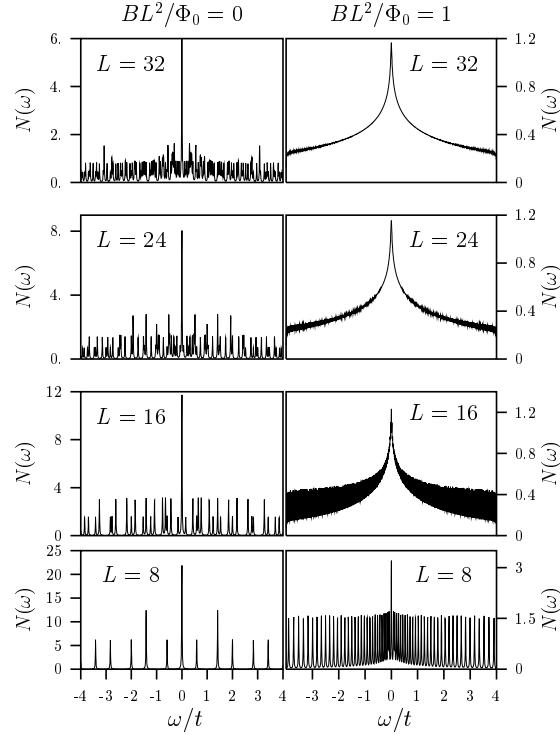


Figure 9. Density of states $N(\omega) = \frac{1}{N} \sum_r N(r, \omega)$ with (right column) and without (left column) magnetic field. Here, we consider $\delta = 0.01t$.

with and without magnetic fields. When B is turned on, the degeneracy of levels is lifted. Each level - apart from the $\epsilon_n = 0$ level which is two fold degenerate - is nondegenerate. This is precisely what one expects for Landau levels which have degeneracy $L^2 B / \Phi_0$ which is unity in our case. This provides an intuitive understanding of why the method works so well. Since each level becomes singly degenerate, the single particle states cover homogeneously the the energy range of the band-width. Clearly this can only be achieved by breaking the lattice symmetry on finite sized systems.

We now turn our attention to the specific heat coefficient $\gamma = C_v/T$ (see Fig. 10 a, b). As apparent, for a given system size, the inclusion of the magnetic field provides a gain of more than one order of magnitude in the temperature scale at which size effects set in. In particular the $\ln(1/T)$ behavior of γ due the the van-Hove singularity becomes apparent already on $L = 6$ lattices.

Upon inspection a similar improvement is obtained for the spin susceptibility (see Fig. 10 c, d). Note that since we are dealing with free electrons the charge and spin susceptibilities are identical.

One crucial question is whether the magnetic field will introduce a sign problem in the numerical simulations. It turns out that in some non-trivial cases it does not and we can hence benefit from the observed dramatic reduction in size effects. This method has been

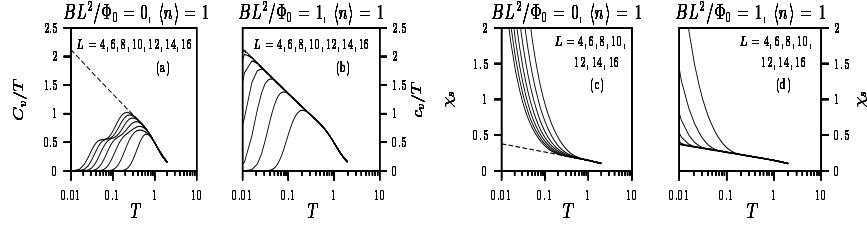


Figure 10. Specific and spin susceptibility versus temperature without (a,c) and with (b,d) magnetic field. The curves from right to left correspond to increasingly large lattices as denoted in the figure. The dashed line corresponds to the exact result.

used successfully in a depleted Kondo lattice model and has opened a whole temperature range where coherence effects may be studied without suffering from size effects.³⁸

Other schemes have been proposed to reduce size effects. In particular, averaging over boundary conditions has been suggested.^{40,41} This method has the advantage of not breaking translation symmetry. However, the averaging requires several simulations and is hence computationally expensive. In contrast with the presented method the improvement in reduction of size effects is obtained within a single simulation.

4.2 The Hubbard Model

The Hubbard model in a magnetic field is given by:

$$H_U = -t \underbrace{\sum_{\langle \vec{i}, \vec{j} \rangle, \sigma} e^{\frac{2\pi i}{\Phi_0} \int_{\vec{i}}^{\vec{j}} \vec{A} \cdot d\vec{l}} c_{\vec{i}, \sigma}^\dagger c_{\vec{j}, \sigma}}_{= \sum_{\vec{i}, \vec{j}, \sigma} c_{\vec{i}, \sigma}^\dagger h_t(\vec{A})_{\vec{i}, \vec{j}} c_{\vec{j}, \sigma}} + U \sum_{\vec{i}} (n_{\vec{i}, \uparrow} - 1/2) (n_{\vec{i}, \downarrow} - 1/2) - \mu \sum_{\vec{i}} n_{\vec{i}}. \quad (113)$$

We start by considering the $SU(2)$ invariant HS decoupling of Eq. (25). Since the HS field couples equivalently to up and down spins we obtain for the FTQMC:

$$\text{Tr} \left[e^{-\beta(H - \mu N)} \right] = \sum_{\vec{s}} C_{\vec{s}} \det (1 + B_{\vec{s}}(\beta, 0))^2 \quad \text{with} \quad (114)$$

$$C_{\vec{s}} = \frac{1}{2^N} e^{\beta UN/4 - i\alpha \sum_{\vec{i}, n} s_{\vec{i}, n}}, \quad B(\beta, 0) = \prod_{n=1}^m e^{h_I(\vec{s}_n)} e^{-\Delta\tau(h_I(\vec{A}) - \mu)}.$$

Here, $h_I(\vec{s}_n)_{\vec{i}, \vec{j}} = \delta_{\vec{i}, \vec{j}} i\alpha s_{\vec{i}, n}$ and $\cos(\alpha) = \exp(-\Delta\tau U/2)$.

For the PQMC a trial wave function which factorizes in spin up and down sectors is usually chosen:

$$|\Psi_T\rangle = |\Psi_T^\uparrow\rangle \otimes |\Psi_T^\downarrow\rangle, \quad |\Psi_T^\sigma\rangle = \prod_{y=1}^{N_p^\sigma} (\bar{c}_\sigma^\dagger P^\sigma)_y |0\rangle. \quad (115)$$

Here, $N_p^\uparrow = N_p^\downarrow$ are the number of particles in the spin up and down sectors. Typically, the trial wave function is chosen to be the solution of the non-interacting system. Hence,

$$P^\uparrow = P^\downarrow = P \quad (116)$$

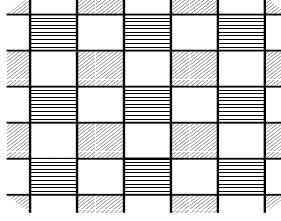


Figure 11. Checkerboard decomposition for the two dimensional square lattice with nearest neighbor hopping. $h_t^{(1)}$ ($h_t^{(2)}$) contains hopping processes along the boundaries of the horizontally (diagonally) shaded squares.

so that:

$$\langle \Psi_T | e^{2\Theta H} | \Psi_T \rangle = \sum_{\vec{s}} C_{\vec{s}} \det(P^\dagger B_{\vec{s}}(2\Theta, 0) P)^2. \quad (117)$$

Replacing β by 2Θ and setting $\mu = 0$ in the above definitions of $B_{\vec{s}}(2\Theta, 0)$ and $C_{\vec{s}}$ yields the form of those quantities for the PQMC.

Before discussing separately the attractive ($U < 0$) and repulsive case ($U > 0$) let us comment on the matrix multiplication $e^{-\Delta\tau h_t(\vec{A})} C$ where C is an $N \times N$ matrix. A straightforward multiplication yields a computational cost scaling as N^3 . To reduce this cost the checkerboard decomposition is often used. The matrix h_t is written as $h_t = h_t^{(1)} + h_t^{(2)}$ where $h_t^{(1)}$ and $h_t^{(2)}$ are sums of commuting four site hopping matrices: $h_t^{(1,2)} = \sum_{j=1}^{N/4} (h_t^{(1,2)})_j$ (See Fig. 11). Thus the the multiplication

$$\begin{aligned} e^{-\Delta\tau h_t} A &= e^{-\Delta\tau h_t^{(1)}} e^{-\Delta\tau h_t^{(2)}} A + \mathcal{O}(\Delta\tau^2) \\ &= \prod_{i=1}^{N/4} e^{-\Delta\tau (h_t^{(1)})_i} \prod_{j=1}^{N/4} e^{-\Delta\tau (h_t^{(2)})_j} A + \mathcal{O}(\Delta\tau^2) \end{aligned} \quad (118)$$

Since $(h_t^{(1,2)})_j$ involves hopping on four sites irrespective of the lattice size, the matrix multiplication $e^{-\Delta\tau (h_t^{(1,2)})_j} A$ scales as N . Hence the overall matrix multiplication scales as N_s^2 . Since the systematic error involved in the Trotter decomposition is already of order $\Delta\tau^2$ the above checkerboard decomposition does not introduce a new source of errors. However we have gained a power in the computational cost.

4.2.1 $U < 0$

The attractive Hubbard model has generated considerable interest since i) in two dimensions it shows a Kosterlitz-Thouless transition to an s-wave superconducting state^{42,43} and ii) in the strong coupling limit pairs form at a temperature scale roughly set by U whereas the Kosterlitz-Thouless transition is expected to scale as t^2/U . Hence, the model offers the possibility of studying the physics of a metallic state whith preformed pairs, a subject of interest in the context of high T_c superconductivity.⁴⁴

When $U < 0$, α in Eq. (114) is a pure imaginary number so that in the absence of magnetic fields, the determinant is real. Since the weight is the square of a real number it

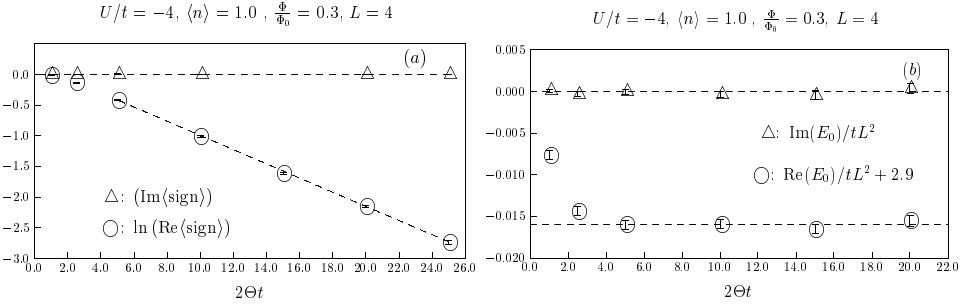


Figure 12. (a) Real and imaginary part of the average sign. Within the error-bars the imaginary part vanishes and the real part decays exponentially with the projection parameter Θ . (b) Real and imaginary part of the energy as a function of Θ . For a single HS configuration, the imaginary part of the energy is non-zero. Only after averaging does it vanish. In this case, the sign problem is not severe, since convergence to the ground state is achieved before the average sign becomes prohibitively small.

is positive and hence no sign problem occurs. In the weak coupling limit or BCS regime, the simulations perform well. However, in the strong coupling limit the method suffers from long autocorrelation times.

In the presence of a magnetic field, a sign problem occurs since the determinant becomes complex. We illustrate this point as well as the behavior of the sign problem when the product of determinants is complex by considering a vector potential $\vec{A}(\vec{x}) = \Phi \vec{e}_x / L$ at $\Phi = 0.3\Phi_0$. Such a vector potential introduces a twist in the boundary conditions which corresponds to nothing but threading a magnetic field with flux Φ through the torus on which lies the electronic system. This may be used to probe for flux quantization as well as a Kosterlitz-Thouless transition.⁴³ Note that insulating states may equally be determined by such a construction.^{45,25} For this choice of \vec{A} , the fermionic determinant is complex. Hence in the framework of the PQMC and following Eq. (13), observables are evaluated with:

$$\langle O \rangle = \frac{\sum_{\vec{s}} \overline{\text{Pr}}_{\vec{s}} \text{sign}(\vec{s}) \langle O \rangle_{\vec{s}}}{\sum_{\vec{s}} \overline{\text{Pr}}_{\vec{s}} \text{sign}(\vec{s})} \quad \text{where} \quad (119)$$

$$\text{sign}(\vec{s}) = \frac{\det(P^\dagger B_{\vec{s}}(2\Theta, 0) P)^2}{|\det(P^\dagger B_{\vec{s}}(2\Theta, 0) P)|}, \quad \overline{\text{Pr}}_{\vec{s}} = \frac{|\det(P^\dagger B_{\vec{s}}(2\Theta, 0) P)|^2}{\sum_{\vec{s}} |\det(P^\dagger B_{\vec{s}}(2\Theta, 0) P)|^2}.$$

For a given HS configuration $\text{sign}(\vec{s}) = e^{-i\phi(\vec{s})}$. After summation over the HS fields, the average sign, $\langle \text{sign} \rangle = \sum_{\vec{s}} \overline{\text{Pr}}_{\vec{s}} \text{sign}(\vec{s})$ is a real quantity. This follows from

$$\langle \text{sign} \rangle = \frac{\langle \Psi_T | e^{-2\Theta H} | \Psi_T \rangle}{\sum_{\vec{s}} |\det(P^\dagger B_{\vec{s}}(2\Theta, 0) P)|^2} + \mathcal{O}(\Delta\tau^2) \quad (120)$$

and the fact that H is hermitian so that the numerator is a real quantity.^e This property gives us a nice internal check for the validity of Monte Carlo sampling. As shown in Fig. (12) the real part of the average sign decays exponentially with the projection parameter Θ .

^eThis property equally holds for finite values of $\Delta\tau$ since the kinetic and potential terms in H as well as the trial wave function are real representable in Fourier space.

4.2.2 $U > 0$

The repulsive Hubbard model is given by Eq. (113) with $U > 0$. At half-filling the Hubbard model describes a Mott insulator with long range antiferromagnetic order. The nature of the metallic state in the vicinity of the Mott insulator as well as the metal-insulator transition as a function of doping has attracted considerable recent interest in particular in conjunction with high T_c superconductors.⁴⁶

For $U > 0$, α in Eq. (114) is a real number so that the determinant is a complex quantity and hence a sign problem occurs. In the special case of half-filling, $\mu = 0$ in Eq. (113), the model is particle-hole symmetric. This symmetry allows us to avoid the sign-problem. The particle-hole transformation reads.

$$\mathcal{P}^{-1} c_{\vec{i}, \sigma}^\dagger \mathcal{P} = (-1)^{i_x + i_y} c_{\vec{i}, \sigma}^\dagger. \quad (121)$$

where $\vec{i} = (i_x, i_y)$. Note that the Hamiltonian (113) at $\mu = 0$ is invariant under a combined time reversal and particle-hole transform due to the presence of the vector potential. Using the notation of Eq. (114):

$$\begin{aligned} \mathcal{P}^{-1} H_t \mathcal{P} &\equiv \mathcal{P}^{-1} \vec{c}_\sigma^\dagger h_t(\vec{A}) \vec{c}_\sigma \mathcal{P} = \vec{c}_\sigma^\dagger \overline{h_t(\vec{A})} \vec{c}_\sigma \text{ and} \\ \mathcal{P}^{-1} H_I(\vec{s}_n) \mathcal{P} &\equiv \mathcal{P}^{-1} \vec{c}_\sigma^\dagger h_I(\vec{s}_n) \vec{c}_\sigma \mathcal{P} = \sum_{\vec{i}} i \alpha s_{\vec{i}, n} + \vec{c}_\sigma^\dagger \overline{h_I(\vec{s}_n)} \vec{c}_\sigma. \end{aligned} \quad (122)$$

Thus

$$\begin{aligned} \det(1 + B_{\vec{s}}(\beta, 0)) &= \text{Tr} \left[\prod_{n=1}^m e^{H_I(\vec{s}_n)} e^{-\Delta \tau H_t} \right] \\ &= \text{Tr} \left[\prod_{n=1}^m e^{\mathcal{P}^{-1} H_I(\vec{s}_n) \mathcal{P}} e^{-\Delta \tau \mathcal{P}^{-1} H_t \mathcal{P}} \right] \\ &= e^{\sum_{\vec{i}} i \alpha s_{\vec{i}, n}} \det(1 + \prod_{n=1}^m e^{\overline{h_I(\vec{s}_n)}} e^{-\Delta \tau \overline{h_t(\vec{A})}}) \\ &= e^{\sum_{\vec{i}} i \alpha s_{\vec{i}, n}} \overline{\det(1 + B_{\vec{s}}(\beta, 0))} \end{aligned} \quad (123)$$

and hence $C_{\vec{s}} \det(1 + B_{\vec{s}}(\beta, 0))^2$ (see Eq. (114)) is a positive quantity even in the presence of a magnetic field. More generally, the above is valid for a half-filled Hubbard model on bipartite lattice^f with inter-sublattice single-electron hopping. Thus, and for example, simulations on a half-filled honnneycomb lattice are sign free and show Mott semi-metal-insulator transitions.⁴⁷

Away from half-filling the sign problem sets in and turns out to be rather severe with the use of the HS transformation of Eq. (25). It is then more efficient to consider the HS which couples to the magnetization (24) to obtain:

$$\begin{aligned} \text{Tr} e^{-\beta(H - \mu N)} &= C \sum_{\vec{s}} \det \left(1 + B_{\vec{s}}^\uparrow(\beta, 0) \right) \det \left(1 + B_{\vec{s}}^\downarrow(\beta, 0) \right) \text{ with} \\ C &= e^{\beta U N / 4} 2^N, \quad B_{\vec{s}}^\sigma(\beta, 0) = \prod_{n=1}^m e^{h_I^\sigma(\vec{s}_n)} e^{-\Delta \tau (h_t(\vec{A}) - \mu)} \end{aligned} \quad (124)$$

^fThe lattice sites of a bipartite lattice may be split into two sublattices, A and B such that the nearest neighbors of any site in A belong to B .

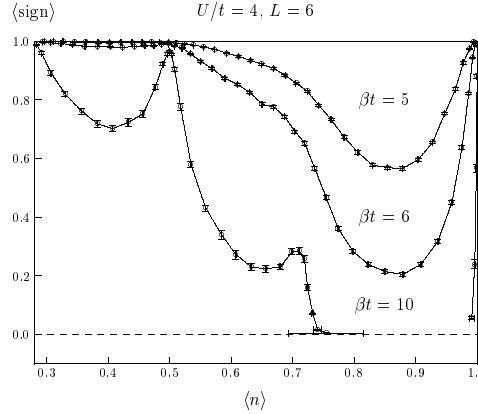


Figure 13. Average sign versus band filling $\langle n \rangle$ on a 6×6 lattice at $U/t = 4$. Filled shells are present at $\langle n \rangle = 26/36$ and $\langle n \rangle = 18/36$. As apparent for those band fillings, the decay of the average sign is slow.

with $(h_I^\sigma(\vec{s}_n))_{\vec{i},\vec{j}} = \delta_{\vec{i},\vec{j}} \sigma \tilde{\alpha} s_{\vec{i},n}$. In a very similar manner as above, one will show that for the particle-hole symmetric case the sign problem does not occur since: $\det(1 + B_{\vec{s}}^\uparrow(\beta, 0)) = e^{\sum_{\vec{i},\sigma} \tilde{\alpha} s_{\vec{i},n}} \det(1 + B_{\vec{s}}^\downarrow(\beta, 0))$. Fig. 13 plots the average sign as a function of electronic density for the HS transformation (24). As apparent, it is most severe at low dopings. In general, when the average sign drops below 0.1 accurate simulations become prohibitively expensive. One will notice that the average sign decays more slowly for special band fillings. Those band fillings correspond to filled shells for which the solution of the non-interacting system is non-degenerate. The dependence of the average sign on different choices of the HS decoupling has been considered in.⁴⁸

For the PQMC, we again assume that the trial wave function factorizes in spin up and spin down sectors (115). With the HS transformation (24) we obtain:

$$\langle \Psi_T | e^{-2\Theta H} | \Psi_T \rangle = C \sum_{\vec{s}} \prod_{\sigma} \det(P^{\sigma,\dagger} B_{\vec{s}}^\sigma(2\Theta, 0) P^\sigma) \quad (125)$$

with similar definitions for $B_{\vec{s}}^\sigma$ as for the FTQMC. At half-filling, the sign problem may again be avoided provided that the trial wave function is appropriately chosen. It is convenient to require the trial wave function to be a non-degenerate ground state of a given single-particle Hamiltonian $H_0 = H_0^\uparrow + H_0^\downarrow$ with,

$$H_0^\sigma = \sum_{\sigma} \vec{c}_{\sigma}^\dagger T_0^\sigma \vec{c}_{\sigma}. \quad (126)$$

Using Eq. (90) to relate the PQMC to the FTQMC and following Eq. (123) the sign problem is absent provided that:

$$P^{-1} \vec{c}_{\uparrow}^\dagger T_0^\uparrow c_{\uparrow} P = \vec{c}_{\uparrow}^\dagger \overline{T_0^\uparrow} \vec{c}_{\uparrow} + C \quad (127)$$

where C is a constant.

The choice of the trial wave function is important. Quick convergence as a function of projection parameter Θ is the measure for the quality of the trial wave function. Away from half-band filling when the sign problem sets in, it is important to reach convergence

quickly before the average sign becomes prohibitively small. In the PQMC, different symmetry sectors may be probed independently by judiciously choosing the trial wave function. An example where symmetry considerations are important is the half-filled Hubbard model. The ground state at this band-filling has long-range antiferromagnetic order. This continuous symmetry breaking is accompanied by gapless spin excitations: spin waves. On finite size lattices, however, spin symmetry is not broken, and the ground state is a spin singlet with an energy gap to spin excitations set by v_s/L where v_s is the spin-wave velocity and L the linear size of the lattice. Choosing a spin-singlet trial wave function is a good starting point since it is orthogonal to the low lying spin excitations. Hence, they do not have to be filtered out at great cost (see Fig. 14). To generate a spin singlet trial wave function we consider the non-interacting Hamiltonian:

$$H_0 = \sum_{\langle \vec{i}, \vec{j} \rangle, \sigma} \left(t_{\vec{i}, \vec{j}} c_{\vec{i}, \sigma}^\dagger c_{\vec{j}, \sigma} + \text{H.c.} \right) \quad (128)$$

where $\langle \vec{i}, \vec{j} \rangle$ is a sum over nearest neighbors and

$$t_{\vec{i}, \vec{i} + \vec{a}_x} = \begin{cases} -t(1 + \delta) & \text{if } i_x = 2n + 1 \\ -t(1 - \delta) & \text{if } i_x = 2n \end{cases}, \quad t_{\vec{i}, \vec{i} + \vec{a}_y} = -t(1 + \delta) \quad (129)$$

with $\delta \ll t$. The dimerization δ leads to a non-degenerate ground state at half band-filling and hence to a spin singlet ground state which we use as trial wave function $|\Psi_T\rangle$:

$$\vec{S}^2 |\Psi_T\rangle = 0, \quad \text{with} \quad \vec{S} = \sum_{\vec{i}} \vec{S}_{\vec{i}} \quad (130)$$

and $\vec{S}_{\vec{i}}$ is the spin operator on site \vec{i} . This trial wave function was used to produce the data of Fig. 6.

Unrestricted Hartree Fock solutions may be used as trial wave functions. In contrast to the above, those trial wave functions have overlaps with all symmetry sectors. However, this choice of trial wave function optimizes the overlap with the ground state. At finite dopings, those trial wave functions have been discussed in.⁴⁹

4.3 Periodic Anderson and Kondo Lattice Models

The Kondo lattice model (KLM) as well as the periodic Anderson model (PAM) are the prototype Hamiltonians to describe heavy fermion materials⁵⁰ and Kondo insulators.⁵¹ The physics under investigation is that of a lattice of magnetic impurities embedded in a metallic host. The symmetric PAM reads:

$$\begin{aligned} H_{PAM} = & \sum_{\vec{k}, \sigma} \varepsilon(\vec{k}) c_{\vec{k}, \sigma}^\dagger c_{\vec{k}, \sigma} - V \sum_{\vec{i}, \sigma} \left(c_{\vec{i}, \sigma}^\dagger f_{\vec{i}, \sigma} + f_{\vec{i}, \sigma}^\dagger c_{\vec{i}, \sigma} \right) \\ & + U_f \sum_{\vec{i}} \left(n_{\vec{i}, \uparrow}^f - 1/2 \right) \left(n_{\vec{i}, \downarrow}^f - 1/2 \right). \end{aligned} \quad (131)$$

The unit cell, denoted by \vec{i} , contains an extended and a localized orbitals. The fermionic operators $c_{\vec{k}, \sigma}^\dagger$ ($f_{\vec{k}, \sigma}^\dagger$) create electrons on extended (localized) orbitals with wave vector \vec{k} and z -component of spin σ . The overlap between extended orbitals generates a conduction

band with dispersion relation $\varepsilon(\vec{k})$. There is a hybridization matrix element, V , between both orbitals in the unit-cell and the Coulomb repulsion- modeled by a Hubbard U_f - is taken into account on the localized orbitals. In the limit of large U_f , charge fluctuations on the localized orbitals are suppressed and the PAM maps onto the KLM:⁵²

$$H_{KLM} = \sum_{\vec{k},\sigma} \varepsilon(\vec{k}) c_{\vec{k},\sigma}^\dagger c_{\vec{k},\sigma} + J \sum_{\vec{i}} \vec{S}_{\vec{i}}^c \cdot \vec{S}_{\vec{i}}^f. \quad (132)$$

Here $\vec{S}_{\vec{i}}^c = \frac{1}{2} \sum_{s,s'} c_{\vec{i},s}^\dagger \vec{\sigma}_{s,s'} c_{\vec{i},s'}$, where $\vec{\sigma}$ are the Pauli $s = 1/2$ matrices. A similar definition holds for $\vec{S}_{\vec{i}}^f$. A magnetic energy scale $J = 8V^2/U$ emerges and there is a constraint of one electron per localized orbital.

It is beyond the scope of this review to discuss the physics of the PAM and KLM. The interested reader is referred to.⁵³ Here we concentrate only on the technical aspects involved in the simulations. Simulations of the PAM are identical to those for the repulsive Hubbard model.⁵⁴ Again, at half-filling, the sign problem is absent due to the underlying particle-hole symmetry.

Simulation of the KLM on the other hand have up to recently been plagued by the sign-problem even in the case of half-filling where the model is invariant under particle-hole transformation.^{55,56} To achieve a sign-free formulation of the problem²⁹ in the half-filled case we first consider the Hamiltonian:

$$H = \sum_{\vec{k},\sigma} \varepsilon(\vec{k}) c_{\vec{k},\sigma}^\dagger c_{\vec{k},\sigma} - \frac{J}{4} \sum_{\vec{i}} \left[\sum_{\sigma} c_{\vec{i},\sigma}^\dagger f_{\vec{i},\sigma} + f_{\vec{i},\sigma}^\dagger c_{\vec{i},\sigma} \right]^2. \quad (133)$$

With the use of the HS transformation of Eq. (25) to decouple the perfect square term a QMC algorithm may readily be formulated for the above model. Without constraints on the Hilbert space, the reader will easily persuade himself that the sign problem does not occur since we are dealing with an attractive interaction ($J > 0$) which couples to an internal symmetry – the spin – with an even number of states. We will see that the constraint of single occupancy on f -sites leads to a sign problem away from half-filling.

To see how H relates to H_{KLM} we compute the square in Eq. (133) to obtain:

$$\begin{aligned} H = & \sum_{\vec{k},\sigma} \varepsilon(\vec{k}) c_{\vec{k},\sigma}^\dagger c_{\vec{k},\sigma} + J \sum_{\vec{i}} \vec{S}_{\vec{i}}^c \cdot \vec{S}_{\vec{i}}^f \\ & - \frac{J}{4} \sum_{\vec{i},\sigma} \left(c_{\vec{i},\sigma}^\dagger c_{\vec{i},-\sigma}^\dagger f_{\vec{i},-\sigma} f_{\vec{i},\sigma} + \text{H.c.} \right) + \frac{J}{4} \sum_{\vec{i}} \left(n_{\vec{i}}^c n_{\vec{i}}^f - n_{\vec{i}}^c - n_{\vec{i}}^f \right). \end{aligned} \quad (134)$$

As apparent, there are only pair-hopping processes between the f - and c -sites. Thus the total number of doubly occupied and empty f -sites is a conserved quantity:

$$[H, \sum_{\vec{i}} (1 - n_{\vec{i},\uparrow}^f)(1 - n_{\vec{i},\downarrow}^f) + n_{\vec{i},\uparrow}^f n_{\vec{i},\downarrow}^f] = 0. \quad (135)$$

If we denote by Q_n the projection onto the Hilbert space with $\sum_{\vec{i}} (1 - n_{\vec{i},\uparrow}^f)(1 - n_{\vec{i},\downarrow}^f) + n_{\vec{i},\uparrow}^f n_{\vec{i},\downarrow}^f = n$ then:

$$HQ_0 = H_{KLM} + \frac{JN}{4} \quad (136)$$

since in the Q_0 subspace the f -sites are singly occupied and hence the pair-hopping term vanishes. Thus, it suffices to choose

$$Q_0|\Psi_T\rangle = |\Psi_T\rangle \quad (137)$$

to ensure that

$$\frac{\langle\Psi_T|e^{-\Theta H}Oe^{-\Theta H}|\Psi_T\rangle}{\langle\Psi_T|e^{-2\Theta H}|\Psi_T\rangle} = \frac{\langle\Psi_T|e^{-\Theta H_{KLM}}Oe^{-\Theta H_{KLM}}|\Psi_T\rangle}{\langle\Psi_T|e^{-2\Theta H_{KLM}}|\Psi_T\rangle}. \quad (138)$$

To obtain a trial wave function which satisfies the requirements $Q_0|\Psi_T\rangle = |\Psi_T\rangle$ we are forced to choose H_0 of the form:

$$H_0 = \sum_{(\vec{i}, \vec{j}), \sigma} \left(t_{\vec{i}, \vec{j}} c_{\vec{i}, \sigma}^\dagger c_{\vec{j}, \sigma} + \text{H.c.} \right) + h_z \sum_{\vec{i}} e^{i\vec{Q}\cdot\vec{i}} \left(f_{\vec{i}, \uparrow}^\dagger f_{\vec{j}, \uparrow} - f_{\vec{i}, \downarrow}^\dagger f_{\vec{j}, \downarrow} \right) \quad (139)$$

which generates a Néel state ($\vec{Q} = (\pi, \pi)$) on the localized orbitals, and the hopping matrix elements satisfy Eq. (129). With this choice of the trial wave function and the definition of the particle-hole transformation

$$\begin{aligned} \mathcal{P}^{-1} c_{\vec{i}, \sigma}^\dagger \mathcal{P} &= (-1)^{i_x + i_y} c_{\vec{i}, \sigma}^\dagger \\ \mathcal{P}^{-1} f_{\vec{i}, \sigma}^\dagger \mathcal{P} &= -(-1)^{i_x + i_y} f_{\vec{i}, \sigma}^\dagger \end{aligned}$$

one may readily show the absence of sign problem at half-filling.

Although attractive, the above approach turns out to be (i) numerically inefficient and (ii) restricted to the PQMC. In the half-filled case, the principle source of inefficiency lies in the coupled constraints, $Q_0|\Psi_T\rangle = |\Psi_T\rangle$ and $|\Psi_T\rangle$ is a Slater determinant factorizable in the spin indices which inhibits the choice of a spin singlet trial wave function. Since the ground state is known to be a spin singlet on finite lattices^{57,58} convergence is bad especially in the small J/t limit for which the ground state has long-range antiferromagnetic order, with small spin-wave velocity.

To alleviate both problems we relax the constraint $Q_0|\Psi_T\rangle = |\Psi_T\rangle$ and add a Hubbard term on the f -sites to the Hamiltonian:

$$\begin{aligned} H = \sum_{\vec{k}, \sigma} \varepsilon(\vec{k}) c_{\vec{k}, \sigma}^\dagger c_{\vec{k}, \sigma} - \frac{J}{4} \sum_{\vec{i}} \left[\sum_{\sigma} c_{\vec{i}, \sigma}^\dagger f_{\vec{i}, \sigma} + f_{\vec{i}, \sigma}^\dagger c_{\vec{i}, \sigma} \right]^2 \\ + U_f \sum_{\vec{i}} (n_{\vec{i}, \uparrow}^f - 1/2)(n_{\vec{i}, \downarrow}^f - 1/2) \end{aligned} \quad (140)$$

This Hamiltonian is again block diagonal in the Q_n subspaces. During the imaginary time propagation, the components $Q_n|\Psi_T\rangle$ of the trial wave function will be suppressed by a factor $e^{-\Theta U_f n/2}$ in comparison to the component $Q_0|\Psi_T\rangle$.

To incorporate the Hubbard term in the QMC simulation the HS transformation of Eq. (25) is recommended at least for the half filled case. Having relaxed the constraint $Q_0|\Psi_T\rangle = |\Psi_T\rangle$ we are now free to choose a spin singlet trial wave function which we generate from:

$$H_0 = \sum_{\vec{k}, \sigma} \varepsilon(\vec{k}) c_{\vec{k}, \sigma}^\dagger c_{\vec{k}, \sigma} - \frac{J}{4} \sum_{\vec{i}, \sigma} (c_{\vec{i}, \sigma}^\dagger f_{\vec{i}, \sigma} + f_{\vec{i}, \sigma}^\dagger c_{\vec{i}, \sigma}) \quad (141)$$

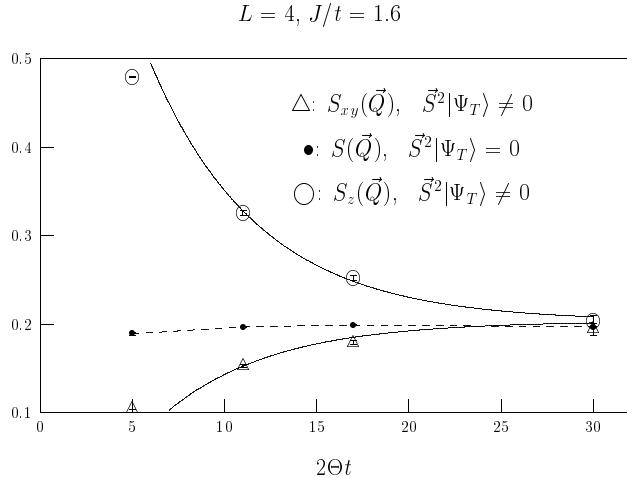


Figure 14. Spin-spin correlations as a function of the projection parameter Θ . Here, $S(\vec{Q}) = \frac{4}{3}\langle\vec{S}^f(\vec{Q}) \cdot \vec{S}^f(-\vec{Q})\rangle$, $S_z^f(\vec{Q}) = 4\langle\vec{S}_z^f(\vec{Q}) \cdot \vec{S}_z^f(-\vec{Q})\rangle$, and $S_{xy}^f(\vec{Q}) = 2\left(\langle\vec{S}_x^f(\vec{Q}) \cdot \vec{S}_x^f(-\vec{Q})\rangle + \langle\vec{S}_y^f(\vec{Q}) \cdot \vec{S}_y^f(-\vec{Q})\rangle\right)$. The trial wave function with $\vec{S}^2|\Psi_T\rangle \neq 0$ ($\vec{S}^2|\Psi_T\rangle = 0$) corresponds to the ground state of the Hamiltonian in Eq. (141) (Eq. (139)). In the *large* Θ limit, the results are independent on the choice of the trial wave function. In particular, starting from a broken symmetry state the symmetry is restored at *large* values of Θt . For this system, the spin gap is given by $\Delta_{sp} = 0.169 \pm 0.004$.²⁹ Starting with a trial wave function with $\vec{S}^2|\Psi_T\rangle \neq 0$, convergence to the ground state follows approximatively the form: $a + be^{-\Delta_{sp}2\Theta}$. The solid lines correspond to a least square fit to this form.

which is nothing but the non-interacting PAM with hybridization $V = J/4$. The ground state at half-filling is clearly a spin singlet. With this choice of the trial wave function, HS transformations of Eqn. (25), (28) as well as the particle-hole transformation of Eq. (140) the absence of sign problem in the half-filled case is readily shown for $J > 0$.

Fig. 14 demonstrates the importance of using a spin singlet trial wave function. Starting from a Néel order for the f-electrons (See Eq. (139)) convergence to the ground state follows approximatively $e^{-\Delta_{sp}2\Theta}$ where Δ_{sp} corresponds to the spin-gap. When the spin gap is small, convergence is poor and the remedy is to consider a spin singlet trial wave function.

It is equally possible to consider the ferromagnetic exchange $J < 0$. To achieve a sign free simulation at least at half-filling we define the particle hole transformation as

$$\begin{aligned} \mathcal{P}^{-1}c_{\vec{i},\sigma}^\dagger \mathcal{P} &= (-1)^{i_1+i_2} c_{\vec{i},\sigma}^\dagger \\ \mathcal{P}^{-1}f_{\vec{i},\sigma}^\dagger \mathcal{P} &= (-1)^{i_1+i_2} f_{\vec{i},\sigma}^\dagger \end{aligned}$$

in conjunction with a trial wave function generated by the non-interacting Hamiltonian

$$H_0 = \sum_{\vec{k},\sigma} \varepsilon(\vec{k}) c_{\vec{k},\sigma}^\dagger c_{\vec{k},\sigma} - \frac{J}{4} \sum_{\langle \vec{i},\vec{j} \rangle,\sigma} (c_{\vec{i},\sigma}^\dagger f_{\vec{j},\sigma} + f_{\vec{j},\sigma}^\dagger c_{\vec{i},\sigma}) \quad (142)$$

leads to a sign free simulation for the ferromagnetic half-filled case.

Having introduced the Hubbard term on the f -sites, the FTQMC may be formulated. For a given temperature U_f has to be large enough so as to eliminate double occupancy on the f -sites and hence ensure projection onto the Q_0 subspace. At this point, it is very convenient to choose the $SU(2)$ -invariant HS decomposition of Eq. (25) since one can take the limit $U_f \rightarrow \infty$ by setting $\alpha = \pi/2$. Hence irrespective of the considered temperature, we are guaranteed to be in the correct Hilbert space. Note that since the Hubbard U_f term commutes with the Hamiltonian of Eq. (133) no uncontrolled Trotter error is involved in taking the limit $U_f \rightarrow \infty$.

It is beyond the scope of this review to describe the results obtained with the above algorithm for the Kondo Lattice model and the reader is referred to the Refs.^{29,59,38}

4.4 Hard-Core Boson Systems and the Heisenberg Hamiltonian

The spin 1/2 Heisenberg model is defined by the Hamiltonian

$$H = J \sum_{\langle \vec{i}, \vec{j} \rangle} \vec{S}_{\vec{i}} \cdot \vec{S}_{\vec{j}}, \quad (143)$$

where the spin operator, $\vec{S}_{\vec{i}} = (1/2) \sum_{\sigma, \sigma'} c_{\vec{i}, \sigma}^\dagger \vec{\sigma}_{\sigma, \sigma'} c_{\vec{i}, \sigma'}$ act on the states, $|\uparrow\rangle$, $|\downarrow\rangle$ and $\vec{\sigma}$ are the Pauli spin 1/2 matrices. Defining the raising and lowering spin operators,

$$S_{\vec{i}}^+ = S_{\vec{i}}^x + i S_{\vec{i}}^y, \quad S_{\vec{i}}^- = S_{\vec{i}}^x - i S_{\vec{i}}^y \quad (144)$$

transforms the Heisenberg model to:

$$H = J \sum_{\langle \vec{i}, \vec{j} \rangle} \frac{1}{2} (S_{\vec{i}}^+ S_{\vec{j}}^- + S_{\vec{i}}^- S_{\vec{j}}^+) + S_{\vec{i}}^z S_{\vec{j}}^z. \quad (145)$$

The raising and lowering operators satisfy the commutation relations:

$$[S_{\vec{i}}^+, S_{\vec{j}}^-] = \delta_{\vec{i}, \vec{j}} 2 S_{\vec{i}}^z. \quad (146)$$

The mapping onto hard core bosons follows from the identification

$$|\uparrow\rangle \rightarrow |1\rangle, \quad |\downarrow\rangle \rightarrow |0\rangle \quad (147)$$

Hard core boson operators, b^\dagger acting on the states $|1\rangle$ and $|0\rangle$ satisfy the commutation rules:

$$[b_{\vec{i}}^\dagger, b_{\vec{j}}] = \delta_{\vec{i}, \vec{j}} 2 \left(b_{\vec{i}}^\dagger b_{\vec{i}} - \frac{1}{2} \right) \quad (148)$$

which follows from the usual bosonic commutation rules with constraint of no double occupancy: $b_{\vec{i}}^\dagger b_{\vec{i}}^\dagger = 0$. The identification of states (147) leads to:

$$S_{\vec{i}}^+ \rightarrow b_{\vec{i}}^\dagger, \quad S_{\vec{i}}^- \rightarrow b_{\vec{i}}, \quad \text{and} \quad S_{\vec{i}}^z \rightarrow \left(b_{\vec{i}}^\dagger b_{\vec{i}} - \frac{1}{2} \right). \quad (149)$$

Since both the S and b operators satisfy identical commutation rules on their respective Hilbert spaces (Eqn. (146) and (148)), the Heisenberg model may be written in its hard core boson form:

$$H = J \sum_{\langle \vec{i}, \vec{j} \rangle} \frac{1}{2} (b_{\vec{i}}^\dagger b_{\vec{j}} + b_{\vec{j}}^\dagger b_{\vec{i}}) + \left(b_{\vec{i}}^\dagger b_{\vec{i}} - \frac{1}{2} \right) \left(b_{\vec{j}}^\dagger b_{\vec{j}} - \frac{1}{2} \right) \quad (150)$$

To carry out simulations of hard core boson Hamiltonians such as the Heisenberg model within the framework of the FTQMC of PQMC, we have to explicitly build the hard core bosons from fermions ($b_i^\dagger = c_{i,\uparrow}^\dagger c_{i,\downarrow}^\dagger$) and restrict the Hilbert space to doubly or empty occupied sites in terms of fermions. One can achieve this goal by considering the Hamiltonian

$$H = -t \sum_{\langle i,j \rangle} \left(\sum_{\sigma} c_{i,\sigma}^\dagger c_{j,\sigma} + H.c. \right)^2 - V \sum_{\langle i,j \rangle} (n_i \pm n_j)^2 - U \sum_i n_{i,\uparrow} n_{i,\downarrow}. \quad (151)$$

Here, $n_i = n_{i,\uparrow} + n_{i,\downarrow}$ and $n_{i,\sigma} = c_{i,\sigma}^\dagger c_{i,\sigma}$. At first glance, we can see that for $t, U, V > 0$ and with the use of the HS transformations (25),(28), the sign is absent since the weight for a given configuration of HS fields is a product of two identical real determinants.

To see how the above Hamiltonian relates to that of hard core boson systems, we expand the squares to obtain (up to a chemical potential term):

$$H = -2t \sum_{\langle i,j \rangle} \left(c_{i,\uparrow}^\dagger c_{i,\downarrow}^\dagger c_{j,\downarrow} c_{j,\uparrow} + H.c. \right) + (t \mp 2V) \sum_{\langle i,j \rangle} n_i n_j + 4t \sum_{\langle i,j \rangle} \vec{S}_i \vec{S}_j - (U + 8V) \sum_i n_{i,\uparrow} n_{i,\downarrow}. \quad (152)$$

As apparent, the Hamiltonian conserves the number of singly occupied sites. That is, the operator

$$P = \sum_{\vec{i}} n_{\vec{i}} (2 - n_{\vec{i}}). \quad (153)$$

which counts the singly occupied sites is a conserved quantity:

$$[P, H] = 0. \quad (154)$$

In particular, if one projects onto the subspace with $P = 0$ (P_0) then in this subspace the spin-spin term as well as the Hubbard interaction vanish:

$$HP_0 = -2t \sum_{\langle i,j \rangle} \left(c_{i,\uparrow}^\dagger c_{i,\downarrow}^\dagger c_{j,\downarrow} c_{j,\uparrow} + H.c. \right) + (t \mp 2V) \sum_{\langle i,j \rangle} n_i n_j - 3(t + V) \sum_{\vec{i}} n_{\vec{i}}. \quad (155)$$

To enforce this constraint (i.e. projection on the P_0 subspace on the imaginary time propagation) within the PQMC one just has to appropriately choose the trial wave function:

$$P_0 |\Psi_T\rangle = |\Psi_T\rangle \quad (156)$$

A possible choice is:

$$|\Psi_T\rangle = |\Psi_T^\uparrow\rangle \otimes |\Psi_T^\downarrow\rangle, \quad |\Psi_T^\sigma\rangle = \prod_{n=1}^{N_p^\sigma} c_{i_n, \sigma}^\dagger |0\rangle \quad (157)$$

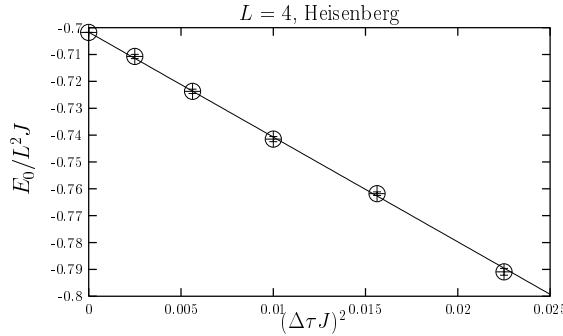


Figure 15. Ground state energy of the Heisenberg model on a 4×4 lattice as a function of the imaginary time discretization $\Delta\tau$. The data point at $\Delta\tau = 0$ corresponds to the exact result for this lattice size.

where $N_p^\uparrow = N_p^\downarrow$ corresponds to the number of electrons in the spin σ sector. Alternatively and in analogy to the Kondo lattice model, the constraint may be imposed with the attractive U term. The inclusion of this term allows the use of the FTQMC.

In the P_0 subspace

$$b_i^\dagger = c_{i,\uparrow}^\dagger c_{i,\downarrow}^\dagger \text{ and } n_i = 2b_i^\dagger b_i, \quad (158)$$

where b_i^\dagger are the desired hard-core bosons. Thus,

$$\begin{aligned} HP_0 = & -2t \sum_{\langle i,j \rangle} (b_i^\dagger b_j + H.c.) + 4(t \mp 2V) \sum_{\langle i,j \rangle} b_i^\dagger b_i b_j^\dagger b_j \\ & - 6(t + V) \sum_{\vec{i}} b_i^\dagger b_i \end{aligned} \quad (159)$$

which is nothing but a model of hard core bosons with nearest neighbor density-density interaction of arbitrary sign.

We can test the method at the Heisenberg point: $V = 0$. (The explicit mapping, up to a chemical potential term, to Eq. (150) is achieved after the canonical transformation: $b_i^\dagger \rightarrow (-1)^{i_x+i_y} b_i^\dagger$.) Fig. 15 plots the ground state energy as a function of $\Delta\tau$. As apparent, the QMC converges to the exact result. In this formulation, frustrating interactions would amount in inserting terms of the form $t (c_i^\dagger c_j + H.c.)^2$ for i, j within the same sublattice and with $t > 0$. The reader will easily convince himself that this leads to a sign problem. Finally we note that the use of the auxiliary field QMC for non-frustrated Heisenberg models should be seen as an amusing curiosity since it is not competitive with the loop algorithm approach.

5 The Hirsch-Fye Impurity Algorithm

A very much related algorithm to the FTQMC and PQMC is the Hirsch-Fye impurity algorithm⁶⁰ which is extensively used in the framework of dynamical mean field theories.^{15,16} As its name suggests, this algorithm is triggered at solving impurity problems and allows

simulations directly in the *thermodynamic* limit. However as will be apparent below the CPU time scales as the third power of the inverse temperature β . For simplicity we will consider the single impurity Anderson model for the formulation of the problem and the Kondo model for the example applications.

The Anderson impurity model reads:

$$H = \overbrace{\sum_{\vec{k}, \sigma} \epsilon(\vec{k}) c_{\vec{k}, \sigma}^\dagger c_{\vec{k}, \sigma} + V \sum_{\vec{k}, \sigma} c_{\vec{k}, \sigma}^\dagger f_\sigma + f_\sigma^\dagger c_{\vec{k}, \sigma} + \epsilon_f \sum_\sigma f_\sigma^\dagger f_\sigma}^{\equiv H_0} + U (f_\uparrow^\dagger f_\uparrow - 1/2) (f_\downarrow^\dagger f_\downarrow - 1/2). \quad (160)$$

Following the procedure introduced for the Hubbard model, the partition function is given by:

$$Z \equiv \text{Tr} [e^{-\beta(H - \mu N)}] = \sum_{\vec{s}} \left[\prod_\sigma \det [1 + B_m^\sigma B_{m-1}^\sigma \cdots B_1^\sigma] \right]. \quad (161)$$

With the notation

$$H_0 = \sum_{\sigma, \vec{i}} a_{\vec{i}, \sigma}^\dagger (h_0)_{\vec{i}, \vec{j}} a_{\vec{j}, \sigma} \quad (162)$$

where $a_{\vec{j}, \sigma}$ denotes c - of f -operators, \vec{i}, \vec{j} running over all orbitals, and the HS transformation of Eq. (24),

$$\begin{aligned} B_n^\sigma &= e^{-\Delta \tau h_0} e^{V_n^\sigma} \\ (V_n^\sigma)_{i,j} &= \delta_{i,j} \begin{cases} 0 & \text{if } \vec{i} \neq \text{impurity site} \\ \tilde{\alpha} \sigma s_n & \text{if } \vec{i} = \text{impurity site} \end{cases} \end{aligned} \quad (163)$$

Note that since we are considering a single impurity, we require only a single Hubbard Stratonovich field per time slice, s_n . Finally, $m\Delta\tau = \beta$.

The determinant in a given spin sector may be written as

$$\det [1 + B_m^\sigma B_{m-1}^\sigma \cdots B_1^\sigma] = \det O^\sigma \text{ with} \quad (164)$$

$$O^\sigma = \begin{pmatrix} 1 & 0 & . & . & 0 & B_1^\sigma \\ -B_2^\sigma & 1 & 0 & . & . & 0 \\ 0 & -B_3^\sigma & 1 & . & . & 0 \\ . & 0 & -B_4^\sigma & . & . & . \\ . & . & 0 & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & 0 & -B_m^\sigma & 1 \end{pmatrix}. \quad (165)$$

From Eq. (84) we identify:

$$(O^\sigma)^{-1} \equiv g^\sigma = \begin{pmatrix} G^\sigma(1, 1) & G^\sigma(1, 2) & \dots & G^\sigma(1, m) \\ G^\sigma(2, 1) & G^\sigma(2, 2) & \dots & G^\sigma(2, m) \\ \vdots & \vdots & \ddots & \vdots \\ G^\sigma(m, 1) & G^\sigma(m, 2) & \dots & G^\sigma(m, m) \end{pmatrix} \quad (166)$$

where $G^\sigma(n_1, n_2)$ are the time displaced Green functions as defined in Eqn. (81) and (82). Given a HS configuration \vec{s} and \vec{s}' and associated matrices V and V' the Green functions g and g' satisfy the following Dyson equation.

$$g' = g + (g - 1)(e^{V' - V} - 1)g' \quad (167)$$

To demonstrate the above, we consider

$$\tilde{O}^\sigma = O^\sigma e^{-V^\sigma} = \begin{pmatrix} e^{-V_1^\sigma} & 0 & . & . & 0 & e^{-\Delta\tau h_0} \\ e^{-\Delta\tau h_0} & e^{-V_2^\sigma} & 0 & . & . & 0 \\ 0 & e^{-\Delta\tau h_0} & e^{-V_3^\sigma} & . & . & 0 \\ . & 0 & e^{-\Delta\tau h_0} & . & . & . \\ . & . & 0 & . & . & . \\ . & . & . & . & 0 & e^{-\Delta\tau h_0} \\ 0 & . & . & . & 0 & e^{-V_m^\sigma} \end{pmatrix} \quad (168)$$

so that (omitting the spin index σ)

$$\tilde{g}' \equiv \tilde{O}'^{-1} = [\tilde{O} + \underbrace{\tilde{O}' - \tilde{O}}_{\equiv e^{-V'} - e^{-V}}]^{-1} = \tilde{g} - \tilde{g} (e^{-V'} - e^{-V}) \tilde{g}'. \quad (169)$$

Substitution, $\tilde{g} = e^V g$, leads to the Dyson Eq. (167).

The Green function matrix has dimensions $mN \times mN$ where N is the total number of orbitals and m the number of Trotter slices. Let $x = (\tau_x, i_x)$ with Trotter slice τ_x and orbital i_x , and the site index of the impurity to 0. Since

$$(e^{V' - V} - 1)_{x,y} = (e^{V' - V} - 1)_{x,x} \delta_{x,y} \delta_{i_x,0} \quad (170)$$

we can use the Dyson equation only for the impurity Green function:

$$g'_{f,f'} = g_{f,f'} + \sum_{f''} (g - 1)_{f,f''} (e^{V' - V} - 1)_{f'',f''} g'_{f'',f'} \quad (171)$$

with indices $f \equiv (\tau, 0)$ running from $1 \dots m$.

We now have all the ingredients to carry out a simulation. (Note that the algorithm is free of numerical instabilities.) Let us start from a random HS configuration \vec{s}' . For this configuration, we have to compute the impurity Green function. From the $U = 0$ solution, g_0^f (which can be obtained analytically⁶¹), we compute the impurity Green function for the HS fields \vec{s}' with the use of Eq. (171). This is readily achieved at the cost of an $m \times m$ matrix inversion.

To upgrade a single HS field with the Metropolis method, we have to calculate the ratio of the determinants (see Eq. (161)) as shown in section 3.3. This ratio only requires the knowledge of the equal time impurity Green function which is already at hand. If the proposed spin-flip is accepted the impurity Green function may readily be upgraded with the use of Eq. (171). Note that since we are considering a single spin flip, sum over f'' in Eq. (171) has only one non-vanishing contribution.

The attractive feature of the Hirsch-Fye impurity algorithm is that the conduction electrons may be considered directly in the thermodynamic limit. This is not be possible within the previously discussed FTQMC since the dimension of the matrices involved scale as the total number of orbitals. The Hirsch-Fye algorithm is not limited to impurity models. However, when applied to lattice models it is less efficient than the FTQMC and PQMC.

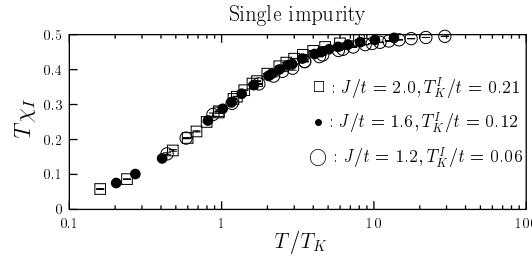


Figure 16. Impurity spin susceptibility of the Kondo model as computed with the Hirsch-Fye impurity algorithm.³⁸

The Hirsch-Fye impurity algorithm may equally be applied to the Kondo model,

$$H = \sum_{\vec{k},\sigma} \varepsilon(\vec{k}) c_{\vec{k},\sigma}^\dagger c_{\vec{k},\sigma} + J \vec{S}_{\vec{I}}^c \cdot \vec{S}_{\vec{I}}^f \quad (172)$$

where the same formulation as for the lattice problem is used. Fig. 16 plots the impurity spin susceptibility

$$\chi_I = \int_0^\beta \langle \vec{S}_{\vec{I}}^f(\tau) \cdot \vec{S}_{\vec{I}}^f \rangle \quad (173)$$

for various values of J/t for a half-filled conduction band. As apparent and at low energies the data collapse to the universal form $\chi_I = \frac{1}{T} f(T/T_K^I)$ where T_K^I is the Kondo temperature.

6 Conclusion

We have presented in all details the auxiliary field (or determinantal) Quantum Monte Carlo method for lattice problems. The ground state, finite temperature as well as Hirsch-Fye impurity algorithms were considered. The formulation of these algorithms for a range of fermionic as well as bosonic models was discussed. When the sign problem is avoidable the computational effort scales as the volume to the cubed times inverse temperature. For the Hirsch-Fye impurity algorithm – formulated directly in the thermodynamic limit – the computational cost scales as the cubed of the inverse temperature. The algorithms produce thermodynamic, dynamical -in conjunction with the Maximum Entropy method- as well as ground state properties of the model under consideration. They are unique in the sense that the sign problem turns out to be avoidable for particle-hole symmetric models as well as for models with attractive interactions which couple independently to an internal symmetry with an even number of states. It is not clear at present if other symmetries may be put to use so as to avoid the sign problem. Clearly, the sign problem remains the central issue and calls for new numerical approaches to the correlated electron problem.

Acknowledgments Most of the work presented here originates from collaborations with S. Capponi, M. Feldbacher, N. Furukawa, W. Hanke, M. Imada, A. Muramatsu, D.J. Scalapino, S. Sorella, M. Troyer and D. Würtz to which I express my warmest thanks. Financial support of the DFG in terms of a Heisenberg fellowship (grant number AS 120/1-1) is acknowledged.

Appendix

A The Monte Carlo Method

In this appendix, we briefly outline the Monte Carlo method. Our aim is to compute:

$$\langle O \rangle_P = \int_{\Omega} d\vec{x} P(\vec{x}) O(\vec{x}) \quad (174)$$

where Ω denotes the integration space, and $P(\vec{x})$ is a probability distribution,

$$\int_{\Omega} d\vec{x} P(\vec{x}) = 1 \text{ and } P(\vec{x}) \geq 0 \quad \forall \vec{x} \in \Omega. \quad (175)$$

For simplicity, we will assume Ω to be a subspace of \mathcal{R}^d with volume V . Here, d is the dimension. In practice, one does not have access to P , but to the unnormalized probability distribution $g(\vec{x})$ which satisfies

$$P(\vec{x}) = \frac{g(\vec{x})}{Z}, \quad Z = \int_{\Omega} d\vec{x} g(\vec{x}). \quad (176)$$

In the terminology of statistical mechanics, $g(\vec{x})$ corresponds to the Boltzmann weight and Z to the partition function. Hence the evaluation of Eq. (174) boils down to the calculating the quotient of two integrals. One may break up Ω into hypercubes of linear dimension h and approximate the integrals by sums. Depending upon the method used the systematic error will be proportional to h^k . The required number of function evaluations N is of the order V/h^d so that the systematic error scales a $N^{-k/d}$. Clearly, when d is large poor results are obtained. As we will now see, in the large d limit, the Monte Carlo method becomes attractive.

A.1 The Central Limit Theorem

Suppose that we have a set of independent points $\{\vec{x}_i\}$, $i : 1 \cdots N$ distributed according to the probability distribution $P(\vec{x})$ we can approximate $\langle O \rangle_P$ by:

$$\langle O \rangle_P \sim \frac{1}{N} \sum_{\substack{i=1 \\ \vec{x}_i \in P}}^N O(\vec{x}_i) = X. \quad (177)$$

Clearly, X will depend upon the chosen series of $\{\vec{x}_i\}$. The central limit theorem, tells us that in the large N limit the probability of obtaining a given value of X , $\mathcal{P}(X)$ reads

$$\mathcal{P}(X) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp \left[-\frac{(X - \langle O \rangle_P)^2}{2\sigma^2} \right] \quad \text{with } \sigma^2 = \frac{1}{N} (\langle O^2 \rangle_P - \langle O \rangle_P^2). \quad (178)$$

Thus independently of the dimension d , the convergence to the exact result scales as $1/\sqrt{N}$. The width of the above normal distribution, σ , corresponds to the statistical error. For practical purposes, one estimates σ by

$$\sigma^2 \approx \frac{1}{N} \left(\frac{1}{N} \sum_{\substack{i=1 \\ \vec{x}_i \in P}}^N O(\vec{x}_i)^2 - \left(\frac{1}{N} \sum_{\substack{i=1 \\ \vec{x}_i \in P}}^N O(\vec{x}_i) \right)^2 \right) \quad (179)$$

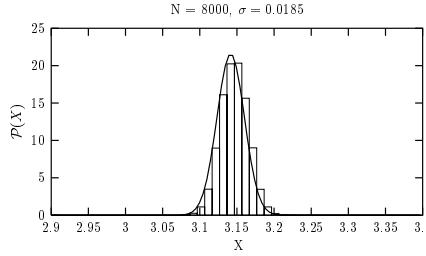


Figure 17. Boxes correspond to the distribution results obtained after 10000 simulations. For each simulation we draw $N = 8000$ points. For a single simulation, we obtain $\sigma = 0.0185$. The solid line corresponds to the result of central limit theorem with above value of σ .

Instead of demonstrating the central limit theorem, we give a simple example: the evaluation of the number π obtained via:

$$\pi = 4 \int_0^1 dx \int_0^1 dy \Theta(1 - x^2 + y^2), \quad (180)$$

where Θ is the Heavyside function, $\Theta(x) = 1$ for $x > 0$ and vanishes otherwise. In this example we $P(x, y) \equiv 1$. To generate the a sequence of N points $(x, y)_i$ from this probability distribution, we draw random numbers, x, y , in the interval $[0, 1]$. For $N = 8000$ we obtain an error $\sigma = 0.0185$ with the use of Eq. (179). To check the central limit theorem, we repeat the simulation 10000 times with different random numbers. Fig. (17) shows the thus obtained distribution which compares well to the result of the central limit theorem.

The jackknife and bootstrap methods⁶² provide alternative ways of estimating the error (179). These methods become particularly useful, if not essential, when one wishes to estimate the error on $f(\langle O_1 \rangle, \dots, \langle O_1 \rangle)$ where f is an arbitrary function of n variables. For a given sequence $\vec{x}_i \in P(\vec{x})$, $i : 1 \dots N$, the jackknife focuses on the samples that leaves out one configuration at a time:

$$f_i^J = f \left(\frac{1}{N-1} \sum_{j \neq i} O_1(\vec{x}_j), \dots, \frac{1}{N-1} \sum_{j \neq i} O_n(\vec{x}_j) \right). \quad (181)$$

The error estimate on f is then given by:

$$(\sigma_f^J)^2 \approx N \left(\frac{1}{N} \sum_{i=1}^N (f_i^J)^2 - \left(\frac{1}{N} \sum_{i=1}^N f_i^J \right)^2 \right) \quad (182)$$

One may verify explicitly that for $n = 1$ and $f(x) = x$ Eq. (182) reduces to Eq. (179) up to a factor $(N/(N-1))^2$ which tends to unity in the large N limit.

An alternative method for determining errors of f is the bootstrap algorithm. For a given sample of N configurations $\{\vec{x}_1 \dots \vec{x}_N\}$ drawn from the probability distribution $P(\vec{x})$, we can construct N^N sets of N configurations, $\{\vec{x}_{i_1} \dots \vec{x}_{i_N}\}$ with $i_1 : 1 \dots N$, $i_2 : 1 \dots N, \dots, i_N : 1 \dots N$, which correspond to the ideal bootstrap samples. For a

given bootstrap sample, defined by the vector $\vec{i} = (i_1, \dots, i_N)$,

$$f_i^B = f \left(\frac{1}{N} \sum_{k=1}^N O_1(\vec{x}_{i_k}), \dots, \frac{1}{N} \sum_{k=1}^N O_n(\vec{x}_{i_k}) \right). \quad (183)$$

The bootstrap estimate of the error is given by:

$$(\sigma_f^B)^2 \approx \frac{1}{N^N} \sum_{i_1, \dots, i_N=1}^N (f_{\vec{i}}^B)^2 - \left(\frac{1}{N^N} \sum_{i_1, \dots, i_N=1}^N f_{\vec{i}}^B \right)^2. \quad (184)$$

Again, one may check that for the special case, $n = 1$ and $f(x) = x$ Eq. (184) reduces to Eq. (179). Clearly, when N is large, it is numerically out of reach to generate all of the N^N bootstrap samples. Typically, to estimate the right hand side of Eq. (184) 200 or more bootstrap samples are generated stochastically. Since each bootstrap sample is equally probable we can generate them with: $i_k = \text{trunc}(N * \xi_k + 1)$ where ξ_k is a random number in the interval $[0, 1]$ and the function `trunc` returns an integer by truncating the numbers after the decimal point.

A.2 Generating Markov Chains

Our task is now to generate a set of points \vec{x} distributed according to $P(x)$. We introduce a Monte-Carlo time t , and a time dependent probability distribution $P_t(\vec{x})$ which evolves in time according to a Markov process: the future ($t + 1$) depends only on the present (t). Our aim is to obtain: $P_{t \rightarrow \infty}(\vec{x}) = P(\vec{x})$. To define the Markov chain, we introduce a matrix $T_{\vec{y}, \vec{x}}$ which corresponds to the probability of going from \vec{x} to \vec{y} . The time evolution of $P_t(\vec{x})$ is given by:

$$P_{t+1}(\vec{y}) = \sum_x T_{\vec{y}, \vec{x}} P_t(\vec{x}) \quad (185)$$

T has to satisfy the following properties.

$$\sum_x T_{\vec{y}, \vec{x}} = \sum_y T_{\vec{y}, \vec{x}} = 1, \quad T_{\vec{y}, \vec{x}} \geq 0 \quad (186)$$

That is, the probability of reaching a given \vec{y} from any \vec{x} or of landing anywhere in Ω given a initial \vec{x} is of unit. T has to be ergodic:

$$\forall \vec{x}, \vec{y} \in \Omega \exists s | (T^s)_{\vec{y}, \vec{x}} > 0. \quad (187)$$

Thus, we are assured to sample the whole phase space provided the above is satisfied. Lastly, the requirement of stationarity:

$$\sum_{\vec{x}} T_{\vec{y}, \vec{x}} P(\vec{x}) = P(\vec{y}). \quad (188)$$

Once we have reached the desired distribution, $P(\vec{x})$, we wish to stay there. Stationarity is automatically satisfied if

$$T_{\vec{y}, \vec{x}} P(\vec{x}) = T_{\vec{x}, \vec{y}} P(\vec{y}) \quad (189)$$

as may be seen by summing on sides over \vec{y} or \vec{x} . This relation is referred to as detailed balance or microreversibility. However, one has to keep in mind that stationarity and not detailed balance is essential.

That $P_t(\vec{x})$ approaches the equilibrium distribution $P(\vec{x})$ may be seen with:

$$\begin{aligned} \|P_{t+1} - P\| &\equiv \sum_{\vec{y}} |P_{t+1}(\vec{y}) - P(\vec{y})| \\ &= \sum_{\vec{y}} \left| \sum_{\vec{x}} T_{\vec{y}, \vec{x}} P_t(\vec{x}) - \sum_{\vec{x}} T_{\vec{y}, \vec{x}} P(\vec{x}) \right| \\ &\leq \sum_{\vec{y}} \sum_{\vec{x}} T_{\vec{y}, \vec{x}} |P_t(\vec{x}) - P(\vec{x})| \\ &= \sum_{\vec{x}} |P_t(\vec{x}) - P(\vec{x})| \equiv \|P_t - P\|. \end{aligned} \quad (190)$$

Under the assumption of ergodicity, the strict equality holds only when $P_t = P$. Due to Eq. (186), the right eigenvectors of T have eigenvalues $|\lambda| \leq 1$, $\lambda = 1$ corresponding to the stationary distribution P . Starting with an arbitrary distribution $P_0(\vec{x})$ convergence to P will be determined by the eigenvalue of T with second largest absolute value, λ_1 . The rate of convergence of $P_t(\vec{x})$ to $P(\vec{x})$ will then scale as $\exp -t/\tau$ where $\tau = -1/\log(|\lambda_1|) > 0$.

Having defined T , we now have to construct it explicitly. Let $T_{\vec{y}, \vec{x}}^0$ the probability of proposing a move from \vec{x} to \vec{y} and $a_{\vec{y}, \vec{x}}$ the probability of accepting it. $1 - a_{\vec{y}, \vec{x}}$ corresponds to the probability of rejecting the move. T_0 is required to satisfy Eq. (186). Since in general we want to propose moves which change the initial configuration, $T_{\vec{x}, \vec{x}}^0 = 0$. With $a_{\vec{y}, \vec{x}}$ and $T_{\vec{y}, \vec{x}}^0$ we build $T_{\vec{y}, \vec{x}}$ with:

$$T_{\vec{y}, \vec{x}} = \begin{cases} T_{\vec{y}, \vec{x}}^0 a_{\vec{y}, \vec{x}} & \text{if } \vec{y} \neq \vec{x} \\ \sum_{\vec{y} \neq \vec{x}} T_{\vec{y}, \vec{x}}^0 (1 - a_{\vec{y}, \vec{x}}) & \text{if } \vec{y} = \vec{x} \end{cases} \quad (191)$$

Clearly $T_{\vec{y}, \vec{x}}$ satisfies Eq. (186). To satisfy the stationarity, we impose the detailed balance condition to obtain the equality:

$$T_{\vec{y}, \vec{x}}^0 a_{\vec{y}, \vec{x}} P_{\vec{x}} = T_{\vec{x}, \vec{y}}^0 a_{\vec{x}, \vec{y}} P_{\vec{y}}. \quad (192)$$

Let us set:

$$a_{\vec{y}, \vec{x}} = \mathcal{F}(Z) \quad \text{with} \quad Z = \frac{T_{\vec{x}, \vec{y}}^0 P_{\vec{y}}}{T_{\vec{y}, \vec{x}}^0 P_{\vec{x}}} \quad (193)$$

with $\mathcal{F}(Z) : [0 : \infty[\rightarrow [0, 1]$. Since $a_{\vec{x}, \vec{y}} = \mathcal{F}(1/Z)$, \mathcal{F} has to satisfy:

$$\frac{\mathcal{F}(Z)}{\mathcal{F}(1/Z)} = Z. \quad (194)$$

There are many possible choices. The Metropolis algorithm is based on the choice:

$$\mathcal{F}(Z) = \min(Z, 1). \quad (195)$$

Thus, one proposes a move from \vec{x} to \vec{y} and accepts it with probability $Z = \frac{T_{\vec{x}, \vec{y}}^0 P_{\vec{y}}}{T_{\vec{y}, \vec{x}}^0 P_{\vec{x}}}$. In the practical implementation, one picks a random number r in the interval $[0 : 1]$. If $r < Z$

$(r > Z)$ one accepts (rejects) the move. Alternative choices of $\mathcal{F}(Z)$ are for example:

$$\mathcal{F}(Z) = \frac{Z}{1+Z} \quad (196)$$

which is referred to as the Heat bath method.

That the so constructed T matrix is ergodic depends upon the choice of T^0 . In many cases, one will wish to combine different types of moves to achieve ergodicity. For a specific move, i we construct $T^{(i)}$ as shown above so that $T^{(i)}$ conditions (186) and (189). The moves may be combined in two ways:

$$T = \sum_i \lambda_i T^{(i)}, \quad \sum_i \lambda_i = 1 \quad (197)$$

which is referred to as random upgrading since one picks with probability λ_i the move $T^{(i)}$. Clearly, T equally satisfies (186), (189) and if the moves have to be chosen appropriately to satisfy the ergodicity condition. Another choice is sequential upgrading. A deterministic ordering of the moves is chosen to obtain:

$$T = \prod_i T^{(i)}. \quad (198)$$

This choice does not satisfy detailed balance condition, but does satisfy stationarity (188) as well as (186). Again ergodicity has to be *checked* on a case to case basis.

In principle, we could now start with an arbitrary probability distribution $P_0(\vec{x})$ and propagate it along Monte Carlo time t ($P_{t+1} = TP_t$). Convergence to the equilibrium distribution will occur on time scales set by τ . This procedure involves handling many configurations \vec{x} at a given time t . Alternatively one can start with a single configuration \vec{x} and propagate it according to T . That is the probability of having the configuration \vec{y} at the next MC time is given by $T_{\vec{y}, \vec{x}}$. This procedure generates a sequence of configuration in MC time. For *large* values of N (see below) $\vec{x}_{t=1} \cdots \vec{x}_{t=N}$ will be distributed according to P . The observable O may now be estimated with:

$$\langle O \rangle_P \approx \frac{1}{N} \sum_{t=1}^N O(\vec{x}_t). \quad (199)$$

The required value of N depends autocorrelation time of the observable O :

$$C_O(t) = \frac{\frac{1}{N} \sum_{s=1}^N O(\vec{x}_s)O(\vec{x}_{s+t}) - \left(\frac{1}{N} \sum_{s=1}^N O(\vec{x}_s) \right)^2}{\frac{1}{N} \sum_{s=1}^N O(\vec{x}_s)^2 - \left(\frac{1}{N} \sum_{s=1}^N O(\vec{x}_s) \right)^2} \quad (200)$$

One expects $C_O(t) \sim e^{-t/\tau_O}$ where τ_O corresponds to the MC time scale on which memory of the initial configuration is lost. Hence, to obtain meaningful results, $N \gg \tau_O$. Note that one should equally take into account a *warm up* time by discarding at least the first τ_O configurations in the MC sequence. Naively, one would expect $\tau_O = \tau$. However, this depends on the overlap of the observable with the slowest mode in the MC dynamics which relaxes as $e^{-t/\tau}$. In particular in a model with spin rotation symmetry the slowest mode may correspond to the rotation of the total spin. In this case, observables which are invariant under a spin rotation will not be effected by the slowest mode of the MC dynamics. Hence in this case $\tau_O < \tau$.

We now consider the estimation of the error. To apply the central limit theorem, we need a set of independent estimates of $\langle O \rangle_P$. This may be done by regrouping the data into *bins* of size $n\tau_O$.

$$\tilde{O}_n(t) = \frac{1}{n\tau_O} \sum_{s=1}^{n\tau_O} O(\vec{x}_{(t-1)n\tau_O+s}) \quad (201)$$

with $t = 1 \cdots N/(n\tau_O)$. If n is large enough (i.e. $n \approx 10 - 20$) then $\tilde{O}_n(t)$ may be considered as an independent estimate, and the error is given by:

$$\sigma_n = \sqrt{\frac{1}{M} \left(\frac{1}{M} \sum_{t=1}^M \tilde{O}_n(t)^2 - \left(\frac{1}{M} \sum_{i=1}^M \tilde{O}_n(t) \right)^2 \right)} \quad (202)$$

where $M = N/(n\tau_O)$. If n is large enough the error σ_n should be n independent.

We conclude this section with an example of error analysis for the one-dimensional Ising model:

$$H(\{\sigma\}) = -J \sum_{i=1}^L \sigma_i \sigma_{i+1} \quad \sigma_{L+1} = \sigma_1 \quad (203)$$

where $\sigma_i = \pm 1$. This model may easily be solved exactly with the transfer matrix method and thus produces a useful testing ground for the MC approach. In particular at zero temperature, $T = 0$, a phase transition to a ferromagnetically ordered phase ($J > 0$) occurs.⁷ Spin-spin correlations are given by:

$$g(r) = \frac{1}{L} \sum_{i=1}^L \langle \sigma_i \sigma_{i+r} \rangle \quad \text{with} \quad \langle \sigma_i \sigma_{i+r} \rangle = \frac{\sum_{\{\sigma\}} e^{-\beta H(\{\sigma\})} \sigma_i \sigma_{i+r}}{\sum_{\{\sigma\}} e^{-\beta H(\{\sigma\})}} \quad (204)$$

where β corresponds to the inverse temperature. To simulate the model, we use a simple random site upgrading method: a site (i) is randomly chosen, and the spin is flipped ($\sigma_i \rightarrow -\sigma_i$) with the heat bath algorithm. The MC time unit corresponds to a single sweep meaning that L sites are randomly chosen before a measurement is carried out.

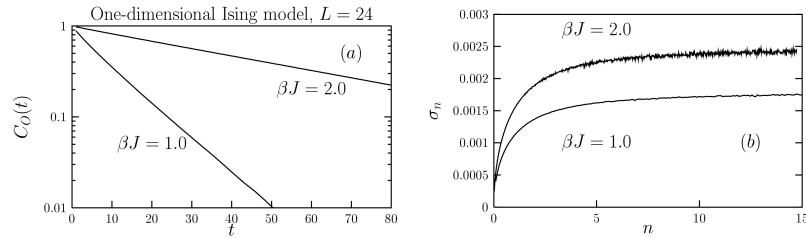


Figure 18. One dimensional Ising model on an $L=24$ site lattice. (a) Autocorrelation time (see Eq. (200)) for $g(r = L/2)$. The time unit is corresponds to a single sweep. (b) Estimate of the error (see Eq. (202)). Here, n corresponds to the size of the bins in units of the autocorrelation time. As apparent $n \sim 10$ is sufficient to obtain a reliable estimate of the error. After 2×10^6 sweeps, our results yield $g(r = L/2) = 0.076 \pm 0.0018$ and 0.909 ± 0.0025 for $\beta t = 1$ and 2 respectively. The exact result reads $g(r = L/2) = 0.0760$ and 0.9106 at $\beta t = 1$ and 2 respectively.

Fig. 18 plots the autocorrelation time for $g(r = L/2)$ on an $L = 24$ site lattice at $\beta J = 1$ and $\beta J = 2$. From Fig. 18a one can extract the autocorrelation time: $\tau_O \approx 11, 54$ for for $\beta J = 1, 2$ respectively. Fig. 18b plots the error as function of bin size in units of the τ_O (see Eq. (202)). As apparent, $n \approx 10$ is sufficient to get a reliable estimate of the error.

References

1. S. R. White. *Physics Reports*, 301:187, 1998.
2. H. G. Evertz, G. Lana, and M. Marcu. *Phys. Rev. Lett.*, 70:875, 1993.
3. R. Blankenbecler, D. J. Scalapino, and R. L. Sugar. *Phys. Rev. D*, 24:2278, 1981.
4. J. E. Hirsch, D. J. Scalapino, R. L. Sugar, and R. Blankenbecler. *Phys. Rev. B*, 26:5033, 1981.
5. M. Barma and B. S. Shastry. *Phys. Rev. B*, 18:3351, 1978.
6. F. F. Assaad and D. Würtz. *Phys. Rev. B*, 44:2681, 1991.
7. R. J. Baxter. *Exactly solved models in statistical mechanics*. Academic Press Limited, London, 1989.
8. M. Troyer, M. Imada, and K. Ueda. *J. Phys. Soc. Jpn.*, 66:2957, 1997.
9. J. E. Hirsch. *Phys. Rev. B*, 31:4403, 1985.
10. S. R. White, D. J. Scalapino, R. L. Sugar, E. Y. Loh, J. E. Gubernatis, and R. T. Scalettar. *Phys. Rev. B*, 40:506, 1989.
11. E. Loh and J. Gubernatis. In W. Hanke and Y. V. Kopaev, editors, *Modern Problems of Condensed Matter Physics*, volume 32, page 177. North Holland, Amsterdam, 1992.
12. G. Sugiyama and S.E. Koonin. *Analys of Phys.*, 168:1, 1986.
13. S. Sorella, S. Baroni, R. Car, and M. Parrinello. *Europhys. Lett.*, 8:663, 1989.
14. S. Sorella, E. Tosatti, S. Baroni, R. Car, and M. Parrinello. *Int. J. Mod. Phys. B*, 1:993, 1989.
15. M. Jarrell. *Phys. Rev. Lett.*, 69:168, 1992.
16. A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg. *Rev. of. Mod. Phys.*, 68:13, 1996.
17. R. M. Fye. *Phys. Rev. B*, 33:6271, 1986.
18. B.B. Beard and U.J. Wiese. *Phys. Rev. Lett.*, 77:5130, 1996.
19. H. G. Evertz. The loop algorithm. *cond-mat/9707221*, 1997.
20. M. Brunner, F. F. Assaad, and A. Muramatsu. *Phys. Rev. B*, 62:12395, 2000.
21. M. Brunner, S. Capponi, F. F. Assaad, and A. Muramatsu. *Phys. Rev. B*, 63:R180511, 2001.
22. M. Troyer, F. F. Assaad, and D. Würtz. *Helv. Phys. Acta.*, 64:942, 1991.
23. J. E. Hirsch. *Phys. Rev. B*, 28:4059, 1983.
24. F. F. Assaad. In E. Krause and W. Jäger, editors, *High performance computing in science and engineering*, page 105. Springer, Berlin, 1998. [[cond-mat/9806307](#)].
25. F. F. Assaad, M. Imada, and D. J. Scalapino. *Phys. Rev. B*, 56:15001, 1997.
26. F. F. Assaad, M. Imada, and D. J. Scalapino. *Phys. Rev. Lett.*, 77:4592, 1996.
27. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1992.
28. F. F. Assaad and M. Imada. *J. Phys. Soc. Jpn.*, 65:189, 1996.
29. F. F. Assaad. *Phys. Rev. Lett.*, 83:796, 1999.

30. M. Jarrell and J.E. Gubernatis. *Physics Reports*, 269:133, 1996.
31. W. von der Linden. *Applied Physics A*, 60:155, 1995.
32. J. E. Hirsch. *Phys. Rev. B*, 38:12023, 1988.
33. M. Feldbacher and F. F. Assaad. *Phys. Rev. B*, 63:73105, 2001.
34. S. Fahy and D. R. Hamann. *Phys. Rev. Lett.*, 65:3437, 1990.
35. S. Fahy and D. R. Hamann. *Phys. Rev. B*, 43:765, 1991.
36. S. Zhang, J. Carlson, and J. E. Gubernatis. *Phys. Rev. Lett.*, 74:3652, 1995.
37. S. Zhang, J. Carlson, and J. E. Gubernatis. *Phys. Rev. B*, 55:7464, 1997.
38. F. F. Assaad. Depleted kondo lattices: mean-field and Quantum Monte Carlo calculations. *cond-mat/0104126*, to appear in *Phys. Rev. B*.
39. E. Fradkin. *Field Theories of condensed matter systems*. Frontiers in Physics. Addison-Wesley Publishing Company, Redwood City, 1991.
40. D. Poilblanc. *Phys. Rev. B*, 44:9562, 1991.
41. C. Gross. *Z. Phys. B*, 86:359, 1992.
42. R. T. Scalettar, E. Y. Loh, J. E. Gubernatis, A. Moreo, S. R. White, D. J. Scalapino, and R. L. Sugar. *Phys. Rev. Lett.*, 62:1407, 1989.
43. F. F. Assaad, W. Hanke, and D. J. Scalapino. *Phys. Rev. B*, 50:12835, 1994.
44. N. Trivedi and M. Randeria. *Phys. Rev. Lett.*, 75:312, 1995.
45. W. Kohn. *Phys. Rev. A*, 133:171, 1964.
46. M. Imada, A. Fujimori, and Y. Tokura. *Rev. Mod. Phys.*, 70:1039, 1998.
47. S. Sorella and E. Tosatti. *Europhys. Lett.*, 19:699, 1992.
48. G. G. Batrouni and R. T. Scalettar. *Phys. Rev. B*, 42:2282, 1990.
49. N. Furukawa and M. Imada. *J. Phys. Soc. Jpn.*, 60:3669, 1991.
50. P. A. Lee, T. M. Rice, J. W. Serene, L. J. Sham, and J. W. Wilkins. *Comm. Condens. Matter Phys.*, 12:99, 1986.
51. G. Aeppli and Z. Fisk. *Comm. Condens. Matter Phys.*, 16:155, 1992.
52. J. R. Schrieffer and P. A. Wolff. *Phys. Rev.*, 149:491, 1966.
53. H. Tsunetsugu, M. Sigrist, and K. Ueda. *Rev. Mod. Phys.*, 69:809, 1997.
54. M. Vekic, J. W. Cannon, D. J. Scalapino, R. T. Scalettar, and R. L. Sugar. *Phys. Rev. Lett.*, 74:2367, 1995.
55. R. M. Fye and D. J. Scalapino. *Phys. Rev. Lett.*, 65:3177, 1990.
56. R. M. Fye and D. J. Scalapino. *Phys. Rev. B*, 44:7486, 1991.
57. S. Q. Shen. *Phys. Rev. B*, 53:14252, 1996.
58. H. Tsunetsugu. *Phys. Rev. B*, 55:3042, 1997.
59. S. Capponi and F. F. Assaad. *Phys. Rev. B*, 63:155113, 2001.
60. J. E. Hirsch and R. M. Fye. *Phys. Rev. Lett.*, 56:2521, 1986.
61. A. C. Hewson. *The Kondo Problem to Heavy Fermions*. Cambridge Studies in Magnetism. Cambridge Universiy Press, Cambridge, 1997.
62. B. Efron. *The jackknife, the bootstrap and other resampling plans*. CBMS-NSF conference series in applied mathematics. J. W. Arrowsmith Ltd., Bristol, England, 1982.

Effective Hamiltonian Approach for Strongly Correlated Lattice Models

Sandro Sorella^{1,2}

¹ Istituto Nazionale per la Fisica della Materia, 34014 Trieste, Italy

² SISSA, Via Beirut n.2-4, 34014 Trieste, Italy
E-mail: sorella@sissa.it

We review a recent approach for the simulation of many-body interacting systems based on an efficient generalization of the Lanczos method for Quantum Monte Carlo simulations. This technique allows to perform systematic corrections to a given variational wavefunction, that allow to estimate exact energies and correlation functions, whenever the starting variational wavefunction is a qualitatively correct description of the ground state. The stability of the variational wavefunction against possible phases, not described at the variational level can be tested by using the “effective Hamiltonian” approach. In fact Monte Carlo methods, such as the “fixed node approximation” and the present “generalized Lanczos technique” (Phys. Rev. B 64, 024512, 2001) allow to obtain exact ground state properties of an effective Hamiltonian, chosen to be as close as possible to the exact Hamiltonian, thus yielding the most reasonable estimates of correlation functions. We also describe a simplified one-parameter scheme that improve substantially the efficiency of the generalized Lanczos method. This is tested on the $t - J$ model, with a special effort to obtain accurate pairing correlations, and provide a possible non-phonon mechanism for High temperature superconductivity.

1 Introduction

Despite the tremendous progress of computer performances the general task of determining the ground state wavefunction of a many-electron system is still far from being settled. For instance, even for simplified models on a lattice, there is no general consensus on the ground state properties of a system of about 100 electrons on $L \approx 100$ sites. The most striking example is the so called $t - J$ model: This model is still a subject of intense numerical studies, due to its possible relevance for High Tc superconductivity.^{1,2} The Hamiltonian reads:

$$\hat{H} = J \sum_{\langle i,j \rangle} \left(\hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j - \frac{1}{4} \hat{n}_i \hat{n}_j \right) - t \sum_{\langle i,j \rangle, \sigma} \hat{c}_{i,\sigma}^\dagger \tilde{c}_{j,\sigma}, \quad (1)$$

where $\tilde{c}_{i,\sigma}^\dagger = \hat{c}_{i,\sigma}^\dagger (1 - \hat{n}_{i,\bar{\sigma}})$, $\hat{n}_i = \sum_{\sigma} \hat{n}_{i,\sigma}$ is the electron density on site i , $\hat{\mathbf{S}}_i = \sum_{\sigma, \sigma'} \hat{c}_{i,\sigma}^\dagger \tau_{\sigma, \sigma'} \tilde{c}_{i,\sigma'}$ is the spin operator and $\tau_{\sigma, \sigma'}$ are Pauli matrices. In the following we consider N electrons on L sites, with periodic boundary conditions,(PBC), in order to minimize size effects.

After many years of intense numerical and theoretical efforts there is no general consensus on the properties of this simple Hamiltonian and of the related Hubbard model. In particular according to density matrix renormalization group (DMRG) studies,⁴ d -wave superconductivity is not stable in this model, whereas a ground state non uniform in density (with so called “stripes”) is found. Several QMC studies provide controversial results,

most of them indicating a superconducting behavior, and some of them,⁵ indicating the opposite.

The reason of the above controversy, can be easily explained within the straightforward variational approach. Whenever a model Hamiltonian cannot be solved exactly either numerically (with no sign problem) or analytically, the most general and reasonable approach is an approximate minimization of the energy within a particular class of wavefunctions, for instance also DMRG can be considered a variational approach with a particularly complicated variational wavefunction obtained by DMRG iterations. However, within the variational approach, one faces the following problem: for large system size L the gap to the first excited state scales generally to zero quite rapidly with L . Thus between the ground state energy and the variational energy there maybe a very large number of states with completely different correlation functions. In this way one can generally obtain different variational wavefunctions with almost similar energy per site, but with completely different correlation functions. It is easily understood that, within a straightforward variational technique, there is no hope to obtain sensible results for large system size, unless for a system with a finite gap to all excitations, such as spin liquid,⁶ or band insulators.

In the following we are trying to argue that a possible solution to the previous limitation of the variational technique is provided by what we call in the following “the effective Hamiltonian approach”.

This approach relies on the following assumption:

“Among similar Hamiltonians with local interactions the ground state correlation functions depend weakly on the details of the Hamiltonian, in a sense that similar Hamiltonians should provide similar correlation functions”. In this way the ground state of an effective Hamiltonian (such as the fixed node Hamiltonian⁸) that can be solved exactly by Quantum Monte Carlo schemes can be used as a variational state of the desired Hamiltonian, in this way providing not only a good variational energy but the most reasonable estimate of correlation functions, as long as the variational energy obtained is close – but not terribly close as in the straightforward variational approach – to the exact ground state energy.

The paper is based therefore on the recent numerical advances for solving approximately model Hamiltonians on a lattice: the fixed node,⁸ and the “generalized Lanczos technique”,³ that allows to improve systematically the variational energy provided by the effective Hamiltonian approach, by combining in an efficient way the power of the Lanczos variational technique with the “effective Hamiltonian approach”. Through all the paper and pictures we will use “FN” to indicate the “fixed node approach”, whereas “SR” will indicate the “stochastic reconfiguration method” used to apply the “generalized Lanczos” scheme. In the first part we describe the Lanczos technique, then we derive the effective Hamiltonian approach in a slightly more general way than the standard “fixed node” method. Finally we show that the mentioned “generalized Lanczos method” represents a very efficient implementation of both the previous techniques – Lanczos and fixed node – on a lattice. We also point out some slight but important improvements and simplifications to the most recent formulation of the “generalized Lanczos scheme”.³ In the last section before the conclusion we show some example on the t-J model, where the “effective Hamiltonian approach” is clearly useful, as the pairing correlation functions appear to be rather independent from the initial variational guess, even for large system size $L \simeq 50$ and small J/t .

2 The Lanczos Technique

The Lanczos technique represents a remarkable improvement of the power method used to filter out systematically the ground state component of a given initial wavefunction ψ_G by an iterative technique. The power method is based on the following equation:

$$|\psi_0\rangle \simeq (\Lambda I - H)^p |\psi_G\rangle \quad (2)$$

where Λ is a suitable large shift to ensure convergence to the ground state for large p , I is the identity matrix and $|\psi_0\rangle$ the ground state of H . At a given iteration p , after applying just p powers of the Hamiltonian, a much better wavefunction ψ_p can be obtained by combining, with proper coefficients α_k , the states obtained with the power method in the previous iterations:

$$|\psi_p\rangle = \left(1 + \sum_{k=1}^p \alpha_k H^k\right) |\psi_G\rangle \quad (3)$$

with parameters $\{\alpha_k\}$ for $k = 1, \dots, p$ minimizing the energy expectation value $\langle\psi_p|\hat{H}|\psi_p\rangle/\langle\psi_p|\psi_p\rangle$. For any p it is simple to show that the wavefunction (3) corresponds exactly to apply p Lanczos step iterations to the initial wavefunction $|\psi_G\rangle$. The H -polynomial of degree p which is applied to the initial state ψ_G , can be generally factorized in terms of its roots z_i :

$$\left(1 + \sum_{k=1}^p \alpha_k \hat{H}^k\right) = \prod_{i=1}^p (1 - H/z_i) \quad (4)$$

This decomposition will be particularly important for applying statistically the Lanczos technique with the Stochastic Reconfiguration (see later). As it is clear from Fig. (1), the Lanczos method converges very quickly to the ground state wavefunction especially when a particularly good “guess” is used for ψ_G .

Whenever the ground state wavefunction is approached $|\langle\psi_0|\psi_p\rangle|^2/\langle\psi_p|\psi_p\rangle^2 = 1 - \epsilon_p$, with $\epsilon_p \rightarrow 0$ for larger p , with the energy approaching the exact value with corrections $\simeq \epsilon_p$. On the other hand, the variance σ_p^2 of the Hamiltonian on the approximate state ψ_p

$$\sigma_p^2 = \langle\psi_p|H^2|\psi_p\rangle - \langle\psi_p|H|\psi_p\rangle^2 = O(\epsilon_p)$$

is going to zero in the limit when ψ_p is the exact eigenstate ψ_0 with the same corrections proportional to ϵ_p . It is clear therefore that a very stable evaluation of the energy can be done by using few Lanczos steps values of the energy and the corresponding variance. Then, by performing simple extrapolation (linear or even polynomial), the exact ground state result is easily estimated provided the energy-variance values are close to the linear regime (see Fig. 1). The same scheme can be applied even for correlation functions,³ and represents one of the most simple and effective methods to estimate exact correlation functions with few Lanczos steps (i.e. with a minor computational effort) whenever the variational wavefunction ψ_G is particularly good, i.e. is close to the linear energy vs. variance regime. Such property of the variational wavefunction can be satisfied even for system size $L \simeq 100$.³

The initial wavefunction to which the Lanczos and the following techniques will be applied can be written as follows:¹⁰

$$|\psi_G\rangle = |\psi_{p=0}\rangle = \hat{P}_0 \hat{P}_N \hat{J}|D\rangle. \quad (5)$$

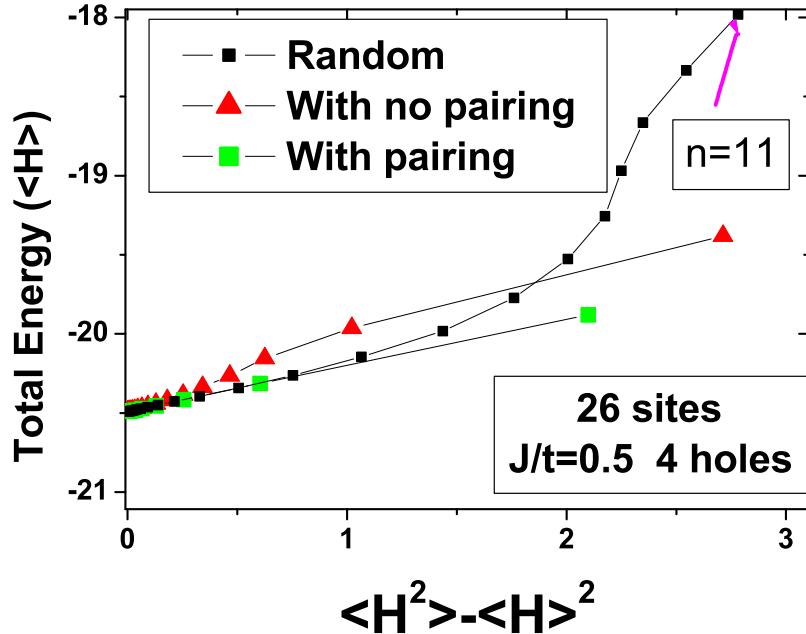


Figure 1. Energy $\langle H \rangle$ vs. variance $\langle H^2 \rangle - \langle H \rangle^2$ of the Lanczos technique for different initial wavefunction ψ_G . Here n represents the number of iterations. Lower variance is always obtained for larger n . The zero variance limit is the exact results.

where $|D\rangle$ is a BCS wavefunction, which is an exact eigenstate of the following Hamiltonian:

$$\hat{H}_{BCS} = \hat{H}_0 + \frac{\Delta_{BCS}}{2}(\hat{\Delta}^\dagger + \hat{\Delta}) \quad (6)$$

$$\hat{\Delta}^\dagger = \sum_{\langle i,j \rangle} M_{i,j} (\tilde{c}_{i,\uparrow}^\dagger \tilde{c}_{j,\downarrow}^\dagger + \tilde{c}_{j,\uparrow}^\dagger \tilde{c}_{i,\downarrow}^\dagger) \quad (7)$$

where $\hat{H}_0 = \sum_{k,\sigma} \epsilon_k \tilde{c}_{k,\sigma}^\dagger \tilde{c}_{k,\sigma}$ is the free electron tight binding nearest-neighbor Hamiltonian, $\epsilon_k = -2(\cos k_x + \cos k_y) - \mu$, μ is the free-electron chemical potential and $\hat{\Delta}^\dagger$ creates all possible singlet bonds with d-wave symmetry being $M_{i,j}$, $M_{i,j}$ not restricted to nearest neighbors, but exhaustively parametrized with a reasonable number of variational parameters as described in.³ \hat{P}_N and \hat{P}_0 are the projectors over the subspaces with a fixed number N of particles and no doubly occupied states. Finally the Jastrow factor $\hat{J} = \exp \left(1/2 \sum_{i,j} v(i-j) \hat{n}_i \hat{n}_j \right)$ couples the holes via the density operators \hat{n}_i and contains other few variational parameters. We note here that by performing a particle-hole transformation on the spin down $\tilde{c}_{i,\downarrow}^\dagger \rightarrow (-1)^i \tilde{c}_{i,\downarrow}^\dagger$, the ground state of the BCS Hamil-

tonian is just a Slater-determinant with $N = L$ particles.¹¹ This is the reason why this variational wavefunction can be considered of the generic Jastrow-Slater form, a standard variational wavefunction used in QMC. All the mentioned variational parameters are obtained by minimizing the energy expectation value of H over ψ_G .³

Using the particle-hole transformation, it is also possible to control exactly the spurious finite system divergences related to the nodes of the d-wave order parameter.

3 The Effective Hamiltonian Approach

In a discrete Hilbert space defined for instance by configurations x of electrons with definite positions and spins we consider *any* Hamiltonian H with real matrix elements $H_{x',x}$ and *any* real wavefunction $\psi_G(x)$ assumed to be non zero for each configuration x .

By means of the wavefunction ψ_G – hereafter called the guiding wavefunction – we can define a two parameter class of Hamiltonians H_{FN}^γ depending on γ and r :

$$H_{FN}^\gamma = \begin{cases} H_{x,x} + (1 + \gamma)\mathcal{V}_{sf}(x) + r(1 + \gamma)e_L(x) & \text{for } x' = x \\ H_{x',x} & \text{if } x' \neq x \quad \text{and } \psi_G(x')H_{x',x}/\psi_G(x) < 0 \\ -\gamma H_{x',x} & \text{if } x' \neq x \quad \text{and } \psi_G(x')H_{x',x}/\psi_G(x) > 0 \end{cases} \quad (8)$$

where the local energy $e_L(x)$ is defined by:

$$e_L(x) = \sum_{x'} \psi_G(x')H_{x',x}/\psi_G(x) \quad (9)$$

and the so called sign-flip term $\mathcal{V}_{sf}(x)$ introduced in⁸ is given by considering the sum of all the *positive* off-diagonal matrix elements appearing in the local energy. The effective Hamiltonian H^γ has the *same* matrix elements of the Hamiltonian H for all off-diagonal matrix elements that do not frustrate the guiding function signs, the other ones are taken into account by proper modification of the diagonal term.

The following properties are almost an immediate consequence of the above definitions:

- i) for $\gamma = -1$ $H = H_{FN}^\gamma$,
- ii) for $r = -1/(1 + \gamma)$ and $\gamma \neq -1$ the ground state of H_{FN}^γ is the guiding wavefunction itself with zero ground state energy, namely $H_{FN}^\gamma|\psi_G\rangle = 0$.
- iii) $H = H_{FN}^\gamma - (1 + \gamma)\frac{dH_{FN}^\gamma}{d\gamma}$
- iv) $E_L(x) = \sum_{x'} \psi_G(x')H_{FN}^\gamma/\psi_G(x) = e_L(x)(1 + r(1 + \gamma))$ where $E_L(x)$ is the local energy of the effective Hamiltonian H_{FN}^γ , whereas $e_L(x) = \sum_{x'} \psi_G(x')H/\psi_G(x)$, the corresponding one for H . Moreover:
- (v) for $\gamma \geq 0$ the ground state $\psi_0^{FN}(x)$ of H_{FN}^γ may be chosen to have the same signs of the guiding wavefunction, namely $\psi_G(x)\psi_{FN}(x) \geq 0$ for any configuration x . This follows by doing a unitary transformation of the basis $|\bar{x}\rangle = \text{Sign}[\psi_G(x)]|x\rangle$, in which the off-diagonal matrix elements of the Hamiltonian $H_{FN,\bar{x}',\bar{x}}^\gamma < 0$ are non-positive. Thus the Perron-Frobenius theorem holds implying that a ground state wavefunction (in principle there maybe degeneracy) can be chosen to satisfy $\psi_0^{FN}(\bar{x}) \geq 0$ in the new basis, which finally proves (v) in the original basis. The statement (v) suggests that the effective Hamiltonian H_{FN}^γ represents the lattice counterpart of the fixed node (FN) hamiltonian,

a well known approximation for continuous models.⁹ Furthermore, provided the matrix elements of the hamiltonian H or H^{FN} satisfy an ergodicity property (namely that any two arbitrary configurations x and x' can be always connected by a suitable large number M of hamiltonian powers $\langle x' | H^M | x \rangle \neq 0$), then a more restrictive property holds: the ground state is unique for any $\gamma \geq 0$. This implies immediately that:

(vi) the ground state energy $E(\gamma)$ of the fixed node hamiltonian H_{FN}^γ is an analytic function of γ , due to the finite size gap separating the unique ground state from the first excited state. We assume in the following that this very general property holds for the given hamiltonian a condition which is not restrictive, also considering that if ergodicity is not satisfied, all previous and the following considerations hold in all the subspaces of configurations x ergodically connected by the powers of the hamiltonian.

By using Green Function Monte Carlo the ground state energy $E(\gamma)$ can be very efficiently computed for $\gamma > 0$ as all the matrix elements of the importance sampled Green function $G_{x',x}^{FN} = \psi_G(x') [\Lambda \delta_{x',x} - (H_{FN}^\gamma)_{x',x}] / \psi_G(x)$ are all positive for large enough constant shift Λ . This is obtained by averaging the local energy $\langle E_L(x) \rangle$ over the configurations x generated statistically by the Green function G^{FN} with a standard algorithm.^{7,12,13} Notice also that, by property (iv), the local energy E_L of this fixed node hamiltonian is proportional to the local energy e_L of H and therefore this computation satisfy the so called zero variance property: both E_L and e_L have zero statistical variance if ψ_G is an exact eigenstate of H .

For $r = 0$ H_{FN}^γ reduces to the standard fixed node hamiltonian defined in⁸ ($\gamma = 0$) and extended to $\gamma \neq 0$ in.¹⁴ Thus a rigorous theorem holds relating the ground state energy $E(\gamma)$ of the fixed node ground state ψ_{FN}^γ of H_{FN}^γ , to its variational expectation value $E^{FN}(\gamma) = \langle \psi_{FN}^\gamma | H | \psi_{FN}^\gamma \rangle$ on the hamiltonian H :

$$E^{FN}(\gamma) \leq E(\gamma) \leq \langle \psi_G | H | \psi_G \rangle \quad (10)$$

Using property (i) we therefore notice that by increasing the value of r from the variational value $r = -1/(1 + \gamma)$ up to $r = 0$ the ground state of the fixed node hamiltonian H_{FN}^γ becomes a variational state with lower energy expectation value. This implies immediately that the fixed node effective hamiltonian is more appropriate to describe the ground state of H .

In the continuous case r cannot be extended to positive values because the local energy e_L may assume arbitrary large negative values close to the nodes, and the best variational energy can be actually obtained just for $r = 0$ (since for $r = 0$ the fixed node gives the lowest possible energy compatible with the nodes of the guiding function). In a lattice case such a theorem is missing, and there is no reason to expect that $r = 0$ is just the optimal value.

A simple and efficient scheme to compute a variational upper bound of the energy for any r is described in the following paragraphs. Using property (iii)

$$E_{FN}(\gamma) = \langle \psi_{FN}^\gamma | H^\gamma - (1 + \gamma) \frac{dH_{FN}^\gamma}{d\gamma} | \psi_{FN}^\gamma \rangle = E(\gamma) - (1 + \gamma) \frac{dE(\gamma)}{d\gamma} \quad (11)$$

where in the latter equality the Hellmann-Feynmann theorem has been used. By using that

H_{FN}^γ depends linearly on γ , the well known convexity property of $E(\gamma)$ holds¹⁵ :

$$\frac{d^2E(\gamma)}{d\gamma^2} \leq 0 \quad (12)$$

Therefore the expectation value $E_{FN}(\gamma)$ of the hamiltonian H on the fixed node state is a monotonically increasing function of γ , as clearly $\frac{dE_{FN}(\gamma)}{d\gamma} = -(1 + \gamma) \frac{d^2E(\gamma)}{d\gamma^2} \geq 0$. The best variational estimate is obtained therefore for $\gamma = 0$, as in the conventional scheme.

The extension to finite γ is however convenient to provide better variational estimates of $E_{FN}^{\gamma=0}$, which in fact maybe sizable lower than the standard estimate $E_{FN}(0) \leq E(0)$ for $r = 0$. This extension allows also to make a rigorous upper bound of E_{FN}^γ also in the case $r > 0$, without missing the zero variance property. In fact, always by the convexity property of $E(\gamma)$,

$$-\frac{dE(\gamma)}{d\gamma}|_{\gamma=0} \leq -\frac{E(\gamma) - E(0)}{\gamma} \quad (13)$$

we finally get that at the best variational condition $\gamma = 0$

$$E_{FN}(0) \leq E(0) - (E(\gamma) - E(0))/\gamma. \quad (14)$$

For $r = 0$ the above upper bound improves also the previously known value (10), at least for γ small enough where the above inequality becomes a strict equality.

In practice, since the energy as a function of γ is almost linear a very good estimate can be obtained using the above inequality even for $\gamma = 1$, as shown in Fig.(2) for a test example on the $t - J$ model, where it is also clear that the variational energy can be improved by turning on the parameter r .

4 The Generalized Lanczos

The optimization of the parameter r is rather problematic within the scheme of the previous section especially when few Lanczos steps are applied to the guiding function and the dependence of the energy as a function of r cannot be resolved within available statistical errors. Though the energy maybe rather insensitive to r , the behavior of correlation functions, may strongly depend on it, especially when the guiding function shows some instability towards different phases not described at the variational level. Within this approach the instability of the guiding function is characterized by the existence of a considerable number of configurations x with local energy $e_L(x)$ much below the average and with correlation properties much different than the average. By increasing r these configurations will have larger and larger weight in the fixed node ground state ψ_{FN}^γ (since they have much lower-energy diagonal term) and will display clearly the possible instabilities of the variational wavefunction ψ_G .

The sign-flip term $\mathcal{V}_{sf}(x)$ is divergent whenever the guiding function is exceedingly small (i.e. close to the nodes or finite-size lattice pseudo-nodes of ψ_G), thus requiring an infinite shift Λ ,¹⁴ because for the statistical implementation of the power method the diagonal term $\Lambda - (H_{FN}^\gamma)_{x,x} = \Lambda - H_{x,x} - (1 + \gamma)\mathcal{V}_{sf}(x) - r(1 + \gamma)e_L(x)$ (see Eq. 8) has to be non negative. For $r = -1/(1 + \gamma)$, in the variational case, a better approach, but similar in spirit, is obtained by sampling¹⁶ the square of the variational wavefunction

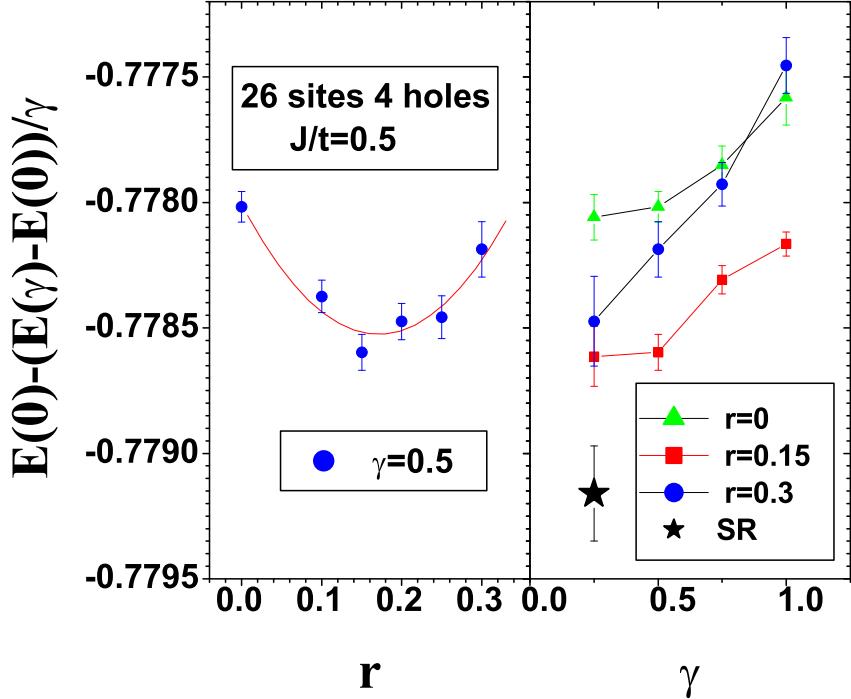


Figure 2. Variational energy of the t-J hamiltonian as a function of the parameters r and γ , for the BCS-guiding function (5), without any Lanczos improvement. The $\gamma \rightarrow 0$ limit in the right panel corresponds to the expectation value $E^{FN}(\gamma) = \langle \psi_{FN}^\gamma | H | \psi_{FN}^\gamma \rangle$ for $\gamma = 0$ where ψ_{FN}^γ is the ground state of the effective hamiltonian H_{FN}^γ . Each point, due to inequality (14), represents an upper bound for $E^{FN}(\gamma = 0)$ and, clearly, for the ground state of H . All the estimates reported here are much better than the standard $r = 0$ lattice fixed node upper bound $E(\gamma = 0)^8$ for $E^{FN}(\gamma = 0)$: $E(\gamma = 0) = -0.77580(2)$ much above the upper energy scale. The value (SR) obtained with the “generalized Lanczos” described in the following sections is also shown for comparison.

ψ_G with a different Green function. This following importance sampled Green function is used for the statistical implementation of the power method:

$$G_{x',x}^\gamma = \begin{cases} \frac{1}{z_{x'}}(\Lambda - H_{x,x}) & \text{for } x' = x \\ -\frac{1}{z_{x'}}\psi_G(x')(H_{FN}^\gamma)_{x',x}/\psi_G(x) & \text{for } x' \neq x \end{cases} \quad (15)$$

where in H_{FN}^γ (Eq. 8) appearing in the above equation the parameter r is set to the variational value $r = -1/(1 + \gamma)$, z_x is a normalization factor obtained by setting $\sum_{x'} z_{x'} G_{x',x}^\gamma = z_x$, namely:

$$z_x = \Lambda - e_L(x) + (1 + \gamma)\mathcal{V}_{sf}(x) \quad (16)$$

In this way it is straightforward to show that:

$$\sum_x G_{x',x}^\gamma |\psi_G(x)|^2 = |\psi_G(x)|^2 \quad (17)$$

Thus the importance-sampled Green function G^γ maybe used to generate configurations that sample the variational wavefunction square. The advantage of the present approach is evident since the diagonal term of the Green function does not contain the sign-flip term, and a finite reasonable Λ can be used. For instance in the $t - J$ model Λ can be set to zero. Instead a zero shift is not allowed for the importance sampled Green function of the effective hamiltonian itself:

$$G_{FN} = \psi_G(x') [\Lambda - (H_{FN}^\gamma)_{x',x}] / \psi_G(x) \quad (18)$$

which performs the same task for $r = -1/(1 + \gamma)$, but with a less efficient infinite Λ scheme.¹⁴

In the following, within the spirit of the “effective hamiltonian approach”, the variational wavefunction is improved by tuning a parameter r proportional to the local energy, in order to modify and improve the effective hamiltonian H_{FN}^γ , whose ground state is just ψ_G for $r = -1/(1 + \gamma)$. This parameter is then changed in order to be as close as possible to the true hamiltonian for $\gamma \geq 0$, when computations free of sign problem are possible. Indeed in order to improve H_{FN}^γ it is very useful to notice that $H_{FN}^\gamma = H$, the exact hamiltonian, for $\gamma = -1$ and any non-zero r . Thus at finite positive γ an optimal variational parameter r can be used, that on a lattice, maybe significantly different from the fixed node value $r = 0$, since this value represents the optimal one only in a continuous model, when there exists a rigorous proof that $r = 0$ provides the minimum possible energy.

In order to determine a feasible scheme for the optimization of r in the lattice case, we need to implement small modifications of the Green function (15). We notice that there are two important changes of this Green function that are easily implemented.

4.1 One Lanczos Step Improvement

In this case the Green function (15) is modified by:

$$G_{1LS}^\gamma = r_{x'} G_{x',x}^\gamma / r_x \quad (19)$$

where $r_x = 1 + \alpha e_L(x)$. After applying statistically the above Green function, after a large number of iterations the configurations x , will be distributed according to the weight (not necessarily positive):

$$\psi_G(x)\psi_1(x)$$

where

$$\psi_1 = (1 + \alpha H)|\psi_G\rangle = \sum_x r_x \psi_G(x)|x\rangle \quad (20)$$

is the first Lanczos step wavefunction as described in Eq. (1). Since the Lanczos iteration improves the wavefunction and the factor r_x has not a definite sign on each configuration x , it is clear that the phases of the ground state wavefunction are much better represented

by the signs of $r_x \psi_G(x)$ rather than by the ones corresponding to $\psi_G(x)$. The parameter $\alpha = \alpha_1/\alpha_0$ can be determined by satisfying the SR conditions:³

$$\begin{aligned}\langle \psi_G | H(\alpha_0 + \alpha_1 e_L) | \psi_G \rangle &= \langle \psi_G | H(\Lambda - H) | \psi_n \rangle \\ \langle \psi_G | (\alpha_0 + \alpha_1 e_L) | \psi_G \rangle &= \langle \psi_G | (\Lambda - H) | \psi_n \rangle\end{aligned}\quad (21)$$

where $\alpha_i, i = 0, 1$ are computed statistically at any given iteration n in order to improve the SR state $r_x \psi_n(x)$, until convergence is reached for large n . In this case $\psi_n(x)$ is independent of n and statistically equal to ψ_G , whereas α will converge (statistically) to the exact one Lanczos step value. Once this value is determined the energy expectation value over ψ_1 can be evaluated by statistically averaging the local energy $e_L(x)$ corresponding to ψ_G (and not to ψ_1), providing a substantial reduction of computational effort. In this case, since the value of γ is immaterial for the statistical averages, it is more convenient to use $\gamma = 1$, that minimizes statistical fluctuations.

In general, the use of the SR conditions³ allows to obtain the energy and correlation expectation values of the p -Lanczos step wavefunction ψ_p , by using a guiding function ψ_G containing only $p - 1$ powers of the Hamiltonian, e.g. $|\psi_G\rangle \rightarrow |\psi_{p-1}\rangle$. The use of $|\psi_{p-1}\rangle$ as a guiding function for sampling ψ_p may not be the optimal choice. In the following we describe a guiding function with better nodes than ψ_{p-1} but with the same number $p - 1$ of hamiltonian powers, that will be used in the following sections whenever the method SR will be applied,

Using the root decomposition (4) of the H -polynomial defining the p -Lanczos step wavefunction $|\psi_p\rangle$, we can single out any real root z_k and similarly to the first Lanczos step case:

$$\begin{aligned}\psi_p(x) &= r_x \psi_G(x) \quad \text{with} \\ r_x &= 1 - e_L(x)/z_k \\ |\psi_G\rangle &\rightarrow \prod_{i \neq k} (1 - H/z_i) |\psi_G\rangle\end{aligned}\quad (22)$$

The new local energy $e_L(x)$, obtained with the new guiding function, will keep into account the phases of the p -Lanczos step wavefunction exactly. In this way, within this decomposition, it is clear that the best guiding function ψ_G of the previous form, is obtained by choosing the real root z_k such that:

$$<1 - e_L(x)/z_k> \quad (23)$$

is as far as possible (on average over ψ_G) from the zero value. This condition (23) will minimize the sign changes of $\psi_G(x)$ to obtain $\psi_p(x) = (1 - e_L(x)/z_k) \psi_G(x)$, thus providing the best possible phases that we can safely obtain with $p - 1$ powers of the hamiltonian applied to the bare ψ_G .

4.2 Fixed Node Improvement

In this case the Green function is modified similarly:

$$G'_{FN} = r_{x'} G_{x',x}^\gamma / Sgn(r_x) \quad (24)$$

It is easily obtained that for $r_x = 1 - \frac{1+r(1+\gamma)}{\Lambda} e_L(x)$ and large shift Λ , the effective hamiltonian H_{FN}^γ (8) is indeed considered, as for $\Lambda \rightarrow \infty$ the matrix elements of G_{FN} (18) coincide with the ones defined above for $\Lambda G'_{FN}$, up to $O(\frac{1}{\Lambda})$.

In particular for $r = 0$, and $\gamma = 0$ we recover the standard fixed node.⁸ Notice also that, if the hamiltonian is free of sign problem $\mathcal{V}_{sf}(x) = 0$ and the Fixed node is exact. Then the choice $r = 0$ provides the exact sampling of the ground state of H even for finite Λ , as the factor r_x is proportional to z_x (16) and simplifies in (18,15).

4.3 Generalized Lanczos

Using the above Green function (24), the parameter $r = \frac{-(\Lambda \alpha_1 / \alpha_0) - 1}{1 + \gamma}$, a single parameter at any order p of the Lanczos iterations, is optimized using the SR conditions(21) with ψ_n now depending explicitly on n and differing from the initial guiding function ψ_G : $r_x \psi_n(x) = (G'_{FN})^n \psi_G$. These conditions provide, as mentioned before, α_0, α_1 statistically.³ However, in this case, the parameter r , determined by the SR condition, may not coincide with the lowest possible energy condition. A further modification of the Green function³

$$G'_\eta = r_{x'} G_{x',x}^\gamma / |r_x|^{1-\eta} Sgn(r_x) \quad (25)$$

that interpolates between the Lanczos limit (19) for $\eta = 0$ (when the SR conditions coincide with the Euler condition of minimum energy) and the Fixed node limit (24) for $\eta = 1$ allows to overcome this difficulty, as we get closer but not exactly equal to the Lanczos limit, and one can obtain even lower variational energies.³

For the $t - J$ model we avoid to consider here this extra-complication, since the SR conditions (21) have been tested to coincide almost exactly with the Euler conditions of minimum energy (see Fig. 2) even for $\eta = 1$ at least for $\Lambda = 0$. As shown in the same figure the SR may also provide a slightly lower energy than the corresponding one obtained by the best r effective hamiltonian H_{FN}^γ , because for small Λ the factor r_x in Eq. (24) may change sign and can correct also the phases of the wavefunction and not only the amplitudes. This is also the reason to work with the minimum possible shift Λ . In principle it is possible to further improve the variational energy and the nodes of the sampled wavefunction, by performing the reconfiguration scheme each k_p steps, with an effective Green function:

$$G'_{k_p} = r_{x'} (G^\gamma)_{x',x}^{k_p} / Sgn(r_x) \quad (26)$$

For $\gamma = 1$, it is possible to work with $k_p > 1$ and with reasonable statistical fluctuations (that increase obviously with k_p). By increasing k_p the factor r_x provides non trivial changes to the phase of the wavefunction with corresponding improvement in energy expectation value. We have not systematically studied this possible modification of the method so far. This extension to $k_p > 1$ should be clearly useful for model hamiltonians, such as the Hubbard model at strong coupling, when a large shift Λ is required for the convergence of the method.

For $\Lambda = 0$ or finite, the coefficient r in the factor r_x may have little to do with the coefficient appearing in H_{FN}^γ , but, even at finite Λ , an effective hamiltonian can be still defined,³ which is qualitatively similar to H_{FN}^γ . In the following discussions we will not consider the difference between the finite Λ effective hamiltonian and the infinite Λ one (8) because it is irrelevant for our purposes.

At each iteration p of the generalized Lanczos the special guiding function described in Eq. (22) is used, yielding optimal phases as close as possible to the p -Lanczos step wavefunction. As far as the remaining parameter γ , this is restricted to be positive for statistical reasons (no sign problem). Clearly from property (12), the smaller is γ , the better is the variational energy but increased fluctuations occurs for computing the SR conditions (21). On the other hand, the Green-function shift Λ has to be taken as small as possible, compatibly with $\Lambda - H_{x,x} > 0$ for any x , in order to further improve the efficiency of the power method. Within the SR method by minimizing at best the parameters γ and Λ (or increasing k_p) we can further improve this technique, in a practical scheme. The optimization of the parameter r , since it affects a change in the effective hamiltonian H^γ is particularly important for correlation functions. Instead all the other parameters (including η or k_p for instance) may help to obtain slightly lower variational energies, but are in general much less important. The variational SR results for the $t - J$ model, described in the following sections, are obtained with $\gamma = 1/4$ and $\Lambda = 0$ and refer to the fixed node Green function (24), whereas the symbol FN will always refer to the standard fixed-node case $\Lambda \rightarrow \infty$, $\gamma = r = 0$.

5 Results on the t-J Model

We consider the pairing correlations in the $t - J$ model for square clusters with periodic boundary conditions:

$$\begin{aligned} P_{i,j;k,l} &= \langle \Delta_{i,j}^\dagger \Delta_{k,l} \rangle \\ \Delta_{i,j}^\dagger &= c_{i,\uparrow}^\dagger c_{j,\downarrow}^\dagger + i \leftrightarrow j \end{aligned} \quad (27)$$

$\Delta_{i,j}^\dagger$ creates a singlet pair in the sites i, j . On each lattice we take the first nearest neighbor pair i, j fixed and move k, l parallel or perpendicular to the direction i, j . In all cases studied the parallel correlations are positive and the perpendicular ones are negative, consistent with a d -wave symmetry of the pairing. The existence of phase coherence in the thermodynamic limit is obtained whenever $P_{i,j;k,l}$ remains finite for large distance separation between the pairs i, j and k, l . A systematic study has been reported in.¹⁷ Here we focus only on few test cases to show the power of the method, and the importance to work with an effective hamiltonian H_{FN}^γ with a single variational parameter r as described in the previous section. For all cluster used the distance between pair i, j and pair k, l refers to the minimum one between $|R_i - R_k|$, $|R_i - R_l|$, $|R_j - R_k|$ and $|R_j - R_l|$. Only for the 6×6 we use the so called Manhattan distance $|(x, y)| = |x| + |y|$, since the pair (k, l) in this case is moved in both perpendicular directions. First the pair (k, l) is translated parallel to the x -axis up to the maximum distance allowed by PBC, and then (for the 6×6) the pair (k, l) is moved parallel to the y -axis.

First of all, whenever the initial variational wavefunction used is qualitatively correct (5), few Lanczos iterations are really enough to obtain exact ground state properties. This is clearly shown in Fig. (3) where the exact results coincide within few error bars with the variance extrapolated results, that in turn are very close to the $p = 2$ Lanczos wavefunction results. However for larger system when the solution is not known, few Lanczos iterations, though systematically improving the energy, cannot change qualitatively the pairing correlations of the initial wavefunction, and in general the variational approach is not reliable.

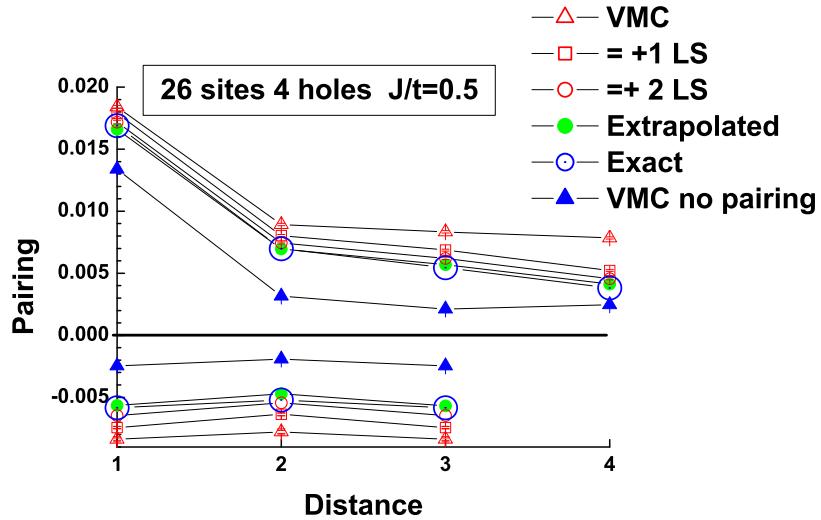


Figure 3. Pairing correlations in the 26 lattice for 4 holes in the $J/t = 0.5$ $t-J$ model for the variational Lanczos technique as compared with the exact result obtained with exact diagonalization. The variance extrapolated values are obtained using only the $p = 0, 1, 2$ results available with the statistical algorithm also for much larger system size.

In order to show this effect, we have used two different variational wavefunctions on a 6×6 4-holes $J/t = 0.5$ cluster, and improved both initializations with the methods described in the previous section: the pure variational Lanczos technique, the standard Fixed node (FN) and the “generalized Lanczos method” (SR), within the simplified scheme considered before. For one wavefunction initialization, the BCS variational parameters are optimized by minimizing the energy, for the other one we have reduced to a very small values $\simeq 10^{-4}$ the corresponding variational parameter Δ_{BCS} in (6), just in order to remove the degeneracy of the free-electron determinant in the 6×6 . This choice yields a variational wavefunction with definite quantum numbers and with small pairing correlations.

We see in Fig. (4), top panels, that the Lanczos technique is very much dependent on the two different initial choices, even though the energy is in both cases very much improved by few Lanczos iterations. As shown in Fig. (5), the variance extrapolated results of the energy are consistent for both initial wavefunctions. On the other hand the pairing correlations remain inconsistent for about a factor two at large distance.

In this example we clearly see the limitation of the straightforward variational technique: within a very similar energy (e.g. the extrapolated ones) the pairing correlations maybe even qualitatively different.

A completely different behavior is obtained as soon as the FN is applied (middle panels in Fig. 4). The energy improvement within this technique is apparently marginal compared to the standard Lanczos technique (see Fig. 5). Instead the behavior of pairing correlations

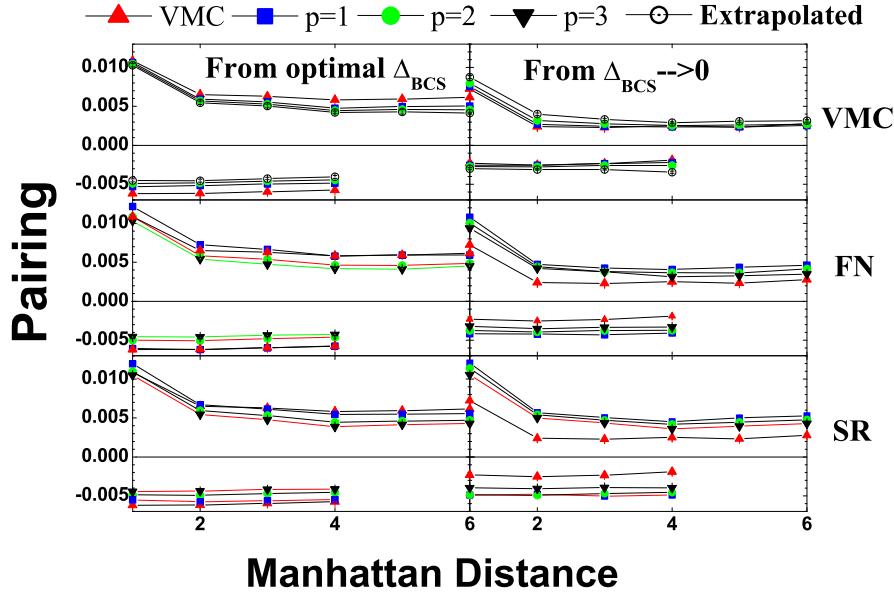


Figure 4. Pairing correlations in the 6×6 lattice for 4 holes in the $J/t = 0.5 t - J$ model. Left panels and right panels refer to two different initial guiding functions with or with vanishing small d-wave order parameter respectively. The latter is used in order to remove the degeneracy of the free electron Slater-determinant. The panels at different rows refer to different methods, as a function p of the hamiltonian powers used to evaluate the local energy e_L , required by all the methods: the larger is p , the more (L^p for $p \geq 2$) computationally demanding is the calculation. The VMC values (red triangles) are plotted in all panels for comparison.

is much better, and already the simple fixed node approximation applied to the pairing correlations is rather independent of the initial wavefunction. The only drawback of this technique is that when systematic improvements to the variational wavefunction are implemented (larger p in the figure), the convergence properties are not behaving so accurately, as one could expect from the convergence of the energy reported in Fig. (5). In particular, even at the most accurate level of this fixed-node approximation – namely the fixed node over the two Lanczos step wavefunction – the two different initializations give pairing correlations differing by about 20% at the largest distance. This is much better than the straightforward Lanczos variational technique (this difference was about 70% for the corresponding two Lanczos step wavefunctions) but is not satisfactory enough.

The reason of such behavior is easily understood in terms of the effective hamiltonian approach. In a lattice case it appears really important for correlation functions to optimize the parameter r appearing in the effective hamiltonian (8) and not just taking the FN ansatz $r = 0$. This optimization scheme is particularly important whenever some correlations that are not included at the variational level (or much depressed as in the case studied) are increasing as we go down in energy with the help of the improved $p - 1$ ($p > 1$)

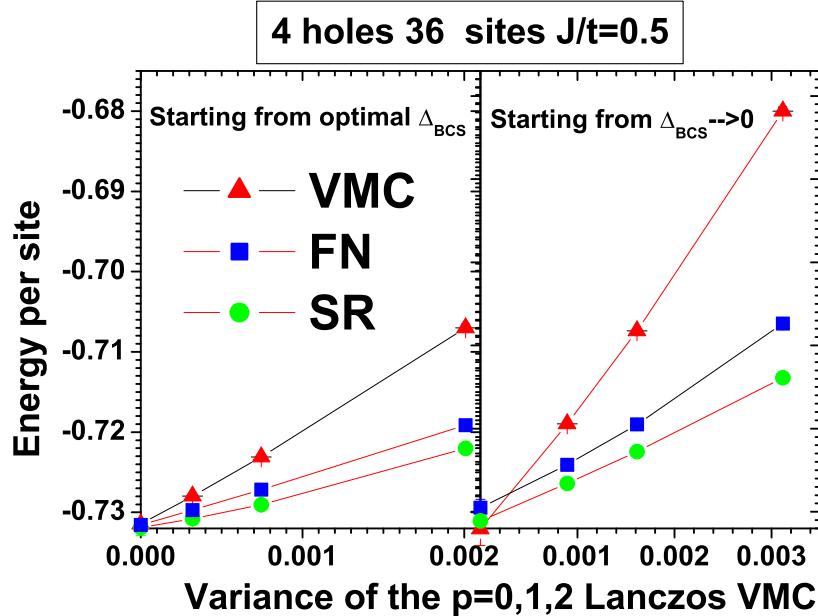


Figure 5. Variational energies obtained with various methods as a function of the variance per site σ^2 / L^2 of the p -Lanczos step wavefunction (VMC), which is improved either with standard Fixed node (FN) or the generalized Lanczos (SR), with the simplified and efficient scheme described in the previous section. The values at zero variance are extrapolations with a quadratic least square fit.

Lanczos step guiding function. In general for larger p the parameter r increases, thus the SR scheme provides correlation functions substantially different and more accurate than the FN. In the bottom panels it is remarkable that, after applying only 3–steps of the SR technique, both initializations *provide the same results within error bars ($\leq 3\%$) at the largest distance*. These results can be considered benchmark accurate calculations of pairing correlations in the 6×6 cluster. These pairing correlations clearly indicate a robust d -wave superconducting ground state in the $t - J$ model, at least for this J/t ratio. In this example we notice that correlation functions, in the effective hamiltonian approach, begin to be consistent within 5% whenever the variational energy is accurate within $\sim 1\%$, that is at least one order of magnitude better than a straightforward variational technique like the Lanczos one.

Of course for larger size, consistent correlation functions, i.e. independent from the initial wavefunction with or without Δ_{BCS} , can be obtained for a larger number p of SR-iterations. Here we report a sample case for a 50 site cluster at small $J/t = 0.1$. We see in Fig. (6) that the sizable pairing correlations present in the variational wavefunction with $\Delta_{BCS} \neq 0$, represents just an artifact of the variational calculation. At the third step, of the SR technique, when, as shown in Fig. (7) we reach an accuracy in energy below 1%

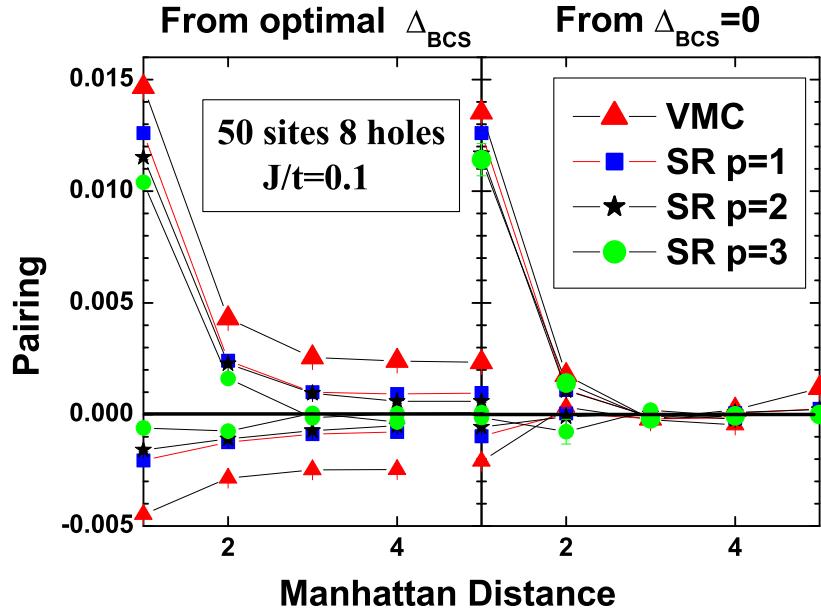


Figure 6. Pairing correlations in the 50 site lattice for 8 holes in the $J/t = 0.1$ $t - J$ model. Left panels and right panels refers to different initial guiding function with or without d-wave order parameter respectively. The pairing correlations for both calculations are consistently small at the most accurate level of approximation (SR $p = 3$).

(assuming that the variance extrapolated energies-both consistent- are exact), the pairing correlations are again consistent within few error bars, and clearly vanishingly small at large distance.

6 Conclusions

We have shown that within a brute force variational technique, such as the Lanczos method for few iterations, it is hard to obtain accurate values of correlation functions unless the energy accuracy is far from the present possibilities, at least in two dimensions. An accuracy of about one part over 10^4 in the energy would be probably possible with at least 10 Lanczos steps or 100000 states in DMRG 2D calculations for systems of about 100 sites with periodic boundary conditions. This kind of accuracy maybe enough to obtain consistent correlation functions even within these two variational methods, but is far from being possible at present.

We have shown instead that a qualitatively different and very promising approach, based on the optimization of an effective hamiltonian, rather than adding more variational parameters in a brute force variational scheme, appears to be very useful to control correla-

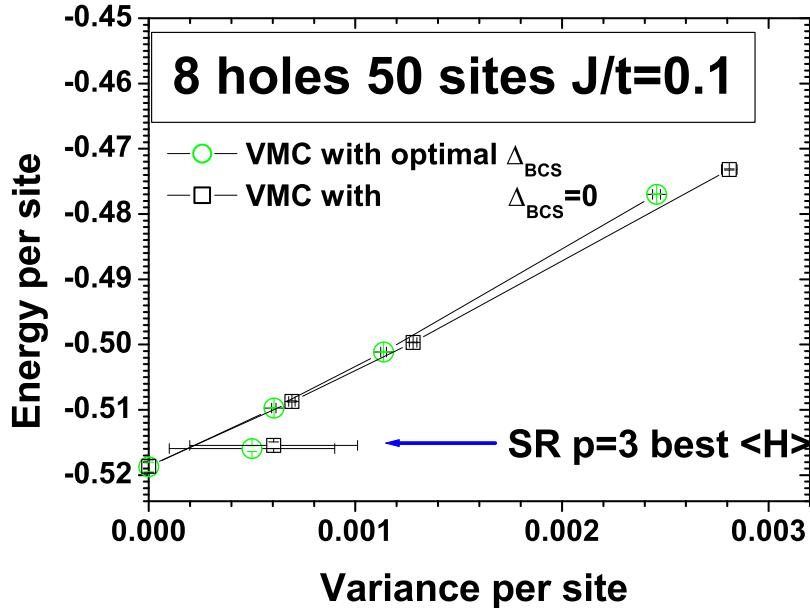


Figure 7. Variational energy as a function of the variance per site σ^2/L^2 for the p -Lanczos step wavefunction (VMC), which is improved by the “generalized Lanczos method” (SR). The best variational SR $p = 3$ energies are indicated by the arrows.

tion functions. The idea is based on the “effective hamiltonian approach” described in the introduction. In this scheme it is assumed that between similar Hamiltonians, the correlation functions of their ground states should be also similar. The SR technique, allows to systematically improve the effective hamiltonian considered even compared to the lattice fixed node one,⁸ with an iterative scheme very similar to the Lanczos one, thus the name “generalized Lanczos”.

Within this scheme it is clear that there are robust pairing correlations in the $t - J$ model at sizable but not unphysical value of J/t .¹⁷ However there exists a critical value $(J/t)_c \geq 0.1$ below which pairing correlations are clearly suppressed. The existence of such a critical $(J/t)_c$ is clearly understood because at $J/t = 0$, the ferromagnetic instability takes place even at large doping.¹⁸

Acknowledgments

We are indebted to F. Becca, L. Capriotti, A. Parola and E. Dagotto for many useful discussions. This work was partially supported by MIUR, (COFIN-2001) and INFM (Pais-Malodi).

References

1. G. Baskaran, Z. Zou, and P. W. Anderson, Solid State Comm. **63**, 973–975 (1987).
2. F. C. Zhang and T. M. Rice, Phys. Rev. B **37**, 3759,3763 (1988).
3. S. Sorella, *Generalized Lanczos algorithm for variational quantum Monte Carlo*, Phys. Rev. B **64**, 024512 1–16 (2001).
4. S. R. White and D. Scalapino, Phys. Rev. Lett. **80**, 1272-1275 (1998); ibidem Phys. Rev. B **60**, 753 (1999).
5. C. Y. Shih, Y. C. Chen, and H. Q. Lin Phys. Rev. Lett. **81**, 1294-1297 (1998).
6. L. Capriotti, F. Becca, A. Parola and S. Sorella *Resonating Valence Bond Wavefunctions for Strongly Frustrated Spin Systems* Phys. Rev. Lett. **87**, 098201 (2001).
7. N. Trivedi and D. M. Ceperley, Phys. Rev. B **41**, 4552 (1990).
8. D. F. B. ten Haaf, J. M. J. van Leeuwen, W. van Saarloos, and D. M. Ceperley, Phys. Rev. B **51**, 13039 (1995).
9. *see e.g.* D. M. Ceperley, J. Stat. Phys. **63**, 1237 (1991).
10. C. Gros, Phys. Rev. B **38**, 931 (1988).
11. H. Yokoyama and H. Shiba, J. Phys. Soc. Jpn. **57**, 2482 (1988).
12. K. Runge, Phys. Rev. B **45**, 12292 (1992); ibidem 7229.
13. M. Calandra and S. Sorella, *Numerical study of the two dimensional Heisenberg model using a Green function Monte Carlo technique with a fixed number of walkers*, Phys. Rev. B **57**, 11446-11456 (1998).
14. S. Sorella and L. Capriotti, Phys. Rev. B **61**, 2599 (2000).
15. This follows by writing $\gamma = p\gamma_1 + (1 - p)\gamma_2$, for any γ_1, γ_2 and $0 \leq p \leq 1$, thus finding a variational lower bound for $E(\gamma) \geq pE(\gamma_1) + (1 - p)E(\gamma_2)$ because the ground state energy $E(\gamma)$ of H^γ is certainly bounded by the minimum possible energy that can be obtained by each of the two terms in the RHS of the following equation: $H^\gamma = pH^{\gamma_1} + (1-p)H^{\gamma_2}$. The above inequality represents just the convexity property of $E(\gamma)$.
16. C. S. Hellberg and E. Manousakis, Phys. Rev. B **61**, 11787 (2000).
17. S. Sorella, G. Martins, F. Becca, C. Gazza, A. Parola and E. Dagotto, *Superconductivity in the two dimensional t-J model* cond-mat/0110460 , (2001).
18. F. Becca and S. Sorella, *Nagaoka Ferromagnetism in the Two-Dimensional Infinite-U Hubbard Model*, Phys. Rev. Lett. **86**, 3396-3399 (2001).

The LDA+DMFT Approach to Materials with Strong Electronic Correlations

Karsten Held¹, Igor A. Nekrasov², Georg Keller³, Volker Eyert⁴, Nils Blümer⁵, Andrew K. McMahan⁶, Richard T. Scalettar⁷, Thomas Pruschke³, Vladimir I. Anisimov² and Dieter Vollhardt³

¹ Physics Department, Princeton University, Princeton, NJ 08544, USA

² Institute of Metal Physics, Russian Academy of Sciences-Ural Division
Yekaterinburg GSP-170, Russia

³ Theoretical Physics III, Center for Electronic Correlations and Magnetism
Institute for Physics, University of Augsburg, 86135 Augsburg, Germany
E-mail: Dieter.Vollhardt@Physik.Uni-Augsburg.DE

⁴ Institute for Physics, Theoretical Physics II, University of Augsburg
86135 Augsburg, Germany

⁵ Institute for Physics, Johannes Gutenberg University, 55099 Mainz, Germany

⁶ Lawrence Livermore National Laboratory, University of California
Livermore, CA 94550, USA

⁷ Physics Department, University of California, Davis, CA 95616, USA

LDA+DMFT is a novel computational technique for *ab initio* investigations of real materials with strongly correlated electrons, such as transition metals and their oxides. It combines the strength of conventional band structure theory in the local density approximation (LDA) with a modern many-body approach, the dynamical mean-field theory (DMFT). In the last few years LDA+DMFT has proved to be a powerful tool for the realistic modeling of strongly correlated electronic systems. In this paper the basic ideas and the set-up of the LDA+DMFT(X) approach, where X is the method used to solve the DMFT equations, are discussed. Results obtained with X=QMC (quantum Monte Carlo) and X=NCA (non-crossing approximation) are presented and compared. By means of the model system $\text{La}_{1-x}\text{Sr}_x\text{TiO}_3$ we show that the method X matters qualitatively and quantitatively. Furthermore, we discuss recent results on the Mott-Hubbard metal-insulator transition in the transition metal oxide V_2O_3 and the α - γ transition in the 4f-electron system Ce.

1 Introduction

The calculation of physical properties of electronic systems by controlled approximations is one of the most important challenges of modern theoretical solid state physics. In particular, the physics of transition metal oxides – a singularly important group of materials both from the point of view of fundamental research and technological applications – may only be understood by explicit consideration of the strong effective interaction between the conduction electrons in these systems. The investigation of electronic many-particle systems is made especially complicated by quantum statistics, and by the fact that the investigation of many phenomena require the application of non-perturbative theoretical techniques.

From a microscopic point of view theoretical solid state physics is concerned with the investigation of interacting many-particle systems involving electrons and ions. However,

it is an established fact that many electronic properties of matter are well described by the purely electronic Hamiltonian

$$\hat{H} = \sum_{\sigma} \int d^3r \hat{\Psi}^{+}(\mathbf{r}, \sigma) \left[-\frac{\hbar^2}{2m_e} \Delta + V_{\text{ion}}(\mathbf{r}) \right] \hat{\Psi}(\mathbf{r}, \sigma) \\ + \frac{1}{2} \sum_{\sigma\sigma'} \int d^3r d^3r' \hat{\Psi}^{+}(\mathbf{r}, \sigma) \hat{\Psi}^{+}(\mathbf{r}', \sigma') V_{ee}(\mathbf{r}-\mathbf{r}') \hat{\Psi}(\mathbf{r}', \sigma') \hat{\Psi}(\mathbf{r}, \sigma), \quad (1)$$

where the crystal lattice enters only through an ionic potential. The applicability of this approach may be justified by the validity of the Born and Oppenheimer approximation.¹ Here, $\hat{\Psi}^{+}(\mathbf{r}, \sigma)$ and $\hat{\Psi}(\mathbf{r}, \sigma)$ are field operators that create and annihilate an electron at position \mathbf{r} with spin σ , Δ is the Laplace operator, m_e the electron mass, e the electron charge, and

$$V_{\text{ion}}(\mathbf{r}) = -e^2 \sum_i \frac{Z_i}{|\mathbf{r} - \mathbf{R}_i|} \quad \text{and} \quad V_{ee}(\mathbf{r} - \mathbf{r}') = \frac{e^2}{2} \sum_{\mathbf{r} \neq \mathbf{r}'} \frac{1}{|\mathbf{r} - \mathbf{r}'|} \quad (2)$$

denote the one-particle potential due to all ions i with charge eZ_i at given positions \mathbf{R}_i , and the electron-electron interaction, respectively.

While the *ab initio* Hamiltonian (1) is easy to write down it is impossible to solve exactly if more than a few electrons are involved. Numerical methods like Green's Function Monte Carlo and related approaches have been used successfully for relatively modest numbers of electrons. Even so, however, the focus of the work has been on jellium and on light atoms and molecules like H, H₂, ³He, ⁴He, see, e.g., the articles by Anderson, Bernu, Ceperley *et al.* in the present Proceedings of the *NIC Winter School 2002*. Because of this, one generally either needs to make substantial approximations to deal with the Hamiltonian (1), or replace it by a greatly simplified model Hamiltonian. At present these two different strategies for the investigation of the electronic properties of solids are applied by two largely separate groups: the density functional theory (DFT) and the many-body community. It is known for a long time already that DFT, together with its local density approximation (LDA), is a highly successful technique for the calculation of the electronic structure of many real materials.² However, for strongly correlated materials, i.e., *d*- and *f*-electron systems which have a Coulomb interaction comparable to the band-width, DFT/LDA is seriously restricted in its accuracy and reliability. Here, the model Hamiltonian approach is more general and powerful since there exist systematic theoretical techniques to investigate the many-electron problem with increasing accuracy. These many-body techniques allow one to describe qualitative tendencies and understand the basic mechanism of various physical phenomena. At the same time the model Hamiltonian approach is seriously restricted in its ability to make quantitative predictions since the input parameters are not accurately known and hence need to be adjusted. One of the most successful techniques in this respect is the dynamical mean-field theory (DMFT) – a non-perturbative approach to strongly correlated electron systems which was developed during the past decade.^{3–11} The LDA+DMFT approach, which was first formulated by Anisimov *et al.*,^{12,13} combines the strength of DFT/LDA to describe the weakly correlated part of the *ab initio* Hamiltonian (1), i.e., electrons in *s*- and *p*-orbitals as well as the long-range interaction of the *d*- and *f*-electrons, with the power of DMFT to describe the strong correlations induced by the local Coulomb interaction of the *d*- or *f*-electrons.

Starting from the *ab initio* Hamiltonian (1), the LDA+DMFT approach is presented in Section 2, including the DFT in Section 2.1, the LDA in Section 2.2, the construction of a model Hamiltonian in Section 2.3, and the DMFT in Section 2.4. As methods used to solve the DMFT we discuss the quantum Monte Carlo (QMC) algorithm in Section 2.5 and the non-crossing approximation (NCA) in Section 2.6. A simplified treatment for transition metal oxides is introduced in Section 2.7, and the scheme of a self-consistent LDA+DMFT in Section 2.8. As a particular example, the LDA+DMFT calculation for $\text{La}_{1-x}\text{Sr}_x\text{TiO}_3$ is discussed in Section 3, emphasizing that the method X to solve the DMFT matters on a quantitative level. Our calculations for the Mott-Hubbard metal-insulator transition in V_2O_3 are presented in Section 4, in comparison to the experiment. Section 5 reviews our recent calculations of the Ce α - γ transition, in the perspective of the models referred to as Kondo volume collapse and Mott transition scenario. A discussion of the LDA+DMFT approach and its future prospects in Section 6 closes the presentation.

2 The LDA+DMFT Approach

2.1 Density Functional Theory

The fundamental theorem of DFT by Hohenberg and Kohn¹⁴ (see, e.g., the review by Jones and Gunnarsson²) states that the ground state energy is a functional of the electron density which assumes its minimum at the ground state electron density. Following Levy,¹⁵ this theorem is easily proved and the functional even constructed by taking the minimum (infimum) of the energy expectation value w.r.t. all (many-body) wave functions $\varphi(\mathbf{r}_1\sigma_1, \dots \mathbf{r}_N\sigma_N)$ at a given electron number N which yield the electron density $\rho(\mathbf{r})$:

$$E[\rho] = \inf \left\{ \langle \varphi | \hat{H} | \varphi \rangle \mid \langle \varphi | \sum_{i=1}^N \delta(\mathbf{r} - \mathbf{r}_i) | \varphi \rangle = \rho(\mathbf{r}) \right\}. \quad (3)$$

However, this construction is of no practical value since it actually requires the evaluation of the Hamiltonian (1). Only certain contributions like the Hartree energy $E_{\text{Hartree}}[\rho] = \frac{1}{2} \int d^3 r' d^3 r V_{\text{ee}}(\mathbf{r} - \mathbf{r}') \rho(\mathbf{r}') \rho(\mathbf{r})$ and the energy of the ionic potential $E_{\text{ion}}[\rho] = \int d^3 r V_{\text{ion}}(\mathbf{r}) \rho(\mathbf{r})$ can be expressed directly in terms of the electron density. This leads to

$$E[\rho] = E_{\text{kin}}[\rho] + E_{\text{ion}}[\rho] + E_{\text{Hartree}}[\rho] + E_{\text{xc}}[\rho], \quad (4)$$

where $E_{\text{kin}}[\rho]$ denotes the kinetic energy, and $E_{\text{xc}}[\rho]$ is the unknown exchange and correlation term which contains the energy of the electron-electron interaction except for the Hartree term. Hence all the difficulties of the many-body problem have been transferred into $E_{\text{xc}}[\rho]$. While the kinetic energy E_{kin} cannot be expressed explicitly in terms of the electron density one can employ a trick to determine it. Instead of minimizing $E[\rho]$ with respect to ρ one minimizes it w.r.t. a set of one-particle wave functions φ_i related to ρ via

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\varphi_i(\mathbf{r})|^2. \quad (5)$$

To guarantee the normalization of φ_i , the Lagrange parameters ε_i are introduced such that the variation $\delta\{E[\rho] + \varepsilon_i[1 - \int d^3 r |\varphi_i(\mathbf{r})|^2]\}/\delta\varphi_i(\mathbf{r}) = 0$ yields the Kohn-Sham¹⁶

equations:

$$\left[-\frac{\hbar^2}{2m_e} \Delta + V_{\text{ion}}(\mathbf{r}) + \int d^3 r' V_{\text{ee}}(\mathbf{r} - \mathbf{r}') \rho(\mathbf{r}') + \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})} \right] \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}). \quad (6)$$

These equations have the same form as a one-particle Schrödinger equation which, *a posteriori*, justifies to calculate the kinetic energy by means of the one-particle wave-function ansatz. The kinetic energy of a *one-particle* ansatz which has the ground state density is, then, given by $E_{\text{kin}}[\rho_{\text{min}}] = -\sum_{i=1}^N \langle \varphi_i | \hbar^2 \Delta / (2m_e) | \varphi_i \rangle$ if the φ_i are the self-consistent (spin-degenerate) solutions of Eqs. (6) and (5) with lowest “energy” ε_i . Note, however, that the one-particle potential of Eq. (6), i.e.,

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ion}}(\mathbf{r}) + \int d^3 r' V_{\text{ee}}(\mathbf{r} - \mathbf{r}') \rho(\mathbf{r}') + \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})}, \quad (7)$$

is only an auxiliary potential which artificially arises in the approach to minimize $E[\rho]$. Thus, the wave functions φ_i and the Lagrange parameters ε_i have no physical meaning at this point. Altogether, these equations allow for the DFT/LDA calculation, see the flow diagram Fig. 1.

2.2 Local Density Approximation

So far no approximations have been employed since the difficulty of the many-body problem was only transferred to the unknown functional $E_{\text{xc}}[\rho]$. For this term the local density approximation (LDA) which approximates the functional $E_{\text{xc}}[\rho]$ by a function that depends on the local density only, i.e.,

$$E_{\text{xc}}[\rho] \rightarrow \int d^3 r E_{\text{xc}}^{\text{LDA}}(\rho(\mathbf{r})), \quad (8)$$

was found to be unexpectedly successful. Here, $E_{\text{xc}}^{\text{LDA}}(\rho(\mathbf{r}))$ is usually calculated from the perturbative solution¹⁷ or the numerical simulation¹⁸ of the jellium problem which is defined by $V_{\text{ion}}(\mathbf{r}) = \text{const.}$

In principle DFT/LDA only allows one to calculate static properties like the ground state energy or its derivatives. However, one of the major applications of LDA is the calculation of band structures. To this end, the Lagrange parameters ε_i are interpreted as the physical (one-particle) energies of the system under consideration. Since the true ground-state is not a simple one-particle wave-function, this is an approximation beyond DFT. Actually, this approximation corresponds to the replacement of the Hamiltonian (1) by

$$\hat{H}_{\text{LDA}} = \sum_{\sigma} \int d^3 r \hat{\Psi}^{+}(\mathbf{r}, \sigma) \left[-\frac{\hbar^2}{2m_e} \Delta + V_{\text{ion}}(\mathbf{r}) + \int d^3 r' \rho(\mathbf{r}') V_{\text{ee}}(\mathbf{r} - \mathbf{r}') + \frac{\delta E_{\text{xc}}^{\text{LDA}}[\rho]}{\delta \rho(\mathbf{r})} \right] \hat{\Psi}(\mathbf{r}, \sigma). \quad (9)$$

For practical calculations one needs to expand the field operators w.r.t. a basis Φ_{ilm} , e.g., a linearized muffin-tin orbital (LMTO)¹⁹ basis (i denotes lattice sites; l and m are orbital

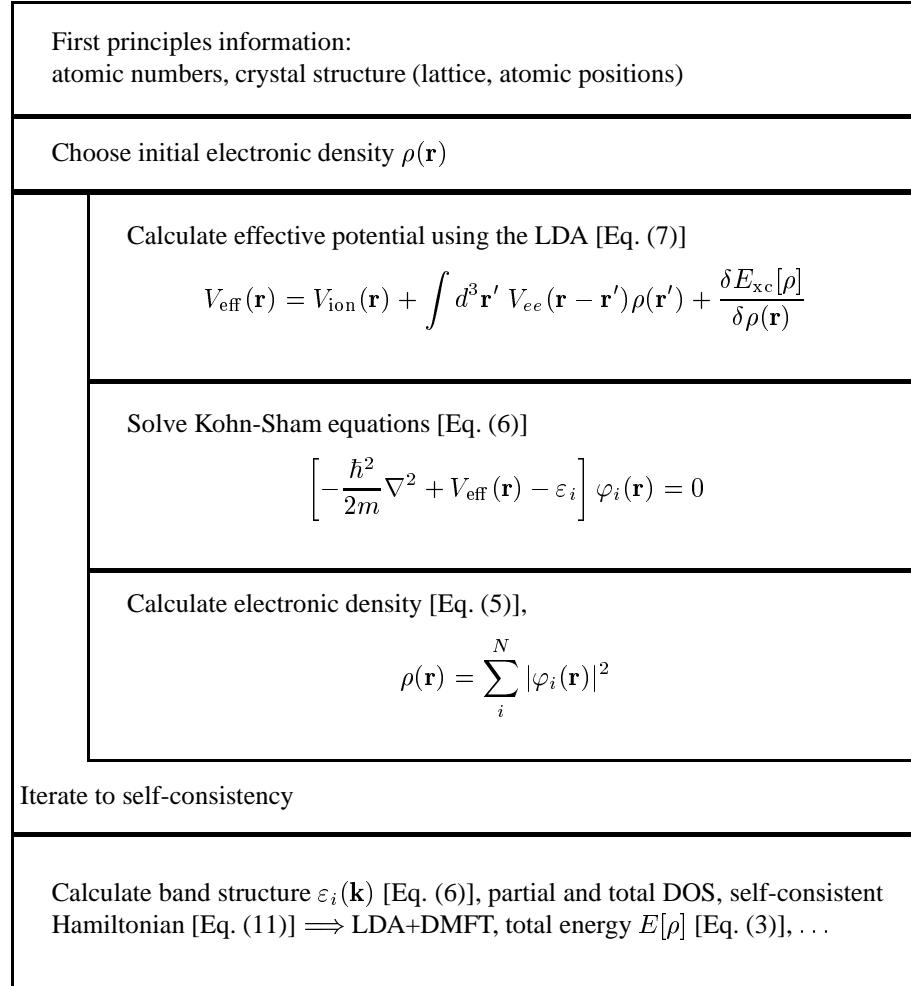


Figure 1. Flow diagram of the DFT/LDA calculations.

indices). In this basis,

$$\hat{\Psi}^+(\mathbf{r}, \sigma) = \sum_{ilm} \hat{c}_{ilm}^{\sigma\dagger} \Phi_{ilm}(\mathbf{r}) \quad (10)$$

such that the Hamiltonian (9) reads

$$\hat{H}_{\text{LDA}} = \sum_{ilm,jl'm',\sigma} (\delta_{ilm,jl'm'} \varepsilon_{ilm} \hat{n}_{ilm}^\sigma + t_{ilm,jl'm'} \hat{c}_{ilm}^{\sigma\dagger} \hat{c}_{jl'm'}^\sigma). \quad (11)$$

Here, $\hat{n}_{ilm}^\sigma = \hat{c}_{ilm}^{\sigma\dagger} \hat{c}_{ilm}^\sigma$,

$$t_{ilm,jl'm'} = \left\langle \Phi_{ilm} \left| -\frac{\hbar^2 \Delta}{2m_e} + V_{\text{ion}}(\mathbf{r}) + \int d^3\mathbf{r}' \rho(\mathbf{r}') V_{ee}(\mathbf{r} - \mathbf{r}') + \frac{\delta E_{\text{xc}}^{\text{LDA}}[\rho]}{\delta \rho(\mathbf{r})} \right| \Phi_{jl'm'} \right\rangle \quad (12)$$

for $ilm \neq jl'm'$ and zero otherwise; ε_{ilm} denotes the corresponding diagonal part.

As for static properties, the LDA approach based on the self-consistent solution of Hamiltonian (11) together with the calculation of the electronic density Eq. (5) [see the flow diagram Fig. 1] has also been highly successful for band structure calculations – but only for weakly correlated materials.² It is not reliable when applied to correlated materials and can even be completely wrong because it treats electronic *correlations* only very rudimentarily. For example, it predicts the antiferromagnetic insulator La₂CuO₄ to be a non-magnetic metal²⁰ and also completely fails to account for the high effective masses observed in 4f-based heavy fermion compounds.

2.3 Supplementing LDA with Local Coulomb Correlations

Of prime importance for correlated materials are the local Coulomb interactions between *d*- and *f*-electrons on the same lattice site since these contributions are largest. This is due to the extensive overlap (w.r.t. the Coulomb interaction) of these localized orbitals which results in strong correlations. Moreover, the largest non-local contribution is the nearest-neighbor density-density interaction which, to leading order in the number of nearest-neighbor sites, yields only the Hartree term (see Ref. 4 and, also, Ref. 21) which is already taken into account in the LDA. To take the local Coulomb interactions into account, one can supplement the LDA Hamiltonian (11) with the local Coulomb matrix approximated by the (most important) matrix elements $U_{mm'}^{\sigma\sigma'}$ (Coulomb repulsion and Z-component of Hund's rule coupling) and $J_{mm'}$ (spin-flip terms of Hund's rule coupling) between the localized electrons (for which we assume $i = i_d$ and $l = l_d$):

$$\begin{aligned} \hat{H} = & \hat{H}_{\text{LDA}} - \hat{H}_{\text{LDA}}^U + \frac{1}{2} \sum_{i=i_d, l=l_d} \sum_{m\sigma, m'\sigma'}' U_{mm'}^{\sigma\sigma'} \hat{n}_{ilm\sigma} \hat{n}_{ilm'\sigma'} \\ & - \frac{1}{2} \sum_{i=i_d, l=l_d} \sum_{m\sigma, m'}' J_{mm'} \hat{c}_{ilm\sigma}^\dagger \hat{c}_{ilm'\bar{\sigma}}^\dagger \hat{c}_{ilm'\sigma} \hat{c}_{ilm\bar{\sigma}}. \end{aligned} \quad (13)$$

Here, the prime on the sum indicates that at least two of the indices of an operator have to be different, and $\bar{\sigma} = \downarrow (\uparrow)$ for $\sigma = \uparrow (\downarrow)$. A term \hat{H}_{LDA}^U is subtracted to avoid double-counting of those contributions of the local Coulomb interaction already contained in \hat{H}_{LDA} . Since there does not exist a direct microscopic or diagrammatic link between the model Hamiltonian approach and LDA it is not possible to express \hat{H}_{LDA}^U rigorously in terms of U , J and ρ . A commonly employed approximation for \hat{H}_{LDA}^U assumes the LDA energy E_{LDA}^U of \hat{H}_{LDA}^U to be²²

$$E_{\text{LDA}}^U = \frac{1}{2} \bar{U} n_d (n_d - 1) - \frac{1}{2} J \sum_{\sigma} n_{d\sigma} (n_{d\bar{\sigma}} - 1). \quad (14)$$

Here, $n_{d\sigma} = \sum_m n_{il_d m\sigma} = \sum_m \langle \hat{n}_{il_d m\sigma} \rangle$ is the total number of interacting electrons per spin, $n_d = \sum_{\sigma} n_{d\sigma}$, \bar{U} is the average Coulomb repulsion and J the average exchange or Hund's rule coupling. In typical applications we have $U_{mm'}^{\uparrow\downarrow} \equiv U$, $J_{mm'} \equiv J$, $U_{mm'}^{\sigma\sigma'} = U - J - J\delta_{\sigma\sigma'}$ for $m \neq m'$ (here, the first term J is due to the reduced Coulomb repulsion between different orbitals and the second term $J\delta_{\sigma\sigma'}$ directly arises from the Z-component

of Hund's rule coupling), and (with the number of interacting orbitals M)

$$\bar{U} = \frac{U + (M - 1)(U - J) + (M - 1)(U - 2J)}{2M - 1}.$$

Since the one-electron LDA energies can be obtained from the derivatives of the total energy w.r.t. the occupation numbers of the corresponding states, the one-electron energy level for the *non-interacting, paramagnetic* states of (13) is obtained as²²

$$\varepsilon_{il,dm}^0 \equiv \frac{d}{dn_{il,dm}}(E_{\text{LDA}} - E_{\text{LDA}}^U) = \varepsilon_{il,dm} - \bar{U}(n_d - \frac{1}{2}) + \frac{J}{2}(n_d - 1) \quad (15)$$

where $\varepsilon_{il,dm}$ is defined in (11) and E_{LDA} is the total energy calculated from \hat{H}_{LDA} (11). Furthermore we used $n_{d\sigma} = n_d/2$ in the paramagnet.

This leads to a new Hamiltonian describing the non-interacting system

$$\hat{H}_{\text{LDA}}^0 = \sum_{ilm,jl'm',\sigma} (\delta_{ilm,jl'm'} \varepsilon_{ilm}^0 \hat{n}_{ilm}^\sigma + t_{ilm,jl'm'} \hat{c}_{ilm}^{\sigma\dagger} \hat{c}_{jl'm'}^\sigma), \quad (16)$$

where $\varepsilon_{il,dm}^0$ is given by (15) for the interacting orbitals and $\varepsilon_{ilm}^0 = \varepsilon_{ilm}$ for the non-interacting orbitals. While it is not clear at present how to systematically subtract \hat{H}_{LDA}^U one should note that the subtraction of a Hartree-type energy does not substantially affect the *overall* behavior of a strongly correlated paramagnetic metal in the vicinity of a Mott-Hubbard metal-insulator transition (see also Section 2.7).

In the following, it is convenient to work in reciprocal space where the matrix elements of \hat{H}_{LDA}^0 , i.e., the LDA one-particle energies without the local Coulomb interaction, are given by

$$(H_{\text{LDA}}^0(\mathbf{k}))_{qlm,q'l'm'} = (H_{\text{LDA}}(\mathbf{k}))_{qlm,q'l'm'} - \delta_{qlm,q'l'm'} \delta_{ql,q_d l_d} \left[\bar{U}(n_d - \frac{1}{2}) - \frac{J}{2}(n_d - 1) \right]. \quad (17)$$

Here, q is an index of the atom in the elementary unit cell, $(H_{\text{LDA}}(\mathbf{k}))_{qlm,q'l'm'}$ is the matrix element of (11) in \mathbf{k} -space, and q_d denotes the atoms with interacting orbitals in the unit cell. The non-interacting part, \hat{H}_{LDA}^0 , supplemented with the local Coulomb interaction forms the (approximated) *ab initio* Hamiltonian for a particular material under investigation:

$$\begin{aligned} \hat{H} = & \hat{H}_{\text{LDA}}^0 + \frac{1}{2} \sum_{i=i_d, l=l_d} \sum'_{m\sigma, m'\sigma'} U_{mm'}^{\sigma\sigma'} \hat{n}_{ilm\sigma} \hat{n}_{ilm'\sigma'} \\ & - \frac{1}{2} \sum_{i=i_d, l=l_d} \sum'_{m\sigma, m'} J_{mm'} \hat{c}_{ilm\sigma}^\dagger \hat{c}_{ilm'\bar{\sigma}}^\dagger \hat{c}_{ilm'\sigma} \hat{c}_{ilm\bar{\sigma}} \end{aligned} \quad (18)$$

To make use of this *ab initio* Hamiltonian it is still necessary to determine the Coulomb interaction U . To this end, one can calculate the LDA ground state energy for different numbers of interacting electrons n_d ('constrained LDA'²³) and employ Eq. (14) whose second derivative w.r.t. n_d yields U . However, one should keep in mind that, while the total LDA spectrum is rather insensitive to the choice of the basis, the calculation of U strongly depends on the shape of the orbitals which are considered to be interacting. E.g., for LaTiO₃ at a Wigner Seitz radius of 2.37 a.u. for Ti a LMTO-ASA calculation²⁴ using

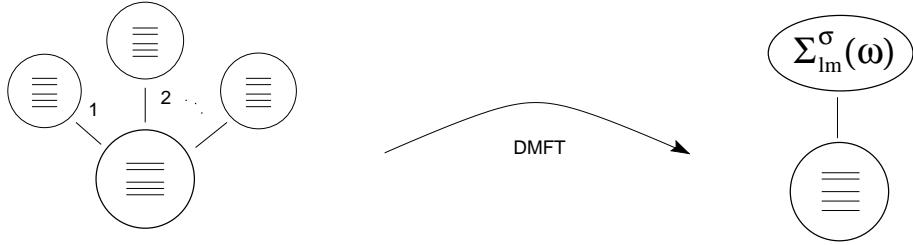


Figure 2. If the number of neighboring lattice sites goes to infinity, the central limit theorem holds and fluctuations from site-to-site can be neglected. This means that the influence of these neighboring sites can be replaced by a mean influence, the dynamical mean-field described by the self energy $\Sigma_{lm}^{\sigma}(\omega)$. This DMFT problem is equivalent to the self-consistent solution of the \mathbf{k} -integrated Dyson equation (21) and the multi-band Anderson impurity model Eq. (20).

the TB-LMTO-ASA code¹⁹ yielded $U = 4.2$ eV in comparison to the value $U = 3.2$ eV calculated by ASA-LMTO within orthogonal representation.²⁵ Thus, an appropriate basis like LMTO is mandatory and, even so, a significant uncertainty in U remains.

2.4 Dynamical Mean-Field Theory

The many-body extension of LDA, Eq. (18), was proposed by Anisimov *et al.*²² in the context of their LDA+U approach. Within LDA+U the Coulomb interactions of (18) are treated within the Hartree-Fock approximation. Hence, LDA+U does not contain true many-body physics. While this approach is successful in describing long-range ordered, insulating states of correlated electronic systems it fails to describe strongly correlated *paramagnetic* states. To go beyond LDA+U and capture the many-body nature of the electron-electron interaction, i.e., the frequency dependence of the self-energy, various approximation schemes have been proposed and applied recently.^{12,26-30} One of the most promising approaches, first implemented by Anisimov *et al.*,¹² is to solve (18) within DMFT³⁻¹¹ ("LDA+DMFT"). Of all extensions of LDA only the LDA+DMFT approach is presently able to describe the physics of *strongly* correlated, paramagnetic metals with well-developed upper and lower Hubbard bands and a narrow quasiparticle peak at the Fermi level. This characteristic three-peak structure is a signature of the importance of many-body effects.^{7,8}

During the last ten years, DMFT has proved to be a successful approach to investigate strongly correlated systems with local Coulomb interactions.¹¹ It becomes exact in the limit of high lattice coordination numbers^{3,4} and preserves the dynamics of local interactions. Hence, it represents a *dynamical* mean-field approximation. In this non-perturbative approach the lattice problem is mapped onto an effective single-site problem (see Fig. 2) which has to be determined self-consistently together with the \mathbf{k} -integrated Dyson equation connecting the self energy Σ and the Green function G at frequency ω :

$$G_{qlm,q'l'm'}(\omega) = \frac{1}{V_B} \int d^3k \left([\omega 1 + \mu 1 - H_{\text{LDA}}^0(\mathbf{k}) - \Sigma(\omega)]^{-1} \right)_{qlm,q'l'm'}. \quad (19)$$

Here, 1 is the unit matrix, μ the chemical potential, the matrix $H_{\text{LDA}}^0(\mathbf{k})$ is defined in (17), $\Sigma(\omega)$ denotes the self-energy matrix which is non-zero only between the interacting

orbitals, $[...]^{-1}$ implies the inversion of the matrix with elements $n (=qlm)$, $n' (=q'l'm')$, and the integration extends over the Brillouin zone with volume V_B .

The DMFT single-site problem depends on $\mathcal{G}(\omega)^{-1} = G(\omega)^{-1} + \Sigma(\omega)$ and is equivalent^{7,8} to an Anderson impurity model (the history and the physics of this model is summarized by Anderson in Ref. 31) if its hybridization $\Delta(\omega)$ satisfies $\mathcal{G}^{-1}(\omega) = \omega - \int d\omega' \Delta(\omega') / (\omega - \omega')$. The local one-particle Green function at a Matsubara frequency $i\omega_\nu = i(2\nu + 1)\pi/\beta$ (β : inverse temperature), orbital index m ($l = l_d$, $q = q_d$), and spin σ is given by the following functional integral over Grassmann variables ψ and ψ^* :

$$G_{\nu m}^\sigma = -\frac{1}{Z} \int \mathcal{D}[\psi] \mathcal{D}[\psi^*] \psi_{\nu m}^\sigma \psi_{\nu m}^{\sigma*} e^{\mathcal{A}[\psi, \psi^*, \mathcal{G}^{-1}]}. \quad (20)$$

Here, $Z = \int \mathcal{D}[\psi] \mathcal{D}[\psi^*] \exp(\mathcal{A}[\psi, \psi^*, \mathcal{G}^{-1}])$

is the partition function and the single-site action \mathcal{A} has the form (the interaction part of \mathcal{A} is in terms of the “imaginary time” τ , i.e., the Fourier transform of ω_ν)

$$\begin{aligned} \mathcal{A}[\psi, \psi^*, \mathcal{G}^{-1}] = & \sum_{\nu, \sigma, m} \psi_{\nu m}^{\sigma*} (\mathcal{G}_{\nu m}^\sigma)^{-1} \psi_{\nu m}^\sigma \\ & - \frac{1}{2} \sum'_{m\sigma, m\sigma'} U_{mm'}^{\sigma\sigma'} \int_0^\beta d\tau \psi_m^{\sigma*}(\tau) \psi_m^\sigma(\tau) \psi_{m'}^{\sigma'*}(\tau) \psi_{m'}^{\sigma'}(\tau) \\ & + \frac{1}{2} \sum'_{m\sigma, m} J_{mm'} \int_0^\beta d\tau \psi_m^{\sigma*}(\tau) \psi_m^{\bar{\sigma}}(\tau) \psi_{m'}^{\bar{\sigma}*}(\tau) \psi_{m'}^\sigma(\tau). \end{aligned} \quad (21)$$

This single-site problem (20) has to be solved self-consistently together with the \mathbf{k} -integrated Dyson equation (19) to obtain the DMFT solution of a given problem, see the flow diagram Fig. 3.

Due to the equivalence of the DMFT single-site problem and the Anderson impurity problem a variety of approximative techniques have been employed to solve the DMFT equations, such as the iterated perturbation theory (IPT)^{7,11} and the non-crossing approximation (NCA),^{32–34} as well as numerical techniques like quantum Monte Carlo simulations (QMC),³⁵ exact diagonalization (ED),^{36,11} or numerical renormalization group (NRG).³⁷ QMC and NCA will be discussed in more detail in Section 2.5 and 2.6, respectively. IPT is non-self-consistent second-order perturbation theory in U for the Anderson impurity problem (20) at half-filling. It represents an ansatz that also yields the correct perturbational U^2 -term and the correct atomic limit for the self-energy off half-filling,³⁸ for further details see Refs. 38, 12, 26. ED directly diagonalizes the Anderson impurity problem at a limited number of lattice sites and orbitals. NRG first replaces the conduction band by a discrete set of states at $D\Lambda^{-n}$ (D : bandwidth; $n = 0, \dots, \mathcal{N}_s$) and then diagonalizes this problem iteratively with increasing accuracy at low energies, i.e., with increasing \mathcal{N}_s . In principle, QMC and ED are exact methods, but they require an extrapolation, i.e., the discretization of the imaginary time $\Delta\tau \rightarrow 0$ (QMC) or the number of lattice sites of the respective impurity model $N_s \rightarrow \infty$ (ED), respectively.

In the context of LDA+DMFT we refer to the computational schemes to solve the DMFT equations discussed above as LDA+DMFT(X) where X=IPT,¹² NCA,³⁰ QMC²⁴

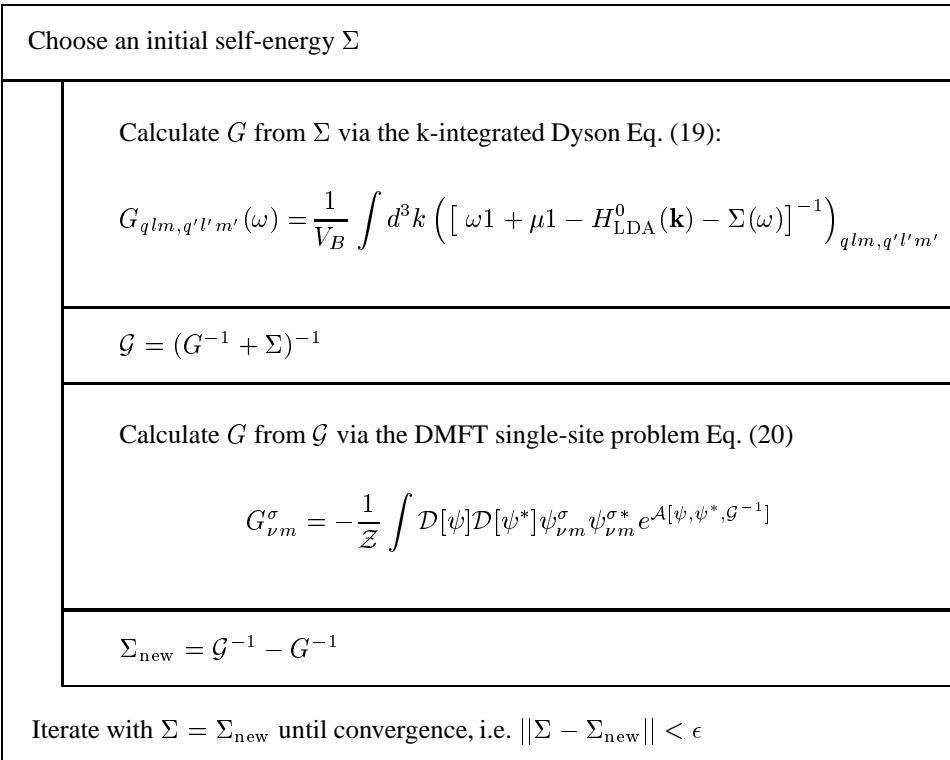


Figure 3. Flow diagram of the DMFT self-consistency cycle.

have been investigated in the case of $\text{La}_{1-x}\text{Sr}_x\text{TiO}_3$. The same strategy was formulated by Lichtenstein and Katsnelson²⁶ as one of their LDA++ approaches. Lichtenstein and Katsnelson applied LDA+DMFT(IPT),⁴² and were the first to use LDA+DMFT(QMC),⁴³ to investigate the spectral properties of iron. Recently, also V_2O_3 ,⁴⁴ $\text{Ca}_{2-x}\text{Sr}_x\text{RuO}_4$,^{45,46} Ni,⁴⁷ Fe,⁴⁷ Pu,^{48,49} and Ce^{50,51} have been studied by LDA+DMFT. Realistic investigations of itinerant ferromagnets (e.g., Ni) have also recently become possible by combining density functional theory with multi-band Gutzwiller wave functions.⁵²

2.5 QMC Method to Solve DMFT

The self-consistency cycle of the DMFT (Fig. 3) requires a method to solve for the dynamics of the single-site problem of DMFT, i.e., Eq. (20). The QMC algorithm by Hirsch and Fye³⁵ is a well established method to find a numerically exact solution for the Anderson impurity model and allows one to calculate the impurity Green function G at a given \mathcal{G}^{-1} as well as correlation functions. In essence, the QMC technique maps the interacting electron problem Eq. (20) onto a sum of non-interacting problems where the single particle moves in a fluctuating, time-dependent field and evaluates this sum by Monte Carlo sampling, see the flow diagram Fig. 4 for an overview. To this end, the imaginary time interval

$[0, \beta]$ of the functional integral Eq. (20) is discretized into Λ steps of size $\Delta\tau = \beta/\Lambda$, yielding support points $\tau_l = l\Delta\tau$ with $l = 1 \dots \Lambda$. Using this Trotter discretization, the integral $\int_0^\beta d\tau$ is transformed to the sum $\sum_{l=1}^\Lambda \Delta\tau$ and the exponential terms in Eq. (20) can be separated via the Trotter-Suzuki formula for operators \hat{A} and \hat{B} ⁵³

$$e^{-\beta(\hat{A}+\hat{B})} = \prod_{l=1}^\Lambda e^{-\Delta\tau\hat{A}}e^{-\Delta\tau\hat{B}} + \mathcal{O}(\Delta\tau), \quad (22)$$

which is exact in the limit $\Delta\tau \rightarrow 0$. The single site action \mathcal{A} of Eq. (21) can now be written in the discrete, imaginary time as

$$\begin{aligned} \mathcal{A}[\psi, \psi^*, \mathcal{G}^{-1}] &= \Delta\tau^2 \sum_{\sigma m l, l'=0}^{\Lambda-1} \psi_{ml}^* \mathcal{G}_m^{\sigma -1} (l\Delta\tau - l'\Delta\tau) \psi_{ml'}^* \\ &\quad - \frac{1}{2} \Delta\tau \sum'_{m\sigma, m'\sigma'} U_{mm'}^{\sigma\sigma'} \sum_{l=0}^{\Lambda-1} \psi_{ml}^* \psi_{ml}^* \psi_{m'l}^* \psi_{m'l}^*, \end{aligned} \quad (23)$$

where the first term was Fourier-transformed from Matsubara frequencies to imaginary time. In a second step, the $M(2M-1)$ interaction terms in the single site action \mathcal{A} are decoupled by introducing a classical auxiliary field $s_{lmm'}^{\sigma\sigma'}$:

$$\begin{aligned} &\exp \left\{ \frac{\Delta\tau}{2} U_{mm'}^{\sigma\sigma'} (\psi_{ml}^* \psi_{ml}^* - \psi_{m'l}^* \psi_{m'l}^*)^2 \right\} = \\ &\frac{1}{2} \sum_{s_{lmm'}^{\sigma\sigma'} = \pm 1} \exp \left\{ \Delta\tau \lambda_{lmm'}^{\sigma\sigma'} s_{lmm'}^{\sigma\sigma'} (\psi_{ml}^* \psi_{ml}^* - \psi_{m'l}^* \psi_{m'l}^*) \right\}, \end{aligned} \quad (24)$$

where $\cosh(\lambda_{lmm'}^{\sigma\sigma'}) = \exp(\Delta\tau U_{mm'}^{\sigma\sigma'} / 2)$ and M is the number of interacting orbitals. This so-called discrete Hirsch-Fye-Hubbard-Stratonovich transformation can be applied to the Coulomb repulsion as well as the Z-component of Hund's rule coupling.⁵⁴ It replaces the interacting system by a sum of $\Lambda M(2M-1)$ auxiliary fields $s_{lmm'}^{\sigma\sigma'}$. The functional integral can now be solved by a simple Gauss integration because the Fermion operators only enter quadratically, i.e., for a given configuration $\mathbf{s} = \{s_{lmm'}^{\sigma\sigma'}\}$ of the auxiliary fields the system is non-interacting. The quantum mechanical problem is then reduced to a matrix problem

$$G_{\tilde{m}l_1l_2}^{\sigma} = \frac{1}{Z} \frac{1}{2} \sum_l \sum_{m'\sigma', m''\sigma''}^l \sum_{s_{l'm'm'}^{\sigma''\sigma'} = \pm 1} [(M_{\tilde{m}}^{\sigma\mathbf{s}})^{-1}]_{l_1l_2} \prod_{m\sigma} \det \mathbf{M}_m^{\sigma\mathbf{s}} \quad (25)$$

with the partition function Z , the matrix

$$\mathbf{M}_{\tilde{m}}^{\sigma\mathbf{s}} = \Delta\tau^2 [\mathbf{G}_m^{\sigma -1} + \Sigma_m^{\sigma}] e^{-\tilde{\lambda}_m^{\sigma\mathbf{s}}} + \mathbf{1} - e^{-\tilde{\lambda}_m^{\sigma\mathbf{s}}} \quad (26)$$

and the elements of the matrix $\tilde{\lambda}_m^{\sigma\mathbf{s}}$

$$\tilde{\lambda}_{mll'}^{\sigma\mathbf{s}} = -\delta_{ll'} \sum_{m'\sigma'} \lambda_{mm'}^{\sigma\sigma'} \tilde{\sigma}_{mm'}^{\sigma\sigma'} s_{lmm'}^{\sigma\sigma'}. \quad (27)$$

Here $\tilde{\sigma}_{mm'}^{\sigma\sigma'} = 2\Theta(\sigma' - \sigma + \delta_{\sigma\sigma'}[m' - m] - 1)$ changes sign if $(m\sigma)$ and $(m'\sigma')$ are exchanged. For more details, e.g., for a derivation of Eq. (26) for the matrix \mathbf{M} , see Refs. 35, 11.

Since the sum in Eq. (25) consists of $2^{\Lambda M(2M-1)}$ addends, a complete summation for large Λ is computationally impossible. Therefore the Monte Carlo method, which is often an efficient way to calculate high-dimensional sums and integrals, is employed for importance sampling of Eq. (25). In this method, the integrand $F(x)$ is split up into a normalized probability distribution P and the remaining term O :

$$\int dx F(x) = \int dx O(x) P(x) \equiv \langle O \rangle_P \quad (28)$$

with

$$\int dx P(x) = 1 \quad \text{and} \quad P(x) \geq 0. \quad (29)$$

In statistical physics, the Boltzmann distribution is often a good choice for the function P :

$$P(x) = \frac{1}{Z} \exp(-\beta E(x)). \quad (30)$$

For the sum of Eq. (25), this probability distribution translates to

$$P(\mathbf{s}) = \frac{1}{Z} \prod_{m\sigma} \det \mathbf{M}_m^{\sigma\mathbf{s}} \quad (31)$$

with the remaining term

$$O(\mathbf{s})_{\vec{m}l_1l_2}^{\sigma} = [(M_{\vec{m}}^{\sigma\mathbf{s}})^{-1}]_{l_1l_2}. \quad (32)$$

Instead of summing over all possible configurations, the Monte Carlo simulation generates configurations x_i with respect to the probability distribution $P(x)$ and averages the observable $O(x)$ over these x_i . Therefore the relevant parts of the phase space with a large Boltzmann weight are taken into account to a greater extent than the ones with a small weight, coining the name importance sampling for this method. With the central limit theorem one gets for \mathcal{N} statistically independent addends the estimate

$$\langle O \rangle_P = \frac{1}{\mathcal{N}} \sum_{\substack{i=1 \\ x_i \in P(x)}}^{\mathcal{N}} O(x_i) \pm \frac{1}{\sqrt{\mathcal{N}}} \sqrt{\langle O^2 \rangle_P - \langle O \rangle_P^2}. \quad (33)$$

Here, the error and with it the number of needed addends \mathcal{N} is nearly independent of the dimension of the integral. The computational effort for the Monte Carlo method is therefore only rising polynomially with the dimension of the integral and not exponentially as in a normal integration. Using a Markov process and single spin-flips in the auxiliary fields, the computational cost of the algorithm in leading order of Λ is

$$2aM(2M-1)\Lambda^3 \times \text{number of MC-sweeps}, \quad (34)$$

where a is the acceptance rate for a single spin-flip.

The advantage of the QMC method (for the algorithm see the flow diagram Fig. 4) is that it is (numerically) exact. It allows one to calculate the one-particle Green function as well as two-particle (or higher) Green functions. On present workstations the QMC approach is able to deal with up to seven *interacting* orbitals and temperatures above about room temperature. Very low temperatures are not accessible because the numerical effort grows like $\Lambda^3 \propto 1/T^3$. Since the QMC approach calculates $G(\tau)$ or $G(i\omega_n)$ with a statistical error, it also requires the maximum entropy method⁵⁶ to obtain the Green function $G(\omega)$ at real (physical) frequencies ω .

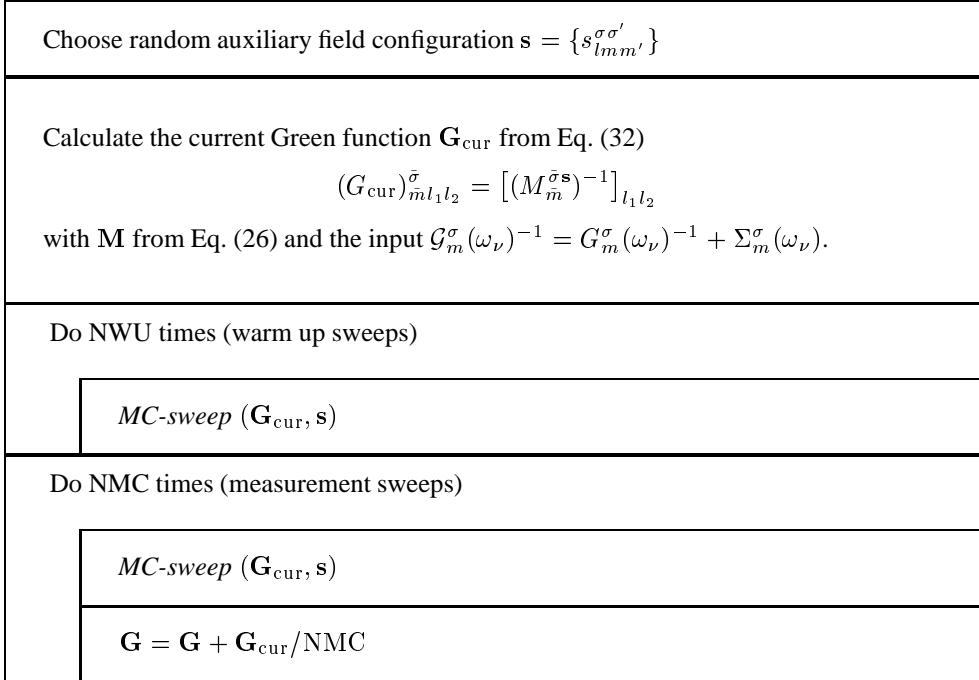


Figure 4. Flow diagram of the QMC algorithm to calculate the Green function matrix \mathbf{G} using the procedure *MC-sweep* of Fig. 5.

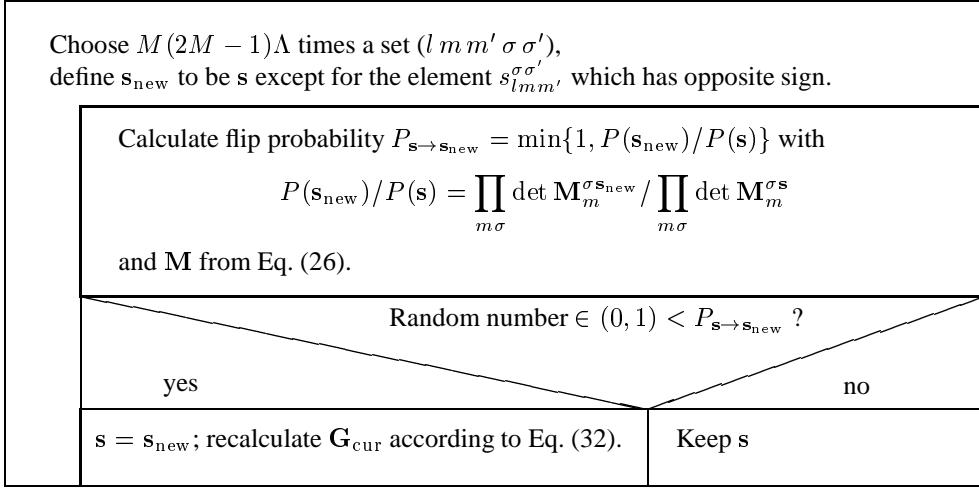


Figure 5. Procedure *MC-sweep* using the Metropolis⁵⁵ rule to change the sign of $s_{lmm'}^{\sigma\sigma'}$. The recalculation of \mathbf{G}_{cur} , i.e., the matrix \mathbf{M} of Eq. (26), simplifies to $\mathcal{O}(\Lambda^2)$ operations if only one $s_{lmm'}^{\sigma\sigma'}$ changes sign.^{35,11}

2.6 NCA Method to Solve DMFT

The NCA approach is a resolvent perturbation theory in the hybridization parameter $\Delta(\omega)$ of the effective Anderson impurity problem.³² Thus, it is reliable if the Coulomb interaction U is large compared to the band-width and also offers a computationally inexpensive approach to check the general spectral features in other situations.

To see how the NCA can be adapted for the DMFT, let us rewrite Eq. (19) as

$$G_\sigma(z) = \frac{1}{N_k} \sum_{\mathbf{k}} [z - H_{LDA}^0(\mathbf{k}) - \Sigma(z)]^{-1} \quad (35)$$

where $z = \omega + i0^+ + \mu$. Again, $H_{LDA}^0(\mathbf{k})$, $\Sigma(z)$ and hence $G_\sigma^0(\zeta)$ and $G_\sigma(z)$ are matrices in orbital space. Note that $\Sigma(z)$ has nonzero entries for the correlated orbitals only.

On quite general grounds, Eq. (35) can be cast into the form

$$G_\sigma(z) = \frac{1}{z - E^0 - \Sigma_\sigma(z) - \Delta_\sigma(z)} \quad (36)$$

where

$$E^0 = \frac{1}{N_k} \sum_{\mathbf{k}} H_{LDA}^0(\mathbf{k}) \quad (37)$$

with the number of \mathbf{k} points N_k and

$$\lim_{\omega \rightarrow \pm\infty} \Re e\{\Delta_\sigma(\omega + i\delta)\} = 0 \quad . \quad (38)$$

Given the the matrix E^0 , the Coulomb matrix U and the hybridization matrix $\Delta_\sigma(z)$, we are now in a position to set up a resolvent perturbation theory with respect to $\Delta_\sigma(z)$. To this end, we first have to diagonalize the local Hamiltonian

$$\begin{aligned} H_{\text{local}} &= \sum_{\sigma} \sum_{qml} \sum_{q'm'l'} c_{qlm\sigma}^\dagger E_{qlm,q'l'm'}^0 c_{qlm\sigma} \\ &\quad + \frac{1}{2} \sum_{m\sigma} \sum_{m'\sigma'} U_{mm'}^{\sigma\sigma'} n_{q_d l_d m\sigma} n_{q_d l_d m'\sigma'} \\ &\quad - \frac{1}{2} \sum_{m\sigma} \sum_{m'} J_{mm'} c_{q_d l_d m\sigma}^\dagger c_{q_d l_d m'\bar{\sigma}} c_{q_d l_d m'\sigma} c_{q_d l_d m\bar{\sigma}} \\ &= \sum_{\alpha} E_{\alpha} |\alpha\rangle\langle\alpha| \end{aligned} \quad (39)$$

with local eigenstates $|\alpha\rangle$ and energies E_{α} . In contrast to the QMC, this approach allows one to take into account the full Coulomb matrix plus spin-orbit coupling.

With the states $|\alpha\rangle$ defined above, the fermionic operators with quantum numbers $\kappa = (q, l, m)$ are expressed as

$$\begin{aligned} c_{\kappa\sigma}^\dagger &= \sum_{\alpha,\beta} (D_{\beta\alpha}^{\kappa\sigma})^* |\alpha\rangle\langle\beta| , \\ c_{\kappa\sigma} &= \sum_{\alpha,\beta} D_{\alpha\beta}^{\kappa\sigma} |\alpha\rangle\langle\beta| . \end{aligned} \quad (40)$$

The key quantity for the resolvent perturbation theory is the resolvent $R(z)$, which obeys a Dyson equation³²

$$R(z) = R^0(z) + R^0(z)S(z)R(z) , \quad (41)$$

where $R_{\alpha\beta}^0(z) = 1/(z - E_\alpha)\delta_{\alpha\beta}$ and $S_{\alpha\beta}(z)$ denotes the self-energy for the local states due to the coupling to the environment through $\Delta(z)$.

The self-energy $S_{\alpha\beta}(z)$ can be expressed as power series in the hybridization $\Delta(z)$.³² Retaining only the lowest-, i.e. 2nd-order terms leads to a set of self-consistent integral equations

$$\begin{aligned} S_{\alpha\beta}(z) = & \sum_{\sigma} \sum_{\kappa\kappa'} \sum_{\alpha'\beta'} \int \frac{d\varepsilon}{\pi} f(\varepsilon) (D_{\alpha'\alpha}^{\kappa\sigma})^* \Gamma_{\sigma}^{\kappa\kappa'}(\varepsilon) R_{\alpha'\beta'}(z + \varepsilon) D_{\beta'\beta}^{\kappa'\sigma} \\ & + \sum_{\sigma} \sum_{\kappa\kappa'} \sum_{\alpha'\beta'} \int \frac{d\varepsilon}{\pi} (1 - f(\varepsilon)) D_{\alpha'\alpha}^{\kappa\sigma} \Gamma_{\sigma}^{\kappa\kappa'}(\varepsilon) R_{\alpha'\beta'}(z - \varepsilon) (D_{\beta'\beta}^{\kappa'\sigma})^* \end{aligned} \quad (42)$$

to determine $S_{\alpha\beta}(z)$, where $f(\varepsilon)$ denotes Fermi's function and $\Gamma(\varepsilon) = -\Im m \{ \Delta(\varepsilon + i0^+) \}$. The set of equations (42) are in the literature referred to as non-crossing approximation (NCA), because, when viewed in terms of diagrams, they contain no crossing of band-electron lines. In order to close the cycle for the DMFT, we still have to calculate the true local Green function $G_\sigma(z)$. This, however, can be done within the same approximation with the result

$$G_\sigma^{\kappa\kappa'}(i\omega) = \frac{1}{Z_{\text{local}}} \sum_{\alpha,\alpha'} \sum_{\nu,\nu'} D_{\alpha\alpha'}^{\kappa\sigma} (D_{\nu\nu'}^{\kappa'\sigma})^* \oint \frac{dz e^{-\beta z}}{2\pi i} R_{\alpha\nu}(z) R_{\alpha'\nu'}(z + i\omega) . \quad (43)$$

Here, $Z_{\text{local}} = \sum_{\alpha} \oint \frac{dz e^{-\beta z}}{2\pi i} R_{\alpha\alpha}(z)$ denotes the local partition function and β is the inverse temperature.

Like any other technique, the NCA has its merits and disadvantages. As a self-consistent resummation of diagrams it constitutes a conserving approximation to the Anderson impurity model. Furthermore, it is a (computationally) fast method to obtain dynamical results for this model and thus also within DMFT. However, the NCA is known to violate Fermi liquid properties at temperatures much lower than the smallest energy scale of the problem and whenever charge excitations become dominant.^{57,34} Hence, in some parameter ranges it fails in the most dramatic way and must therefore be applied with considerable care.³⁴

2.7 Simplifications for Transition Metal Oxides with Well Separated e_g - and t_{2g} -Bands

Many transition metal oxides are cubic perovskites, with only a slight distortion of the cubic crystal structure. In these systems the transition metal d -orbitals lead to strong Coulomb interactions between the electrons. The cubic crystal-field of the oxygen causes the d -orbitals to split into three degenerate t_{2g} - and two degenerate e_g -orbitals. This splitting is often so strong that the t_{2g} - or e_g -bands at the Fermi energy are rather well separated from all other bands. In this situation the low-energy physics is well described by taking only the degenerate bands at the Fermi energy into account. Without symmetry

breaking, the Green function and the self-energy of these bands remain degenerate, i.e., $G_{qlm,q'l'm'}(z) = G(z)\delta_{qlm,q'l'm'}$ and $\Sigma_{qlm,q'l'm'}(z) = \Sigma(z)\delta_{qlm,q'l'm'}$ for $l = l_d$ and $q = q_d$ (where l_d and q_d denote the electrons in the interacting band at the Fermi energy). Downfolding to a basis with these degenerate q_d - l_d -bands results in an effective Hamiltonian $H_{\text{LDA}}^{0\,\text{eff}}$ (where indices $l = l_d$ and $q = q_d$ are suppressed)

$$G_{mm'}(\omega) = \frac{1}{V_B} \int d^3k \left([\omega 1 + \mu 1 - H_{\text{LDA}}^{0\,\text{eff}}(\mathbf{k}) - \Sigma(\omega)]^{-1} \right)_{mm'}. \quad (44)$$

Due to the diagonal structure of the self-energy the degenerate interacting Green function can be expressed via the non-interacting Green function $G^0(\omega)$:

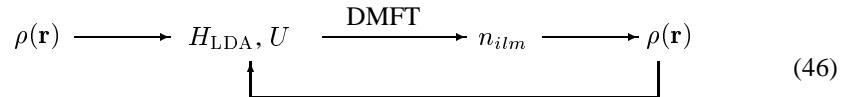
$$G(\omega) = G^0(\omega - \Sigma(\omega)) = \int d\epsilon \frac{N^0(\epsilon)}{\omega - \Sigma(\omega) - \epsilon}. \quad (45)$$

Thus, it is possible to use the Hilbert transformation of the unperturbed LDA-calculated density of states (DOS) $N^0(\epsilon)$, i.e., Eq. (45), instead of Eq. (19). This simplifies the calculations considerably. With Eq. (45) also some conceptual simplifications arise: (i) the subtraction of \hat{H}_{LDA}^U in (45) only results in an (unimportant) shift of the chemical potential and, thus, the exact form of \hat{H}_{LDA}^U is irrelevant; (ii) Luttinger's theorem of Fermi pinning holds, i.e., the interacting DOS at the Fermi energy is fixed at the value of the non-interacting DOS at $T = 0$ within a Fermi liquid; (iii) as the number of electrons within the different bands is fixed, the LDA+DMFT approach is automatically self-consistent.

In this context it should be noted that the approximation Eq. (45) is justified only if the overlap between the t_{2g} orbitals and the other orbitals is rather weak.

2.8 Extensions of the LDA+DMFT Scheme

In the present form of the LDA+DMFT scheme the band-structure input due to LDA and the inclusion of the electronic correlations by DMFT are performed as successive steps without subsequent feedback. In general, the DMFT solution will result in a change of the occupation of the different bands involved. This changes the electron density $\rho(\mathbf{r})$ and, thus, results in a new LDA-Hamiltonian \hat{H}_{LDA} (11) since \hat{H}_{LDA} depends on $\rho(\mathbf{r})$. At the same time also the Coulomb interaction U changes and needs to be determined by a new constrained LDA calculation. In a *self-consistent* LDA+DMFT scheme, H_{LDA} and U would define a new Hamiltonian (18) which again needs to be solved within DMFT, etc., until convergence is reached:



Without Coulomb interaction ($U = 0$) this scheme reduces to the self-consistent solution of the Kohn-Sham equations. A self-consistency scheme similar to Eq. (46) was employed by Savrasov and Kotliar⁴⁹ in their calculation of Pu. An *ab initio* DMFT scheme formulated directly in the continuum was recently proposed by Chitra and Kotliar.⁵⁸

3 Comparison of Different Methods to Solve DMFT: The Model System $\text{La}_{1-x}\text{Sr}_x\text{TiO}_3$

The stoichiometric compound LaTiO_3 is a cubic perovskite with a small orthorhombic distortion ($\angle \text{Ti} - \text{O} - \text{Ti} \approx 155^\circ$)⁵⁹ and is an antiferromagnetic insulator⁶⁰ below $T_N = 125$ K.⁶¹ Above T_N , or at low Sr-doping x , and neglecting the small orthorhombic distortion (i.e., considering a cubic structure with the same volume), LaTiO_3 is a strongly correlated, but otherwise simple paramagnet with only *one* 3d-electron on the trivalent Ti sites. This makes the system a perfect trial candidate for the LDA+DMFT approach.

The LDA band-structure calculation for undoped (cubic) LaTiO_3 yields the DOS shown in Fig. 6 which is typical for early transition metals. The oxygen bands, ranging from -8.2 eV to -4.0 eV, are filled such that Ti is three-valent. Due to the crystal-field splitting, the Ti 3d-bands separates into two empty e_g -bands and three degenerate t_{2g} -bands. Since the t_{2g} -bands at the Fermi energy are well separated also from the other bands we employ the approximation introduced in section 2.5 which allows us to work with the LDA DOS [Eq. (45)] instead of the full one-particle Hamiltonian H_{LDA}^0 of [Eq. (19)]. In the LDA+DMFT calculation, Sr-doping x is taken into account by adjusting the chemical potential to yield $n = 1 - x = 0.94$ electrons within the t_{2g} -bands, neglecting effects disorder and the x -dependence of the LDA DOS (note, that Sr and Ti have a very similar band structure within LDA). There is some uncertainty in the LDA-calculated Coulomb interaction parameter $U \sim 4 - 5$ eV (for a discussion see Ref. 24) which is here assumed to be spin- and orbital-independent. In Fig. 7, results for the spectrum of $\text{La}_{0.94}\text{Sr}_{0.06}\text{TiO}_3$

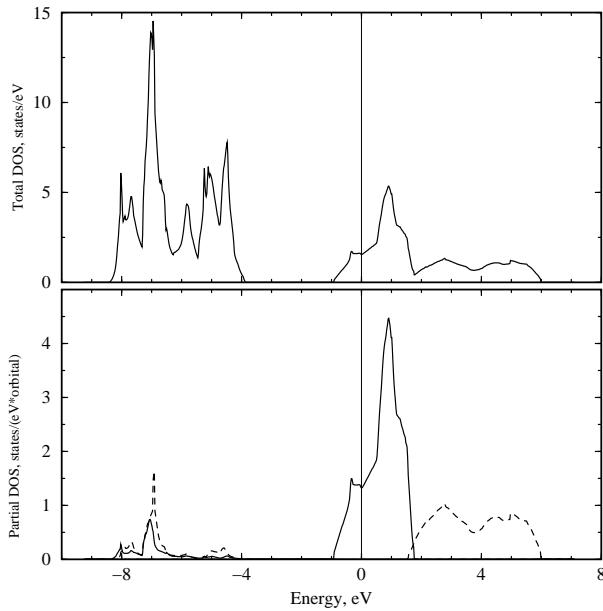


Figure 6. Densities of states of LaTiO_3 calculated with LDA-LMTO. Upper figure: total DOS; lower figure: partial t_{2g} (solid lines) and e_g (dashed lines) DOS [reproduced from Ref.24].

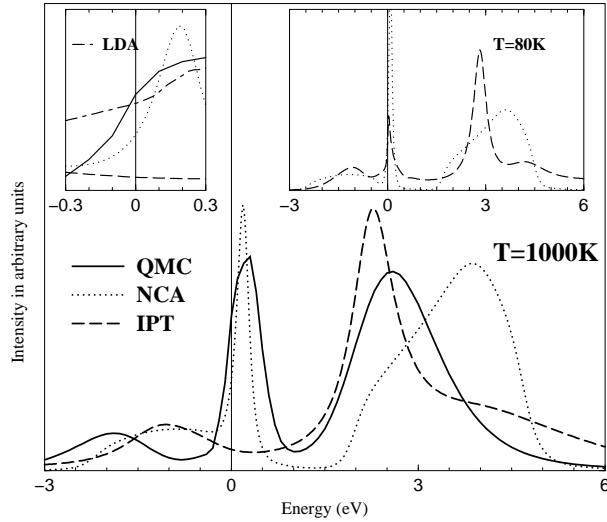


Figure 7. Spectrum of $\text{La}_{0.94}\text{Sr}_{0.06}\text{TiO}_3$ as calculated by LDA+DMFT(X) at $T = 0.1$ eV (≈ 1000 K) and $U = 4$ eV employing the approximations X=IPT, NCA, and numerically exact QMC. Inset left: Behavior at the Fermi level including the LDA DOS. Inset right: X=IPT and NCA spectra at $T = 80$ K [reproduced from Ref.24].

as calculated by LDA+DMFT(IPT, NCA, QMC) for the same LDA DOS at $T \approx 1000$ K and $U = 4$ eV are compared.²⁴ In Ref. 24 the formerly presented IPT¹² and NCA³⁰ spectra were recalculated to allow for a comparison at exactly the same parameters. All three methods yield the typical features of strongly correlated metallic paramagnets: a lower Hubbard band, a quasi-particle peak (note that IPT produces a quasi-particle peak only below about 250K which is therefore not seen here), and an upper Hubbard band. By contrast, within LDA the correlation-induced Hubbard bands are missing and only a broad central quasi-particle band (actually a one-particle peak) is obtained (Fig. 6).

While the results of the three evaluation techniques of the DMFT equations (the approximations IPT, NCA and the numerically exact method QMC) agree on a qualitative level, Fig. 7 reveals considerable quantitative differences. In particular, the IPT quasi-particle peak found at low temperatures (see right inset of Fig. 7) is too narrow such that it disappears already at about 250 K and is, thus, not present at $T \approx 1000$ K. A similarly narrow IPT quasi-particle peak was found in a three-band model study with Bethe-DOS by Kajueter and Kotliar.³⁸ Besides underestimating the Kondo temperature, IPT also produces notable deviations in the shape of the upper Hubbard band. Although NCA comes off much better than IPT it still underestimates the width of the quasiparticle peak by a factor of two. Furthermore, the position of the quasi-particle peak is too close to the lower Hubbard band. In the left inset of Fig. 7, the spectra at the Fermi level are shown. At the Fermi level, where at sufficiently low temperatures the interacting DOS should be pinned at the non-interacting value, the NCA yields a spectral function which is almost by a factor of two too small. The shortcomings of the NCA-results, with a too small low-energy scale and too much broadened Hubbard bands for multi-band systems, are well understood

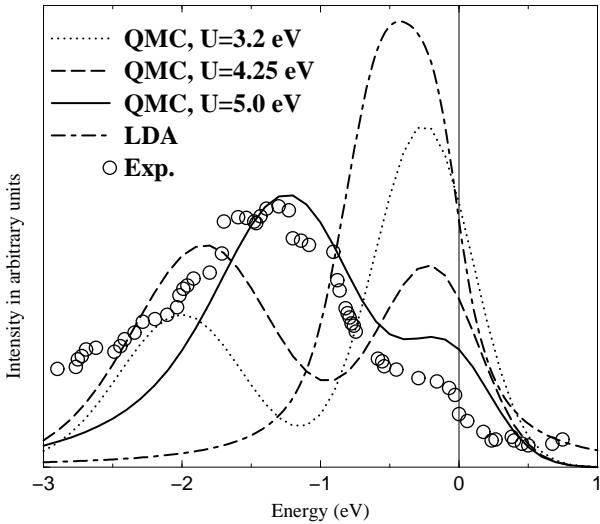


Figure 8. Comparison of the experimental photoemission spectrum,⁶⁴ the LDA result, and the LDA+DMFT(QMC) calculation for $\text{La}_{0.94}\text{Sr}_{0.06}\text{TiO}_3$ (i.e., 6% hole doping) and different Coulomb interaction $U = 3.2, 4.25$, and 5 eV [reproduced from Ref.24].

and related to the neglect of exchange type diagrams.⁶³ Similarly, the deficiencies of the IPT-results are not entirely surprising in view of the semi-phenomenological nature of this approximation, especially for a system off half filling.

This comparison shows that the choice of the *method* used to solve the DMFT equations is indeed *important*, and that, at least for the present system, the approximations IPT and NCA differ quantitatively from the numerically exact QMC. Nevertheless, the NCA gives a rather good account of the qualitative spectral features and, because it is fast and can often be applied to comparatively low temperatures, can yield an overview of the physics to be expected.

Photoemission spectra provide a direct experimental tool to study the electronic structure and spectral properties of electronically correlated materials. A comparison of LDA+DMFT(QMC) at 1000 K⁶⁵ with the experimental photoemission spectrum⁶⁴ of $\text{La}_{0.94}\text{Sr}_{0.06}\text{TiO}_3$ is presented in Fig 8. To take into account the uncertainty in U ,²⁴ we present results for $U = 3.2, 4.25$ and 5 eV . All spectra are multiplied with the Fermi step function and are Gauss-broadened with a broadening parameter of 0.3 eV to simulate the experimental resolution.⁶⁴ LDA band structure calculations, the results of which are also presented in Fig. 8, clearly fail to reproduce the broad band observed in the experiment at 1-2 eV below the Fermi energy.⁶⁴ Taking the correlations between the electrons into account, this lower band is easily identified as the lower Hubbard band whose spectral weight originates from the quasi-particle band at the Fermi energy and which increases with U . The best agreement with experiment concerning the relative intensities of the Hubbard band and the quasi-particle peak and, also, the position of the Hubbard band is found for $U = 5 \text{ eV}$. The value $U = 5 \text{ eV}$ is still compatible with the *ab initio* calculation of this parameter within LDA.²⁴ One should also bear in mind that photoemission experiments are sensitive

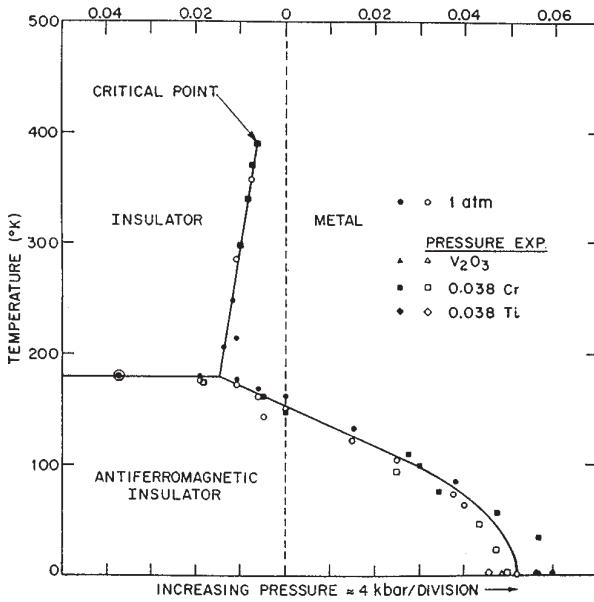


Figure 9. Experimental phase diagram of V_2O_3 doped with Cr and Ti [reproduced from Ref. 68]. Doping V_2O_3 effects the lattice constants in a similar way as applying pressure (generated either by a hydrostatic pressure P , or by changing the V -concentration from V_2O_3 to $V_{2-y}O_3$) and leads to a Mott-Hubbard transition between the *paramagnetic* insulator (PI) and metal (PM). At lower temperatures, a Mott-Heisenberg transition between the *paramagnetic* metal (PM) and the *antiferromagnetic* insulator (AFI) is observed.

to surface properties. Due to the reduced coordination number at the surface the bandwidth is likely to be smaller, and the Coulomb interaction less screened, i.e., larger. Both effects make the system more correlated and, thus, might also explain why better agreement is found for $U = 5$ eV. Besides that, also the polycrystalline nature of the sample, as well as spin and orbital⁶⁶ fluctuation not taken into account in the LDA+DMFT approach, will lead to a further reduction of the quasi-particle weight.

4 Mott-Hubbard Metal-Insulator Transition in V_2O_3

One of the most famous examples of a cooperative electronic phenomenon occurring at intermediate coupling strengths is the transition between a paramagnetic metal and a paramagnetic insulator induced by the Coulomb interaction between the electrons – the Mott-Hubbard metal-insulator transition. The question concerning the nature of this transition poses one of the fundamental theoretical problems in condensed matter physics.⁶⁷ Correlation-induced metal-insulator transitions (MIT) are found, for example, in transition metal oxides with partially filled bands near the Fermi level. For such systems band theory typically predicts metallic behavior. The most famous example is V_2O_3 doped with Cr as shown in Fig. 9. While at low temperatures V_2O_3 is an antiferromagnetic insulator with monoclinic crystal symmetry, it has a corundum structure in the high-temperature paramagnetic phase. All transitions shown in the phase diagram are of first order. In the case

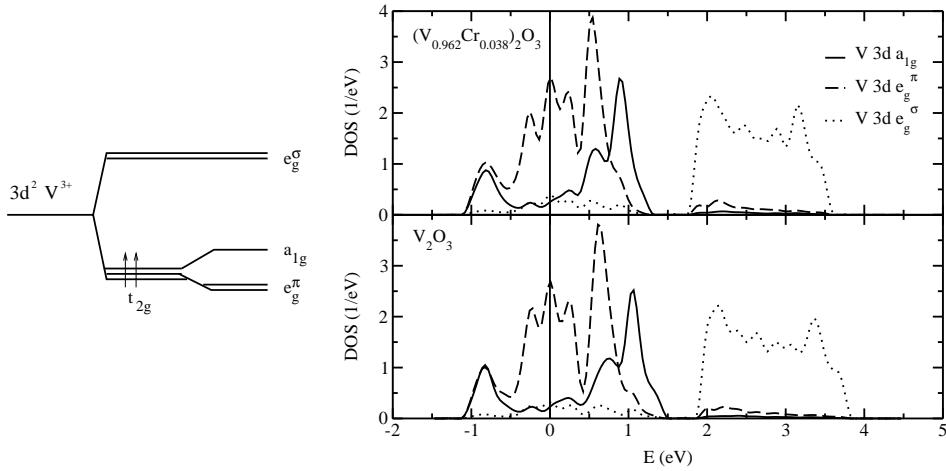


Figure 10. Left: Scheme of the 3d levels in the corundum crystal structure. Right: Partial LDA DOS of the 3d bands for paramagnetic metallic V_2O_3 and insulating $(V_{0.962}Cr_{0.038})_2O_3$ [reproduced from Ref.44].

of the transitions from the high-temperature paramagnetic phases into the low-temperature antiferromagnetic phase this is naturally explained by the fact that the transition is accompanied by a change in crystal symmetry. By contrast, the crystal symmetry across the MIT in the paramagnetic phase remains intact, since only the ratio of the c/a axes changes discontinuously. This may be taken as an indication for the predominantly electronic origin of this transition which is not accompanied by any conventional long-range order. From a models point of view the MIT is triggered by a change of the ratio of the Coulomb interaction U relative to the bandwidth W . Originally, Mott considered the extreme limits $W = 0$ (when atoms are isolated and insulating) and $U = 0$ where the system is metallic. While it is simple to describe these limits, the crossover between them, i.e., the metal-insulator transition itself, poses a very complicated electronic correlation problem. Among others, this metal-insulator transition has been addressed by Hubbard in various approximations⁶⁹ and by Brinkman and Rice within the Gutzwiller approximation.⁷⁰ During the last few years, our understanding of the MIT in the one-band Hubbard model has considerably improved, in particular due to the application of dynamical mean-field theory.⁷¹

Both the paramagnetic *metal* V_2O_3 and the paramagnetic *insulator* $(V_{0.962}Cr_{0.038})_2O_3$ have the same corundum crystal structure with only slightly different lattice parameters.^{72,73} Nevertheless, within LDA both phases are found to be metallic (see Fig. 10). The LDA DOS shows a splitting of the five Vanadium d-orbitals into three t_{2g} states near the Fermi energy and two e_g^σ states at higher energies. This reflects the (approximate) octahedral arrangement of oxygen around the vanadium atoms. Due to the trigonal symmetry of the corundum structure the t_{2g} states are further split into one a_{1g} band and two degenerate e_g^π bands, see Fig. 10. The only visible difference between $(V_{0.962}Cr_{0.038})_2O_3$ and V_2O_3 is a slight narrowing of the t_{2g} and e_g^σ bands by ≈ 0.2 and 0.1 eV, respectively as well as a weak downshift of the centers of gravity of both groups of bands for V_2O_3 . In particular, the insulating gap of the Cr-doped system is seen to be missing in the LDA DOS. Here we will employ LDA+DMFT(QMC) to show

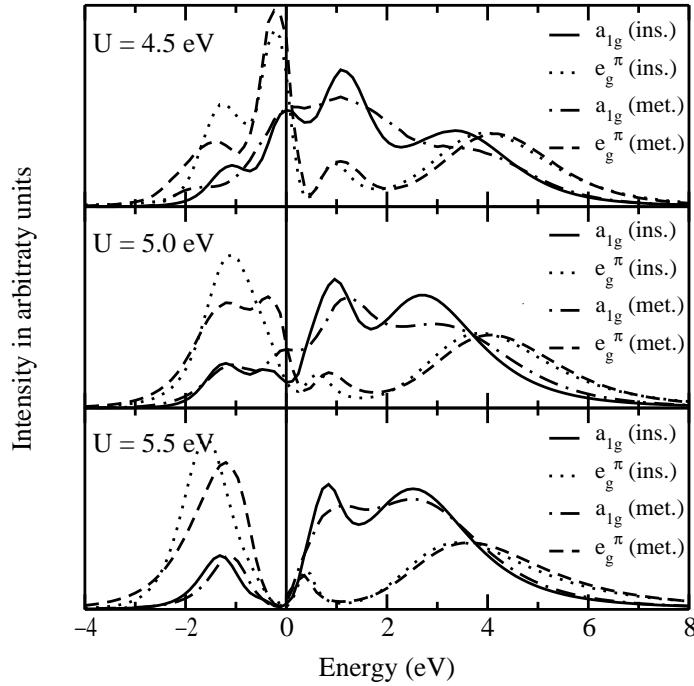


Figure 11. LDA+DMFT(QMC) spectra for paramagnetic $(V_{0.962}Cr_{0.038})_2O_3$ (“ins.”) and V_2O_3 (“met.”) at $U = 4.5, 5$ and 5.5 eV, and $T = 0.1$ eV ≈ 1000 K [reproduced from Ref.44].

explicitly that the insulating gap is caused by the electronic correlations. In particular, we make use of the simplification for transition metal oxides described in Section 2.7 and restrict the LDA+DMFT(QMC) calculation to the three t_{2g} bands at the Fermi energy, separated from the e_g^σ and oxygen bands.

While the Hund’s rule coupling J is insensitive to screening effects and may, thus, be obtained within LDA to a good accuracy ($J = 0.93$ eV²⁵), the LDA-calculated value of the Coulomb repulsion U has a typical uncertainty of at least 0.5 eV.²⁴ To overcome this uncertainty, we study the spectra obtained by LDA+DMFT(QMC) for three different values of the Hubbard interaction ($U = 4.5, 5.0, 5.5$) in Fig. 11. All QMC results presented were obtained for $T = 0.1$ eV. However, simulations for V_2O_3 at $U = 5$ eV, $T = 0.143$ eV, and $T = 0.067$ eV suggest only a minor smoothing of the spectrum with increasing temperature. From the results obtained we conclude that the critical value of U for the MIT is at about 5 eV: At $U = 4.5$ eV one observes pronounced quasiparticle peaks at the Fermi energy, i.e., characteristic metallic behavior, even for the crystal structure of the insulator $(V_{0.962}Cr_{0.038})_2O_3$, while at $U = 5.5$ eV the form of the calculated spectral function is typical for an insulator for both sets of crystal structure parameters. At $U = 5.0$ eV one is then at, or very close to, the MIT since there is a pronounced dip in the DOS at the Fermi energy for both a_{1g} and e_g^π orbitals for the crystal structure of $(V_{0.962}Cr_{0.038})_2O_3$, while for pure V_2O_3 one still finds quasiparticle peaks. (We note that at $T \approx 0.1$ eV one only observes metallic-like and insulator-like behavior, with a rapid but smooth crossover

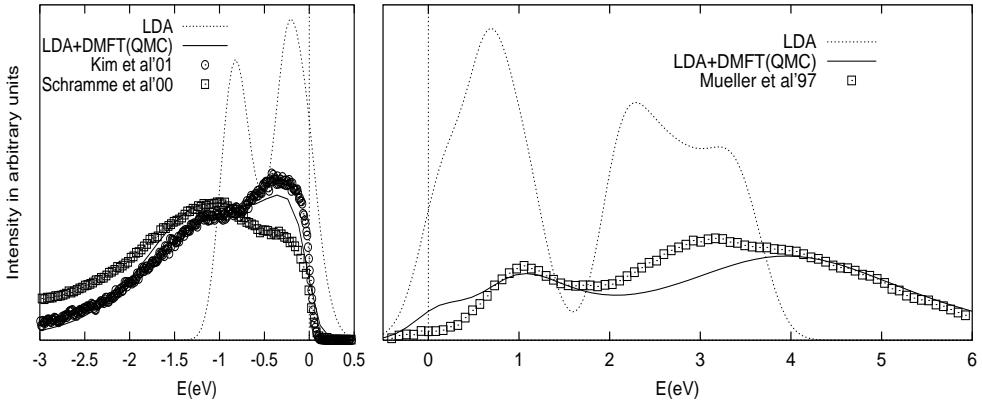


Figure 12. Comparison of the LDA+DMFT(QMC) spectrum⁴⁴ at $U = 5$ eV and $T = 0.1$ eV ≈ 1000 K below (left Figure) and above (right Figure) the Fermi energy (at 0 eV) with the LDA spectrum⁴⁴ and the experimental spectrum (left: photoemission spectrum of Schramme *et al.*⁷⁴ and Kim *et al.*,⁷⁵ right: X-ray absorption spectrum of Müller *et al.*⁷⁶).

between these two phases, since a sharp MIT occurs only at lower temperatures^{39,71}). The critical value of the Coulomb interaction $U \approx 5$ eV is in reasonable agreement with the values determined spectroscopically by fitting to model calculations, and by constrained LDA, see⁴⁴ for details.

To compare with the photoemission spectrum of V_2O_3 spectrum by Schramme *et al.*⁷⁴ and by Kim *et al.*⁷⁵ as well as with the X-ray absorption data by Müller *et al.*,⁷⁶ the LDA+DMFT(QMC) spectrum of Fig. 11 is multiplied with the Fermi function at $T = 0.1$ eV and Gauss-broadened by 0.05 eV to account for the experimental resolution. The theoretical result for $U = 5$ eV is seen to be in good agreement with experiment (Fig. 12). In contrast to the LDA results, our results not only describe the different bandwidths above and below the Fermi energy (≈ 6 eV and $\approx 2 - 3$ eV, respectively), but also the position of two (hardly distinguishable) peaks below the Fermi energy (at about -1 eV and -0.3 eV) as well as the pronounced two-peak structure above the Fermi energy (at about 1 eV and 3-4 eV). While LDA also gives two peaks below and above the Fermi energy, their position and physical origin is quite different. Within LDA+DMFT(QMC) the peaks at -1 eV and 3-4 eV are the incoherent Hubbard bands induced by the electronic correlations whereas in the LDA the peak at 2-3 eV is caused by the e_g^σ states and that at -1 eV is the band edge maximum of the a_{1g} and e_g^π states (see Fig. 10). Note that the theoretical and experimental spectrum is highly *asymmetric* w.r.t the Fermi energy. This high *asymmetry* which is caused by the orbital degrees of freedom is missing in the one-band Hubbard model which was used by Rozenberg *et al.*⁷⁷ to describe the optical spectrum of V_2O_3 .

The comparison between theory and experiment for Cr-doped *insulating* V_2O_3 is not as good as for metallic V_2O_3 , see Ref. 75. This might be, among other reasons, due to the different Cr-doping of experiment and theory, the difference in temperatures (which is important because the insulating gap of a Mott insulator is filled when increasing the temperature⁷¹), or the fact that every V ion has a unique neighbor in one direction, i.e., the LDA supercell calculation has *a pair* of V ions per unit cell. The latter aspect has so far not

been included but arises naturally when one goes from the simplified calculation scheme described in Section 2.7 (and employed in the present Section with different self-energies for the a_{1g} and e_g^π bands) to a full Hamiltonian calculation.

Particularly interesting are the spin and the orbital degrees of freedom in V_2O_3 . From our calculations,⁴⁴ we conclude that the spin state of V_2O_3 is $S = 1$ throughout the Mott-Hubbard transition region. This agrees with the measurements of Park *et al.*⁷⁸ and also with the data for the high-temperature susceptibility.⁷⁹ But, it is at odds with the $S = 1/2$ model by Castellani *et al.*⁸⁰ and with the results for a one-band Hubbard model which corresponds to $S = 1/2$ in the insulating phase and, contrary to our results, shows a substantial change of the local magnetic moment at the MIT.⁷¹ For the orbital degrees of freedom we find a predominant occupation of the e_g^π orbitals, but with a significant admixture of a_{1g} orbitals. This admixture decreases at the MIT: in the metallic phase we determine the occupation of the (a_{1g} , e_{g1}^π , e_{g2}^π) orbitals as (0.37, 0.815, 0.815), and in the insulating phase as (0.28, 0.86, 0.86). This should be compared with the experimental results of Park *et al.*⁷⁸ From their analysis of the linear dichroism data the authors concluded that the ratio of the configurations $e_g^\pi e_g^\pi : e_g^\pi a_{1g}$ is equal to 1:1 for the paramagnetic metallic and 3:2 for the paramagnetic insulating phase, corresponding to a one-electron occupation of (0.5, 0.75, 0.75) and (0.4, 0.8, 0.8), respectively. Although our results show a somewhat smaller value for the admixture of a_{1g} orbitals, the overall behavior, including the tendency of a *decrease* of the a_{1g} admixture across the transition to the insulating state, are well reproduced.

In the study above, the experimental crystal parameters of V_2O_3 and $(V_{0.962}Cr_{0.038})_2O_3$ have been taken from the experiment. This leaves the question unanswered whether a change of the lattice is the driving force behind the Mott transition, or whether it is the electronic Mott transition which causes a change of the lattice. For another system, Ce, we will show in Section 5 that the energetic changes near a Mott transition are indeed sufficient to cause a first-order volume change.

5 The Cerium Volume Collapse: An Example for a 4f-Electron System

Cerium exhibits a transition from the γ - to the α -phase with increasing pressure or decreasing temperature. This transition is accompanied by an unusually large volume change of 15%,⁸¹ much larger than the 1-2% volume change in V_2O_3 . The γ -phase may also be prepared in metastable form at room temperature in which case the reverse γ - α transition occurs under pressure.⁸² Similar volume collapse transitions are observed under pressure in Pr and Gd (for a recent review see Ref. 83). It is widely believed that these transitions arise from changes in the degree of 4f electron correlation, as is reflected in both the Mott transition⁸⁴ and the Kondo volume collapse (KVC)⁸⁵ models.

The Mott transition model envisions a change from itinerant, bonding character of the 4f-electrons in the α -phase to non-bonding, localized character in the γ -phase, driven by changes in the 4f-4f inter-site hybridization. Thus, as the ratio of the 4f Coulomb interaction to the 4f-bandwidth increases, a Mott transition occurs to the γ -phase, similar to the Mott-Hubbard transition of the 3d-electrons in V_2O_3 (Section 4).

The Kondo volume collapse⁸⁵ scenario ascribes the collapse to a strong change in the energy scale associated with the screening of the local 4f-moment by conduction electrons

(Kondo screening), which is accompanied by the appearance of an Abrikosov-Suhl-like quasiparticle peak at the Fermi level. In this model the $4f$ -electron spectrum of Ce would change across the transition in a fashion very similar to the Mott scenario, i.e., a strong reduction of the spectral weight at the Fermi energy should be observed in going from the α - to the γ -phase. The subtle difference comes about by the γ -phase having metallic f -spectra with a strongly enhanced effective mass as in a heavy fermion system, in contrast to the f -spectra characteristic of an insulator in the case of the Mott scenario. The f -spectra in the Kondo picture also exhibit Hubbard side-bands not only in the γ -phase, but in the α -phase as well, at least close to the transition. While local-density and static mean-field theories correctly yield the Fermi-level peaks in the f -spectra for the α -phase, they do not exhibit such additional Hubbard side-bands, which is sometimes taken as characteristic of the “ α -like” phase in the Mott scenario.⁸⁴ However, this behavior is more likely a consequence of the static mean-field treatment, as correlated solutions of both Hubbard and periodic Anderson models exhibit such residual Hubbard side-bands in the α -like regimes.⁸⁶

Typically, the Hubbard model and the periodic Anderson model are considered as paradigms for the Mott and KVC model, respectively. Although both models describe completely different physical situations it was shown recently that one can observe a surprisingly similar behavior at finite temperatures: the evolution of the spectrum and the local magnetic moment with increasing Coulomb interaction show very similar features as well as, in the case of a periodic Anderson model with nearest neighbor hybridization, the phase diagram and the charge compressibility.^{86,87} From this point of view the distinction between the two scenarios appears to be somewhat artificial, at least at temperatures relevant for the description of the α - γ transition.

For a realistic calculation of the Cerium α - γ transition, we employ the full Hamiltonian calculation described in Sections 2.2, 2.3, and 2.4 where the one-particle Hamiltonian was calculated by LDA and the $4f$ Coulomb interaction U along with the associated $4f$ site energy shift by a constrained LDA calculation (for details of the the two independent calculations presented in the current Section see Refs. 83,51 and Ref. 50). We have not included the spin-orbit interaction which has a rather small impact on LDA results for Ce, nor the intra-atomic exchange interaction which is less relevant for Ce as occupations with more than one $4f$ -electron on the same site are rare. Furthermore, the $6s$ -, $6p$ -, and $5d$ -orbitals are assumed to be non-interacting in the formalism of Eq. (13), Section 2.3. Note, that the $4f$ orbitals are even better localized than the $3d$ orbitals and, thus, uncertainties in U are relatively small and would only translate into a possible volume shift for the α - γ -transition.

The LDA+DMFT(QMC) spectral evolution of the Ce $4f$ -electrons is presented in Fig. 13. It shows similarities to V_2O_3 (Fig. 11, Section 4): At a volume per atom $V = 20 \text{ \AA}^3$, Fig. 13 shows that almost the entire spectral weight lies in a large quasiparticle peak with a center of gravity slightly above the chemical potential. This is similar to the LDA solution, however, a weak upper Hubbard band is also present even at this small volume. At the volumes 29 \AA^3 and 34 \AA^3 which approximately bracket the α - γ transition, the spectrum has a three peak structure. Finally, by $V = 46 \text{ \AA}^3$, the central peak has disappeared leaving only the lower and upper Hubbard bands. However, an important difference to V_2O_3 is that the spd -spectrum shows metallic behavior and, thus, Cerium remains a metal throughout this transition monitored by a vanishing $4f$ quasiparticle resonance.

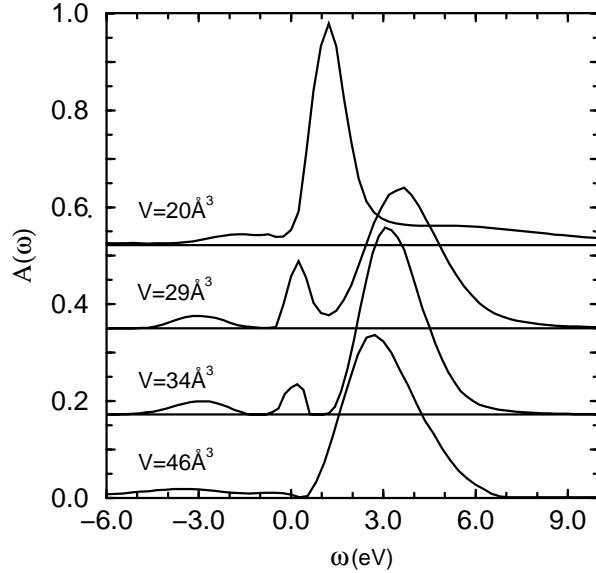


Figure 13. Evolution of the 4f spectral function $A(\omega)$ with volume at $T = 0.136$ eV ($\omega = 0$ corresponds to the chemical potential; curves are offset as indicated; $\Delta\tau = 0.11\text{eV}^{-1}$). Coinciding with the sharp anomaly in the correlation energy (Fig. 14), the central quasiparticle resonance disappears, at least at finite temperatures [reproduced from Ref. 51].

To study the energetic changes associated with the rapid change of the quasiparticle weight at the Fermi energy, we calculate the DMFT energy per site for the model Hamiltonian (13)

$$E_{\text{DMFT}} = \frac{T}{N_k} \sum_{n\mathbf{k}\sigma} \text{Tr}(H_{\text{LDA}}^0(\mathbf{k})G_{\mathbf{k}}(i\omega_n))e^{i\omega_n 0^+} + U_f d. \quad (47)$$

Here, Tr denotes the trace over the 16×16 matrices, T the temperature, N_k the number of \mathbf{k} points, $G_{\mathbf{k}}$ the Green function matrix w.r.t. the orbital indices, $H_{\text{LDA}}^0(\mathbf{k})$ the LDA one-particle matrix Eq. (17), and

$$d = \frac{1}{2} \sum'_{m\sigma, m'\sigma'} \langle \hat{n}_{ifm\sigma} \hat{n}_{ifm'\sigma'} \rangle \quad (48)$$

is a generalization of the one-band double occupation for multi-band models.

Fig. 14a shows our calculated DMFT(QMC) energies E_{DMFT} as a function of atomic volume at three temperatures *relative* to the paramagnetic Hartree Fock (HF) energies E_{PMHF} [of the Hamiltonian (13)], i.e., the energy contribution due to *electronic correlations*. Similarly given are the polarized HF energies which reproduce E_{DMFT} at large volumes and low temperatures. With decreasing volume, however, the DMFT energies bend away from the polarized HF solutions. Thus, at $T = 0.054$ eV ≈ 600 K, a region of negative curvature in $E_{\text{DMFT}} - E_{\text{PMHF}}$ is evident within the observed two phase region (arrows).

Fig. 14b presents the calculated LDA+DMFT total energy $E_{\text{tot}}(T) = E_{\text{LDA}}(T) + E_{\text{DMFT}}(T) - E_{\text{mLDA}}(T)$ where E_{mLDA} is the energy of an LDA-like solution of the Hamil-

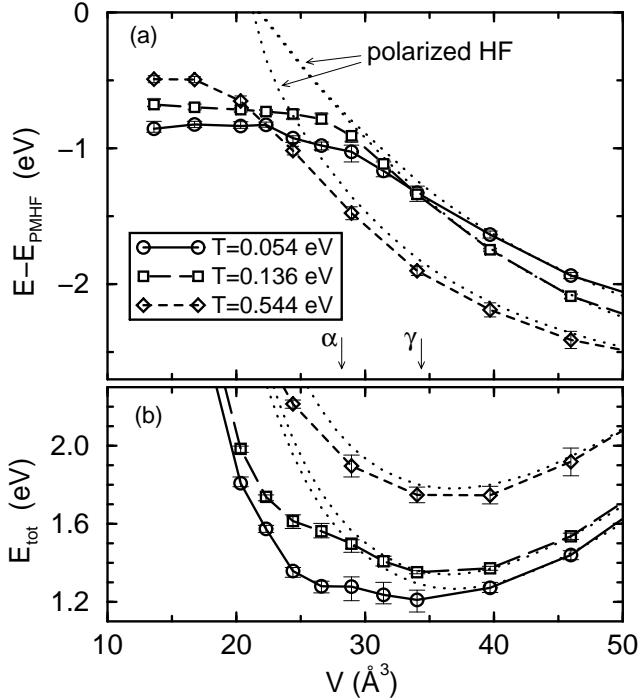


Figure 14. (a) Correlation energy $E_{\text{DMFT}} - E_{\text{PMHF}}$ as a function of atomic volume (symbols) and polarized HF energy $E_{\text{AHF}} - E_{\text{PMHF}}$ (dotted lines which, at large V , approach the DMFT curves for the respective temperatures); arrows: observed volume collapse from the α - to the γ -phase. The correlation energy sharply bends away from the polarized HF energy in the region of the transition. (b) The resultant negative curvature leads to a growing depression of the total energy near $V = 26\text{--}28 \text{ \AA}^3$ as temperature is decreased, consistent with an emerging double well at still lower temperatures and thus the $\alpha\text{-}\gamma$ transition. The curves at $T = 0.544 \text{ eV}$ were shifted downwards in (b) by -0.5 eV to match the energy range [reproduced from Ref. 51].

tonian (13).⁸⁸ Since both E_{LDA} and $E_{\text{PMHF}} - E_{\text{mLDA}}$ have positive curvature throughout the volume range considered, it is the negative curvature of the correlation energy in Fig. 14a which leads to the dramatic depression of the LDA+DMFT total energies in the range $V = 26\text{--}28 \text{ \AA}^3$ for decreasing temperature, which contrasts to the smaller changes near $V = 34 \text{ \AA}^3$ in Fig. 14b. This trend is consistent with a double well structure emerging at still lower temperatures (prohibitively expensive for QMC simulations), and with it a first-order volume collapse. This is in reasonable agreement with the experimental volume collapse given our use of energies rather than free energies, the different temperatures, and the LDA and DMFT approximations. A similar scenario has been proposed recently for the $\delta\text{-}\alpha$ transition in Pu on the basis of LDA+DMFT calculations,⁴⁸ which solves DMFT by an ansatz inspired by IPT and includes a modification of the DFT/LDA step to account for the density changes introduced by the DMFT.⁴⁹

In a separate LDA+DMFT(NCA) calculation for Ce, we have obtained a number of physical quantities for both phases which may be compared to experimental values.⁵⁰ Various static properties extracted from the calculations⁵⁰ and their counterparts from experi-

	$\alpha\text{-Ce}^{\text{Theo}}$	$\alpha\text{-Ce}^{89,90}$	$\gamma\text{-Ce}^{\text{Theo}}$	$\gamma\text{-Ce}^{89,90}$
P_0	0.126	0.1558	0.0150	0.0426
P_1	0.829	0.8079	0.9426	0.9444
P_2	0.044	0.0264	0.0423	0.0131
n_f	0.908	0.8...0.861	1.014	0.971...1
$T_K, [\text{K}]$	1000	945...2000	30	60...95
$\chi, [10^{-3}\text{emu/mol}]$	1.08	0.53...0.70	24	8.0...12

Table 1. Comparison between LDA+DMFT(NCA) calculated parameters for both α - and γ -phase at $T = 580 \text{ K}$ and experimental values^{89,90} [reproduced from Ref. 50]. P_0 , P_1 and P_2 are partial probabilities for an empty, singly and doubly occupied $4f$ -state, n_f is the f -electron occupancy, T_K the estimated Kondo temperature, and χ the magnetic susceptibility.

ments are collected in Table 1 and show an overall fair to good agreement in the tendencies and, except for the susceptibility, the absolute values. Since the calculation of the magnetic susceptibility χ in Ref. 50 was based on simplifying assumptions, the absolute numbers cannot be expected to match experiment. However, the general tendency and especially the ratio between α - and γ -Ce is in good agreement with experiment. It is interesting to note that the experiments predict a finite Kondo screening-scale for both phases, which actually would point toward the KVC scenario. Finally, let us compare spectral functions for the $4f$ -states calculated with the LDA+DMFT(NCA) approach to experimental data.⁹¹ The photoemission spectrum for $\alpha\text{-Ce}$ (upper part of Fig. 15) shows a main structure between 3 eV and 7 eV, which is attributed to $4f^2$ final state multiplets. In the calculated spectrum all excitations to $4f^2$ states are described by the featureless upper Hubbard band. As a consequence of the simplified interaction model all doubly occupied states are degenerate. This shortcoming in our calculation is responsible for the sharply peaked main structure. The neglected exchange interaction would produce a multiplet structure, which would be closer to the experiment. The experimental peak at about 0.5 eV is attributed to two $4f^1$ final states, which are split by spin-orbit coupling. The calculated f -spectrum shows a sharp quasiparticle or Kondo resonance slightly above the Fermi energy, which is the result of the formation of a singlet state between f - and conduction states. We thus suggest that the spectral weight seen in the experiment is a result of this quasiparticle resonance. Since we did not yet include spin-orbit coupling in our model, we cannot observe the mentioned splitting of the resonance. However, as it is well known,⁹⁴ the introduction of such a splitting would eventually split the Kondo resonance. If we used the experimentally determined value of about 0.3 eV for the spin-orbit splitting,⁹² the observed resonance of width 0.5 eV would indeed occur in the calculations. In the lower part of Fig. 15, a comparison between experiment and our calculation for $\gamma\text{-Ce}$ is shown. The most striking difference between lower and upper part of Fig. 15 is the absence of the Kondo resonance in the high temperature phase ($\gamma\text{-Ce}$; transition temperature 141 K ⁸¹) which is in agreement with our calculations.

In the insets of Fig. 15, our results for the non-occupied states in the f -density are compared with RIPES data.⁹³ The calculated f -spectra were multiplied by the Fermi-step function and broadened with an Lorentzian of the width 0.1 eV in order to mimic the experimental resolution in the theoretical curves. Here, as above the theoretical overestimation of the sharpness of the upper Hubbard band is a consequence of the simplified local inter-

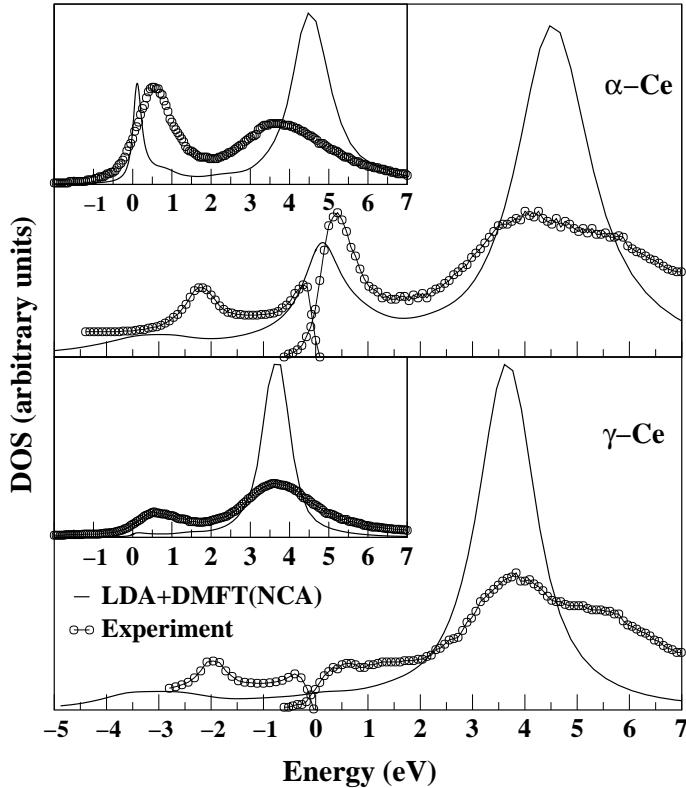


Figure 15. Comparison between combined photoemission⁹¹ and BIS⁹² experimental (circles) and theoretical (solid line) f -spectra for α - (upper part) and γ -Ce (lower part) at $T = 580$ K. The relative intensities of the BIS and photoemission portions are roughly for one $4f$ electron. The experimental and theoretical spectra were normalized and the theoretical curve was broadened with resolution width of 0.4 eV. In the insets a comparison between RIPES⁹³ experimental (circles) and theoretical (solid line) f -spectra is given. The experimental and theoretical data were normalized and the theoretical curve was broadened with broadening coefficient of 0.1 eV [reproduced from Ref. 50].

action and thus of the missing multiplet structure of the $4f^2$ -final states. The main feature of the experimental spectra, i.e., a strong decrease of the intensity ratio for Kondo resonance and upper Hubbard band peaks from α - to γ -Ce, can also be seen in the theoretical curves of Fig. 15 as well as in the study presented in Fig. 13. A more thorough comparison of these two independent LDA+DMFT(NCA) and LDA+DMFT(QMC) studies remains to be done.

6 Conclusion and Outlook

In this paper we discussed the set-up of the computational scheme LDA+DMFT which merges two non-perturbative, complementary investigation techniques for many-particle systems in solid state physics. LDA+DMFT allows one to perform *ab initio* calculations

of real materials with strongly correlated electrons. Using the band structure results calculated within local density approximation (LDA) as input, the missing electronic correlations are introduced by dynamical mean-field theory (DMFT). On a technical level this requires the solution of an effective self-consistent, multi-band Anderson impurity problem by some numerical method (e.g. IPT, NCA, QMC). Comparison of the photoemission spectrum of $\text{La}_{1-x}\text{Sr}_x\text{TiO}_3$ calculated by LDA+DMFT using IPT, NCA, and QMC reveal that the choice of the evaluation method is of considerable importance. Indeed, only with the numerically exact QMC quantitatively reliable results are obtained. The results of the LDA+DMFT(QMC) approach were found to be in very good agreement with the experimental photoemission spectrum of $\text{La}_{0.94}\text{Sr}_{0.06}\text{TiO}_3$.

We also presented results of a LDA+DMFT(QMC) study⁴⁴ of the Mott-Hubbard metal-insulator transition (MIT) in the paramagnetic phase of (doped) V_2O_3 . These results showed a Mott-Hubbard MIT at a reasonable value of the Coulomb interaction $U \approx 5\text{eV}$ and are in very good agreement with the experimentally determined photoemission and X-ray absorption spectra for this system, i.e., above *and* below the Fermi energy. In particular, we find a spin state $S = 1$ in the paramagnetic phase, and an orbital admixture of $e_g^\pi e_g^\pi$ and $e_g^\pi a_{1g}$ configurations, which both agree with recent experiments. Thus, LDA+DMFT(QMC) provides a remarkably accurate microscopic theory of the strongly correlated electrons in the paramagnetic metallic phase of V_2O_3 .

Another material where electronic correlations are considered to be important is Cerium. We reviewed our recent investigations of the Ce α - γ transition, based on LDA+DMFT(QMC)⁵¹ and LDA+DMFT(NCA)⁵⁰ calculations. The spectral results and susceptibilities show the same tendency as seen in the experiment, namely a dramatic reduction in the size of the quasiparticle peak at the Fermi level when passing from the α - to the γ -phase. While we do not know at the moment whether the zero-temperature quasiparticle peak will completely disappear at an even larger volume (i.e., in a rather Mott-like fashion) or simply fade away continuously with increasing volume (i.e., in a more Kondo-like fashion), an important aspect of our results is that the rapid reduction in the size of the peak seems to coincide with the appearance of a negative curvature in the correlation energy and a shallow minimum in the total energy. This suggest that the electronic correlations responsible for the reduction of the quasiparticle peak are associated with energetic changes that are strong enough to cause a volume collapse in the sense of the Kondo volume collapse model,⁸⁵ or a Mott transition model⁸⁴ including electronic correlations.

At present LDA+DMFT is the only available *ab initio* computational technique which is able to treat correlated electronic systems close to a Mott-Hubbard MIT, heavy fermions, and *f*-electron materials. The physical properties of such systems are characterized by the correlation-induced generation of small, Kondo-like energy scales which require the application of genuine many-body techniques. The appearance of Kondo-like energy scales in strongly correlated systems leads to several experimentally relevant consequences. One of the most important features is the enhancement of the quasiparticle mass m^* (i.e., the decrease of the quasiparticle residue Z). This phenomenon can be observed as an enhancement of the coefficient γ in the specific heat. Another important characteristic is the Wilson ratio between γ and the Pauli spin susceptibility χ . Future LDA+DMFT investigations will determine these quantities for real systems, as well as the optical conductivity, phase-diagrams, the local vertex function, and various susceptibilities.

LDA+DMFT provides, at last, a powerful tool for *ab initio* investigations of real mate-

rials with strong electronic correlations. Indeed, LDA+DMFT depends on the input from both band structure theory *and* many-body approaches. Hence, for this computational scheme to be entirely successful in the future two strong and vital communities will finally have to join forces.

Acknowledgments

We are grateful to J. W. Allen, P. W. Anderson, R. Bulla, R. Claessen, U. Eckern, G. Esirgen, A. Georges, K.-H. Höck, S. Horn, M. Jarrell, J. Keller, H.-D. Kim, D. E. Kondakov, G. Kotliar, J. Lægsgaard, A. Lichtenstein, D. van der Marel, T. M. Rice, G. A. Sawatzky, J. Schmalian, M. Schramme, M. Sigrist, M. Ulmke, and M. Zölfli for helpful discussions. We thank A. Sandvik for making available his maximum entropy code. The QMC code of Ref. 11 App. D was modified for use for some of the results of Section 5. This work was supported in part by the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 484 (DV,GK,VE), Forschergruppe HO 955/2 (VE), and project Pr 298/5-1 & 2 (TP), the Russian Foundation for Basic Research by RFFI-01-02-17063 (VA,IN), the U.S. Department of Energy by University California LLNL under contract No. W-7405-Eng-48. (AM), the U.S. National Science Foundation by DMR-9985978 (RS), a Feodor-Lynen grant of the Alexander von Humboldt foundation (KH), the Lorentz Center in Leiden, the Leibniz-Rechenzentrum, München, and the John v. Neumann Institute for Computing, Jülich.

References

1. M. Born und R. Oppenheimer, Ann. Phys. (Leipzig) **84**, 457 (1927).
2. R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).
3. W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
4. E. Müller-Hartmann, Z. Phys. B **74**, 507 (1989); *ibid.* B **76**, 211 (1989).
5. U. Brandt und C. Mielsch, Z. Phys. B **75**, 365 (1989); *ibid.* B **79**, 295 (1989); *ibid.* B **82**, 37 (1991).
6. V. Janiš, Z. Phys. B **83**, 227 (1991); V. Janiš and D. Vollhardt, Int. J. Mod. Phys. **6**, 731 (1992).
7. A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992).
8. M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992).
9. D. Vollhardt, in *Correlated Electron Systems*, edited by V. J. Emery, World Scientific, Singapore, 1993, p. 57.
10. Th. Pruschke, M. Jarrell, and J. K. Freericks, Adv. in Phys. **44**, 187 (1995).
11. A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
12. V. I. Anisimov, A. I. Poteryaev, M. A. Korotin, A. O. Anokhin, and G. Kotliar, J. Phys. Cond. Matter **9**, 7359 (1997).
13. An introduction to LDA+DMFT less complete than the present Proceedings has been given in K. Held, I. A. Nekrasov, N. Blümer, V. I. Anisimov, and D. Vollhardt, Int. J. Mod. Phys. B **15**, 2611 (2001).
14. P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).

15. M. Levy, Proc. Natl. Acad. Sci. (USA), **76**, 6062 (1979).
16. W. Kohn and L. J. Sham, Phys. Rev. **140**, 4A, A1133 (1965); W. Kohn and L. J. Sham, Phys. Rev. A - Gen.Phys. **140**, 1133 (1965); L. J. Sham and W. Kohn, Phys. Rev. **145 N 2**, 561 (1966).
17. L. Hedin and B. Lundqvist, J. Phys. C: Solid State Phys. **4**, 2064 (1971); U. von Barth and L. Hedin, J. Phys. C: Solid State Phys. **5**, 1629 (1972).
18. D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
19. O. K. Andersen, Phys. Rev. B **12**, 3060 (1975); O. Gunnarsson, O. Jepsen, and O. K. Andersen, Phys. Rev. B **27**, 7144 (1983); O. K. Andersen and O. Jepsen, Phys. Rev. Lett. **53**, 2571 (1984).
20. T. C. Leung, X. W. Wang, and B. N. Harmon, Phys. Rev. B **37**, 384 (1988); W. E. Pickett, Rev. Mod. Phys. **61**, 433 (1989).
21. J. Wahle, N. Blümer, J. Schlipf, K. Held, and D. Vollhardt, Phys. Rev. B **58**, 12749 (1998).
22. V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991); V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, J. Phys. Cond. Matter **9**, 767 (1997).
23. O. Gunnarsson, O. K. Andersen, O. Jepsen, and J. Zaanen, Phys. Rev. B **39**, 1708 (1989).
24. I. A. Nekrasov, K. Held, N. Blümer, A. I. Poteryaev, V. I. Anisimov, and D. Vollhardt, Euro. Phys. J. B **18**, 55 (2000).
25. I. Solovyev, N. Hamada, and K. Terakura, Phys. Rev. B **53**, 7158 (1996).
26. A. I. Lichtenstein and M. I. Katsnelson, Phys. Rev. B **57**, 6884 (1998).
27. V. Drchal, V. Janiš, and J. Kudrnovský, in *Electron Correlations and Material Properties*, edited by A. Gonis, N. Kioussis, and M. Ciftan, Kluwer/Plenum, New York, 1999, p. 273.
28. J. Lægsgaard and A. Svane, Phys. Rev. B **58**, 12817 (1998).
29. Th. Wolenski, *Combining bandstructure and dynamical mean-field theory: A new perspective on V_2O_3* , Ph.D. Thesis, Universität Hamburg 1998 (Shaker Verlag, Aachen, 1999).
30. M. B. Zölfli, Th. Pruschke, J. Keller, A. I. Poteryaev, I. A. Nekrasov, and V. I. Anisimov, Phys. Rev. B **61**, 12810 (2000).
31. P. W. Anderson, in *Moment formation in solids*, edited by W. J. L. Buyers, Plenum Press, New York and London, 1984, p. 313.
32. H. Keiter and J. C. Kimball, Phys. Rev. Lett. **25**, 672 (1970); N. E. Bickers, D. L. Cox, and J. W. Wilkins, Phys. Rev. B **36**, 2036 (1987).
33. Th. Pruschke and N. Grewe, Z. Phys. B **74**, 439 (1989).
34. Th. Pruschke, D. L. Cox, and M. Jarrell, Phys. Rev. B **47**, 3553 (1993).
35. J. E. Hirsch and R. M. Fye, Phys. Rev. Lett. **56**, 2521 (1986); M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992); M. Rozenberg, X. Y. Zhang, and G. Kotliar, Phys. Rev. Lett. **69**, 1236 (1992); A. Georges and W. Krauth, Phys. Rev. Lett. **69**, 1240 (1992); M. Jarrell, in *Numerical Methods for Lattice Quantum Many-Body Problems*, edited by D. Scalapino, Addison Wesley, 1997.
36. M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994).
37. R. Bulla, Adv. Sol. State Phys. **46**, 169 (2000).
38. H. Kajueter and G. Kotliar, Int. J. Mod. Phys. **11**, 729 (1997).

39. M. J. Rozenberg, Phys. Rev. B **55**, R4855 (1997).
40. J. E. Han, M. Jarrell, and D. L. Cox, Phys. Rev. B **58**, R4199 (1998).
41. K. Held and D. Vollhardt, Euro. Phys. J. B **5**, 473 (1998).
42. M. I. Katsnelson and A. I. Lichtenstein, J. Phys. Cond. Matter **11**, 1037 (1999).
43. M. I. Katsnelson and A. I. Lichtenstein, Phys. Rev. B **61**, 8906 (2000).
44. K. Held, G. Keller, V. Eyert, V. I. Anisimov, and D. Vollhardt, Phys. Rev. Lett. **86**, 5345 (2001).
45. A. Liebsch and A. I. Lichtenstein, Phys. Rev. Lett. **84**, 1591 (2000).
46. V. I. Anisimov, I. A. Nekrasov, D. E. Kondakov, T. M. Rice, and M. Sigrist, cond-mat/0011460, ibid. cond-mat/0107095 (2001).
47. A. I. Lichtenstein, M. I. Katsnelson, and G. Kotliar **87**, 67205 (2001).
48. S. Y. Savrasov, G. Kotliar, and E. Abrahams, Nature **410**, 793 (2001).
49. S. Y. Savrasov and G. Kotliar, cond-mat/0106308.
50. M. B. Zölf, I. A. Nekrasov, Th. Pruschke, V. I. Anisimov, J. Keller, Phys. Rev. Lett. (in press), cond-mat/0101280.
51. K. Held, A. K. McMahan, and R. T. Scalettar, Phys. Rev. Lett. (in press), cond-mat/0106599; A. K. McMahan, K. Held, and R. T. Scalettar, in preparation.
52. W. Weber, J. Büinemann, and F. Gebhard, in *Band-Ferromagnetism*, edited by K. Baberschke, M. Donath, and W. Nolting, Lecture Notes in Physics, Vol. 580 (Springer, Berlin, 2001), p. 9; J. Büinemann, F. Gebhard, W. Weber, Phys. Rev. B **57**, 6896 (1998).
53. M. Suzuki, Prog. Theor. Phys. **56**, 1454 (1976).
54. One limitation of QMC is that it is very difficult to deal with the spin-flip term of the Hund's rule coupling because of a "minus-sign problem" which arises in a Hubbard-Stratonovich decoupling of this spin-flip term, see K. Held, Ph.D. thesis Universität Augsburg 1999 (Shaker Verlag, Aachen, 1999). In the particle-hole symmetric case another decoupling scheme which includes the spin-flip term is possible without "minus-sign problem", see Y. Motome and M. Imada, J. Phys. Soc. Jap. **66**, 1872 (1997).
55. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
56. M. Jarrell and J. E. Gubernatis, Physics Reports **269**, 133 (1996).
57. E. Müller-Hartmann, Z. Phys. B **57**, 281 (1984).
58. G. Kotliar, Physica B 259-261, 711 (1999). R. Chitra and G. Kotliar, Phys. Rev. B **62**, 12715 (2000).
59. D. A. MacLean, H.-N. Ng, and J. E. Greidan, J. Solid State Chem. **30**, 35 (1979).
60. M. Eitel and J. E. Greidan, Journal of the Less-Common Metals **116**, 95 (1986).
61. J. P. Gopel, J. E. Greidan, and D. A. MacLean, J. Solid State Chem. **43**, 244 (1981).
62. Y. Okimoto, T. Katsufuji, Y. Okada, T. Arima, and Y. Tokura, Phys. Rev. B **51**, 9581 (1995).
63. W. Heindl, Th. Pruschke, and J. Keller, J. Phys. – Condens. Matter **12**, 2245 (2000).
64. A. Fujimori *et al.*, Phys. Rev. Lett. **69**, 1796 (1992). A. Fujimori *et al.*, Phys. Rev. B **46**, 9841 (1992). A qualitatively *and* quantitatively similar spectrum was obtained recently by Yoshida *et al.*, cond-mat/9911446, with an experimental energy resolution of only 30 meV. One may suppose, however, that broadening effects due to the polycrystalline nature of the sample and surface effects are, then, larger than the in-

strumental resolution.

65. At present, QMC simulations of the DMFT equations are not feasible at the experimental temperature (80K). We note, however, that no intrinsic temperature dependence was observed in the experiment,⁶⁴ at least up to room temperature.
66. B. Keimer, D. Casa, A. Ivanov, J.W. Lynn, M. v. Zimmermann, J.P. Hill, D. Gibbs, Y. Taguchi, and Y. Tokura, Phys. Rev. Lett. **85**, 3946 (2000).
67. N. F. Mott, Rev. Mod. Phys. **40**, 677 (1968); *Metal-Insulator Transitions* (Taylor & Francis, London, 1990); F. Gebhard, *The Mott Metal-Insulator Transition* (Springer, Berlin, 1997).
68. D. B. McWahn *et al.*, Phys. Rev. B **7**, 1920 (1973).
69. J. Hubbard, Proc. Roy. Soc. London Ser. A **276**, 238 (1963); **277**, 237 (1963); **281**, 401 (1964).
70. W. F. Brinkman and T. M. Rice, Phys. Rev. B **2**, 4302 (1970).
71. G. Moeller, Q. Si, G. Kotliar, and M. Rozenberg, Phys. Rev. Lett. **74**, 2082 (1995); J. Schlipf, M. Jarrell, P. G. J. van Dongen, N. Blümer, S. Kehrein, Th. Pruschke, and D. Vollhardt, Phys. Rev. Lett. **82**, 4890 (1999); M. J. Rozenberg, R. Chitra, and G. Kotliar, Phys. Rev. Lett. **83**, 3498 (1999); R. Bulla, Phys. Rev. Lett. **83**, 136 (1999); R. Bulla, T. A. Costi, and D. Vollhardt, Phys. Rev. B **64**, 45103 (2001).
72. P. D. Dernier, J. Phys. Chem. Solids **31**, 2569 (1970).
73. Use of the crystal structure of Cr-doped V₂O₃ for the insulating phase of pure V₂O₃ is justified by the observation that Cr-doping is equivalent to the application of (negative) pressure.
74. M. Schramme, Ph.D. thesis, Universität Augsburg 2000 (Shaker Verlag, Aachen, 2000); M. Schramme *et al.* (unpublished).
75. H.-D. Kim, J.-H. Park, J. W. Allen, A. Sekiyama, A. Yamasaki, K. Kadono, S. Suga, Y. Saitoh, T. Muro, and P. Metcalf, cond-mat/0108044.
76. O. Müller, J. P. Urbach, E. Goering, T. Weber, R. Barth, H. Schuler, M. Klemm, S. Horn, and M. L. denBoer, Phys. Rev. B **56**, 15056 (1997).
77. M. J. Rozenberg, G. Kotliar, H. Kajueter, G. A. Thomas, D. H. Rapkine, J. M. Honig, and P. Metcalf, Phys. Rev. Lett. **75**, 105 (1995).
78. J.-H. Park, L. H. Tjeng, A. Tanaka, J. W. Allen, C. T. Chen, P. Metcalf, J. M. Honig, F. M. F. de Groot, and G. A. Sawatzky, Phys. Rev. B **61**, 11 506 (2000).
79. D. J. Arnold and R. W. Mires, J. Chem. Phys. **48**, 2231 (1968).
80. C. Castellani, C. R. Natoli, and J. Ranninger, Phys. Rev. B **18**, 4945 (1978); **18**, 4967 (1978); **18**, 5001 (1978).
81. *Handbook on the Physics and Chemistry of Rare Earths*, edited by K. A. Gschneider Jr. and L. R. Eyring (North-Holland, Amsterdam, 1978); in particular, D. G. Koskenmaki and K. A. Gschneider Jr., *ibid*, p.337.
82. J. S. Olsen, L. Gerward, U. Benedict, and J.-P. Itié, Physica **133B**, 129 (1985).
83. A. K. McMahan, C. Huscroft, R. T. Scalettar, and E. L. Pollock, J. Comput.-Aided Mater. Design **5**, 131 (1998).
84. B. Johansson, Philos. Mag. **30**, 469 (1974); B. Johansson, I.A. Abrikosov, M. Aldén, A. V. Ruban, and H.L. Skriver Phys. Rev. Lett. **74**, 2335 (1995).
85. J. W. Allen and R. M. Martin, Phys. Rev. Lett. **49**, 1106, (1982); J. W. Allen and L. Z. Liu, Phys. Rev. B **46**, 5047, (1992); M. Lavagna, C. Lacroix, and M. Cyrot, Phys. Lett. **90A**, 210 (1982).

86. K. Held, C. Huscroft, R.T. Scalettar, and A.K. McMahan, Phys. Rev. Lett. **85**, 373 (2000); see also C. Huscroft, A. K. McMahan, and R. T. Scalettar, Phys. Rev. Lett. **82**, 2342 (1999).
87. K. Held and R. Bulla, Eur. Phys. J. B **17**, 7 (2000).
88. We solve self-consistently for n_f using a 4f self energy $\Sigma = U_f(n_f - \frac{1}{2})$, and then remove this contribution from the eigenvalue sum to get the kinetic energy. The potential energy is taken to be $\frac{1}{2}U_f n_f(n_f - 1)$.
89. L. Z. Liu, J. W. Allen, O. Gunnarsson, N. E. Christensen, and O. K. Andersen, Phys. Rev. B **45**, 8934 (1992).
90. A. P. Murani, Z. A. Bowden, A. D. Taylor, R. Osborn, and W. G. Marshall, Phys. Rev. B **48**, 13981 (1993).
91. D. M. Wieliczka, C. G. Olson, and D. W. Lynch, Phys. Rev. B **29**, 3028 (1984).
92. E. Wuilloud, H. R. Moser, W. D. Schneider, and Y. Baer, Phys. Rev. B **28**, 7354 (1983).
93. M. Grioni, P. Weibel, D. Malterre, Y. Baer, and L. Duo, Phys. Rev. B **55**, 2056 (1997).
94. T. A. Costi, Phys. Rev. Lett. **85**, 1504 (2000).

Classical Molecular Dynamics

Godehard Sutmann

John von Neumann Institute for Computing
Central Institute for Applied Mathematics
Research Centre Jülich, 52425 Jülich, Germany
E-mail: g.sutmann@fz-juelich.de

An introduction to classical molecular dynamics simulation is presented. In addition to some historical notes, an overview is given over particle models, integrators and different ensemble techniques. In the end, methods are presented for parallelisation of short range interaction potentials. The efficiency and scalability of the algorithms on massively parallel computers is discussed with an extended version of Amdahl's law.

1 Introduction

Computer simulation methods have become a very powerful tool to attack the many-body problem in statistical physics, physical chemistry and biophysics. Although the theoretical description of complex systems in the framework of statistical physics is rather well developed and the experimental techniques for detailed microscopic information are rather sophisticated, it is often only possible to study specific aspects of those systems in great detail via the simulation. On the other hand, simulations need specific input parameters that characterize the system in question, and which come either from theoretical considerations or are provided by experimental data. Having characterized a physical system in terms of model parameters, simulations are often used both to solve theoretical models beyond certain approximations and to provide a hint to experimentalists for further investigations. In the case of big experimental facilities it is even often required to prove the potential outcome of an experiment by computer simulations. In that way one can say that the field of computer simulations has developed into a very important branch of science, which on the one hand helps theorists and experimentalists to go beyond their *inherent limitations* and on the other hand is a scientific field on its own.

The traditional simulation methods for many-body systems can be divided into two classes of stochastic and deterministic simulations, which are largely covered by the Monte Carlo (MC) method and the molecular dynamics (MD) method, respectively. Monte Carlo simulations probe the configuration space by trial moves of particles. Within the so-called Metropolis algorithm, the energy change from step n to $n + 1$ is used as a trigger to accept or reject the new configuration. Paths towards lower energy are always accepted, those to higher energy are accepted with a probability governed by Boltzmann statistics. In that way, properties of the system can be calculated by averaging over all Monte Carlo moves (where one move means that every degree of freedom is probed once on average). By contrast, MD methods are governed by the system's Hamiltonian and consequently Hamilton's equations of motion

$$\dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} \quad , \quad \dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i} \quad (1)$$

are integrated to move particles to new positions and to get new velocities at these new positions. This is an advantage of MD simulations with respect to MC, since not only is the configuration space probed but the whole phase space which gives additional information about the dynamics of the system. Both methods are complementary in nature but they lead to the same averages of static quantities, given that the system under consideration is ergodic and the same statistical ensemble is used.

Although there are different methods to obtain information about complex systems, particle simulations always require a model for the interaction between system constituents. This model has to be tested against experimental results, i.e. it should reproduce or approximate experimental findings like distribution functions or phase diagrams, and theoretical constraints, i.e. it should obey certain fundamental or limiting laws like energy conservation.

Concerning MD simulations the ingredients for a program are basically threefold:

- (i) As already mentioned, a model for the interaction between system constituents (atoms, molecules, surfaces etc.) is needed. Often, it is assumed that particles interact only pairwise, which is exact e.g. for particles with fixed partial charges. This assumption greatly reduces the computational effort and the work to implement the model into the program.
- (ii) An integrator is needed, which propagates particle positions and velocities from time t to $t + \delta t$. It is a finite difference scheme which moves trajectories discretely in time. The time step δt has to be chosen properly to guarantee stability of the integrator, i.e. there should be no drift in the system's energy.
- (iii) A statistical ensemble has to be chosen, where thermodynamic quantities like pressure, temperature or the number of particles are controlled. The natural choice of an ensemble in MD simulations is the microcanonical ensemble (NVE), since the system's Hamiltonian without external potentials is a conserved quantity. Nevertheless, there are extensions to the Hamiltonian which also allow to simulate different statistical ensembles.

These steps essentially define an MD simulation. Having this tool at hand, it is possible to obtain *exact* results within numerical precision. Results are only correct with respect to the model which enters into the simulation and they have to be tested against theoretical predictions and experimental findings. If the simulation results differ from the *real system* properties or are incompatible with *solid* theoretical manifestations, the model has to be refined. This procedure can be understood as an adaptive refinement which leads in the end to an approximation of a model of the *real world* at least for certain properties. The model itself may be constructed from plausible considerations, where parameters are chosen from neutron diffraction or NMR measurements. It may also result from first principle investigations, like quantum *ab initio* calculations. Although the electronic distribution of the particles is calculated very accurately, this type of model building contains also some approximations, since many-body interactions are mostly neglected (this would increase the parameter space in the model calculation enormously). However, it often provides a good starting point for a realistic model.

An important issue of simulation studies is the accessible time- and length-scale coverable by microscopic simulations. Fig.1 shows a schematic representation for different types of simulations in a *length-time-diagram*. It is clear that the more detailed a simulation technique operates, the smaller is the accessibility of long times and large length scales. Therefore quantum simulations, where fast motions of electrons are taken into account, are located in the lower left corner of the diagram and typical length and time

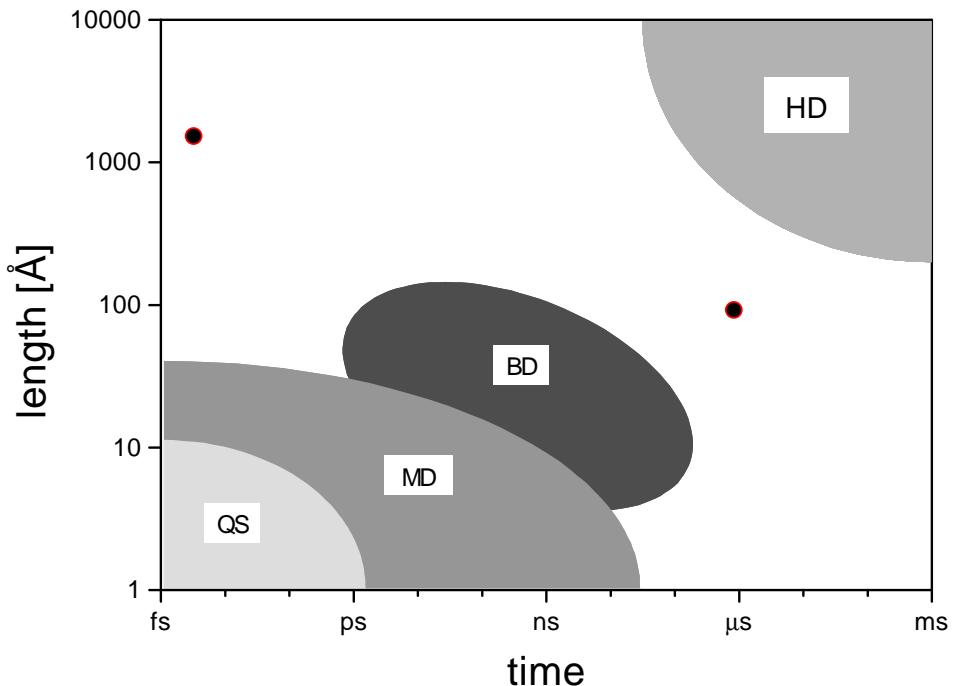


Figure 1. Schematic comparison of time- and length-scales, accessible to different types of simulation techniques (quantum simulations (QM), molecular dynamics (MD), Brownian dynamics (BD) and hydrodynamics/fluid dynamics (HD)). The black dots mark the longest ($\approx 1 \mu s$) and the biggest ($N > 5 \times 10^9$, $L \approx 0.4 \mu m$ molecular dynamics simulations by Duan & Kollman¹ and Roth² respectively).

scales are of order of Å and ps. Classical molecular dynamics approximates electronic distributions in a rather coarse-grained fashion by putting either fixed partial charges on interaction sites or by adding an approximate model for polarization effects. In both cases, the time scale of the system is not dominated by the motion of electrons, but the time of intermolecular collision events, rotational motions or intramolecular vibrations, which are orders of magnitude slower than those of electron motions. Consequently, the time step of integration is larger and trajectory lengths are of order ns and accessible lengths of order 10 – 100 Å. If one considers tracer particles in a solvent medium, where one is not interested in a detailed description of the solvent, one can apply Brownian dynamics, where the effect of the solvent is hidden in average quantities. Since collision times between tracer particles is very long, one may apply larger timesteps. Furthermore, since the solvent is not simulated explicitly, the lengthscales may be increased considerably. Finally, if one is interested not in a microscopic picture of the simulated system but in macroscopic quantities, the concepts of hydrodynamics may be applied, where the system properties are hidden in effective numbers, e.g. density, viscosity, sound velocity.

It is clear that the performance of particle dynamics simulations strongly depends on the computer facilities at hand. The first studies using MD simulation techniques were performed in 1957 by B. J. Alder and T. E. Wainright³ who simulated the phase transition

of a system of hard spheres. The general method, however, was presented two years later.⁴ In this early simulation, which was run on an IBM-704, up to 500 particles could be simulated, for which 500 collisions per hour could be calculated. Taking into account 200000 collisions for a production run, these simulations lasted for more than two weeks. The propagation of hard spheres in a simulation is determined by the collision events between two particles. Therefore, the propagation is not based on an integration of the equations of motion, but rather the calculation of the time of the next collision, which results in a variable time step in the calculations.

The first MD simulation which was applied to atoms interacting via a continuous potential was performed by A. Rahman in 1964. In this case, a model system for Argon was simulated and not only binary collisions were taken into account but the interactions were modeled by a Lennard-Jones potential and the equations of motion were integrated with a finite difference scheme. This work may be considered as seminal for dynamical calculations. It was the first work where an exact method (within numerical precision) was used to calculate dynamical quantities like autocorrelation functions and transport coefficients like the diffusion coefficient for a realistic system. Also more involved topics like the dynamic van Hove function and non-Gaussian corrections to diffusion were evaluated. The calculations were performed for 864 particles on a CDC 3600, where the propagation of all particles for one time step took $\approx 45\text{ s}$. The calculation of 50000 timesteps then took more than three weeks!^a

With the development of faster and bigger massively parallel architectures the accessible time and length scales are increasing. In the case of classical MD simulations it was demonstrated by J. Roth in 1999 on the CRAY T3E-1200 in Jülich that it is possible to simulate more than 5×10^9 particles, corresponding to a length scale of several 1000 Å. This was possible with the highly memory optimised MD program IMD,^{5,2} which used the 512 nodes with 256 MB memory each, quite efficiently. However, the limits of such a demonstration became rather obvious, since for a usual production run of 10000 time steps a simulation time of a quarter of a year would be required (given that the whole machine is dedicated to one user). In another demonstration run Y. Duan and P. A. Kollman extended the time scale of an all atom MD simulation to 1 μs , where they simulated the folding process of the subdomain HP-36 from the villin headpiece.^{6,1} The protein was modelled with a 596 interaction site model dissolved in a system of 3000 water molecules. Using a timestep of integration of $2 \times 10^{-15}\text{ s}$, the program was run for 5×10^8 steps. In order to perform this type of calculation, it was necessary to run the program several months on a CRAY T3D and CRAY T3E with 256 processors. It is clear that such kind of simulation is exceptional due to the large amount of computer resources needed, but is nonetheless a kind of milestone pointing to future simulation practices.

Classical molecular dynamics methods are nowadays applied to a huge class of problems, e.g. properties of liquids, defects in solids, fracture, surface properties, friction, molecular clusters, polyelectrolytes and biomolecules. Due to the large area of applicability, simulation codes for molecular dynamics were developed by many groups. On the internet homepage of the Collaborative Computational Project No.5 (CCP5)⁷ a lot of computer codes are assembled for condensed phase dynamics. During the last years several programs were designed for parallel computers. Among them, which are partly avail-

^aOn a standard PC this calculation may be done within one hour nowadays!

able free of charge, are, e.g., Amber/Sander,⁸ CHARMM,⁹ NAMD,¹⁰ NWChem¹¹ and LAMMPS.¹²

2 Models for Particle Interactions

A system is completely determined through its Hamiltonian $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, where \mathcal{H}_0 is the *internal* part of the Hamiltonian, given as

$$\mathcal{H}_0 = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + \sum_{i < j}^N u(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i < j}^N u^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots \quad (2)$$

where \mathbf{p} is the momentum, m the mass of the particles and u and $u^{(3)}$ are pair and three-body interaction potentials. \mathcal{H}_1 is an external part, which can include time dependent effects or external sources for a force. All simulated objects are defined within a model description. Often a precise knowledge of the interaction between atoms, molecules or surfaces are not known and the model is constructed in order to describe the main features of some observables. Besides boundary conditions, which are imposed, it is the model which completely determines the system from the physical point of view. In classical simulations the *objects* are most often described by point-like centers which interact through pair- or multibody interaction potentials. In that way the highly complex description of electron dynamics is abandoned and an effective picture is adopted where the main features like the hard core of a particle, electric multipoles or internal degrees of freedom of a molecules are modeled by a set of parameters and (most often) analytical functions which depend on the mutual position of particles in the configuration. Since the parameters and functions give a complete information of the system's energy as well as the force acting on each particle through $\mathbf{F} = -\nabla U$, the combination of parameters and functions is also called a *force field*. Different types of force field were developed during the last ten years. Among them are e.g. MM3,¹³ MM4,¹⁴ Dreiding,¹⁵ SHARP,¹⁶ VALBON,¹⁷ UFF,¹⁸ CFF95,¹⁹ AMBER²⁰ CHARMM,²¹ OPLS²² and MMFF.²³ Typical examples for force field functions are summarized in Fig. 2.

There are major differences to be noticed for the potential forms. The first distinction is to be made between pair- and multibody potentials. In systems with no constraints, the interaction is most often described by pair potentials, which is simple to implement into a program. In the case where multibody potentials come into play, the counting of interaction partners becomes increasingly more complex and dramatically slows down the execution of the program. Only for the case where interaction partners are known in advance, e.g. in the case of torsional or bending motions of a molecule can the interaction be calculated efficiently by using neighbor lists or by an intelligent way of indexing the molecular sites.

A second important difference between interactions is the spatial extent of the potential, classifying it into short and long range interactions. If the potential drops down to zero faster than r^{-d} , where r is the separation between two particles and d the dimension of the problem, it is called short ranged, otherwise it is long ranged. This becomes clear by considering the integral

$$I = \int \frac{dr^d}{r^n} = \begin{cases} \infty & : n \leq d \\ \text{finite} & : n > d \end{cases} \quad (3)$$

i.e. a particles' potential energy gets contributions from *all particles of the universe* if $n \leq d$, otherwise the interaction is bound to a certain region, which is often modeled by a spherical interaction range. The long range nature of the interaction becomes most important for potentials which only have potential parameters of the same sign, like the gravitational potential where no screening can occur. For Coulomb energies, where positive and negative charges may compensate each other, long range effects may be of minor importance in some systems like molten salts. In the following two examples shall illustrate the different treatment of short- and long range interactions.

2.1 Short Range Interactions

Short range interactions offer the possibility to take into account only neighbored particles up to a certain distance for the calculation of interactions. In that way a cutoff radius is introduced beyond of which mutual interactions between particles are neglected. As an approximation one may introduce *long range corrections* to the potential in order to compensate for the neglect of explicit calculations. The whole short range potential may then be written as

$$U = \sum_{i < j}^N u(r_{ij} | r_{ij} < R_c) + U_{lrc} \quad (4)$$

The long-range correction is thereby given as

$$U_{lrc} = 2\pi N \rho_0 \int_{R_c}^{\infty} dr r^2 g(r) u(r) \quad (5)$$

where ρ_0 is the number density of the particles in the system and $g(r) = \rho(r)/\rho_0$ is the radial distribution function. For computational reasons, $g(r)$ is most often only calculated up to R_c , so that in practice it is assumed that $g(r) = 1$ for $r > R_c$, which makes it possible for many types of potentials to calculate U_{lrc} analytically.

Besides internal degrees of freedom of molecules, which may be modeled with short range interaction potentials (c.f. Fig.2), it is first of all the excluded volume of a particle which is of importance. A finite diameter of a particle may be represented by a steep repulsive potential acting at short distances. This is either described by an exponential function or an algebraic form, $\propto r^{-n}$, where $n \geq 9$. Another source of short range interaction is the van der Waals interaction. For neutral particles these are the London forces arising from induced dipole interactions. Fluctuations of the electron distribution of a particle give rise to fluctuating dipole moments, which on average compensate to zero. But the instantaneous created dipoles induce also dipoles on neighbored particles which attract each other $\propto r^{-6}$. Two common forms of the resulting interactions are the Buckingham potential

$$u_{\alpha\beta}^B(r_{ij}) = A_{\alpha\beta} e^{-B_{\alpha\beta} r_{ij}} - \frac{D_{\alpha\beta}}{r_{ij}^6} \quad (6)$$

and the Lennard-Jones potential

$$u_{\alpha\beta}^{LJ}(r_{ij}) = 4\epsilon_{\alpha\beta} \left(\left(\frac{\sigma_{\alpha\beta}}{r_{ij}} \right)^1 2 - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}} \right)^6 \right) \quad (7)$$

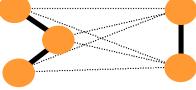
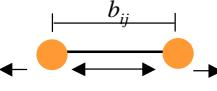
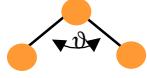
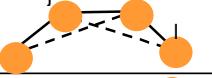
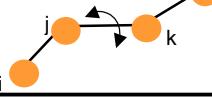
6-9 van derWaals		$u_{ij}(r_{ij}) = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^9 - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$
6-12 van derWaals		$u_{ij}(r_{ij}) = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$
Electrostatic		$u_{ij}(r_{ij}) = \frac{q_i q_j}{r_{ij}}$
Quadratic bond stretching		$u_{ij}^s(r_{ij}) = \frac{1}{2} k_{ij} (r_{ij} - b_{ij})^2$
Morse bond stretching		$u_{ij}(r_{ij}) = k \left(1 - e^{-a(r_{ij} - r_0)} \right)^2$
Bond-bending		$u_{ij}^b(\vartheta_{ijk}) = \frac{1}{2} k_{ijk} (\vartheta_{ijk} - \vartheta_{ijk}^0)^2$
Improper dihedrals		$u_{ij}^{id}(\xi_{ijkl}) = \frac{1}{2} k_{ijkl} (\xi_{ijkl} - \xi_0)^2$
Proper dihedrals		$u_{ij}^{pd}(\varphi_{ijkl}) = k_\varphi (1 + \cos(n\varphi_{ijkl} - \varphi_0))$

Figure 2. Typical examples for potential terms as used in common force-fields.

which are compared in Fig.3. In Eqs.6,7 the indices α, β indicate the species of the particles, i.e. there are parameters A, B, D in Eq.6 and ϵ, σ in Eq.7 for intra-species interactions ($\alpha = \beta$) and cross species interactions ($\alpha \neq \beta$). For the Lennard-Jones potential the parameters have a simple physical interpretation: ϵ is the minimum potential energy, located at $r = 2^{1/6}\sigma$ and σ is the diameter of the particle, since for $r < \sigma$ the potential becomes repulsive. Often the Lennard-Jones potential gives a reasonable approximation of a *true* potential. However, from exact quantum ab initio calculations an exponential type repulsive potential is often more appropriate. Especially for dense systems the too steep repulsive part often leads to an overestimation of the pressure in the system. Since computationally the Lennard-Jones interaction is quite attractive the repulsive part is sometimes replaced by a weaker repulsive term, like $\propto r^{-9}$. The Lennard-Jones potential has another advantage over the Buckingham potential, since there are combining rules for the parameters. A common choice are the Lorentz-Berelot combining rules

$$\sigma_{\alpha\beta} = \frac{\sigma_{\alpha\alpha} + \sigma_{\beta\beta}}{2} \quad , \quad \epsilon_{\alpha\beta} = \sqrt{\epsilon_{\alpha\alpha} \epsilon_{\beta\beta}} \quad (8)$$

This combining rule is, however, known to overestimate the well depth parameter. Two

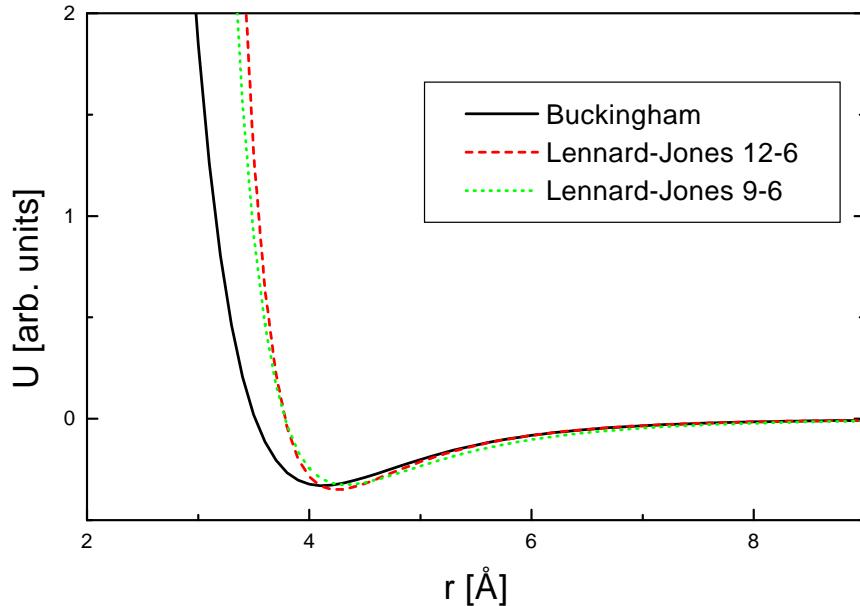


Figure 3. Comparison between a Buckingham-, Lennard-Jones (12-6) and Lennard-Jones (9-6) potential.

other commonly known combining rules try to correct this effect, which are Kong²⁴ rules

$$\sigma_{\alpha\beta} = \left[\frac{1}{2^{13}} \frac{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^{12}}{\sqrt{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^6\epsilon_{\beta\beta}\sigma_{\beta\beta}^6}} \left(1 + \left(\frac{\epsilon_{\beta\beta}\sigma_{\beta\beta}^{12}}{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^{12}} \right)^{1/13} \right)^{13} \right]^{1/6} \quad (9)$$

$$\epsilon_{\alpha\beta} = \frac{\sqrt{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^6\epsilon_{\beta\beta}\sigma_{\beta\beta}^6}}{\sigma_{\alpha\beta}^6} \quad (10)$$

and the Waldman-Kagler²⁵ rule

$$\sigma_{\alpha\beta} = \left(\frac{\sigma_{\alpha\alpha}^6 + \sigma_{\beta\beta}^6}{2} \right)^{1/6}, \quad \epsilon_{\alpha\beta} = \frac{\sqrt{\epsilon_{\alpha\alpha}\sigma_{\alpha\alpha}^6\epsilon_{\beta\beta}\sigma_{\beta\beta}^6}}{\sigma_{\alpha\beta}^6} \quad (11)$$

In a recent study²⁶ of Ar-Kr and Ar-Ne mixtures, these combining rules were tested and it was found that the Kong rules give the best agreement between simulated and experimental pressure-density curves. An illustration of the different combining rules is shown in Fig.4 for the case of an Ar-Ne mixture.

2.2 Long Range Interactions

In the case of long range potentials, like the Coulomb potential, interactions between all particles in the system must be taken into account, if treated without any approximation.

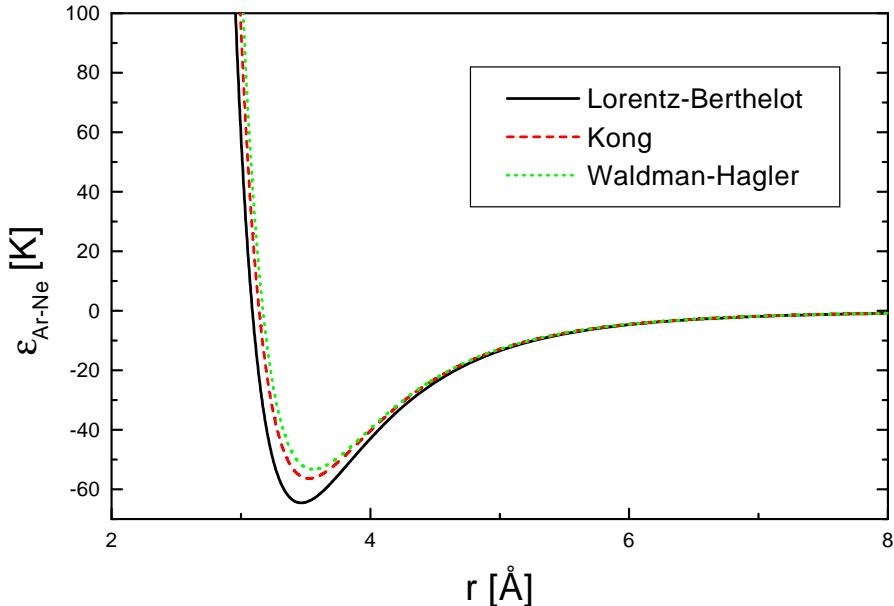


Figure 4. Resulting cross-terms of the Lennard-Jones potential for an Ar-Ne mixture. Shown is the effect of different combining rules (Eqs.8-11). Parameters used are $\sigma_{Ar} = 3.406 \text{ \AA}$, $\epsilon_{Ar} = 119.4 \text{ K}$ and $\sigma_{Ne} = 2.75 \text{ \AA}$, $\epsilon_{Ne} = 35.7 \text{ K}$.

This leads to an $\mathcal{O}(N^2)$ problem, which increases considerably the execution time of a program for larger systems. For systems with open boundary conditions (like liquid droplets), this method is straightforwardly implemented and reduces to a double sum over all pairs of particles. In the case when periodic boundary conditions are applied, not only the interactions with particles in the *central cell* but also with all periodic images must be taken into account and formally a lattice sum has to be evaluated

$$U = \frac{1}{2} \sum_{i,j=1}^N \sum'_{\mathbf{n}} \frac{q_i q_j}{|\mathbf{r}_{ij} - \mathbf{n}L|} \quad (12)$$

where \mathbf{n} is a lattice vector and $\sum'_{\mathbf{n}}$ means that for $\mathbf{n} = 0$ it is $i \neq j$. It is, however, a well known problem that this type of lattice sum is conditionally convergent, i.e. the result depends on the sequence of evaluating the sum (see e.g.²⁷). A method to overcome this limitation was invented by Ewald.²⁸ The idea is to introduce a convergence factor into the sum of Eq.12 which depends on a parameter s , evaluate the sum and put $s \rightarrow 0$ in the end. A characterization of the convergence factors was given in Ref.^{29,30} A form which leads to the Ewald sum is an exponential $e^{-s\mathbf{n}^2}$, transforming Eq.12 into

$$U(s) = \frac{1}{2} \sum_{i,j=1}^N \sum'_{\mathbf{n}} \frac{q_i q_j}{|\mathbf{r}_{ij} - \mathbf{n}L|} e^{-s\mathbf{n}^2} \quad (13)$$

The evaluation and manipulation of this equation proceeds now by using the definition for the Γ -function and the Jacobi imaginary transform. A very instructive way of the derivation

of the Ewald sum may be found in Ref.,^{29,30} a heuristic derivation is given Ref.³¹ For brevity, only the final form of the sum is given here

$$U = \frac{1}{2} \left\{ \underbrace{\sum_{i,j=1}^N \sum_{\mathbf{n}} \frac{q_i q_j \operatorname{erfc}(\alpha |\mathbf{r}_{ij} - \mathbf{n}L|/L)}{|\mathbf{r}_{ij} - \mathbf{n}L|}}_{U_{real}} + \underbrace{\frac{4\pi q_i q_j}{L^3} \sum_{\mathbf{k}} \frac{1}{k^2} e^{i\mathbf{k}\cdot\mathbf{r}_{ij}} e^{-k^2/4\alpha^2}}_{U_{reciprocal}} \right. \\ \left. + \underbrace{\frac{1}{L} \left[\sum_{\mathbf{n} \neq 0} \frac{\operatorname{erfc}(\alpha \mathbf{n})}{|\mathbf{n}|} + \frac{e^{-\pi^2 \mathbf{n}^2}/\alpha^2}{\pi \mathbf{n}^2} - \frac{2\alpha}{\sqrt{\pi}} \right] \sum_{i=1}^N q_i^2 + \underbrace{\frac{4\pi}{L^3} \left| \sum_{i=1}^N q_i \right|^2}_{U_{surface}} \right\} \quad (14)$$

The evaluation of the potential thus splits into four different terms, where the so called *self*- and *surface-terms* are constant and may be calculated in the beginning of a simulation. The first two sums, however, depend on the interparticle separations \mathbf{r}_{ij} , which need to be evaluated in each time step. It is seen that the lattice sum is essentially split into a sum which is evaluated in real space and a sum over reciprocal space-vectors, $\mathbf{k} = 2\pi\mathbf{n}/L$. The parameter α appears formally in the derivation as a result of an integral splitting but it has a very intuitive physical meaning. The first sum gives the potential of a set of point charges which are screened by an opposite charge of the same magnitude but with a Gaussian form factor where the width of the Gaussian is given by α . The second sum subtracts this screening charge, but the sum is evaluated in reciprocal space. Since $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ decays as e^{-x^2} for large x , the first sum contains mainly short range contributions. On the other side, the second sum decays strongly for large k -vectors and thus contains mainly long range contributions. Most often, the parameter α is chosen in way to reduce the evaluation of the real space sum to the central simulation cell. Often, a spherical cutoff is then applied for this term, i.e. contributions of particle pairs, separated in a distance $|\mathbf{r}_{ij}| > R_c$ are neglected. On the other hand, the reciprocal space sum is conventionally truncated after a maximum wave-vector \mathbf{k}_{max} . All three parameters $\alpha, R_c, \mathbf{k}_{max}$ may be chosen in an optimal way to balance the truncation error in each sum and the number of operations. This balancing even leads to the effect that the Ewald sum may be tuned^{32,33} to scale with $\mathcal{O}(N^{3/2})$ (for fast methods which have better scaling characteristics, see Ref.³¹). A detailed analysis of the individual errors occurring in the different sums was given in Ref.³⁴ An alternative derivation of the Ewald sum starts directly by assuming a Gaussian form factor for the screening charge. This gives the opportunity to investigate also form factors, differing from a Gaussian. In these cases the convergence function is in general not known but it is assumed to exist. Different form factors were studied systematically in Ref.³⁵

The present form of the Ewald sum gives an exact representation of the potential energy of point like charges in a system with periodic boundary conditions. Sometimes the charge distribution in a molecule is approximated by a point dipole or higher multipole moments. A more general form of the Ewald sum, taking into account arbitrary point multipoles was given in Ref.³⁶ The case, where also electronic polarizabilities are considered is given in Ref.³⁷

In certain systems, like in molten salts or electrolyte solutions, the interaction between charged species may approximated by a screened Coulomb potential, which has a Yukawa-

like form

$$U = \frac{1}{2} \sum_{i,j=1}^N q_i q_j \frac{e^{-\kappa |\mathbf{r}_{ij}|}}{|\mathbf{r}_{ij}|} \quad (15)$$

The parameter κ is the inverse Debye length, which gives a measure of screening strength in the system. If $\kappa < 1/L$ the potential is short ranged and usual cut-off methods may be used. Instead, if $\kappa > 1/L$, or generally if $u(r = L/2)$ is larger than the prescribed uncertainties in the energy, the minimum image convention in combination with truncation methods fails and the potential must be treated in a more rigorous way, which was proposed in Ref.,³⁸ where an extension of the Ewald sum for such Yukawa type potentials was developed.

3 The Integrator

For a given potential model which characterizes the physical system, it is the integrator which is responsible for the accuracy of the simulation results. If the integrator would work without any error the simulation would provide *exact model results* within the errors occurring due to a finite number representation. However, any finite difference integrator is naturally an approximation for a system developing continuously in time. The requirements for the integrator are therefore to be

- accurate, in the sense that it approximates the *true* trajectory very well (this may be checked with simple model systems for which analytical solutions exist)
- stable, in the sense that it conserves energy and that small perturbations do not lead to instabilities
- robust, in the sense that it allows for large time steps in order to propagate the system efficiently through phase space

In the following different types of integration schemes are presented. First, simple integrators based on Taylor expansions are shown. Later on integrators based on an operator splitting method are discussed which provide the possibility to introduce in an elegant way the integrations of motion on different time scales. Finally, attention is given to more complex situations where molecules with orientational degrees of freedom are considered.

3.1 Expansion Based Integrators

The simplest and most straight forward way to construct an integrator is by expanding the positions and velocities in a Taylor series. The class of integrators which may be obtained in that way are called Verlet-Störmer integrators.^{39,40} For a small enough time step δt the expansion gives

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t) \delta t + \frac{1}{2} \mathbf{a}(t) \delta t^2 + \frac{1}{6} \mathbf{b}(t) \delta t^3 + \dots \quad (16)$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \mathbf{a}(t) \delta t + \frac{1}{2} \mathbf{b}(t) \delta t^2 + \frac{1}{6} \mathbf{c}(t) \delta t^3 + \dots \quad (17)$$

where \mathbf{a} , \mathbf{b} , \mathbf{c} are the 2nd, 3rd and 4th time derivative of the coordinates. In the same way the expansion may be performed for $\delta t \rightarrow -\delta t$, which gives

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \mathbf{v}(t) \delta t + \frac{1}{2} \mathbf{a}(t) \delta t^2 - \frac{1}{6} \mathbf{b}(t) \delta t^3 \pm \dots \quad (18)$$

$$\mathbf{v}(t - \delta t) = \mathbf{v}(t) - \mathbf{a}(t) \delta t + \frac{1}{2} \mathbf{b}(t) \delta t^2 - \frac{1}{6} \mathbf{c}(t) \delta t^3 \pm \dots \quad (19)$$

Adding up Eqs.16,18 and Eqs.17,19 gives for the new positions and velocities

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \mathbf{a}(t) \delta t^2 + \mathcal{O}(\delta t^4) \quad (20)$$

$$\mathbf{v}(t + \delta t) = 2\mathbf{v}(t) - \mathbf{v}(t - \delta t) + \mathbf{b}(t) \delta t^2 + \mathcal{O}(\delta t^4) \quad (21)$$

A method whose truncation varies as $\delta t^{(n+1)}$ is called an n-th order method. Eqs.20,21 are therefore of 3rd order. The drawback of Eq.21 is, however, that it requires the 3rd derivative of the coordinates with respect with to time which is not routinely calculated in MD simulations and thus introduces some additional computational and storage overhead. To overcome this drawback one can simply subtract Eq.18 from Eq.16, giving the central difference scheme for the velocity

$$\mathbf{v}(t) = \frac{1}{2\delta t} (\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)) + \mathcal{O}(\delta t^3) \quad (22)$$

This is, however, one order less in accuracy than Eq.21 and also provides velocities at timestep t , not at $t + \delta t$. Since this information is not required by Eq.20 to calculate accurately the positions, one may take Eq.22 as an estimate for the velocities from which the kinetic energy of the system is calculated.

From the point of view of storage requirements, Eqs.20,22 are not optimal, since information is required from positions not only at time t but also at time $t - \delta t$. An equivalent algorithm, which stores only information from one timestep is the so called *velocity Verlet* algorithm, which reads

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t) \delta t + \frac{1}{2} \mathbf{a}(t) \delta t^2 \quad (23)$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \frac{1}{2} \delta t (\mathbf{a}(t) + \mathbf{a}(t + \delta t)) \quad (24)$$

This scheme, however, requires the knowledge of the accelerations, \mathbf{a} , at timestep $t + \delta t$. One may therefore decompose Eq.24 into two steps. First calculate

$$\mathbf{v}(t + \delta t/2) = \mathbf{v}(t) + \frac{1}{2} \delta t \mathbf{a}(t) \quad (25)$$

then compute the actual forces on the particles at time $t + \delta t$ and finish the velocity calculation with

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t + \delta t/2) + \frac{1}{2} \mathbf{a}(t + \delta t) \delta t \quad (26)$$

At this point the kinetic energy may be calculated without a time delay of δt , as it was in Eq.22. Several other schemes have been proposed in literature, such as the leap-frog⁴¹

scheme or Beeman's⁴² algorithm. They all have the same accuracy and should produce identical trajectories in coordinate space ^b

3.2 Operator Splitting Methods

A more rigorous derivation, which in addition leads to the possibility of splitting the propagator of the phase space trajectory into several time scales, is based on the phase space description of a classical system. The time evolution of a point in the $6N$ dimensional phase space is given by the Liouville equation

$$\Gamma(t) = e^{i\mathcal{L}t} \Gamma(0) \quad (27)$$

where $\Gamma = (\mathbf{q}, \mathbf{p})$ is the $6N$ dimensional vector of generalized coordinates, $\mathbf{q} = \mathbf{q}_1, \dots, \mathbf{q}_N$, and momenta, $\mathbf{p} = \mathbf{p}_1, \dots, \mathbf{p}_N$. The Liouville operator, \mathcal{L} , is defined as

$$i\mathcal{L} = \{\dots, \mathcal{H}\} = \sum_{j=1}^N \left(\frac{\partial \mathbf{q}_j}{\partial t} \frac{\partial}{\partial \mathbf{q}_j} + \frac{\partial \mathbf{p}_j}{\partial t} \frac{\partial}{\partial \mathbf{p}_j} \right) \quad (28)$$

In order to construct a discrete timestep integrator, the Liouville operator is split into two parts, $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$, and a Trotter expansion⁴³ is performed

$$e^{i\mathcal{L}\delta t} = e^{i(\mathcal{L}_1 + \mathcal{L}_2)\delta t} \quad (29)$$

$$= e^{i\mathcal{L}_1\delta t/2} e^{i\mathcal{L}_2\delta t} e^{i\mathcal{L}_1\delta t/2} + \mathcal{O}(\delta t^3) \quad (30)$$

The partial operators can be chosen to act only on positions and momenta. Assuming usual cartesian coordinates for a system of N free particles, this can be written as

$$i\mathcal{L}_1 = \sum_{j=1}^N \mathbf{F}_j \frac{\partial}{\partial \mathbf{p}_j} \quad (31)$$

$$i\mathcal{L}_2 = \sum_{j=1}^N \mathbf{v}_j \frac{\partial}{\partial \mathbf{r}_j} \quad (32)$$

Applying Eq.29 to the phase space vector Γ and using the property $e^{a\partial/\partial x} f(x) = f(x+a)$ for any function f , where a is independent of x , gives

$$\mathbf{v}_i(t + \delta t/2) = \mathbf{v}(t) + \frac{\mathbf{F}_i(t)}{m_i} \frac{\delta t}{2} \quad (33)$$

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \delta t/2)\delta t \quad (34)$$

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}_i(t + \delta t/2) + \frac{\mathbf{F}_i(t + \delta t)}{m_i} \frac{\delta t}{2} \quad (35)$$

which is the velocity Verlet algorithm, Eqs.23,25,26.

^bThis statement is derived from the point of view of accuracy. Since numerical operations are in general not associative a different implementation of an algorithm will have different round off errors and therefore the accumulation of the roundoff error will accumulate which will lead in practice to a deviation from the above statement.

In the same spirit, another algorithm may be derived by simply changing the definitions for $\mathcal{L}_1 \rightarrow \mathcal{L}_2$ and $\mathcal{L}_2 \rightarrow \mathcal{L}_1$. This gives the so called *position Verlet algorithm*

$$\mathbf{r}_i(t + \delta t/2) = \mathbf{r}_i(t) + \mathbf{v}(t) \frac{\delta t}{2} \quad (36)$$

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}(t) + \frac{\mathbf{F}_i(t + \delta t/2)}{m_i} \quad (37)$$

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t + \delta t/2) + (\mathbf{v}(t) + \mathbf{v}_i(t + \delta t)) \frac{\delta t}{2} \quad (38)$$

Here the forces are calculated at intermediate positions $\mathbf{r}_i(t + \delta t/2)$. The equations of both the velocity Verlet and the position Verlet algorithms have the property of propagating velocities or positions on half time steps. Since both schemes decouple into an applied force term and a *free flight* term, the three steps are often called *half-kick/drift/half kick* for the velocity Verlet and correspondingly *half-drift/kick/half-drift* for the position Verlet algorithm.

Both algorithms, the velocity and the position Verlet method, are examples for symplectic algorithms, which are characterized by a volume conservation in phase space. This is equivalent to the fact that the Jacobian matrix of a transform $x' = f(x, p)$ and $p' = g(x, p)$ satisfies

$$\begin{pmatrix} f_x & f_p \\ g_x & g_p \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} f_x & f_p \\ g_x & g_p \end{pmatrix} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \quad (39)$$

Any method which is based on the splitting of the Hamiltonian, is symplectic. This does not yet, however, guarantee that the method is also time reversible, which may be also be considered as a strong requirement for the integrator. This property is guaranteed by symmetric methods, which also provide a better numerical stability.⁴⁴ Methods, which try to enhance the accuracy by taking into account the particles' history (multi-step methods) tend to be incompatible with symplecticness,^{45,46} which makes symplectic schemes attractive from the point of view of data storage requirements. Another strong argument for symplectic schemes is the so called *backward error analysis*.^{47–49} This means that the trajectory produced by a discrete integration scheme, may be expressed as the solution of a perturbed ordinary differential equation whose *rhs* can formally be expressed as a power series in δt . It could be shown that the system, described by the ordinary differential equation is Hamiltonian, if the integrator is symplectic.^{50,51} In general, the power series in δt diverges. However, if the series is truncated, the trajectory will differ only as $\mathcal{O}(\delta t^p)$ of the trajectory, generated by the symplectic integrator on timescales $\mathcal{O}(1/\delta t)$.⁵²

3.3 Multiple Time Step Methods

It was already mentioned that the rigorous approach of the decomposition of the Liouville operator offers the opportunity for a decomposition of time scales in the system. Supposing that there are different time scales present in the system, e.g. fast intramolecular vibrations and slow domain motions of molecules, then the factorization of Eq.29 may be written in

a more general way

$$e^{i\mathcal{L}\Delta t} = e^{i\mathcal{L}_1^{(s)}\Delta t/2} e^{i\mathcal{L}_1^{(f)}\Delta t/2} e^{i\mathcal{L}_2\delta t} e^{i\mathcal{L}_1^{(f)}\Delta t/2} e^{i\mathcal{L}_1^{(s)}\Delta t/2} \quad (40)$$

$$= e^{i\mathcal{L}_1^{(s)}\Delta t/2} \left\{ e^{i\mathcal{L}_1^{(f)}\delta t/2} e^{i\mathcal{L}_2\delta t} e^{i\mathcal{L}_1^{(f)}\delta t/2} \right\}^p e^{i\mathcal{L}_1^{(s)}\Delta t/2} \quad (41)$$

where the time increment is $\Delta t = p\delta$. The decomposition of the Liouville operator may be chosen in the convenient way

$$i\mathcal{L}_1^{(s)} = \mathbf{F}_i^{(s)} \frac{\partial}{\partial \mathbf{p}_i} \quad , \quad i\mathcal{L}_1^{(f)} = \mathbf{F}_i^{(f)} \frac{\partial}{\partial \mathbf{p}_i} \quad , \quad i\mathcal{L}_2 = \mathbf{v}_i \frac{\partial}{\partial \mathbf{q}_i} \quad (42)$$

where the superscript (s) and (f) mean slow and fast contributions to the forces. The idea behind this decomposition is simply to take into account contributions from slowly varying components only every p 'th timestep with a large time interval. Therefore, the force computation may be considerably speeded up in the the $p - 1$ intermediate force computation steps. In general, the scheme may be extended to account for more time scales. Examples for this may be found in Refs.^{53–55} One obvious problem, however, is to separate the timescales in a proper way. The scheme of Eq.41 is *exact* if the time scales decouple completely. This, however, is very rarely found and most often timescales are coupled due to nonlinear effects. Nevertheless, for the case where Δt is not very much larger than δt ($p \approx 10$), the separation may be often justified and lead to stable results. Another criteria for the separation is to distinguish between long range and short range contributions to the force. Since the magnitude and the fluctuation frequency is very much larger for the short range contributions this separation makes sense for speeding up computations including long range interactions.⁵⁶

The method has, however, its limitations.^{57,58} As described, a particle gets every n 'th timestep a *kick* due to the slow components. It was reported in literature that this may excite a system's resonance which will lead to strong artifacts or even instabilities.^{59,60} Recently different schemes were proposed to overcome these resonances by keeping the property of symplecticness.^{61–67}

3.4 Constraint Dynamics

The methods discussed so far are quite general for the cases of free particles. If constraints are applied, e.g. a fixed bond length between particles in a molecule or a fixed bond angle, the integration scheme has either to be modified or extended. A modification of the integrator means that the equations of motion have to be formulated for rotational and translational degrees of freedoms. On the other hand an extension of the integrator means that constraints have to be taken into account when moving an individual particle via the integration scheme. The first method is mainly applied to molecules which are modeled as rigid body, i.e. the motion may easily be described as the translation of the center of mass and a rotation around its principle axis of inertia. In the case of large molecules, where not all bond lengths and angles are fixed and which exhibit internal degrees of freedom such as side chain motions or rotation of atomic groups, constraint methods have to be applied. In the following the motion of rigid bodies and the constraint dynamics will be described.

3.4.1 Rigid Body Motion

The natural way of describing a rigid body is to specify the coordinates and the moment of the center of mass as well as the orientation with respect to a space fixed coordinate system and the angular velocity around the principle molecular axis. The translational motion is thereby described by the total force acting on the molecule and the integration schemes, described earlier may be applied. The rotational motion requires a description of the orientation of the molecule and a calculation of the torque. As a first choice for the orientational description one could use the Euler angles $(\varphi, \vartheta, \psi)$ to build up the rotation matrix. However, a numerical problem appears with solving the equations of motions

$$\frac{\partial \varphi}{\partial t} = -\omega_x \frac{\sin \varphi \cos \vartheta}{\sin \vartheta} + \omega_y \frac{\cos \varphi \cos \vartheta}{\sin \vartheta} + \omega_z \quad (43)$$

$$\frac{\partial \vartheta}{\partial t} = \omega_x \cos \varphi + \omega_y \sin \varphi \quad (44)$$

$$\frac{\partial \psi}{\partial t} = \omega_x \frac{\sin \varphi}{\sin \vartheta} - \omega_y \frac{\cos \varphi}{\sin \vartheta} \quad (45)$$

It is obvious that for the case, when ϑ is close or equal zero, the terms in $\partial_t \varphi$ and $\partial_t \psi$ diverge and lead to numerical instabilities. Since the orientation where $\vartheta = 0$ is physically not a special case but only related to the special convention of a chosen coordinate system, one can in principle switch the description into another coordinate system when ϑ approaches zero.⁶⁸ This is, however, not very efficient since variables have to be calculated and stored in different coordinate systems at the same time.

An elegant method which avoids these problems is the orientational description in terms of quaternions, $\mathbf{q} = (q_1, \dots, q_4)$, originally introduced by Hamilton in order to extend the complex numbers.⁶⁹ They are defined by algebraic relations and have the property $\sum_i q_i^2 = 1$. Also, they can be expressed in terms of Eulerangles

$$q_1 = \cos \frac{\vartheta}{2} \cos \frac{\varphi + \psi}{2} \quad (46)$$

$$q_2 = \sin \frac{\vartheta}{2} \cos \frac{\varphi - \psi}{2} \quad (47)$$

$$q_3 = \sin \frac{\vartheta}{2} \sin \frac{\varphi - \psi}{2} \quad (48)$$

$$q_4 = \cos \frac{\vartheta}{2} \sin \frac{\varphi + \psi}{2} \quad (49)$$

so that they can completely describe the orientation of a fixed body in space. The equations of motion for \mathbf{q} are given by

$$\frac{\partial \mathbf{q}}{\partial t} = \frac{1}{2} \mathbf{Q} \boldsymbol{\omega}^b \quad (50)$$

where

$$\mathbf{Q} = \begin{pmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & -q_4 & q_3 \\ q_3 & q_4 & q_1 & -q_2 \\ q_4 & -q_3 & q_2 & q_1 \end{pmatrix} \quad (51)$$

and

$$(\boldsymbol{\omega}^b)^T = (0, \omega_x^b, \omega_y^b, \omega_z^b) \quad (52)$$

where the superscript b denotes that the angular velocities are evaluated in the body fixed frame. From Eq.50 it becomes obvious that the divergence problems have disappeared. The transformation between a fixed body frame ($\hat{\mathbf{x}}$) and a space fixed description (\mathbf{x}) is then provided by $\hat{\mathbf{x}} = \mathbf{R}\mathbf{x}$, where the rotation matrix \mathbf{R} is given by

$$\mathbf{R} = \begin{pmatrix} q_1^2 + q_2^2 - q_3^2 - q_4^2 & 2(q_2q_3 + q_1q_4) & 2(q_2q_4 - q_1q_3) \\ 2(q_2q_3 - q_1q_4) & q_1^2 - q_2^2 + q_3^2 - q_4^2 & 2(q_3q_4 + q_1q_2) \\ 2(q_2q_4 + q_1q_3) & 2(q_3q_4 - q_1q_2) & q_1^2 - q_2^2 - q_3^2 + q_4^2 \end{pmatrix} \quad (53)$$

The integration scheme for the rotational part is more involved than it is for its translational counterpart. Since the calculation of angular momenta is most easily done in the fixed body (molecular) frame, where the moment of inertia tensor is diagonal, a transformation from the space fixed (laboratory) frame to the molecular frame is required.

In the following a leap-frog *like* scheme is described, which uses information of half-step results.⁷⁰ In a first step the torque on the molecule is calculated by $\mathbf{T}_i = \sum_{\alpha} \mathbf{d}_{\alpha} \times \mathbf{F}_{\alpha}$, where α denotes the molecular sites and $\mathbf{d}_{\alpha} = \mathbf{R}(\mathbf{Q})\hat{\mathbf{d}}_{\alpha}$ is the vector pointing from the center of mass of the molecule to site α , written in the laboratory frame. Having the torque, the angular momentum \mathbf{j} and the angular velocity $\boldsymbol{\omega}$ in the molecular frame can be obtained via

$$\mathbf{j}_i(t) = \mathbf{j}_i(t - \delta t/2) + \frac{1}{2}\mathbf{T}_i\delta t \quad (54)$$

$$\hat{\boldsymbol{\omega}}_i = \hat{\mathbf{I}}^{-1}\mathbf{R}^T(\mathbf{Q}_i)\mathbf{j}_i(t) \quad (55)$$

where $\hat{\mathbf{I}}^{-1}$ is the inverse of the diagonal moment of inertia tensor. A similar step is performed to calculate the quaternions at time $t + \delta t/2$

$$\mathbf{q}_i(t + \delta t/2) = \mathbf{q}_i(t) + \frac{\delta t}{2}\mathbf{Q}(\mathbf{q}_i(t))\hat{\boldsymbol{\omega}}_i \quad (56)$$

In the next step the angular momenta are propagated from where the angular velocity can be obtained, in order to complete the intergation step

$$\hat{\mathbf{j}}_i(t + \delta t/2) = \mathbf{R}^T(\mathbf{Q}_i(t + \delta t/2))(\mathbf{j}_i(t - \delta t/2) + \delta t\mathbf{T}_i(t)) \quad (57)$$

$$\hat{\boldsymbol{\omega}}_i(t + \delta t/2) = \hat{\mathbf{I}}^{-1}\hat{\mathbf{j}}_i(t + \delta t/2) \quad (58)$$

$$\mathbf{q}_i(t + \delta t) = \mathbf{q}_i(t) + \frac{\delta t}{2}\mathbf{Q}_i(t + \delta t/2)\hat{\boldsymbol{\omega}}_i(t + \delta t/2) \quad (59)$$

Examples for rigid-body algorithms, based on a splitting method, which conserve the symplectic structure and are time reversible can be found in Refs.^{71,72} The method presented in Ref.⁷¹ is implemented in the downloadable research program ORIENT.⁷³

3.4.2 Constrained Motion

Describing large molecules as rigid bodies is often a poor approximation. Especially, if conformational changes of a molecule are expected, it is not possible to freeze all internal degrees of freedom. However, if one is not interested in the high frequency intramolecular vibrational motions, which require a rather small timestep of integration (and therefore will slow down a simulation considerably), one can fix the bond lengths between neighbored sites and allow for bending and dihedral motions. The control of the bond length can be achieved by introducing Lagrangian multipliers. If the constraints are formulated as holonomic constraints the equations of motion are modified according to

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} \quad (60)$$

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial U(\mathbf{q})}{\partial \mathbf{q}_i} + \mathbf{g}'(\mathbf{q})\lambda \quad (61)$$

$$g(\mathbf{q}) = 0 \quad (62)$$

The *advantage* of this method is that the atoms of the molecule may be treated individually with a simple integrator scheme. No transformation from a laboratory to molecular frame has to be performed. Also the integration of angular degrees of freedom and the calculation of torques is not required. Solving for the Lagrange multiplier, generally leads to a diagonalization of an $P \times P$ matrix, where P is the number of constraints. However, since constraints are applied most often only to nearby atoms the matrix is sparse and fast methods may be applied.⁷⁴ An alternative to this direct method is to fulfill the constraints in an iterative way one by one up to a given precision. First an unconstrained motion of the atoms of a molecule is performed which leads in general to positions which do not satisfy the constraints. The correction of the positions is then achieved via

$$\mathbf{r}_i(t + \delta t) \rightarrow \mathbf{r}_i(t + \delta t) + \frac{\delta t^2}{2m_i} \sum_{\gamma} \mathbf{F}_{\gamma}^c \quad (63)$$

where γ runs over all constraints and the constraint forces are given by

$$\mathbf{F}^c = \frac{\mu}{2\delta t^2} \frac{(\mathbf{d}_0^2 - \mathbf{d}^2)}{|\mathbf{d}_0 \mathbf{d}|} \mathbf{d}_0 \quad (64)$$

where \mathbf{d}_0 is the constrained bond vector at the start of the integration step and \mathbf{d} is the bond vector after the unconstrained integration step, $\mu = m_i m_j / (m_i + m_j)$ is the reduced mass of the atom pair i, j . For a molecule with multiple constraints, Eq.63 is a first order correction , which can be applied in an iterative way up to a given precision, $|\mathbf{d}_0 - \mathbf{d}| / |\mathbf{d}_0| < 10^{-k}$, where k is the desired precision which is often chosen as $k \geq 4$ in order to conserve energy. This method is used in the algorithms SHAKE,⁷⁵ invented by Ryckaert et al. and in RATTLE⁷⁶ invented by Andersen. The latter one was proven to

be symplectic and time reversible.⁵⁹ Refinements of the SHAKE algorithm were proposed in Ref.⁷⁷ Schemes, where rotations of linked bodies are taken into account by quaternion methods, were proposed in Refs.^{78,79}

4 Simulating in Different Ensembles

In MD simulations it is possible to realize different types of thermodynamic ensembles which are characterized by the control of certain thermodynamic quantities. If one knows how to calculate a thermodynamic quantity, e.g. the temperature or pressure, it is often possible to formulate an algorithm which fixes this property to a desired value. However, it is not always clear whether this algorithm describes the properties of a given thermodynamic ensemble.

One can distinguish four different types of control mechanisms:

Differential control : the thermodynamic quantity is fixed to the prescribed value and no fluctuations around an average value occur.

Proportional control : the variables, coupled to the thermodynamic property f , are corrected in each iteration step through a coupling constant towards the prescribed value of f . The coupling constant determines the strength of the fluctuations around $\langle f \rangle$.

Integral control : the system's Hamiltonian is extended and variables are introduced which represent the effect of an external system which fix the state to the desired ensemble. The time evolution of these variables is determined by the equations of motion derived from the extended Hamiltonian.

Stochastic control : the values of the variables coupled to the thermodynamic property f are propagated according to modified equations of motion, where certain degrees of freedom are additionally modified stochastically in order to give the desired mean value of f .

In the following, different statistical ensembles are presented and all methods will be discussed via examples.

4.1 The Microcanonical Ensemble

The microcanonical ensemble (NVE) may be considered as the *natural* ensemble for molecular dynamics simulations (as it is the canonical ensemble (NVT) for Monte Carlo simulations). If no time dependent external forces are considered, the system's Hamiltonian is constant, implying that the system's dynamics evolves on a constant energy surface. The corresponding probability density in phase space is therefore given by

$$\rho(\mathbf{q}, \mathbf{p}) = \delta(\mathcal{H}(\mathbf{q}, \mathbf{p}) - E) \quad (65)$$

In a computer simulation this theoretical condition is generally violated, due to limited accuracy in integrating the equations of motion and due to roundoff errors resulting from a limited precision of number representation. In Ref.⁸⁰ a numerical experiment was performed showing that tiny perturbations of the initial positions of a trajectory are doubled

about every picosecond. This would mean even for double precision arithmetic that after about 50 ps roundoff errors will be dominant.⁵⁹ This is, however, often not a too serious restriction, since most time correlation functions drop to zero on a much shorter time scale. Only for the case where long time correlations are expected one does have to be very careful in generating trajectories.

4.2 The Canonical Ensemble

The simplest extension to the microcanonical ensemble is the canonical one (N,V,T), where the number of particles, the volume and the temperature are fixed to prescribed values. The temperature T is, in contrast to N and V , an intensive parameter. The extensive counterpart would be the kinetic energy of the system. In the following, different control mechanisms, introduced in Sec. 4 are described.

4.2.1 The Differential Thermostat

Different methods were proposed to fix the temperature to a fixed value during a simulation without allowing fluctuations of T . The first method was introduced by Woodcock,⁸¹ where the velocities were scaled according to $\mathbf{p}_i \rightarrow \sqrt{T_0/T} \mathbf{p}_i$, where T_0 is the reference temperature and T the actual temperature, calculated from the velocity of the particles. This method leads to discontinuities in the momentum part of the phase space trajectory due to the rescaling procedure.

An extension of this method implies a constraint of the equations of motion to keep the temperature fixed.^{82–84} The principle of least constraint by Gauss states that a force added to restrict a particle motion on a constraint hypersurface should be normal to the surface in a realistic dynamics. From this principle the equations of motion are derived

$$\frac{\partial \mathbf{q}_i}{\partial t} = \mathbf{p}_i \quad (66)$$

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial V}{\partial \mathbf{q}_i} - \zeta \mathbf{p}_i \quad (67)$$

where ζ is a constraint force term, calculated as

$$\zeta = -\frac{\sum_{i=1}^N \frac{\mathbf{p}_i}{m_i} \frac{\partial V}{\partial \mathbf{q}_i}}{\sum_{i=1}^N \frac{\mathbf{p}_i^2}{m_i}} \quad (68)$$

Since the principle of least constraint by Gauss is used, this algorithm is also called *Gaussian thermostat*. It may be shown for this method that the configurational part of the phase space density is of canonical form, i.e.

$$\rho(\mathbf{q}, \mathbf{p}) = \delta(T - T_0) e^{-\beta U(\mathbf{q})} \quad (69)$$

4.2.2 The Proportional Thermostat

The proportional thermostat tries to correct deviations of the actual temperature T from the prescribed one T_0 by multiplying the velocities by a certain factor λ in order to move the system dynamics towards one corresponding to T_0 . The difference with respect to the differential control is that the method allows for fluctuations of the temperature, thereby not fixing it to a constant value. In each integration step it is insured that the T is corrected to a value more close to T_0 . A thermostat of this type was proposed by Berendsen et al.^{85,86} who introduced *weak coupling methods to an external bath*. The weak coupling thermostat was motivated by the minimization of local disturbances of a stochastic thermostat while keeping the global effects unchanged. This leads to a modification of the momenta $\mathbf{p}_i \rightarrow \lambda \mathbf{p}_i$, where

$$\lambda = \left[1 + \frac{\delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right) \right]^{\frac{1}{2}} \quad (70)$$

The constant τ_T , appearing in Eq.70, is a so called coupling time constant which determines the time scale on which the desired temperature is reached. It is easy to show that the proportional thermostat conserves a Maxwell distribution. However, the method cannot be mapped onto a specific thermodynamic ensemble. In Ref.⁸⁷ the phase space distribution could be shown to be

$$\rho(\mathbf{q}, \mathbf{p}) = f(\mathbf{p}) e^{-\beta(U(\mathbf{q}) - \alpha\beta\delta U(\mathbf{q})^2/3N)} \quad (71)$$

where $\alpha \simeq (1 - \delta E/\delta U)$ and δU , δE are the mean fluctuations of the potential and total energy. $f(\mathbf{p})$ is in general an unknown function of the momenta, so that the full density cannot be determined. For $\alpha = 0$, which corresponds in Eq.70 to $\tau_T = \delta t$, the fluctuations in the kinetic energy vanish and Eq.71 reduces to Eq.69, i.e. it represents the canonical distribution. The other extreme of $\tau_T \rightarrow \infty$ corresponds to an isolated system and the energy should be conserved, i.e. $\delta E = \delta K + \delta U = 0$ and $\alpha = 1$. In this case, Eq.71 corresponds to the microcanonical distribution.⁸⁷ Eq.71 may therefore be understood as an interpolation between the canonical and the microcanonical ensemble.

4.2.3 The Stochastic Thermostat

In the case of a stochastic thermostat, all or a subset of the degrees of freedom of the system are subject to collisions with *virtual* particles. This method can be motivated by a Langevin stochastic differential equation which describes the motion of a particle due to the thermal agitation of a heat bath

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial U}{\partial \mathbf{q}_i} - \gamma \mathbf{p}_i + \mathbf{F}^+ \quad (72)$$

where γ is a friction constant and \mathbf{F}^+ a Gaussian random force. The amplitude of \mathbf{F}^+ is determined by the second fluctuation dissipation theorem

$$\langle \mathbf{F}_i^+(t_1) \mathbf{F}_j^+(t_2) \rangle = 2\gamma k_B T \delta_{ij} \delta(t_1 - t_2) \quad (73)$$

A larger value for γ will increase thermal fluctuations, while $\gamma = 0$ reduces to the microcanonical ensemble. This method was applied to molecular dynamics in Ref.⁸⁸ A more

direct way was followed in Refs.^{89,90} where particles collide occasionally with virtual particles from a Maxwell distribution corresponding to a temperature T_0 and after collisions loose their memory completely, i.e. the motion is totally randomized and the momenta become discontinuous. In order not to disturb the phase space trajectory too much, the collision frequency has to be chosen not too high. Since a large collision frequency will lead to a strong loss of the particle's memory, it will lead to a fast decay of dynamic correlation functions.⁹¹ The characteristic decay time of correlation functions should therefore be a measure for the collision time. It was proved for the stochastic thermostat⁸⁹ that it leads to a canonical distribution function.

A slightly different method which is able to control the coupling to an external bath was suggested in Refs.^{92,93} In this approach the memory of the particle is not completely destroyed but the new momenta are chosen to be

$$\mathbf{p}_{i,n} = \sqrt{1 - \alpha^2} \mathbf{p}_{i,o} + \alpha \mathbf{p}_r \quad (74)$$

where \mathbf{p}_r is chosen a momentum, drawn from a Maxwell distribution corresponding to T_0 . Similar to the proportional thermostat, the parameter α may be tuned to give distributions ranging from the microcanonical to the canonical ensemble.

4.2.4 The Integral Thermostat

The integral method is also often called *extended system method* as it introduces additional degrees of freedom into the system's Hamiltonian for which equations of motion can be derived. They are integrated in line with the equations for the spatial coordinates and momenta. The idea of the method invented by Nosé,^{94,95} is to reduce the effect of an external system acting as heat reservoir to keep the temperature of the system constant, to one additional degree of freedom. The thermal interactions between a heat reservoir and the system result in a change of the kinetic energy, i.e. the velocity of the particles in the system. Formally it may therefore be expressed a scaling of the velocities. Nosé introduced two sets of variables: real and so called virtual ones. The virtual variables are consistently derived from a Sundman transformation⁹⁶ $d\tau/dt = s$, where τ is a virtual time and s is a resulting scaling factor, which is treated as dynamical variable. The transformation from virtual to real variables is then performed as

$$\mathbf{p}_i = \pi_i s \quad , \quad \mathbf{q}_i = \rho_i \quad (75)$$

The introduction of the *effective mass*, M_s , connects also a momentum to the additional degree of freedom, π_s . The resulting Hamiltonian, expressed in virtual coordinates reads

$$\mathcal{H}^* = \sum_{i=1}^N \frac{\pi_i^2}{2m_i s^2} + U(\boldsymbol{\rho}) + \frac{\pi_s^2}{2M_s} + gk_B T \ln s \quad (76)$$

where $g = 3N + 1$ is the number of degrees of freedom (system of N free particles). The Hamiltonian in Eq.76 was shown to lead to a probability density in phase space, corresponding to the canonical ensemble.

The equations of motion drawn from this Hamiltonian are

$$\frac{\partial \rho_i}{\partial \tau} = \frac{\pi_i}{s^2} \quad (77)$$

$$\frac{\partial \pi_i}{\partial \tau} = -\frac{\partial U(\rho)}{\partial \rho_i} \quad (78)$$

$$\frac{\partial s}{\partial \tau} = \frac{\pi_s}{M_s} \quad (79)$$

$$\frac{\partial \pi_s}{\partial \tau} = \frac{1}{s^3} \sum_{i=1}^N \frac{\pi_i^2}{m_i} - \frac{gk_B T}{s} \quad (80)$$

If one transforms these equations back into real variables, it is found⁹⁷ that they can be simplified by introducing the new variable $\zeta = \partial s / \partial t = sp_s / M_s$ (p_s is *real* momentum connected to the heat bath)

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} \quad (81)$$

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial U(\mathbf{q})}{\partial \mathbf{q}_i} - \zeta \mathbf{p}_i \quad (82)$$

$$\frac{\partial \ln s}{\partial t} = \zeta \quad (83)$$

$$\frac{\partial \zeta}{\partial t} = \frac{1}{M_s} \left(\sum_{i=1}^N \frac{\mathbf{p}_i^2}{m_i} - gk_B T \right) \quad (84)$$

These equations describe the so called Nosé-Hoover thermostat.

4.3 The Constant-Pressure Constant-Enthalpy Ensemble

In order to control the pressure in an MD simulation cell, it is necessary to allow for volume variations. A simple picture for a constant pressure system is a box the walls of which are coupled to a piston which controls the pressure. In contrast to the case where the temperature is controlled, no coupling to the dynamics of the particles (timescales) is performed but the length scales of the system will be modified. In the following, different algorithms are described for a constant pressure ensemble. The conserved quantity will not be the system's energy, since there will be an energy transfer to or from the *external* system (piston etc.), but the enthalpy H will be constant. In line with the constant temperature methods there are also differential, proportional, integral and stochastic methods to achieve a constant pressure situation in simulations. The differential method, however, is not discussed here, since there are problems with that method related to the *correct initial* pressure.^{98,99} A scheme for the calculation of the pressure in MD simulations for various model systems is given in the appendix.

4.3.1 The Proportional Barostat

The proportional thermostat in Sec. 4.2.2 was introduced as an extension for the equation of the momentum, since it couples to the kinetics of the particles. Since the barostat acts on a volume change, which may be expressed in a scaling of particles' positions, a phenomenological extension for the equation of motion of the coordinates may be formulated⁸⁵

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} + \alpha \mathbf{q}_i \quad (85)$$

while a change in volume is postulated as

$$\dot{V} = 3\alpha V \quad (86)$$

A change in pressure is related to the isothermal compressibility κ_T

$$\dot{P} = -\frac{1}{\kappa_T V} \frac{\partial V}{\partial t} = -\frac{3\alpha}{\kappa_T} \quad (87)$$

which is approximated as

$$\frac{(P_0 - P)}{\tau_P} = -\frac{3\alpha}{\kappa_T} \quad (88)$$

and therefore Eq.85 can be written as

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} - \frac{\kappa_T}{3\tau_P} (P_0 - P) \quad (89)$$

which corresponds to a scaling of the boxlength and coordinates $\mathbf{q} \rightarrow s\mathbf{q}$ and $L \rightarrow sL$, where

$$s = 1 - \frac{\kappa_T \delta t}{3\tau_P} (P_0 - P) \quad (90)$$

The time constant τ_P was introduced into Eq.88 as a characteristic timescale on which the system pressure will approach the desired pressure P_0 . It also controls the strength of the coupling to the barostat and therefore the strength of the volume/pressure fluctuations. If the isothermal compressibility, κ_T , is not known for the system, the constant $\tau'_P = \tau_P/\kappa_T$ may be considered as a phenomenological coupling time which can be adjusted to the system under consideration. As for the proportional thermostat, a drawback for this method is that the statistical ensemble is not known. In analog to the thermostat, it may be assumed to *interpolate* between the microcanonical and the constant-pressure/constant-enthalpy ensemble, depending on the coupling constant τ_P .

4.3.2 The Integral Barostat

In line with the integral thermostat one can introduce a new degree freedom into the systems Hamiltonian which controls volume fluctuations. This method was first proposed by Andersen.⁸⁹ The idea is to include the volume as an additional degree of freedom and to write the Hamiltonian in a scaled form, where lengths are expressed in units of the boxlength $L = V^{1/3}$, i.e. $\mathbf{q}_i = L \boldsymbol{\rho}_i$ and $\mathbf{p}_i = L \boldsymbol{\pi}_i$. Since L is also a dynamical

quantity, the momentum is not related to the simple time derivative of the coordinates but $\partial_t \mathbf{q}_i = L \partial_t \boldsymbol{\rho}_i + \boldsymbol{\rho}_i \partial_t L$. The extended system Hamiltonian is then written as

$$\mathcal{H}^* = \frac{1}{V^{2/3}} \sum_{i=1}^N \frac{\pi_i}{2m_i} + U(V^{1/3} \boldsymbol{\rho}) + P_{ex} V + \frac{\pi_V^2}{2M_V} \quad (91)$$

where P_{ex} is the prescribed external pressure and π_V and M_V are a momentum and a mass associated with the fluctuations of the volume.

The equations of motion which are derived from this Hamiltonian are

$$\frac{\partial \boldsymbol{\rho}_i}{\partial t} = \frac{1}{V^{2/3}} \frac{\pi_i}{m_i} \quad (92)$$

$$\frac{\partial \pi_i}{\partial t} = \frac{\partial U(V^{1/3} \boldsymbol{\rho})}{\partial \boldsymbol{\rho}_i} \quad (93)$$

$$\frac{\partial V}{\partial t} = \frac{\pi_V}{M_V} \quad (94)$$

$$\frac{\partial \pi_V}{\partial t} = \frac{1}{3V} \left(\frac{1}{V^{2/3}} \sum_{i=1}^N \frac{\pi_i}{m_i} - V^{1/3} \boldsymbol{\rho}_i \frac{\partial U(\mathbf{q})}{\partial \mathbf{q}_i} \right) \quad (95)$$

A transformation to real variables then gives

$$\frac{\partial \mathbf{q}_i}{\partial t} = \frac{\mathbf{p}_i}{m_i} + \frac{1}{3V} \frac{\partial V}{\partial t} \mathbf{q}_i \quad (96)$$

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial U(\mathbf{q})}{\partial \mathbf{q}_i} - \frac{1}{3V} \frac{\partial V}{\partial t} \mathbf{p}_i \quad (97)$$

$$\frac{\partial V}{\partial t} = \frac{\mathbf{p}_V}{M_V} \quad (98)$$

$$\frac{\partial \mathbf{p}_V}{\partial t} = \underbrace{\frac{1}{3V} \left(\sum_{i=1}^N \frac{\mathbf{p}_i}{m_i} - \mathbf{q}_i \frac{\partial U(\mathbf{q})}{\partial \mathbf{q}_i} \right)}_P - P_{ex} \quad (99)$$

In the last equation the term in brackets corresponds to the pressure, calculated from the virial theorem (cf. Appendix A). The associated volume force, introducing fluctuations of the box volume is therefore controlled by the internal pressure, originating from the particle dynamics and the external pressure, P_{ex} .

5 Parallel Molecular Dynamics

With the advent of massively parallel computers, where thousands of processors may work on a single task, it has become possible to increase the size of the numerical problems considerably. As has been already mentioned in Sec.1 it is in principle possible to treat

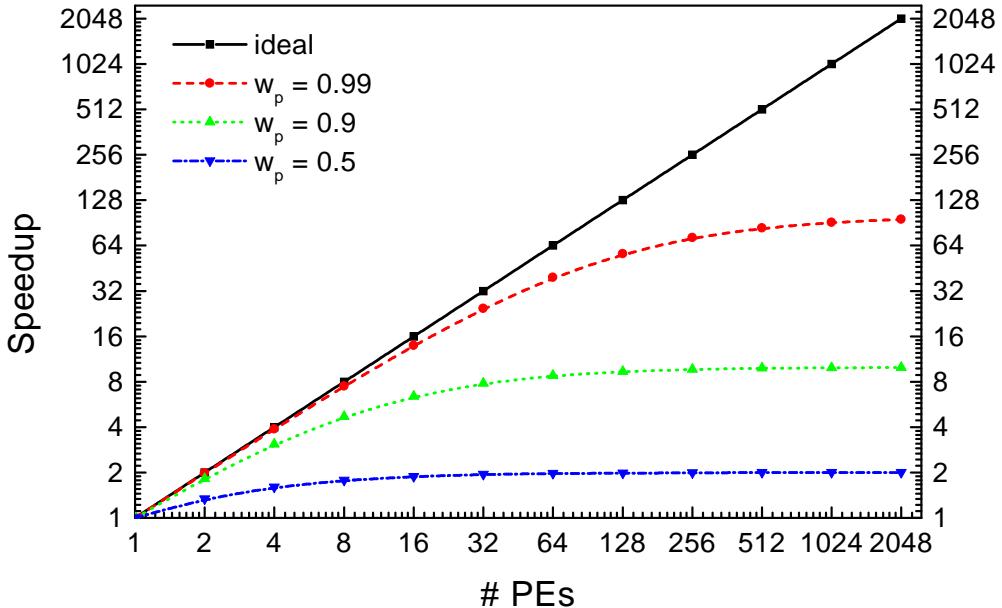


Figure 5. The ideal speedup for parallel applications with 50%, 90%, 99% and 100% (ideal) parallel work as a function of the number of processors.

multi-billion particle systems. However, the whole success of parallel computing strongly depends both on the underlying problem to be solved and the optimization of the computer program. The former point is related to a principle problem which is manifested in the so called Amdahl's law.¹⁰⁰ If a problem has inherently certain parts which can be solved only in serial, this will give an upper limit for the parallelization which is possible. The speedup σ , which is a measure for the gain of using multiple processors with respect to a single one, is therefore bound

$$\sigma = \frac{N_p}{w_p + N_p w_s}. \quad (100)$$

Here, N_p is the number of processors, w_p and w_s is the amount of work, which can be executed in parallel and in serial, i.e. $w_p + w_s = 1$. From Eq.100 it is obvious that the maximum efficiency is obtained when the problem is completely parallelizable, i.e. $w_p = 1$ which gives an N_p times faster execution of the program. In the other extreme, when $w_s = 1$ there is no gain in program execution at all and $\sigma = 1$, independent of N_p . In Fig.5 this limitation is illustrated for several cases, where the relative amount for the serial work was modified. If the parallel work is 50%, the maximum speedup is bound to $\sigma = 2$. If one aims to execute a program on a real massively parallel computer with hundreds or thousands of processors, the problem at hand must be inherently parallel for 99.99...%. Therefore, not only big parallel computers guarantee a fast execution of programs, but the problem itself has to be chosen properly.

Concerning MD programs there are only a few parts which have to be analysed for parallelization. As was shown, an MD program consists essentially of the force routine, which

Machine	CPU	Network	Latency	Bandwidth
CRAY T3E-1200 (ZAM Jülich)	DEC 21164 (600 MHz)	CRAY T3E interconnect	$\approx 8\mu s$ ($2\mu s$)	≈ 350 MB/s
ZAMpano (ZAM Jülich)	Intel Pentium III Xeon (550 MHz)	Myrinet	$\approx 80\mu s$ ($15\mu s$)	≈ 65 MB/s
MPCB (CNRS Orléans)	Intel Pentium III (550 MHz)	Fast Ethernet	$\approx 470\mu s$	≈ 10 MB/s

Table 1. Interprocessor communication networks for a massively parallel machine (CRAY T3E-1200) with 512 processors and two Linux based PC clusters with 32 (ZAMpano) and 40 (MPCB) processors.

costs usually more than 90% of the execution time. If one uses neighbor lists, these may be also rather expensive while reducing the time for the force evaluation. Other important tasks are the integration of motion, the parameter setup at the beginning of the simulation and the file input/output (I/O). In the next chapter it will be shown how to parallelize the force routine. The integrator may be naturally parallelized, since the loop over N particles may be subdivided and performed on different processors. The parameter setup has either to be done in serial so that every processor has information about relevant system parameters, or it may be done in parallel and information is distributed from every processor via a broadcast. The file I/O is a more complicated problem. The message passing interface MPI I does not offer a parallel I/O operation. In this case, if every node writes some information to the same file there is, depending on the configuration of the system, often only one node for I/O, to which internally the data are sent from the other nodes. The same applies for reading data. Since on this node the data from/for the nodes are written/read sequentially, this is a serial process which limits the speedup of the execution. The new MPI II standard offers parallel read/write operations, which lead to a considerable efficiency gain with respect to MPI. However, the efficiency obtained depend strongly on the implementation on different architectures.

Another serious point is the implementation into the computer code. A problem which is inherently 100% parallel will not be solved with maximum speed if the program is not 100% mapped onto this problem. Implementation details for parallel algorithms will be discussed in the following sections. Independent of the implementation of the code, Eq.100 gives only an upper theoretical limit which will only be reached in very rare cases. For most problems it is necessary to communicate data from one processor to another or even to all other processors in order to take into account data dependencies. This implies an overhead which depends on the latency and the bandwidth of the interprocessor network. This strongly depends on the hardware implementation, as is shown in Table 1.

It is shown that the latency time (time which is used to initialize a communication) differs by a factor of about 50, while bandwidths differ by a factor of about 35. The effect of the data exchange will be included into Amdahl's law later on and more realistic speed up curves will be obtained.

5.1 Particle Decomposition

In order to share the work between N_p processors one can distribute N particles in the beginning of the simulation homogenously onto the processors. If a particle is assigned

permanently to a certain processor, the method is called particle decomposition (PD). If particles are distributed in the beginning of the simulation according to their spatial arrangement, topologically neighbored processors will contain neighbored particles in space. This remains almost true if the system is very viscous or it is a solid. However, in the case of liquids or gases, the particles will mix after a short time due to diffusive motions and neighbored processors may store data of particles which are spatially far apart from each other and *vice versa*. In order to calculate the interparticle forces, a global communication between processors is necessary.

5.1.1 Replicated Data Algorithm

The simplest way to parallelize a serial program or to write a new parallel code is the so called replicated data algorithm. Every node stores the coordinates of all N particles in the system. However, in contrast to a serial program, each processor computes the forces only of a certain subset of particles N/N_p , the so called *local particles*, i.e. the number of loop iterations is N^2/N_p . Having computed the forces on the local particles, the integration step is performed and the positions and velocities are updated. The next step consists in broadcasting the new positions to the other $N_p - 1$ processors. This involves a global communication which can either be implemented in a loop over all processors or can be realized by optimized library operations, e.g. MPI_AllToAll, which sends a subarray to all other processors, where it is sorted into the whole array, and receives subarrays from all other processors which are sorted into the local whole array. A straightforward implementation of this algorithm thus requires $N_p - 1$ send/receive operations.

A more efficient way makes use of a tree-like communication pattern, which needs only $\log_2(N_p)$ send/receive operations. This algorithm was proposed by Fox et al.¹⁰¹ Assuming a linear processor topology with periodic boundary conditions this algorithm works as follows: in a first step all processors receive the particle coordinates from their neighbored processor to their right. Each processor element (PE) now stores the updated coordinates from 2 PEs. In the next step the updated coordinates are received from the second PE to the right, i.e. in this step the information of already four PEs is obtained. One proceeds in step n to receive the coordinate vector from the $2^{(n-1)}$ th neighbor to the right and to send the vector from the local PE to the $2^{(n-1)}$ th neighbor node to the left. For the case of 128 PEs the whole communication is finished after 7 steps instead of 127 steps in the case of an all-to-all communication scheme. Note, however, that the total amount of data which has to be sent is unaffected by the algorithm. Minimized is only the latency time of the send/receive operations. Furthermore, for the tree-like communication one has to introduce an temporary array which stores the incoming/outgoing data from/to neighbored PEs which have to be sorted into the *global* coordinate vector.

The scheme described up to now does not take into account the principle of action and counteraction (Newton's 3rd law). Implemented in an optimal way, this may speed up a computation by nearly a factor of two. An easy way to account for this symmetry relation is to divide the matrix of particle interactions from PE p_i and PE p_j ($p_i \neq p_j$) into a checkerboard scheme. Now, interactions are calculated on p_i when the coordinate index of the particles obey the property $i > j$ and $i + j$ is an even number. On the other hand, interactions on p_j are taken into account when $i < j$ and $i + j$ is an odd number. In a next step the locally computed forces have to be sent to the $N_p - 1$ processors and Newton's

3rd law can be applied. This scheme implies two global communications with coordinates and forces. It has to checked which method is more preferential. If communication is of minor importance, i.e. in the case of fast network, the 2nd variant will be faster.

5.1.2 Distributed Data Algorithm - Systolic Loop

The replicated data algorithm is easy and fast to implement. However, the disadvantage is the storage of a large coordinate vector. This may become difficult when very large numbers of particles are to be simulated and/or the computer memory is small.

Distributed data algorithms are then favorable. One type of this strategy is the systolic loop algorithm. In this algorithm a local PE stores only the coordinates, velocities and forces of the local particles.

Forces on the local PE are calculated in the usual way in a $N_L(N_L - 1)/2$ -loop, where N_L is the number of local particles. The communication between the PEs are organized in the following way. The processor indices are extended periodically, i.e. $\text{PE}_i = \text{PE}_{i+P}$, where $i = 1, \dots, P$ is the index of the processor. Coordinates are sent from PE_j to PE_i , with $i < j$, where the forces between the particles are calculated explicitly. The coordinates are stored in a temporary array of length $3N_L$. Applying the principle of action/counteraction the calculated force vectors are to be sent from $\text{PE } p_i$ to $\text{PE } p_j$. The total force evaluation on a tagged particle is completed after $(N_p - 1)/2$ send/receive operations of the coordinates in the first step and $(N_p - 1)/2$ send/receive operations of the forces in the second step.

The algorithm exhibits a special feature when N_p is an even number. In that case half of the processors do not work in the last loop over processors, since redundant information would be communicated^c. This implies a nonlinear scaling of the algorithm with increasing number of PEs, e.g. from $P = 1$ to $P = 2$ the maximum speed-up is $4/3$. Only for large numbers of PEs, the effect is reduced and the speed-up approaches a linear behavior. This effect may be avoided if one refers to the procedure discussed before (cf. Sec.5.1.1). In the last step of the loop over processors the send operations of coordinates are $p_1 \rightarrow p_{N_p/2-1}, \dots, p_{N_p/2} \rightarrow p_{N_p}$. In addition one communicates now coordinates also from $p_{N_p/2-1} \rightarrow p_1, \dots, p_{N_p} \rightarrow p_{N_p/2}$. In the following the force matrix is subdivided and its elements are only calculated explicitly if the particle index pair $i > j$ and the sum $i + j$ is even or if $i < j$ and the sum $i + j$ is odd. In the next step the force vector elements are send from $\text{PE } p_j$ to $\text{PE } p_i$ so that Newton's third law may be applied.

5.1.3 An Intermediate Algorithm - Hypersystolic Loop

In the case of the systolic loop algorithm, coordinates and forces are sent from one PE to the next and are stored only temporarily. On the other hand, if these coordinates would be stored on each PE, the information to calculate all forces in the system would be distributed in less than $N_p - 1$ send operations. This is illustrated in Table 2 for the case of 6 and 8 processors. In the case of 6 PEs the cycle is finished after 2 communication steps. As is seen, the interactions $p_1 \rightarrow p_2$ and $p_1 \rightarrow p_4$ are calculated on PE 1, $p_1 \rightarrow p_5$ is calculated

^cIn the last communication step PE N_p sends data to PE $N_p/2$ where forces are calculated and send back to PE N_p using the principle of action/counteraction. If PE $N_p/2$ would also send its coordinates to PE N_p and would receive the forces back from the same PE this results in a double calculation of forces leading to a wrong result.

PE	1	2	3	4	5	6		PE	1	2	3	4	5	6	7	8	
Step	0	1	2	3	4	5	6	Step	0	1	2	3	4	5	6	7	8
	1	2	3	4	5	6	1		1	2	3	4	5	6	7	8	1
	2	4	5	6	1	2	3		2	4	5	6	7	8	1	2	3

Table 2. Hypersystolic matrices for the cases of 6 and 8 processors. Bold numbers indicate the location of coordinates, which are used to calculate interactions with particles on processor 1.

on PE 4 and $p_1 \rightarrow p_3$ and $p_1 \rightarrow p_6$ are calculated on PE 6^d. This algorithm of sending and receiving data is called hypersystolic loop^{102, 103} and the scheme of communicating processors (cf. Table 2) is called the hypersystolic matrix. As a further example the case of 8 processors is also shown in Table 2, which finishes after 3 communication steps. For large numbers of processors this scheme is very promising as a compromise between the memory intensive replicated data algorithm and the communication intensive systolic loop algorithm. Unfortunately, however, a general scheme is not yet known how to build up the hypersystolic matrix, which makes the method unsortable for an arbitrary number of processors.

5.2 Force Decomposition

Another parallelization strategy aims to distribute the $N \times N$ force matrix onto the N_p processors. Two different implementations will be explained in the following.

5.2.1 Data Replicated Algorithm

This algorithm is closely related to the one discussed in Sec. 5.1.1. Particles are distributed uniformly on the processors and each PE stores the whole information of the N particles. The algorithm aims to use Newton's 3rd law, i.e. the calculation of the force matrix is reduced to the upper triangular matrix \mathbf{M}^u , and to distribute the work homogeneously, i.e. each PE should have the same number of force loop iterations. This is achieved in the following way.¹⁰⁴ Each PE is assigned a fixed number of rows of \mathbf{M}^u having the same area A on each PE. The total number of force iterations on each PE is given by

$$A(L_k) = \left(N - \sum_{j=1}^{k-1} L_j - \frac{L_k + 1}{2} \right) \quad (101)$$

with $k \in [1, N_p]$ and L_k being the number of rows in \mathbf{M}^u being assigned to PE p_k , i.e. $\sum_{k=1}^{N_p} L_k = N$. As is required for an equal workload, all areas $A(L_k)$ should be equal. This leads to

$$L_k = \frac{1}{2} \left(Q_k - \sqrt{Q_k^2 - \frac{4N(N-1)}{N_p}} \right) \quad (102)$$

^dThe notation $p_i \rightarrow p_j$ means thereby the interaction between particles on processor p_i with those on processor p_j .

where

$$Q_k = 2N - 1 - 2 \sum_{j=1}^{k-1} k - 1 L_j \quad (103)$$

This subdivision of forces guarantees equal lengths of force loops on each PE. In order to get the complete force on every particle, a global reduction of the force vector has to be done, which can, in analogy to Sec. 5.1.1, also be performed in a tree-like structure, requiring $\log_2(N_p)$ steps. In order to save a global reduction of positions, the propagation of positions and velocities from step n to $n + 1$ is done for all particles on every PE.

5.2.2 Low-Communication Version

Another type of decomposition of the force matrix requires a quadratic number of processors $N_p = n^2$, which are arranged in a $\sqrt{N_p} \times \sqrt{N_p}$ matrix P_{ij} .¹⁰⁵ The particle distribution on the processors is performed according to the following rule (cf. Fig.6)

$$I_{ij} = \begin{cases} [\frac{N}{N_p}(j - \frac{1}{2}) + 1; \frac{N}{N_p}j] & : i < j \\ [\frac{N}{N_p}(i - 1) + 1; \frac{N}{N_p}i] & : i = j \\ [\frac{N}{N_p}(i - 1) + 1; \frac{N}{N_p}(i - \frac{1}{2})] & : i > j \end{cases} \quad (104)$$

where I_{ij} denotes the interval of particle indices, stored on processor P_{ij} . Calculation of particle interactions is then performed on each local processor without exchange of particle coordinates with other PEs. In the following the symmetry of the transpose matrix is used and forces are exchanged between processor P_{ij} and P_{ji} ($i \neq j$). Now, in every row i of the matrix is the whole information for the forces which is necessary for the diagonal element P_{ii} . Consequently a reduction step is performed in the following to sum up the

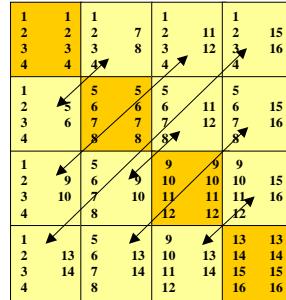


Figure 6. Communication pattern between 16 processors for the case of 16 particles (indicated numbers).

forces row wise on the diagonal elements of the matrix. In a next the particle coordinates and velocities are propagated only on the processors belonging to the diagonal elements of P_{ij} . In a last step, the updated positions are distributed according to Eq. 104. This scheme is less communication intensive as the algorithm discussed in Sec.5.2.1, since it requires only the row and column wise replication of data ($2(\sqrt{N_p} - 1)$ operations),

the transpose exchange (one communication operation for every *non-diagonal* PE) and the force reduction ($(\sqrt{N_p} - 1)$ operations). This reduction in communication is, however, achieved by a very much more complicated implementation and a less balanced workload, since not all processors have to calculate the same number of pair forces (cf. Fig.6).

5.3 Domain Decomposition

The principle of spatial decomposition methods is to assign geometrical domains to different processors. This implies that particles are no longer bound to a certain processor but will be transferred from one PE to another, according to their spatial position. This algorithm is especially designed for systems with short range interactions or to any other algorithm where a certain cut-off in space may be applied. Since neighbored processors contain all relevant data needed to compute forces on particles located on a given PE, this algorithm avoids the problem of global communications. Given that the range of interaction between particles is a cut-off radius of size R_c , the size, D of the domains is preferentially chosen to be $D > R_c$, so that only the $3^d - 1$ neighbored processors have to communicate data (d is the dimension of the problem). Whether this can be fulfilled depends on the interplay between size of the system and the numbers of processors. If a small system is treated with a large number of processors, the domains will be small and $D < R_c$. In this case not only the next but also the second or even higher order neighbor PEs have to send their coordinates to a given PE. For simplicity we assume here $D > R_c$.

The algorithm then works as follows. Particles are distributed in the beginning of the simulation to a geometrical region. The domains are constructed to have a rather homogeneous distribution of particles on each processor, e.g. for homogeneous bulk liquids the domains can be chosen as equally sized cuboids which fill the simulation box. In order to calculate forces between particles on different processors, coordinates of the so called *boundary particles* (those which are located in the outer region of size $R_b \geq R_c$ of the domains) have to be exchanged. Two types of lists are constructed for this purpose. The one contains all particle indices, which have left the local domain and which have consequently to be transferred to the neighbored PE. The other one contains all particle indices, which lie in the outer region of size R_b of a domain. The first list is used to update the particles' *address*, i.e. all information like positions, velocities, forces etc. are sent to the neighbored PE and are erased in the old domain. The second list is used to send temporarily position coordinates which are only needed for the force computation. The calculation of forces then operates in two steps. First, the forces due to local particles are computed using Newton's 3rd law. In a next step, forces due to the boundary particles are calculated. The latter forces are thus calculated twice: on the local PE and the neighbored PE. This extra computation has the advantage that there is no communication step for forces. A more elaborate scheme has nevertheless been proposed which includes also Newton's 3rd law for the boundary particles and thus the communication of forces.^{106,107} Having finished the evaluation of forces, the new positions and velocities are evaluated only for local particles.

A naive method would require $3^d - 1$ send/receive operations. However, this may be reduced to $2 \log_d(3^d - 1)$ operations with a similar tree-like method, as described in Sec.5.1.1. The method is described here for the case of $d = 2$. It may be generalized rather easily. The 4 processors, located directly at the edges of a given one are labeled as

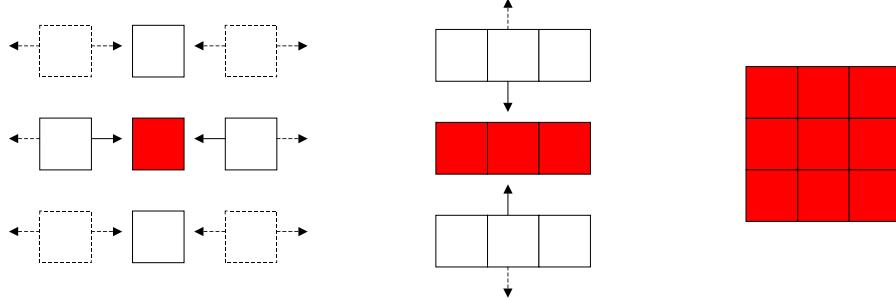


Figure 7. Communication pattern for the domain decomposition algorithm in 2 dimensions.

left/right and up/down. Then in a first step, information is sent/received to/from the left and the right PE, i.e. each processor now stores the coordinates of three PEs (including local information). The next step proceeds in sending/receiving the data to the up and down PEs. This step finishes already the whole communication process.

The updating process is not necessarily done in each time step. If the width of the boundary region is chosen as $R_b = R_c + \delta r$, it is possible to trigger the update automatically via the criterion $\max(|\mathbf{x}(t_0 + t) - \mathbf{x}(t_0)|) \leq \delta r$, which is the maximum change in distance of any particle in the system, measured from the last update.

A special feature of this algorithm is the fact that it shows a theoretical superlinear speed-up if Verlet neighbor lists are used. The construction of the Verlet list requires $N'(N' - 1)/2 + N'\delta N$ operations, where δN is the number of boundary particles and N' is the number of particles on a PE. If the number of PEs is increased as twice as large, there are $N'/2$ particles on each processor which therefore requires $N'/2(N'/2 - 1)/2 + N'/2\delta N$ operations. If $N' \gg \delta N$ and $N'^2 \gg N'$ one gets a speed-up factor of $\approx 4!$

5.4 Performance Estimations

In order to estimate the performance of the different algorithms on a theoretical basis it is useful to extend the ideal Amdahl's law to a more realistic case. The ideal law only takes into account the degree of parallel work. From that point of view all parallel algorithms for a given problem should work in the same way. However the communication between the processors is also a limiting factor in parallel applications and so it is natural to extend Amdahl's law in the following way

$$\sigma = \frac{1}{w_p/N_p + w_s + c(N_p)} \quad (105)$$

where $c(N_p)$ is a function of the number of processors which will characterize the different parallel algorithms. The function will contain both communication work, which depends on the bandwidth of the network and the effect of the latency time, i.e. how fast the network responds to the communication instruction. The function $c(N_p)$ expresses the relative portion of communication with respect to computation. Therefore it will depend in general also on the number of particles which are simulated.

In the following an analysis for three different types of parallel algorithms is presented. It is always assumed that the work is strictly parallel, i.e. $w_p = 1$.

5.4.1 All-to-All Communication

An example for the all-to-all communication was given with the systolic loop algorithm. Every processor sends to all other $N_p - 1$ PEs. The amount of data which is sent reduces with increasing N_p , since only the data stored on every PE are sent. The work for communication may therefore be expressed as

$$c(N_p) = (N_p - 1) \left(\lambda + \frac{\chi}{N_p} \right) \quad (106)$$

In Fig.8a the speedup, calculated on basis of Eq.106, is shown for some values of λ and χ . It is found that the slope of the speedup may become negative! Since the latency time λ is in general a small number, the third term in the denominator is small compared with the first. Therefore the speedup curve for small N_p is nearly linear. However, since for larger N_p the first term in c_p grows nearly linear with N_p , it becomes dominant for the behavior of σ . An interesting observation is that the behavior for large N_p is mainly dominated by the latency time than the amount of data which is sent. With an increase of N_p the number of particles which is stored on each PE is reduced and consequently the ratio of computation to communication becomes smaller and smaller. For very few data on each PE, it resembles somebody who wants to give a telephone call but after every spoken word he has to dial again. One can imagine that even with modern telephones the time for finishing a message becomes longer and longer.

5.4.2 Tree-Like Communication

A parallel algorithm which uses a tree-like structure to distribute the data was discussed for the force-decomposition algorithm. In this case the number of message passing calls reduces to $\log_2(N_p)$ whereas the amount of data to be send to the next PE is doubled in each communication step. The expression for c_p may therefore be written as

$$c(N_p) = \log_2(N_p)\lambda + \sum_{n=1}^{\log_2(N_p)} \frac{2^{(n-1)}\chi}{N_p} \quad (107)$$

In Fig.8b the speedup behavior is shown. It is found that it is decreased with respect to the ideal behavior. Due to the slow logarithmic increase of the latency time part no decreasing behavior of σ is observed. For the unrealistic case where the communication part is neglected, the speedup is rather close to the ideal line. A considerable deviation is only found for $N_p > 500$. The result for this algorithm is therefore that it is communication limited rather than latency time limited.

5.4.3 Local Communication

The spatial decomposition algorithm is an example for the case of local communication. As was described in Sec.5.3, only six communication steps are required to distribute the

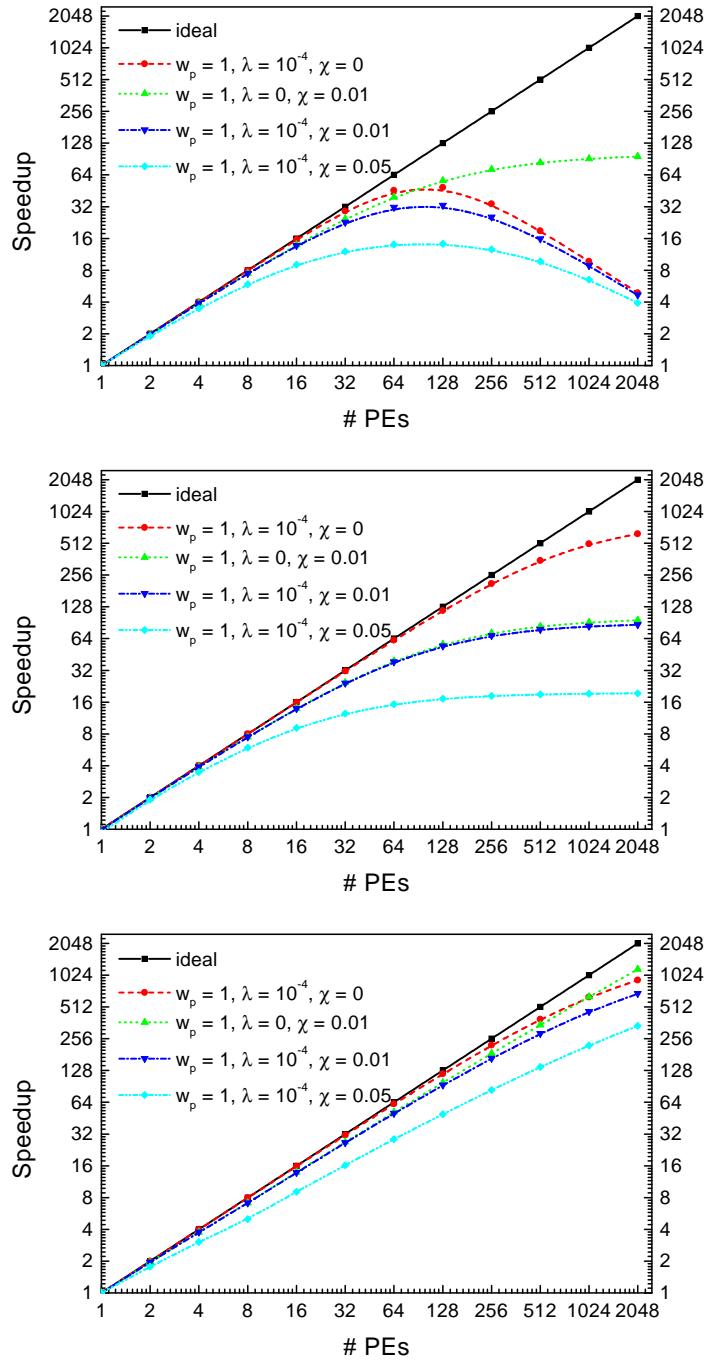


Figure 8. Estimations of realistic speedup curves if one includes the latency time and bandwidth of the processor interconnect. It is assumed that the problem can be parallelized for 100%. Different parameter values are compared for the latency time λ and bandwidth χ for the all-to-all communication (top), a tree like communication (middle) and local nearest neighbor communications (bottom). The ideal curve neglects communication completely.

data to neighbored PEs. Therefore the latency time part is constant whereas the amount of data to be sent and consequently the communication part is decreased with larger N_p . The communication function reads therefore

$$c(N_p) = f(N_p) \left(\lambda + \frac{\chi}{N_p^{2/3}} \right), \quad f(N_p) = \begin{cases} 0 & N_p = 1 \\ 2 & N_p = 2 \\ 4 & N_p = 4 \\ 6 & N_p \leq 8 \end{cases} \quad (108)$$

Here the function $f(N_p)$ was introduced to cover also the cases for small numbers of PEs, where a data exchange is not necessary in each spatial direction. As seen from Fig.8c the speedup curves are nearly linear with a slightly smaller slope than unity. However, for very large numbers of PEs the curves will also flatten. Nevertheless, the local communication model provides the best speedup behavior from all parallel algorithms and seems to be best suited for large parallel architectures.

5.4.4 Final Remark

Note that the local communication model in its present form is only valid for short range interaction potentials. If the potential is longer ranged than one spatial domain, the function $f(N_p)$ has to be modified. For long range interactions, all-to-all communications are generally required. In that case the tree-method may be mostly preferred.

This theoretical analysis demonstrates the importance of a fast interconnect between processors for the case of molecular dynamics simulations. Not included in the communication function $c(N_p)$ is the bandwidth function of the network. This, however, will only slightly change Figs.8a-c.

Appendix

A Calculating the Pressure

As is well known from thermodynamics the pressure may be calculated via the virial theorem. However, there are problems in deriving the expression for the pressure when working in periodic boundary conditions,¹⁰⁸ due to the missing walls on which the pressure acts. The usual derivation of the virial theorem where a gas or a liquid is bound in a certain volume, fails in this case. However, using a fictitious surface, where particles may cross over, it is possible to obtain a similar expression for the pressure in periodic boundary conditions (PBC) as in finite volumes, which reads

$$P = \frac{2}{3V} \left\langle \sum_{i=1}^N \frac{1}{2} m_i v_i^2 \right\rangle - \left\langle \frac{\partial U}{\partial \mathbf{r}_{ij}} \mathbf{r}_{ij} \right\rangle \quad (109)$$

$$= \frac{NkT}{V} + \frac{1}{3V} \left\langle \sum_{\alpha} \sum_{i,j} \mathbf{F}(\mathbf{r}_{ij} - \alpha L)(\mathbf{r}_{ij} - \alpha L) \right\rangle \quad (110)$$

where L is the length of the simulation box and the parameter α accounts for the periodic images. If the range of interaction between particles is smaller than $L/2$, this translation corresponds to the so called *minimum image convention*.

A different derivation of the pressure, which is rather convenient for the case of molecular systems, starts from the thermodynamic definition

$$P = - \left(\frac{\partial A}{\partial V} \right)_T \quad (111)$$

where A is the Helmholtz free energy of the system, which can be expanded to give

$$P = \frac{1}{\beta Q} \frac{\partial Q}{\partial V} \quad (112)$$

where Q is the partition function

$$Q(V, T) = \frac{1}{N! h^{3N}} \int d\mathbf{r} d\mathbf{p} e^{-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})} \quad (113)$$

and \mathcal{H} is the system's Hamiltonian. In order to introduce the volume as an independent parameter, the following change in variables is performed

$$\mathbf{r} = V^{1/3} \boldsymbol{\rho} \quad ; \quad \mathbf{p} = V^{-1/3} \boldsymbol{\pi} \quad ; \quad \boldsymbol{\pi} = m V^{1/3} \partial_t \boldsymbol{\rho} \quad (114)$$

Therefore the expression for the pressure can be written as

$$P = \frac{1}{N! h^{3N}} \frac{1}{\beta Q} \frac{\partial}{\partial V} \int \int d\boldsymbol{\rho} d\boldsymbol{\pi} e^{-\beta \mathcal{H}(V^{1/3} \boldsymbol{\rho}, V^{-1/3} \boldsymbol{\pi})} \quad (115)$$

$$= - \frac{1}{N! h^{3N}} \frac{1}{Q} \int \int d\boldsymbol{\rho} d\boldsymbol{\pi} \frac{\partial}{\partial V} \mathcal{H}(V^{1/3} \boldsymbol{\rho}, V^{-1/3} \boldsymbol{\pi}) e^{-\beta \mathcal{H}(V^{1/3} \boldsymbol{\rho}, V^{-1/3} \boldsymbol{\pi})} \quad (116)$$

$$= - \left\langle \frac{\partial}{\partial V} \mathcal{H}(V^{1/3} \boldsymbol{\rho}, V^{-1/3} \boldsymbol{\pi}) \right\rangle_T \quad (117)$$

In the following two examples are given how to calculate the pressure for specific systems.

A.1 Monatomic Systems

The Hamiltonian for a monatomic system, interacting via pair-forces may be written in scaled variables as

$$\mathcal{H} = \frac{1}{V^{\frac{2}{3}}} \sum_{i=1}^N \frac{\boldsymbol{\pi}_i^2}{2m_i} + \sum_{i,j=1; i < j}^N U(V^{\frac{1}{3}} \rho_{ij}) \quad (118)$$

where $\rho_{ij} = |\boldsymbol{\rho}_i - \boldsymbol{\rho}_j|$. Differentiating with respect to the volume gives

$$\frac{\partial \mathcal{H}}{\partial V} = - \frac{2}{3V^{\frac{5}{3}}} \sum_{i=1}^N \frac{\boldsymbol{\pi}_i^2}{2m_i} + \frac{1}{3V^{\frac{2}{3}}} \sum_{i,j=1; i < j}^N U'(V^{\frac{1}{3}} \rho_{ij}) \rho_{ij} \quad (119)$$

Transforming back this expression into *real variables*

$$\frac{\partial \mathcal{H}}{\partial V} = - \frac{2}{3V} \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + \frac{1}{3V} \sum_{i,j=1; i < j}^N U'(r_{ij}) r_{ij} \quad (120)$$

leads to the equation for the pressure

$$P = \frac{1}{3V} \left\langle \sum_{i=1}^N \frac{\mathbf{p}_i^2}{m_i} - \sum_{i,j=1; i \neq j}^N U'(r_{ij}) r_{ij} \right\rangle_T \quad (121)$$

which is the same expression predicted by the Virial theorem.¹⁰⁹

A.2 Rigid Nonlinear Molecules

As outlined in Sec.3.4.1 the motion of a rigid body can be described as a translation of the center-of-mass (COM) and a rotation around the principal axis. The Hamiltonian is therefore described in terms of positional and orientational coordinates. The scaling procedure, Eq.114, is applied again only on the COM coordinates and momenta, since orientational degrees of freedom have no inherent length scale. The Hamiltonian in scaled variables can therefore be written as

$$\mathcal{H} = \frac{1}{V^{\frac{2}{3}}} \sum_{i=1}^N \frac{\pi_i^2}{2M_i} + \frac{1}{2} \sum_{i=1}^N \omega_i^T \mathbf{I} \omega_i + \sum_{i,j=1; i < j}^N \sum_{a,b=1}^{N_s} U(V^{\frac{1}{3}} \rho_{ij}^{ab}) \quad (122)$$

Here π_i denotes the scaled momentum of the COM of molecule i , M_i the molecular mass, N_s the number of molecular sites and $\rho_{ij}^{ab} = |\rho_i^a - \rho_j^b|$ is the scaled distance between site a on molecule i and site b on molecule j . The position of site a is thereby given as

$$\rho_i^a = \rho_i + \mathbf{d}^a V^{-\frac{1}{3}} \quad (123)$$

where ρ_i is the position of the COM of molecule i and \mathbf{d}^a the distance vector from the COM to site a . Since only the center of mass vector, \mathbf{R} , is scaled with the volume term, the distance vector, \mathbf{d} , is multiplied by $V^{-1/3}$. Differentiating Eq.122 with respect to the volume gives

$$P = \frac{1}{3V} \left\langle \frac{\pi_i^2}{2M_i} - \sum_{i,j=1; i < j}^N \sum_{a,b=1}^{N_s} U'(R_{ij}^{ab}) \frac{1}{R_{ij}^{ab}} (\mathbf{R}_{ij} - \mathbf{d}^{ab}) \mathbf{R}_{ij} \right\rangle_T \quad (124)$$

$$= \frac{1}{3V} \left\langle \frac{\pi_i^2}{2M_i} - \sum_{i,j=1; i < j}^N \sum_{a,b=1}^{N_s} \mathbf{F}_{ij}^{ab} \mathbf{R}_{ij} + \sum_{i=1}^N \sum_{a=1}^{N_s} \mathbf{F}_i^a \mathbf{d}_a \right\rangle_T \quad (125)$$

where \mathbf{F}_{ij}^{ab} is the force, acting from site a of molecule i on site b of molecule j and \mathbf{F}_i^a the total force on site a of molecule i . In order to write Eq.125 in the present form, it was used that the mean value $\langle \mathbf{F}_i^a \mathbf{d}_b \rangle$ vanishes. The actual values of the scalar product, however, may give contributions to the fluctuations of the pressure. The last term in Eq.125 acts as a kind of correction, which reduces the pressure with respect to a simple superposition of pairforces in the expression of the virial theorem. This fact may be interpreted as arising from the constraint forces which keep the molecule rigid.

References

1. Y. Duan and P. A. Kollman. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282:740, 1998.
2. J. Roth, F. Gähler, and H.-R. Trebin. A molecular dynamics run with 5.180.116.000 particles. *Int. J. Mod. Phys. C*, 11:317–322, 2000.
3. B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27:1208–1209, 1957.
4. B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. I. General method. *J. Chem. Phys.*, 31:459, 1959.
5. J. Stadler, R. Mikulla, and H.-R. Trebin. IMD: A software package for molecular dynamics studies on parallel computers. *Int. J. Mod. Phys. C*, 8:1131–1140, 1997.
6. Y. Duan, L. Wang, and P. A. Kollman. The early stage of folding of villin headpiece subdomain observed in 200-nanosecond fully solvated molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA*, 95:9897, 1998.
7. <http://wserv1.dl.ac.uk/CCP/CCP5/>.
8. <http://www.amber.ucsf.edu/amber/amber.html>.
9. <http://www.scripps.edu/brooks/c27docs/Charmm27.Html>.
10. <http://www.ks.uiuc.edu/Research/namd/>.
11. <http://www.emsl.pnl.gov:2080/docs/nwchem/nwchem.html>.
12. <http://www.cs.sandia.gov/sjplimp/lammps.html>.
13. W. L. Cui, F. B. Li, and N. L. Allinger. *J. Amer. Chem. Soc.*, 115:2943, 1993.
14. N. Nevins, J. H. Lii, and N. L. Allinger. *J. Comp. Chem.*, 17:695, 1996.
15. S. L. Mayo, B. D. Olafson, and W. A. Goddard. *J. Phys. Chem.*, 94:8897, 1990.
16. M. J. Bearpark, M. A. Robb, F. Bernardi, and M. Olivucci. *Chem. Phys. Lett.*, 217:513, 1994.
17. T. Cleveland and C. R. Landis. *J. Amer. Chem. Soc.*, 118:6020, 1996.
18. A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. *J. Amer. Chem. Soc.*, 114:10024, 1992.
19. Z. W. Peng, C. S. Ewig, M.-J. Hwang, M. Waldman, and A. T. Hagler. Derivation of class ii force fields. 4. van der Waals parameters of Alkali metal cations and Halide anions. *J. Phys. Chem.*, 101:7243–7252, 1997.
20. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Amer. Chem. Soc.*, 117:5179–5197, 1995.
21. A. D. Mackerell, J. Wiorkiewicz-Kuczera, and M. Karplus. *J. Amer. Chem. Soc.*, 117:11946, 1995.
22. W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Amer. Chem. Soc.*, 118:11225–11236, 1996.
23. T. A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comp. Chem.*, 17:490–519, 1996.
24. J. Kong. *J. Chem. Phys.*, 59:2464, 1973.
25. M. Waldman and A. T. Hagler. *J. Comp. Chem.*, 14:1077, 1993.
26. J. Delhommelle and P. Millié. Inadequacy of the lorentz-bertelot combining rules

- for accurate predictions of equilibrium properties by molecular simulation. *Molec. Phys.*, 99:619–625, 2001.
27. E. L. Pollock and B. J. Alder. Static dielectric properties of stockmayer fluids. *Physica*, 102A:1, 1980.
 28. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys.*, 64:253, 1921.
 29. S. W. de Leeuw, J. M. Perram, and E. R. Smith. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc. R. Soc. London*, A373:27, 1980.
 30. S. W. de Leeuw, J. M. Perram, and E. R. Smith. Simulation of electrostatic systems in periodic boundary conditions. II. Equivalence of boundary conditions. *Proc. R. Soc. London*, A373:57, 1980.
 31. P. Gibbon. Long range interactions in many-particle simulation. (this volume).
 32. J. W. Perram, H. G. Petersen, and S. W. de Leeuw. An algorithm for the simulation of condensed matter which grows as the $3/2$ power of the number of particles. *Molec. Phys.*, 65:875–893, 1988.
 33. D. Fincham. Optimisation of the Ewald sum for large systems. *Molec. Sim.*, 13:1–9, 1994.
 34. J. Kolafa and J. W. Perram. Cutoff errors in the Ewald summation formulae for point charge systems. *Molec. Sim.*, 9:351–368, 1992.
 35. D. M. Heyes. Electrostatic potentials and fields in infinite point charge lattices. *J. Chem. Phys.*, 74:1924–1929, 1980.
 36. W. Smith. Point multipoles in the Ewald sum. *CCP5 Newsletter*, 46:18–30, 1998.
 37. T. M. Nymand and P. Linse. Ewald summation and reaction field methods for potentials with atomic charges, dipoles and polarizabilities. *J. Chem. Phys.*, 112:6152–6160, 2000.
 38. G. Salin and J. P. Caillol. Ewald sums for yukawa potentials. *J. Chem. Phys.*, 113:10459–10463, 2000.
 39. L. Verlet. Computer experiments on classical fluids. I. Thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98, 1967.
 40. C. W. Gear. *Numerical initial value problems in ordinary differential equations*. Prentice Hall, Englewood Cliffs, NJ, 1971.
 41. R. W. Hockney. The potential calculation and some applications. *Meth. Comput. Phys.*, 9:136–211, 1970.
 42. D. Beeman. Some multistep methods for use in molecular dynamics calculations. *J. Comp. Phys.*, 20:130–139, 1976.
 43. H. F. Trotter. On the product of semi-groups of operators. *Proc. Am. Math. Soc.*, 10:545–551, 1959.
 44. O. Buneman. Time-reversible difference procedures. *J. Comp. Phys.*, 1:517–535, 1967.
 45. E. Hairer and P. Leone. Order barriers for symplectic multi-value methods. In D. Griffiths, D. Higham, and G. Watson, editors, *Pitman Research Notes in Mathematics*, volume 380, pages 133–149, 1998.
 46. D. Okunbor and R. D. Skeel. Explicit canonical methods for Hamiltonian systems. *Math. Comput.*, 59:439–455, 1992.
 47. E. Hairer. Backward error analysis of numerical integrators and symplectic methods.

Ann. Numer. Math., 1:107–132, 1994.

48. E. Hairer and C. Lubich. The lifespan of backward error analysis for numerical integrators. *Numer. Math.*, 76:441–462, 1997.
49. S. Reich. Backward error analysis for numerical integrators. *SIAM J. Numer. Anal.*, 36:1549–1570, 1999.
50. D. M. Stoffer. *Some geometrical and numerical methods for perturbed integrable systems*. PhD thesis, Swiss Federal Institute of Technology, Zürich, 1988.
51. J. M. Sanz-Serna M. Calvo. *Numerical Hamiltonian Problems*. Chapman and Hall, London, 1994.
52. R. D. Skeel. Integration schemes for molecular dynamics and related applications. In M. Ainsworth, J. Levesley, and M. Marletta, editors, *The Graduate Student's Guide to Numerical Analysis*, pages 119–176, New York, 1999. Springer.
53. M. E. Tuckerman and W. Langel. Multiple time scale simulation of a flexible model of co₂. *J. Chem. Ohys.*, 100:6368, 1994.
54. P. Procacci, T. Darden, and M. Marchi. A very fast Molecular Dynamics method to simulate biomolecular systems with realistic electrostatic interactions. *J. Phys. Chem.*, 100:10464–10468, 1996.
55. P. Procacci, M. Marchi, and G. L. Martyna. Electrostatic calculations and multiple time scales in molecular dynamics simulation of flexible molecular systems. *J. Chem. Phys.*, 108:8799–8803, 1998.
56. P. Procacci and M. Marchi. Taming the Ewald sum in molecular dynamics simulations of solvated proteins via a multiple time step algorithm. *J. Chem. Phys.*, 104:3003–3012, 1996.
57. J. J. Biesiadecki and R. D. Skeel. Dangers of multiple time step methods. *J. Comp. Phys.*, 109:318–328, 1993.
58. J. L. Scully and J. Hermans. Multiple time steps: limits on the speedup of molecular dynamics simulations of aqueous systems. *Molec. Sim.*, 11:67–77, 1993.
59. B. J. Leimkuhler and R. D. Skeel. Symplectic numerical integrators in constrained Hamiltonian systems. *J. Comp. Phys.*, 112:117–125, 1994.
60. T. Schlick. Some failures and success of long timestep approaches to biomolecular simulations. In P. Deuflhard, J. Hermans, B. J. Leimkuhler, A. Mark, S. Reich, and R. D. Skeel, editors, *Lecture notes in computational science and engineerung. Algorithms for macromolecular modelling*, volume 4, pages 221–250, New York, 1998. Springer.
61. E. Barth and T. Schlick. Overcoming stability limitations in biomolecular danamics. I. Combining force splitting via extrapolation with Langevin dynamics. *J. Chem. Phys.*, 109:1617–1632, 1998.
62. E. Barth and T. Schlick. Extrapolation versus impulse in multiple-timestepping schemes. II. Linear analysis and applications to Newtonian and Langevin dynamics. *J. Chem. Phys.*, 109:1633–1642, 1998.
63. B. Garcia-Archilla, J. M. Sanz-Serna, and R. D. Skeel. Long-time-step methods for oscillatory differential equations. *SIAM J. Sci. Comp.*, 20:930–963, 1998.
64. B. Garcia-Archilla, J. M. Sanz-Serna, and R. D. Skeel. The mollified impulse method for oscillatory differential equations. In D. F. Griffiths and G. A. Watson, editors, *Numerical analysis 1997*, pages 111–123, London, 1998. Pitman.
65. B. Garcia-Archilla, J. M. Sanz-Serna, and R. D. Skeel. The mollified impulse method

- for oscillatory differential equations. *SIAM J. Sci. Comp.*, 20:930–963, 1998.
- 66. J. A. Izaguirre. *Longer time steps for molecular dynamics*. PhD thesis, University of Illinois at Urbana-Champaign, 1999.
 - 67. J. A. Izaguirre, S. Reich, and R. D. Skeel. Longer time steps for molecular dynamics. *J. Chem. Phys.*, 110:9853, 1999.
 - 68. J. Barojas, D. Levesque, and B. Quentrec. Simulation of diatomic homonuclear liquids. *Phys. Rev. A*, 7:1092–1105, 1973.
 - 69. J. B. Kuipers. *Quaternions and rotation sequences*. Princeton University Press, Princeton, New Jersey, 1998.
 - 70. D. Fincham. Leapfrog rotational algorithms. *Molec. Simul.*, 8:165, 1992.
 - 71. A. Dullweber, B. Leimkuhler, and R. McLachlan. Symplectic splitting methods for rigid body molecular dynamics. *J. Chem. Phys.*, 107:5840–5851, 1997.
 - 72. A. Kol, B. B. Laird, and B. J. Leimkuhler. A symplectic method for rigid-body molecular simulation. *J. Chem. Phys.*, 107:2580–2588, 1997.
 - 73. A. J. Stone, A. Dullweber, M. P. Hodges, P. L. A. Popelier and D. J. Wales. ORIENT: A program for studying interactions between molecules, Version 3.2, University of Cambridge (1995–1997), available at <http://www-stone.ch.cam.ac.uk/programs.html>.
 - 74. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran*. Cambridge University Press, Cambridge, 1992.
 - 75. J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-Alkanes. *J. Comp. Phys.*, 23:327–341, 1977.
 - 76. H. C. Andersen. Rattle: a velocity version of the Shake algorithm for molecular dynamics calculations. *J. Comp. Phys.*, 52:24–34, 1982.
 - 77. B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: a linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–1472, 1997.
 - 78. T. R. Forester and W. Smith. SHAKE, Rattle and Roll: efficient constraint algorithms for linked rigid bodies. *J. Comp. Chem.*, 19:102–111, 1998.
 - 79. X.-W. Wu and S.-S. Sung. Constraint dynamics algorithm for simaultion of semiflexible macromolecules. *J. Comp. Chem.*, 19:1555–1566, 1998.
 - 80. M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*. Oxford Science Publications, Oxford, 1987.
 - 81. L. V. Woodcock. Isothermal molecular dynamics calculations for liquid salt. *Chem. Phys. Lett.*, 10:257–261, 1971.
 - 82. W. G. Hoover, A. J. C. Ladd, and B. Moran. High strain rate plastic flow studied via nonequilibrium molecular dynamics. *Phys. Rev. Lett.*, 48:1818–1820, 1982.
 - 83. D. J. Evans, W. G. Hoover, B. H. Failor, B. Moran, and A. J. C. Ladd. Nonequilibrium molecular dynamics via Gauss’s principle of least constraint. *Phys. Rev. A*, 28:1016–1021, 1983.
 - 84. D. J. Evans. Computer experiment for nonlinear thermodynamics of Couette flow. *J. Chem. Phys.*, 78:3298–3302, 1983.
 - 85. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684, 1984.
 - 86. H. J. C. Berendsen. Transport properties computed by linear response through weak

- coupling to a bath. In M. Meyer and V. Pontikis, editors, *Computer Simulation in Materials Science*, pages 139–155, Amsterdam, 1991. Kluwer Academic Publishers.
87. T. Morishita. Fluctuation formulas in molecular dynamics simulations with the weak coupling heat bath. *J. Chem. Phys.*, 113:2976–2982, 2000.
 88. T. Schneider and E. Stoll. Molecular dynamics study of a three dimensional one-component model for distortive phase transitions. *Phys. Rev. B*, 17:1302–1322, 1978.
 89. H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384, 1980.
 90. E. Bonomi. *J. Stat. Phys.*, 39:167, 1985.
 91. H. Tanaka, K. Nakanishi, and N. Watanabe. *J. Chem. Phys.*, 78:2626, 1983.
 92. M. E. Riley, M. E. Coltrin, and D. J. Diestler. A velocity reset method of simulating thermal motion and damping in gas-solid collisions. *J. Chem. Phys.*, 88:5934–5942, 1988.
 93. G. Sutmann and B. Steffen. Correction of finite size effects in molecular dynamics simulation applied to friction, 2001. submitted to Comp. Phys. Comm.
 94. S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511–519, 1984.
 95. S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molec. Phys.*, 52:255–268, 1984.
 96. K. Zare and V. Szebehely. Time transformations for the extended phase space. *Celestial Mech.*, 11:469, 1975.
 97. W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.
 98. D. J. Evans and G. P. Morris. The isothermal isobaric molecular dynamics ensemble. *Phys. Lett. A*, 98:433–436, 1983.
 99. D. J. Evans and G. P. Morris. Isothermal isobaric molecular dynamics. *Chem. Phys.*, 77:63–66, 1983.
 100. G. M. Amdahl. Validity of the single-processor approach to achieving large scale computing capabilities. In *AFIPS Conference Proceedings*, volume 30, pages 483–485, Reston, Va., 1967. AFIPS Press.
 101. G. C. Fox, M. A. Johnson, G. A. Lyzenga, S. W. Otto, J. K. Salmon, and D. W. Walker. *Solving problems on concurrent processors: Volume 1*. Prentice Hall, Englewood Cliffs, NJ, 1988.
 102. T. Lippert, H. Hoeber, G. Ritzenhöfer, and K. Schilling. Hyper-Systolic Processing on APE100/Quadratics N²-Loop Computations. Technical Report HLRZ 95-45, WUB 95-21, 1995.
 103. T. Lippert, A. Seyfried, A. Bode, and Schilling K. Hyper-systolic parallel computing. *IEEE Trans. Paral. Distr. Syst.*, 9:97–108, 1998.
 104. R. Murty and D. Okunbor. Efficient parallel algorithms for molecular dynamics simulations. *Parall. Comp.*, 25:217–230, 1999.
 105. V. E. Taylor, R. L. Stevens, and K. E. Arnold. Parallel molecular dynamics: Communication requirements for parallel machines. In *Proc. of the fifth Symposium on the Frontiers of Massively Parallel Computation*, pages 156–163, 1994.
 106. D. Brown, J. H. R. Clarke, M. Okuda, and T. Yamazaki. A domain decomposition parallel processing algorithm for molecular dynamics simulations of polymers. *Comp.*

- Phys. Comm.*, 83:1, 1994.
- 107. M. Pütz and A. Kolb. Optimization techniques for parallel molecular dynamics using domain decomposition. *Comp. Phys. Comm.*, 113:145–167, 1998.
 - 108. J. J. Erpenbeck and W. W. Wood. Molecular dynamics techniques for hard core systems. In B. J. Berne, editor, *Statistical mechanics B: Modern theoretical chemistry*, volume 6, pages 1–40, New York, 1977. Plenum.
 - 109. H. Goldstein. *Classical Mechanics*. Addison Wesley, Reading, Massachusetts, 1950.

Static and Time-Dependent Many-Body Effects via Density-Functional Theory

Heiko Appel and Eberhard K. U. Gross

Institut für Theoretische Physik, Freie Universität Berlin
Arnimallee 14, 14195 Berlin, Germany
E-mail: {appel, hardy}@physik.fu-berlin.de

After introducing the basic concepts of static and time-dependent density-functional theory we focus on numerical algorithms for the propagation of the time-dependent Kohn-Sham equations. Two different methods, based on modifications of the Crank-Nicholson and the split-operator propagation schemes, respectively, are presented. We discuss some strategies for the parallelization of the Kohn-Sham propagation using state-of-the-art message-passing protocols. Finally, some results for atoms in strong laser fields are presented.

1 Introduction

The non-relativistic treatment of quantum-mechanical problems in solid-state physics or quantum chemistry requires, in principle, the solution of the full many-body Schrödinger equation for the combined system of electrons and nuclei

$$(\hat{H} - E) |\Psi\rangle = 0. \quad (1)$$

Here the Hamiltonian is given by

$$\hat{H} = \hat{T}_n + \hat{W}_{nn} + \hat{V}_{\text{ext},n} + \hat{T}_e + \hat{W}_{ee} + \hat{V}_{\text{ext},e} + \hat{W}_{en}, \quad (2)$$

where \hat{T}_n , \hat{T}_e denote the kinetic-energy operators of the nuclei and electrons, respectively, \hat{W}_{nn} , \hat{W}_{ee} and \hat{W}_{en} contain the interparticle Coulomb interactions, and $\hat{V}_{\text{ext},n}$, $\hat{V}_{\text{ext},e}$ represent the external potentials acting on the system. Using the solution of eq. (1) all observables of interest are readily evaluated from the corresponding expectation value

$$A = \langle \Psi | \hat{A} | \Psi \rangle. \quad (3)$$

While eq. (1) provides the correct starting point for the quantum mechanical treatment of any many-body system, its numerical solution becomes exceedingly difficult with increasing particle number. To see how this comes about even for relatively small finite systems consider, as an example, the nitrogen atom. Suppose that we want to store the values of the electronic ground-state wave function in a rough table containing only 10 entries for each Cartesian coordinate of the 7 electrons. This results in $10^{(7 \times 3)}$ entries for the table. Furthermore, reserving only one byte of memory per entry the data of the table will require 10^{11} DVD's for storage. Here we have assumed an ample capacity of 10^{10} bytes per DVD. Turning from finite to extended systems the situation becomes even worse. In the case of solids with particle numbers of the order of 10^{23} the task of solving eq. (1) becomes completely non-feasible. This simple example illustrates that, for sufficiently large systems, the direct numerical determination of the many-body wave function is neither possible nor desirable.

In the last decades density-functional theory (DFT) has become a very popular approach for the quantum-mechanical treatment of many-particle systems. Instead of using the complicated many-body wave function, traditional DFT deals with the electronic ground-state density

$$\rho(\mathbf{r}) = N \int d^3 r_2 \int d^3 r_3 \cdots \int d^3 r_N |\psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 \quad (4)$$

as the basic variable. Within the framework of DFT it can be shown that all observables are functionals of the density only. The next section introduces briefly the basic notions of static and time-dependent DFT (TDDFT). Having provided the formal background for the discussion we then turn to numerical aspects of the propagation in section 3. We first review the well known Crank-Nicholson^{1,2} and split-operator spectral method³ and then introduce the modifications necessary for the propagation of the time-dependent Kohn-Sham equations. Section 3.3 is devoted to a discussion on the parallelization of the time-dependent Kohn-Sham equations. Finally, in section 4, we present some results for atoms in strong laser pulses.

2 Basic Concepts of DFT

2.1 Static Density Functional Theory

Traditional ground-state DFT^{4,5} assumes the Born-Oppenheimer approximation, i.e., the nuclear motion is frozen and the nuclear centers are kept at fixed positions. The Coulomb interaction between the fixed nuclei and the electrons is described by an external potential \hat{v} acting on the electrons only. Assuming this framework, the fundamental Hohenberg-Kohn theorem of density-functional theory can be summarized by the following three statements

- (i) The external potential \hat{v} (usually due to the nuclei) is uniquely determined by the electronic ground-state density. With the knowledge of \hat{v} the complete electronic Hamiltonian, including the kinetic energy \hat{T}_e and the Coulomb repulsion of the electrons \hat{W}_{ee} , is known

$$\hat{H} = \hat{T}_e + \hat{W}_{ee} + \hat{v}. \quad (5)$$

A formal solution of the many-body Schrödinger equation can then be used in principle to evaluate expectation values of any observable of interest. Hence, any observable of a static many-body system is a functional of its ground-state density.

- (ii) Consider now a given system with a given (fixed) external potential \hat{v}_0 . Then the total-energy functional

$$E_{v_0}[\rho] = \langle \psi[\rho] | \hat{T}_e + \hat{W}_{ee} + \hat{v}_0 | \psi[\rho] \rangle \quad (6)$$

obeys the Hohenberg-Kohn variational principle: The exact ground-state energy of the interacting electronic system is obtained if and only if the exact ground-state density ρ_0 is inserted in eq. (6). For densities ρ differing from ρ_0 the following inequality holds

$$E_0 = E_{v_0}[\rho_0] < E_{v_0}[\rho]. \quad (7)$$

Therefore, the ground-state energy E_0 and density ρ_0 can be determined by minimizing the functional $E_{v_0}[\rho]$. In practice this can be achieved by solving the Euler equation

$$\frac{\delta E_{v_0}[\rho]}{\delta \rho(\mathbf{r})} = 0. \quad (8)$$

(iii) The density dependence of the functional $F[\rho]$

$$F[\rho] = \langle \psi[\rho] | \hat{T}_e + \hat{W}_{ee} | \psi[\rho] \rangle \quad (9)$$

is universal, i.e. it is the same for all systems with a fixed particle-particle interaction \hat{W}_{ee} .

For a proof of these statements the interested reader is referred to the literature.^{4,6} Since the statements of the Hohenberg-Kohn theorem are independent of the specific form of the particle-particle interaction they hold in particular for the special case of noninteracting particles, where $\hat{W}_{ee} = 0$. This leads directly to the Kohn-Sham theorem:

Given the ground-state density $\rho(\mathbf{r})$ of an interacting system, the local, i.e., multiplicative single-particle potential $v_s[\rho]$ reproducing $\rho(\mathbf{r})$ as ground-state density of a non-interacting system is uniquely determined. Hence, $\rho(\mathbf{r})$ can be calculated from the effective single-particle equations (atomic units are used throughout)

$$\left(-\frac{\nabla^2}{2} + v_s[\rho](\mathbf{r}) - \epsilon_j \right) \varphi_j(\mathbf{r}) = 0, \quad j = 1, \dots, N, \quad (10)$$

where the ground-state density ρ is obtained from the Kohn-Sham orbitals

$$\rho(\mathbf{r}) = \sum_{j=1}^N |\varphi_j(\mathbf{r})|^2. \quad (11)$$

The Kohn-Sham eqns. (10) together with statement (i) of the Hohenberg-Kohn theorem provide an efficient practical scheme for the calculation of observables of static interacting-electron systems. In this way the solution of the full many-body Schrödinger equation can be circumvented: The Kohn-Sham equations are solved with some approximation for the functional $v_s[\rho]$, the density is calculated from the Kohn-Sham orbitals and finally the density is inserted in the corresponding functionals for the observables of interest.

Conventionally, the effective single-particle potential is decomposed in the following way

$$v_s[\rho](\mathbf{r}) = v_0(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r' + \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}. \quad (12)$$

Here v_0 is the external potential, the second term describes the classical electrostatic interaction between the electrons and $v_{xc}(\mathbf{r}) = \delta E_{xc}[\rho]/\delta \rho(\mathbf{r})$ contains all exchange-correlation effects. Viewed historically, the most popular approximation for E_{xc} is the local-density approximation

$$E_{xc}[\rho] = \int \rho(\mathbf{r}) e_{xc}^{\text{unif}}(\rho(\mathbf{r})) d^3 r. \quad (13)$$

Here $e_{xc}^{\text{unif}}(\rho)$ is the exchange-correlation energy per particle of the uniform electron gas with density ρ . The solution of the Kohn-Sham equations (10) involves a self-consistency cycle. From an initial guess for the orbitals, the density is calculated using (11). Inserting the density in (12) yields an approximation for the effective single-particle potential. In the next step the Kohn-Sham equations (10) are solved, resulting in a new set of orbitals. This cycle is repeated until self-consistency is reached.

Practical implementations of the Kohn-Sham equations follow different strategies. In quantum chemistry the Kohn-Sham orbitals are usually expanded in a basis set. This turns the effective single-particle equation into a simple eigenvalue problem. Typical choices for basis functions are Gaussian-type orbitals (GTO's) as implemented in GAUSSIAN 98⁷ or Slater-type orbitals (STO's) as implemented in the Amsterdam density-functional program ADF.⁸ For infinite periodic solids, the Kohn-Sham orbitals are Bloch functions. The latter can be expanded in plane waves (usually combined with a pseudopotential treatment of the core regions) as, e.g., in the FHI⁹ code, in linearized augmented plane-waves (LAPW's) as in the WIEN2k¹⁰ or the FLEUR¹¹ codes, in linear muffin-tin orbitals (LMTO's)¹² or in local orbitals as in the SIESTA¹³ code.

2.2 Time-Dependent Density-Functional Theory

To describe interacting many-electron systems in time-dependent external fields an extension of the traditional ground-state theory is required. Recall that the ground-state theory establishes a one-to-one correspondence between ground-state densities and external potentials. In a time-dependent theory the question arises if there is also a one-to-one correspondence between time-dependent densities and time-dependent external potentials. The answer is positive and given by the Runge-Gross theorem,^{14,15} the time-dependent analogue of the Hohenberg-Kohn theorem:

Two densities $\rho(\mathbf{r}, t)$ and $\rho'(\mathbf{r}, t)$ evolving from a common initial state $\Psi_0 = \Psi(t_0)$ under the influence of two potentials $v(\mathbf{r}, t)$ and $v'(\mathbf{r}, t)$ are always different provided that the potentials differ by more than a purely time-dependent function

$$v(\mathbf{r}, t) \neq v'(\mathbf{r}, t) + c(t). \quad (14)$$

The proof of this theorem assumes that the potentials $v(\mathbf{r}, t)$ and $v'(\mathbf{r}, t)$ are both Taylor expandable in the time coordinate around the initial time t_0 .

Similar to the static case the one-to-one correspondence between time-dependent densities and time-dependent potentials can be established for arbitrary particle-particle interaction, in particular for a vanishing interaction. This ensures the uniqueness of a density-dependent single-particle potential $v_s(\mathbf{r}, t)$ which reproduces a given time-dependent density of an interacting system of interest. The time-dependent single-particle equations containing the effective potential $v_s(\mathbf{r}, t)$ are called the time-dependent Kohn-Sham (TDKS) equations

$$-i\partial_t \varphi_j(\mathbf{r}, t) = \left(-\frac{\nabla^2}{2} + v_s(\mathbf{r}, t) \right) \varphi_j(\mathbf{r}), \quad j = 1, \dots, N. \quad (15)$$

Again the density is obtained from the orbitals

$$\rho(\mathbf{r}, t) = \sum_{j=1}^N |\varphi_j(\mathbf{r}, t)|^2. \quad (16)$$

It is customary to partition the effective time-dependent potential as

$$v_s[\rho](\mathbf{r}, t) = v_0(\mathbf{r}, t) + \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d^3 r' + v_{xc}[\rho](\mathbf{r}, t). \quad (17)$$

The second term on the right is the time-dependent Hartree potential and $v_{xc}[\rho](\mathbf{r}, t)$ is the time-dependent exchange-correlation potential. The simplest approximation possible for $v_{xc}[\rho](\mathbf{r}, t)$ is the adiabatic local-density approximation

$$v_{xc}[\rho](\mathbf{r}, t)^{\text{ALDA}} = \left. \frac{d}{dn} e_{xc}^{\text{unif}}(\rho) \right|_{\rho=\rho(\mathbf{r}, t)}. \quad (18)$$

From the way of its construction it can be seen directly that this approximation is local, both in space and time. More sophisticated approximations such as the optimized-effective potential have been suggested¹⁶ which are non-local in space and time.

3 Propagation Methods for the TDKS Equations

Mathematically the solution of the time-dependent Kohn-Sham equations is an initial value problem. A given set of initial orbitals $\varphi_j(t_0)$ is propagated forward in time. No self-consistent iterations are required as in the static case. In terms of the time-evolution operator the orbitals at $t > t_0$ can be expressed as

$$\varphi_j(t) = \hat{U}(t, t_0) \varphi_j(t_0) \quad j = 1, \dots, N, \quad (19)$$

where

$$\hat{U}(t, t_0) = \hat{T} \exp \left(-i \int_{t_0}^t \hat{H}_{KS}(\tau) d\tau \right). \quad (20)$$

Note that due to the Hartree and exchange-correlation contributions the Kohn-Sham Hamiltonian is explicitly time-dependent even in the absence of a time-dependent external field. Because of this explicit dependence we have to keep the time-ordered exponential in the time-evolution operator.

The numerical task is now to find a discretized form of eq. (19). First of all let us consider the spatial representation of the orbitals. There are several discretizations possible:

- The values of the orbital are sampled on a 3D uniform Cartesian grid

$$\begin{aligned} \varphi_j(\mathbf{r}, t) &= \varphi_j(x, y, z, t) \\ &\rightarrow \varphi_j(x_k, y_l, z_m, t_n), \end{aligned} \quad (21)$$

where

$$x_k = x_0 + k \Delta x, \quad k_{\min} < k < k_{\max}, \quad (22)$$

y_l and z_m are treated similarly and

$$t_n = n \Delta t, \quad n = 0, \dots, n_{\max}. \quad (23)$$

This is the most flexible but also the computationally most expensive approach.

- If the applied external field has certain symmetries, such as a laser field linearly polarized along the z -direction, a representation in cylindrical coordinates can be advantageous. In the case of linearly polarized lasers the angular quantum number m of the orbitals is preserved so that a possible representation of the orbitals is

$$\varphi_j(\mathbf{r}, t) = \xi(\rho, z, t) e^{im\phi}. \quad (24)$$

In this case the variables ρ, z and t are discretized.

- Another frequently used approach employs spherical coordinates. Here the orbital is expanded in spherical harmonics and only the radial functions are treated on a uniform grid

$$\varphi_j(\mathbf{r}, t) = \sum_{l=0}^L \sum_{m=-l}^l R_{lm}(r, t) Y_l^m(\theta, \phi). \quad (25)$$

Adaptive grids are also possible, require however a better bookkeeping than uniform grids.

Having chosen a discretized representation for the orbitals $\varphi_j(t_0)$ we have to approximate the time-ordered exponential in eq. (20). As a first simplification we drop the time-ordering over the time step Δt

$$\varphi_j(t + \Delta t) = \exp(-i\hat{H}_{KS}(t + \Delta t/2)\Delta t) \varphi_j(t), \quad j = 1, \dots, N. \quad (26)$$

Although this seems to be an ad hoc approximation it can be shown rigorously¹⁷ that the discretization error introduced by this step is of the same order in Δt as the error introduced by the propagation schemes considered in this article. Only this fact justifies the omission of the time-ordering.

In the next two sections we review the Crank-Nicholson and split-operator schemes respectively and show how they have to be modified for the propagation of the Kohn-Sham equations.

3.1 Crank-Nicholson Propagator

Assuming a time-independent Hamiltonian the Crank-Nicholson (CN) scheme utilizes the so called Caley approximation to the time-evolution operator

$$\exp(-i\hat{H}\Delta t) = \frac{1 - i\hat{H}\Delta t/2}{1 + i\hat{H}\Delta t/2} + \mathcal{O}(\Delta t^3). \quad (27)$$

This approximation is accurate up to second order in Δt , unconditionally stable and unitary. Inserting (27) in (26) results in an implicit approximation for the unknown orbital at $t + \Delta t$. In other words a set of linear equations has to be solved in each time step

$$[1 + i\hat{H}\Delta t/2] \varphi_j(t + \Delta t) = [1 - i\hat{H}\Delta t/2] \varphi_j(t). \quad (28)$$

So far we have just considered the standard CN propagation. However, in a Kohn-Sham propagation the Hamiltonian is time-dependent. To account for this we can evaluate the Hamiltonian midway between two time steps as in (26)

$$[1 + i\hat{H}_{KS}(t + \Delta t/2)\Delta t/2] \varphi_j(t + \Delta t) = [1 - i\hat{H}_{KS}(t + \Delta t/2)\Delta t/2] \varphi_j(t). \quad (29)$$

Now a further complication appears.¹⁸ The Kohn-Sham Hamiltonian $\hat{H}_{KS}(t + \Delta t/2)$ depends via the Hartree and exchange-correlation potentials on the still unknown solutions $\varphi_j(t + \Delta t/2)$. To obtain an approximation for the Hamiltonian $\hat{H}_{KS}(t + \Delta t/2)$ we have to propagate with a two-step predictor-corrector approach. In the predictor step we use in (29) instead of $\hat{H}_{KS}(t + \Delta t/2)$ the retarded Hamiltonian $\hat{H}_{KS}(t)$

$$[1 + i\hat{H}_{KS}(t)\Delta t/2] \varphi'_j(t + \Delta t) = [1 - i\hat{H}_{KS}(t)\Delta t/2] \varphi_j(t). \quad (30)$$

From the solution $\varphi'_j(t + \Delta t)$ the corresponding Hartree and exchange-correlation potentials are constructed, leading to the Hamiltonian $\hat{H}'_{KS}(t + \Delta t)$. In the corrector step we use the average

$$\hat{H}_{KS}(t + \Delta t/2) = \frac{\hat{H}'_{KS}(t + \Delta t) + \hat{H}_{KS}(t)}{2} \quad (31)$$

as approximation to $\hat{H}_{KS}(t + \Delta t/2)$. The numerical effort introduced by this predictor-corrector scheme is doubled compared to the ordinary CN propagation. Two linear systems have to be solved in each time step. This extra effort cannot be avoided since the approximation $\hat{H}_{KS}(t + \Delta t/2) \approx \hat{H}_{KS}(t)$ would cause a continuous decrease in energy due to the use of a somewhat retarded potential. In contrast the propagation with $\hat{H}_{KS}(t + \Delta t/2) \approx \hat{H}'_{KS}(t + \Delta t)$ causes an increase in energy due to the use of an advanced potential.

3.2 Split-Operator Scheme

The split-operator (SPO) technique exploits the fact that the total Hamiltonian can be split into two parts such that each part is diagonal in either configuration or momentum space. In general one finds for two non-commuting operators \hat{A} and \hat{B} the following splittings

$$\exp((\hat{A} + \hat{B})\lambda) = \exp(\hat{A}\lambda) \exp(\hat{B}\lambda) + \mathcal{O}(\lambda^2) \quad (32)$$

$$\exp((\hat{A} + \hat{B})\lambda) = \exp\left(\hat{A}\frac{\lambda}{2}\right) \exp(\hat{B}\lambda) \exp\left(\hat{A}\frac{\lambda}{2}\right) + \mathcal{O}(\lambda^3). \quad (33)$$

This suggests the following approximations for the short-time propagator

$$\exp(-i\hat{H}\Delta t) = \exp(-i\hat{V}\Delta t) \exp(-i\hat{T}\Delta t) + \mathcal{O}(\Delta t^2) \quad (34)$$

and

$$\exp(-i\hat{H}\Delta t) = \exp\left(-i\hat{T}\frac{\Delta t}{2}\right) \exp(-i\hat{V}\Delta t) \exp\left(-i\hat{T}\frac{\Delta t}{2}\right) + \mathcal{O}(\Delta t^3). \quad (35)$$

The propagation is now performed with the following steps

$$\begin{aligned} \varphi(\mathbf{r}, t) &\xrightarrow{FT} \varphi(\mathbf{q}, t) \xrightarrow{\exp(-i\hat{T}\frac{\Delta t}{2})} \varphi'(\mathbf{q}, t) \xrightarrow{FT} \varphi'(\mathbf{r}, t) \xrightarrow{\exp(-i\hat{V}\Delta t)} \varphi''(\mathbf{r}, t) \\ &\xrightarrow{FT} \varphi''(\mathbf{q}, t) \xrightarrow{\exp(-i\hat{T}\frac{\Delta t}{2})} \varphi'''(\mathbf{q}, t) \xrightarrow{FT} \varphi(\mathbf{r}, t + \Delta t). \end{aligned}$$

By switching between momentum and configuration space each of the exponentials can be evaluated in its diagonal representation causing only multiplications with phase factors.

The propagation with the SPO can be efficiently implemented by using Fast-Fourier Transforms. As in the case of the CN the SPO is unconditionally stable and unitary.

Let us now turn to a modification of the SPO scheme which allows for the propagation of the TDKS equations. Similar as in the last section we face the same problem. The time-dependent Hamiltonian has to be evaluated midway between two time steps and depends on the unknown solutions $\varphi_j(t + \Delta t/2)$. There is an elegant way to account for this lack of information in a SPO propagation. Consider again the short-time propagator

$$\exp\left(-i\hat{H}_{KS}(t + \frac{\Delta t}{2})\Delta t\right) = \exp\left(-i\hat{T}\frac{\Delta t}{2}\right) \exp\left(-i\hat{V}_s(t + \frac{\Delta t}{2})\Delta t\right) \underbrace{\exp\left(-i\hat{T}\frac{\Delta t}{2}\right)}_{1.} + \mathcal{O}(\Delta t^3) \quad (36)$$

and think of the exponential (1.) in (36) as the first part of the low order splitting in (34) for half the time step

$$\exp\left(-i\hat{H}_{KS}(t + \frac{\Delta t}{4})\frac{\Delta t}{2}\right) = \underbrace{\exp\left(-i\hat{V}_s(t + \frac{\Delta t}{4})\frac{\Delta t}{2}\right)}_{2.} \underbrace{\exp\left(-i\hat{T}\frac{\Delta t}{2}\right)}_{1.} + \mathcal{O}(\Delta t^2). \quad (37)$$

Since we only need the orbital densities to construct the Hartree and exchange-correlation potentials we have to evaluate the absolute value squared of the orbitals in configuration space

$$\begin{aligned} \varphi_j(\mathbf{r}, t) &\xrightarrow{FT} \varphi_j(\mathbf{q}, t) \xrightarrow{\exp(-i\hat{T}\frac{\Delta t}{2})} \varphi'_j(\mathbf{q}, t) \xrightarrow{FT} \varphi'_j(\mathbf{r}, t) \xrightarrow{\exp(-i\hat{V}_s\frac{\Delta t}{2})} \varphi''_j(\mathbf{r}, t) \longrightarrow |\varphi''_j(\mathbf{r}, t)|^2 \\ \varphi'_j(\mathbf{r}, t) &\longrightarrow |\varphi'_j(\mathbf{r}, t)|^2 = |\varphi''_j(\mathbf{r}, t)|^2. \end{aligned}$$

Because the second exponential in (37) constitutes just a phase in configuration space it cancels when evaluating the absolute value. Therefore, it is sufficient to apply only the first exponential in (37). Also note that due to this cancellation the time argument of the Hamiltonian in (37) has no effect. Going back to (36) this is exactly what we have reached after the evaluation of the first exponential. Thus, we can construct a low-order approximation to $\varphi_j(t + \Delta t/2)$ or similarly to $\hat{V}(t + \Delta t/2)$ by calculating the orbital densities after the first exponential in (36). This approximation is then used in the second exponential in (36) for the unknown $\hat{V}(t + \Delta t/2)$.

Although we are reducing the order of the propagation error from $\mathcal{O}(\Delta t^3)$ to $\mathcal{O}(\Delta t^2)$ there is no extra effort required to obtain the Kohn-Sham potential midway between two time steps. The operator splitting generates the required information on the fly. This is in contrast to the adapted CN scheme of the last section where we have to double the numerical effort to be consistent in the time step. As a drawback of the SPO approach remains only the reduced order in the propagation error.

3.3 Strategies for Parallelization

One can think of several starting points for a parallelization of the Kohn-Sham time propagation using standard message-passing protocols such as the MPI standard. Depending

on the time-evolution algorithm this could be parallel FFT's in the case of the SPO or a distribution of the numerical grid over the available nodes in the case of a CN propagation. However, in both cases the implementation has to be done with great care. Grid boundaries have to be communicated between the nodes and a high information traffic is caused automatically. In contrast, the simplest and at the same time most efficient parallelization can be achieved by a distribution of the orbitals. Each node is assigned a fixed number of orbitals. The node is then computing the time evolution of the orbitals and is evaluating the partial orbital densities. In each time step the total density has to be evaluated only once from the partial densities and sent back to the nodes in order to calculate the effective single-particle potential. This approach reduces the communication traffic between the nodes to a minimum. Running such a parallelization on a cluster of modern workstations or PC's elapses on the order of 5 min per node for a time step of a single orbital. The traffic caused by the evaluation and forwarding of the total density requires on the other hand only a fraction of a second. Considering this ratio of traffic and computational load it is already sufficient to connect the nodes with a cheap 100 Mbit LAN. A further argument for this approach is the simplicity of the implementation. It took only 12 MPI commands in a code with more than 10.000 lines. This simplifies the debugging of such a code considerably.

4 Examples for the Solution of the TDKS Equations

For a given initial state the Runge-Gross theorem ensures a one-to-one mapping between time-dependent densities and time-dependent potentials. Hence, any observable of a time-dependent electronic system is a functional of the time-dependent density and the corresponding initial state. For most observables it is difficult to write down explicit approximations for the functional dependence on the time-dependent density and the initial state. However, in some cases the exact functional is known.

From a practical point of view any calculation within TDDFT is performed in two successive steps

- (i) First the TDKS equations are solved for a given initial state and the time-dependent density is evaluated from the orbitals via eq (16).
- (ii) Using the time-dependent density from step (i) and the initial state, the functional for the observable is evaluated.

Usually both steps involve approximations. To solve the TDKS equations some approximation for the exchange-correlation part of the effective single-particle potential has to be employed. Unless the exact functional for the observable is known, the second approximation enters in step (ii), where some approximate functional form for the density dependence of the observable of interest has to be assumed.

To illustrate the steps we review in the following sections two prototypical examples. First we discuss a density-functional treatment of high-harmonic generation (HHG). This constitutes a case where the exact functional is known, i.e. only approximations to $v_{xc}[\rho](\mathbf{r}, t)$ are required. The second example is the double-ionization of the Helium atom. In this case it is difficult to find explicit density functionals for the ion yield of singly and doubly ionized Helium, so that approximations in both steps (i),(ii) are involved.

4.1 High-Harmonic Generation

Even 35 years after the discovery of third-harmonic generation in a rare gas medium by New and Ward¹⁹ the subject of high-harmonic generation is still a field of active research. The interest in HHG originates mainly from the perspectives for possible applications. The process is considered as a possible candidate for the generation of coherent VUV or soft X-ray pulsed sources.

For a density-functional treatment of HHG we follow the two-step procedure described above. Taking the optimized effective potential¹⁶ as approximation for $v_{xc}[\rho](\mathbf{r}, t)$ the TDKS equations are solved for the Helium atom in a strong laser pulse

$$i \frac{\partial}{\partial t} \psi(\mathbf{r}, t) = \left(-\frac{\nabla^2}{2} - \frac{2}{r} + 2 \int \frac{|\psi(\mathbf{r}', t)|^2}{|\mathbf{r} - \mathbf{r}'|} d^3 r + v_{xc}^{\text{OEP}}(\mathbf{r}, t) + E_0 f(t) z \sin(\omega_0 t) \right) \psi(\mathbf{r}, t). \quad (38)$$

Here the laser field is treated in dipole approximation, has a frequency ω_0 , a peak intensity E_0 and is taken to be linearly polarized in z-direction. The envelope $f(t)$ of the laser pulse describes a linear ramp over the first three cycles and is then held constant for the following 15 cycles.

In the second step of the calculation we have to evaluate the functionals for the observables of interest. In the case of harmonic spectra this can be done without any further approximation. Considering only the response of a single atom and neglecting propagation effects of the generated radiation in the medium, it can be shown²⁰ that the Fourier transform of the induced dipole moment

$$d(t) = \int z \rho(\mathbf{r}, t) d^3 r \quad (39)$$

is proportional to the experimentally observed harmonic distribution. Thus, the density functional for the harmonic spectra can be written down exactly

$$S[\rho](\omega) = |d(\omega)|^2 = \left| \int \exp(i\omega t) \left(\int z \rho(\mathbf{r}, t) d^3 r \right) dt \right|^2. \quad (40)$$

Note that the functional for harmonic spectra depends only on the density. Since the calculation is started in the ground state, the dependence on the initial state drops out. For this particular choice of initial state the initial Kohn-Sham orbitals can be obtained uniquely from the ground-state density by virtue of the traditional static Hohenberg-Kohn theorem. Together with the Kohn-Sham equation in (38) the functional $S[\rho](\omega)$ in (40) provides an efficient practical scheme for the systematic exploration of harmonic spectra. Repeating the computational procedure for different laser parameters ω_0 and E_0 or different envelopes $f(t)$ optimal conditions for the generation of high harmonics can be found. For example, by running different simulations, it turned out that two-color laser fields

$$E(t) = f(t)[E_0 \sin(\omega_0 t) + E_1 \sin(\omega_1 t + \delta)] \quad (41)$$

increase the efficiency of high-harmonic generation considerably. Typically the field strengths E_0 and E_1 are chosen to be of the same order and ω_1 is taken to be an integer multiple of ω_0 . Possible phase shifts between the two fields are taken into account by the constant δ . In Fig. 1 we show the harmonic spectra for intensities $E_0 = E_1 = 0.01$ a.u., where the laser frequency ω_1 was adjusted to the second or third harmonic of the fundamental frequency $\omega_0 = 0.0740$ a.u. which corresponds to a wavelength of $\lambda = 616$ nm. The calculation shows that the harmonics from a two-color laser pulse can be more intense up to two orders of magnitude compared to a single color pulse.²¹ Such results may be used to guide the experimental work in the search for coherent soft X-ray sources.

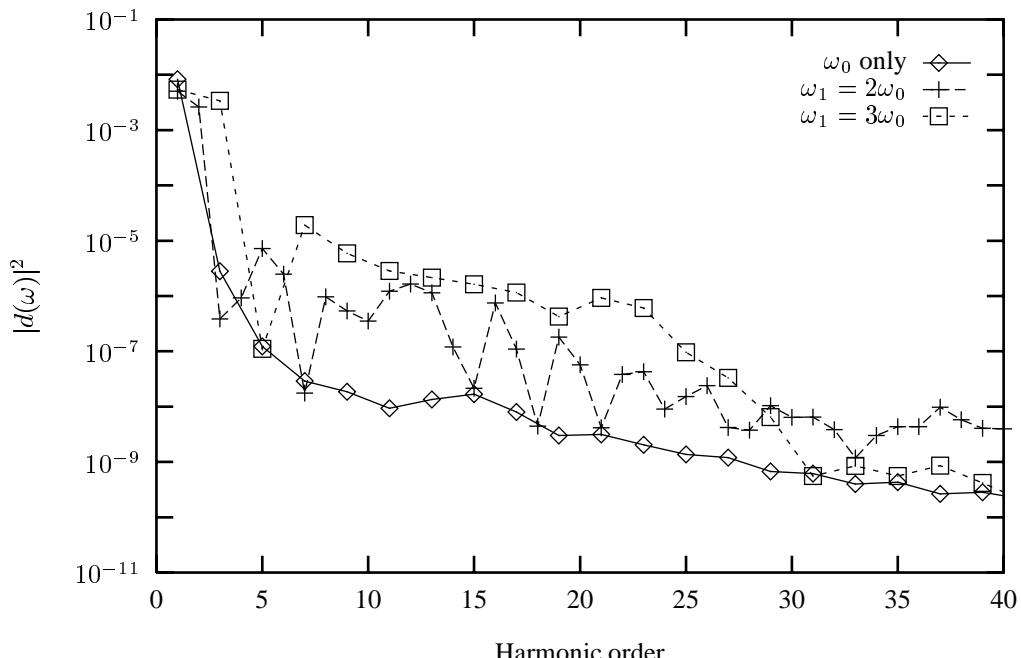


Figure 1. Harmonic spectra of helium calculated for one-color and two-color laser pulses with a total intensity of $I = 7.0 \cdot 10^{14}$ W/cm² respectively. The frequency of the second color has been adjusted to the second or third harmonic of the fundamental frequency $\omega_0 = 0.0740$ a.u. ($\lambda = 616$ nm).

4.2 Helium Double Ionization

When atoms or molecules are exposed to strong laser pulses there are three basic routes to analyze the laser-matter interaction. The first way is the observation of emitted harmonic light as discussed in the last section. The second route is to measure the kinetic energy and the angular distribution of the ionized electrons (photoelectrons) leaving the laser focus and the last possibility is to count the produced ions or to measure their momentum distribution. In this section we discuss a density-functional calculation of ion-yields for singly and doubly ionized atoms or molecules.

Considering laser pulses with a relatively long rise time it is well known²² that the ionization dynamics is dominated by a “sequential” emission of electrons. With growing intensity the ionization yield of the singly ionized species grows according to the power law of the lowest non-vanishing order of perturbation theory. Only when a depletion of the neutral species starts to arise, an appreciable yield of doubly-ionized species can be detected. These properties of the ion yields can be described successfully with the single-active electron approximation:^{23,24} a stepwise scenario is assumed where the ionization occurs by a sequential emission of the electrons. The situation is different in the regime of ultra-short laser pulses. Experimentally it has been demonstrated in high-precision measurements²⁵ that double-ionization yields are up to six orders of magnitude larger than the rates expected in a sequential process. This is a clear manifestation of electron-electron correlation. Single-active electron calculations, by construction, will fail to describe the correlations involved in the ionization process. Time-dependent density-functional theory on the other hand is in principle capable to give the exact ionization yields, provided the exact functionals in the two-step procedure from above are known. For step (i) the familiar ALDA or TDOEP functionals can be used, so that only approximations for the ion-yields of singly P^{+1} and doubly P^{+2} ionized species have to be found. Using a geometrical concept that relies on the spatial partitioning of the wave function it is possible^{26,27} to find approximations for P^{+1} and P^{+2} in terms of the time-dependent density

$$\begin{aligned} P[\rho]^{+1}(t) &= \int_A d^3r \rho(\mathbf{r}, t) - \int_A d^3r_1 \int_A d^3r_2 \rho(\mathbf{r}_1, t) \rho(\mathbf{r}_2, t) g[\rho](\mathbf{r}_1, \mathbf{r}_2, t) \\ P[\rho]^{+2}(t) &= 1 - \int_A d^3r \rho(\mathbf{r}, t) + \frac{1}{2} \int_A d^3r_1 \int_A d^3r_2 \rho(\mathbf{r}_1, t) \rho(\mathbf{r}_2, t) g[\rho](\mathbf{r}_1, \mathbf{r}_2, t). \end{aligned} \quad (42)$$

Here it was assumed that the Kohn-Sham propagation starts in the ground-state so that the dependence on the initial state drops out, similar to the case of harmonic spectra. In eq. (42), $g[\rho]$ denotes the pair-correlation function. This quantity is a density functional which, in practice, needs to be approximated. Following the same procedure as in section 4.1, the TDKS equation (38) is solved for the Helium atom in a strong laser pulse. From the resulting time-dependent density the functionals (42) are evaluated. Although this scheme provides a considerable improvement over sequential ionization yields it shows still a discrepancy of about two orders of magnitude when the results of the calculation are compared directly to experiment.²⁶

To test the relative importance of the two approximations involved in steps (i) and (ii), Lappas and van Leeuwen²⁸ have performed numerically exact time-propagations for a 1D soft core model of the Helium atom. Using the correlated Helium wave function, obtained in their simulation, exact reference values have been obtained for the ionization yields P^{+1} and P^{+2} of this model system. Since it is also possible to obtain the exact time-dependent density from the correlated Helium wave function, approximations involved in the first step (i) of the computational procedure can be circumvented. By evaluating the functionals for P^{+1} and P^{+2} (42) with the exact density the resulting approximate yields can be compared with the exact reference values and the performance of the functionals P^{+1} and P^{+2} can be tested directly. Lappas and van Leeuwen find that, although the double-ionization yields obtained in this way still show discrepancies, they reproduce the well known knee structure^{25,29} known from experiment. Since the knee cannot be obtained from the same

functional P^{+2} when approximate densities from TDOEP calculations are inserted, the approximations employed for the effective single particle potential $v_s[\rho]$ appear to have a bigger impact on the results.

References

1. *A comparison of different propagation schemes for the time-dependent Schrödinger equation*, C. Leforestier et al., J. Comput. Phys. **94**, 59-80 (1991).
2. *Quantum dynamics of the collinear (H, H_2) reaction*, E. A. McCullough, Jr. and R. E. Wyatt, J. Chem. Phys. **51**, 1253 (1969); **54**, 3592 (1971).
3. *Solution of the Schrödinger equation by a spectral method*, M. D. Feit, J. A. Fleck, Jr., and A. Steiger, J. Comput. Phys. **47**, 412-433 (1982).
4. *Inhomogeneous electron gas*, P. Hohenberg and W. Kohn, Phys. Rev. **136**, B 864 (1964).
5. *Self-consistent equations including exchange and correlation effects*, W. Kohn and L. J. Sham, Phys. Rev. **140**, A 1133 (1965).
6. *Density Functional Theory*, R. M. Dreizler and E. K. U. Gross, (Springer-Verlag, Berlin, 1990).
7. <http://www.gaussian.com>
8. <http://www.scm.com>
9. <http://www.fhi-berlin.mpg.de/th/fhimd/>
10. <http://www.wien2k.at>
11. <http://www.flapw.de>
12. <http://www.mpi-stuttgart.mpg.de/andersen/LMTODOC/LMTODOC.html>
13. <http://www.uam.es/departamentos/ciencias/fismateriac/siesta/>
14. *Density-functional theory for time-dependent systems*, E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 997 (1984).
15. *Density-functional theory of time-dependent phenomena*, E. K. U. Gross, J.F.Dobson and M.Petersilka in: *Topics in Current Chemistry: Density Functional Theory*. (ed.) R.F.Nalewajski, Springer, (1996).
16. *Time-dependent optimized effective potential*, C. A. Ullrich, U. J. Gossman, and E. K. U. Gross, Phys. Rev. Lett. **74**, 872 (1995).
17. *On the exponential form of time-displacement operators in quantum mechanics*, P. Pechukas and J. C. Light, J. Chem. Phys. **44**, 3897 (1966).
18. *One-dimensional nuclear dynamics in the time-dependent Hartree-Fock approximation* P. Bonche, S. Koonin, and J. W. Negele, Phys. Rev. C **13**, 1226 (1976).
19. *Optical third-harmonic generation in gases by a focused laser beam*, J. F. Ward and G. H. C. New, Phys. Rev. A **185**, 57 (1969).
20. *High-order harmonic generation: Simplified model and relevance of single-atom theories to experiment*, B. Sundaram, P. Milonni, Phys. Rev. A **41**, 6571 (1990).
21. *High harmonic generation in hydrogen and helium atoms subject to one- and two-color laser pulses*, S. Erhard and E. K. U. Gross, in: *Multiphoton Processes 1996*, P. Lambropoulos and H. Walther, ed(s), (IOP, 1997), p 37.
22. *Mechanisms for Multiple Ionization of Atoms by Strong Pulsed Lasers*, P. Lambropoulos, Phys. Rev. Lett. **55**, 2141 (1985).

23. *High-order harmonic generation from atoms and ions in the high intensity regime*, J. L. Krause, K. J. Schafer, and K. C. Kulander, Phys. Rev. Lett. **68**, 3535 (1992).
24. *Calculation of photoemission from atoms subject to intense laser fields*, J. L. Krause, K. J. Schafer, and K. C. Kulander, Phys. Rev. A **45**, 4998 (1992).
25. *Precision Measurement of Strong Field Double Ionization of Helium*, B. Walker, B. Sheehy, L. F. DiMauro, P. Agostini, K. J. Schafer, and K. C. Kulander, Phys. Rev. Lett. **73**, 1227 (1994).
26. *Strong-field double ionization of Helium, a density functional perspective*, M. Petersilka and E.K.U. Gross, Laser Physics **9**, 105 (1999).
27. *Ten topical questions in time-dependent density functional theory*, N. T. Maitra, K. Burke, H. Appel, E. K. U. Gross, and R. van Leeuwen, to appear in Reviews in Modern Quantum Chemistry: A Celebration of the Contributions of R. G. Parr, ed. K. D. Sen. (World Scientific,2001).
28. *Electron correlation effects in the double-ionization of He*, D. G. Lappas and R. van Leeuwen, J. Phys. B: At. Mol. Opt. Phys. **31**, L249-L256 (1998).
29. *Observation of nonsequential double ionization of helium with optical tunneling*, D. N. Fittinghof, P. R. Bolton, B. Chang and K. C. Kulander, Phys. Rev. Lett. **69**, 2642 (1992).

Path Integration via Molecular Dynamics

Mark E. Tuckerman

Department of Chemistry and Courant Institute of Mathematics
New York University, New York, NY 10003, USA
E-mail: mark.tuckerman@nyu.edu

1 Introduction

The formulation by Feynman of quantum statistical mechanics in terms of path integrals¹ has had a considerable impact on our ability to analyze the properties of quantum many-body systems at finite temperature. Not only do path integrals possess mathematical elegance, but they can be rendered into a computationally tractable form with an inherent structure perfectly suited for implementation on modern day parallel computing architectures. Thus, it is possible to calculate numerous equilibrium observable properties, including both thermodynamic and structural quantities, in a computationally efficient manner compared to other quantum approaches. Moreover, path integrals can be formulated in a variety of statistical ensembles, including the canonical (NVT), isothermal-isobaric (NPT), grand-canonical (μ VT), and others, allowing a variety of external conditions to be investigated.

In spite of the power of path integrals, several important outstanding issues remain unresolved. First is the difficulty associated with the calculation of dynamical properties, such as time correlation functions (from which transport properties and spectra are obtained) and rates. Second is the problem of treating many-fermion systems. Both of these involve what is referred to as the “sign problem” which arises from a need to compute averages of rapidly oscillating phase functions. Several of the lectures in this series will describe this problem and progress that has been made (see lectures by D. M. Ceperley, R. Ramirez, R. Egger, and S. Sorella). See also, for example, references 2-5 .

Path integration involves the calculation of many-dimensional integrals. Traditionally, path integration has been performed using Monte Carlo (MC) techniques.⁶⁻⁹ In principle, it is also possible to use molecular dynamics (MD), i.e. the numerical solution of Newtonian type equations of motion, to compute a path integral. However, as was shown by Hall and Berne,¹⁰ the evaluation of path integrals by MD is beset with difficulties related to the non-ergodic nature of the trajectories generated. (The source of the ergodicity problem will be made manifestly clear later when we explore the mathematical structure of the discretized path integral.) Only recently has an MD algorithm that explicitly treats the ergodicity problem become available,^{11,12} making the evaluation of path integrals by MD methods almost as efficient as an optimized MC approach. There are a number of reasons to prefer MD over MC in the evaluation of path integrals. First, MD moves are complete system moves, and, therefore, MD approaches are considerably easier to implement on parallel computing architectures. Second, as will be seen in the second lecture, the combination of path integration with the Car-Parrinello *ab initio* molecular dynamics,¹³ in which internuclear

forces are computed “on the fly” from electronic structure calculations, requires the use of an MD propagation algorithm. Third, MD allows the use of efficient adiabatic propagation schemes in the calculation of approximate quantum dynamical properties via the so called “centroid dynamics” approach,^{14–17} to be discussed in this lecture.

In this lecture, the path integral formulation of the canonical density matrix and quantum partition function will be introduced and a MD based algorithm for performing the path integral will be discussed in detail. In particular, it will be shown that the introduction of a variable transformation in the path integral in conjunction with an appropriate thermostating scheme can yield a very efficient MD approach. Applications to a number of illustrative example problems will be presented.

2 The Density Matrix and Quantum Statistical Mechanics

In quantum statistical mechanics, one considers an ensemble of systems described by Hamiltonian H , with each member of the ensemble possessing a state vector $|\Psi^{(\kappa)}\rangle$, $\kappa = 1, \dots, Z$ in the Hilbert space that evolves in time according to the Schrödinger equation. Z represents the number of members in the ensemble. Let $\{|\phi_i\rangle\}$ be a complete orthonormal set of vectors in the Hilbert space. Then each state vector can be expanded in this set according to

$$|\Psi^{(\kappa)}\rangle = \sum_i C_i^{(\kappa)} |\phi_i\rangle \quad (1)$$

In order to compute the expectation value of any observable represented by an operator A , one must compute the expectation value of A for each system in the ensemble $\langle\Psi^{(\kappa)}|A|\Psi^{(\kappa)}\rangle$ and average over the members of the ensemble. Thus,

$$\langle A \rangle = \frac{1}{Z} \sum_{\kappa=1}^Z \langle \Psi^{(\kappa)} | A | \Psi^{(\kappa)} \rangle \quad (2)$$

Substituting Eq.(1) into Eq. (2) gives

$$\langle A \rangle = \sum_{i,j} \left(\frac{1}{Z} \sum_{\kappa=1}^Z C_j^{(\kappa)*} C_i^{(\kappa)} \right) \langle \phi_j | A | \phi_i \rangle \quad (3)$$

where the term in parentheses represents the ensemble average of the product of expansion coefficients. Inspection of this formula reveals that the expectation value can be written as the trace of a matrix product:

$$\langle A \rangle = \sum_{i,j} \rho_{ij} A_{ji} = \text{Tr}(\rho A) \quad (4)$$

where $A_{ji} = \langle \phi_j | A | \phi_i \rangle$, and the matrix ρ_{ij} , known as the *density matrix*, is given by

$$\rho_{ij} = \frac{1}{Z} \sum_{\kappa=1}^Z C_i^{(\kappa)} C_j^{(\kappa)*} \quad (5)$$

The corresponding operator ρ with matrix elements $\rho_{ij} = \langle \phi_i | \rho | \phi_j \rangle$ can be written in the form

$$\rho = \frac{1}{Z} \sum_{\kappa=1}^Z |\Psi^{(\kappa)}\rangle \langle \Psi^{(\kappa)}| \quad (6)$$

i.e., the ensemble average of the projection operator onto the state vector of each member of the ensemble. From Eq. (5), it is clear that ρ is hermitian, so that it can be diagonalized with a set of real eigenvalues w_i and a complete set of orthonormal vectors $|w_i\rangle$. Conditions on the eigenvalues w_i can be derived from Eq. (4) by setting $A = 1$. Then,

$$\text{Tr}(\rho) = \sum_i w_i = \frac{1}{Z} \sum_{\kappa=1}^Z \langle \Psi^{(\kappa)} | \Psi^{(\kappa)} \rangle = 1 \quad (7)$$

since $|\Psi^{(\kappa)}\rangle$ is normalized to 1 for each member of the ensemble. Also, by letting $A = |w_j\rangle \langle w_j|$, we find that

$$\text{Tr}(\rho |w_j\rangle \langle w_j|) = \sum_{i,k} w_i \langle w_k | w_i \rangle \langle w_i | w_j \rangle \langle w_j | w_k \rangle = w_j \quad (8)$$

and also that

$$\langle A \rangle = \frac{1}{Z} \sum_{\kappa=1}^Z \langle \Psi^{(\kappa)} | w_j \rangle \langle w_j | \Psi^{(\kappa)} \rangle = \frac{1}{Z} \sum_{\kappa=1}^Z |\langle w_j | \Psi^{(\kappa)} \rangle|^2 \geq 0 \quad (9)$$

Therefore, $w_j \geq 0, \forall j$. Combining these two results gives the following properties satisfied by the eigenvalues of ρ

$$\begin{aligned} 0 &\leq w_i \leq 1, & \forall i \\ \sum_i w_i &= 1 \end{aligned} \quad (10)$$

i.e., the eigenvalues w_i of ρ have the properties of probabilities.

An equation of motion for the density matrix can be derived by introducing the quantum time evolution operator $e^{-iHt/\hbar}$ and using it to evolve the state vectors in time:

$$|\Psi^{(\kappa)}(t)\rangle = e^{-iHt/\hbar} |\Psi^{(\kappa)}(0)\rangle \quad (11)$$

Thus,

$$\begin{aligned} \rho(t) &= \frac{1}{Z} \sum_{\kappa=1}^Z e^{-iHt/\hbar} |\Psi^{(\kappa)}(0)\rangle \langle \Psi^{(\kappa)}(0)| e^{iHt/\hbar} \\ &= e^{-iHt/\hbar} \rho e^{iHt/\hbar} \end{aligned} \quad (12)$$

Differentiating both sides with respect to time gives the equation of motion:

$$\frac{\partial \rho}{\partial t} = -\frac{i}{\hbar} (H\rho - \rho H) = -\frac{i}{\hbar} [H, \rho(t)] \quad (13)$$

Note the difference in sign from the usual Heisenberg equation $dA/dt = i[H, A]/\hbar$.

In equilibrium, the density matrix must have no explicit time dependence, i.e., $\partial\rho/\partial t = 0$. Eq. (13) implies that $[H, \rho] = 0$, so that ρ can be expressed as a pure function of H and can therefore be diagonalized simultaneously with H . Thus,

$$\rho = f(H) = \sum_i f(E_i) |E_i\rangle\langle E_i| \quad (14)$$

where E_i and $|E_i\rangle$ are the eigenvalues and eigenvectors of the Hamiltonian, respectively. Thus, the eigenvalues w_i of ρ are pure functions of the eigenvalues of H , however, they must still satisfy the properties Eq. (10). The particular form of the function $f(E_i)$ determines which particular ensemble ρ represents. For the canonical, or NVT, ensemble, one of the most commonly used ensembles, $f(E_i)$ takes the form

$$f(E_i) = \frac{e^{-\beta E_i}}{Z} \quad (15)$$

where $\beta = 1/k_B T$. The normalization constant Z insures that $\text{Tr } \rho = 1$ and is given by

$$Z = \sum_i e^{-\beta E_i} = \text{Tr } [e^{-\beta H}] \quad (16)$$

Z is the canonical partition function, which determines the thermodynamics of the ensemble. The canonical density matrix therefore takes the form

$$\rho = \frac{e^{-\beta H}}{Z} \quad (17)$$

and the expectation value of an operator A , according to Eq. (4), is computed from

$$\langle A \rangle = \frac{1}{Z} \text{Tr } [A e^{-\beta H}] \quad (18)$$

We see, therefore, that in order to study the properties of systems in the canonical ensemble, we need to be able to evaluate traces such as occur in Eqs. (16) and (18). In the next section, we shall see how such traces can be expressed in terms of path integrals.

3 Path Integral Formulation of the Canonical Density Matrix and Partition Function

Consider a single quantum particle of mass m , with momentum, p and coordinate, x , in a one-dimensional potential $\phi(x)$ described by a Hamiltonian:

$$H = \frac{p^2}{2m} + \phi(x) \equiv K + \Phi \quad (19)$$

where K and Φ are the kinetic and potential operators, respectively. In statistical mechanics, one is interested in an ensemble of such systems, and, if the ensemble is characterized by a constant temperature, T , and volume (length in one dimension), then the ensemble is the canonical ensemble for which the density matrix is given by

$$\rho = e^{-\beta H} \quad (20)$$

where $\beta = 1/kT$. The partition function, $Z(\beta)$, is the trace of ρ . The starting point for the derivation of the path integral is the evaluation of this trace in the coordinate basis,

which is valid since the trace is independent of the basis in which it is evaluated. Since the coordinate basis is a continuous basis, the trace is calculated as an integral:

$$\begin{aligned} Z(\beta) &= \text{Tr}(e^{-\beta H}) = \int dx \langle x | e^{-\beta H} | x \rangle \\ &= \int dx \langle x | e^{-\beta(K+\Phi)} | x \rangle \end{aligned} \quad (21)$$

In general, the operators, K and Φ do not commute, so that the exponential, $\exp[-\beta(K + \Phi)]$ cannot be evaluated directly. However, use can be made of the Trotter theorem, which states that for any two operators, A and B , which, in general do not commute,

$$e^{\lambda(A+B)} = \lim_{P \rightarrow \infty} \left[e^{\frac{\lambda}{2P}B} e^{\frac{\lambda}{P}A} e^{\frac{\lambda}{2P}B} \right]^P \quad (22)$$

The proof of the Trotter theorem is rather involved and will not be given here, however, the interested reader is referred, for example, to the book by L. Schulman,¹⁸ where a discussion of the proof is given. Substituting the Trotter theorem into Eq. (21) yields

$$Z(\beta) = \lim_{P \rightarrow \infty} \int dx \langle x | \left[e^{-\frac{\beta}{2P}\Phi} e^{-\frac{\beta}{P}K} e^{-\frac{\beta}{2P}\Phi} \right]^P | x \rangle \quad (23)$$

Define an operator, Ω

$$\Omega = e^{-\frac{\beta}{2P}\Phi} e^{-\frac{\beta}{P}K} e^{-\frac{\beta}{2P}\Phi} \quad (24)$$

so that the partition function can be written

$$Z(\beta) = \lim_{P \rightarrow \infty} \int dx \langle x | \Omega^P | x \rangle = \lim_{P \rightarrow \infty} \int dx \langle x | \Omega \cdot \Omega \cdots \Omega | x \rangle \quad (25)$$

Equation (25) involves a product of P factors of the operator, Ω . Since Ω involves a product of three separate exponentials, it is possible (as we shall see shortly) to evaluate the coordinate-space matrix elements $\langle x | \Omega | x' \rangle$ of this operator analytically. Therefore, in order to obtain an expression that involves these matrix elements, we introduce an identity operator between each pair of factors of Ω in Eq. (25) in the form of a closure or completeness relation of the coordinate-space eigenvectors, $|x\rangle$:

$$I = \int dx |x\rangle\langle x| \quad (26)$$

Since there P operators in the product in Eq. (25), $P - 1$ such insertions are possible. Labeling the integrations as x_2, \dots, x_P and changing the integration variable "x" in Eq. (25) to " x_1 " yields

$$\begin{aligned} Z(\beta) &= \lim_{P \rightarrow \infty} \int dx_1 \cdots dx_P \langle x_1 | \Omega | x_2 \rangle \langle x_2 | \Omega | x_3 \rangle \langle x_3 | \cdots | x_P \rangle \langle x_P | \Omega | x_1 \rangle \\ &= \lim_{P \rightarrow \infty} \int dx_1 \cdots dx_P \left[\prod_{i=1}^P \langle x_i | \Omega | x_{i+1} \rangle \right]_{x_{P+1}=x_1} \end{aligned} \quad (27)$$

where the condition, $x_{P+1} = x_1$ is a result of the trace.

We now need to evaluate the coordinate space matrix elements of Ω :

$$\langle x_i | \Omega | x_{i+1} \rangle = \langle x_1 | e^{-\frac{\beta}{2P}\Phi} e^{-\frac{\beta}{P}K} e^{-\frac{\beta}{2P}\Phi} | x_{i+1} \rangle \quad (28)$$

Note that the potential operators are functions of x alone, and they are acting directly on coordinate eigenstates. Thus, pulling out the corresponding eigenvalues, we have

$$\langle x_i | \Omega | x_{i+1} \rangle = e^{-\frac{\beta}{2P}\phi(x_i)} \langle x_i | e^{-\frac{\beta}{P}K} | x_{i+1} \rangle e^{-\frac{\beta}{2P}\phi(x_{i+1})} \quad (29)$$

The coordinate space matrix elements of $\exp(-\beta K/P)$ can be evaluated by introducing another completeness relation for momentum eigenstates,

$$I = \int dp |p\rangle \langle p| \quad (30)$$

which allows the matrix elements to be written as

$$\langle x_i | e^{-\frac{\beta}{P}K} | x_{i+1} \rangle = \int dp \langle x_i | p \rangle \langle p | e^{-\frac{\beta}{P}K} | x_{i+1} \rangle \quad (31)$$

In the above expression, $K = p^2/2m$ now acts on one of its eigenstates from the left, yielding:

$$\langle x_i | e^{-\frac{\beta}{P}K} | x_{i+1} \rangle = \int dp e^{-\beta p^2/2mP} \langle x_i | p \rangle \langle p | x_{i+1} \rangle \quad (32)$$

Using the following relation for the inner product of coordinate and momentum eigenstates:

$$\langle x | p \rangle = \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar} \quad (33)$$

we find

$$\langle x_i | e^{-\frac{\beta}{P}K} | x_{i+1} \rangle = \frac{1}{2\pi\hbar} \int dp e^{-\beta p^2/2mP} e^{ip(x_i - x_{i+1})/\hbar} \quad (34)$$

Performing the momentum integral by completing the square, the above matrix element becomes

$$\langle x_i | e^{-\frac{\beta}{P}K} | x_{i+1} \rangle = \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{1/2} \exp \left[-\frac{mP}{2\beta\hbar^2} (x_{i+1} - x_i)^2 \right] \quad (35)$$

Substituting Eq. (35) into Eq. (29) gives the following expression for the matrix elements of Ω :

$$\langle x_i | \Omega | x_{i+1} \rangle = \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{1/2} \exp \left[-\frac{mP}{2\beta\hbar^2} (x_{i+1} - x_i)^2 - \frac{\beta}{2P} (\phi(x_i) + \phi(x_{i+1})) \right] \quad (36)$$

Finally, substituting Eq. (36) into Eq. (27) yields, for the canonical partition function

$$\begin{aligned} Z(\beta) &= \lim_{P \rightarrow \infty} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \int dx_1 \cdots dx_P \\ &\quad \exp \left\{ - \sum_{i=1}^P \left[\frac{mP}{2\beta\hbar^2} (x_{i+1} - x_i)^2 + \frac{\beta}{P} \phi(x_i) \right] \right\}_{x_{P+1}=x_1} \end{aligned} \quad (37)$$

where the fact that $(\beta/2P) \sum_{i=1}^P (\phi(x_i) + \phi(x_{i+1})) = (\beta/P) \sum_{i=1}^P \phi(x_i)$ since $x_{P+1} = x_1$. Eq. (37) involves the limit of a P -dimensional integral known as the *discretized path integral* for the partition function, often denoted simply as $Z_P(\beta)$:

$$Z_P(\beta) = \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \int dx_1 \cdots dx_P \exp \left\{ - \sum_{i=1}^P \left[\frac{mP}{2\beta\hbar^2} (x_{i+1} - x_i)^2 + \frac{\beta}{P} \phi(x_i) \right] \right\}_{x_{P+1}=x_1} \quad (38)$$

so that $Z(\beta) = \lim_{P \rightarrow \infty} Z_P(\beta)$. Before going on to describe how to evaluate Eq. (38) by MD methods, some discussion on the discretized path integral is in order.

4 The Continuous Limit

The analysis of the preceding section showed how to obtain the partition function as a discrete path integral. Here we shall show how to obtain the $P \rightarrow \infty$ limit and provide an interpretation of the result. Let us begin by extending the result of Eq. (38) to a general density matrix element. By a similar procedure, it can be shown that general matrix elements of the density matrix, $\langle x | \exp(-\beta H) | x' \rangle$ are given by

$$\begin{aligned} \langle x | e^{-\beta H} | x' \rangle &= \lim_{P \rightarrow \infty} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} e^{-\frac{\beta}{2P}(\phi(x)+\phi(x'))} \\ &\times \int dx_2 \cdots dx_P \exp \left\{ -\frac{mP}{2\beta\hbar^2} \sum_{i=1}^P (x_{i+1} - x_i)^2 - \frac{\beta}{P} \sum_{i=2}^P \phi(x_i) \right\}_{x_1=x, x_{P+1}=x'} \end{aligned} \quad (39)$$

Interestingly, an expression for the matrix elements of another similar exponential operator, the quantum time evolution operator $\exp(-iHt/\hbar)$ can be obtained from Eq. (39) by setting $\beta = it/\hbar$:

$$\begin{aligned} \langle x | e^{-iHt/\hbar} | x' \rangle &= \lim_{P \rightarrow \infty} \left(\frac{mP}{2\pi i t \hbar} \right)^{P/2} e^{-\frac{it}{2P\hbar}(\phi(x)+\phi(x'))} \\ &\times \int dx_2 \cdots dx_P \exp \left\{ \frac{imP}{2t\hbar} \sum_{i=1}^P (x_{i+1} - x_i)^2 - \frac{it}{P\hbar} \sum_{i=2}^P \phi(x_i) \right\}_{x_1=x, x_{P+1}=x'} \end{aligned} \quad (40)$$

Let us focus our attention on Eq. (40) for a short while. We shall return to the density matrix and partition function afterward. In order to obtain the continuous limit of Eq. (40), we introduce a parameter

$$\epsilon = \frac{t}{P} \quad (41)$$

so that the above path integral expression becomes

$$\begin{aligned} \langle x | e^{-iHt/\hbar} | x' \rangle &= \lim_{P \rightarrow \infty} \left(\frac{m}{2\pi i \epsilon \hbar} \right)^{P/2} e^{-\frac{i\epsilon}{2\hbar}(\phi(x)+\phi(x'))} \\ &\times \int dx_2 \cdots dx_P \exp \left\{ \frac{im}{2\epsilon\hbar} \sum_{i=1}^P (x_{i+1} - x_i)^2 - \frac{i\epsilon}{\hbar} \sum_{i=2}^P \phi(x_i) \right\}_{x_1=x, x_{P+1}=x'} \end{aligned} \quad (42)$$

or, multiplying and dividing by ϵ in the first term in the exponential,

$$\begin{aligned} \langle x | e^{-iHt/\hbar} | x' \rangle &= \lim_{P \rightarrow \infty} \left(\frac{m}{2\pi i \epsilon \hbar} \right)^{P/2} e^{-\frac{i\epsilon}{2\hbar}(\phi(x) + \phi(x'))} \\ &\times \int dx_2 \cdots dx_P \exp \left\{ \frac{i\epsilon}{\hbar} \sum_{i=1}^P \frac{m}{2} \left(\frac{x_{i+1} - x_i}{\epsilon} \right)^2 - \frac{i\epsilon}{\hbar} \sum_{i=2}^P \phi(x_i) \right\}_{x_1=x, x_{P+1}=x'} \end{aligned} \quad (43)$$

The limit $P \rightarrow \infty$ is equivalent to the limit $\epsilon \rightarrow 0$. Now, the points, x_1, \dots, x_P can be thought of as specific points of a continuous function $x(s)$, $s \in [0, t]$ such that $x_i = x((i-1)\epsilon)$ with $x(0) = x$ and $x(t) = x'$ as illustrated in the figure below: In

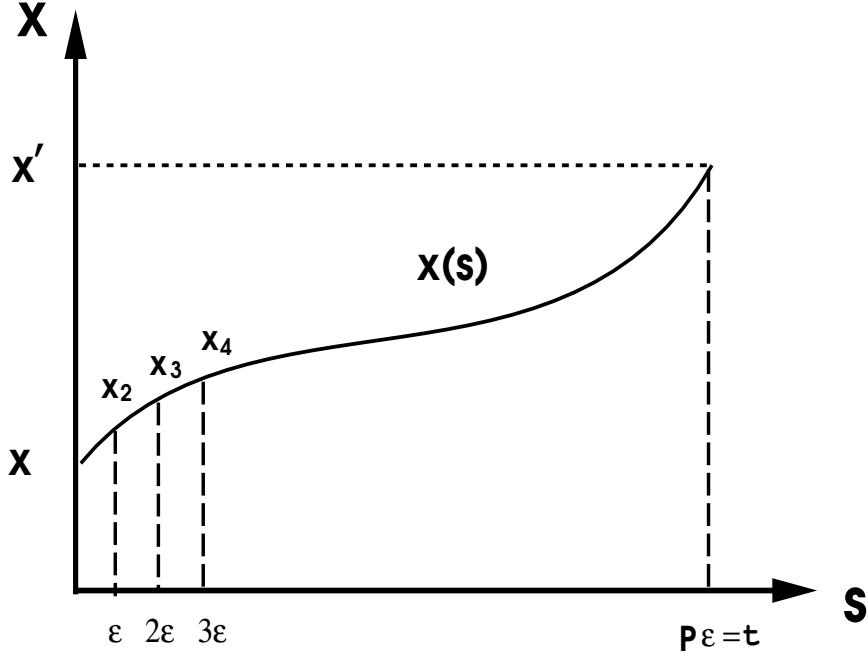


Figure 1. Illustration of the discretized path integral

in this picture, several quantities appearing in Eq. (43) have limits that can be easily recognized. For example, the first term in the exponential will be recognized as the derivative with respect to s of $x(s)$:

$$\lim_{\epsilon \rightarrow 0} \frac{x_{i+1} - x_i}{\epsilon} = \lim_{\epsilon \rightarrow 0} \frac{x(i\epsilon) - x((i-1)\epsilon)}{\epsilon} = \frac{dx}{ds} \quad (44)$$

In addition, when the $\epsilon \rightarrow 0$ limit is taken of the sum appearing in the exponential, the result will be recognized as a Riemann sum or trapezoidal rule type of expression for a

continuous integral over s :

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} & \left[\frac{i\epsilon}{\hbar} \sum_{i=1}^P \frac{m}{2} \left(\frac{x_{i+1} - x_i}{\epsilon} \right)^2 - \frac{i\epsilon}{2\hbar} (\phi(x) + \phi(x')) - \frac{i\epsilon}{\hbar} \sum_{i=2}^P \phi(x_i) \right] \\ & = \frac{i}{\hbar} \int_0^t ds \left[\frac{m}{2} \left(\frac{dx}{ds} \right)^2 - \phi(x(s)) \right]. \end{aligned} \quad (45)$$

Finally, consider the integration measure

$$\left(\frac{m}{2\pi i\epsilon\hbar} \right)^{P/2} dx_1 \cdots dx_P.$$

As $P \rightarrow \infty$ and $\epsilon \rightarrow 0$, the number of points becomes infinite, and they become infinitely closely spaced. As Fig. 1 suggests, then integration becomes an integration over *all* continuous functions, $x(s)$ that begin at x and end at x' . A special notation is introduced to represent this “integration over all functions”:

$$\lim_{P \rightarrow \infty, \epsilon \rightarrow 0} \left(\frac{m}{2\pi i\epsilon\hbar} \right)^{P/2} dx_1 \cdots dx_P \equiv \mathcal{D}x(s). \quad (46)$$

Combining Eq. (45) and Eq. (46) gives the following expression for the matrix element of the propagator:

$$\langle x | e^{-iHt/\hbar} | x' \rangle = \int_x^{x'} \mathcal{D}x(s) \exp \left\{ \frac{i}{\hbar} \int_0^t ds \left[\frac{m}{2} \left(\frac{dx}{ds} \right)^2 - \phi(x(s)) \right] \right\} \quad (47)$$

Eq. (47) is known as the *functional integral representation* of the path integral. In effect, it represents an integration over all functions, $x(s)$ with the boundary conditions $x(0) = x$ and $x(t) = x'$ and a weight given by the exponential appearing in Eq. (47). Such functions, $x(s)$, can also be regarded as “paths” between x and x' , hence, Eq. (47) is also referred to as the continuous *path integral*, being an integration over all paths between x and x' .

The integrand appearing in the exponential in Eq. (47) has a special name in classical mechanics. It is known as the *Lagrangian*:

$$L(x, \dot{x}) = \frac{m}{2} \dot{x}^2 - \phi(x) \quad (48)$$

where $\dot{x} \equiv dx/ds$. The Lagrangian is simply the difference between the kinetic and potential energies, expressed as a function of the velocity, \dot{x} and position, x . The reader may easily verify that the following *Euler-Lagrange* equation:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) - \frac{\partial L}{\partial x} = 0 \quad (49)$$

is equivalent to the Newton equation of motion $m\ddot{x} = -d\phi/dx$. Moreover, the integral of the Lagrangian over a specific path, $x(s)$, plays an important role in classical mechanics. It is known as the *action* integral:

$$A[x(s)] = \int_0^t ds L(x(s), \dot{x}(s)) \quad (50)$$

Again, the reader may verify directly that the Euler-Lagrange equation, Eq. (49) results from extremization of the action

$$\delta A = 0 \quad (51)$$

with respect to the path. Paths that satisfy this extremization condition are known as *classical paths*. Thus, in terms of the action, the continuous path integral for the propagator can be written:

$$\langle x | e^{-iHt/\hbar} | x' \rangle = \int_x^{x'} \mathcal{D}x(s) e^{\frac{i}{\hbar} A[x(s)]} \quad (52)$$

The content of Eq. (52) is that the complete propagator is constructed by “summing” (integrating) over all paths between x and x' weighted by the complex exponential of the action for each path divided by \hbar . The sum over paths is illustrated in Fig. 2 below:

By virtue of Eq. (51), it is clear that classical paths will have the most significant contribution to the path integral, as a small change from a classical path will only cause small changes in the action, hence only small variations in the oscillating integrand of Eq. (52). For paths that differ substantially from classical paths, small changes will cause large changes in the action, and, hence, the oscillatory functions in Eq. (52) will fluctuate wildly, leading to positive and negative contributions that largely cancel out.

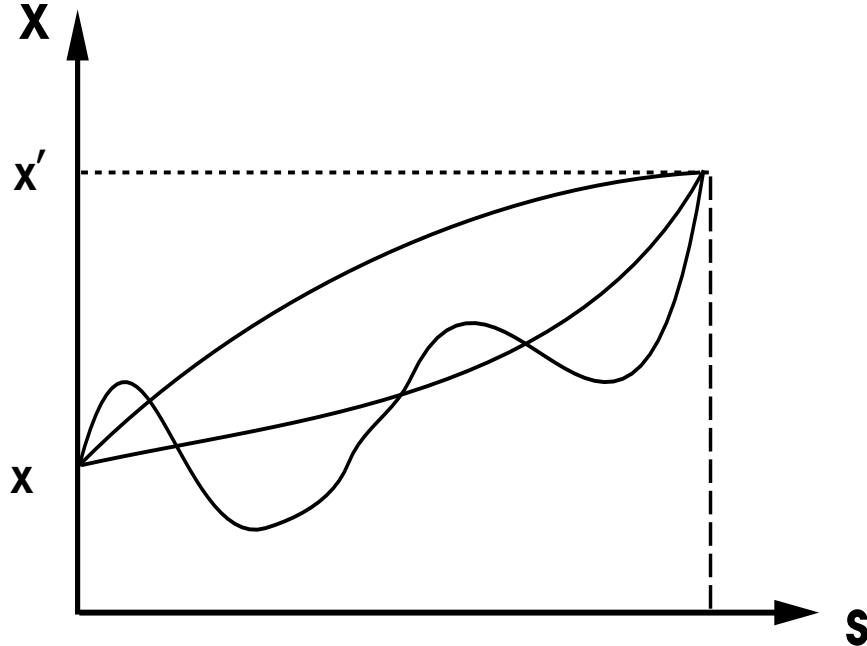


Figure 2. Illustration of the sum over paths

Turning, once again, to the density matrix, we noted above that an expression for the propagator could be obtained from the density matrix by letting $\beta = it/\hbar$. Conversely, the density matrix can be obtained from the propagator by letting $t = -i\beta\hbar$, i.e., by letting time be imaginary. For this reason, the canonical density matrix is often referred to as an *imaginary time* or *Euclidean* propagator. Taking the continuous limit of Eq. (39) leads to the following path integral expression for the density matrix:

$$\langle x|e^{-\beta H}|x'\rangle = \int_x^{x'} \mathcal{D}x(\tau) \exp \left\{ -\frac{1}{\hbar} \int_0^{\beta\hbar} d\tau \left[\frac{m}{2} \left(\frac{dx}{d\tau} \right)^2 + \phi(x(\tau)) \right] \right\} \quad (53)$$

where τ is an imaginary time integration variable, $\tau \in [0, \beta\hbar]$, and the integration is now over all imaginary time paths, $x(\tau)$ beginning at x and ending at x' . Eq. (53) is known as a continuous *imaginary time path integral*. The quantity in the integrand of the exponential in Eq. (53) has the form of the kinetic plus potential energies expressed as a function of x and $\dot{x} = dx/d\tau$. This is known as the *Euclidean Lagrangian*

$$\mathcal{L}(x, \dot{x}) = \frac{m}{2}\dot{x}^2 + \phi(x) \quad (54)$$

and its integral over a continuous path $x(\tau)$ is called the *imaginary time* or *Euclidean* action:

$$S[x(\tau)] = \int_0^{\beta\hbar} d\tau \mathcal{L}(x(\tau), \dot{x}(\tau)) \quad (55)$$

Minimization of this action via $\delta S = 0$ leads to the same Euler-Lagrange equation of motion for the classical paths:

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{x}} \right) - \frac{\partial \mathcal{L}}{\partial x} = 0 \quad (56)$$

However, when the Euclidean Lagrangian is substituted into Eq. (56), the resulting equation of motion for the classical paths is

$$m\ddot{x} = \frac{d\phi}{dx} \quad (57)$$

which leads to motion on the inverted potential surface, $-\phi(x)$. Again, Eq. (57) must be subject to the condition, $x(0) = x$ and $x(\beta\hbar) = x'$ in order to obtain the appropriate classical paths. Thus, the density matrix can be expressed as an imaginary time path integral involving the Euclidean action:

$$\langle x|e^{-\beta H}|x'\rangle = \int_x^{x'} \mathcal{D}x(\tau) e^{-\frac{1}{\hbar}S[x(\tau)]} \quad (58)$$

Since $S[x(\tau)]$ is a minimum along the classical paths, paths satisfying Eq. (57) subject to the boundary conditions constitute the dominant contribution to the imaginary time path integral. Paths far from the classical paths will have large Euclidean actions and, therefore, will be severely damped out by the damped exponential appearing in Eq. (58).

Finally, returning to the partition function, recall that $Z(\beta)$ is the trace of the density matrix, which can now be expressed in terms of an imaginary time path integral:

$$\begin{aligned} Z(\beta) &= \int dx \langle x | e^{-\beta H} | x \rangle \\ &= \int dx \int_x^x \mathcal{D}x(\tau) e^{-\frac{1}{\hbar} S[x(\tau)]} \end{aligned} \quad (59)$$

The imaginary time path integral expression for $Z(\beta)$ states that one must first calculate the diagonal density matrix elements by performing a sum over all imaginary time paths, $x(\tau)$ that begin and end at the same point, x , via the boundary condition, $x(0) = x(\beta\hbar) = x$, and then integrate over all values of x . This sum is illustrated in Fig. 3 below: Since these

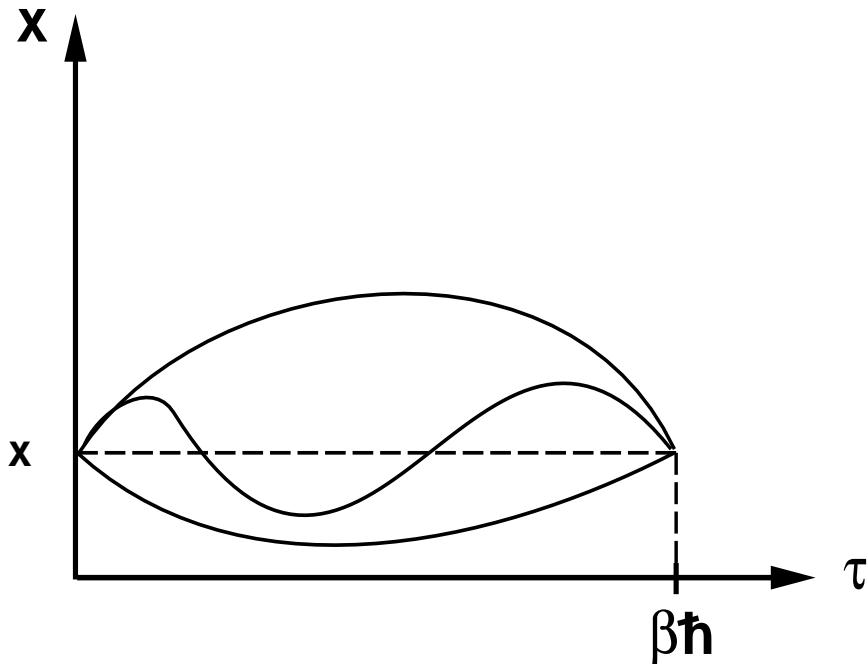


Figure 3. Illustration of the sum over cyclic imaginary time paths

imaginary time paths are cyclic, in that they return to their starting values in imaginary time $\tau = \beta\hbar$, Eq. (59) is often written in the shorthand form:

$$Z(\beta) = \oint \mathcal{D}x(\tau) e^{-\frac{1}{\hbar} S[x(\tau)]} \quad (60)$$

where \oint indicates integration over all cyclic paths that satisfy the condition $x(0) = x(\beta\hbar)$.

5 Thermodynamics and Expectation Values in Terms of Path Integrals

Suppose we wish to compute the expectation value of an operator, \hat{O} that is purely a function of the operator, x , $\hat{O} = \hat{O}(x)$. By definition, the expectation value of \hat{O} is

$$\langle \hat{O} \rangle = \frac{\text{Tr} [\hat{O} e^{-\beta H}]}{\text{Tr} [e^{-\beta H}]} = \frac{1}{Z(\beta)} \text{Tr} [\hat{O} e^{-\beta H}] \quad (61)$$

Once again, carrying out the trace in the coordinate basis gives

$$\begin{aligned} \langle \hat{O} \rangle &= \frac{1}{Z(\beta)} \int dx \langle x | \hat{O} e^{-\beta H} | x \rangle \\ &= \frac{1}{Z(\beta)} \int dx o(x) \langle x | e^{-\beta H} | x \rangle \end{aligned} \quad (62)$$

where $o(x)$ is the corresponding eigenvalue of \hat{O} obtained by acting on a coordinate eigenstate. Substituting in the continuous imaginary time path integral expression for the diagonal density matrix element, the expectation value of \hat{O} can be written as a continuous path integral of the form:

$$\langle \hat{O} \rangle = \frac{1}{Z(\beta)} \oint \mathcal{D}x(s) o(x(s)) e^{-\frac{1}{\hbar} S[x(s)]} \quad (63)$$

It can also be written as the limit of a discrete path integral:

$$\begin{aligned} \langle \hat{O} \rangle &= \frac{1}{Z(\beta)} \lim_{P \rightarrow \infty} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \\ &\times \int dx_1 \cdots dx_P o(x_1) \exp \left\{ - \sum_{i=1}^P \left[\frac{mP}{2\beta\hbar^2} (x_{i+1} - x_i)^2 + \frac{\beta}{P} \phi(x_i) \right] \right\}_{x_{P+1}=x_1} \end{aligned} \quad (64)$$

Recognizing, however, that the integral in Eq. (64) is invariant under a cyclic relabeling of all of the path integration variables, $x_1 \rightarrow x_2, x_2 \rightarrow x_3, \dots$, such a cyclic relabeling can be performed P times, the resulting expressions added together and then divided by P to yield:

$$\begin{aligned} \langle \hat{O} \rangle &= \frac{1}{Z(\beta)} \lim_{P \rightarrow \infty} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \\ &\times \int dx_1 \cdots dx_P \frac{1}{P} \sum_{i=1}^P o(x_i) \exp \left\{ - \sum_{i=1}^P \left[\frac{mP}{2\beta\hbar^2} (x_{i+1} - x_i)^2 + \frac{\beta}{P} \phi(x_i) \right] \right\}_{x_{P+1}=x_1} \end{aligned} \quad (65)$$

A finite P expression for $\langle \hat{O} \rangle$ can be obtained by substituting $Z_P(\beta)$ for $Z(\beta)$ and writing

$$\begin{aligned} \langle \hat{O} \rangle_P &= \frac{1}{Z_P(\beta)} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \\ &\times \int dx_1 \cdots dx_P \frac{1}{P} \sum_{i=1}^P o(x_i) \exp \left\{ - \sum_{i=1}^P \left[\frac{mP}{2\beta\hbar^2} (x_{i+1} - x_i)^2 + \frac{\beta}{P} \phi(x_i) \right] \right\}_{x_{P+1}=x_1} \end{aligned} \quad (66)$$

so that the true expectation value is obtained in the limit $P \rightarrow \infty$:

$$\langle \hat{O} \rangle = \lim_{P \rightarrow \infty} \langle \hat{O} \rangle_P \quad (67)$$

Note that the quantity

$$\begin{aligned} f(x_1, \dots, x_P) &= \frac{1}{Z_P(\beta)} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \\ &\exp \left\{ - \sum_{i=1}^P \left[\frac{mP}{2\beta\hbar^2} (x_{i+1} - x_i)^2 + \frac{\beta}{P} \phi(x_i) \right] \right\}_{x_{P+1}=x_1} \end{aligned} \quad (68)$$

satisfies the properties of a P -dimensional probability distribution function, i.e. it is positive definite and integrates to unity. Defining the quantity

$$o_P(x_1, \dots, x_P) = \frac{1}{P} \sum_{i=1}^P o(x_i) \quad (69)$$

the expectation value can be written as an average with respect to the probability distribution function $f(x_1, \dots, x_P)$:

$$\langle \hat{O} \rangle_P = \langle o_P(x_1, \dots, x_P) \rangle_f \equiv \int dx_1 \cdots dx_P o_P(x_1, \dots, x_P) f(x_1, \dots, x_P) \quad (70)$$

The quantity $o_P(x_1, \dots, x_P)$ is called an *estimator* for the operator \hat{O} . In the limit $P \rightarrow \infty$, the average of $o_P(x_1, \dots, x_P)$ with respect to the probability distribution $f(x_1, \dots, x_P)$ will yield the true expectation value of \hat{O} :

$$\langle \hat{O} \rangle = \lim_{P \rightarrow \infty} \langle o_P(x_1, \dots, x_P) \rangle_f \quad (71)$$

Estimators play an important role in path integral calculations. Any observable quantity will have a corresponding estimator, i.e. a function of x_1, \dots, x_P whose average gives, in the limit $P \rightarrow \infty$, the true expectation value of that observable.

Estimators for thermodynamic quantities can be derived as well. These are often obtained as derivatives of the partition function. Consider the total internal energy, E , given by

$$E = -\frac{\partial}{\partial \beta} \ln Z(\beta) = -\frac{1}{Z(\beta)} \frac{\partial Z(\beta)}{\partial \beta} \quad (72)$$

The estimator is obtained by substituting $Z_P(\beta)$ for the true $Z(\beta)$ and computing the required derivative. The reader may easily verify that the resulting expression is

$$E_P = \langle \varepsilon_P(x_1, \dots, x_P) \rangle_f \quad (73)$$

where the estimator, $\varepsilon_P(x_1, \dots, x_P)$ is given by

$$\varepsilon_P(x_1, \dots, x_P) = \frac{P}{2\beta} - \sum_{i=1}^P \left[\frac{mP}{2\beta^2 \hbar^2} (x_{i+1} - x_i)^2 - \frac{1}{P} \phi(x_i) \right] \quad (74)$$

Equation (74) is known as the *primitive energy estimator*. The true total thermodynamic internal energy will be given as the $P \rightarrow \infty$ limit of the average of this estimator with respect to the distribution function $f(x_1, \dots, x_P)$. However, owing to the quadratic term in Eq. (74) and its linearly P -dependent prefactor, this estimator is somewhat difficult to work with numerically, especially as P becomes large, as was originally shown by Hermann Bruiskin and Berne.¹⁹ These authors showed that the convergence error in this estimator grows with P . In order to rectify this, they derived an equivalent form of the estimator using the virial theorem that involves only the potential and its first derivative. The alternative estimator, known as the *virial estimator* is given by

$$\varepsilon_P^{(\text{vir})}(x_1, \dots, x_P) = \frac{1}{2\beta} + \frac{1}{P} \sum_{i=1}^P \left[\frac{1}{2} (x_i - x_c) \frac{\partial \phi}{\partial x_i} + \phi(x_i) \right] \quad (75)$$

where x_c is a quantity known as the *path centroid* and is simply the geometric center of the path given by

$$x_c = \frac{1}{P} \sum_{i=1}^P x_i \quad (76)$$

or, in the continuous limit,

$$x_c = \frac{1}{\beta \hbar} \int_0^{\beta \hbar} d\tau x(\tau) \quad (77)$$

This quantity is of central importance in the development of semi-classical effective potentials via the approach of Feynman and Kleinert²⁰ and can also be used to obtain approximate quantum dynamics via the so called centroid dynamics method.¹⁴⁻¹⁷

Other thermodynamic estimators can be derived, for example, for the pressure²¹ or heat capacity,²² however, details of the derivations will not be given here. Rather, we will next turn our attention to the problem of computing path integrals and quantum observable properties via molecular dynamics.

6 Path Integral Molecular Dynamics

The continuous functional integral representation of the path integral is mathematically elegant and can be used for formal manipulations, however, it is not suitable for direct numerical evaluation. The latter requires the discrete, finite P form of the partition function given by

$$Z_P(\beta) = \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \int dx_1 \cdots dx_P \exp \left\{ - \sum_{i=1}^P \left[\frac{mP}{2\beta\hbar^2} (x_{i+1} - x_i)^2 + \frac{\beta}{P} \phi(x_i) \right] \right\}_{x_{P+1}=x_1} \quad (78)$$

Introducing a “chain frequency”

$$\omega_P = \frac{\sqrt{P}}{\beta\hbar} \quad (79)$$

and an effective potential

$$U_{\text{eff}}(x_1, \dots, x_P) = \sum_{i=1}^P \left[\frac{1}{2} m \omega_P^2 (x_{i+1} - x_i)^2 + \frac{1}{P} \phi(x_i) \right]_{x_{P+1}=x_1} \quad (80)$$

Eq. (78) can be written as

$$Z_P(\beta) = \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \int dx_1 \cdots dx_P e^{-\beta U_{\text{eff}}(x_1, \dots, x_P)} \quad (81)$$

When written in this manner, the quantum partition function looks like a classical configurational partition function for a P -particle system, where the P particles are discrete points along a cyclic path. This is illustrated in Fig. 4 below: Since the discrete cyclic

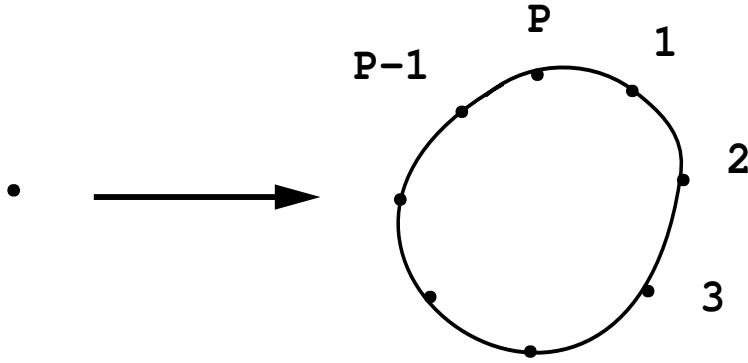


Figure 4. Illustration of the connection between a single quantum particle and the discretized P -point (“bead”) path

path resembles a beaded necklace, the P particles are often referred to as “beads.” Owing to the quadratic term in Eq. (80), the beads are coupled to each other by nearest neighbor harmonic springs with frequency ω_P , as the figure illustrates. Moreover, each bead is separately subject to the external potential, $\phi(x)$, i.e. $\phi(x)$ acts on only one bead at a time and gives rise to *no additional* coupling between the beads.

In order to make the connection between the quantum partition function and the fictitious P -particle classical system more manifest, consider supplementing Eq. (81) by a set of P Gaussian integrals and writing

$$Z_P(\beta) = \mathcal{N} \int dp_1 \cdots dp_P \int dx_1 \cdots dx_P \exp \left\{ -\beta \left[\sum_{i=1}^P \frac{p_i^2}{2\tilde{m}_i} + U_{\text{eff}}(x_1, \dots, x_P) \right] \right\} \quad (82)$$

The new Gaussian variables are regarded as fictitious classical “momenta” so that the constants, \tilde{m}_i have units of mass. Note that these Gaussian integrals are uncoupled and can be performed analytically to yield, $\prod_{i=1}^P (2\pi\tilde{m}_i/\beta)^{P/2}$, and the overall constant \mathcal{N} can be chosen so as to reproduce the correct prefactor in Eq. (81). Thus, we have complete freedom to choose the \tilde{m}_i as we like. Writing the partition function in this manner, however, gives it the form of a P -dimensional classical phase space integral for the fictitious classical system consisting of P beads. This connection between the quantum system and the P -particle fictitious classical system is known as the *classical isomorphism*.²³ The true quantum system is recovered as the number of fictitious classical beads, P , becomes infinite.

Having connected the quantum partition function to a fictitious classical partition function, Eq. (82) can be evaluated, at least in principle, using classical molecular dynamics (MD) based on equations of motion derived from a fictitious classical Hamiltonian of the form

$$\begin{aligned} H(p, x) &= \sum_{i=1}^P \frac{p_i^2}{2\tilde{m}_i} + U_{\text{eff}}(x_1, \dots, x_P) \\ &= \sum_{i=1}^P \left[\frac{p_i^2}{2\tilde{m}_i} + \frac{1}{2} m\omega_P^2 (x_{i+1} - x_i)^2 + \frac{1}{P} \phi(x_i) \right]_{x_{P+1}=x_1} \end{aligned} \quad (83)$$

where $p \equiv \{p_1, \dots, p_P\}$ and $x \equiv \{x_1, \dots, x_P\}$. Such an approach is doomed to failure, however, as was established in 1984 by Hall and Berne.¹⁰ The reason for this is due to the quadratic term in Eq. (83). First, as P becomes large, the effective force constant between the beads, $m\omega_P^2$ grows linearly with P . At the same time, the contribution from the external potential is attenuated by a factor of $1/P$ so that the harmonic forces dominate. This causes MD trajectories to remain very close to the invariant tori described by the $\phi = 0$ system and not explore the full available phase space. Second, since the quadratic term couples all of the beads, there are many time scales buried in this term, and the highest frequency of these will limit the time step that can be used. This means that lower frequency modes will be inadequately sampled, leading to very slow convergence of observable quantities. Third, ordinary MD will generate a microcanonical distribution of H , i.e. a distribution function of the form $\delta(H(p, x) - E)$, where E is the conserved energy. Clearly, this is not the form appearing in Eq. (82), which requires a canonical distribution of the form $\exp(-\beta H)$.

In order to solve these problems, we introduce a new MD based approach which includes the following features:

- i. a change of integration variables in Eq. (82), which serves to uncouple the harmonic term, and a corresponding reformulation of the fictitious classical Hamiltonian;
- ii. a multiple time scale numerical integration algorithm that treats the inherent time scales of the harmonic and external potential terms with appropriate time steps;
- iii. a highly efficient canonical MD method that rigorously generates a canonical phase space distribution.

First, since the integration variables in Eq. (82) are completely arbitrary, we are free to change them with the aim of uncoupling the quadratic term. To this end, there are several

possibilities. Consider, first, a simple change of variables of the form:

$$\begin{aligned} u_1 &= x_1 \\ u_i &= x_i - x_i^* \end{aligned} \quad (84)$$

where

$$x_i^* = \frac{(i-1)x_{i+1} + x_1}{i} \quad (85)$$

This transformation is known as a *staging transformation* as it is based on the *staging* Monte Carlo approach of Ceperley and Pollock.⁶ Note that the inverse of this transformation can be expressed in a convenient recursive fashion:

$$\begin{aligned} x_1 &= u_1 \\ x_i &= u_i + \frac{i-1}{i} x_{i+1} + \frac{1}{i} u_1 \end{aligned} \quad i = 2, \dots, P \quad (86)$$

where the $i = P$ term is used to start the recursion. It can also be expressed directly by

$$\begin{aligned} x_1 &= u_1 \\ x_i &= \sum_{j=i}^P \frac{i-1}{j-1} u_j \end{aligned} \quad i = 2, \dots, P \quad (87)$$

If Eq. (86) is substituted into Eq. (82), the resulting expression for the partition function is

$$Z_P(\beta) = \mathcal{N} \int dp_1 \cdots dp_P \int du_1 \cdots du_P \exp \left\{ -\beta \left[\sum_{i=1}^P \frac{p_i^2}{2\tilde{m}_i} + U_{\text{eff}}^{(\text{stage})}(u_1, \dots, u_P) \right] \right\} \quad (88)$$

where the transformed effective potential is

$$U_{\text{eff}}^{(\text{stage})}(u_1, \dots, u_P) = \sum_{i=1}^P \left[\frac{1}{2} m_i \omega_P^2 u_i^2 + \frac{1}{P} \phi(x_i(u_1, \dots, u_P)) \right] \quad (89)$$

and $x_i(u_1, \dots, u_P)$ are the linear transformation functions defined by Eq. (86) or Eq. (87). In Eq. (89), the masses, m_i are given by

$$\begin{aligned} m_1 &= 0 \\ m_i &= \frac{i}{i-1} m \end{aligned} \quad (90)$$

and are known as the *staging masses*. The important feature of Eq. (89) is the fact that the quadratic term is completely uncoupled in terms of the u variables. Note, also, that the variable u_1 does not appear in the transformed harmonic term, since $m_1 = 0$. This uncoupled mode represents collective motion of the entire cyclic chain. This suggests that an effective MD scheme can be obtained based on the equations of motion derived from the effective fictitious classical Hamiltonian

$$\begin{aligned} H^{(\text{stage})}(p, u) &= \sum_{i=1}^P \frac{p_i^2}{2\tilde{m}_i} + U_{\text{eff}}^{(\text{stage})}(u_1, \dots, u_P) \\ &= \sum_{i=1}^P \left[\frac{p_i^2}{2\tilde{m}_i} + \frac{1}{2} m_i \omega_P^2 u_i^2 + \frac{1}{P} \phi(x_i(u_1, \dots, u_P)) \right] \end{aligned} \quad (91)$$

Note that this Hamiltonian is not equivalent to the Hamiltonian in Eq. (83), nor should it be. The Hamiltonian in Eq. (91) generates a different dynamics that will sample the full configuration space more effectively. The Hamiltonian in Eq. (91) suggests that the optimal choice of the masses, \tilde{m}_i is

$$\begin{aligned}\tilde{m}_1 &= m \\ \tilde{m}_i &= m_i\end{aligned}\tag{92}$$

since, with this choice, all of the staging modes, u_2, \dots, u_P will move on the *same time scale*, thereby leading to efficient sampling of all modes in an MD scheme. The equations of motion derived from Eq. (91) are

$$\begin{aligned}\dot{u}_i &= \frac{p_i}{\tilde{m}_i} \\ \dot{p}_i &= -\frac{\partial U_{\text{eff}}^{(\text{stage})}}{\partial u_i} = -m_i \omega_P^2 u_i - \frac{\partial \phi}{\partial u_i}\end{aligned}\tag{93}$$

The forces, $\partial \phi / \partial u_i$, are obtained via the chain rule using Eq. (86) which gives a convenient recursive form for the partial derivatives:

$$\begin{aligned}\frac{\partial \phi}{\partial u_1} &= \frac{1}{P} \sum_{i=1}^P \frac{\partial \phi}{\partial x_i} \\ \frac{\partial \phi}{\partial u_i} &= \frac{\partial \phi}{\partial x_i} + \frac{i-2}{i-1} \frac{\partial \phi}{\partial x_{i-1}}\end{aligned}\tag{94}$$

This form is especially convenient since the bead forces, $\partial \phi / \partial x_i$ can be computed directly given the form of the potential.

Equations (93) alone are not sufficient to give a satisfactory path integral MD scheme, however, since they still only generate a microcanonical distribution, $\delta(H^{(\text{stage})}(p, u) - E)$. In order to ensure that a proper canonical distribution is generated, Eqs. (93) must be coupled to a thermostating method. Although many such approaches exist, our experience has been that the extended system methods such as the Nosé-Hoover chain²⁴ or generalized Gaussian moment²⁵ are the most effective. Both of these schemes can be rigorously shown to generate a canonical distribution in the physical phase space variables, (p, u) . It is also possible to write the partition function in the form

$$\begin{aligned}Z_P(\beta) &= \mathcal{N} \int dp_1 \cdots dp_P \int du_1 \cdots du_P \delta \left(\sum_{i=1}^P \frac{p_i^2}{2\tilde{m}_i} - K_0 \right) \\ &\times \exp \left\{ -\beta U_{\text{eff}}^{(\text{stage})}(u_1, \dots, u_P) \right\}\end{aligned}\tag{95}$$

and employ an isokinetic method following the approach recently introduced by Minary and Tuckerman,²⁶ which has also been shown to be highly effective and offers the advantage of simplicity. Here, we shall show how to employ the Nosé-Hoover chain (NHC) approach. In the NHC approach, a set of M additional “heat bath” variables, η_1, \dots, η_M is introduced along with a set of corresponding momenta, $p_{\eta_1}, \dots, p_{\eta_M}$, such that the k th variables control the kinetic energy fluctuations in the $(k-1)$ st variables, where $k=0$ corresponds to the physical momenta. For example, for a particle moving in a one-dimensional

potential described by equations of motion of them form $\dot{x} = p/m$, $\dot{p} = -d\phi/dx$, the thermostatted equations take the form

$$\begin{aligned}\dot{x} &= \frac{p}{m} \\ \dot{p} &= -\frac{d\phi}{dx} - \frac{p_{\eta_1}}{Q_1}p \\ \dot{\eta}_k &= \frac{p_{\eta_k}}{Q_k} \\ \dot{p}_{\eta_1} &= \frac{p^2}{m} - kT - \frac{p_{\eta_2}}{Q_2}p_{\eta_1} \\ \dot{p}_{\eta_k} &= \frac{p_{\eta_{k-1}}^2}{Q_{k-1}} - kT - \frac{p_{\eta_{k+1}}}{Q_{k+1}}p_{\eta_k} \\ \dot{p}_{\eta_M} &= \frac{p_{\eta_{M-1}}^2}{Q_{M-1}} - kT\end{aligned}\quad (96)$$

where Q_1, \dots, Q_M is a set of thermostat mass parameters (having units of energy \times time²) that control the time scale on which these variables evolve. Equations (96) conserve the following energy:

$$H' = \frac{p^2}{2m} + \phi(x) + \sum_{k=1}^M \left[\frac{p_{\eta_k}^2}{2Q_k} + kT\eta_k \right] \quad (97)$$

Note that H' is *not* a Hamiltonian for Eqs. (96). The proof that Eqs. (96) generates a

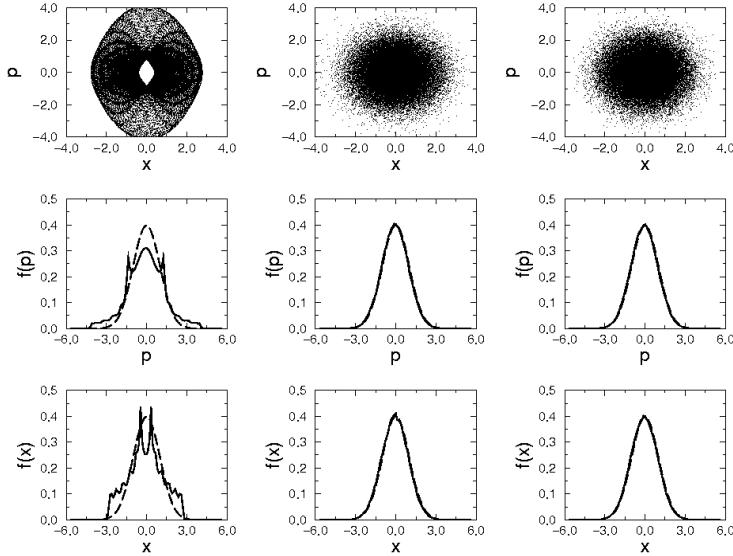


Figure 5. Poincaré sections, probability distributions, $f(x)$ and $f(p)$ of position and momentum for a harmonic oscillator with potential $\phi(x) = m\omega^2x^2/2$ with $m = 1$, $\omega = 1$. Left column shows results for $M = 1$ (Nosé-Hoover thermostat), middle column shows, $M = 4$, and right column shows, $M = 6$.

correct canonical distribution can be found in the recent work of Tuckerman, *et al.*²⁷ Figure 5 shows that these equations are capable of generating a correct canonical distribution for a simple harmonic oscillator described by $\phi(x) = m\omega^2 x^2/2$ for $M > 1$ (note that $M = 1$ corresponds to the more familiar Nosé-Hoover thermostat, which fails for a harmonic oscillator).

Path integral MD is dominated by harmonic motion. Therefore, it is *absolutely crucial* to couple a thermostat to *each degree of freedom* in the system. This means *each* Cartesian direction of *each* mode variable. Failure to do so will result in a scheme that does not converge. Therefore, the complete path integral MD equations take the form:

$$\begin{aligned}\dot{u}_i &= \frac{p_i}{\tilde{m}_i} \\ \dot{p}_i &= -m_i \omega_P^2 u_i - \frac{\partial \phi}{\partial u_i} - \frac{p_{\eta_{1,i}}}{Q_1} p_i \\ \dot{\eta}_{k,i} &= \frac{p_{\eta_{k,i}}}{Q_k} \\ \dot{p}_{\eta_{1,i}} &= \frac{p_i^2}{\tilde{m}_i} - kT - \frac{p_{\eta_{2,i}}}{Q_2} p_{\eta_{1,i}} \\ \dot{p}_{\eta_{k,i}} &= \frac{p_{\eta_{k-1,i}}^2}{Q_{k-1}} - kT - \frac{p_{\eta_{k+1,i}}}{Q_{k+1}} p_{\eta_{k,i}} \\ \dot{p}_{\eta_{M,i}} &= \frac{p_{\eta_{M-1,i}}^2}{Q_{M-1}} - kT\end{aligned}\tag{98}$$

The conserved energy of these equations is

$$H' = \sum_{i=1}^P \left\{ \frac{p_i^2}{2\tilde{m}_i} + \frac{1}{P} \phi(x_i) + \sum_{k=1}^M \left[\frac{p_{\eta_{k,i}}^2}{2Q_k} + kT \eta_{k,i} \right] \right\}\tag{99}$$

The thermostat mass parameters are chosen to evolve on the time scale of the harmonic forces, and, therefore are assigned values according to:

$$Q_k = \frac{1}{\beta \omega_P^2} \quad \forall k\tag{100}$$

Figure 6 illustrates the performance of the path integral MD scheme for a harmonic oscillator, $\phi(x) = m\omega^2 x^2/2$ with

$$\frac{m\omega}{\hbar} = 0.03 \quad \beta\hbar\omega = 15.8$$

and $P = 400$. The figure shows how the virial energy estimator converges if straightforward microcanonical MD is used compared to the algorithm in Eqs. (98) and compared to staging Monte Carlo. It can be seen that the simple microcanonical MD performs extremely poorly, however, when Eqs. (98) are used, the algorithm is almost as efficient as the staging Monte Carlo, as measured by the error bar as a function of block size.²⁸ Figure 7 also shows what happens if a single global thermostat is used instead of a thermostat on each degree of freedom. This comparison serves to underscore the warning given above that each degree of freedom requires its own thermostat.

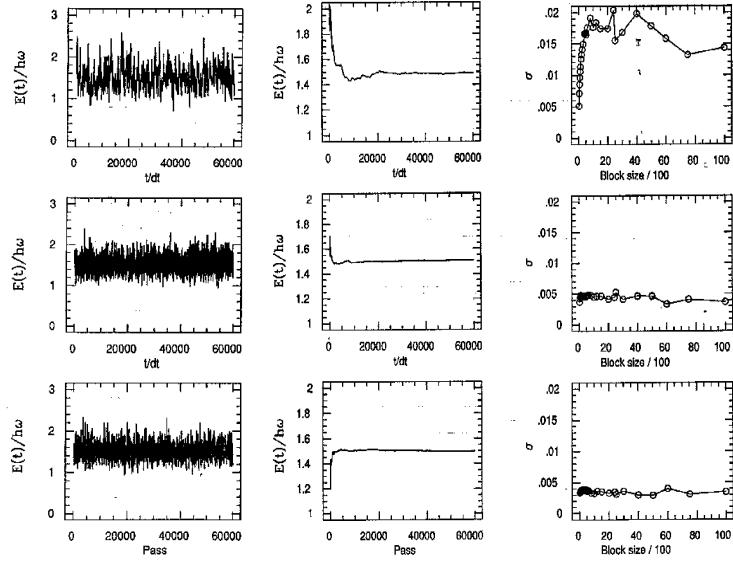


Figure 6. Instantaneous and cumulative value of the virial energy estimator and the associated error bar as a function of block size for standard microcanonical MD (top row), the staging MD algorithm presented (middle row), and staging Monte Carlo using the Ceperley and Pollock algorithm for the harmonic oscillator example .

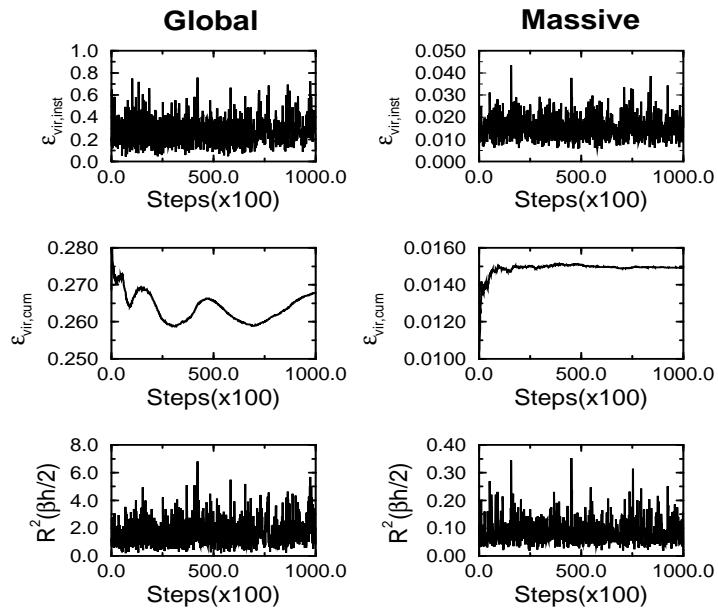


Figure 7. Illustration of the problem inherent with the use of a global thermostat in path integral MD by examination of the instantaneous and cumulative values of the virial energy estimator and the spread of the path, defined as the distance between furthest points on the path integral chain.

For completeness, we mention another possible transformation that can be used to uncouple the harmonic potential term. This is the so called *normal mode* transformation obtained by performing a Fourier expansion of the discrete cyclic path:

$$x_j = \sum_{l=1}^P c_l e^{2\pi i(j-1)(l-1)/P} \quad (101)$$

The coefficients, c_l are complex, so that the normal mode variables, u_l are given by

$$\begin{aligned} u_1 &= c_1 \\ u_P &= c_{(P+2)/2} \\ u_{2l-2} &= \text{Re}(c_l) \\ u_{2l-1} &= \text{Im}(c_l) \end{aligned} \quad (102)$$

Associated with the normal mode transformation is a set of normal mode frequencies:

$$\lambda_{2l-1} = \lambda_{2l-2} = 2P \left[1 - \cos \left(\frac{2\pi(l-1)}{P} \right) \right] \quad (103)$$

from which a set of normal mode masses can be obtained by

$$m_l = m\lambda_l \quad (104)$$

The path integral MD equations, Eqs. (98) are equally valid for the normal mode transformation, and simply require using Eqs. (102) to determine the forces and Eq. (104) in place of the staging masses. The normal mode method has the advantage that the mode, u_1 is given, in terms of the bead variables, by

$$u_1 = \frac{1}{P} \sum_{i=1}^P x_i \quad (105)$$

i.e. the path centroid. Therefore, the normal mode method should be used when one wishes to compute quantum free energy profiles or perform approximate quantum dynamics via the centroid dynamics approach.

In order to complete the path integral MD scheme, a reversible multiple time scale integration method is needed. In order to illustrate how such a scheme is constructed, consider, first a simple one-particle, one-dimensional classical system described by a Hamiltonian $H = p^2/2m + \phi(x)$. The equations of motion

$$\dot{x} = \frac{p}{m} \quad \dot{p} = F(x) \quad (106)$$

where $F(x) = -\phi'(x)$, can be cast in an operator form:

$$\dot{x} = iLx \quad \dot{p} = iLp \quad (107)$$

where iL is the *Liouville operator* given by

$$iL = \frac{p}{m} \frac{\partial}{\partial x} + F(x) \frac{\partial}{\partial p} \quad (108)$$

The equations of motion can now be solved formally by

$$x(t) = e^{iLt}x(0) \quad p(t) = e^{iLt}p(0) \quad (109)$$

where the operator $\exp(iLt)$ is called the *classical propagator*. Although we cannot evaluate its action directly on an initial condition, $\{x(0), p(0)\}$, we can approximate it for a small time interval, Δt , using the Trotter theorem. Note that iL is composed of two terms, iL_1 and iL_2 given by

$$iL_1 = \frac{p}{m} \frac{\partial}{\partial x} \quad iL_2 = F(x) \frac{\partial}{\partial p} \quad (110)$$

If Δt is small, then we may approximate

$$e^{iL\Delta t} = e^{iL_2\Delta t/2} e^{iL_1\Delta t} e^{iL_2\Delta t/2} + \mathcal{O}(\Delta t^3) \quad (111)$$

Note that, although the error in one step is of order Δt^3 , the error accumulated by applying the operator in Eq. (111) N times to generate a trajectory of real time length t will be of order $N\Delta t^3 = t\Delta t^2 = t^3/N^2$. Now, each operator appearing in Eq. (111) can be shown to generate a simple translation of the variable on which it acts:

$$\begin{aligned} e^{iL_2/\Delta t/2} p &= p + \frac{\Delta t}{2} F(x) \\ e^{iL_2/\Delta t/2} x &= x \end{aligned} \quad (112)$$

as can be easily shown by Taylor expanding the exponentials. Similarly,

$$\begin{aligned} e^{iL_1/\Delta t} p &= p \\ e^{iL_1/\Delta t} x &= x + \Delta t \frac{p}{m} \end{aligned} \quad (113)$$

Therefore, the action of the operator in Eq. (111) on an initial condition, $\{p(0), x(0)\}$, generates three sequential updates:

$$\begin{aligned} p' &= p(0) + \frac{\Delta t}{2} F(x(0)) \\ x(\Delta t) &= x + \Delta t \frac{p'}{m} \\ &\text{Compute new force} \\ p(\Delta t) &= p' + \frac{\Delta t}{2} F(x(\Delta t)) \end{aligned} \quad (114)$$

or as a pseudocode, beginning with values $x = x(0)$ and $p = p(0)$:

$$\begin{aligned} p &\leftarrow p + \frac{\Delta t}{2} F(x) \\ x &\leftarrow x + \Delta t \frac{p}{m} \\ &\text{Compute new force} \\ p &\leftarrow p + \frac{\Delta t}{2} F(x) \end{aligned} \quad (115)$$

At the output, x and p hold the values of $x(\Delta t)$ and $p(\Delta t)$, respectively. The schemes, (114) and (115) can easily be shown to be equivalent to the familiar velocity verlet algorithm

$$\begin{aligned} x(\Delta t) &= x(0) + \Delta t \frac{p(0)}{m} + \frac{\Delta t^2}{2m} F(x(0)) \\ p(\Delta t) &= p(0) + \frac{\Delta t}{2} [F(x(0)) + F(x(\Delta t))] \end{aligned} \quad (116)$$

but derived in a powerful way using operator calculus. The technique of writing down a factorization of the classical propagator based on terms in the Liouville operator and then translating each operator into an update step in a pseudocode is a useful technique in developing numerical integrators in that it avoids the need to write down explicit finite difference equations such as Eqs. (116), which can be very cumbersome for complex sets of equations such as the path integral MD equations, Eqs. (98).

In order to illustrate the power of the operator approach, we will develop a simple multiple time scale integrator, which will form the basis of the algorithm to be used for the path integral MD equations. Suppose the potential, $\phi(x)$, contains a dominant fast term, $\phi_{\text{fast}}(x)$ and a slower term, $\phi_{\text{slow}}(x)$,

$$\phi(x) = \phi_{\text{fast}}(x) + \phi_{\text{slow}}(x) \quad (117)$$

These will give rise to fast and slow forces, $F_{\text{fast}}(x) = -\phi'_{\text{fast}}(x)$ and $F_{\text{slow}}(x) = -\phi'_{\text{slow}}(x)$, respectively. Let us construct a reference system from the fast part of the potential with a Liouville operator

$$iL^{(\text{ref})} = \frac{p}{m} \frac{\partial}{\partial x} + F_{\text{fast}}(x) \frac{\partial}{\partial p} = iL_1^{(\text{ref})} + iL_2^{(\text{ref})} \quad (118)$$

and a correction

$$iL' = F_{\text{slow}}(x) \frac{\partial}{\partial p} \quad (119)$$

Using the Trotter theorem, we break up the classical propagator as follows:

$$e^{iL\Delta t} = e^{iL'\Delta t/2} e^{iL^{(\text{ref})}\Delta t} e^{iL'\Delta t/2} \quad (120)$$

where Δt is chosen to be appropriate to the slow motion. However, in order to integrate the fast reference system accurately, we need a smaller time step. Let $\delta t = \Delta t/n$. The idea of multiple time scale integration is to integrate the reference system for n steps using the time step, δt according to the propagator:

$$\begin{aligned} e^{iL\Delta t} &= e^{iL'\Delta t/2} \\ &\times \left[e^{iL_2^{(\text{ref})}\delta t/2} e^{iL_1^{(\text{ref})}\delta t} e^{iL_2^{(\text{ref})}\delta t/2} \right]^n \\ &\times e^{iL'\Delta t/2} \end{aligned} \quad (121)$$

The action of this operator on an initial condition, $x = x(0)$, $p = p(0)$ can be written as a

pseudocode:

```


$$p \leftarrow p + \frac{\Delta t}{2} F_{\text{slow}}(x)$$

    for istep = 1 to n
        
$$p \leftarrow p + \frac{\delta t}{2} F_{\text{fast}}(x)$$

        
$$x \leftarrow x + \delta t \frac{p}{m}$$

        Compute new fast force
        
$$p \leftarrow p + \frac{\delta t}{2} F_{\text{fast}}(x)$$

    endfor
    Compute new slow force

$$p \leftarrow p + \frac{\Delta t}{2} F_{\text{slow}}(x) \quad (122)$$

```

The advantage of such a scheme in path integral MD is that the fast forces are very simple harmonic oscillator forces while the slow forces will usually be very expensive interparticle forces. Hence, the costly evaluation of the slow forces need only be done every n steps, thereby saving considerable CPU time. This scheme, known as the *reversible reference system propagator algorithm* (r-RESPA) was first introduced by Tuckerman, Martyna and Berne in 1992²⁹ and proves to be very effective in path integral MD calculations. Incorporation of the NHC thermostat coupling into the scheme is somewhat more involved and will not be discussed in detail here. Suffice it to say that an additional operator iL_{NHC} is introduced which governs the evolution of the heat bath variables, and this operator is incorporated into the fast reference system. Details of how to factorize this operator and incorporate it into the r-RESPA algorithm are given in Ref.³⁰

7 Many-Body Path Integrals

In this final section, we discuss the formulation of many-body quantum systems in terms of path integrals. Here, we shall focus on the case in which spin statistics can be neglected. This is generally an acceptable approximation at most temperatures of interest. However, a discussion of spin statistics in path integrals will be provided by other lectures in this series.

Consider an N -particle quantum system in a volume, V at temperature, T , described by a Hamiltonian of the form

$$H = \sum_{I=1}^N \frac{\mathbf{p}_I^2}{2M_I} + \phi(\mathbf{R}_1, \dots, \mathbf{R}_N) \quad (123)$$

If the derivation of Sec. 3 is followed for the N -particle Hamiltonian, the resulting discrete

path integral expression for the partition function is

$$Z(N, V, T) = \lim_{P \rightarrow \infty} \left[\prod_{I=1}^N \left(\frac{M_I P}{2\pi\beta\hbar^2} \right)^{3P/2} \int_{D(V)} d\mathbf{R}_I^{(1)} \cdots d\mathbf{R}_I^{(P)} \right] \\ \times \exp \left\{ -\beta \left[\sum_{i=1}^P \left(\sum_{I=1}^N \frac{1}{2} M_I \omega_P^2 (\mathbf{R}_I^{(i+1)} - \mathbf{R}_I^{(i)})^2 + \frac{1}{P} \phi(\mathbf{R}_1^{(i)}, \dots, \mathbf{R}_N^{(i)}) \right) \right] \right\} \quad (124)$$

where the integral is defined over the domain, $D(V)$ defined by the containing volume. The partition function now describes a fictitious classical system consisting of N cyclic chains, each containing, P beads. Again, we note that the interaction potential, $\phi(\mathbf{R}_1, \dots, \mathbf{R}_N)$, only acts between beads with the same bead index, i . This is illustrated in Fig. 8 below.

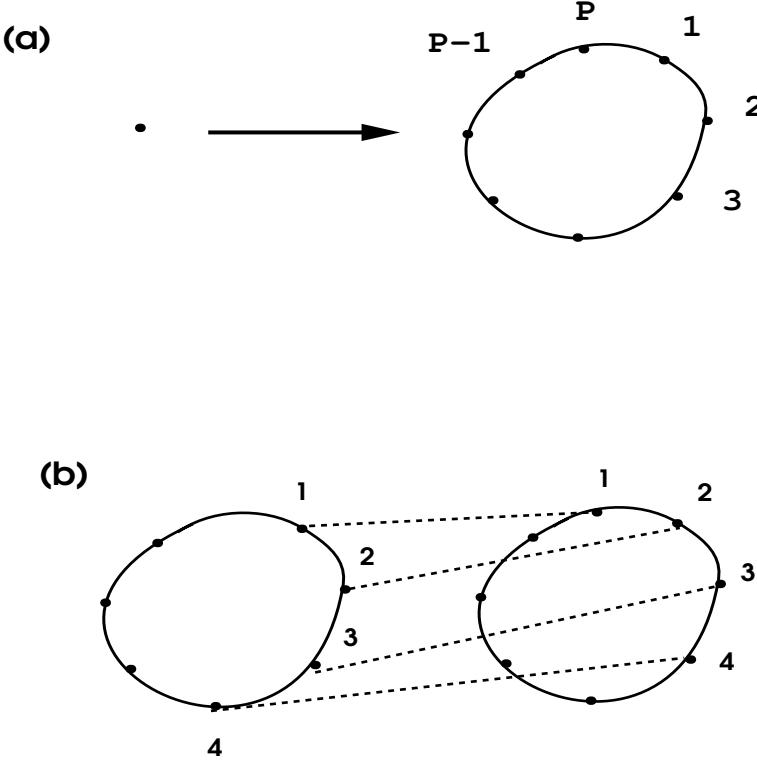


Figure 8. Illustration of the interaction between two discrete paths in a many-body path integral. The path integral MD scheme of the previous section is easily extended to the N -particle

system. For each cyclic chain, a staging or normal mode transformation is made and the complete set of mode equations of motion is written down by extending Eqs. (98). Note that a separate thermostat is still coupled to *each* degree of freedom, for a total of $3NP$

thermostat chains!! Although this may seem like a large number of extra variables, since each chain contains $2M$ variables, the CPU time required to integrate these variables is still negligible compared to the time required to evaluate the potential and forces. Fig. 9 below shows the performance of the method on a simple system, a single water molecule in a box of volume, 150 \AA^3 . The potential comes from “on the fly” electronic structure calculations

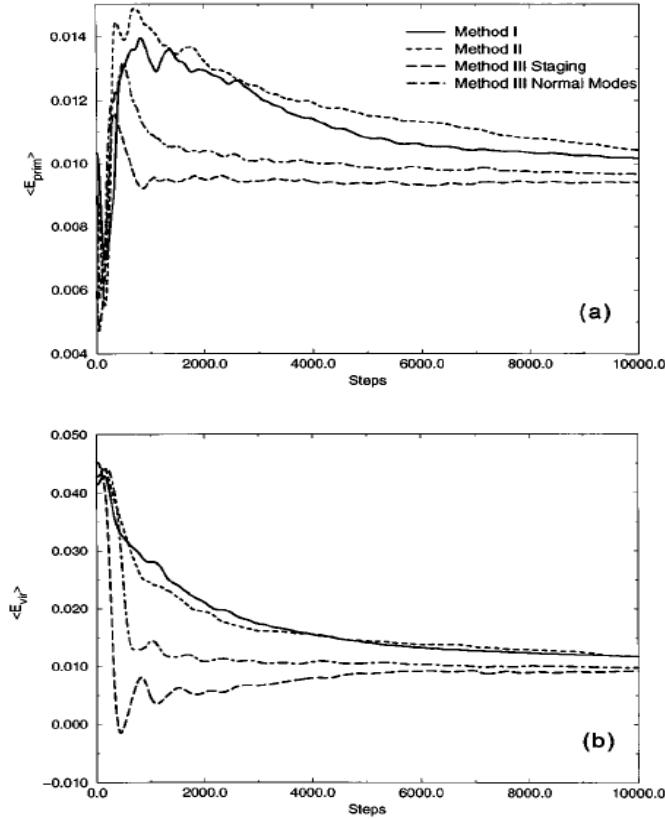


Figure 9. Performance of the staging (Method III – staging), normal mode (Method III – normal modes) compared to a path integral MD simulation with no variable transformations and no thermostats (Method I) and to a path integral MD simulation with a thermostat on each time slice (Method II). The figure compares the convergence of the virial and primitive energy estimators including equilibration segments.

coupled to the path integral (to be discussed in the lecture on *ab initio* path integrals) and shows, again, the convergence of the primitive and virials energy estimators for standard microcanonical MD, path integral MD with a single thermostat on each time slice, and the recommended thermostat on each degree of freedom. Clearly, the recommended scheme outperforms the other methods rather dramatically.

8 Summary

In summary, we have introduced the path integral formulation of quantum statistical mechanics in both the discrete and continuous formulations and shown how to compute path integrals using molecular dynamics. In particular, we have shown that a variable transformation that uncouples the harmonic bead-bead interaction term is necessary together with the coupling of the system to a thermostat on each degree of freedom in order to have an efficient scheme. We have introduced an efficient multiple time scale integrator for the path integral MD equations and, finally, we have extended the scheme to many-body systems in which the spin statistics can be neglected.

References

1. R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York (1965).
2. D. F. B. Tenhaaf, H. J. M. Vanbemmel, J. M. J. Vanleeuwen, M. Vanssarlooos and D. M. Ceperley, *Phys. Rev. B* **51**, 13039 (1995).
3. M. Boninsegni, C. Pierleoni and D. M. Ceperley, *Phys. Rev. Lett.* **72**, 1854 (1994).
4. G. Krilov, E. Sim and B. J. Berne, *Chem. Phys.* **268**, 21 (2001).
5. G. Krilov, E. Sim and B. J. Berne, *J. Chem. Phys.* **114**, 1075 (2001).
6. E. L. Pollock and D. M. Ceperley, *Phys. Rev. B* **30**, 2555 (1984).
7. M. Sprik, M. L. Klein and D. Chandler, *J. Chem. Phys.* **83**, 3942 (1985).
8. D. F. Coker, D. Thirumalai, and B. J. Berne, *J. Chem. Phys.* **86**, 5689 (1987).
9. D. M. Ceperley, *Rev. Mod. Phys.* **67**, 279 (1995).
10. R. W. Hall and B. J. Berne, *J. Chem. Phys.* **81**, 3641 (1984).
11. M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein, *J. Chem. Phys.* **99**, 2796 (1993).
12. M. E. Tuckerman, D. Marx, M. L. Klein and M. Parrinello, *J. Chem. Phys.* **104**, 5579 (1996).
13. R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
14. J. Cao and G. A. Voth, *J. Chem. Phys.* **99**, 10070 (1993).
15. J. Cao and G. A. Voth, *J. Chem. Phys.* **100**, 5106 (1994).
16. J. Cao and G. J. Martyna, *J. Chem. Phys.* **104**, 2028 (1994).
17. D. Marx, M. E. Tuckerman and G. J. Martyna, *Comp. Phys. Comm.* **118**, 166 (1999).
18. L. Schulman, *Techniques and Applications of Path Integration* John Wiley & Sons, New York (1981).
19. M. Herman, E. F. Bruskin, and B. J. Berne, *J. Chem. Phys.* **76**, 1347 (1982).
20. R. P. Feynman and H. Kleinert, *Phys. Rev. A* **34**, 5080 (1986).
21. G. J. Martyna, A. Hughes, and M. E. Tuckerman, *J. Chem. Phys.* **110**, 3275 (1999).
22. J. P. Neirotti, D. L. Freeman and J. D. Doll, *J. Chem. Phys.* **112**, 3990 (2000).
23. D. Chandler and P. G. Wolynes, *J. Chem. Phys.* **74**, 4078 (1981).
24. G. J. Martyna, M. E. Tuckerman and M. L. Klein, *J. Chem. Phys.* **97**, 2635 (1992).
25. Y. Liu and M. E. Tuckerman, *J. Chem. Phys.* **112**, 1685 (2000).
26. P. Minary and M. E. Tuckerman (to be submitted).
27. M. E. Tuckerman, Y. Liu, G. Ciccotti and G. J. Martyna, *J. Chem. Phys.* **115**, 1678 (2001).

28. J. Cao and B. J. Berne, *J. Chem. Phys.* **91**, 6359 (1989).
29. M. E. Tuckerman, G. J. Martyna and B. J. Berne, *J. Chem. Phys.* **97**, 1990 (1992).
30. G. J. Martyna, M. E. Tuckerman, D. J. Tobias and M. L. Klein, *Mol. Phys.* **87**, 1117 (1996).

Ab Initio Molecular Dynamics and Ab Initio Path Integrals

Mark E. Tuckerman

Department of Chemistry and Courant Institute of Mathematics
New York University, New York, NY 10003, USA
E-mail: mark.tuckerman@nyu.edu

1 Introduction

Modern theoretical methodology, aided by the advent of high speed computing, has advanced to a level that the microscopic details of chemical events can now be treated on a routine basis. One of the most important development in this area has been the so called *ab initio* molecular dynamics (AIMD) method,¹ which combines finite temperature atomistic molecular dynamics with internuclear forces obtained from accurate electronic structure calculations performed “on the fly” as the MD simulation proceeds. Finally, the combination of AIMD with the discrete path integral gives rise to a very powerful technique for studying chemical processes in which nuclear quantum effects play an important role.

AIMD has been used to study a wide variety of chemically interesting systems. These include, but are certainly not limited to, calculations of the structure and infrared spectroscopy of water,^{2,3} proton transport in aqueous acidic and basic environments,^{4–7} proton order/disorder and infrared spectroscopy of ice,^{8,9} structure of liquid silicates,^{10,11} structure and ionic solvation in liquid ammonia,^{12,13} calculation of NMR in proteins¹⁴ and structure of nucleic acids,¹⁵ Ziegler-Natta catalysis,¹⁶ and proton transport in methanol and methanol/water mixtures,¹⁷ to name just a few examples. The wide variety of applications attests to the power and flexibility of the AIMD and *ab initio* path integral (AIPi) approaches.

For reasons of computational efficiency, the most commonly employed approach to AIMD is based on a density functional representation of the electronic structure and expansion of the electronic orbitals in a plane wave basis set, and this is the approach on which we shall focus in this lecture. However, it is important to note that AIMD is a general approach, and a number of examples exist in the literature which employ more accurate or more empirical electronic structure methods^{18,19} as well as different basis sets.^{20–22}

This lecture is organized as follows. We shall begin with a brief review of the Born-Oppenheimer approximation and show how the AIMD method naturally emerges. We shall then describe the adiabatic dynamics approach of Car and Parrinello¹ and arrive at the Car-Parrinello equations of motion. We shall then show how to derive a path integral version of the Born-Oppenheimer approximation and describe the incorporation of the AIMD methodology into the path integral scheme. Finally, we shall show how the Car-Parrinello method can be employed to yield an efficient AIPi algorithm.

2 The Born-Oppenheimer Approximation and *Ab Initio* Molecular Dynamics

Consider a system of N nuclei described by coordinates, $\mathbf{R}_1, \dots, \mathbf{R}_N \equiv \mathbf{R}$, and momenta, $\mathbf{P}_1, \dots, \mathbf{P}_N \equiv \mathbf{P}$ and N_e electrons described by coordinates, $\mathbf{r}_1, \dots, \mathbf{r}_{N_e} \equiv \mathbf{r}$, momenta, $\mathbf{p}_1, \dots, \mathbf{p}_{N_e} \equiv \mathbf{p}$, and spin variables, $s_1, \dots, s_{N_e} \equiv s$ with a Hamiltonian of the form

$$H = \sum_{I=1}^N \frac{\mathbf{P}_I^2}{2M_I} + \sum_{i=1}^{N_e} \frac{\mathbf{p}_i^2}{2m} + \sum_{i>j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I>J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|}$$

$$\equiv T_N + T_e + V_{ee}(\mathbf{r}) + V_{NN}(\mathbf{R}) + V_{eN}(\mathbf{r}, \mathbf{R}) \quad (1)$$

where m is the mass of the electron, and $Z_I e$ is the charge on the I th nucleus. In the second line, T_N , T_e , V_{ee} , V_{NN} , and V_{eN} represent the nuclear and electron kinetic energy operators and electron-electron, electron-nuclear, and nuclear-nuclear interaction potential operators, respectively. We begin by looking for the eigenfunctions and eigenvalues of this Hamiltonian, given by:

$$[T_N + T_e + V_{ee}(\mathbf{r}) + V_{NN}(\mathbf{R}) + V_{eN}(\mathbf{r}, \mathbf{R})] \Psi(\mathbf{x}, \mathbf{R}) = E\Psi(\mathbf{x}, \mathbf{R}) \quad (2)$$

where $\mathbf{x} \equiv (\mathbf{r}, s)$ is the full collection of position and spin variables. Clearly, an exact solution of Eq. (2) is not possible. The Born-Oppenheimer approximation consists in the recognition that there is a strong separation of time scales between the electronic and nuclear motion, since the electrons are lighter than the nuclei by three orders of magnitude. In order to exploit this fact, we assume a solution of the form

$$\Psi(\mathbf{x}, \mathbf{R}) = \phi(\mathbf{x}, \mathbf{R})\chi(\mathbf{R}) \quad (3)$$

where $\chi(\mathbf{R})$ is a nuclear wavefunction and $\phi(\mathbf{x}, \mathbf{R})$ is an electronic wavefunction that depends parametrically on the nuclear positions. Note that

$$T_N\phi(\mathbf{x}, \mathbf{R})\chi(\mathbf{R})$$

$$= \frac{\hbar^2}{2} \sum_{I=1}^N \frac{1}{M_I} [\phi(\mathbf{x}, \mathbf{R})\nabla_I^2\chi(\mathbf{R}) + \chi(\mathbf{R})\nabla_I^2\phi(\mathbf{x}, \mathbf{R}) + 2\nabla_I\phi(\mathbf{x}, \mathbf{R}) \cdot \nabla_I\chi(\mathbf{R})] \quad (4)$$

The Born-Oppenheimer consists in neglecting $\nabla_I\phi(\mathbf{x}, \mathbf{R})$ terms with the justification that the nuclear wavefunction $\chi(\mathbf{R})$ is more localized than the electronic wavefunction, hence, we expect $\nabla_I\chi(\mathbf{R}) \gg \nabla_I\phi(\mathbf{x}, \mathbf{R})$. Substitution of Eq. (3) with the above approximation into Eq. (2) gives

$$[T_e + V_{ee}(\mathbf{r}) + V_{eN}(\mathbf{r}, \mathbf{R})] \phi(\mathbf{x}, \mathbf{R})\chi(\mathbf{R}) + \phi(\mathbf{x}, \mathbf{R})T_N\chi(\mathbf{R}) + V_{NN}(\mathbf{R})\phi(\mathbf{x}, \mathbf{R})\chi(\mathbf{R})$$

$$= E\phi(\mathbf{x}, \mathbf{R})\chi(\mathbf{R}) \quad (5)$$

Dividing by $\phi(\mathbf{x}, \mathbf{R})\chi(\mathbf{R})$ then gives:

$$\frac{[T_e + V_{ee}(\mathbf{r}) + V_{eN}(\mathbf{r}, \mathbf{R})] \phi(\mathbf{x}, \mathbf{R})}{\phi(\mathbf{x}, \mathbf{R})} = E - \frac{[T_N + V_{NN}(\mathbf{R})] \chi(\mathbf{R})}{\chi(\mathbf{R})} \quad (6)$$

From the above, it is clear that the left side can only be a function of \mathbf{R} alone. Let this function be denoted, $\varepsilon(\mathbf{R})$. Thus,

$$\frac{[T_e + V_{ee}(\mathbf{r}) + V_{eN}(\mathbf{r}, \mathbf{R})] \phi(\mathbf{x}, \mathbf{R})}{\phi(\mathbf{x}, \mathbf{R})} = \varepsilon(\mathbf{R})$$

$$[T_e + V_{ee}(\mathbf{r}) + V_{eN}(\mathbf{r}, \mathbf{R})] \phi(\mathbf{x}, \mathbf{R}) = \varepsilon(\mathbf{R}) \phi(\mathbf{x}, \mathbf{R}) \quad (7)$$

Eq. (7) is an electronic eigenvalue equation for an electronic Hamiltonian, $H_e(\mathbf{R}) = T_e + V_{ee}(\mathbf{r}) + V_{eN}(\mathbf{r}, \mathbf{R})$ which will yield a set of eigenfunction, $\varphi_n(\mathbf{x}, \mathbf{R})$ and eigenvalues, $\varepsilon_n(\mathbf{R})$, which depend parametrically on the nuclear positions, \mathbf{R} . For each solution, there will be a nuclear eigenvalue equation:

$$[T_N + V_{NN}(\mathbf{R}) + \varepsilon_n(\mathbf{R})] \chi(\mathbf{R}) = E \chi(\mathbf{R}) \quad (8)$$

Moreover, each electronic eigenvalue, $\varepsilon_n(\mathbf{R})$ will give rise to an electronic surface on which the nuclear dynamics described by the time-dependent Schrödinger equation for the time-dependent nuclear wave function $X(\mathbf{R}, t)$:

$$[T_N + V_{NN}(\mathbf{R}) + \varepsilon_n(\mathbf{R})] X(\mathbf{R}, t) = i\hbar \frac{\partial}{\partial t} X(\mathbf{R}, t) \quad (9)$$

will evolve. The physical interpretation of Eq. (9) is that the electrons respond instantaneously to the nuclear motion, therefore, it is sufficient to obtain a set of instantaneous electronic eigenvalues and eigenfunctions at each nuclear configuration, \mathbf{R} (hence the parametric dependence of $\varphi_n(\mathbf{x}, \mathbf{R})$ and $\varepsilon_n(\mathbf{R})$ on \mathbf{R}). The eigenvalues, in turn, give a family of (uncoupled) potential surfaces on which the nuclear wavefunction can evolve. These surfaces can become coupled by so called non-adiabatic effects, contained in the terms that have been neglected in the above derivation.

In many cases, non-adiabatic effects can be neglected, and we may consider motion *only* on the ground electronic surface described by:

$$[T_e + V_{ee}(\mathbf{r}) + V_{eN}(\mathbf{r}, \mathbf{R})] \varphi_0(\mathbf{x}, \mathbf{R}) = \varepsilon_0(\mathbf{R}) \varphi_0(\mathbf{x}, \mathbf{R})$$

$$[T_N + \varepsilon_0(\mathbf{R}) + V_{NN}(\mathbf{R})] X(\mathbf{R}, t) = i\hbar \frac{\partial}{\partial t} X(\mathbf{R}, t) \quad (10)$$

Moreover, if nuclear quantum effects can be neglected, then we may arrive at classical nuclear evolution by assuming $X(\mathbf{R}, t)$ is of the form

$$X(\mathbf{R}, t) = A(\mathbf{R}, t) e^{iS(\mathbf{R}, t)/\hbar} \quad (11)$$

and neglecting all terms involving \hbar , which yields an approximate equation for $S(\mathbf{R}, t)$:

$$H_N^{(n)}(\nabla_1 S, \dots, \nabla_N S, \mathbf{R}_1, \dots, \mathbf{R}_N) + \frac{\partial S}{\partial t} = 0 \quad (12)$$

which is just the classical Hamiltonian-Jacobi equation with

$$H_N^{(n)}(\mathbf{P}_1, \dots, \mathbf{P}_N, \mathbf{R}_1, \dots, \mathbf{R}_N) = \sum_{I=1}^N \frac{\mathbf{P}_I^2}{2M_I} + V_{nn}(\mathbf{R}) + \varepsilon_n(\mathbf{R}) \quad (13)$$

denoting the classical nuclear Hamiltonian. The Hamilton-Jacobi equation is equivalent to classical motion on the ground-state surface, $E_0(\mathbf{R}) = \varepsilon_0(\mathbf{R}) + V_{\text{NN}}(\mathbf{R})$ given by

$$M_I \ddot{\mathbf{R}}_I = -\nabla_I E_0(\mathbf{R}) \quad (14)$$

Note that the force $\nabla_I E_0(\mathbf{R})$ contains a term from the nuclear-nuclear repulsion and a term from the derivative of the electronic eigenvalue, $\varepsilon_0(\mathbf{R})$. Because of the Hellman-Feynman theorem, this term is equivalent to:

$$\nabla_I \varepsilon_0(\mathbf{R}) = \langle \varphi_0(\mathbf{R}) | \nabla_I H_e(\mathbf{R}) | \varphi_0(\mathbf{R}) \rangle \quad (15)$$

Finally, we need to specify how the electronic equation is to be solved to obtain the ground state eigenvalue, $\varepsilon_0(\mathbf{R})$. Again, an exact solution to the electronic problem is, in general, not possible. However, a useful approximation is given by the density functional theory, which states, via the Hohenberg-Kohn theorem, that the ground state energy, $\varepsilon_0(\mathbf{R})$ at a given nuclear configuration, \mathbf{R} is obtained by minimizing a certain functional, $\varepsilon[n]$, over all electronic densities

$$n(\mathbf{r}) = \sum_{s, s_2, \dots, s_{N_e}} \int d\mathbf{r}_2 \cdots d\mathbf{r}_{N_e} |\varphi_0(\mathbf{r}, s, \mathbf{r}_2, s_2, \dots, \mathbf{r}_{N_e}, s_{N_e})|^2 \quad (16)$$

(Here, \mathbf{r} and s represent a single position and spin variable, respectively.) A convenient form for this functional is given by the scheme of Kohn and Sham, in which a set of doubly occupied single-particle states, $\psi_i(\mathbf{r})$, $i = 1, \dots, N_e/2$, each containing a spin-up and a spin-down electron, is introduced. These are known as the Kohn-Sham (KS) orbitals. In terms of these orbitals, the density is given by

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 \quad (17)$$

and the functional takes the form

$$\varepsilon[\{\psi_i\}] = -\frac{\hbar^2}{2m} \sum_i \langle \psi_i | \nabla^2 | \psi_i \rangle + \frac{e^2}{2} \int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \varepsilon_{\text{xc}}[n] + \int d\mathbf{r} n(\mathbf{r}) V_{\text{en}}(\mathbf{r}, \mathbf{R}) \quad (18)$$

The first term in the functional represents the quantum kinetic energy, the second is the direct Coulomb term from Hartree-Fock theory, the third term is the exact (unknown) exchange and correlation energies, and the fourth term is the interaction of the electron density with the external potential due to the nuclei. This functional is minimized over the set of single-particle orbitals subject to an orthogonality condition

$$\langle \psi_i | \psi_j \rangle = \delta_{ij} \quad (19)$$

Moreover, in order to combine this minimization with the nuclear dynamics of Eq. (14), it is necessary to carry out the minimization at each nuclear configuration. Thus, if Eq. (14) is integrated in a MD calculation using a numerical integrator, then the minimization would need to be carried out at each step of the MD simulation and the forces computed using the orbitals thus obtained.

In 1985, Car and Parrinello showed that this coupling between nuclear time evolution and electronic minimization could be treated efficiently via an implicit adiabatic dynamics approach.¹ In their scheme, a fictitious dynamics for the electronic orbitals is invented which, given orbitals initially at the minimum for an initial nuclear configuration, would

allow them to follow the nuclear motion adiabatically, and thus, be automatically at their approximately minimized configuration at each step of the MD evolution. This dynamics is controlled by introducing a set of orbitals “velocities” $\{\dot{\psi}_i(\mathbf{r})\}$ and a fictitious electronic “kinetic energy” (not to be confused with the true quantum kinetic energy) given by

$$K_{\text{fict}} = \mu \sum_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle \quad (20)$$

where μ is a fictitious mass parameter (having units of energy \times time²) that controls the time scale on which the electrons “evolve” and introducing a Lagrangian that includes the orbitals as fictitious dynamical degrees of freedom:

$$L = \mu \sum_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle + \frac{1}{2} \sum_{I=1}^N M_I \dot{\mathbf{R}}_I^2 - E[\{\psi\}, \mathbf{R}] + \sum_{i,j} [\Lambda_{ij} (\langle \psi_i | \psi_j \rangle - \delta_{ij})] \quad (21)$$

where $E[\{\psi\}, \mathbf{R}] = \varepsilon[\{\psi\}, \mathbf{R}] + V_{\text{nn}}(\mathbf{R})$. The matrix Λ_{ij} is a set of Lagrange multipliers introduced in order to ensure that the condition $\langle \psi_i | \psi_j \rangle = \delta_{ij}$ is satisfied dynamically as a constraint. The Euler-Lagrange equations

$$\begin{aligned} \frac{d}{dt} \left(\frac{\delta L}{\delta \dot{\psi}_i^*(\mathbf{r})} \right) - \frac{\delta L}{\delta \psi_i^*(\mathbf{r})} &= 0 \\ \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{R}}_I} \right) - \frac{\partial L}{\partial \mathbf{R}_I} &= 0 \end{aligned} \quad (22)$$

gives the following coupled dynamical equations of motion:

$$\begin{aligned} M_I \ddot{\mathbf{R}}_I &= -\nabla_I E[\{\psi\}, \mathbf{R}] \\ \mu \ddot{\psi}_i(\mathbf{r}) &= -\frac{\delta}{\delta \psi_i^*(\mathbf{r})} E[\{\psi\}, \mathbf{R}] + \sum_j \Lambda_j \psi_j(\mathbf{r}) \end{aligned} \quad (23)$$

These are known as the Car-Parrinello (CP) equations, and they form the basis of the AIMD method. The electronic equation can also be written in an abstract bra-ket form as:

$$\mu |\ddot{\psi}_i\rangle = -\frac{\partial E}{\partial \langle \psi_i |} + \sum_j \Lambda_{ij} |\psi_j\rangle \quad (24)$$

Below, we present an algorithm for integrating the CP equations subject to the orthogonality constraint based on the velocity Verlet scheme derived from the Liouville operator formalism in the path integral MD lecture.³²

Beginning with an initially minimized set of Kohn-Sham orbitals, $\{|\psi_i(0)\rangle\}$ corresponding to an initial nuclear configuration, $\mathbf{R}(0)$ and initial velocities, $\{\dot{\psi}_i(0)\}$, $\dot{\mathbf{R}}(0)$, the first step is a velocity update:

$$\begin{aligned} |\dot{\psi}_i^{(1)}(0)\rangle &= |\dot{\psi}_i(0)\rangle + \frac{\Delta t}{2\mu} |\varphi_i(0)\rangle \quad i = 1, \dots, \frac{N_e}{2} \\ \dot{\mathbf{R}}_I(\Delta t/2) &= \dot{\mathbf{R}}_I(0) + \frac{\Delta t}{2M_I} \mathbf{F}_I(0) \quad I = 1, \dots, N \end{aligned} \quad (25)$$

followed by a position/orbital update:

$$|\tilde{\psi}_i\rangle = |\psi_i(0)\rangle + \Delta t |\dot{\psi}_i^{(1)}\rangle \quad i = 1, \dots, \frac{N_e}{2}$$

$$\mathbf{R}_I(\Delta t) = \mathbf{R}_I(0) + \Delta t \dot{\mathbf{R}}_I(\Delta t/2) \quad I = 1, \dots, N \quad (26)$$

where $|\varphi_i(0)\rangle = (\partial E / \partial \langle \psi_i |)|_{t=0}$ is the initial force on the orbital, $|\psi_i\rangle$. At this point, we do not yet have the orbitals at $t = \Delta t$ or orbital velocities at $t = \Delta t/2$ because the constraint force $\Lambda_{ij}|\psi_j\rangle$ needs to be applied to both the orbitals and orbital velocities. In order to do this, we need to determine the Lagrange multiplier matrix, which is accomplished by enforcing the orthogonality constraint on the orbitals at $t = \Delta t$:

$$\langle \psi_i(\Delta t) | \psi_j(\Delta t) \rangle = \delta_{ij} \quad (27)$$

where

$$|\psi_i(\Delta t)\rangle = |\tilde{\psi}_i\rangle + \sum_j X_{ij} |\psi_j(0)\rangle \quad (28)$$

where $X_{ij} = (\Delta t^2 / 2\mu) \Lambda_{ij}$. Substituting Eq. (28) into Eq. (27) yields a matrix equation for the Lagrange multipliers:

$$\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}\mathbf{B} + \mathbf{B}^\dagger\mathbf{X}^\dagger + \mathbf{A} = \mathbf{I} \quad (29)$$

where $A_{ij} = \langle \tilde{\psi}_i | \tilde{\psi}_j \rangle$ and $B_{ij} = \langle \psi_i(0) | \tilde{\psi}_j \rangle$. Noting that $\mathbf{A} = \mathbf{I} + \mathcal{O}(\Delta t^2)$ and $\mathbf{B} = \mathbf{I} + \mathcal{O}(\Delta t)$, the matrix equation can be solved iteratively via

$$\mathbf{X}_{n+1} = \frac{1}{2} [\mathbf{I} - \mathbf{A} + \mathbf{X}_n(\mathbf{I} - \mathbf{B}) + (\mathbf{I} - \mathbf{B}^\dagger)\mathbf{X}_n^\dagger - \mathbf{X}_n^2] \quad (30)$$

starting from an initial guess

$$\mathbf{X}_0 = \frac{1}{2}(\mathbf{I} - \mathbf{A}) \quad (31)$$

Once the matrix X_{ij} is obtained, the orbitals are updated using Eq. (28) and an orbital velocity update

$$|\dot{\psi}_i^{(2)}\rangle = |\dot{\psi}_i^{(1)}\rangle + \frac{1}{\Delta t} \sum_j X_{ij} |\psi_j(0)\rangle \quad (32)$$

is performed.

At this point, the new orbitals and nuclear forces, $|\varphi_i(\Delta t)\rangle$ and $\mathbf{F}_I(\Delta t)$ are calculated, and a velocity update of the form

$$|\dot{\psi}_i^{(3)}\rangle = |\dot{\psi}_i^{(2)}\rangle + \frac{\Delta t}{2\mu} |\varphi_i(\Delta t)\rangle \quad i = 1, \dots, \frac{N_e}{2}$$

$$\dot{\mathbf{R}}_I(\Delta t) = \dot{\mathbf{R}}_I(\Delta t/2) + \frac{\Delta t}{2M_I} \mathbf{F}_I(\Delta t) \quad (33)$$

is performed. Again, we do not have the final orbital velocities until an appropriate constraint force is applied. For the velocities, the appropriate force is the first time derivative of the orthogonality constraint:

$$\langle \psi_i(\Delta t) | \dot{\psi}_j(\Delta t) \rangle + \langle \dot{\psi}_i(\Delta t) | \psi_j(\Delta t) \rangle = 0 \quad (34)$$

where

$$|\dot{\psi}_i(\Delta t)\rangle = |\dot{\psi}_i^{(3)}\rangle + \sum_j Y_{ij} |\psi_i(\Delta t)\rangle \quad (35)$$

and Y_{ij} are a new set of Lagrange multipliers for enforcing the condition Eq. (34). Substituting Eq. (35) into Eq. (34) gives a simple solution for Y_{ij} :

$$Y = -\frac{1}{2} (C + C^\dagger) \quad (36)$$

where $C_{ij} = \langle \psi_i(\Delta t) | \dot{\psi}_i^{(3)} \rangle$. Given the matrix, Y_{ij} , the final orbital velocities are obtained via Eq. (35).

3 Plane Wave Basis Sets

In the traditional CP approach, periodic boundary conditions are employed, and the orbitals, $\{\psi_i(\mathbf{r})\}$ become Bloch functions, $\{\psi_{i\mathbf{k}}(\mathbf{r})\}$, where \mathbf{k} samples the first Brillouin zone. These Bloch functions are expanded in a plane wave basis:

$$\psi_{i\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{g}} c_{\mathbf{g}}^i e^{i\mathbf{g}\cdot\mathbf{r}} \quad (37)$$

where c_g^i is a set of expansion coefficients, Ω is the system volume, $\mathbf{g} = 2\pi\mathbf{h}^{-1}\hat{\mathbf{g}}$ is a reciprocal lattice vector, \mathbf{h} is the cell matrix whose columns are the cell vectors, and $\hat{\mathbf{g}}$ is a vector of integers. An advantage of plane waves is that the sums needed to go back and forth between reciprocal space and real space can be performed efficiently using fast Fourier transforms (FFTs). For the applications to be considered herein, which are largely concerned with nonmetallic systems, it is sufficient to consider only the Γ -point ($\mathbf{k} = (0, 0, 0)$), so that the plane wave expansion becomes

$$\psi_i(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{g}} c_{\mathbf{g}}^i e^{i\mathbf{g}\cdot\mathbf{r}} \quad (38)$$

In this case, the coefficients become dynamical variables, and the CP equations take the form:

$$\begin{aligned} M_I \ddot{\mathbf{R}}_I &= -\frac{\partial E}{\partial \mathbf{R}_I} \\ \mu \ddot{c}_{\mathbf{g}}^i &= -\frac{\partial E}{\partial c_{\mathbf{g}}^{i*}} + \sum_j \Lambda_{ij} c_{\mathbf{g}}^j \end{aligned} \quad (39)$$

At the Γ -point, the orbitals can always be chosen to be real functions. Therefore, the plane-wave expansion coefficients satisfy the following property

$$c_{\mathbf{g}}^{i*} = c_{-\mathbf{g}}^i \quad (40)$$

which requires keeping only half of the full set of plane-wave expansion coefficients. In actual applications, plane waves up to a given cutoff, $|\mathbf{g}|^2/2 < E_{\text{cut}}$, only are kept. Similarly, the density $n(\mathbf{r})$ given by Eq. (17) can also be expanded in a plane wave basis:

$$n(\mathbf{r}) = \frac{1}{\Omega} \sum_{\mathbf{g}} n_{\mathbf{g}} e^{i\mathbf{g}\cdot\mathbf{r}} \quad (41)$$

However, since $n(\mathbf{r})$ is obtained as a square of the KS orbitals, the cutoff needed for this expansion is $4E_{\text{cut}}$ for consistency with the orbital expansion.

Using Eqs. (37) and (41) and the orthogonality of the plane waves, it is straightforward to compute the various energy terms. For example, the kinetic energy can be easily shown to be

$$\varepsilon_{\text{KE}} = -\frac{1}{2} \sum_i \int d\mathbf{r} \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) = \frac{1}{2} \sum_i \sum_{\mathbf{g}} g^2 |c_{\mathbf{g}}^i|^2 \quad (42)$$

where $g = |\mathbf{g}|$. Similarly, the Hartree energy becomes

$$\varepsilon_{\text{H}} = \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} = \frac{1}{\Omega} \sum_{\mathbf{g}} \frac{4\pi}{g^2} |n_{\mathbf{g}}|^2 \quad (43)$$

where the summation excludes the $\mathbf{g} = (0, 0, 0)$ term.

The exchange and correlation energy, $\varepsilon_{\text{xc}}[n]$ is generally treated within a local density (LDA) or generalized gradient approximations (GGA) wherein it is assumed to take the approximate form

$$\varepsilon_{\text{xc}}[n] \approx \int d\mathbf{r} f(n(\mathbf{r}), \nabla n(\mathbf{r}), \nabla^2 n(\mathbf{r})) = \frac{\Omega}{N_{\text{grid}}} \sum_{\mathbf{r}} f(n(\mathbf{r}), \nabla n(\mathbf{r}), \nabla^2 n(\mathbf{r})) \quad (44)$$

where the sum is taken over a set of N_{grid} real-space grid points. The gradient and (if needed) the Laplacian of the density can be computed efficiently using FFTs:

$$\begin{aligned} \nabla n(\mathbf{r}) &= \sum_{\mathbf{g}} i\mathbf{g} e^{i\mathbf{g}\cdot\mathbf{r}} \sum_{\mathbf{r}'} n(\mathbf{r}') e^{-i\mathbf{g}\cdot\mathbf{r}'} \\ \nabla^2 n(\mathbf{r}) &= - \sum_{\mathbf{g}} g^2 e^{i\mathbf{g}\cdot\mathbf{r}} \sum_{\mathbf{r}'} n(\mathbf{r}') e^{-i\mathbf{g}\cdot\mathbf{r}'} \end{aligned} \quad (45)$$

The external energy is made somewhat complicated by the fact that, in a plane wave basis, very large basis sets are needed to treat the rapid spatial fluctuations of core electrons. Therefore, core electrons are often replaced by atomic pseudopotentials^{23–25} are augmented plane wave techniques.²⁶ Here, we shall discuss the former. In the atomic pseudopotential scheme, the nucleus plus the core electrons are treated in a frozen core type approximation as an ion carrying only the valence charge. In order to make this approximation, the valence orbitals, which, in principle must be orthogonal to the core orbitals, must see a different pseudopotential for each angular momentum component in the core, which means that the pseudopotential must be nonlocal. To see how this comes about, we consider a potential operator of the form

$$\hat{V}_{\text{pseud}} = \sum_{l=0}^{\infty} \sum_{m=-l}^l v_l(r) |lm\rangle \langle lm| \quad (46)$$

where r is the distance from the ion, and $|lm\rangle \langle lm|$ is a projection operator onto each angular momentum component. In order to truncate the infinite sum over l in Eq. (46), we assume that for some $l \geq \bar{l}$, $v_l(r) = v_{\bar{l}}(r)$ and add and subtract the function $v_{\bar{l}}(r)$ in Eq.

(46):

$$\begin{aligned}
\hat{V}_{\text{pseud}} &= \sum_{l=0}^{\infty} \sum_{m=-l}^l (v_l(r) - v_{\bar{l}}(r)) |lm\rangle \langle lm| + v_{\bar{l}}(r) \sum_{l=0}^{\infty} \sum_{m=-l}^l |lm\rangle \langle lm| \\
&= \sum_{l=0}^{\infty} \sum_{m=-l}^l (v_l(r) - v_{\bar{l}}(r)) |lm\rangle \langle lm| + v_{\bar{l}}(r) \\
&\approx \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^l \Delta v_l(r) |lm\rangle \langle lm|
\end{aligned} \tag{47}$$

where the second line follows from the fact that the sum of the projection operators is unity, $\Delta v_l(r) = v_l(r) - v_{\bar{l}}(r)$, and the sum in the third line is truncated before $\Delta v_l(r) = 0$. The complete pseudopotential operator will be

$$\hat{V}_{\text{pseud}}(r; \mathbf{R}_1, \dots, \mathbf{R}_N) = \sum_{I=1}^N \left[v_{\text{loc}}(|\mathbf{r} - \mathbf{R}_I|) + \sum_{l=0}^{\bar{l}-1} \Delta v_l(|\mathbf{r} - \mathbf{R}_I|) |lm\rangle \langle lm| \right] \tag{48}$$

where $v_{\text{loc}}(r) \equiv v_{\bar{l}}(r)$ is known as the local part of the pseudopotential (having no projection operator attached to it). Now, the external energy, being derived from the ground-state expectation value of a one-body operator, will be given by

$$\varepsilon_{\text{ext}} = \sum_i \langle \psi_i | \hat{V}_{\text{pseud}} | \psi_i \rangle \tag{49}$$

The first (local) term gives simply a local energy of the form

$$\varepsilon_{\text{loc}} = \sum_{I=1}^N \int d\mathbf{r} n(\mathbf{r}) v_{\text{loc}}(|\mathbf{r} - \mathbf{R}_I|) \tag{50}$$

which can be evaluated in reciprocal space as

$$\varepsilon_{\text{loc}} = \frac{1}{\Omega} \sum_{I=1}^N \sum_{\mathbf{g}} n_{\mathbf{g}}^* \tilde{v}_{\text{loc}}(\mathbf{g}) e^{-i\mathbf{g} \cdot \mathbf{R}_I} \tag{51}$$

where $\tilde{V}_{\text{loc}}(\mathbf{g})$ is the Fourier transform of the local potential. Note that at $\mathbf{g} = (0, 0, 0)$, only the nonsingular part of $\tilde{v}_{\text{loc}}(\mathbf{g})$ contributes. In the evaluation of the local term, it is often convenient to add and subtract a long-range term of the form $Z_I \text{erf}(\alpha_I r)/r$, where $\text{erf}(x)$ is the error function, for each ion in order to obtain the nonsingular part explicitly and a residual short-range function $\bar{v}_{\text{loc}}(|\mathbf{r} - \mathbf{R}_I|) = v_{\text{loc}}(|\mathbf{r} - \mathbf{R}_I|) - Z_I \text{erf}(\alpha_I |\mathbf{r} - \mathbf{R}_I|)/|\mathbf{r} - \mathbf{R}_I|$ for each ionic core. For the nonlocal contribution, Eq. (38) is substituted into Eq. (48), an expansion of the plane waves in terms of spherical Bessel functions and spherical harmonics is made, and, after some algebra, one obtains

$$\varepsilon_{\text{NL}} = \sum_i \sum_I \sum_{\mathbf{g}, \mathbf{g}'} e^{-i\mathbf{g} \cdot \mathbf{R}_I} c_{\mathbf{g}}^{i*} v_{\text{NL}}(\mathbf{g}, \mathbf{g}') c_{\mathbf{g}'}^i e^{i\mathbf{g}' \cdot \mathbf{R}_I} \tag{52}$$

where

$$v_{\text{NL}}(\mathbf{g}, \mathbf{g}') = (4\pi)^2 \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^l \int dr r^2 j_l(gr) j_l(g'r) \Delta v_l(\mathbf{r}) Y_{lm}(\theta_{\mathbf{g}}, \phi_{\mathbf{g}}) Y_{lm}^*(\theta_{\mathbf{g}'}, \phi_{\mathbf{g}'}) \quad (53)$$

where $\theta_{\mathbf{g}}$ and $\phi_{\mathbf{g}}$ are the spherical polar angles associated with the vector \mathbf{g} , and same for $\theta_{\mathbf{g}'}$ and $\phi_{\mathbf{g}'}$. Eq. (53) shows that the evaluation of the nonlocal energy can be quite computationally expensive. It also shows, however, that the matrix element is *almost* separable in \mathbf{g} and \mathbf{g}' dependent terms. A fully separable approximation can be obtained by writing

$$\begin{aligned} v_{\text{NL}}(\mathbf{g}, \mathbf{g}') &= (4\pi)^2 \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^l \int dr r^2 \int dr' r'^2 j_l(gr) j_l(g'r') \Delta v_l(\mathbf{r}) \frac{\delta(r - r')}{rr'} \\ &\times Y_{lm}(\theta_{\mathbf{g}}, \phi_{\mathbf{g}}) Y_{lm}^*(\theta_{\mathbf{g}'}, \phi_{\mathbf{g}'}) \end{aligned} \quad (54)$$

where a radial δ -function has been introduced. Now, the δ -function is expanded in terms of a set of radial eigenfunctions (usually taken to be those of the Hamiltonian from which the pseudopotential is obtained) for each angular momentum channel

$$\frac{\delta(r - r')}{rr'} = \sum_{n=0}^{\infty} \phi_{nl}^*(r) \phi_{nl}(r') \quad (55)$$

If this expansion is now substituted into Eq. (54), the result is

$$\begin{aligned} v_{\text{NL}}(\mathbf{g}, \mathbf{g}') &= (4\pi)^2 \sum_{n=0}^{\infty} \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^l \left[\int dr r^2 j_l(gr) \Delta v_l(\mathbf{r}) \phi_{nl}^*(r) Y_{lm}(\theta_{\mathbf{g}}, \phi_{\mathbf{g}}) \right] \\ &\times \left[\int dr' r'^2 j_l(g'r') Y_{lm}^*(\theta_{\mathbf{g}'}, \phi_{\mathbf{g}'}) \phi_{nl}(r') \right] \end{aligned} \quad (56)$$

which is now fully separable at the expense of another infinite sum that needs to be truncated. Although the sum over n can be truncated after any number of terms, the so called *Kleinman-Bylander approximation*²⁷ is the result of truncating it at just a single term. The result of this truncation can be shown to yield the approximation form:

$$\begin{aligned} v_{\text{NL}}(\mathbf{g}, \mathbf{g}') &\approx (4\pi)^2 \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^l N_{lm}^{-1} \left[\int dr r^2 j_l(gr) \Delta v_l(\mathbf{r}) \phi_l^*(r) Y_{lm}(\theta_{\mathbf{g}}, \phi_{\mathbf{g}}) \right] \\ &\times \left[\int dr' r'^2 \Delta v_l(r') j_l(g'r') Y_{lm}^*(\theta_{\mathbf{g}'}, \phi_{\mathbf{g}'}) \phi_l(r') \right] \end{aligned} \quad (57)$$

where

$$N_{lm} = \int dr r^2 \phi_l^*(r) \Delta v_l(r) \phi_l(r) \quad (58)$$

and $\phi_l(r) \equiv \phi_{0l}(r)$. Finally, substituting Eq. (57) into Eq. (52) gives the nonlocal energy as

$$\varepsilon_{\text{NL}} = \sum_{i=1}^{N_e} \sum_{I=1}^N \sum_{l=0}^{\bar{l}-1} \sum_{m=-l}^l Z_{iIlm}^* Z_{iIlm} \quad (59)$$

where

$$Z_{Ilm} = \sum_{\mathbf{g}} c_{\mathbf{g}}^i e^{i\mathbf{g}\cdot\mathbf{R}_I} \tilde{F}_{lm}(\mathbf{g}) \quad (60)$$

and

$$\tilde{F}_{lm}(\mathbf{g}) = 4\pi N_{lm}^{-1/2} \int dr r^2 j_l(gr) \Delta v_l(ur) \phi_l(r) Y_{lm}(\theta_{\mathbf{g}}, \phi_{\mathbf{g}}) \quad (61)$$

Having specified all of the energy terms in terms of the plane wave expansion, these expressions can be differentiated in order to obtain the forces on the ions and coefficients needed for the CP equations of motion.

The last issue on which we shall touch briefly is that of boundary conditions within the plane wave description. Plane waves naturally describe a situation in which three-dimensional periodic boundary conditions are to be used, such as in solids and liquids. What if we wish to treat systems, such as cluster, surface, or wire, in which one or more boundaries is *not* periodic? It turns out that such situations can be treated rather easily within the plane wave description using a technique developed by Martyna and Tuckerman,^{28,29} which involves the use of a screening function in the long-range energy terms, i.e. the Hartree and local pseudopotential terms. The idea is to use the so called first image form of the average energy in order to form an approximation to a cluster, wire, or surface system, whose error can be controlled by the dimensions of the simulation cell. Thus, given any density, $n(\mathbf{r})$ and any interaction potential, $\phi(\mathbf{r} - \mathbf{r}')$, the average potential energy in this approximation is given by

$$\langle \phi \rangle^{(1)} = \frac{1}{2\Omega} \sum_{\mathbf{g}} |n_{\mathbf{g}}|^2 \bar{\phi}(-\mathbf{g}) \quad (62)$$

where $\bar{\phi}(\mathbf{g})$ is a Fourier expansion coefficient of the potential given by

$$\begin{aligned} \bar{\phi}(\mathbf{g}) &= \int_{-L_c/2}^{L_c/2} dz \int_{-L_b/2}^{L_b/2} dy \int_{-L_a/2}^{L_a/2} dx \phi(\mathbf{r}) e^{-i\mathbf{g}\cdot\mathbf{r}} && \text{(Cluster)} \\ \bar{\phi}(\mathbf{g}) &= \int_{-L_c/2}^{L_c/2} dz \int_{-L_b/2}^{L_b/2} dy \int_{-\infty}^{\infty} dx \phi(\mathbf{r}) e^{-i\mathbf{g}\cdot\mathbf{r}} && \text{(Wire)} \\ \bar{\phi}(\mathbf{g}) &= \int_{-L_c/2}^{L_c/2} dz \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \phi(\mathbf{r}) e^{-i\mathbf{g}\cdot\mathbf{r}} && \text{(Surface)} \end{aligned} \quad (63)$$

Here, L_a , L_b , and L_c are the dimensions of the simulation cell (assumed to be orthorhombic for simplicity) in the x , y , and z directions. In order to have an expression that is easily computed within the plane wave description, consider two functions $\phi^{(\text{long})}(\mathbf{r})$ and $\phi^{(\text{short})}(\mathbf{r})$, which are assumed to be the long and short range contributions to the total potential, such that

$$\phi(\mathbf{r}) = \phi^{(\text{long})}(\mathbf{r}) + \phi^{(\text{short})}(\mathbf{r})$$

$$\bar{\phi}(\mathbf{g}) = \bar{\phi}^{(\text{long})}(\mathbf{g}) + \bar{\phi}^{(\text{short})}(\mathbf{g}). \quad (64)$$

We require that $\phi^{(\text{short})}(\mathbf{r})$ vanish exponentially quickly at large distances from the center of the parallelepiped and that $\phi^{(\text{long})}(\mathbf{r})$ contain the long range dependence of the full

potential, $\phi(\mathbf{r})$. With these two requirements, it is possible to write

$$\begin{aligned}\bar{\phi}^{(\text{short})}(\mathbf{g}) &= \int_{D(\Omega)} d\mathbf{r} \exp(-i\mathbf{g} \cdot \mathbf{r}) \phi^{(\text{short})}(\mathbf{r}) \\ &= \int_{\text{all space}} d\mathbf{r} \exp(-i\mathbf{g} \cdot \mathbf{r}) \phi^{(\text{short})}(\mathbf{r}) + \epsilon(\mathbf{g}) \\ &= \tilde{\phi}^{(\text{short})}(\mathbf{g}) + \epsilon(\mathbf{g})\end{aligned}\quad (65)$$

with exponentially small error, $\epsilon(\mathbf{g})$, provided the range of $\phi^{(\text{short})}(\mathbf{r})$ is chosen small compared size of the parallelepiped, i.e. Eq. (65) defines the properties that the heretofore arbitrary function, $\phi^{(\text{short})}(\mathbf{r})$, must satisfy. Therefore, $\phi^{(\text{short})}(\mathbf{r})$, will be made a function of a convergence parameter, α , which can be used to adjust the range of $\phi^{(\text{short})}(\mathbf{r})$ such that $\epsilon(\mathbf{g}) \sim 0$ and the error, $\epsilon(\mathbf{g})$, will be neglected in the following.

The function, $\tilde{\phi}^{(\text{short})}(\mathbf{g})$, is the Fourier transform of $\phi^{(\text{short})}(\mathbf{r})$ evaluated at the quantized g-vector. Therefore,

$$\begin{aligned}\bar{\phi}(\mathbf{g}) &= \bar{\phi}^{(\text{long})}(\mathbf{g}) + \tilde{\phi}^{(\text{short})}(\mathbf{g}) \\ &= \bar{\phi}^{(\text{long})}(\mathbf{g}) - \tilde{\phi}^{(\text{long})}(\mathbf{g}) + \tilde{\phi}^{(\text{short})}(\mathbf{g}) + \tilde{\phi}^{(\text{long})}(\mathbf{g}) \\ &= \hat{\phi}^{(\text{screen})}(\mathbf{g}) + \tilde{\phi}(\mathbf{g})\end{aligned}\quad (66)$$

where $\tilde{\phi}(\mathbf{g}) = \tilde{\phi}^{(\text{short})}(\mathbf{g}) + \tilde{\phi}^{(\text{long})}(\mathbf{g})$ is the Fourier transform of the full potential, $\phi(\mathbf{r}) = \phi^{(\text{short})}(\mathbf{r}) + \phi^{(\text{long})}(\mathbf{r})$, evaluated at the quantized g-vector and

$$\hat{\phi}^{(\text{screen})}(\mathbf{g}) = \bar{\phi}^{(\text{long})}(\mathbf{g}) - \tilde{\phi}^{(\text{long})}(\mathbf{g}). \quad (67)$$

This result, Eqs. (67) and Eqs. (66), leads to

$$\langle \phi \rangle = \frac{1}{2\Omega} \sum_{\mathbf{g}} |\bar{n}(\mathbf{g})|^2 [\tilde{\phi}(-\mathbf{g}) + \hat{\phi}^{(\text{screen})}(-\mathbf{g})] \quad (68)$$

The new function appearing in the average potential energy, Eq. (68), is the difference between the Fourier series and Fourier transform form of the long range part of the potential energy evaluated at the quantized g-vector (cf Eq.(67)) and will be referred to as the screening function because it is constructed to “screen” the interaction of the system with an infinite array of periodic images. The specific case of the Coulomb potential,

$$\phi(\mathbf{r}) = \frac{1}{r} \quad (69)$$

can be separated into short and long range components via

$$\frac{1}{r} = \frac{\text{erf}(\alpha r)}{r} + \frac{\text{erfc}(\alpha r)}{r} \quad (70)$$

where the first term is long range. Here, α is an arbitrary convergence parameter. The screening function for the cluster case is easily computed by introducing an FFT grid and performing the integration numerically. For the wire and surface cases, analytical expressions can be worked out and are given by

$$\begin{aligned}
\bar{\phi}^{(\text{screen})}(\mathbf{g}) = & \frac{4\pi}{g^2} e^{-g^2/4\alpha^2} - \frac{4\pi}{g^2} \left\{ \cos\left(\frac{g_c L_c}{2}\right) \right. \\
& \times \left[\exp\left(-\frac{g_s L_c}{2}\right) - \frac{1}{2} \exp\left(-\frac{g_s L_c}{2}\right) \operatorname{erfc}\left(\frac{\alpha^2 L_c - g_s}{2\alpha}\right) \right. \\
& - \frac{1}{2} \exp\left(\frac{g_s L_c}{2}\right) \operatorname{erfc}\left(\frac{\alpha^2 L_c + g_s}{2\alpha}\right) \left. \right] \\
& \left. + \exp\left(-\frac{g^2}{4\alpha^2}\right) \operatorname{Re} \left[\operatorname{erfc}\left(\frac{\alpha^2 L_c + ig_c}{2\alpha}\right) \right] \right\} - \frac{4\pi}{g^2} e^{-g^2/4\alpha^2}
\end{aligned} \tag{71}$$

(Surface)

$$\begin{aligned}
\bar{\phi}^{(\text{screen,Coul})}(\mathbf{g}) = & \frac{4\pi}{g^2} [\exp(-g^2/4\alpha^2) E(\alpha, L_b, g_b) E(\alpha, L_c, g_c) \\
& + \cos\left(\frac{g_b L_b}{2}\right) \frac{4\sqrt{\pi}}{\alpha L_b} \exp(-g_c^2/4\alpha^2) I(\alpha, L_b, L_c, g_c) \\
& + \cos\left(\frac{g_c L_c}{2}\right) \frac{4\sqrt{\pi}}{\alpha L_c} \exp(-g_b^2/4\alpha^2) I(\alpha, L_c, L_b, g_b)] \\
& - \frac{4\pi}{g^2} e^{-g^2/4\alpha^2}
\end{aligned} \tag{72}$$

(Wire)

where

$$I(\alpha, L_1, L_2, g) = \int_0^{\alpha L_1/2} dx x e^{-g_a^2 L_1^2/16x^2} e^{-x^2} E\left(\frac{2x}{L_1}, L_2, g\right) \tag{73}$$

and

$$E(\lambda, L, g) = \operatorname{erf}\left(\frac{\lambda^2 L + ig}{2\lambda}\right) \tag{74}$$

where $\mathbf{g} = (g_a, g_b, g_c)$ and $g_s = \sqrt{g_a^2 + g_b^2}$. The one-dimensional integrals in Eq. (73) are well suited to be performed by Gaussian quadrature techniques.

4 The Path Integral Born-Oppenheimer Approximation and *Ab Initio* Path Integral Molecular Dynamics

The Born-Oppenheimer approximation described above can be recast in terms of imaginary time path integrals. Using the continuous notation previously introduced, the partition function can be written as

$$\begin{aligned}
Z(N, V, T) = & \oint \mathcal{D}\mathbf{R}(\tau) \oint \mathcal{D}\mathbf{r}(\tau) \\
& \times \exp \left\{ -\frac{1}{\hbar} \int_0^{\beta\hbar} d\tau \left[T_N(\dot{\mathbf{R}}(\tau) + T_e(\dot{\mathbf{r}}(\tau)) + V_{ee}(\mathbf{r}(\tau)) + V_{NN}(\mathbf{R}(\tau)) + V_{eN}(\mathbf{r}(\tau), \mathbf{R}(\tau)) \right] \right\}
\end{aligned} \tag{75}$$

If the above path integral is approached using the idea of influence functionals,³⁰ then Eq. (75) is written as

$$Z(N, V, T) = \oint \mathcal{D}\mathbf{R}(\tau) \exp \left\{ -\frac{1}{\hbar} \int_0^{\beta\hbar} [T_N(\dot{\mathbf{R}}(\tau)) + V_{NN}(\mathbf{R}(\tau))] \right\} F[\mathbf{R}(\tau)]$$

$$F[\mathbf{R}(\tau)] = \oint \mathcal{D}\mathbf{r}(\tau) \exp \left\{ -\frac{1}{\hbar} \int_0^{\beta\hbar} [T_e(\dot{\mathbf{r}}(\tau)) + V_{ee}(\mathbf{r}(\tau)) + V_{eN}(\mathbf{r}(\tau), \mathbf{R}(\tau))] \right\} \quad (76)$$

where $F[\mathbf{R}(\tau)]$ is known as the *influence functional*. From the form of $F[\mathbf{R}(\tau)]$, it is clear that this quantity is a partition function for the electronic subsystem along a given nuclear path $\mathbf{R}(\tau)$. At a fixed nuclear configuration, \mathbf{R} , this partition function could be computed from the electronic eigenvalues, $\varepsilon_n(\mathbf{R})$ and would be related to the negative exponential of the free energy:

$$F(\mathbf{R}) = \sum_n e^{-\beta\varepsilon_n(\mathbf{R})} = e^{-\beta A(\mathbf{R})} \quad (77)$$

where $A(\mathbf{R})$ is the free energy at nuclear configuration, \mathbf{R} . In the Born-Oppenheimer approximation, adiabaticity must be assumed along a nuclear path, i.e. at each τ , the electronic eigenvalue problem is solved for the specific configuration, $\mathbf{R}(\tau)$, at this imaginary time point. This means that the influence functional, $F[\mathbf{R}(\tau)]$, and hence the free energy, must be local in τ . This leads to a convenient expression for the path integral:

$$Z(N, V, T) = \oint \mathcal{D}\mathbf{R}(\tau) \exp \left\{ -\frac{1}{\hbar} \int_0^{\beta\hbar} [T_N(\dot{\mathbf{R}}(\tau)) + V_{NN}(\mathbf{R}(\tau)) + A(\mathbf{R}(\tau))] \right\} \quad (78)$$

known as the free energy Born-Oppenheimer path integral approximation introduced by Cao and Berne.³¹

On the other hand, if we simply consider the nuclear eigenvalue problem in Eq. (8), we would lead to a path integral expression in which we perform a separate path integral on each electronic surface, $\varepsilon_n(\mathbf{R})$ and then sum over the surfaces, i.e.

$$Z(N, V, T) = \sum_n \oint \mathcal{D}\mathbf{R}(\tau) \exp \left\{ -\frac{1}{\hbar} \int_0^{\beta\hbar} [T_N(\dot{\mathbf{R}}(\tau)) + V_{NN}(\mathbf{R}(\tau)) + \varepsilon_n(\mathbf{R}(\tau))] \right\} \quad (79)$$

Differences between these two path integral expressions are discussed in detail by Cao and Berne.³¹ It is important to note that when only the electronic ground state is important, the two expressions are equivalent, since, then

$$F[\mathbf{R}(\tau)] = \exp \left[-\frac{1}{\hbar} \int_0^{\beta\hbar} d\tau \varepsilon_0(\mathbf{R}(\tau)) \right] \quad (80)$$

and the path integral expression reduces to

$$Z(N, V, T) = \oint d\mathbf{R}(\tau) \exp \left\{ -\frac{1}{\hbar} \int_0^{\beta\hbar} [T_N(\dot{\mathbf{R}}(\tau)) + V_{NN}(\mathbf{R}(\tau)) + \varepsilon_0(\mathbf{R}(\tau))] \right\} \quad (81)$$

which is the form on which we shall focus here.

The Born-Oppenheimer path integral form in Eq. (81) can be written as a discrete path integral:

$$\begin{aligned} Z_P(N, V, T) &= \left[\prod_{I=1}^N \left(\frac{M_I P}{2\pi\beta\hbar^2} \right)^{3P/2} \int d\mathbf{R}_I^{(1)} \cdots d\mathbf{R}_I^{(P)} \right] \\ &\times \exp \left\{ -\beta \left[\sum_{i=1}^P \left(\sum_{I=1}^N \frac{1}{2} M_I \omega_P^2 (\mathbf{R}_I^{(i+1)} - \mathbf{R}_I^{(i)})^2 \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{1}{P} \varepsilon_0(\mathbf{R}_1^{(i)}, \dots, \mathbf{R}_N^{(i)}) \right) \right] \right\} \end{aligned} \quad (82)$$

where the true partition function $Z(N, V, T) = \lim_{P \rightarrow \infty} Z_P(N, V, T)$ by virtue of the Trotter Theorem. We have already seen that discrete path integrals of this type can be evaluated by molecular dynamics (MD) using the staging or normal mode transformations and a Nosé-Hoover chain thermostat on each degree of freedom. However, there is a key point: the ground state electronic eigenvalue, $\varepsilon_0(\mathbf{R})$ must be evaluated at each imaginary time slice, $i = 1, \dots, P$, i.e. at P different nuclear configurations. As dictated by the Born-Oppenheimer approximation, this requires P separate electronic structure calculations, thus, P sets of Kohn-Sham orbitals! This is somewhat unfortunate, as electronic structure calculations are already expensive enough for a single nuclear configuration. Now, each path discrete configuration requires P electronic structure calculations! The one saving grace is that the P electronic structure calculations are entirely independent of each other and can, therefore, be performed in parallel with no communication overhead. Thus, combining the staging or normal mode path integral method with the Car-Parrinello equations of motion for each set of Kohn-Sham orbitals, one arrives at the complete set of *ab initio* path integral equations of motion:

$$\begin{aligned} \mu |\ddot{\psi}_i^{(s)}\rangle &= |\tilde{\varphi}_i^{(s)}\rangle + \sum_{\alpha} \lambda_{\alpha}^{(s)} - \sum_j \Lambda_{ij} |\psi_i^{(s)}\rangle - \mu \dot{\eta}_1^{(s)} |\dot{\psi}_i^{(s)}\rangle \\ M_I'^{(s)} \ddot{u}_I^{(s)} &= -M_I^{(s)} \omega_P^2 u_I^{(s)} - \frac{1}{P} \frac{\partial \tilde{E}^{(s)}}{\partial u_I^{(s)}} - M_I'^{(s)} \dot{\xi}_{1I}^{(s)} \dot{u}_I^{(s)} \\ Q_R \ddot{\xi}_{I,\nu}^{(s)} &= \tilde{G}_{I,\nu}^{(s)} - Q_R \dot{\xi}_{I,\nu}^{(s)} \dot{\xi}_{I,\nu+1}^{(s)} \quad \nu = 1, \dots, M-1 \\ Q_R \ddot{\xi}_{I,M}^{(s)} &= \tilde{G}_{I,M}^{(s)} \\ Q_{\kappa} \ddot{\eta}_{\kappa}^{(s)} &= \gamma_{\kappa}^{(s)} - Q_{\kappa} \dot{\eta}_{\kappa}^{(s)} \dot{\eta}_{\kappa+1}^{(s)} \quad \kappa = 1, \dots, M-1 \\ Q_M \ddot{\eta}_M^{(s)} &= \gamma_M^{(s)} - Q_M \dot{\eta}_{M-1}^{(s)} \dot{\eta}_M^{(s)} \end{aligned} \quad (83)$$

where s indexes the imaginary time slices, and $Q_R = 1/(\beta\omega_P^2)$. The ionic thermostat forces are given by

$$\begin{aligned}\tilde{G}_{I,1}^{(s)} &= \tilde{M}_I (\dot{u}_I^{(s)})^2 - kT \\ \tilde{G}_{I,\nu}^{(s)} &= Q_R (\dot{\xi}_{I,M-2}^{(s)})^2 - kT\end{aligned}\quad (84)$$

The electronic thermostat forces are given by

$$\begin{aligned}\gamma_1^{(s)} &= \mu \sum_i \langle \dot{\psi}_i^{(s)} | \dot{\psi}_i^{(s)} \rangle - E_e \\ \gamma_{M-1}^{(s)} &= \left[Q_{M-2} \left(\dot{\eta}_{M-2}^{(s)} \right)^2 - \frac{1}{\beta_e} \right] + \left[Q_M \left(\dot{\eta}_M^{(s)} \right)^2 - \frac{1}{\beta_e} \right] \\ \gamma_\kappa^{(s)} &= Q_{\kappa-1} \left(\dot{\eta}_{\kappa-1}^{(s)} \right)^2 - \frac{1}{\beta_e} \quad \kappa = 2, \dots, M-2, M\end{aligned}\quad (85)$$

E_e is the desired electronic kinetic energy, and $1/\beta_e = E_e/N_e$, where N_e is the number of fictitious dynamical degrees of freedom. The thermostat masses are given by $Q_1 = 2E_e/\omega_e^2$, $Q_\kappa = 1/(\beta_e\omega_e^2)$, $\kappa = 2, \dots, M$. ω_e is a characteristic frequency for the fictitious electron dynamics and is usually taken to be several times larger than the highest frequency of the nuclear motion. In practice, one might choose N_e to be somewhat smaller than the actual number of electronic degrees of freedom in order to avoid a large thermostat mass disparity.³² Eqs(83) are sufficiently general to allow inclusion of ultrasoft pseudopotentials in a straightforward manner, however, they are not limited to this choice.

5 Illustrative Applications

5.1 Structure of Liquid Ammonia at 273 K

Ammonia is an important weakly hydrogen-bonded liquid that is employed as a solvent in many common organic reactions and in solutions with metals. Its structure has recently been determined experimentally by neutron diffraction³³ so that experimental partial structure factors and radial distribution functions are now available. As a study of chemical processes in ammonia solvent is a prime application for AIMD techniques, it is important to validate the approach by studying the properties of the neat liquid and making a detailed comparison with experiment.

To this end, AIMD simulations based on the CP equations of motion have been carried out on a sample of 32 ammonia molecules in a box of length 11.27 Å with periodic boundary conditions.^{12,13} Exchange and correlation were treated using the B-LYP GGA functional^{34,35} and the KS orbitals were expanded in a plane wave basis up to a cutoff of 70 Ry. Core electrons were treated using the pseudopotentials of Troullier and Martins.²⁴ The system was allowed to equilibrate for 2.2 ps and then a production run of 6.0 ps was carried out using a time step of 5 a.u.

Figure 1 shows the computed neutron scattering partial structure factors, $H_{NN}(q)$, $H_{HH}(q)$, and $H_{NH}(q)$ together with the experimental results. As can bee seen, very good

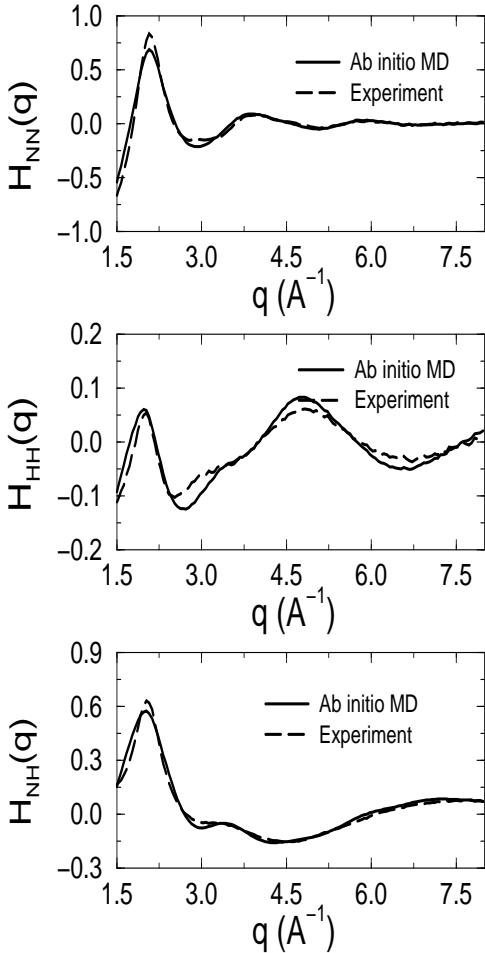


Figure 1. Computed and experimental neutron scattering partial structure factors for liquid ammonia at 273 K. Agreement with experiment is obtained. In addition, we show the computed and experimentally determined radial distribution functions in Fig. 2. In addition, the self diffusion constant was determined from the calculation to be $1.1 \times 10^{-4} \text{ cm}^2/\text{s}$, which compares favorably with the experimental value of $1.0 \times 10^{-4} \text{ cm}^2/\text{s}$.

5.2 Structure of Liquid Methanol at 300 K

Another important hydrogen-bonded liquid is methanol (CH_3OH). Like ammonia, methanol is also used as a solvent in many common organic reactions. It is also an industrially important liquid because of its role in emerging fuel-cell technologies. The structure of liquid methanol has also been determined recently by neutron diffraction,^{36,37} again, making partial structure factors and radial distribution functions readily available for com-

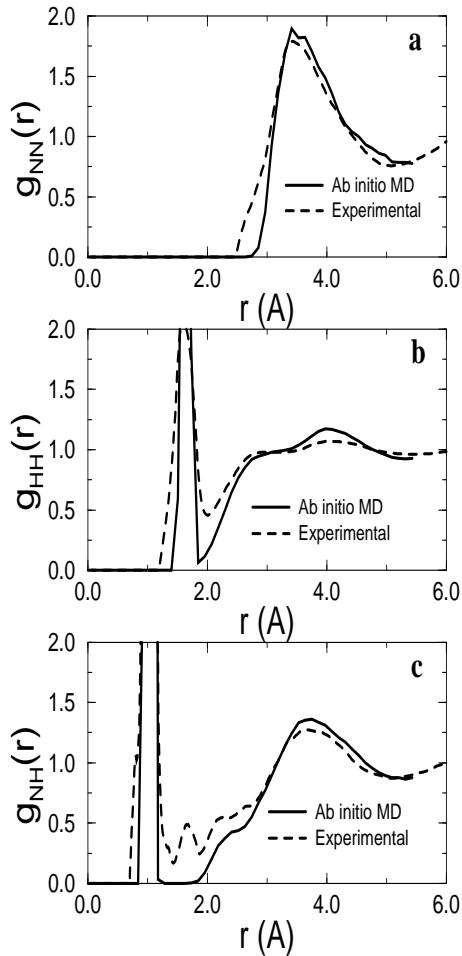


Figure 2. Computed and experimental radial distribution functions for liquid ammonia at 273 K

parison with AIMD calculations. We have recently carried out studies of proton transport in liquid methanol and methanol/water mixtures,¹⁷ and as part of these studies, we carried out a simulation of the pure liquid in order to compare the structural properties with those determined experimentally.

The AIMD simulation protocol employed 32 methanol molecules in a periodic box of 12.93 Å. Exchange and correlation were, again, treated using the B-LYP GGA functional, and, because of the large system size (the system consists of 192 atoms and 224 electronic states), core electrons were treated using the ultrasoft pseudopotential approach of Vanderbilt.²⁵ This allowed the plane wave basis set expansion to be truncated at a cutoff of 25 Ry. The system was equilibrated for 4 ps, and a production run of 20 ps was then carried out using a time step of 5 a.u.

Figure 3 shows the computed and experimentally determined partial structure factors. The heavy-atom partial structure factor is a linear combination of individual partial struc-

ture factors:

$$H_{XX}(q) = 0.042H_{CC}(q) + 0.073H_{CO}(q) + 0.253H_{CM}(q) + 0.032H_{OO}(q) \quad (86)$$

where M denotes a methyl hydrogen. Again, it can be seen that good agreement is

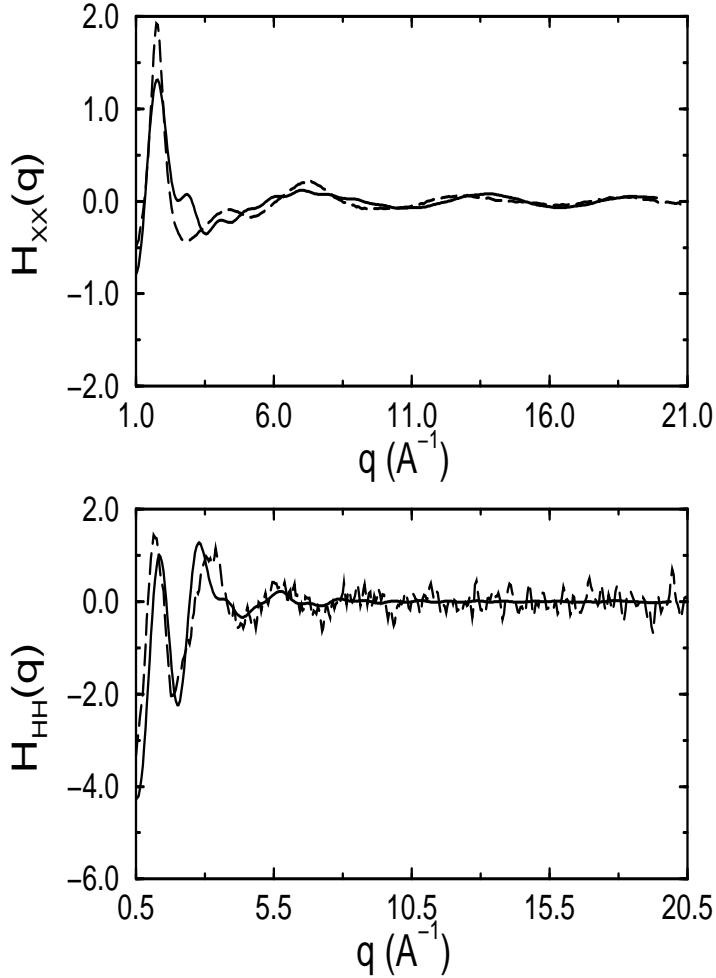


Figure 3. Computed (solid line) and experimental (dashed line) neutron scattering partial structure factors for liquid methanol at 300 K

obtained, although, for this system, the experimental data is somewhat noisier. Figure 4 shows the 10 different radial distribution functions plotted against the experimentally determined radial distribution functions. In all cases, good agreement is obtained. This

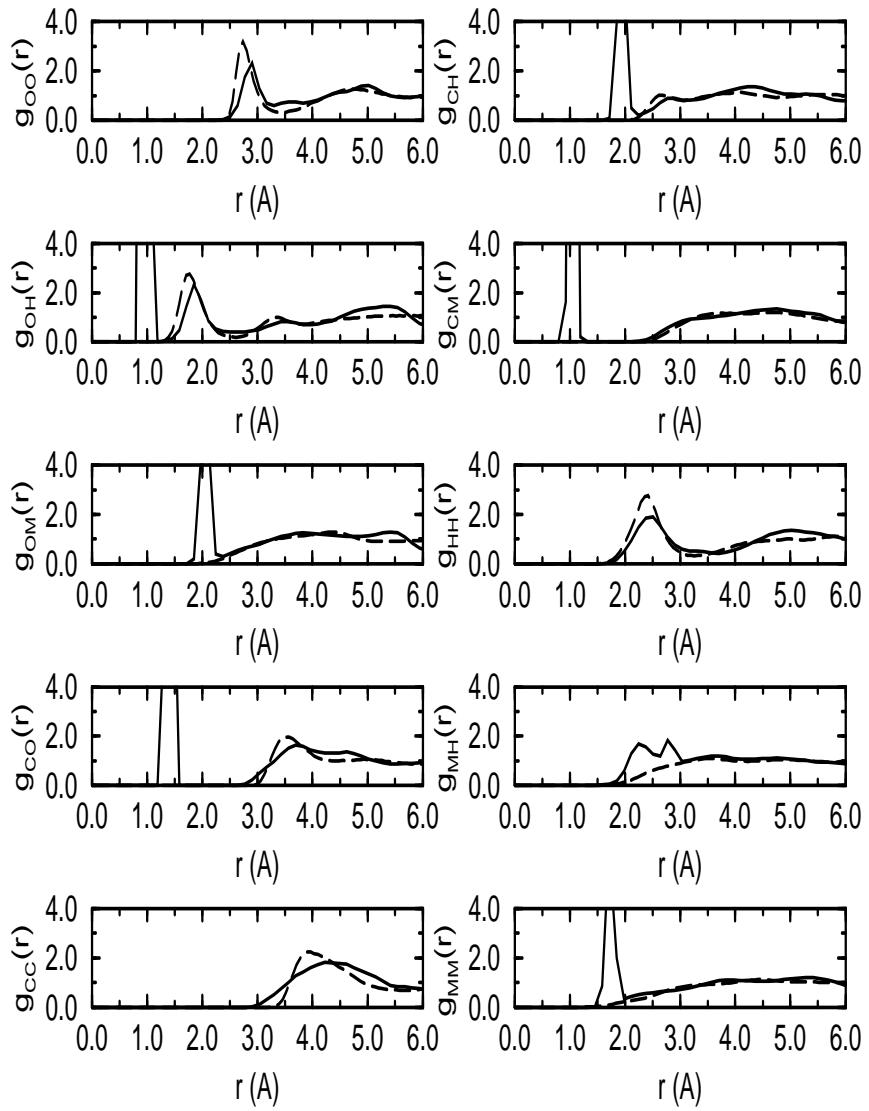


Figure 4. Computed (solid line) and experimental (dashed line) radial distribution functions for liquid methanol at 300 K

and the previous study, as well as numerous studies of liquid water^{2–7} show that the AIMD method is capable of treating a variety of hydrogen-bonded systems accurately.

5.3 Proton Transfer in Malonaldehyde

One of the main advantages of the AIMD technique is that it allows the study of chemical bond-breaking and forming events, for which reliable empirical potential models generally do not exist. As an illustration of the AIMD and AIPI methods, we investigate the role of

nuclear quantum effects on a very common chemical reaction, the proton transfer reaction. Here, a simple example, proton transfer through the internal hydrogen bond in the malonaldehyde molecule is explored.³⁸ The process is illustrated in Fig. 5 below. As the figure

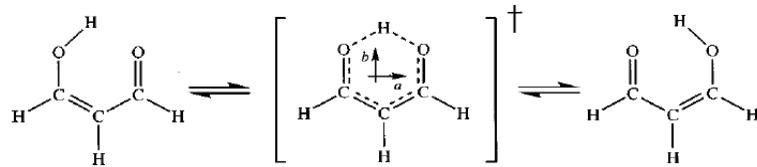


Figure 5. Illustration of the proton transfer process in malonaldehyde

makes clear, the transfer of the proton between the two oxygens gives rise to a change in the chemical bonding pattern around the ring. At zero temperature, the barrier to proton transfer, as computed by high level *ab initio* calculations ranges from 3.1 kcal/mol to 4.6 kcal/mol.

Our AIMD and AIPI simulations are based on the use of the Becke GGA exchange³⁴ and Perdew-Zunger⁴⁰ correlation functionals. The malonaldehyde molecule was placed in an 8.0 Å periodic box, and a plane-wave cutoff of 70 Ry was employed. Core electrons were treated using the Troullier-Martins pseudopotentials.²⁴ This DFT scheme gives a zero-temperature proton transfer barrier of 3.5 kcal/mol, in good agreement with the aforementioned estimates. In addition, for path integral calculations, a discretization of $P = 16$ imaginary time slices was employed. In this study, three different types of calculations were performed: 1) all nuclei treated as classical point particles; 2) all nuclei treated as quantum particles; 3) quantization of *only* the transferring proton (called *classical skeleton* calculations). In each case, thermodynamic integration in conjunction with the bluemoon ensemble approach³⁹ is used to obtain the proton transfer free energy profile at 300 K.

Figure 6 shows the free energy profiles thus obtained for each of the three simulation types. The reaction coordinate on the x -axis is $\nu = d_{O_1H} - d_{O_2H}$, the difference between the distances of each oxygen to the shared proton. The most striking features of these profiles are i) that inclusion of only thermal fluctuations, via the classical nuclei simulation, gives rise to very little difference between the free energy and zero-temperature barriers. When the transferring proton only is quantized, the barrier is considerably lower (approximately 2.1 kcal/mol). Finally, when all nuclei are properly quantized, the barrier is further lowered to approximately 1.6 kcal/mol. This implies that there is a nontrivial quantum effect due to the heavy-atom skeleton. Failure to include this quantum effect leads to an overestimation of the free energy barrier of 31%. (and, hence, underestimation of the proton transfer rate by a factor of roughly 2 in a transition state theory picture⁴¹). We note that our classical skeleton approximation is a comparatively mild one in comparison to approximations that are commonly made in the modeling of such processes. The latter usually completely disregard the structure of the heavy-atom skeleton or attempt to reduce the dimensionality of the problem to a few relevant degrees of freedom. Therefore, the classical skeleton approximation most likely leads to a lower bound estimate of the amount by which more severe approximations would tend to overestimate the free energy

barrier and underestimate the rate.

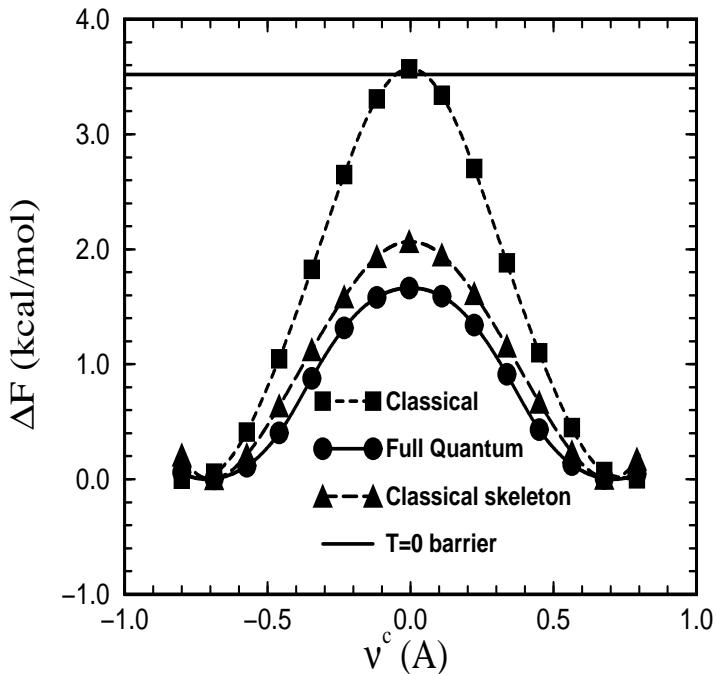


Figure 6. Proton transfer free energy profiles in malonaldehyde

5.4 Proton Transport in Water

Aqueous proton transport is a fundamentally important process in the chemistry of acids and bases and in many biologically important systems. In water, protonic defects (hydronium, H_3O^+ , and hydroxide, OH^- , ions) have an anomalously high mobility that cannot be explained by an ordinary hydrodynamic diffusion picture. In fact, the commonly accepted mobility mechanism is the so called “structural diffusion” or “Grotthuss” mechanism, in which solvation structures diffuse through the hydrogen-bond network via sequential proton transfer reactions. However, the microscopic details of the Grotthuss mechanism for different situations remain largely unelucidated. Here, we describe AIMD and AIFI simulations⁴⁻⁷ that have lead to a clear picture of the structural diffusion mechanism of the hydronium ion in water, a picture that has since been shown to be consistent with all available experimental data.⁴² In addition, one of the key controversial issues, concerning the dominant solvation structures, is resolved. Briefly, one school of thought, put forth by Eigen, considers the dominant solvation structure to be that of a H_3O^+ core surrounded

by three water molecules consisting of a H_9O_4^+ cation. The other school of thought, due to Zundel, favors a picture in which the dominant structure consists of a protonated water dimer or H_5O_2^+ cation, in which the proton is equally shared between two water molecules.

The simulation protocol consists of 31 water molecules and one hydronium ion in a 10 Å Aperiodic box. Exchange and correlation are, again, treated using the B-LYP functional, and a plane wave basis set truncated at a cutoff of 70 Ry was employed. Core electrons were treated using the Troullier-Martins pseudopotentials. For path integral simulations, a discretization of 8 imaginary time slices was employed. AIMD and AIPI trajectories of length 20 ps using a time step of 7 a.u. were generated.

Figure 7 shows schematically the structural diffusion mechanism that is uncovered in these simulations. As the figure shows, the process involves the breaking of a hydrogen

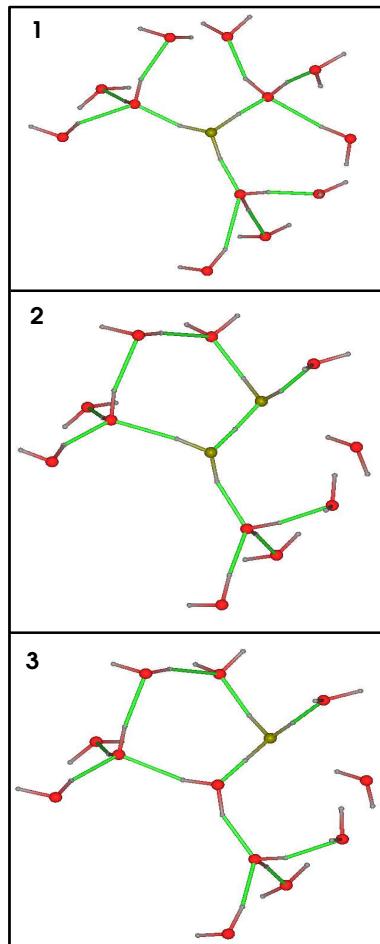


Figure 7. Schematic of the hydronium transport mechanism in water. The figure shows the hydronium and its first two solvation shells.

bond between the first and second solvation shell members of H_3O^+ , i.e., a second solvation shell fluctuation.^{4,5} Following this hydrogen-bond breaking event, a first solvation shell water is left in a state in which its coordination is 3 instead of the usual average value of 4 for water. In this state, it is coordinated more like hydronium than water, and it is, therefore, “prepared” to become a properly solvated hydronium via proton transfer. When its coordination number changes, the oxygen-oxygen distance between the hydronium and the undercoordinated water shrinks by approximately 0.1 Å, and the proton moves to the middle of the bond, forming an intermediate H_5O_2^+ cation state. The proton can then either return to the original hydronium or continue to cross the hydrogen bond to the new oxygen site. If the latter occurs, then there is a new H_9O_4^+ cation formed with a new hydronium core. Thus, the solvation structure has migrated through the hydrogen bond network via the proton transfer step. The rate-limiting process is the hydrogen bond-breaking event, which requires approximately 1.5 ps to occur. This number is in good agreement with the experimentally determined rate of structural diffusion from NMR measurements.⁴³ In addition, the activation enthalpy, approximately 3 kcal/mol, can be explained by this mechanism, which requires approximately 2.5 kcal/mol to break the hydrogen bond and another 0.5 kcal/mol to shrink the oxygen-oxygen distance after the hydrogen bond is broken.⁴⁴

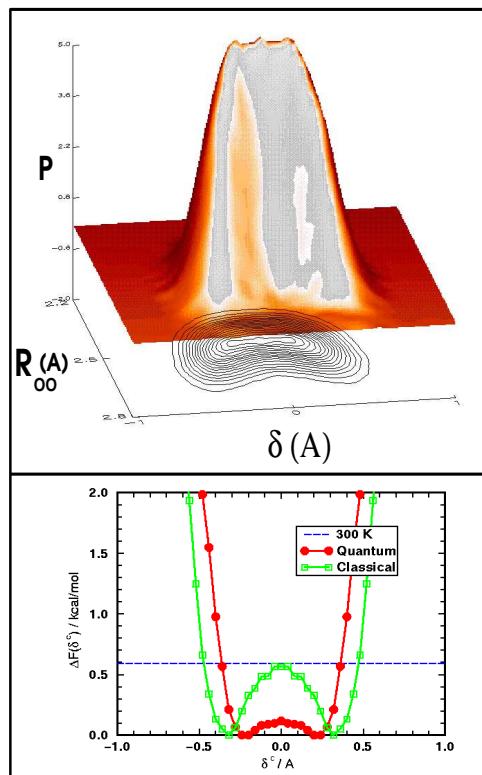


Figure 8. Quantum probability distribution of the proton transfer reaction coordinate, δ and the oxygen-oxygen distance, R_{OO} .

Inclusion of nuclear quantum effects via the path integral can resolve the controversy of the solvation structures.^{6,7} In particular, if we plot the probability distribution of the oxygen-oxygen distance, R_{OO} and the proton transfer coordinate, δ (similar to the coordinate, ν , described above) for the hydrogen bond in which proton transfer is “most likely to occur” (defined as the hydrogen bond with the smallest value of δ), which is shown in Fig. 8, we see that the probability that the solvation complex is $H_9O_4^+$ or $H_5O_2^+$ or any complex in between these two ideal, limiting structures is approximately the same. This is also confirmed by studying the free energy profile along the coordinate, δ also shown in Fig. 8. The fact that there is a broad flat minimum in this free energy confirms the notion that there is no single dominant solvation structure. Rather, the defect is best described as a “fluxional” defect, that can take on, with same probability, the characteristics of the $H_9O_4^+$ or $H_5O_2^+$ cations and all structures in between these. Interestingly, a purely classical treatment of the predicts that the $H_9O_4^+$ is considerably more stable than the $H_5O_2^+$ cation by approximately 0.6 kcal/mol. This proton transfer barrier is completely washed out by nuclear zero-point motion, leading to the fluxional defect picture proposed in Ref.⁷

References

1. R. Car and M. Parrinello, *Phys. Rev.Lett.* **55**, 2471 (1985).
2. P. L. Silvestrelli and M. Parrinello, *J. Chem. Phys.* **111**, 3572 (1999).
3. P. L. Silvestrelli, M. Bernasconi and M. Parrinello, *Chem. Phys. Lett.* **277**, 478 (1997).
4. M. E. Tuckerman, K. Laasonen, M. Sprik and M. Parrinello, *J. Chem. Phys.* **103**, 150 (1995).
5. M. E. Tuckerman, K. Laasonen, M. Sprik and M. Parrinello, *J. Phys. Chem.* **99**, 5749 (1995).
6. D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, *Nature* **397**, 601 (1999).
7. D. Marx, M. E. Tuckerman and M. Parrinello, *J. Phys. Condens. Matt.* **12**, A153 (2000).
8. M. Bernasconi, P. L. Silvestrelli and M. Parrinello, *Phys. Rev. Lett.* **81**, 1235 (1998).
9. M. Benoit, D. Marx and M. Parrinello, *Nature* **392**, 258 (1998).
10. J. Sarnthein, A. Pasquarello and R. Car, *Science* **275**, 1925 (1997).
11. M. Benoit, S. Ispas and M. E. Tuckerman, *Phys. Rev. B* **64**, 224205 (2991).
12. M. Diraison, G. J. Martyna and M. E. Tuckerman, *J. Chem. Phys.* **111**, 1096 (1999).
13. Y. Liu and M. E. Tuckerman, *J. Phys. Chem. B* **105**, 6598 (2001).
14. S. Piana, D. Sebastiani, P. Carloni and M. Parrinello, *J. Am. Chem. Soc.* **123**, 8730 (2001).
15. J. Hutter, P. Carloni and M. Parrinello, *J. Am. Chem. Soc.* **118**, 871 (1996).
16. M. Boero and M. Parrinello, *J. Am. Chem. Soc.* **122**, 501 (2000).
17. J. A. Morrone and M. E. Tuckerman (to be submitted).
18. Z. H. Liu, L. E. Carter and E. A. Carter, *J. Phys. Chem.* **99**, 4355 (1995).
19. B. Di Martino, M. Celino and V. Rosato, *Comp. Phys. Comm.* **120**, 255 (1999).
20. R. A. Friesner, *Chem. Phys. Lett.* **116**, 39 (1985).
21. G. Lippert, J. Hutter and M. Parrinello, *Mol. Phys.* **92**, 477 (1997).
22. Y. Liu and M. E. Tuckerman (to be submitted).
23. G. B. Bachelet, D. R. Hamann, and M. Schlüter, *Phys. Rev. B* **26**, 4199 (1982).

24. N. Troullier and J. Martins, *Phys. Rev. B* **43**, 6796 (1991).
25. D. Vanderbilt, *Phys. Rev. B* **41**, 7892 (1990).
26. P. E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).
27. L. Kleinman and D. M. Bylander, *Phys. Rev. Lett.* **48**, 1425 (1982).
28. G. J. Martyna and M. E. Tuckerman, *J. Chem. Phys.* **110**, 2810 (1999).
29. P. Minary, M. E. Tuckerman, K. A. Pihakari and G. J. Martyna, *J. Chem. Phys.* (in press).
30. R. P. Feynman and F. L. Vernon, *Ann. Phys.* **24**, 118 (1963).
31. J. Cao and B. J. Berne, *J. Chem. Phys.* **99**, 2902 (1993).
32. M. E. Tuckerman and M. Parrinello, *J. Chem. Phys.* **101**, 1302 (1994).
33. M. A. Ricci, N. Nardone, F. P. Ricci, C. Andreani and A. K. Soper, *J. Chem. Phys.* **102**, 7650 (1995).
34. A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
35. C. Lee, W. Yang, and R. C. Parr, *Phys. Rev. B* **37**, 785 (1988).
36. T. Yamaguchi, K. Hidaka and A. K. Soper, *Mol. Phys.* **96**, 1159 (1999).
37. S. Dixit, W. C. K. Poon, and J. Crain, *J. Phys. Condens. Matt.* **12**, L323 (2000).
38. M. E. Tuckerman and D. Marx, *Phys. Rev. Lett.* **86**, 4946 (2001).
39. M. Sprik and G. Ciccotti, *J. Chem. Phys.* **109**, 7737 (1998).
40. J. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).
41. G. A. Voth, *J. Phys. Chem.* **97**, 8365 (1993).
42. N. Agmon, *Chem. Phys. Lett.* **244**, 456 (1995).
43. S. Meiboom, *J. Chem. Phys.* **34**, 375 (1961).
44. N. Agmon, *J. Molec. Liq.* **73,74**, 513 (1997).

Dynamic Properties via Fixed Centroid Path Integrals

Rafael Ramírez and Telesforo López-Ciudad

Instituto de Ciencia de Materiales
Consejo Superior de Investigaciones Científicas (C.S.I.C.)
Cantoblanco, 28049 Madrid, Spain
E-mail: {ramirez, tito.lopez}@icmm.csic.es

Computer simulations of complex many-body systems by Centroid Molecular Dynamics (CMD) are a practical route for the calculation of time dependent properties, i.e., time correlation functions, of quantum systems in thermodynamic equilibrium. The CMD approach can be readily implemented into standard codes for Monte Carlo or molecular dynamics path integral simulations of static properties. However, the advantageous property of being readily implementable is not a prerequisite for the even more important property of predicting correct results. In this context, the rigorous formulation of CMD is an important goal to understand its capability to describe real time quantum dynamics. Our purpose in this contribution is fourfold: (i) to present a derivation of CMD and a related dynamic approach within the Schrödinger formulation, as an alternative to the usual route based on the path integral formulation; (ii) to analyze the capability of these approaches in the study of simple exact solvable models; (iii) to review the CMD applications done on a variety of systems and properties; (iv) to summarize the most important open problems for CMD applications.

1 Introduction

The formulation of statistical mechanics using path integrals (PI) provides a practical computational route for the simulation of *static properties* of quantum many-body systems in thermodynamic equilibrium.^{1–5} One of the most suggestive pictures derived from the PI approach is the so-called *quantum-classical isomorphism*. This isomorphism states that the statistical behavior of a set of quantum particles can be mapped onto a classical model of interacting “ring polymers”. In more precise terms, the canonical partition function, Z , of N quantum particles results to be identical to the *classical* partition function of N interacting “ring polymers”,

$$Z \equiv Z_{\text{ring-polymers}}^{\text{cla}} . \quad (1)$$

The isomorphism opens the possibility to extend classical simulation methods such as Monte Carlo (MC) and molecular dynamics (MD) to the quantum domain. Thus, static properties of quantum systems, i.e., those that can be derived from the partition function Z , can be readily computed by classical simulations of ring polymers.

A second version of the quantum-classical isomorphism was worked out by Feynman and Hibbs in 1965.¹ These authors realized that the classical partition function of the N rings polymers can be alternatively rewritten in terms of the N *centroid positions* or center of masses of the ring polymers. As a result, Feynman and Hibbs defined an *effective classical potential*, V_{ecp} , that allowed them to formulate a new isomorphism: The partition function of the N quantum particles, Z , turns out to be identical to the *classical* partition function of N particles moving in the effective classical potential, V_{ecp} ,

$$Z \equiv Z_{\text{ecp}}^{\text{cla}} . \quad (2)$$

The computation of V_{ecp} is a difficult task that implies solving *fixed centroid path integrals*, that are defined as standard path integrals with a geometrical constraint that fixes the centroid position. There exists an important body of literature on methods based on the effective classical potential and fixed centroid path integrals. Many interesting works focus on the formulation of variational approximations for V_{ecp} , that can be applied to compute static properties.^{1,6-9} Other relevant application is the formulation of the PI quantum transition state theory (QTST), a kinetic approach to compute rate constants of thermally activated quantum processes.^{4,10-15}

Based on the success of the quantum-classical isomorphism for the simulation of static properties, one may wonder if there exists some kind of *dynamic* quantum-classical isomorphism, that would allow us to derive *exact* quantum dynamics from *classical dynamics* of either the “ring polymers” or their centroids. Unfortunately, the answer is not. The reason is that dynamic properties, like time correlation functions, can not be derived from the knowledge of the partition function. Thus, an identity in the partition functions of the quantum system and its classical isomorph, does not imply a coincidence in the corresponding dynamic properties. However, the suggestive picture provided by the quantum-classical isomorphism has motivated the sought of an *approximate* description of real time quantum dynamics in terms of classical dynamics of the corresponding classical isomorph.

The most practical approximation developed so far is centroid molecular dynamics (CMD), formulated by Cao and Voth in 1994.¹⁶⁻²⁰ CMD solves the dynamic equations of classical particles moving in the effective classical potential, V_{ecp} . For the simple case of a particle of mass m , whose centroid position and momentum are (X, P) , the CMD equations read

$$\dot{X} = \frac{P}{m}, \quad (3)$$

$$\dot{P} = f_m. \quad (4)$$

where the dot indicates a time derivative, and f_m is the *mean force*, defined as the position derivative of the effective classical potential

$$f_m = -\frac{dV_{ecp}}{dX}. \quad (5)$$

The capability of CMD to describe real time quantum dynamics lies on the *assumption* that the effective classical potential, V_{ecp} , provides some kind of average of dynamic properties of the quantum system, so that classical dynamics using V_{ecp} reproduces quantitatively some dynamic results. More precisely, Cao and Voth suggested that classical time correlation functions of position or momentum coordinates, calculated from trajectories generated by the CMD equations, are a well-defined approximation to the Kubo transform of the corresponding quantum time correlation functions.

One appealing characteristic of the CMD equations is their simplicity. However, the derivation of CMD is cumbersome and not easy to understand. From the original CMD derivation,¹⁸ it was not clear why the centroid coordinate should play such a prominent role in the determination of time correlation functions. A more rigorous formulation of CMD was presented by Jang and Voth in 1999.²¹⁻²³ The essential step for this improved derivation was the definition of an operator, called the quasi-density operator, whose representation in a position basis corresponds to a fixed centroid path integral. The main result

derived by Jang and Voth was to demonstrate that Kubo transformed time correlation functions of position or momentum operators can be *exactly* represented in terms of averages of operators calculated using the quasi-density operator. The operator formalism presented by Jang and Voth²¹ is an essential prerequisite for a sound theoretical formulation of the CMD approximation, and represents a substantial improvement with respect to the original CMD formulation.¹⁸

The specialized PI concepts that are needed for the definition of CMD, like centroid coordinates, fixed centroid path integrals, and the effective classical potential, cannot be easily translated into physical quantities defined outside the PI formulation. This circumstance is unfortunate, because one who is not familiar with these specialized concepts will have serious difficulties in understanding the physical meaning of CMD and other applications involving centroid coordinates. Paradoxically, the increasing number of these applications over the last 40 years strongly suggest that these concepts have a precise physical significance. This significance should be made clearer if the formulation of CMD were presented without reference to the PI formulation.

One of the main goals of the present chapter is to present a derivation of CMD, and a related improved dynamic approach, by working within the Schrödinger formulation. The Schrödinger formulation of CMD is complementary to the PI formulation and offers some distinct advantages besides. In particular, the Schrödinger formulation shows how CMD is related to the time evolution of canonical density operators and also clarifies the relation of CMD to *linear response theory* and the *fluctuation-dissipation theorem*. Both topics are of considerable importance for a full understanding of the CMD approximations. A short account of the Schrödinger formulation of CMD has been presented elsewhere.²⁴

This chapter is organized as follows. In Section 2 we review the basic relations that characterize a quantum system in thermodynamic equilibrium. Some auxiliary quantities, that will be used in the Schrödinger formulation of fixed centroid path integrals, are also introduced. The standard PI formulation of fixed centroid path integrals is presented in Section 3, while their Schrödinger formulation is presented in Section 4. The derivation of CMD and a related dynamic approximation is the main goal of Sec. 5. Numerical tests of the capability of these dynamic approximations to treat some exact solvable models are given in Sec. 6. A review of already published CMD simulations is presented in Sec. 7. Some open problems for CMD applications are summarized in Sec. 8. The chapter ends with the conclusions in Sec. 9.

2 Definition of Auxiliary Quantities

For the sake of clarity, we consider a canonical ensemble of independent particles of mass m moving in one dimension. The extension of most of the present study to a many-body problem is straightforward. The Hamiltonian of the particle is assumed to be

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x}) , \quad (6)$$

where \hat{x} , \hat{p} , and $V(\hat{x})$ are the position, momentum, and potential energy operators, respectively. The unnormalized canonical density operator of the ensemble of particles is

$$\hat{\rho} = e^{-\beta \hat{H}} , \quad (7)$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature and k_B is the Boltzmann constant. The partition function is the trace of the density operator,

$$Z = \text{Tr}[\hat{\rho}] . \quad (8)$$

It is convenient to define a normalized density operator, whose trace is unity, by dividing the unnormalized one by its trace. Normalized density operators are represented by a tilde $\tilde{\rho}$,

$$\tilde{\rho} = Z^{-1} \hat{\rho} . \quad (9)$$

The canonical average of an arbitrary quantum mechanical operator, \hat{A} , is the following trace

$$\langle \hat{A} \rangle \equiv \langle \hat{A} \rangle_\rho = \text{Tr}[\hat{A} \hat{\rho}] . \quad (10)$$

Some auxiliary quantities will be needed for the Schrödinger formulation of fixed centroid path integrals. The first one is the *auxiliary Hamiltonian*, $\hat{H}_a(f, v)$, defined by adding linear terms in \hat{x} and \hat{p} to the Hamiltonian \hat{H}

$$\hat{H}_a(f, v) = \hat{H} - f\hat{x} - v\hat{p} . \quad (11)$$

The parameter f represents a constant external force acting upon the quantum particle and the parameter v has dimensions of velocity. Note the following equivalence: $\hat{H}_a(0, 0) \equiv \hat{H}$. The *auxiliary canonical density operator*, $\hat{\rho}_a(f, v)$, and the *auxiliary partition function*, $Z_a(f, v)$ are defined using the auxiliary Hamiltonian, $\hat{H}_a(f, v)$ with the help of Eqs. (7) and (8). The average of an operator \hat{A} , calculated with the auxiliary density operator, $\hat{\rho}_a(f, v)$, is

$$\langle \hat{A} \rangle_{\rho_a(f, v)} = \text{Tr}[\hat{A} \hat{\rho}_a(f, v)] , \quad (12)$$

where the normalized auxiliary canonical density operator is

$$\hat{\rho}_a(f, v) = [Z_a(f, v)]^{-1} \hat{\rho}_a(f, v) . \quad (13)$$

3 Definition of Fixed Centroid Path Integrals

In this Section, we present the PI formulation of the canonical density matrix and the definition of fixed centroid path integrals. An important relation between fixed centroid path integrals and the auxiliary canonical density matrix is derived. This relation will provide the starting point for the Schrödinger formulation of fixed centroid path integrals.

3.1 PI Formulation of the Canonical Density Matrix

The position representation of the unnormalized canonical density operator, $\hat{\rho}$, is the matrix

$$\rho(x, x') \equiv \langle x | \hat{\rho} | x' \rangle , \quad (14)$$

where $|x\rangle$ is an eigenfunction of the operator \hat{x} . The matrix elements $\rho(x, x')$ are *propagators* representing the probability amplitude associated to the movement of the particle

from position x' to x in an Euclidean time $\beta\hbar$. The phase space PI formulation of these matrix elements is²⁵

$$\rho(x, x') = \int_{x'}^x D[x(u), p(u)] e^{-\frac{S[x(u), p(u)]}{\hbar}}. \quad (15)$$

$[x(u), p(u)]$ are the position and momentum defining a particle path in phase space. The Euclidean time action is defined by the following functional of the path

$$S[x(u), p(u)] = \int_0^{\beta\hbar} du \left\{ \frac{p(u)^2}{2m} + V[x(u)] - ip(u)\dot{x}(u) \right\}, \quad (16)$$

where $\dot{x}(u)$ is the derivative of $x(u)$ with respect to u . The integral measure is given by

$$D[x(u), p(u)] = \lim_{N \rightarrow \infty} \prod_{k=1}^{N-1} dx_k \prod_{k=0}^{N-1} dp_k \left(\frac{1}{2\pi\hbar} \right)^N, \quad (17)$$

where the path $[x(u), p(u)]$ has been discretized in phase space as

$$(x', p_0), (x_1, p_1), (x_2, p_2), \dots, (x_{N-1}, p_{N-1}), (x, p_0). \quad (18)$$

The Euclidean time action in Eq. (16) displays a quadratic dependence on the momentum coordinates $p(u)$. Therefore, all Gaussian integrals in $p(u)$ that appear in Eq. (15) can be evaluated analytically. The resulting path integral is^{1,25}

$$\rho(x, x') = \int_{x'}^x D_x[x(u)] e^{-\frac{S_x[x(u)]}{\hbar}}, \quad (19)$$

where the Euclidean time action is now

$$S_x[x(u)] = \int_0^{\beta\hbar} du \left\{ \frac{m}{2} \dot{x}(u)^2 + V[x(u)] \right\}, \quad (20)$$

and the integral measure is given by

$$D_x[x(u)] = \lim_{N \rightarrow \infty} \prod_{k=1}^{N-1} dx_k \left(\frac{mN}{2\pi\beta\hbar^2} \right)^{\frac{N}{2}}. \quad (21)$$

3.2 PI Formulation of Constrained Propagators

The *centroid position* and *centroid momentum* of a given path $[x(u), p(u)]$ are defined as the average position and momentum of the path

$$x_c = \frac{1}{\beta\hbar} \int_0^{\beta\hbar} du x(u), \quad (22)$$

$$p_c = \frac{1}{\beta\hbar} \int_0^{\beta\hbar} du p(u). \quad (23)$$

For the uncountable set of paths contributing to the path integral in Eq. (15), the property of having the same average point x_c and p_c is an equivalence relation that allows us to classify the paths into equivalence classes.⁹ Each value of (x_c, p_c) labels a different class

of paths. A *fixed centroid path integral* or *constrained propagator* is defined by a path integral like Eq. (15), that includes only the paths of a given equivalence class (X, P)

$$\sigma(x, x'; X, P) = \int_{x'}^x D[x(u), p(u)] \delta(X - x_c) \delta(P - p_c) e^{-\frac{S[x(u), p(u)]}{\hbar}} , \quad (24)$$

where δ is the Dirac delta function. The path integral in Eq. (15) can be recovered by integrating over all equivalence classes

$$\rho(x, x') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP \sigma(x, x'; X, P) . \quad (25)$$

It is important to formulate the relationship between fixed centroid path integrals, $\sigma(x, x'; X, P)$, and the auxiliary canonical density matrix, $\rho_a(x, x'; f, v)$. Considering the definition of the Euclidean time action in Eq. (16), one readily derives that the Euclidean time action corresponding to the auxiliary Hamiltonian $\hat{H}_a(f, v)$, is related to that one of the original Hamiltonian, \hat{H} , by

$$S_a[x(u), p(u); f, v] = S[x(u), p(u)] - \beta \hbar f X - \beta \hbar v P , \quad (26)$$

where (X, P) are the centroid position and momentum of the path $[x(u), p(u)]$. This result implies that the PI formulation of the auxiliary density matrix can be written as

$$\rho_a(x, x'; f, v) = \int_{x'}^x D[x(u), p(u)] e^{-\frac{S[x(u), p(u)]}{\hbar}} e^{\beta f X} e^{\beta v P} . \quad (27)$$

Considering the definition of the fixed centroid path integral in Eq. (24), the previous relation can be alternatively written as

$$\rho_a(x, x'; f, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP \sigma(x, x'; X, P) e^{\beta f X} e^{\beta v P} . \quad (28)$$

This equation shows that the canonical propagator, $\rho_a(x, x'; f, v)$, and the constrained propagator, $\sigma(x, x'; X, P)$, are related by a two-sided Laplace transform.²⁶ The variables $(X, \beta f)$ and $(P, \beta v)$ form pairs of conjugate variables in this integral transform. This relation is the essential link to define fixed centroid path integrals within the Schrödinger formulation.

4 The Schrödinger Formulation of Fixed Centroid Path Integrals

We have chosen to present the results of this Section in a way that is formally analogous to the formulation of the Wigner representation of canonical density operators.²⁷ We recall that with the help of an *integral transform* of the canonical density matrix, $\rho(x, x')$, one defines the Wigner representation, $\rho^W(x, p)$. The coordinates (x, p) of the Wigner matrix defines a phase space, and the statical and dynamic properties of the canonical ensemble can be represented within this phase space.

In this Section, we will define by an integral transform of the auxiliary density operator, $\hat{\rho}_a(f, v)$, a new representation, $\hat{\sigma}(X, P)$. The coordinates (X, P) define a phase space and we will show that the statical and dynamic properties of the canonical ensemble can be represented within this phase space.

Some results of this section have been derived by Jang and Voth using the PI formulation.²¹ However, our derivation is presented without reference to the PI formulation and offers new physical insight. A short account of the present derivation has been presented elsewhere.²⁴

4.1 Definition of the Static Response Phase Space

As Eq. (28) is valid for arbitrary x and x' , one can omit these variables to get a relation between quantum mechanical operators

$$\hat{\rho}_a(f, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP \hat{\sigma}(X, P) e^{\beta f X} e^{\beta v P}. \quad (29)$$

The auxiliary density operator, $\hat{\rho}_a(f, v)$, is related to the operator $\hat{\sigma}(X, P)$ by a two-sided Laplace transform. This integral transform is an operator relation that can be used as starting point to define $\hat{\sigma}(X, P)$, without referring to the PI formulation.

It is convenient to visualize that the coordinates (X, P) define a phase space associated with the canonical ensemble: The *static response (SR) phase space*. Each phase space point, (X, P) , is associated to a quantum mechanical operator, $\hat{\sigma}(X, P)$. The name “SR phase space” has been chosen because the auxiliary density operator $\hat{\rho}_a(f, v)$ obviously depends on the *static response* of the original canonical ensemble to arbitrary linear modifications of its Hamiltonian \hat{H} . The SR phase space has interesting properties, in particular for the study of the *linear response* of the canonical ensemble defined by the Hamiltonian \hat{H} (see below).

By taking the trace of the operators in Eq. (29), and considering the linear property of two-sided Laplace transforms, one readily derives

$$Z_a(f, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP C(X, P) e^{\beta f X} e^{\beta v P}, \quad (30)$$

where $C(X, P)$ is the following trace

$$C(X, P) = \text{Tr}[\hat{\sigma}(X, P)]. \quad (31)$$

We refer to $C(X, P)$ as the *SR phase space density* associated to the point (X, P) . The *SR phase space average* of an arbitrary function, $g(X, P)$, is defined as

$$\{g(X, P)\} = Z^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP C(X, P) g(X, P). \quad (32)$$

Note the use of braces $\{\dots\}$ to represent SR phase space averages. The basic elements of the SR phase space are illustrated in Fig. 1.

For the formulation of operator averages it is convenient to define a normalized operator

$$\hat{\sigma}(X, P) = [C(X, P)]^{-1} \hat{\sigma}(X, P). \quad (33)$$

We call $\hat{\sigma}(X, P)$ the *generalized SR density operator*. The word *generalized* indicates that $\hat{\sigma}(X, P)$ is a generalization of a true density operator, in the sense that it describes mixed states where the “probabilities”, w_n , associated to the eigenfunctions of the operator, may be not only numbers satisfying $0 \leq w_n \leq 1$, but also numbers greater than one and lower than zero (see below). For a true density operator, as $\hat{\rho}(f, v)$, these probabilities are necessarily numbers in the range $0 \leq w_n \leq 1$.

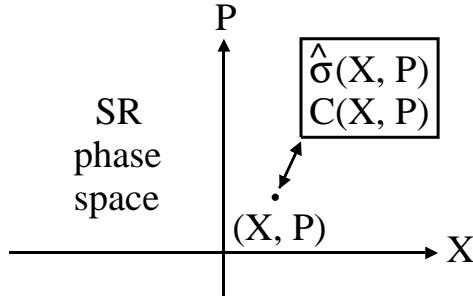


Figure 1. Schematic representation of the SR phase space. Each phase space point (X, P) is associated with a quantum operator, $\hat{\sigma}(X, P)$, whose trace defines the phase space density, $C(X, P)$.

By setting $f = 0$ and $v = 0$ in Eq. (29) and after dividing by Z , one derives the definition of the normalized canonical density operator as a SR phase space average

$$\hat{\rho} \equiv \{ \hat{\sigma}(X, P) \}. \quad (34)$$

To avoid confusion in the nomenclature, let us summarize the relationship between the phase space concepts introduced in this section, and fixed centroid path integral concepts, that are currently used in the literature:

- The SR phase space coordinates (X, P) are the centroid position X and momentum P associated to fixed centroid path integrals.
- The position representation of the unnormalized operator, $\hat{\sigma}(X, P)$, corresponds to the fixed centroid path integral given in Eq. (24).
- The normalized operator $\hat{\rho}(X, P)$ is identical to the quasi-density operator defined by Jang and Voth²¹ as the inverse two-sided Laplace transform of Eq. (29).

The main advantage of our formulation,²⁴ with respect to the similar work by Jang and Voth,²¹ is that we make explicit use of the integral transform relation in Eq. (29) to derive the properties of the operator $\hat{\sigma}(X, P)$. For example, given the dynamic and Bloch equations for the auxiliary canonical density operator, $\hat{\rho}_a(f, v)$, Eq. (29) allows us to derive the corresponding transformed equations for $\hat{\sigma}(X, P)$.

In the next Subsections, the most important properties of the SR phase space are derived. We focus on the four following topics:

- the dynamic equation for $\hat{\sigma}(X, P)$,
- the Bloch equation for $\hat{\sigma}(X, P)$,
- the formulation of canonical averages as SR phase space averages,
- the formulation of time correlation functions as SR phase space averages.

4.2 Dynamic Equation for the Operator $\hat{\sigma}(X, P)$

The dynamic equation of the auxiliary canonical density operator, $\hat{\rho}_a(f, v)$, evolving in time under the action of the Hamiltonian \hat{H} , is

$$i\hbar \frac{\partial \hat{\rho}_a(f, v)}{\partial t} = [\hat{H}, \hat{\rho}_a(f, v)] , \quad (35)$$

where $[\hat{H}, \hat{\rho}_a(f, v)]$ is the commutator of the operators inside the square brackets. This relation is the starting point to derive the dynamic equation for $\hat{\sigma}(X, P)$. The time derivative of the two-sided Laplace transform in Eq. (29) is

$$\frac{\partial \hat{\rho}_a(f, v)}{\partial t} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dXdP \frac{\partial \hat{\sigma}(X, P)}{\partial t} e^{\beta f X} e^{\beta v P} . \quad (36)$$

By applying (left and right) the Hamiltonian operator \hat{H} to Eq. (29), and taking into account the linear property of the operator \hat{H} , one gets

$$\hat{H}\hat{\rho}_a(f, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dXdP \hat{H}\hat{\sigma}(X, P) e^{\beta f X} e^{\beta v P} , \quad (37)$$

$$\hat{\rho}_a(f, v)\hat{H} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dXdP \hat{\sigma}(X, P) \hat{H} e^{\beta f X} e^{\beta v P} . \quad (38)$$

Subtracting the last two Eqs., one derives

$$[\hat{H}, \hat{\rho}_a(f, v)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dXdP [\hat{H}, \hat{\sigma}(X, P)] e^{\beta f X} e^{\beta v P} . \quad (39)$$

The results of Eqs. (35), (36), and (39) imply that the dynamic equation for $\hat{\sigma}(X, P)$ is identical to that one corresponding to a canonical density operator

$$i\hbar \frac{\partial \hat{\sigma}(X, P)}{\partial t} = [\hat{H}, \hat{\sigma}(X, P)] . \quad (40)$$

This dynamic equation can be integrated to give

$$\hat{\sigma}(X, P; t) = e^{-i\frac{\hat{H}t}{\hbar}} \hat{\sigma}(X, P) e^{i\frac{\hat{H}t}{\hbar}} . \quad (41)$$

By taking the trace of the last equation and considering the cyclic invariance of the trace of a product of operators, one readily derives that the trace, $C(X, P)$, of this operator is a conserved quantity

$$\text{Tr}[\hat{\sigma}(X, P; t)] = \text{Tr}[\hat{\sigma}(X, P)] . \quad (42)$$

4.3 Bloch Equation for the Operator $\hat{\sigma}(X, P)$

The Bloch equation, defining the temperature dependence of the operator $\hat{\sigma}(X, P)$, can be derived as the two-sided Laplace transform of the Bloch equation of the auxiliary density operator $\hat{\rho}_a(f, v)$

$$\frac{\partial \hat{\rho}_a(f, v)}{\partial \beta} = -\hat{H}_a(f, v) \hat{\rho}_a(f, v) . \quad (43)$$

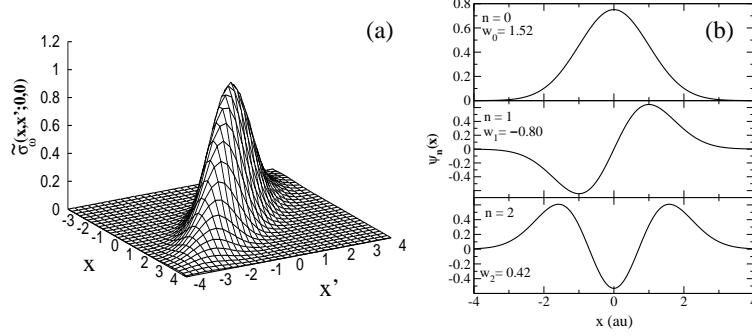


Figure 2. (a) Three-dimensional plot of the SR density matrix $\tilde{\sigma}_\omega(x, x'; 0, 0)$ corresponding to a harmonic oscillator with $m = \omega = 1$ a.u. and $\beta = 2$ a.u.. (b) Eigenfunctions, $\psi_n(x)$, and eigenvalues, w_n , of $\tilde{\sigma}_\omega(x, x'; 0, 0)$ for $n \leq 2$. The eigenfunctions are identical to the energy eigenfunctions of a harmonic oscillator. The eigenvalues form an alternating series of positive and negative real numbers.

This derivation is straightforward and it has been published elsewhere,²⁸ therefore we present the final result

$$\frac{\partial \hat{\sigma}(X, P)}{\partial \beta} = - \left[\hat{H} + \frac{\hat{x} - X}{\beta} \frac{\partial}{\partial X} + \frac{\hat{p} - P}{\beta} \frac{\partial}{\partial P} - \frac{2}{\beta} \right] \hat{\sigma}(X, P) \quad (44)$$

One interesting point to be stressed here is that this Bloch equation can be solved analytically for the particular case of a harmonic oscillator.²⁸ The spectral decomposition of the harmonic SR density operator reads

$$\hat{\tilde{\sigma}}_\omega(X, P) = \sum_{n=0}^{\infty} w_n |\psi_n(X, P)\rangle \langle \psi_n(X, P)|, \quad (45)$$

where w_n are the eigenvalues and $|\psi_n(X, P)\rangle$ are the eigenvectors of the operator.

At temperature $T = 0$ all eigenvalues are zero except $w_0 = 1$.²⁸ This result means that $\hat{\tilde{\sigma}}_\omega(X, P)$ is a pure quantum state at $T = 0$. However, at finite temperature ($T > 0$), the eigenvalues w_n form an alternating series of positive and negative real numbers. Then, the operator, $\hat{\tilde{\sigma}}_\omega(X, P)$, is a generalization of a true density operator, in the sense that the “probabilities” or occupation numbers associated to some eigenvectors are allowed to be negative real numbers or even larger than unity. For a harmonic oscillator the eigenvalues w_n satisfy the conditions²⁸

$$\sum_{n=0}^{\infty} w_n = 1, \quad |w_n| \leq 2. \quad (46)$$

The SR density matrix associated to the SR phase space point $(0, 0)$,

$$\tilde{\sigma}_\omega(x, x'; 0, 0) = \langle x | \hat{\tilde{\sigma}}_\omega(0, 0) | x' \rangle, \quad (47)$$

is shown in Fig. 2a. The first eigenfunctions and eigenvalues of this operator are shown in Fig. 2b.

4.4 Formulation of Canonical Averages as SR Phase Space Averages

We want to formulate the SR phase space representation of the canonical average of an arbitrary operator, \hat{A} . By applying \hat{A} to Eq. (29) and by taking into account the linear property of quantum mechanical operators, one gets

$$\hat{A}\hat{\rho}_a(f, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP \hat{A}\hat{\sigma}(X, P) e^{\beta f X} e^{\beta v P}. \quad (48)$$

From the linear property of the two-sided Laplace transform, one derives

$$\text{Tr}[\hat{A}\hat{\rho}_a(f, v)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP \text{Tr}[\hat{A}\hat{\sigma}(X, P)] e^{\beta f X} e^{\beta v P}. \quad (49)$$

Dividing both members by $Z_a(f, v)^{-1}$, and multiplying and dividing the integrand by $C(X, P)$, we obtain

$$\langle \hat{A} \rangle_{\rho_a(f, v)} = [Z_a(f, v)]^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP C(X, P) \langle \hat{A} \rangle_{\sigma(X, P)} e^{\beta f X} e^{\beta v P}, \quad (50)$$

where the average of \hat{A} using the SR density operator is

$$\langle \hat{A} \rangle_{\sigma(X, P)} = \text{Tr}[\hat{A}\hat{\sigma}(X, P)]. \quad (51)$$

By setting $f = 0$ and $v = 0$ in Eq. (50), one gets

$$\langle \hat{A} \rangle = \{ \langle \hat{A} \rangle_{\sigma(X, P)} \}. \quad (52)$$

This relation shows that arbitrary canonical averages can be represented as SR phase space averages.

4.5 Formulation of Time Correlation Functions as SR Phase Space Averages

The *time correlation function* of the position operator \hat{x} and an arbitrary operator \hat{A} is defined by²⁹

$$C_{xA}(t) \equiv \langle \hat{x}(0)\hat{A}(t) \rangle = Z^{-1} \text{Tr}[e^{-\beta \hat{H}} \hat{x}(0)\hat{A}(t)], \quad (53)$$

where the Heisenberg operator $\hat{A}(t)$ is

$$\hat{A}(t) = e^{i \frac{\hat{H} t}{\hbar}} \hat{A} e^{-i \frac{\hat{H} t}{\hbar}}. \quad (54)$$

The *Kubo transformed time correlation function* is defined as²⁹

$$K_{xA}(t) \equiv \langle \hat{x}(0); \hat{A}(t) \rangle = \int_0^{\beta} \frac{d\lambda}{\beta} \langle \hat{x}(-i\lambda\hbar)\hat{A}(t) \rangle, \quad (55)$$

where the operator $\hat{x}(-i\lambda\hbar)$ is [see Eq. (54)]

$$\hat{x}(-i\lambda\hbar) = e^{\lambda \hat{H}} \hat{x} e^{-\lambda \hat{H}}. \quad (56)$$

The definition in Eq. (55) is written in a form that is not ideal for grasping the physical meaning of Kubo transformed time correlation functions. An alternative definition is based on *linear response theory* and the *fluctuation-dissipation theorem*.²⁹ We are interested in near-equilibrium states driven by the external force, f . Let us assume that the external

perturbation started to work in the infinite past, i.e., at $t \rightarrow -\infty$, when the system with Hamiltonian $\hat{H}_a(f, 0)$ was in equilibrium at a certain temperature, i.e., the density matrix was initially canonical, $\hat{\rho}_a(f, 0)$. At $t = 0$ the external perturbation is set to zero ($f = 0$) and then the system evolves in time under the action of the unperturbed Hamiltonian \hat{H} . The average of an arbitrary operator \hat{A} at $t > 0$ is

$$\langle \hat{A}(t) \rangle_{\rho_a(f,0)} = [Z_a(f, 0)]^{-1} \text{Tr} [\hat{\rho}_a(f, 0) \hat{A}(t)] , \quad (57)$$

where the time dependence of the Heisenberg operator $\hat{A}(t)$ is determined by the unperturbed Hamiltonian \hat{H} [see Eq. (54)].

The Kubo transformed time correlation function can be defined by the following expression, which is a version of the quantum fluctuation-dissipation theorem²⁹

$$\langle \hat{x}(0); \hat{A}(t) \rangle = \frac{1}{\beta} \left[\frac{\partial \langle \hat{A}(t) \rangle_{\rho_a(f,0)}}{\partial f} \right]_{f=0} + \langle \hat{x}(0) \rangle \langle \hat{A}(0) \rangle , \quad (58)$$

This expression displays, more clearly than Eq. (55), the physical meaning of Kubo transformed time correlation functions. Note that the definition of the derivative implies

$$\left[\frac{\partial \langle \hat{A}(t) \rangle_{\rho_a(f,0)}}{\partial f} \right]_{f=0} = \lim_{f \rightarrow 0} \frac{\langle \hat{A}(t) \rangle_{\rho_a(f,0)} - \langle \hat{A}(0) \rangle}{f} , \quad (59)$$

where we have considered that the average of $\hat{A}(t)$ is stationary if the external force vanishes ($f = 0$), i.e.,

$$\langle \hat{A}(t) \rangle_{\rho_a(0,0)} \equiv \langle \hat{A}(t) \rangle = \langle \hat{A}(0) \rangle . \quad (60)$$

Eqs. (58) and (59) imply that the time dependence of the Kubo transformed function $K_{xA}(t)$ is determined by the time dependence of the average $\langle \hat{A}(t) \rangle_{\rho_a(f,0)}$ for a vanishing small value of f .

Our next goal is to derive the representation of Kubo transformed time correlation functions as SR phase space averages. Note that the derivation of Eq. (50), giving the average, $\langle \hat{A} \rangle_{\rho_a(f,v)}$, as an integral over the SR phase space, is also valid if the operator \hat{A} is substituted by $\hat{A}(t)$. Then, by applying Eq. (50) to the operator $\hat{A}(t)$ and after setting $v = 0$ in the resulting expression, one gets

$$\langle \hat{A}(t) \rangle_{\rho_a(f,0)} = [Z_a(f, 0)]^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP C(X, P) \langle \hat{A}(t) \rangle_{\sigma(X,P)} e^{\beta f X} . \quad (61)$$

The derivative of the last expression with respect to f is

$$\begin{aligned} \frac{\partial \langle \hat{A}(t) \rangle_{\rho_a(f,0)}}{\partial f} = [Z_a(f, 0)]^{-1} \beta & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP C(X, P) X \langle \hat{A}(t) \rangle_{\sigma(X,P)} e^{\beta f X} - \\ & \beta \langle \hat{x}(0) \rangle_{\rho_a(f,0)} \langle \hat{A}(t) \rangle_{\rho_a(f,0)} \end{aligned} \quad (62)$$

where the second term comes from the derivative of $[Z_a(f, 0)]^{-1}$ with respect to f by noting that

$$Z_a(f, 0) = e^{-\beta F_a(f,0)} , \quad (63)$$

$$\frac{\partial Z_a(f, 0)}{\partial f} = -Z_a(f, 0)\beta \frac{\partial F_a(f, 0)}{\partial f} = Z_a(f, 0)\beta \langle \hat{x}(0) \rangle_{\rho_a(f, 0)}. \quad (64)$$

$F_a(f, 0)$ is the auxiliary free energy. For the particular case that $f = 0$, Eq. (62) reads

$$\left[\frac{\partial \langle \hat{A}(t) \rangle_{\rho_a(f, 0)}}{\partial f} \right]_{f=0} = Z^{-1}\beta \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dX dP C(X, P) X \langle \hat{A}(t) \rangle_{\sigma(X, P)} - \beta \langle \hat{x}(0) \rangle \langle \hat{A}(0) \rangle, \quad (65)$$

The last equation can be alternatively written as

$$\left[\frac{\partial \langle \hat{A}(t) \rangle_{\rho(f, 0)}}{\partial f} \right]_{f=0} = \beta \{ X \langle \hat{A}(t) \rangle_{\sigma(X, P)} \} - \beta \langle \hat{x}(0) \rangle \langle \hat{A}(0) \rangle. \quad (66)$$

Comparing the last expression with Eq. (58), we see that

$$\langle \hat{x}(0); \hat{A}(t) \rangle \equiv \{ X \langle \hat{A}(t) \rangle_{\sigma(X, P)} \}, \quad (67)$$

i.e., the Kubo transformed time correlation function $\langle \hat{x}(0); \hat{A}(t) \rangle$ can be expressed as a SR phase space average. An analogous derivation using the auxiliary Hamiltonian $\hat{H}(0, v)$ leads to the result

$$\langle \hat{p}(0); \hat{A}(t) \rangle \equiv \{ P \langle \hat{A}(t) \rangle_{\sigma(X, P)} \}. \quad (68)$$

Note that these results can not be generalized to the calculation of a correlation function $\langle \hat{B}(0); \hat{A}(t) \rangle$, if the operator \hat{B} is different from \hat{x} or \hat{p} .

5 Constrained Time Evolution of the Operator $\hat{\tilde{\sigma}}(X, P)$

We have defined the most important *formal relations* that characterize the SR phase space. Our next goal is the formulation of a *practical approximation* aiming at the calculation of the Kubo transformed time correlation function as a SR phase space average using Eqs. (67) or (68). The dynamic information in these expressions is carried by the average

$$\langle \hat{A}(t) \rangle_{\sigma(X, P)} = \text{Tr}[\hat{A}(t)\hat{\tilde{\sigma}}(X, P)] = \text{Tr}[\hat{A}\hat{\tilde{\sigma}}(X, P; t)], \quad (69)$$

where

$$\hat{\tilde{\sigma}}(X, P; t) = [C(X, P)]^{-1} \hat{\sigma}(X, P; t), \quad (70)$$

and $\hat{A}(t)$ and $\hat{\sigma}(X, P; t)$ were defined by Eqs. (41) and (54), respectively.

The exact calculation of $\langle \hat{A}(t) \rangle_{\sigma(X, P)}$ implies [see Eq. (69)] the determination of the time evolution of the SR density operator associated to (X, P)

$$\hat{\tilde{\sigma}}(X, P) \text{ (time }=0) \longrightarrow \hat{\tilde{\sigma}}(X, P; t) \text{ (time }=t). \quad (71)$$

The computation of this time evolution is, for a general quantum system, a difficult problem. Moreover, the calculation of a SR phase space average implies solving this difficult problem for a set of operators associated with different points (X, P) . The only practical route to undertake this calculation lies on the formulation of an approximate dynamics for the SR density operator. CMD should be understood as an approximate dynamics for the SR density operator. One clear reason for the difficulty in understanding the original CMD

formulation is that CMD was formulated in 1994,¹⁸ while the SR density operator was first defined in 1999.^{21,24} In this section, an approximate dynamics for the SR density operator is presented by following these steps:

- discussion of a factorization property of the SR density matrix
- formulation of a *variational constrained dynamics* for the SR density operator,
- study of the harmonic and high temperature limits of the dynamic approximation.

5.1 Factorization Property of the Matrix Elements $\sigma(x, x'; X, P)$

The position representation of the SR density operator has a factorization property, that is valid if the Hamiltonian \hat{H} depends quadratically on the momentum operator \hat{p} . The derivation has been published elsewhere,^{21,28} the final result is that $\sigma(x, x'; X, P)$ factorizes into X - and P -dependent factors

$$\sigma(x, x'; X, P) = \sigma_X(x, x'; X) \sigma_P(x, x'; P). \quad (72)$$

The P -dependent factor is

$$\sigma_P(x, x'; P) = \left(\frac{\beta}{2\pi m} \right)^{\frac{1}{2}} e^{-\frac{\beta P^2}{2m}} e^{\frac{m(x-x')^2}{2\beta \hbar^2}} e^{i \frac{P(x-x')}{\hbar}}. \quad (73)$$

The X -dependent factor, $\sigma_X(x, x'; X)$ is the position representation of the operator $\hat{\sigma}_X(X)$, that is defined as

$$\hat{\sigma}_X(X) = \int_{-\infty}^{\infty} dP \hat{\sigma}(X, P). \quad (74)$$

The matrix elements, $\sigma_X(x, x'; X)$, represent constrained propagators whose PI representation is the following fixed centroid path integral

$$\sigma_X(x, x'; X) = \int_{x'}^x D_x[x(u)] \delta(X - x_c) e^{-\frac{S_x[x(u)]}{\hbar}}. \quad (75)$$

The diagonal elements $\sigma_P(x, x; P)$ are [see Eq. (73)], a constant independent of x . This fact implies that the trace of $\sigma(x, x'; X, P)$ factorizes also into X - and P -dependent terms. Using Eq. (72) to calculate this trace, one gets

$$C(X, P) = \int_{-\infty}^{\infty} dx \sigma_X(x, x; X) \sigma_P(x, x; P), \quad (76)$$

$$C(X, P) = C_P(P) \int_{-\infty}^{\infty} dx \sigma_X(x, x; X) \equiv C_P(P) C_X(X). \quad (77)$$

$C_P(P)$ is the *momentum density* in the SR phase space

$$C_P(P) \equiv \sigma_P(x, x; P) = \left(\frac{\beta}{2\pi m} \right)^{\frac{1}{2}} e^{-\frac{\beta P^2}{2m}}, \quad (78)$$

which has the form of a *classical momentum distribution*. $C_X(X)$ is the trace of the operator $\hat{\sigma}_X(X)$, and it is the *position density* in the SR phase space. This density is used to define the effective classical potential, $V_{ecp}(X)$, as

$$C_X(X) = \left(\frac{m}{2\pi\hbar^2\beta} \right)^{\frac{1}{2}} e^{-\beta V_{ecp}(X)}. \quad (79)$$

Other property derived from the fact that $\sigma_P(x, x; P)$ does not depend on x is the following: The averages $\langle \hat{A} \rangle_{\sigma(X, P)}$ of operators defined as arbitrary functions of \hat{x} do not depend on the variable P . For example, the *mean force*, $f_m(X)$, is the average of the force operator, \hat{f}_r ,

$$\langle \hat{f}_r \rangle_{\sigma(X, P)} = Z^{-1} \int_{-\infty}^{\infty} dx \left[-\frac{dV(x)}{dx} \right] \sigma_X(x, x; X) \equiv f_m(X). \quad (80)$$

f_m is a function of X but not of P . Another important average is the dispersion of the position operator \hat{x}

$$\delta x^2(X) = \langle \hat{x}^2 \rangle_{\sigma(X, P)} - [\langle \hat{x} \rangle_{\sigma(X, P)}]^2, \quad (81)$$

that does not depend on the coordinate P . We note that

$$\langle \hat{x} \rangle_{\sigma(X, P)} = X \quad (82)$$

5.2 A Variational Constrained Dynamic Approximation

An approximate dynamics for the operator $\hat{\sigma}(X, P)$ is defined by the following two conditions:

- (i) *Constrained dynamics*: The dynamic state at an arbitrary time t is constrained to be a SR density operator, $\hat{\sigma}(X, P)$,

$$\hat{\sigma}(X, P; t) \approx \hat{\sigma}(X(t), P(t)). \quad (83)$$

The constrained time evolution of $\hat{\sigma}(X, P; t)$ is then characterized by a trajectory $[X(t), P(t)]$ in the SR phase space. A representation of the constrained dynamics is given in Fig. 3.

- (ii) *Variational short time approximation*: The constrained dynamic equations are derived from the Gauss principle of least constraint, i.e., from the condition that the difference, in a least square sense, between the exact and constrained short time dynamics of $\hat{\sigma}(X, P)$ is minimum.

The derivation of the variational short time approximation is done in the position representation. The algebra is straightforward, but rather lengthy. Therefore, we summarize the main steps of the derivation:

- (1) The exact (e) time derivative of the matrix elements $\rho_a(x, x'; f, v)$ is derived using the time dependent Schrödinger equation. The result is

$$\left[\frac{\partial \rho_a(x, x'; f, v)}{\partial t} \right]_e = \left[\frac{f(x - x')}{i\hbar} - v \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial x'} \right) \right] \rho_a(x, x'; f, v). \quad (84)$$

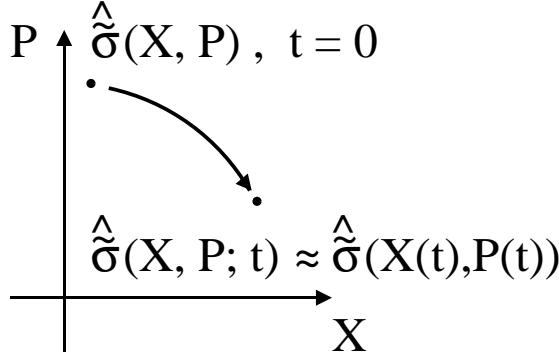


Figure 3. Schematic picture of the constrained dynamics for the operator $\hat{\tilde{\sigma}}(X, P)$, defined at $t = 0$. The dynamic state at time t , $\hat{\tilde{\sigma}}(X, P; t)$, is constrained to be a SR density operator, $\hat{\tilde{\sigma}}(X(t), P(t))$. The arrow represents the trajectory $(X(t), P(t))$ that characterizes the constrained dynamics from $t = 0$ to time t .

(2) The exact time derivative of the matrix elements $\tilde{\sigma}(x, x'; X, P)$ is derived with the help of the two-sided Laplace relation between $\rho_a(x, x'; f, v)$ and $\sigma(x, x'; X, P)$ [see Eq. (28)]. The result is

$$\left[\frac{\partial \tilde{\sigma}(x, x'; X, P)}{\partial t} \right]_e = \left[i \left(\frac{x - x'}{\beta \hbar} \right) \left(\frac{\partial}{\partial X} + \frac{\partial}{\partial x} + \frac{\partial}{\partial x'} \right) - \frac{P}{m} \left(\frac{\partial}{\partial x} + \frac{\partial}{\partial x'} \right) \right] \tilde{\sigma}(x, x'; X, P). \quad (85)$$

(3) The constrained (c) dynamic equation for $\tilde{\sigma}(x, x'; X, P)$ is formulated according to condition (i)

$$\left[\frac{\partial \tilde{\sigma}(x, x'; X, P)}{\partial t} \right]_c = \frac{\partial \tilde{\sigma}(x, x'; X, P)}{\partial X} \dot{X} + \frac{\partial \tilde{\sigma}(x, x'; X, P)}{\partial P} \dot{P}. \quad (86)$$

where

$$\dot{X} = \frac{dX}{dt} ; \dot{P} = \frac{dP}{dt}. \quad (87)$$

(4) The variational short time approximation is derived from minimizing, with respect to \dot{X} and \dot{P} , the following function

$$\mathcal{I}(\dot{X}, \dot{P}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dx' \left| \left[\frac{\partial \tilde{\sigma}(x, x'; X, P)}{\partial t} \right]_e - \left[\frac{\partial \tilde{\sigma}(x, x'; X, P)}{\partial t} \right]_c \right|^2, \quad (88)$$

where $|...|$ indicates the modulus of a complex number.

(5) The minimization results in the following set of equations that define the constrained dynamics

$$\dot{X} = -\frac{P}{m} \frac{x_1(X, P)}{x_2(X, P)}, \quad (89)$$

$$\dot{P} = \frac{p_1(X, P)}{p_2(X, P)}. \quad (90)$$

The functions $x_1(X, P)$ and $x_2(X, P)$ are defined by the following integrals, where the shorthand notation, $\sigma_X \equiv \sigma_X(x, x'; X)$, $\sigma \equiv \sigma(x, x'; X, P)$, and $f_m \equiv f_m(X)$, has been used

$$x_1(X, P) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dx' |\sigma|^2 \left(\frac{\partial \ln \sigma_X}{\partial x} + \frac{\partial \ln \sigma_X}{\partial x'} \right) \left(\frac{\partial \ln \sigma_X}{\partial X} - \beta f_m \right), \quad (91)$$

$$x_2(X, P) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dx' |\sigma|^2 \left(\frac{\partial \ln \sigma_X}{\partial X} - \beta f_m \right)^2. \quad (92)$$

The functions $p_1(X, P)$ and $p_2(X, P)$ are defined by

$$p_1(X, P) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dx' |\sigma|^2 (x - x')^2 \frac{1}{\beta} \left(\frac{\partial \ln \sigma_X}{\partial X} + \frac{\partial \ln \sigma_X}{\partial x} + \frac{\partial \ln \sigma_X}{\partial x'} \right), \quad (93)$$

$$p_2(X, P) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dx' |\sigma|^2 (x - x')^2. \quad (94)$$

The variational equations for the constrained dynamics imply the determination of the diagonal and off diagonal elements of the matrix $\sigma(x, x'; X, P)$, a cumbersome computational task. Therefore, we will introduce an additional approximation to simplify the variational equations.

5.3 The RMD Approximation

The following *factorization approximation* is applied to simplify the variational equations

$$|\sigma(x, x'; X, P)|^2 \approx \sigma(x, x; X, P) \sigma(x', x'; X, P). \quad (95)$$

This relation is *exact* for a pure state. E.g., if $\sigma(x, x')$ denotes a pure state, then

$$|\sigma(x, x')|^2 = |\langle x | \psi \rangle \langle \psi | x' \rangle|^2 = \sigma(x, x) \sigma(x', x'). \quad (96)$$

We recall that the SR density operator is a pure state only at temperature $T = 0$.^{31,32} Therefore, the use of the factorization approximation implies that the variational character of the dynamic equations is lost, except in the limit $T = 0$. The computational advantage of this approximation is that only the diagonal elements of the SR density matrix are needed. We define the function $\phi(x; X)$ as the square root of these diagonal elements

$$\phi(x; X) = [\tilde{\sigma}(x, x; X, P)]^{1/2} \equiv [\tilde{\sigma}_X(x, x; X)]^{1/2}. \quad (97)$$

Applying the factorization approximation to the variational equations in Eqs. (89) and (90), one gets

$$\dot{X} = \frac{P}{m} \gamma(X), \quad (98)$$

$$\dot{P} = f_m(X) + \nu(X). \quad (99)$$

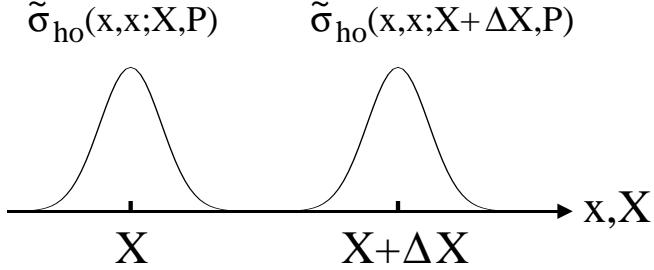


Figure 4. Schematic representation of the translational symmetry of the diagonal elements of the SR density matrix, $\tilde{\sigma}_{ho}(x, x; X, P)$. The diagonal elements associated to two SR phase space positions, X and $X + \Delta X$, are functions of x that differ by an overall translation. This result is valid for a harmonic oscillator at arbitrary temperature and for an arbitrary potential in the high temperature limit.

This dynamic approximation is called the *constrained response matrix dynamics* (RMD). The function $\gamma(X)$ represents a correction factor for the particle mass in the constrained dynamics. It is defined as

$$\gamma(X) = -\frac{\gamma_1(X)}{\gamma_2(X)}, \quad (100)$$

$$\gamma_1(X) = \int_{-\infty}^{\infty} dx \frac{\partial \phi(x; X)}{\partial X} \frac{\partial \phi(x; X)}{\partial x}, \quad (101)$$

$$\gamma_2(X) = \int_{-\infty}^{\infty} dx \left[\frac{\partial \phi(x; X)}{\partial X} \right]^2. \quad (102)$$

The function $\nu(X)$ is a temperature dependent correction term for the mean force $f_m(X)$,

$$\nu(X) = \frac{k_B T}{2} \frac{\partial \ln [\delta x^2(X)]}{\partial X}. \quad (103)$$

Note that $\nu(X)$ vanishes at temperature $T = 0$ and that $\delta x^2(X)$ is the dispersion of the operator \hat{x} defined in Eq. (81).

5.4 The Harmonic and High Temperature Limits of the RMD Approximation

Let us apply the RMD equations to a harmonic oscillator (ho) at arbitrary temperature and to the high temperature limit of an arbitrary potential. For these two cases, the diagonal elements of the SR density matrix satisfy the following equation²⁸

$$\tilde{\sigma}_{ho}(x, x; X + \Delta X, P) = \tilde{\sigma}_{ho}(x - \Delta X, x - \Delta X; X, P). \quad (104)$$

This relation means that the diagonal elements of the SR density matrix associated to the phase space points (X, P) and $(X + \Delta X, P)$ are identical, apart from an overall translation by ΔX (see Fig 4). Thus, the function $\phi_{ho}(x; X)$, defined in Eq. (97), satisfies the condition

$$\phi_{ho}(x; X + \Delta X) = \phi_{ho}(x - \Delta X; X). \quad (105)$$

This relation can be rewritten using a Taylor expansion of the l.h.s. around X and of the r.h.s. around x . The result, to first order in both Δx and ΔX , is

$$\phi_{ho}(x; X) + \Delta X \frac{\partial \phi_{ho}(x; X)}{\partial X} = \phi_{ho}(x; X) - \Delta X \frac{\partial \phi_{ho}(x; X)}{\partial x} . \quad (106)$$

The last equation implies that

$$\frac{\partial \phi_{ho}(x; X)}{\partial X} = - \frac{\partial \phi_{ho}(x; X)}{\partial x} . \quad (107)$$

By introducing this result in the definition of $\gamma(X)$ in Eq. (100), one gets

$$\gamma_{ho}(X) = 1 . \quad (108)$$

It is also easy to show that

$$\nu_{ho}(X) = 0 , \quad (109)$$

as Eq. (105) implies that the dispersion $\delta x^2(X)$ is a constant independent of the coordinate X .

The previous conditions ($\gamma_{ho} = 1$, $\nu_{ho} = 0$) imply that the RMD equations become identical to CMD in the cases where CMD is known to be exact, i.e., for a harmonic oscillator at arbitrary temperature and for an arbitrary potential in the high temperature limit. In the numerical test examples of the following Section, we will see that the correction factors $\gamma(X)$ and $\nu(X)$ have little influence in the resulting dynamics (except in the study of quantum tunneling at temperature $T = 0$). Thus, for practical purposes RMD produces nearly the same dynamic results as CMD and then our derivation of RMD provides new insight into the physical meaning of CMD.

Let us summarize several properties of the RMD and CMD equations:

- At temperature $T = 0$ the RMD approximation is a variational approximation. CMD is, like RMD, a constrained dynamics but it is not a variational approximation. Then, we expect that RMD will provide more accurate results than CMD at temperature $T = 0$.
- At any finite temperature the RMD approximation is not variational, because it includes the factorization approximation given in Eq. (95). One expects that the quality of the RMD and CMD results *decreases* as the temperature increases above $T = 0$, i.e., as the temperature deviates from the variational $T = 0$ limit.
- However, in the high temperature limit, RMD and CMD go over the correct classical limit. Therefore, above some unspecified temperature, the quality of the RMD and CMD results should *increase* as the temperature increases towards the classical limit.
- CMD can be derived from RMD as a harmonic or high temperature limit.
- If $\gamma(X) \neq 1$ or $\nu_{ho}(X) \neq 0$ the RMD equations generate a non-Hamiltonian SR phase space dynamics; i.e., they imply a nonzero phase space compressibility

$$\frac{d}{dt} C[X(t), P(t)] \neq 0 . \quad (110)$$

CMD always conserves the SR phase space probability.

6 Numerical Test on Model Systems

As we stated in the introduction, the capability of CMD (and also RMD) to describe real time quantum dynamics rest on the assumption that the effective classical potential, V_{ecp} , represents some kind of realistic average of dynamic properties. In this Section we apply the CMD and RMD approximations to the study of simple problems that can be solve exactly by mean of numerical techniques.

The main questions to be addressed are:

- In which cases does V_{ecp} carry an accurate dynamic information?
- How does the RMD approximation compare to CMD?

6.1 The Zero Temperature Limit of CMD and RMD

We consider the quantum dynamics of a particle of mass $m = 16$ a.u. moving either on a double-well potential (V_{dw}) or on a quartic potential (V_q), defined as

$$V_{dw}(x) = \frac{1}{4}(x^2 - 1)^2 , \quad (111)$$

$$V_q(x) = 2.2015x^4 , \quad (112)$$

where x is expressed in a.u.. The V_{ecp} has the following simple physical meaning at temperature $T = 0$:^{31,32} It is related to the ground state energy, $E_0(f)$, of the auxiliary Hamiltonian, $\hat{H}_a(f, 0)$,

$$\hat{H}_a(f, 0)|\phi_0(f)\rangle = E_0(f)|\phi_0(f)\rangle . \quad (113)$$

$|\phi_0(f)\rangle$ is the ground state of $\hat{H}_a(f, 0)$. The relation between $V_{ecp}(X)$ and $E_0(f)$ has the form of a Legendre transform

$$V_{ecp}(X) = E_0(f) + fX , \quad (114)$$

where the centroid position X is the average of the position operator \hat{x}

$$X = \langle\phi_0(f)|\hat{x}|\phi_0(f)\rangle = -\frac{dE_0(f)}{df} . \quad (115)$$

The calculation of V_{ecp} at $T = 0$, can be done by solving numerically, for a set of values of the parameter f , the time independent Schrödinger equation for the Hamiltonian $\hat{H}_a(f, 0)$. The function V_{ecp} for the studied model potentials are shown in Fig. 5 as continuous lines. The dotted lines are harmonic approximations at the potential minimum located at $X = 0$.

At temperature $T = 0$, the SR density operators associated to the SR phase space points $(X, 0)$ are pure states identical to the kets $|\phi_o(f)\rangle$.^{31,32} The centroid position X and the force parameter f are related by Eq. (115). Thus, the RMD and CMD approximations are wave packets dynamics in this $T = 0$ limit. The difference between both approaches is determined by the deviation from unity of the RMD parameter $\gamma(X)$ that appears in Eq. (98). The functions $\gamma(X)$ for the two studied model potentials are shown in Fig. 6. The deviation of $\gamma(X)$ from unity is appreciable for V_{dw} , but small for V_q .

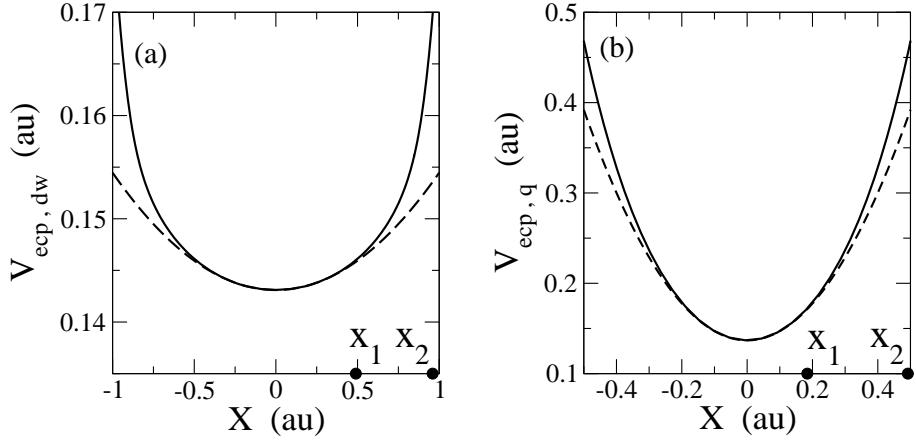


Figure 5. (a) Effective classical potential at temperature $T = 0$ corresponding the double-well potential V_{dw} . The dotted line represents a harmonic approximation around the minimum at $X = 0$. The two positions (X_1 and X_2), marked as filled circles, define two initial ($t = 0$) states used in the dynamic study. (b) The same information is provided for the quartic potential V_q .

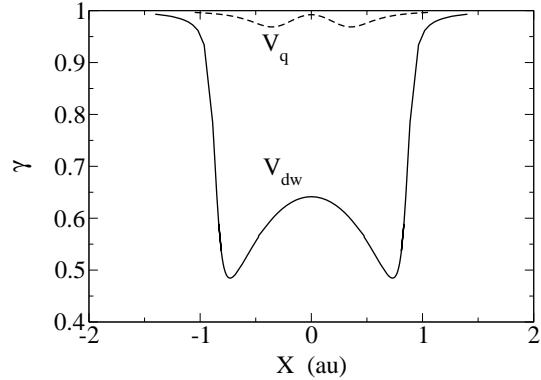


Figure 6. (a) Parameter $\gamma(X)$ for the two studied model potentials at $T = 0$.

For the potential V_{dw} , we study the dynamics of two different initial states corresponding to the SR density operators associated to the points $(X_1, 0)$ and $(X_2, 0)$. The coordinates X_1 and X_2 are shown in Fig. 5a. The coordinate X_1 belongs to a region where V_{ecp} is well described by a harmonic approximation (dotted line in Fig. 5a), while at position X_2 , V_{ecp} deviates clearly from the harmonic limit. The exact phase space trajectory $[X(t), P(t)]$ corresponding to a given initial state, $\hat{\sigma}(X, P)$, is determined by the following averages

$$X(t) = \langle \hat{x}(t) \rangle_{\sigma(X, P)} , \quad (116)$$

$$P(t) = \langle \hat{p}(t) \rangle_{\sigma(X, P)} . \quad (117)$$

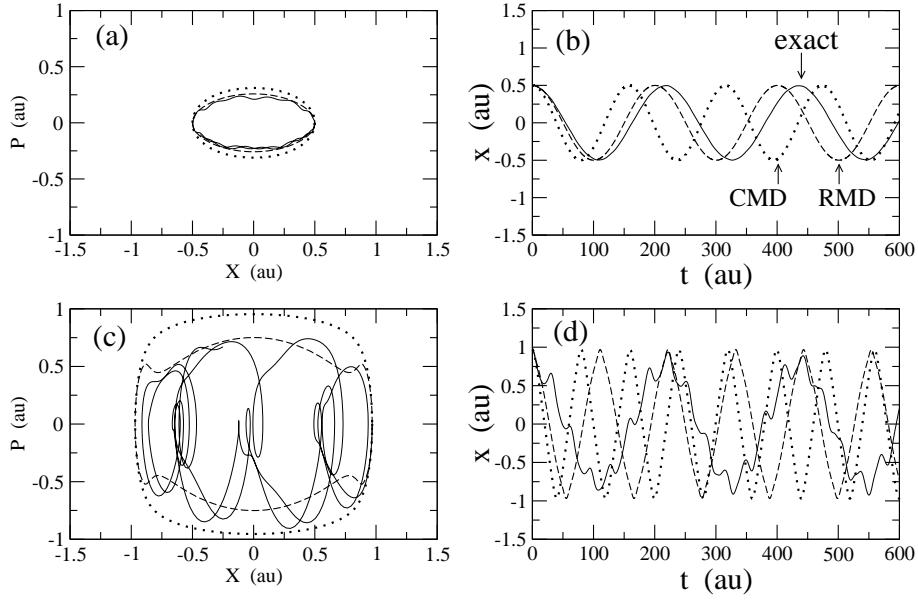


Figure 7. (a) Phase space trajectory for the SR density operator associated to the point $(X_1, 0)$ at time $t = 0$. (b) The average $X(t)$ for SR density operator associated to $(X_1, 0)$ at $t = 0$. (c) Phase space trajectory for the SR density operator associated to the point $(X_2, 0)$ at time $t = 0$. (d) The average $X(t)$ for SR density operator associated to $(X_2, 0)$ at $t = 0$. In the four panels (a)-(d), the exact results are shown by continuous lines, the RMD results by dashed lines, and the CMD results by dotted line. The results corresponds to the double-well model potential, V_{dw} .

The exact phase space trajectories corresponding to the initial states associated to $(X_1, 0)$ and $(X_2, 0)$ are compared in Figs. 7a and 7c with the results derived from the RMD and CMD approximations. The exact result for $X(t)$ is compared to the RMD and CMD expectations in Figs. 7b and 7d. The RMD approximation provides better results than the CMD approximation for the phase space trajectory $[X(t), P(t)]$ and also for the frequency of the oscillation in $X(t)$. The main effect of the factor $\gamma(X)$ in the RMD equations, is the slowing down of this frequency, when compared to the CMD results. The improved RMD results are consequence of its variational character at $T = 0$, a property not shared by CMD.

We observe that the RMD and CMD results derived for the initial state given by $(X_1, 0)$ are more accurate than those one derived for $(X_2, 0)$. This behavior can be rationalized in terms of the dynamical information carried out by the effective classical potential. The initial state defined by $(X_2, 0)$ explores a larger X region along its dynamic evolution than the initial state defined by $(X_1, 0)$, as it is seen by comparing Figs. 7a and 7c. The anharmonicity of V_{ecp} (see Fig. 5) causes that the frequency of the time oscillations in $X(t)$ increases for the initial state defined by $(X_2, 0)$ with respect the result obtained for $(X_1, 0)$ (see Figs. 7b and 7d). This anharmonic effect in V_{ecp} is an *unphysical dynamic result*, because the exact $X(t)$ does not display any change in the oscillation frequency as a function of the initial state. The exact result for $X(t)$ shows that the main dynamic effect associated

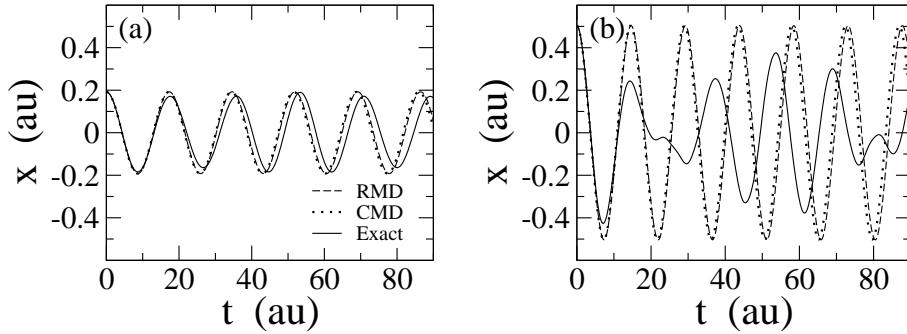


Figure 8. (a) The function $X(t)$ for a initial state defined by the SR density operator associated to the phase state point $(X_1, 0)$. The exact result is compared to the constrained dynamic approximations (RMD and CMD). (b) Exact, RMD, and CMD results for $X(t)$ for an initial state defined by $(X_2, 0)$. The meaning of the plotted lines is the same as for the panel (a). The results correspond to the quartic potential, V_q . Note that the RMD and CMD results are nearly identical.

to the change in the initial condition is a coherent superposition of two different frequencies. Both RMD and CMD are completely unable to reproduce this coherent superposition of frequencies. The final conclusion is that the most meaningful dynamic information of V_{ecp} is contained only in the *harmonic region* around the potential energy minimum at $X = 0$ (see Fig. 5), and that the *anharmonic region* of V_{ecp} provides unphysical dynamic information.

In Fig. 8, we compare the exact result for $X(t)$ with those derived from the RMD and CMD approximations for the quartic potential V_q . We have considered two different initial states defined by the SR density operators associated to the points $(X_1, 0)$ and $(X_2, 0)$. The values X_1 and X_2 are shown in Fig. 5b. We find that for the quartic potential, V_q , the RMD and CMD results are, for practical purposes, indistinguishable. The constrained dynamic approximations (CMD, RMD) are more accurate for the initial state defined by $(X_1, 0)$, when the function $X(t)$ is characterized by oscillations with a unique frequency (see Fig. 8a). For the initial state defined by $(X_2, 0)$, the exact time evolution of $X(t)$ displays a coherent superposition of different frequencies. Again, we find that both RMD and CMD are unable to reproduce this coherent superposition (see Fig. 8b). The main change observed in the CMD and RMD results as a function of the initial condition is an increase in the frequency of the oscillations of $X(t)$. Again, the conclusion to be drawn is that the relevant dynamic information of V_{ecp} at $T = 0$ is reduced to the harmonic part of the potential around the energy minimum. The anharmonic region of V_{ecp} leads to unphysical results in the CMD and RMD approximations.

6.2 Finite Temperature CMD and RMD Results

We present some results for the time correlation function of the position operator at different temperatures

$$C(t) = \langle \hat{x}(0)\hat{x}(t) \rangle = Z^{-1} \text{Tr}[e^{-\beta \hat{H}} \hat{x} e^{i \frac{\hat{H}t}{\hbar}} \hat{x} e^{-i \frac{\hat{H}t}{\hbar}}]. \quad (118)$$

Several general properties of quantum time correlation functions can be readily derived by writing them in a basis of eigenfunctions of the Hamiltonian. Thus, $C(t)$ can be shown to be a *complex* function of t .

$$C(t) = C^R(t) + iC^I(t), \quad (119)$$

The corresponding Kubo transformed correlation function, $K(t)$, is a *real* function of t . By defining the Fourier transform of $K(t)$ as

$$\overline{K}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt K(t) e^{-i\omega t}, \quad (120)$$

one can readily show that^{18,29,30}

$$\overline{C}^R(\omega) = \frac{\beta\hbar\omega}{2} \coth\left(\frac{\beta\hbar\omega}{2}\right) \overline{K}(\omega), \quad (121)$$

$$i\overline{C}^I(\omega) = \frac{\beta\hbar\omega}{2} \overline{K}(\omega), \quad (122)$$

where $\overline{C}^R(\omega)$ and $\overline{C}^I(\omega)$ are the Fourier transforms of $C^R(t)$ and $C^I(t)$, respectively.

The real part, $C^R(t)$, of the time correlation function of the position operator has been studied for the quartic potential, V_q , at three different temperatures ($T_a = 0.03$, $T_b = 0.14$, and $T_c = 0.5$ a.u.). The energy difference between the first excited state and the ground state amounts to $\Delta E = 0.35$ a.u.. Thus, at the two lowest studied temperatures, T_a and T_b , one expects that the time correlation function will be mainly dominated by ground state dynamics, while at T_c one expects dynamic contributions from higher excited states. The exact $C^R(t)$ curves are shown in Fig. 9 by continuous lines. The results corresponding to the CMD (dotted line) and RMD (dashed line) approximations are also displayed in Fig. 9. The first conclusion from these data is that CMD and RMD provide nearly indistinguishable results at all temperatures. The second conclusion is that the time correlation function derived by either CMD or RMD decays to zero too fast as temperature increases. Comparing the CMD results at T_a and T_b , we note that at T_b the time correlation has decayed to zero at times larger than about $t = 40$ a.u.. This is clearly an unphysical behavior related to the anharmonicity of the effective classical potential. Our previous conclusion derived at temperature $T = 0$, that only the harmonic region of V_{ecp} carries meaningful dynamic information seems to be also true at *low enough temperatures*. From the data in Fig. 9, and also from other published numerical work,^{32–34,22} one can establish a temperature range

$$k_b T \lesssim \frac{\Delta E}{4}, \quad (123)$$

where CMD provides the most realistic results for low temperature quantum dynamics. At temperatures above this range the quality of the CMD results *decreases*. This fact is clearly seen in Fig. 9 by comparing the CMD results at T_a and T_b . The CMD results at T_a are a better approximation to the exact data at T_b , than the CMD results derived at T_b (see Fig. 10). Note that $T_a \approx \frac{\Delta E}{12}$ is in the range defined by Eq. (123), while $T_b \approx \frac{\Delta E}{2}$ is outside this range.

As the temperature increases above the low temperature range defined in Eq. (123), e.g., for the temperatures T_b and T_c in Fig. 9, the CMD data are accurate only for short

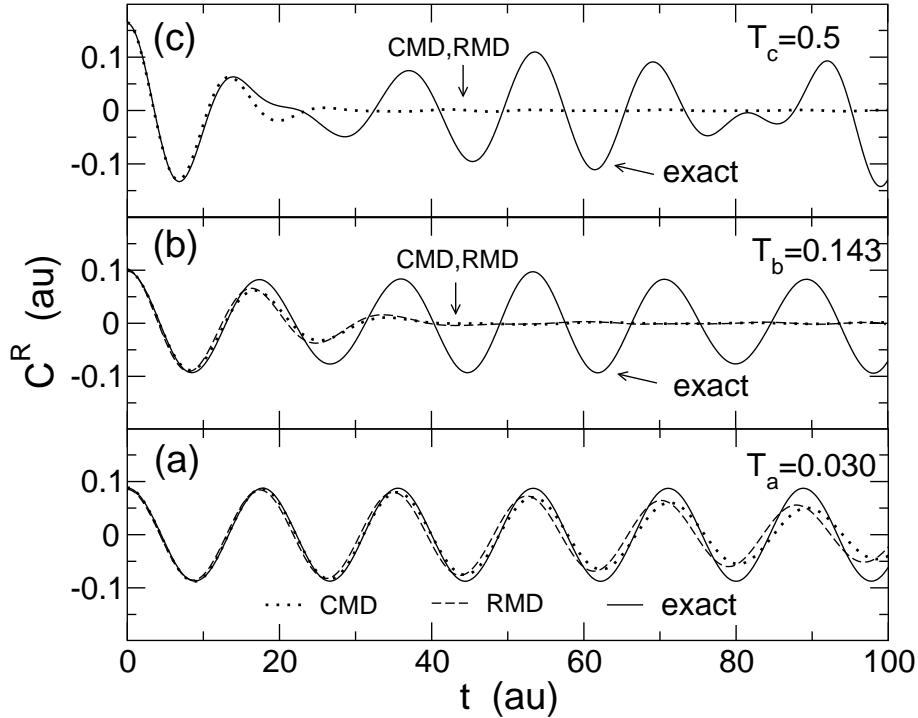


Figure 9. Real part, $C^R(t)$, of the time correlation function of the position operator as calculated for the model potential V_q . The exact results are compared to the *CMD* and *RMD* approximations. The *CMD* and *RMD* curves are nearly identical. Three temperatures were analyzed: (a) $T = 0.03$ au; (b) $T = 0.143$ au; (c) $T = 0.5$ a.u..

times. The value, t_m , defining the upper time where *CMD* is accurate should increase with temperature, because t_m becomes eventually infinity in the high temperature limit when *CMD* becomes exact. An interesting point should be to study how the time t_m depends on temperature.

Other numerical investigations comparing *CMD* and exact results for position time correlation functions has been published by Krilov and Berne,³⁴ Jang and Voth,²² and Cao and Voth.^{18,19} Let us summarize the main conclusions of our numerical investigation at low temperature, that should be applicable to vibrational problems in molecules, solids, impurities in solids, etc.

- The *RMD* approximation provides results nearly identical to *CMD*, except in the study of quantum tunneling, where the *RMD* approximation provides improved results.
- There is a low temperature region, defined as a function of the energy difference between the first excited and ground states [see Eq.(123)], where the significant dynamic information of V_{ecp} is the curvature of the potential around its energy minimum. Long time dynamics derived in this low temperature region by *CMD* is realistic.
- At temperatures above this low temperature region, the long time dynamics derived

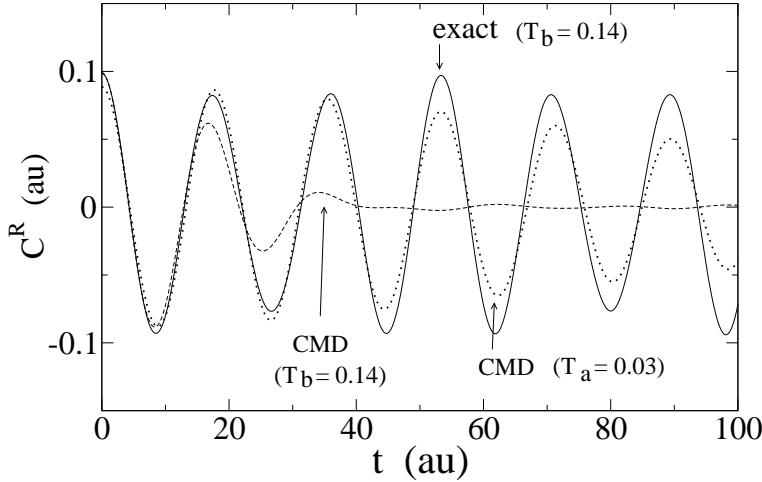


Figure 10. Comparison of the exact real part, $C^R(t)$, of the time correlation function of the position operator for the model potential V_q at temperature $T = 0.14$ a.u. with those results obtained from CMD approximations at temperatures $T = 0.14$ a.u. and $T = 0.03$ a.u.. The exact curve agrees better with the CMD result at low temperature.

by CMD has no physical meaning. However, the short time dynamics is predicted realistically.

6.3 Quantum Transmission Coefficients

The previous numerical results have shown that the RMD correction term, $\nu(X)$, for the mean force has little influence in the dynamics. In this Subsec. we check the influence of this term in the context of PI-QTST. The quantum correction factor to the classical rate constant is given within PI-QTST by^{11,15}

$$\Gamma_{QTST} = \frac{e^{-\beta \Delta V_{ecp}}}{e^{-\beta \Delta V}} \quad (124)$$

where ΔV is the classical potential energy barrier, while ΔV_{ecp} is the corresponding barrier for V_{ecp} . This factor measures the enhancement of the quantum rate with respect to the classical one. If the coordinates X_i and X_b label the reactant and barrier positions, the quantity ΔV_{ecp} can be calculated as an integral of mean force

$$\Delta V_{ecp} = - \int_{X_i}^{X_b} f_m(X) dX . \quad (125)$$

If one includes the RMD correction term [see Eq. (99)], $\nu(X)$, in the integrand of the last equation, one gets a modified effective potential

$$\Delta V_{RMD} = \Delta V_{ecp} - \frac{1}{2\beta} \ln \frac{\delta x^2(X_b)}{\delta x^2(X_i)} . \quad (126)$$

By substituting ΔV_{RMD} by ΔV_{eep} in the expression of Γ_{QTST} , we obtain the following quantum correction factor

$$\Gamma_{RMD} = \left[\frac{\delta x^2(X_b)}{\delta x^2(X_i)} \right]^{\frac{1}{2}} \Gamma_{QTST}. \quad (127)$$

The correction term, $\nu(X)$, leads to a simple dynamic factor in the PI-QTST expression for the rate constant. We have tested the capability of this expression to calculate, for a symmetric Eckart barrier, the transmission coefficient of an incident flux of protons ($m=1836$ a.u.), with initial velocities given by the classical Maxwell-Boltzmann distribution. The Eckart barrier is

$$V(x) = V_b \operatorname{sech}^2(x/a). \quad (128)$$

The parameters defining the potential are the barrier height $V_b = 0.2485$ eV and the distance $a = 0.3491$ Å. These values have been chosen to correspond to the same physical situation as that studied in Refs.^{15,35,36}

The quantum correction factor Γ evaluated exactly for the symmetric Eckart barrier is compared in Tab. 1 to the values of Γ_{QTST} and Γ_{RMD} . The column Γ_v was calculated using a dynamic preexponential factor derived by Gillan,¹¹ and also by Cao and Voth.³⁵ The dynamic factor derived from the RMD equation provides a realistic result.

T (K)	Γ_{QTST}	Γ_v	Γ_{RMD}	Γ
10000	1.00	1.00	1.00	1.00
503.27	1.42	1.62	1.52	1.52
301.96	2.70	3.45	3.12	3.10
150.98	105.3	141.3	177.0	161.9
81.97	1.54×10^7	2.65×10^7	3.97×10^7	3.77×10^7
26.60	1.71×10^{37}	4.28×10^{37}	6.23×10^{37}	7.84×10^{37}

Table 1. Quantum correction factors for the symmetric Eckart barrier at several selected temperatures. Γ_{QTST} is the PI-QTST result. Γ_v was obtained by multiplying the value, Γ_{QTST} , by a previously derived dynamic factor (see text). Γ_{RMD} is calculated using the dynamic factor derived from the RMD approximation. Γ is the exact value.

7 A Review of CMD Applications

The numerical solution of the CMD equations is a classical molecular dynamics problem. However, the computation of the mean force, f_m , is a quantum problem that must be handled numerically by PI simulations and requires the use of efficient algorithms. Although a detailed account of these algorithms is outside the scope of the present chapter, we quote in the next Subsection some relevant literature. The subsequent Subsections present a review of CMD applications on noble gases, liquid $p\text{-H}_2$, water, anharmonic molecule vibrations, and vibrational energy relaxations.

7.1 Numerical Algorithms

The PI computation of the mean force, f_m , can be performed either by MC or MD simulations. However, most CMD applications use the adiabatic or “on the fly” PI MD method.^{20,37,38} The setup of these simulations implies the transformation of the set of cartesian coordinates defining the “ring polymers” into a set of normal modes by a Fourier expansion. The centroid coordinate of each particle corresponds to the normal mode whose wave vector is zero. The centroid modes can be made the slowest moving modes in the MD simulation by assigning a sufficiently small mass to the non-centroid path modes. By virtue of the adiabatic principle, the slow degrees of freedom (the centroids) will move on the effective classical potential that is dynamically produced by the “on the fly” averaging of the non-centroid path modes. To ensure an ergodic exploration by PI MD it is also necessary to modify the dynamics of the normal modes by the attachment of a Nosé-Hoover chain thermostat to each normal mode.^{39,40} A detailed theoretical account of the combination of ab initio Car-Parrinello molecular dynamics with CMD has been presented by Marx *et al.*,⁴¹ while a similar implementation has been published by Pavese *et al.*⁴² The efficiency of CMD algorithms can be largely increased by the parallel implementation of the method.⁴³

7.2 Noble Gases

The calculation of the diffusion constant of Ne at $T = 40K$ was one of the first test of the CMD approximation.¹⁸ The interaction was described by a Lennard-Jones potential and the calculation of V_{ecp} was simplified by using the variational Feynman-Hibbs approximation.¹ The diffusion coefficient of the liquid was calculated as the time integral of the centroid velocity autocorrelation function. The CMD result is 7% lower than the value derived from a classical simulation.¹⁸

Miura *et al.*⁴⁴ have presented an interesting study of nonsuperfluid liquid ^4He at 4 K. The diffusion coefficient and the power spectra of the velocity autocorrelation function were derived by CMD simulations. This work presents also a calculation of the dynamic structure factor, $S_{coh}(k, \omega)$, of the liquid. This is an important quantity that is measured by neutron scattering experiments. In numerical simulations, $S_{coh}(k, \omega)$ can be derived from the relaxation function of the density fluctuations, that is defined as a Kubo transformed correlation function

$$R_{coh}(\mathbf{k}, t) = \frac{1}{N} \int_0^\beta \frac{d\tau}{\beta} \langle \hat{\rho}_{\mathbf{k}}(-i\tau) \hat{\rho}_{-\mathbf{k}}(t) \rangle , \quad (129)$$

where the density operator, $\hat{\rho}_{\mathbf{k}}$, is the following sum over the position operators, $\hat{\mathbf{r}}_i$, of the N system particles

$$\hat{\rho}_{\mathbf{k}} = \sum_{i=1}^N e^{i\mathbf{k}\hat{\mathbf{r}}_i} . \quad (130)$$

In addition to the CMD approximation, Miura *et al.*⁴⁴ make the assumption that the Kubo transformed time correlation function, $R_{coh}(\mathbf{k}, t)$, can be approximated by the correlation function associated to density fluctuations of the centroid positions

$$R_{coh}(\mathbf{k}, t) \approx \frac{1}{N} \{ \rho_{\mathbf{k}}^{(c)}(0) \rho_{-\mathbf{k}}^{(c)}(t) \} , \quad (131)$$

where $\rho_{\mathbf{k}}^{(c)}$ is the number density associated to the centroid positions, \mathbf{R}_i ,

$$\rho_{\mathbf{k}}^{(c)} = \sum_{i=1}^N e^{i\mathbf{k}\mathbf{R}_i}. \quad (132)$$

The assumption given in Eq. (131) has no physical justification, because the density operator $\hat{\rho}_{\mathbf{k}}$ is not a linear function of the position operator [see Eqs. (67) and (68)]. This assumption may lead to unphysical results, e.g., the value of the correlation function, $R_{coh}(\mathbf{k}, 0)$ at $t = 0$, is a static quantity that is incorrectly reproduced using Eq. (131) (see Fig. 3 of the original work).⁴⁴ Interestingly, several results derived for ${}^4\text{He}$ using this assumption show a remarkable agreement with experiment. In particular, the simulated dynamic structure factor, $S_{coh}(k, \omega)$, shows a satisfactory agreement with the experimental spectrum for $0.2 < k < 2.2 \text{ \AA}^{-1}$.

7.3 Liquid *para*-H₂

Liquid *p*-H₂ has been focused of a large number of CMD investigations. The diffusion coefficient of *p*-H₂ was determined by CMD at $T=14$ and 25 K, showing a good agreement to experimental data.³⁸ CMD simulations on liquid *p*-H₂ were also performed to check: (i) the quality of a pairwise approximation to the many-body function, V_{ecp} ;⁴⁵ (ii) the performance of a parallel CMD algorithm.⁴³

The diffusive and vibrational properties of clusters Li(*p*-H₂)_{*n*} (*n*=13, 55, and 180) and of a *p*-H₂ slab containing a lithium impurity has been investigated by CMD.^{46,47} The dynamic simulations focused on the computation of centroid mean square displacements and the power spectra of velocity autocorrelation functions. The generated centroid trajectories were analyzed to characterize the diffusion of the lithium impurity and the melting properties of the *p*-H₂ clusters.

Kinugawa⁴⁸ has presented a study of the dynamic structure factor, $S(k, \omega)$, of liquid *p*-H₂. This simulation is the first calculation of $S(k, \omega)$ by a CMD approach supplemented by the assumption in Eq. (131). The calculated spectral profile indicates the existence of collective dynamics modes in this quantum liquid of Boltzmann particles. The predicted profiles were experimentally confirmed by two different neutron scattering experiments performed *after* the simulations.^{49,50} The agreement between simulation data and experimental results is remarkable, specially because the approximation used to derive the simulation data has not sound theoretical justification.

7.4 Water and Proton Transport in Water

The influence of quantum effects in the dynamic properties of water at $T=300$ K has been studied by CMD using an empirical potential to describe the interatomic interactions.⁵¹ Information on both collective and individual relaxation processes were derived from the study of several time correlation functions. The Debye dielectric relaxation time is the time constant associated to the exponential decay of the collective dipole moment correlation function. The rotational correlation time is associated to a single molecule time correlation function that depends on the molecule orientation with respect to a body-fixed reference frame. The correlation times derived by CMD are lower than the values obtained from a

classical simulation. This faster decay of the quantum correlation functions with respect to the classical ones, is also consistent with the larger value of the self-diffusion coefficient derived under inclusion of quantum effects. The simulation results for the time constants and the self-diffusion coefficient deviate significantly from the experimental values, probably due to deficiencies of the employed empirical potential. The wave numbers of the three intramolecular vibrational modes were derived from the power spectra of the velocity time autocorrelation function, showing a red shift of about 100 cm^{-1} with respect to the results of a classical simulation.⁵¹

Another CMD simulation has been presented using a different empirical potential for water.⁵² In this simulation, the effective classical potential, V_{ecp} was approximated by a simple Feynman-Hibbs approach.¹ Unfortunately, the use of different empirical potentials precludes the comparison of CMD results derived with and without the use of the Feynman-Hibbs approximation.^{52,51}

The proton transport in water has been studied by CMD using a two state⁵³ and a multistate empirical valence model (EVB).⁵⁴ Schmitt *et al.* undergo an analysis of the CMD trajectories in order to determine the rate constant for proton transport in water. This analysis is complicated because of the fluxional character of the excess proton, that can be associated to either the solvated Zundel (H_5O_2^+) and Eigen (H_9O_4^+) cations only as limiting ideal structures.⁵⁵ The rate constant was evaluated using a population autocorrelation function formalism, where selected many-body reaction coordinates were used to define different proton hopping pathways. The quantum rate was found to be two times faster compared to a classical treatment, and in good agreement to the experimental value.⁵⁴ The same simulation methodology was applied to study the kinetic H/D isotope effect in the proton transfer, resulting in an overestimation of this effect by about 25% with respect to the experimental value.⁵⁶

7.5 Anharmonic Molecule Vibrations

CMD can be applied to the determination of vibrational frequencies, by means of the calculation of the power spectra derived from either the velocity or position autocorrelation functions of the atomic nuclei. Calculations of this type has been presented for small molecules, as H_2 ,⁴¹ a linear chain of four water molecules,⁴² and $\text{Cl}^-(\text{H}_2\text{O})_n$ clusters.⁵⁷ While the position of the peaks in the power spectra are expected to be a realistic approximation of the vibrational frequencies, Marx. *et al.* pointed out that the width of the peaks in the power spectra are most likely an artifact inherent to the CMD approach.⁴¹ This conclusion in in agreement with the model study presented in Sec. 6. At temperatures where the molecule is in its vibrational ground state, the width of the CMD power spectra, is due to the anharmonicity of the effective classical. We have already shown that these anharmonic regions in the V_{ecp} lead to unphysical dynamic results. Therefore, one should be aware of the limitations of the CMD procedure to study temperature effects in the vibrational spectra. An alternative to the CMD method for the determination of vibrational frequencies has been presented recently.³³ This method is based on the diagonalization of the covariance tensor of the position centroid fluctuations, as determined by equilibrium (nondynamic) simulations. The capability of this approximation has been tested in the study of the tunneling frequency of a particle in a double-well potential, the vibrational frequencies of molecules (H_2 , C_2H_4 and HOCl), and the phonon frequencies of diamond.^{33,58}

7.6 Vibrational Energy Relaxation

Two different strategies have been employed for the CMD simulation of vibrational energy relaxation. The first, and more direct approach, is to run a CMD simulation with a prepared initial condition, so that the vibrational degree of freedom to be relaxed is initially excited through an internal potential energy boost. The energy dissipation is then monitored along the CMD run. This approach has been applied to study the relaxation rate of a CN^- ion in water.⁵⁹ The second approach is based on the use of a golden rule formula for the thermal rate, which requires the determination of the time autocorrelation function of the bath quantum force operator. This approach has not been yet applied to real systems, the main efforts aimed at the test of different approximations to simplify the calculation of the force autocorrelation function.⁶⁰⁻⁶²

8 Open Problems

In some sense, the CMD approach is not an approximation, but a set of different approximations under a common name. Thus, at temperature $T \rightarrow 0$, CMD is a wave packet dynamics, while at $T \rightarrow \infty$, CMD is Newton dynamics. The dynamic character of the CMD equations changes with temperature because the effective classical potential changes with temperature. The curves shown in Fig. 10 provide a striking example of this change: CMD results are shown at two temperatures, $T_a < T_b$. CMD at temperature T_a is a reasonable approximation to the exact result at T_b , while CMD at T_b deviates more from the exact result at T_b . The physical conditions where the effective classical potential provides meaningful dynamic information in vibrational and diffusive problems need to be further characterized.

A second important problem is that CMD is an approximation for the simulation of time correlation functions of a product of an arbitrary operator with a position or momentum operator. Several recipes for the calculation of time correlation functions of nonlinear operators has been the subject of theoretical investigations.^{19,63} However, from a practical point of view it is not clear which is the most convenient way to deal with nonlinear operators. In particular, it is important to clarify the calculation of the dynamic structure factor, $S(k, \omega)$, of liquids by CMD simulations.

Another interesting problem is the extension of CMD to the study of boson and fermions.^{64,65,69} Several groups are working on this extension following different approaches. Two different methods define so-called permutation potentials in order to mimic the effect of the indistinguishability of the particles.⁶⁶⁻⁶⁸ Thus, the Bose/Fermi system is mapped onto a pseudo-Boltzmann system where CMD is used to approximate Kubo transformed time correlation functions. The role of the permutation potential in the dynamic calculation requires careful study that leads to unexpected results. In this line, Kinugawa has shown that the time correlation function for position operators can be derived from the CMD trajectories in the pseudo-Boltzmann system, but not such relation appears for the momentum operator.⁶⁷ The latest formulation of CMD for Bose/Fermi systems shows the convenience of using an operator formalism without any reference to path integral techniques.⁷⁰

9 Conclusions

The rigorous formulation of CMD turns out to be a cumbersome task. It is much easier to explain how to perform a CMD simulation (see the Introduction) than to explain, in quantum mechanical terms, the physical meaning of this dynamic approach (see Sections 4 and 5). Even more difficult is to specify the cases where CMD leads to correct results for describing real time quantum dynamics of many-body systems at finite temperature. This latter point is surely an important issue for future research.

In this chapter we have focused on the Schrödinger formulation of CMD. Although for historical reasons the original CMD formulation was based on the path integral approach, it has become clear that the most sensible way to formulate this approximation is using an operator formalism in the Schrödinger formulation. However, one can not avoid the path integral formulation, because the only feasible way to perform many-body CMD simulations is via path integrals.

Proceeding along this line we have found a slightly different approach (RMD), that differs from CMD by the presence of two correction terms (one for the mass and the other for the force) that modify the CMD equations. Test calculations performed on simple one dimensional models show that these correction terms improve the dynamic description in some special cases (description of quantum tunneling in the zero temperature limit and calculation of quantum transmission rates), but their influence is negligible in most cases. For this reason, we do not believe that the RMD equations represent a practical alternative to improve CMD.

One conclusion of our work is that it is not realistic to expect that a modification of the CMD equations (as RMD does) will lead to a significant improvement of the dynamic description. The reason is that the main source of error in CMD is not the dynamic equations used to propagate the centroid coordinates. The main limitation of CMD is a consequence of being *a constrained dynamic approximation*, i.e., the dynamic states accessible along the time evolution are severely limited. These states corresponds to fixed centroid path integrals and we do not see any feasible way to avoid the limitation of using fixed centroid path integrals in the constrained dynamics.

Acknowledgments

This work was supported by CICYT (Spain) under contract BFM2000-1318, and by DGESIC through Project No. 1FD97-1358.

References

1. R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
2. D.M. Ceperley, *Path integral theory and methods for 4He* , Rev. Mod. Phys. **67**, 279 (1995).
3. C. Chakravarty, *Path integral simulations of atomic and molecular systems*, Inter. Rev. Phys. Chem. **16**, 421 (1997).

4. M.J. Gillan, *The path-integral simulation of quantum systems*, in *Computer Modelling of Fluids, Polymers, and Solids*, edited by C.R.A. Catlow, S.C. Parker, and M.P. Allen (Kluwer, Dordrecht, 1990).
5. B.J. Berne and D. Thirumalai, *On the simulation of quantum systems: path integral methods*, Ann. Rev. Phys. Chem. **37**, 401 (1987).
6. R. Giachetti and V. Tognetti, *Quantum corrections to the thermodynamics of nonlinear systems*, Phys. Rev. B**33**, 7647 (1986).
7. R.P. Feynman and H. Kleinert, *Effective classical partition functions*, Phys. Rev. A**34**, 5080 (1986).
8. H. Kleinert, *Path Integrals* (World Scientific, Singapore, 1995).
9. A. Cuccoli, R. Giachetti, V. Tognetti, R. Vaia, and P. Verrucchi, *The effective potential and effective Hamiltonian in quantum statistical mechanics*, J. Phys.: Condens. Matter **7**, 7891 (1995).
10. M. J. Gillan, *Quantum simulation of hydrogen in metals* Phys. Rev. Lett. **58**, 563 (1987).
11. M. J. Gillan, *Quantum-classical crossover of the transition rate in the damped double well*, J. Phys. C: Solid State Phys. **20**, 3621 (1987).
12. M. J. Gillan, *The quantum simulation of hydrogen in metals*, Phil. Mag. A **58**, 257 (1988).
13. G.A. Voth, *Path-integral centroid methods in quantum statistical mechanics and dynamics*, in *Advances in Chemical Physics*, Vol. XCIII, edited by I. Prigogine and S.A. Rice (John Wiley & Sons, New York, 1996).
14. G.A. Voth, *Feynman path integral formulation of quantum mechanical transition-state theory*, J. Phys. Chem. **97**, 8365 (1993).
15. G.A. Voth, D. Chandler, and W.H. Miller, *Rigorous formulation of quantum transition state theory and its dynamical corrections*, J. Chem. Phys. **91**, 7749 (1989).
16. J. Cao and G.A. Voth, *A new perspective on quantum time correlation functions*, J. Chem. Phys. **99**, 10070 (1993).
17. J. Cao and G.A. Voth, *The formulation of quantum statistical mechanics based on the Feynman path centroid density. I. Equilibrium properties*, J. Chem. Phys. **100**, 5093 (1994).
18. J. Cao and G.A. Voth, *The formulation of quantum statistical mechanics based on the Feynman path centroid density. II. Dynamical properties*, J. Chem. Phys. **100**, 5106 (1994).
19. J. Cao and G.A. Voth, *The formulation of quantum statistical mechanics based on the Feynman path centroid density. III. Phase space formalism and analysis of centroid molecular dynamics*, J. Chem. Phys. **101**, 6157 (1994).
20. J. Cao and G.A. Voth, *The formulation of quantum statistical mechanics based on the Feynman path centroid density. IV. Algorithms for centroid molecular dynamics*, J. Chem. Phys. **101**, 6168 (1994).
21. S. Jang and G.A. Voth, *Path integral centroid variables and the formulation of their exact real time dynamics*, J. Chem. Phys. **111**, 2357 (1999).
22. S. Jang and G.A. Voth, *A derivation of centroid molecular dynamics and other approximate time evolution methods for path integral centroid variables*, J. Chem. Phys. **111**, 2371 (1999).
23. G.A. Voth, *Feynman path centroid methods for condensed phase quantum dynamics*,

in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B.J. Berne, G. Ciccoti, and D.F. Coker (World Scientific, Singapore, 1998).

24. R. Ramírez and T. López-Ciudad, *Phase-space formulation of thermodynamic and dynamical properties of quantum particles*, Phys. Rev. Lett. **83**, 4456 (1999).
25. M.S. Swanson, *Path Integrals and Quantum Processes* (Academic Press, San Diego, 1992).
26. B. van der Pol and H. Bremmer, *Operational Calculus based on the two-sided Laplace Integral* (Chelsea, New York, 1995).
27. R.P. Feynman, *Statistical Mechanics* (Addison-Wesley, Redwood City, 1972).
28. T. López-Ciudad and R. Ramírez, *Spectral decomposition and Bloch equation of the operators represented by fixed-centroid path integrals*, J. Chem. Phys. **113**, 10849 (2000).
29. R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II* (Springer-Verlag, Berlin, 1985).
30. R. Zwanzig, *Time-correlation functions and transport coefficients in statistical mechanics*, Ann. Rev. Phys. Chem. **16**, 67 (1969).
31. R. Ramírez, T. López-Ciudad, and J.C. Noya, *Feynman effective classical potential in the Schrödinger formulation*, Phys. Rev. Lett. **81**, 3303 (1998).
32. R. Ramírez and T. López-Ciudad, *The Schrödinger formulation of the Feynman path centroid density*, J. Chem. Phys. **111**, 3339 (1999).
33. R. Ramírez and T. López-Ciudad, *Low lying vibrational excitation energies from equilibrium path integral simulations*, J. Chem. Phys. **115**, 103 (2001).
34. G. Krilov and B.J. Berne, *Real time quantum correlation functions. I. Centroid molecular dynamics of anharmonic systems*, J. Chem. Phys. **111**, 9140 (1999).
35. J. Cao and G.A. Voth, *A unified framework for quantum activated rate processes. I. General theory*, J. Chem. Phys. **105**, 6856 (1996).
36. R. Ramírez, *Dynamics of quantum particles by path-integral centroid simulations: The symmetric Eckart barrier*, J. Chem. Phys. **107**, 3550 (1997).
37. G.J. Martyna, *Adiabatic path integral molecular dynamics methods. I. Theory*, J. Chem. Phys. **104**, 2018 (1996).
38. J. Cao and G.J. Martyna, *Adiabatic path integral molecular dynamics methods. II. Algorithms*, J. Chem. Phys. **104**, 2028 (1996).
39. G.J. Martyna, M.L. Klein, and M. Tuckerman, *Nosé-Hoover chains: The canonical ensemble via continuous dynamics*, J. Chem. Phys. **97**, 2635 (1992).
40. M.E. Tuckerman, D. Marx, M.L. Klein, and M. Parrinello, *Efficient and general algorithms for path integral Car-Parrinello molecular dynamics*, J. Chem. Phys. **104**, 5579 (1996).
41. D. Marx, M.E. Tuckerman, and G.J. Martyna, *Quantum Dynamics via adiabatic ab initio centroid molecular dynamics*, Computer Phys. Comm. **118**, 166 (1999).
42. M. Pavese, D.R. Berard, and G.A. Voth, *Ab initio centroid molecular dynamics: a fully quantum method for condensed-phase dynamics simulations*, Chem. Phys. Lett. **300**, 93 (1999).
43. A. Calhoun, M. Pavese, and G.A. Voth, *Hyper-parallel algorithms for centroid molecular dynamics: application to liquid para-hydrogen*, Chem. Phys. Lett. **262**, 415 (1996).
44. S. Miura, S. Okazaki, and K. Kinugawa, *A path integral centroid molecular dynamics*

- study of nonsuperfluid liquid helium-4*, J. Chem. Phys. **110**, 4523 (1999).
45. M. Pavese and G.A. Voth, *Pseudopotentials for centroid molecular dynamics. Application to self-diffusion in liquid para-hydrogen*, Chem. Phys. Lett. **249**, 231 (1996).
 46. K. Kinugawa, P.B. Moore, and M.L. Klein, *Centroid path integral molecular dynamics simulation of lithium para-hydrogen clusters*, J. Chem. Phys. **106**, 1154 (1997).
 47. K. Kinugawa, P.B. Moore, and M.L. Klein, *Centroid path integral molecular-dynamics studies of a para-hydrogen slab containing a lithium impurity*, J. Chem. Phys. **109**, 610 (1998).
 48. K. Kinugawa, *Path integral centroid molecular dynamics study of the dynamic structure factors of liquid para-hydrogen*, Chem. Phys. Lett. **292**, 454 (1998).
 49. F.J. Bermejo, K. Kinugawa, C. Cabrillo, S.M. Bennington, B.Fåk, M.T. Fernández-Díaz, P. Verkerk, J. Dawidowski, and R. Fernández-Perea, *Quantum effects on liquid dynamics as evidenced by the presence of well-defined collective excitations in liquid para-hydrogen*, Phys. Rev. Lett. **84**, 5359 (2000).
 50. M. Zoppi, D. Colognesi, and M. Celli, *Microscopic dynamics of liquid hydrogen*, Europhys. Lett. **53**, 34 (2001).
 51. J. Lobaugh and G. A. Voth, *A quantum model for water: Equilibrium and dynamical properties*, J. Chem. Phys. **106**, 2400 (1997).
 52. B. Guillot and Y. Guissani, *Quantum effects in simulated water by the Feynman-Hibbs approach*, J. Chem. Phys. **108**, 10162 (1998).
 53. J. Lobaugh and G. A. Voth, *The quantum dynamics of an excess proton in water*, J. Chem. Phys. **104**, 2056 (1996).
 54. U.W. Schmitt and G.A. Voth, *The computer simulation of proton transport in water*, J. Chem. Phys. **111**, 9361 (1999).
 55. D. Marx, M.E. Tuckerman, J. Hutter, and M. Parrinello, *The nature of hydrated excess proton in water*, Nature, **397**, 601 (1999).
 56. U.W. Schmitt and G.A. Voth, *The isotope substitution effect on the hydrated proton*, Chem. Phys. Lett. **329**, 36 (2000).
 57. G.K. Schenter, B. C. Garrett, and G.A. Voth, *The quantum vibrational structure of $\text{Cl}^-(\text{H}_2\text{O})_n$ clusters*, J. Chem. Phys. **113**, 5171 (2000).
 58. R. Ramírez, J. Schulte, and M.C. Böhm, *Ground state and excited state properties of ethylene isomers studied by a combined Feynman path integral-ab initio approach*, Mol. Phys. **99**, 1249 (2001).
 59. S. Jang, Y. Pak, and G.A. Voth, *Quantum dynamical simulation of the energy relaxation rate of the CN^- ion in water*, J. Phys. Chem. A **103**, 10289 (1999).
 60. J. Poulsen, S. Keiding, and P.J. Rossky, *Extracting rates of vibrational energy relaxation from centroid molecular dynamics*, Chem. Phys. Lett. **336**, 448 (2001).
 61. J.A. Poulsen and P.J. Rossky, *An ansatz-based variational path integral centroid approach to vibrational energy relaxation in simple liquids*, J. Chem. Phys. **115**, 8014 (2001).
 62. J.A. Poulsen and P.J. Rossky, *Path integral centroid molecular-dynamics evaluation of vibrational energy relaxation in condensed phase*, J. Chem. Phys. **115**, 8024 (2001).
 63. D.R. Reichman, P.-N. Roy, S. Jang, and G.A. Voth, *A Feynman path centroid dynamics approach for the computation of time correlation functions involving nonlinear operators*, J. Chem. Phys. **113**, 919 (2000).
 64. P.-N. Roy and G.A. Voth, *On the Feynman path centroid density for Bose-Einstein*

and Fermi-Dirac statistics, J. Chem. Phys. **110**, 3647 (1999).

65. P.-N. Roy, S. Jang, and G.A. Voth, *Feynman path centroid dynamics for Fermi-Dirac statistics*, J. Chem. Phys. **111**, 5303 (1999).
66. K. Kinugawa, H. Nagao, and K. Ohta, *Path integral centroid molecular dynamics method for Bose and Fermi statistics: formalism and simulation*, Chem. Phys. Lett. **307**, 187 (1999).
67. K. Kinugawa, H. Nagao, and K. Ohta, *A semiclassical approach to the dynamics on many-body Bose/Fermi systems by the path integral centroid molecular dynamics*, J. Chem. Phys. **114**, 1454 (2001).
68. S. Miura and S. Okazaki, *Path integral molecular dynamics method based on a pair density matrix approximation: An algorithm for distinguishable and identical particle systems*, J. Chem. Phys. **115**, 5353 (2001).
69. N.V. Blinov, P.-N. Roy and G.A. Voth, *Path integral formulation of centroid dynamics for systems obeying Bose-Einstein statistics*, J. Chem. Phys. **115**, 4484 (2001).
70. N.V. Blinov and P.-N. Roy *Operator formulation of centroid dynamics for Bose-Einstein and Fermi-Dirac statistics*, J. Chem. Phys. **115**, 7822 (2001).

Quantum Molecular Dynamics with Wave Packets

Uwe Manthe

Theoretische Chemie, Technische Universität München
85747 Garching, Germany
E-mail: manthe@ch.tum.de

Quantum effects are prominent in the dynamics of many molecular systems. Simulating quantum molecular dynamics, the wave packet approach is an efficient tool to solve time-dependent and time-independent Schrödinger equations. The article reviews standard methods employed in wave packet calculations: different type of grid representations for wave functions and propagation schemes for solving the time-dependent Schrödinger equation are described. Iterative diagonalization schemes and filter diagonalization approaches for the time-independent Schrödinger equation are also discussed within the framework of the wave packet dynamics approach. Following the review of the standard methods for wave packet dynamics, an efficient approach for the description of larger systems, the multi-configuration time-dependent Hartree (MCTDH) approach, is presented.

1 Introduction

Quantum mechanics is not only essential for the understanding of the electronic structure of molecules, quantum effects also strongly influence the nuclear motion of many molecular systems. Tunneling is a key issue in the understanding of hydrogen or proton transfer reactions. Vibronic coupling determines the outcome of many photochemical reactions. Zero point energy effects are important for the structure and dynamics of van der Waals-clusters. Many other examples could be found.

In the absence of strong laser fields, molecular system can typically be described by time-independent Hamiltonians. Thus, the system dynamics can be studied by solving the time-independent Schrödinger equation. The solution of a linear eigenvalue problem might therefore be viewed as the most direct approach for studying quantum dynamics of molecular systems: the Hamiltonian is represented in a finite basis set and the resulting matrix is numerically diagonalized. However, two problems associated with this approach should be mentioned. The CPU time required for the (complete) matrix diagonalization is proportional to the cube of the basis size and the required memory is proportional to the square of the basis size. Since the basis set size scales exponentially with the dimensionality of the system, the computation becomes easily infeasible if the system size increases. Second, the interpretation of the numerical results is a formidable task if larger systems are considered. The number of relevant eigenstates is enormous and the spectra typically can not be assigned using simple and physically meaningful patterns.

The wave packet approach, which has become increasingly popular in the last two decades, can reduce both problems. First, the motion of wave packets obtained from the solution of the time-dependent Schrödinger equation can typically be understood using classical-mechanical and semiclassical ideas. This can considerably simplify the interpretation of the numerical results. Second, while in the above diagonalization approach the dynamics of all states are computed at once, the wave packet approach typically describes only the motion of individual wave packets. The initial conditions defining these wave

packets are tailored to the specific experiment in question. This reduction of the required information yields a numerically more efficient scheme. In wave packet dynamics calculations, CPU time and memory requirements scale approximately proportional to the basis set size. It should be noted that these numerical advantages are not limited to time-dependent wave packet calculations. Analogous arguments are valid also for the calculation of individual energy eigenstates by iterative diagonalization or filter diagonalization approaches.

The present article reviews the methods employed in modern wave packet dynamics calculations. Standard schemes for an efficient spatial representation of wavefunctions are described: the Fast Fourier Transform (FFT) approach¹ and the Discrete Variable Representation (DVR).²⁻⁴ Concepts for the efficient temporal integration of the time-dependent Schrödinger equation are discussed and examples of widely used integrators are given.⁵⁻⁷ The connection between these propagation schemes and iterative diagonalization techniques is discussed. The description of the filter diagonalization technique⁸ finally highlights the connection between time-dependent and energy-dependent methods.

Due to the numerical effort, presently standard wave packet calculations are not feasible for systems with more than six dimensions. Even four atom systems can studied in their full dimensionality only under favorable circumstances. Therefore also a scheme which is tailored to the description of multi-dimensional systems will be presented: the multi-configurational time-dependent Hartree (MCTDH) approach.^{9,10} The MCTDH approach employs a two layer scheme for the representation of the wavefunction. The multi-dimensional wavefunction is represented in a time-dependent basis set. The time-dependent basis functions employed are products of one-dimensional wave packets represented in a standard time-independent (FFT or DVR) basis. The MCTDH approach can give an accurate description of multi-dimensional systems which are beyond the scope of standard wave packet calculations. Recent applications include a 24-dimensional calculation on the absorption spectrum of pyrazine^{11,12} and a 12-dimensional investigation of the $H + CH_4 \rightarrow H_2 + CH_3$ reaction.^{14,15}

2 Spatial Representation of Wavefunctions

The dynamics of a wave packet is given by the Schrödinger equation

$$i\frac{\partial}{\partial t}\psi(x_1, \dots, x_f, t) = \hat{H}\psi(x_1, \dots, x_f, t) \quad (1)$$

(atomic units, i.e. $\hbar=1$, are used).

Representing the wavefunction in a finite time-independent basis set,

$$\psi(x_1, \dots, x_f, t) = \sum_n A_n(t)\phi_n(x_1, \dots, x_f), \quad (2)$$

the equations of motions for the time-dependent expansion coefficient are a linear system of first order differential equations:

$$i\frac{\partial}{\partial t}A_n(t) = \sum_m H_{nm}A_m(t). \quad (3)$$

Thus, the solution of the time-dependent Schrödinger equation can be decomposed into two different tasks. The first task is the computation of the matrix elements H_{nm} or, alternatively, the calculation of the action of the Hamiltonian operator on the wavefunction. Second, the resulting differential equation have to be integrated in time. The present section focuses on the first task, while the time propagation will be discussed in the next section.

2.1 Grid and Basis Representations

Wave packet calculations utilize different representations of the wavefunction to evaluate different terms in the Hamiltonian. Consider a Hamiltonian

$$\hat{H} = \hat{H}_0 + V. \quad (4)$$

If the wavefunction represented in the basis $|\phi_n\rangle$ of eigenstates of \hat{H}_0 with the corresponding eigenvalues E_n , the action of \hat{H}_0 on the wavefunction can be evaluated immediately,

$$|\psi\rangle = \sum_n c_n |\phi_n\rangle, \quad (5)$$

$$\langle \phi_n | \hat{H}_0 | \psi \rangle = E_n c_n. \quad (6)$$

Analogously a discrete grid representation based on the grid point states $|X_n\rangle$ can be employed to evaluate the action of the potential on the wavefunction:

$$|\psi\rangle = \sum_n k_n |X_n\rangle, \quad (7)$$

$$\langle X_n | V | \psi \rangle = V(X_n) k_n. \quad (8)$$

If the transformation between the basis and the grid representation, i.e. the matrix $\langle X_n | \phi_m \rangle$, is known, the action of the Hamiltonian can be evaluated by transforming from one representation to the other:

$$\hat{H} = \sum_n |\phi_n\rangle E_n \langle \phi_n| + \sum_n |X_n\rangle V(X_n) \langle X_n|, \quad (9)$$

$$\langle \phi_n | \hat{H} | \psi \rangle = E_n c_n + \sum_m \langle \phi_n | X_m \rangle V(X_m) \left(\sum_j \langle X_m | \phi_j \rangle c_j \right), \quad (10)$$

$$\langle X_n | \hat{H} | \psi \rangle = \sum_m \langle X_n | \phi_m \rangle E_m \left(\sum_j \langle \phi_m | X_j \rangle k_j \right) + V(X_n) k_n. \quad (11)$$

To define this dual representation scheme, the two bases $|\phi_n\rangle$ and $|X_n\rangle$ have to be chosen and the transformation between the two set, $\langle X_n | \phi_m \rangle$, has to be specified. Two different schemes are frequently used for this propose: the discrete variable representation (DVR)⁴ and the fast Fourier transform (FFT)¹ approach. These schemes will be discussed below in more detail.

The above discussion has not explicitly considered the dimensionality of the wavefunction. The transformation between the two bases requires a matrix multiplication. In

principle, the numerical effort of this operation would be proportional to the square of the basis set size. In multi-dimensional calculations the numerical effort can be drastically reduced if the basis sets and grids are direct products of one-dimensional functions:

$$\phi_n(x_1, x_2, \dots, x_f) = \phi_{n_1}^{(1)}(x_1) \cdot \phi_{n_2}^{(2)}(x_2) \cdot \dots \cdot \phi_{n_f}^{(f)}(x_f), \quad (12)$$

$$|X_n\rangle = |(X_1)_{n_1}\rangle |(X_2)_{n_2}\rangle \dots |(X_f)_{n_f}\rangle, \quad (13)$$

where the index n should be read as multi-index n_1, n_2, \dots, n_f and the multi-dimensional coordinate X as X_1, X_2, \dots, X_f . Then the transformation matrix $\langle X_n | \phi_m \rangle$ factorizes:

$$\langle X_n | \phi_m \rangle = \langle (X_1)_{n_1} | \phi_{m_1}^{(1)} \rangle \langle (X_2)_{n_2} | \phi_{m_2}^{(2)} \rangle \dots \langle (X_f)_{n_f} | \phi_{m_f}^{(f)} \rangle. \quad (14)$$

The transformation from the basis grid to the grid representation and vice versa can be calculated for each coordinate separately:

$$\begin{aligned} k_{n_1, n_2, \dots, n_f} &= \langle X_{n_1, n_2, \dots, n_f} | \psi \rangle = \\ &\left(\sum_{m_1} \langle (X_1)_{n_1} | \phi_{m_1}^{(1)} \rangle \left(\sum_{m_2} \langle (X_2)_{n_2} | \phi_{m_2}^{(2)} \rangle \dots \right. \right. \\ &\left. \left. \dots \left(\sum_{m_f} \langle (X_f)_{n_f} | \phi_{m_f}^{(f)} \rangle c_{m_1, m_2, \dots, m_f} \right) \dots \right) \right). \end{aligned} \quad (15)$$

If N_i basis functions or grid points are employed in the i -th coordinate, then the numerical effort of this operation is proportional to $(N_1 + N_2 + \dots + N_f) \prod_{i=1}^f N_i$. Thus, in a multi-dimensional direct product basis the numerical effort of the transformation scales approximately linear with the basis set size $N = \prod_{i=1}^f N_i$.

2.2 DVR

In the discrete variable representation (DVR), the optimally localized grid point states $|X_n\rangle$ are obtained from the eigenstates of the coordinate operator represented in the given finite basis $|\phi_n\rangle$:²

$$\langle \phi_n | x | \phi_m \rangle = \sum_j \langle \phi_n | X_j \rangle X_j \langle X_j | \phi_m \rangle. \quad (16)$$

The eigenvalue X_j are grid points of the one-dimensional grid and the eigenstates $\langle \phi_n | X_j \rangle$ are the grid-to-basis transformation matrix employed in the DVR scheme. Considering all bases $|\chi_n\rangle$ which can be obtained by a unitary transformation of the original $|\phi_n\rangle$ basis, the $|X_n\rangle$ basis minimizes the localization criterion

$$\sum_n (\langle \chi_n | x^2 | \chi_n \rangle - \langle \chi_n | x | \chi_n \rangle^2) \rightarrow \text{minimum}. \quad (17)$$

The evaluation of the potential energy integrals

$$\langle \phi_n | V(x) | \phi_m \rangle = \sum_j \langle \phi_n | X_j \rangle V(X_j) \langle X_j | \phi_m \rangle \quad (18)$$

within the DVR corresponds to a Gaussian quadrature if the basis $|\phi_n\rangle$ consists of orthogonal polynomials multiplied by a weight function.³ Since the wavefunction is typically

more structured than the potential energy function, the numerical inaccuracy resulting from this Gaussian quadrature scheme is usually small compared to the truncation error resulting from the representation of the wavefunction in the finite basis $|\phi_n\rangle$. Thus, the number of grid points can be equal to the number of basis functions without relevant loss of accuracy.⁴

Typical basis sets employed in DVR schemes are the harmonic oscillator eigenfunctions for distance coordinates (Hermite DVR) and Legendre polynomials for angular variables (Legendre DVR). Employing the eigenstates of H_0 operator specifically adjusted to the system under investigation is another interesting possibility (often called potential optimized DVR). Thus, the DVR approach offers maximal freedom for tailoring the basis sets and grids to any specific system.

If the grid representation is employed as a primary representation, explicit transformations to the basis representation can be avoided.⁴ The kinetic energy operator is employed in its grid representation:

$$\hat{T} = \hat{H}_0 + V_0 , \quad (19)$$

$$\langle X_n | \hat{T} | X_m \rangle = \sum_j \langle X_n | \phi_j \rangle E_j \langle \phi_j | X_m \rangle - V_0(X_n) \delta_{nm} . \quad (20)$$

Since usual kinetic energy operators show a simple structure, the application of the total kinetic energy operator can be performed by subsequent application of the different one-dimensional kinetic energy matrices of the respective coordinates. Thus, two matrix multiplies for the forward and backward basis transformations can be replaced by a single matrix multiply with the kinetic energy matrix in coordinate representation. Moreover, these DVR schemes are no longer restricted to direct product type grid. Unnecessary grid points can be dropped in the representation of the wavefunction and the kinetic energy operator^{4,16} which reduces the basis set size.

2.3 FFT

While the DVR approach provides a flexible choice of basis sets and grids, the fast Fourier transform (FFT) scheme¹ focuses on the numerical efficiency of the transformation between the different basis sets employed. The FFT approach employs an evenly spaced grid in a given interval $[x_i, x_f]$ of the coordinate space. The N grid points in this interval are connected via a discrete Fourier transform with an evenly spaced momentum grid in the interval $[p_i, p_i + \frac{2\pi}{\Delta x}]$, where $\Delta x = \frac{x_f - x_i}{N}$ is the grid spacing in coordinate space. The transformation between the discrete coordinate representation $|X_n\rangle$ and the discrete momentum representation $|P_m\rangle$ is given by:

$$|X_n\rangle = \frac{1}{\sqrt{N}} \sum_{m=1}^N e^{i P_m X_n} |P_m\rangle , \quad (21)$$

$$|P_m\rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^N e^{-i P_m X_n} |X_n\rangle . \quad (22)$$

Fast Fourier transform algorithms can be employed to compute these transformation. They provide a numerical efficient scheme if N can be split into many prime factors. N being a power of two is the most favorable case. Then the numerical effort of the transformation is proportional to $\log_2 N \cdot N$.

Connecting to the description of section 2.1, the plane waves $\phi_n(x) = \frac{1}{\sqrt{N}}e^{iP_n x}$ can be viewed as the basis functions employed in the FFT approach. The momentum values P_n are evenly spaced in the interval $[p_i, p_i + \frac{2\pi}{\Delta x}]$ and the grid states are placed on evenly spaced points in the interval $[x_i, x_f]$. The transformation matrix between this basis and the grid representation reads:

$$\langle X_n | P_m \rangle = \frac{1}{\sqrt{N}} e^{i P_m X_n}. \quad (23)$$

Any operator diagonal in the momentum representation, e.g. $\frac{1}{2m} \frac{\partial^2}{\partial x^2}$, can be used as H_0 .

Due to the Fourier representation employed, the resulting wavefunction is $(x_f - x_i)$ -periodic in coordinate space and $\frac{2\pi}{\Delta x}$ -periodic in momentum space. Thus, a converged description of the wavefunction is obtained if it vanishes outside the interval $[x_i, x_f]$ in coordinate space and outside the interval $[p_i, p_i + \frac{2\pi}{\Delta x}]$ in momentum space or obeys the corresponding periodicity requirements.

Also non-evenly spaced grids can be employed in the FFT scheme. To this end, a coordinate transformation $\tilde{x} = f(x)$ is made in the Hamiltonian. The evenly spaced grid of \tilde{x} points then corresponds to an non-evenly spaced grid in the original coordinate x (mapped FFT¹⁷).

3 Propagation of Wave Packets

Solving Eq.(3) requires to evaluate the action of the Hamiltonian on the wavefunction, which has been discussed in the preceding section, and to integrate the set of linear differential equations. Of course, any general purpose integration scheme, e.g. Runge-Kutta or predictor-corrector algorithms, could be used. However, integration schemes specifically developed for this particular type of equations are considerably more efficient. In the following, different integrations schemes widely used in modern computations will be described and compared.

3.1 Split Operator Propagation

The split operator scheme⁵ explicitly utilizes the dual representation approach. The Hamiltonian is splitted into two parts and each is represented in its eigenstate representation (see Sect.2.1).

$$\hat{H} = \hat{H}_0 + V = \sum_n |\phi_n\rangle E_n \langle \phi_n| + \sum_n |X_n\rangle V(X_n) \langle X_n| \quad (24)$$

Employing a Trotter formula

$$e^{-i\hat{H}\Delta t} = e^{-i\hat{V}\Delta t/2} e^{-i\hat{H}_0\Delta t} e^{-i\hat{V}\Delta t/2} + O(\Delta t^3) \quad (25)$$

and evaluating each resulting propagator in its eigenstate representation, a second order short time integrator can be constructed:

$$\begin{aligned} e^{-i\hat{H}\Delta t} &= \left(\sum_n |X_n\rangle e^{-iV(X_n)\Delta t/2} \langle X_n| \right) \cdot \left(\sum_n |\phi_n\rangle e^{-iE_n\Delta t} \langle \phi_n| \right) \cdot \\ &\quad \left(\sum_n |X_n\rangle e^{-iV(X_n)\Delta t/2} \langle X_n| \right) + O(\Delta t^3). \end{aligned} \quad (26)$$

Thus, the operator $\exp(-i\hat{V}\Delta t/2)$ is evaluated in the coordinate representation, where it is diagonal. Then the wavefunction is transformed from the coordinate to the basis representation. The operator $\exp(-i\hat{H}_0\Delta t)$, which is diagonal in this representation, is now applied. After a change back to coordinate representation and application of $\exp(-i\hat{V}\Delta t/2)$, the result of a Δt integration step is obtained. Repeating these steps, the integration can be continued for any required period of time:

$$e^{-i\hat{H}t} = \prod_{n=1}^{\frac{t}{\Delta t}} \left(e^{-i\hat{V}\Delta t/2} e^{-i\hat{H}_0\Delta t} e^{-i\hat{V}\Delta t/2} \right) + 0(\Delta t^2). \quad (27)$$

The integration scheme is unitary and strictly conserves the norm of the wavefunction. Analyzing the numerical propagator in its eigenstate representation,

$$e^{-i\hat{V}\Delta t/2} e^{-i\hat{H}_0\Delta t} e^{-i\hat{V}\Delta t/2} \chi_n = u_n \chi_n, \quad (28)$$

one finds that the discretisation on the finite time step Δt effects only the phase of u_n (since $|u_n| = 1$). This guarantees the long time stability of the integration scheme. However, it is only a low order scheme. Thus, the results tend not to be particularly accurate. High precision results can only be obtained with a prohibitively small time step Δt .

The above discussion has been limited to a specific implementation of the split operator scheme for a simple Hamiltonian. It aimed only on presenting the basic idea. The scheme has been applied to several different type of Hamiltonians and used in different variants.

3.2 Polynomial Expansions

Higher order integration schemes are based on a polynomial expansion of the propagator:

$$e^{-i\hat{H}\Delta t} = \sum_{n=0}^N c_n(\Delta t) \hat{H}^n + 0(\Delta t^{N+1}). \quad (29)$$

The action of the propagator on the wavefunction is computed by successive application of the Hamiltonian,

$$\psi_n = a_n \hat{H} \psi_{n-1} + \sum_{j=0}^{n-1} b_{n,j} \psi_j, \quad (30)$$

where a_n and b_j are coefficients which characterize the particular scheme. Employing orthogonal polynomials, the above recursion relation usually reduces to a three term series: $b_{n,j} = 0$ for $j < n - 2$.

A short polynomial expansion (typically with $N \leq 10$) provides efficient propagators for limited time steps Δt . These integration steps are repeated until the propagation is completed:

$$\psi(M\Delta t) = \left(\prod_{m=1}^M e^{-i\hat{H}\Delta t} \right) \psi(0) + 0(\Delta t^N). \quad (31)$$

The Lanczos scheme is mostly employed to define the coefficients in the polynomial expansion.⁷ In the Lanczos scheme, the Hamiltonian matrix is represented in the Krylov

space $\{\psi, \hat{H}\psi, \dots, \hat{H}^N\psi\}$. The resulting $(N+1)$ -dimensional model Hamiltonian H_{model} is diagonalized,

$$H_{model} \chi_j = E_{model,j} \chi_j , \quad (32)$$

and the propagator is evaluated in the model space:

$$e^{-i\hat{H}\Delta t}\psi = \sum_{j=1}^{N+1} \chi_j \cdot e^{-iE_{model,j}\Delta t} \cdot \langle \chi_j | \psi \rangle + O(\Delta t^{N+1}) . \quad (33)$$

The resulting propagation scheme is numerically unitary. Thus, it provides the same long time stability as the split operator propagation. Its main advantage is the increased accuracy since the error is of higher order: $O(\Delta t^{N+1})$. Practical experience indicates that N should be chosen between 6 and 10 in most applications. The required length of integration step Δt is closely related to the spectrum of the Hamiltonian. Typically $\Delta t/N$ roughly equals the inverse spectral range of Hamiltonian.

Alternatively, the full propagation can be done employing a single polynomial expansion.⁶ Then very high expansion orders N are required. Chebychev polynomials provide an expansion which is stable even for very high orders. Employing Chebychev polynomials T_n of the normalized Hamiltonian

$$\hat{H}_{norm} = \frac{\hat{H} - \bar{E}}{\Delta E} \quad (34)$$

and corresponding n-th order wavefunctions $\psi_n = T_n(\hat{H}_{norm})\psi(0)$ generated by the recursion relation

$$\psi_n = 2\hat{H}_{norm}\psi_{n-1} - \psi_{n-2} , \quad (35)$$

the wavefunction $\psi(t)$ is given as

$$\psi(t) = \sum_{n=0}^N a_n(t)\psi_n . \quad (36)$$

with

$$a_n(t) = (2 - \delta_{n0})e^{-i\bar{E}t}(-i)^n J_n(\Delta Et) . \quad (37)$$

The spectrum of the normalized Hamiltonian should be in the interval $[-1, 1]$. The J_n denote the Bessel functions. The above series converges exponentially if the order N exceeds ΔEt . Thus, the number of Hamiltonian multiplies is directly connected to the spectra range of the Hamiltonian \hat{H} . Due to the very high order of the scheme, the Chebychev method can efficiently produce extremely accurate results. However, extracting information at intermediate times, e.g. the evolution of expectation values with time, is not straightforward within the Chebychev method. The wavefunction would have to be calculated for all desired times simultaneously causing a significant storage problem.

4 Iterative Diagonalization

Considering time-independent Hamiltonians, time-dependent and energy-dependent representation are equivalent. However, direct matrix diagonalization and wave packet propagation schemes employ different computational strategies to solve the Schrödinger equation. In contrast, iterative matrix diagonalization approaches are closely related to wavepacket propagation. To highlight this connection, two examples will be discussed in the following.

The Lanczos scheme is frequently used to directly compute spectra in the energy domain. The first applications to molecular spectra¹⁸⁻²⁰ preceded the development of the integration schemes discussed above. While in the short iterative Lanczos scheme for time propagation only a low order expansion is employed, these computations use high order expansions. Since the Lanczos recursion relation is a three term series, the resulting Hamiltonian matrix is tridiagonal. This tridiagonal matrix can be diagonalized numerically efficient. As a result, the eigenvalues of H_{model} and the overlap of the eigenvalues with the initial vector ψ is obtained:

$$H_{model} \chi_j = E_{model,j} \chi_j , \quad (38)$$

$$\sigma(E) = \sum_j \delta(E_{model,j} - E) \cdot | < \chi_j | \psi > |^2 . \quad (39)$$

The envelop of the absorption spectrum $\sigma(E)$ converges with the increasing Lanczos order. However, due to the numerical instability of the Lanczos recursion at high orders, only convoluted spectra (and not individual eigenvalues) can be converged easily. Computing the spectrum via the autocorrelation function $< \psi | \exp(-i\hat{H}_{model}t) | \psi >$, the connection between the time and energy-dependent approach becomes obvious:

$$\sigma(E) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt < \psi | e^{-i\hat{H}_{model}t} | \psi > e^{iEt} , \quad (40)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \sum_j < \psi | \chi_j > e^{-iE_{model,j}t} < \chi_j | \psi > e^{iEt} , \quad (41)$$

$$= \sum_j < \psi | \chi_j > \delta(E_{model,j} - E) < \chi_j | \psi > . \quad (42)$$

Due to accumulating roundoff errors, high order Lanczos expansions are numerically unstable. In contrast, Chebychev polynomials facilitate arbitrarily accurate expansions at all orders. Thus, virtually any energy-dependent quantity can be directly expanded in Chebychev polynomials:

$$f(\hat{H})\psi = \sum_{n=0}^N c_n T_n(\hat{H})\psi . \quad (43)$$

The expansion coefficient can be obtained from the expansion coefficients $a_n(t)$ of the time propagator. For a normalized Hamiltonian they read

$$c_n = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt a_n(t) \int dE e^{iEt} f(E) . \quad (44)$$

Since a Chebychev expansion of order N yields a converged description of the time-dependent wavefunction for propagation time of about $N/\Delta E$, the expansion index N

could be viewed as a transformed propagation time. In scattering problems the computation of the action of the Greens function

$$\frac{1}{E - (\hat{H} - i\epsilon)} \psi \quad (45)$$

is a central task. This Greens function includes a negative imaginary potential $-i\epsilon$. The modified Hamiltonian $\hat{H} - i\epsilon$ is not hermitian. To obtain a convergent expansion in polynomial of this non-hermitian Hamiltonian, a modified Chebychev recursion has to be used.²¹

5 Filter Diagonalization

Formally time-dependent and energy-dependent representations are equivalent. However, depending on the particular phenomenon, either the time-dependent or the energy-dependent picture can yield a more intuitive interpretation of the process. The energy-dependent description seems favorable if a small number of individual states dominates the process under investigation. Non-overlapping sharp resonances are an example where a energy-dependent picture might be preferable. In contrast, the time-dependent picture is very suitable for the description of rapid dissociation processes showing broad spectra. In many systems both situation are simultaneously present: few resonances are embedded in a moderately structured background. Then the filter diagonalization approach²²⁻²⁴ which combines time-dependent and energy-dependent descriptions yields an effective description of the system.

In the filter diagonalization approach,²² a set of wavefunctions $\psi_1, \psi_2, \dots, \psi_N$ corresponding to N energies $E_n = E_0 + n \Delta E$ in the interval $[E_1, E_N]$ is obtained by wave packet propagation:

$$\psi_n = \int_{-T}^T dt e^{iE_n t} e^{-i\hat{H}t} \psi, \quad (46)$$

$$= \hat{F}(E_n) \psi. \quad (47)$$

The propagation time T should be sufficiently large to separate the different ψ_n , i.e. $\Delta E \cdot T \geq 1$. If the energy grid is sufficiently fine, i.e. N exceeds the number of energy eigenstates in the interval $[E_1, E_N]$, the set of wavefunctions $\psi_1, \psi_2, \dots, \psi_N$ forms a complete basis for the representation of the energy eigenstates in the interval. The eigenstates and eigenvalues can be computed by representing the Hamiltonian in this basis and diagonalizing the resulting N -dimensional Hamiltonian matrix.

In the filter diagonalization approach, the required propagation time is given by the averaged density of states $\rho(E)$:

$$T \geq \frac{N}{E_N - E_1} \geq \frac{1}{E_N - E_1} \int_{E_1}^{E_N} \rho(E) . \quad (48)$$

Since the propagation is followed by an numerically exact diagonalization of the Hamiltonian matrix $\langle \psi_j | \hat{H} | \psi_n \rangle$, accurate energy eigenvalue are obtained. The energy resolution is not limited by the propagation time. In a simple wave packet propagation scheme, the energy resolution would have been Fourier limited to $1/T$. In contrast to a full diagonalization of the Hamiltonian, the filter diagonalization approach reduces the size of the problem

by limiting the representation to a small energy window. The projection into this window is obtained from a wave packet propagation for a limited time T.

The filter diagonalization approach can also directly address the autocorrelation function.^{23,24} Then the explicit construction of the wavefunctions ψ_n can be avoided and the memory requirements are reduced. Also other choices for the filter operators $\hat{F}(E_n)$ are possible. The only principal requirement for $\hat{F}(E_n)$ is to project on states with an energy E_n if N goes to infinity.

6 MCTDH

The wave packet propagation schemes described above employ multi-dimensional time-independent grids or basis sets to represent the wavefunction. The numerical effort of these schemes increases exponentially with the number of degrees of freedom. Given the computational resources presently available, only systems with up to four atoms can be treated accurately. The extension of numerically exact calculations towards larger systems therefore requires other schemes for the solution of the Schrödinger equation. The multi-configurational time-dependent Hartree (MCTDH) approach^{9,10} utilizes optimized time-dependent expansion functions to represent the wavefunction. The numerical effort of the MCTDH approach scales exponentially with the number of degrees of freedom but the effort increases less dramatically with dimensionality than in standard wave packet propagation schemes. Thus, the MCTDH approach facilitates the description of systems with are beyond the range of conventional wave packet propagation. Recent applications include a 24-dimensional calculation on the absorption spectrum of pyrazine^{11,12} and a 12-dimensional investigation of the $H + CH_4 \rightarrow H_2 + CH_4$ reaction.^{14,15}

In the multi-configurational time-dependent Hartree (MCTDH) approach,^{9,10} the wavefunction $\psi(x_1, \dots, x_f, t)$ is represented as

$$\psi(x_1, \dots, x_f, t) = \sum_{j_1=1}^{n_1} \dots \sum_{j_f=1}^{n_f} A_{j_1 \dots j_f}(t) \cdot \phi_{j_1}^{(1)}(x_1, t) \cdot \dots \cdot \phi_{j_f}^{(f)}(x_f, t), \quad (49)$$

The $A_{j_1 \dots j_f}(t)$ are time-dependent expansion coefficients. The time-dependent expansion functions $\phi_{j_\kappa}^{(\kappa)}(x_\kappa, t)$ are called single-particle functions. Standard DVR or FFT-schemes can be used to represent these single-particle functions:

$$\phi_j^{(\kappa)}(x_\kappa, t) = \sum_{l=1}^{N_\kappa} c_{jl}^{(\kappa)}(t) \cdot \chi_l(x_\kappa), \quad (50)$$

where the χ_l denote the time-independent basis functions (or grid points) employed in the DVR or FFT scheme. Based on the above ansatz, equations of motion can be derived from the Dirac-Frenkel variational principle.^{9,10} Analogous to standard wave packet propagation, the expansion coefficients A are propagated by the Hamiltonian represented in the basis employed (which here is time-dependent):

$$i \frac{\partial}{\partial t} A_{l_1 \dots l_f}(t) = \sum_{j_1=1}^{n_1} \dots \sum_{j_f=1}^{n_f} \langle \phi_{l_1}^{(1)} \dots \phi_{l_f}^{(f)} | \hat{H} | \phi_{j_1}^{(1)} \dots \phi_{j_f}^{(f)} \rangle A_{j_1 \dots j_f}(t). \quad (51)$$

The differential equations for the single-particle functions ϕ are more involved:

$$i \frac{\partial}{\partial t} \phi_n^{(\kappa)}(x_\kappa, t) = (1 - \hat{P}_\kappa) \sum_m \rho_{nm}^{(\kappa)-1} \sum_j \langle \psi_m^{(\kappa)} | \hat{H} | \psi_j^{(\kappa)} \rangle \phi_j^{(\kappa)}. \quad (52)$$

The above equations include a projection operator on the space spanned by the single-particle functions,

$$\hat{P}_\kappa = \sum_j |\phi_j^{(\kappa)}\rangle \langle \phi_j^{(\kappa)}|, \quad (53)$$

the matrix of mean-field operators acting only on the coordinate x_κ ,

$$\langle \psi_m^{(\kappa)} | \hat{H} | \psi_j^{(\kappa)} \rangle, \quad (54)$$

which employs the the single-hole functions

$$\begin{aligned} \psi_j^{(\kappa)}(x_1, \dots, x_{\kappa-1}, x_{\kappa+1}, \dots, x_f) &= \sum_{j_1} \dots \sum_{j_{\kappa-1}} \sum_{j_{\kappa+1}} \dots \sum_{j_f} A_{j_1 \dots j_{\kappa-1} j j_{\kappa+1} \dots j_f}(t) \cdot \\ &\cdot \phi_{j_1}^{(1)}(x_1, t) \cdot \dots \cdot \phi_{j_{\kappa-1}}^{(\kappa-1)}(x_{\kappa-1}, t) \cdot \phi_{j_{\kappa+1}}^{(\kappa+1)}(x_{\kappa+1}, t) \cdot \dots \cdot \phi_{j_f}^{(f)}(x_f, t), \end{aligned} \quad (55)$$

and the inverse of the single-particle density matrix $\rho_{ij}^{(\kappa)}$,

$$\rho_{ij}^{(\kappa)} = \langle \psi_i^{(\kappa)} | \psi_j^{(\kappa)} \rangle. \quad (56)$$

The MCTDH-representation (49) of the wavefunction should be compared to the representation of the wavefunction employed in a standard wave packet scheme:

$$\psi(x_1, \dots, x_f, t) = \sum_{l_1=1}^{N_1} \dots \sum_{l_f=1}^{N_f} \tilde{A}_{l_1 \dots l_f}(t) \cdot \chi_{l_1}^{(1)}(x_1) \cdot \dots \cdot \chi_{l_f}^{(f)}(x_f). \quad (57)$$

The standard scheme expands the wavefunction in a time-independent basis while the MCTDH-approach employs an optimized set of time-dependent expansion functions. Thus, n , the number of single-particle functions required, can be much smaller than N , the number of underlying time-independent basis functions. Assuming equal basis set sizes in all f degrees of freedom, the numerical effort of a standard wavepacket propagation is approximately proportional to N^{f+1} . In the MCTDH-approach, there are to different contributions to the numerical effort which scale differently with dimensionality. The numerical effort resulting from the A-coefficients is proportional to n^{f+1} while the effort resulting from the representation of the single-particle function approximately equals $f \cdot n \cdot N^2$. For larger systems, the n^{f+1} component dominates. Thus, the numerical effort of the standard wave packet propagation as well as the effort of the MCTDH scheme scales exponentially with the number of degrees of freedom. However, for multi-dimensional problems, the MCTDH is more efficient than the standard wavepacket propagation since n can be considerably smaller than N . Typically N values are of the order of 10^2 while n is often smaller than 10.

In contrast to standard wave packet propagation, the differential equations (51,52) describing the propagation of the MCTDH wavefunction are nonlinear. Thus, the integration schemes discussed in Sect.3 are not directly applicable. However, instead of resorting to

general purpose integrator, an efficient integration scheme developed particularly for integrating the MCTDH equations^{26,27} can be used. This scheme views eq.(51) as set a linear differential equations with a time-dependent Hamiltonian matrix and employs a short iterative Lanczos scheme to integrate these equations.

A difficulty in the MCTDH approach is the evaluation of the potential energy matrix elements

$$\left\langle \phi_{l_1}^{(1)} \cdot \dots \cdot \phi_{l_f}^{(f)} \right| V \left| \phi_{j_1}^{(1)} \cdot \dots \cdot \phi_{j_f}^{(f)} \right\rangle . \quad (58)$$

The direct integration of these matrix elements using the DVR or FFT grid employed for the representation of the single-particle functions ϕ is prohibitive, since it would require a multi-dimensional grid of the size N^f . The problem can be avoided if the potential can be specified as a sum of products of one-dimensional functions¹⁰

$$V(x_1, x_2, \dots, x_f) = \sum_{j=1}^J c_j \cdot v_j^{(1)}(x_1) \cdot v_j^{(2)}(x_2) \cdot \dots \cdot v_j^{(f)}(x_f) . \quad (59)$$

Then the above multi-dimensional integral can be decomposed into one-dimensional components

$$\begin{aligned} & \left\langle \phi_{l_1}^{(1)} \cdot \dots \cdot \phi_{l_f}^{(f)} \right| V \left| \phi_{j_1}^{(1)} \cdot \dots \cdot \phi_{j_f}^{(f)} \right\rangle = \\ & \sum_{j=1}^J c_j \left\langle \phi_{l_1}^{(1)} | v_j^{(1)} | \phi_{j_1}^{(1)} \right\rangle \cdot \dots \cdot \left\langle \phi_{l_f}^{(f)} | v_j^{(f)} | \phi_{j_f}^{(f)} \right\rangle . \end{aligned} \quad (60)$$

However, for many potentials this decomposition can not be achieved with a reasonable number of terms J.

Alternatively, the problem can be resolved by employing the correlation DVR (CDVR) approach²⁵ to evaluate the above integrals. The CDVR scheme employs time-dependent grids obtained by diagonalizing the coordinate matrix represented in the basis of the time-dependent single-particle functions

$$\langle \phi_n^{(\kappa)} | x_\kappa | \phi_m^{(\kappa)} \rangle = \sum_{l=1}^{n_\kappa} \langle \phi_n^{(\kappa)} | X_l^{(\kappa)} \rangle X_l^{(\kappa)} \langle X_l^{(\kappa)} | \phi_m^{(\kappa)} \rangle . \quad (61)$$

However, these grids can not be used for the quadrature of the potential integrals without an essential modification. In the MCTDH-approach, the coefficients $A_{j_1 \dots j_f}(t)$ describe mainly the correlation between the different degrees of freedom, while the motion of the time-dependent basis functions $\phi_{j_\kappa}^{(\kappa)}(x_\kappa, t)$ accounts for the separable dynamics. Thus, the size of the time-dependent basis depends only on the amount of correlation, it is independent of any separable dynamics. The time-dependent basis is small compared with the size of the underlying primitive grid. Due to the small size of the time-dependent basis, a quadrature based on this basis can properly describe only those parts of the potential which only result in correlations. The number of time dependent grid points $X_j^{(\kappa)}(t)$ is too small to also result in an accurate evaluation of the separable parts of the potential. This problems can be solved by explicitly accounting for separable parts of the potential in the quadrature.²⁵

As noted above, the numerical effort of the MCTDH approach scales exponentially with the dimensionality. Employing directly the ansatz (49), MCTDH calculations are limited to 10-20 degrees of freedom. Even if only two single-particle function per degree of freedom are used, $2^{20} \approx 10^6$ A-coefficients would be required in a 20-dimensional calculation. To overcome this limitation, several physical coordinates have to be grouped together and treated as a single logical coordinate (“mode combination”).^{11,27} If the coordinates $\{x_1, \dots, x_f\}$ are grouped as

$$\{(x_1, x_2, \dots, x_{p_1}), (x_{p_1+1}, \dots, x_{p_2}), \dots, (x_{p_{F-1}+1}, \dots, x_f)\} \quad (62)$$

the corresponding MCTDH wavefunction reads

$$\begin{aligned} \psi(x_1, \dots, x_f, t) = & \sum_{j_1=1}^{n_1} \dots \sum_{j_F=1}^{n_F} A_{j_1 \dots j_p}(t) \cdot \phi_{j_1}^{(1)}(x_1, x_2, \dots, x_{p_1}, t) \cdot \phi_{j_2}^{(2)}(x_{p_1+1}, \dots, x_{p_2}, t) \\ & \dots \cdot \phi_{j_F}^{(F)}(x_{p_{F-1}+1}, \dots, x_f, t). \end{aligned} \quad (63)$$

Employing this scheme, converged 24-dimensional MCTDH wave packet calculations on the $S_0 \rightarrow S_2$ excitation in pyrazine^{11,12} and up to 80-dimensional calculations on the spin-boson model¹³ have been reported. However, this mode combination approach presently can not be combined with the CDVR scheme for potential evaluation. Thus, these calculations can only study wave packet motion on potential energy surfaces which can be represented in the form analogous to eq.(59).

Acknowledgments

The author would like to thank F. Huarte-Larrañaga, W. Tao, A. Lucke, W. Domcke, F. Matzkies, Th. Gerdts, Ch. Schlier, J. Briggs, T. Seideman, W. H. Miller, A. Hammerich, H.-D. Meyer, and L. S. Cederbaum for interesting discussions, coworking on many of the subjects presented in this article, and providing a stimulating working environment. Financial support by the Deutsche Forschungsgemeinschaft, the European commission, and the Fond der Chemischen Industrie is gratefully acknowledged.

References

1. D. Kosloff and R. Kosloff, *J. Comp. Phys.* **52**, 35 (1983).
2. D. O. Harris, G. G. Engerholm, and W. D. Gwinn, *J. Chem. Phys.* **43**, 1515 (1965).
3. A. S. Dickinson and P. R. Certain, *J. Chem. Phys.* **49**, 4209 (1968).
4. J. C. Light, I. P. Hamilton, and J. V. Lill, *J. Chem. Phys.* **82**, 1400 (1985).
5. M. D. Feit, J. A. Fleck, and A. Steiger, *J. Comp. Phys.* **47**, 412 (1982).
6. H. Tal-Ezer and R. Kosloff, *J. Chem. Phys.* **81**, 3967 (1984).
7. T. J. Park and J. C. Light, *J. Chem. Phys.* **85**, 5870 (1986).
8. D. Neuhauser, *J. Chem. Phys.* **95**, 4927 (1991).
9. H.-D. Meyer, U. Manthe, and L. S. Cederbaum, *Chem. Phys. Lett.* **165**, 73 (1990).
10. U. Manthe, H.-D. Meyer, and L. S. Cederbaum, *J. Chem. Phys.* **97**, 3199 (1992).
11. G. A. Worth, H.-D. Meyer, and L. S. Cederbaum, *J. Chem. Phys.* **109**, 3518 (1998).
12. A. Raab, G. A. Worth, H.-D. Meyer, and L. S. Cederbaum, *J. Chem. Phys.* **110**, 936 (1999).

13. H. Wang, *J. Chem. Phys.* **113**, 9948 (2000).
14. F. Huarte-Larranaga and U. Manthe, *J. Chem. Phys.* **113**, 5115 (2000).
15. F. Huarte-Larranaga and U. Manthe, *J. Phys. Chem. A* **105**, 2522 (2001).
16. D. T. Colbert and W. H. Miller, *J. Chem. Phys.* **96**, 1982 (1992).
17. E. Fattal, R. Baer, and R. Kosloff, *Phys. Rev. E* **53**, 1217 (1996).
18. N. Sakamoto and S. Muramatsu, *Phys. Rev. B* **17**, 868 (1978).
19. M. C. M. O'Brien and S. N. Evangelou, *J. Phys. C* **13**, 611 (1980).
20. E. Haller, L. S. Cederbaum, and W. Domcke, *Mol. Phys.* **41**, 1291 (1980).
21. V. A. Mandelshtam and H. S. Taylor, *J. Chem. Phys.* **103**, 2903 (1995).
22. D. Neuhauser, *J. Chem. Phys.* **93**, 2611 (1990).
23. M. R. Wall and D. Neuhauser, *J. Chem. Phys.* **102**, 8011 (1995).
24. V. A. Mandelshtam and H. S. Taylor, *J. Chem. Phys.* **106**, 5085 (1997).
25. U. Manthe, *J. Chem. Phys.* **105**, 6989 (1996).
26. M. H. Beck and H.-D. Meyer, *Z. Phys. D* **42**, 113 (1997).
27. M. H. Beck, A. Jäckle, G. A. Worth, and H.-D. Meyer, *Physics reports* **324**, 1 (2000).

Nonadiabatic Dynamics: Mean-Field and Surface Hopping

Nikos L. Doltsinis

Lehrstuhl für Theoretische Chemie
Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail: nikos.doltsinis@theochem.ruhr-uni-bochum.de

This contribution takes a closer look at the foundations of conventional molecular dynamics simulations such as the Born-Oppenheimer approximation and the treatment of atomic nuclei according to the laws of classical mechanics. Regimes of validity of the adiabatic approximation are defined and models that take into account nonadiabatic effects in situations where the Born-Oppenheimer approximation breaks down are introduced. We focus on two mixed quantum-classical methods that differ only in the way the forces on the — classical — atomic nuclei are determined from the solutions to the time-independent electronic Schrödinger equation. In the Ehrenfest approach, the system moves on a single potential energy surface obtained by weighted averaging over all adiabatic states, whereas the ‘surface hopping’ method allows transitions between pure adiabatic potential energy surfaces according to their weights. In both cases, the weights are the squares of the coefficients of the total electronic wavefunction expanded in terms of the adiabatic state functions.

1 Introduction

Molecular dynamics (MD), in the literal sense, is the simultaneous motion of a number of atomic nuclei and electrons forming a molecular entity. Strictly speaking, a complete description of such a system requires solving the full time-dependent Schrödinger equation including both electronic and nuclear degrees of freedom. This, however, is a formidable computational task which is in fact altogether unfeasible, at present, for systems consisting of more than three atoms and more than one electronic state.¹ In order to study the dynamics of the vast majority of chemical systems, several approximations, therefore, have to be imposed.

Firstly, it is assumed in MD that the motions of slow and fast degrees of freedom are separable (adiabatic or Born-Oppenheimer approximation). In the molecular context this means that the electron cloud adjusts instantly to changes in the nuclear configuration. As a consequence, nuclear motion evolves on a *single* potential energy surface (PES), associated with a *single* electronic quantum state, which is obtained by solving the time-independent Schrödinger equation for a series of fixed nuclear geometries. In practice, most MD simulations are performed on a *ground state* PES.

Moreover, in addition to making the Born-Oppenheimer approximation, MD treats the atomic nuclei as *classical* particles whose trajectories are computed by integrating Newton’s equations of motion.

MD has been applied with great success to study a wide range of systems from biomolecules to condensed phases.^{2,3} Its underlying approximations, on the other hand,

break down in many important physical situations and extensions of the method are needed for those scenarios. An accurate description of hydrogen motion, for instance, requires quantum mechanical treatment. Processes such as charge-transfer reactions and photochemistry are inherently *nonadiabatic*, i.e., they involve (avoided) crossings of different electronic states rendering the Born-Oppenheimer approximation invalid.

Critical assessment of the adiabatic approximation as well as discussion of nonadiabatic extensions will be the subject of the present paper.

Since our focus here is on potential applicability to large-scale systems, we shall retain the classical treatment of the nuclei and only describe the electrons quantum mechanically. We will use the term semiclassical for such mixed quantum-classical models. Both expressions can be frequently found in the literature.

Out of the great many semiclassical approaches to nonadiabatic dynamics that have been proposed two “standard” methods different in philosophy have emerged as the most popular ones. One extreme is the Ehrenfest method,^{1,4-8} where the nuclei move on *one* effective PES which is an average of all adiabatic states involved weighted by their populations (therefore also called mean-field method). The other extreme is the surface hopping approach,^{9,10,7,8,11,12} where the nuclei evolve on pure adiabatic PESs, but switches between adiabatic states are allowed when their populations change.

This article is organised as follows. In Section 2, the Born-Oppenheimer approximation is introduced. Starting from the full time-dependent Schrödinger equation, the uncoupled nuclear equations of motion are derived. Section 3 deals with the semiclassical approach replacing the nuclear wavefunction by a classical trajectory. This will form the basis of all nonadiabatic methods presented in later sections. Conditions for the validity of the Born-Oppenheimer approximation are discussed qualitatively. Two of the most commonly employed nonadiabatic dynamics methods are described in Section 4, namely the Ehrenfest and the surface hopping methods. The section closes by presenting a recent implementation of the surface hopping technique within the framework of Car-Parrinello MD together with an application to the cis-trans photoisomerisation of formaldimine as a case study.

2 Born-Oppenheimer Approximation

A complete, non-relativistic, description of a system of N atoms having the positions $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_K, \dots, \mathbf{R}_N)$ with n electrons located at $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_K, \dots, \mathbf{r}_n)$ is provided by the time-dependent Schrödinger equation

$$\mathcal{H}\Xi(\mathbf{r}, \mathbf{R}; t) = i\hbar \frac{\partial}{\partial t} \Xi(\mathbf{r}, \mathbf{R}; t) \quad , \quad (1)$$

with the total Hamiltonian

$$\mathcal{H}(\mathbf{r}, \mathbf{R}) = \mathcal{T}(\mathbf{R}) + \mathcal{T}(\mathbf{r}) + \mathcal{V}(\mathbf{R}) + \mathcal{V}(\mathbf{r}, \mathbf{R}) + \mathcal{V}(\mathbf{r}) \quad , \quad (2)$$

being the sum of kinetic energy of the atomic nuclei,

$$\mathcal{T}(\mathbf{R}) = -\frac{\hbar^2}{2} \sum_{K=1}^N \frac{\nabla_K^2}{M_K} \quad , \quad (3)$$

kinetic energy of the electrons,

$$\mathcal{T}(\mathbf{r}) = -\frac{\hbar^2}{2m_e} \sum_{k=1}^n \nabla_k^2 , \quad (4)$$

internuclear repulsion,

$$\mathcal{V}(\mathbf{R}) = \frac{e^2}{4\pi\epsilon_0} \sum_{K=1}^{N-1} \sum_{L>K}^N \frac{Z_K Z_L}{|\mathbf{R}_K - \mathbf{R}_L|} , \quad (5)$$

electronic – nuclear attraction,

$$\mathcal{V}(\mathbf{r}, \mathbf{R}) = -\frac{e^2}{4\pi\epsilon_0} \sum_{K=1}^N \sum_{k=1}^n \frac{Z_K}{|\mathbf{r}_k - \mathbf{R}_K|} , \quad (6)$$

and interelectronic repulsion,

$$\mathcal{V}(\mathbf{r}) = \frac{e^2}{4\pi\epsilon_0} \sum_{k=1}^{n-1} \sum_{l>k}^n \frac{1}{|\mathbf{r}_k - \mathbf{r}_l|} . \quad (7)$$

Here, M_K and Z_K denote the mass and atomic number of nucleus K ; m_e and e are the electronic mass and elementary charge, and ϵ_0 is the permittivity of vacuum. The nabla operators ∇_K and ∇_k act on the coordinates of nucleus K and electron k , respectively.

Defining the electronic Hamiltonian (fixed-nuclei approximation of \mathcal{H}) as

$$\mathcal{H}_{\text{el}}(\mathbf{r}, \mathbf{R}) = \mathcal{T}(\mathbf{r}) + \mathcal{V}(\mathbf{R}) + \mathcal{V}(\mathbf{r}, \mathbf{R}) + \mathcal{V}(\mathbf{r}) , \quad (8)$$

we can rewrite the total Hamiltonian as

$$\mathcal{H}(\mathbf{r}, \mathbf{R}) = \mathcal{T}(\mathbf{R}) + \mathcal{H}_{\text{el}}(\mathbf{r}, \mathbf{R}) . \quad (9)$$

Let us suppose the solutions of the time-independent (electronic) Schrödinger equation,

$$\mathcal{H}_{\text{el}}(\mathbf{r}, \mathbf{R})\phi_i(\mathbf{r}, \mathbf{R}) = E_i(\mathbf{R})\phi_i(\mathbf{r}, \mathbf{R}) , \quad (10)$$

are known. Furthermore, the spectrum of $\mathcal{H}_{\text{el}}(\mathbf{r}, \mathbf{R})$ is assumed to be discrete and the eigenfunctions orthonormalised:

$$\int_{-\infty}^{\infty} \phi_i^*(\mathbf{r}, \mathbf{R})\phi_j(\mathbf{r}, \mathbf{R})d\mathbf{r} \equiv \langle \phi_i | \phi_j \rangle = \delta_{ij} . \quad (11)$$

The total wavefunction Ξ can be expanded in terms of the eigenfunctions of \mathcal{H}_{el} since these form a complete set:

$$\Xi(\mathbf{r}, \mathbf{R}; t) = \sum_j \phi_j(\mathbf{r}, \mathbf{R})\chi_j(\mathbf{R}, t) . \quad (12)$$

Insertion of this ansatz into the time-dependent Schrödinger equation (1) followed by multiplication from the left by $\phi_i^*(\mathbf{r}, \mathbf{R})$ and integration over the electronic coordinates leads to a set of coupled differential equations:

$$[\mathcal{T}(\mathbf{R}) + E_i(\mathbf{R})]\chi_i + \sum_j \mathcal{C}_{ij}\chi_j = i\hbar \frac{\partial}{\partial t} \chi_i , \quad (13)$$

where the coupling operator \mathcal{C}_{ij} is defined as

$$\mathcal{C}_{ij} \equiv \langle \phi_i | \mathcal{T}(\mathbf{R}) | \phi_j \rangle - \sum_K \frac{\hbar^2}{M_K} \langle \phi_i | \nabla_K | \phi_j \rangle \nabla_K . \quad (14)$$

The diagonal term \mathcal{C}_{ii} represents a correction to the (adiabatic) eigenvalue E_i of the electronic Schrödinger equation (10). In the case that all coupling operators \mathcal{C}_{ij} are negligible, the set of differential eqns (13) becomes uncoupled:

$$[\mathcal{T}(\mathbf{R}) + E_i(\mathbf{R})] \chi_i = i\hbar \frac{\partial}{\partial t} \chi_i . \quad (15)$$

This means that the nuclear motion proceeds without changes of the quantum state of the electron cloud and, correspondingly, the wavefunction (12) is reduced to a single term (adiabatic approximation):

$$\Xi(\mathbf{r}, \mathbf{R}; t) \approx \phi_i(\mathbf{r}, \mathbf{R}) \chi_i(\mathbf{R}, t) . \quad (16)$$

For a great number of physical situations the Born-Oppenheimer approximation can be safely applied. On the other hand, there are many important chemical phenomena like, for instance, charge transfer and photoisomerisation reactions, whose very existence is due to the inseparability of electronic and nuclear motion. Inclusion of nonadiabatic effects will be the subject of the following sections.

3 Semiclassical Approach

Further simplification of the problem can be achieved by describing nuclear motion by classical mechanics and only the electrons quantum mechanically. In this so-called semiclassical approach,^{13,14} the atomic nuclei follow some trajectory $\mathbf{R}(t)$ while the electronic motion is captured by some time-dependent total wavefunction $\Phi(\mathbf{r}; t)$ satisfying the time-dependent electronic Schrödinger equation,

$$\mathcal{H}_{\text{el}}(\mathbf{r}, \mathbf{R}(t)) \Phi(\mathbf{r}; t) = i\hbar \frac{\partial}{\partial t} \Phi(\mathbf{r}; t) . \quad (17)$$

Again, the total wavefunction is written as a linear combination of adiabatic eigenfunctions $\phi_i(\mathbf{r}, \mathbf{R})$ (solutions of the time-independent Schrödinger equation (10)):

$$\Phi(\mathbf{r}; t) = \sum_j a_j(t) \phi_j(\mathbf{r}, \mathbf{R}) e^{-\frac{i}{\hbar} \int E_j(\mathbf{R}) dt} . \quad (18)$$

Insertion of this ansatz into the time-dependent electronic Schrödinger equation (17) followed by multiplication from the left by $\phi_i^*(\mathbf{r}, \mathbf{R})$ and integration over the electronic coordinates leads to a set of coupled differential equations:

$$\dot{a}_i = - \sum_j a_j C_{ij} e^{-\frac{i}{\hbar} \int (E_j - E_i) dt} , \quad (19)$$

where

$$C_{ij} \equiv \langle \phi_i | \frac{\partial}{\partial t} | \phi_j \rangle \quad (20)$$

are the nonadiabatic coupling elements. Integration of eqns (19) yields the expansion coefficients $a_i(t)$ whose square modulus, $|a_i(t)|^2$, can be interpreted as the probability of finding the system in the adiabatic state i at time t .

We now want to develop a condition for the validity of the Born-Oppenheimer approximation based on qualitative arguments. For this purpose, we shall consider a two-state system. To illustrate the problem, fig. 1 shows the avoided crossing between the covalent and ionic potential energy curves of NaCl.^{15,16} As we can see, the adiabatic wavefunctions ϕ_1 and ϕ_2 change their character as the bond length is varied. The characteristic length, l , over which ϕ_1 and ϕ_2 change significantly clearly depends on the nuclear configuration \mathbf{R} ; in the vicinity of the NaCl avoided crossing, for instance, the character of the wavefunctions varies rapidly, whereas at large separations it remains more or less constant.

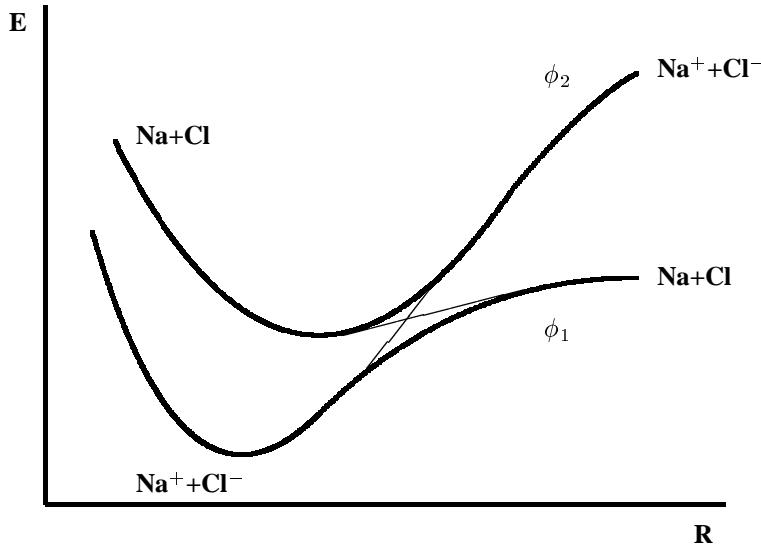


Figure 1. Avoided crossing between the covalent and ionic adiabatic potential curves of NaCl (thin lines: crossing of diabatic states).

Division of the characteristic length l by the velocity of the nuclei, $\dot{\mathbf{R}} = |\dot{\mathbf{R}}|$, at a particular configuration \mathbf{R} defines the time the system needs to travel the distance l around \mathbf{R} :

$$\text{passage time } \tau_p = \frac{l}{\dot{R}} . \quad (21)$$

In order for the Born-Oppenheimer approximation to be valid, the electron cloud has to adjust instantly to the nuclear changes. The time scale characteristic of electronic motion can be obtained from the relation

$$\Delta E = |E_1 - E_2| = \hbar\omega \quad (22)$$

by taking the inverse transition frequency:

$$\tau_e = \frac{1}{\omega} = \frac{\hbar}{\Delta E} . \quad (23)$$

The ratio

$$\xi = \frac{\tau_p}{\tau_e} = \frac{\Delta E l}{\hbar \dot{R}} \quad (24)$$

is the so-called Massay parameter. For values $\xi \gg 1$, i.e. large energy gaps ΔE and small velocities \dot{R} , nonadiabatic effects are negligible. In this case, if the system is prepared in some pure adiabatic state i ($|a_i|^2 = 1$) at time $t = 0$, the rhs of eqn (19) will be zero at all times and the wavefunction expansion (eqn (18)) can be replaced by a single term:

$$\Phi(\mathbf{r}; t) = \phi_i(\mathbf{r}, \mathbf{R}) e^{-\frac{i}{\hbar} \int E_i(\mathbf{R}) dt} . \quad (25)$$

The atomic nuclei are then propagated by solving Newton's equations

$$M_K \ddot{\mathbf{R}}_K = \mathbf{F}_K(\mathbf{R}) , \quad (26)$$

where

$$\mathbf{F}_K(\mathbf{R}) = -\nabla_K E_i(\mathbf{R}) \quad (27)$$

is the force on atom K .

4 Approaches to Nonadiabatic Dynamics

4.1 Mean-Field (Ehrenfest) Method

As we have discussed in the previous section, nonadiabaticity involves changes in the adiabatic state populations $|a_i|^2$ with changing nuclear configuration. Clearly, such a distortion of the electron cloud will, in turn, influence the nuclear trajectory. Although there are situations in which the impact of electronic nonadiabaticity on nuclear motion is negligible (e.g. for high energy collisions or small energy separations between adiabatic states), for many chemical systems it is of prime importance to properly incorporate electronic–nuclear feedback.^{7,8}

The simplest way of doing this is to replace the adiabatic potential energy surface E_i in eqn (27) by the energy expectation value

$$E^{\text{eff}} = \langle \Phi | \mathcal{H}_{\text{el}} | \Phi \rangle = \sum_i |a_i|^2 E_i , \quad (28)$$

where we have used eqn (18). Thus, the atoms evolve on an effective potential representing an average over the adiabatic states weighted by their state populations $|a_i|^2$ (as illustrated in fig. 2). The method is therefore referred to as mean-field (also known as Ehrenfest) approach.

It is instructive to derive an expression for the nuclear forces either from the gradient of eqn (28) or using the Hellmann-Feynman theorem

$$\mathbf{F}_K = -\langle \Phi | \nabla_K \mathcal{H}_{\text{el}} | \Phi \rangle . \quad (29)$$

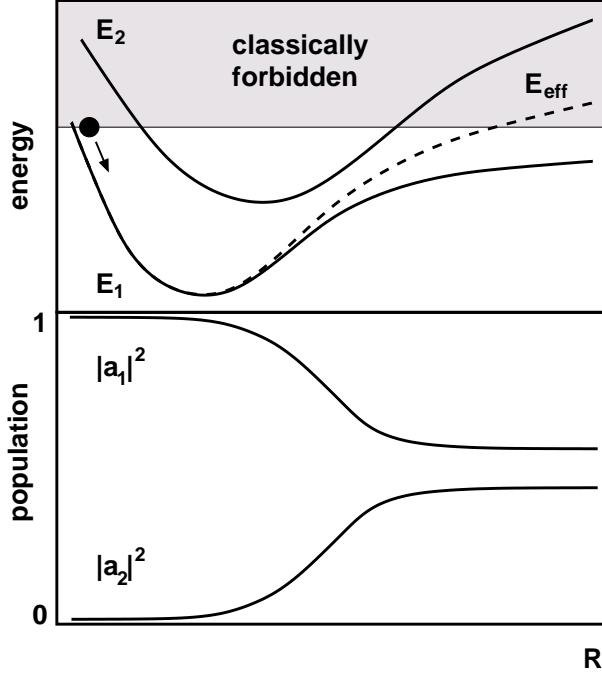


Figure 2. Top: avoided crossing between two adiabatic PES, E_1 and E_2 , and effective potential, E_{eff} , on which the nuclei are propagated in the Ehrenfest method. In the asymptotic region (right) E_{eff} contains contributions from classically forbidden regions of E_2 . Bottom: corresponding adiabatic state populations $|a_1|^2$ and $|a_2|^2$. The system is prepared in state 1 initially with zero kinetic energy. Upon entering the coupling region state 2 is increasingly populated.

Opting for the latter, we start by writing down the relation

$$\nabla_K \langle \phi_i | \mathcal{H}_{\text{el}} | \phi_j \rangle = \nabla_K E_i \delta_{ij} \quad (30)$$

$$= \langle \nabla_K \phi_i | \mathcal{H}_{\text{el}} | \phi_j \rangle + \langle \phi_i | \nabla_K \mathcal{H}_{\text{el}} | \phi_j \rangle + \langle \phi_i | \mathcal{H}_{\text{el}} | \nabla_K \phi_j \rangle \quad (31)$$

$$= \langle \phi_i | \nabla_K \mathcal{H}_{\text{el}} | \phi_j \rangle + (E_j - E_i) \mathbf{d}_{ji} \quad , \quad (32)$$

where we have defined the nonadiabatic coupling vectors, \mathbf{d}_{ji} , as

$$\mathbf{d}_{ji} = \langle \phi_j | \nabla_K | \phi_i \rangle \quad , \quad (33)$$

and used eqn (10) together with the hermiticity of \mathcal{H}_{el} :

$$\langle \phi_i | \mathcal{H}_{\text{el}} | \nabla_K \phi_j \rangle = \langle \nabla_K \phi_j | \mathcal{H}_{\text{el}} | \phi_i \rangle^* = \langle \nabla_K \phi_j | E_j \phi_i \rangle^* = E_i \mathbf{d}_{ij}^* = -E_i \mathbf{d}_{ji} \quad . \quad (34)$$

Note that

$$\mathbf{d}_{ji}^* = -\mathbf{d}_{ij} \quad , \quad (35)$$

because

$$\nabla_K \langle \phi_i | \phi_j \rangle = \nabla_K \delta_{ij} = 0 \quad (36)$$

$$= \langle \nabla_K \phi_i | \phi_j \rangle + \langle \phi_i | \nabla_K \phi_j \rangle = \mathbf{d}_{ji}^* + \mathbf{d}_{ij} \quad . \quad (37)$$

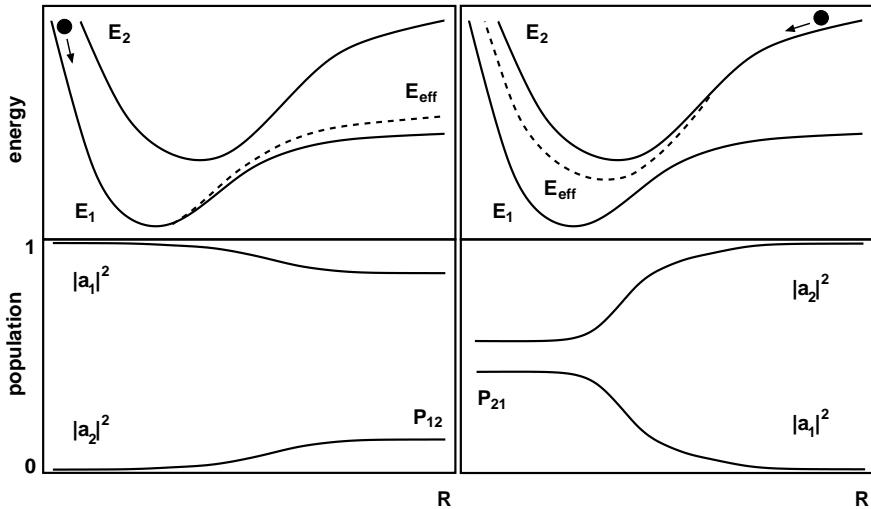


Figure 3. Top left: forward path effective potential, E_{eff} , for two weakly coupled adiabatic PES, E_1 and E_2 . Bottom left: state occupations for a system initially prepared in state 1. The final value of $|a_2|^2$ is equal to the transition probability P_{12} . Top right: backward path effective potential, E_{eff} , for two weakly coupled adiabatic PES, E_1 and E_2 . Bottom left: state occupations for a system initially prepared in state 2. The final value of $|a_1|^2$ is equal to the transition probability P_{21} .

Equating the rhss of eqns (30) and (32) one obtains after rearranging,

$$\langle \phi_i | \nabla_K \mathcal{H}_{\text{el}} | \phi_j \rangle = \nabla_K E_i \delta_{ij} - (E_j - E_i) \mathbf{d}_{ji} \quad . \quad (38)$$

The nuclear forces (29) are thus given by

$$\mathbf{F}_K = - \sum_i |a_i|^2 \nabla_K E_i + \sum_{i,j} a_i^* a_j (E_j - E_i) \mathbf{d}_{ji} \quad . \quad (39)$$

Equation (39) illustrates the two contributions to the nuclear forces; the first term is simply the population-weighted average force over the adiabatic states, while the second term takes into account nonadiabatic changes of the adiabatic state occupations. We would like to point out here that the nonadiabatic contributions to the nuclear forces are in the direction of the nonadiabatic coupling vectors \mathbf{d}_{ji} .

The Ehrenfest method has been applied with great success to a number of chemical problems including energy transfer at metal surfaces.¹⁷ However, due to its mean-field character the method has some serious limitations. A system that was initially prepared in a pure adiabatic state will be in a mixed state when leaving the region of strong nonadiabatic coupling. In general, the pure adiabatic character of the wavefunction cannot be recovered even in the asymptotic regions of configuration space. In cases where the differences in the adiabatic potential energy landscapes are pronounced, it is clear that an average potential will be unable to describe all reaction channels adequately. In particular, if one is interested in a reaction branch whose occupation number is very small, the average path is likely to diverge from the true trajectory. Furthermore, the total wavefunction may contain significant contributions from adiabatic states that are

energetically inaccessible (see fig. 2).

Figure 3 illustrates another severe drawback of the mean-field approach. The principle of microscopic reversibility demands that the forward path probability, $P_{12}^{\text{for}} = |a_2^{\text{final}}|^2$ for a system that was initially prepared in state 1 to end up in state 2 must be equal to the backward path probability, $P_{21}^{\text{back}} = |a_1^{\text{final}}|^2$ for a system that was initially prepared in state 2 to end up in state 1. One can easily think of situations, like the one depicted in fig. 3, for which the effective potentials for the forward and backward paths are very different, resulting also in different populations, $|a_i|^2$. The Ehrenfest method, therefore, violates microscopic reversibility.

It should be noted that the expansion of the total wavefunction in terms of (adiabatic) basis functions (eqn (18)) is not a necessary requirement for the Ehrenfest method; the wavepacket Φ can be propagated numerically using eqn (17). However, projection of Φ onto the adiabatic states facilitates interpretation. Knowledge of the expansion coefficients, a_i , is also the key to refinements of the method such as the surface hopping technique.

4.2 Surface Hopping

We have argued above that after exiting a well localised nonadiabatic coupling region it is unphysical for nuclear motion to be governed by a mixture of adiabatic states. Rather it would be desirable that in asymptotic regions the system evolves on a pure adiabatic PES. This idea is fundamental to the surface hopping approach. Instead of calculating the 'best' (i.e., state-averaged) path like in the Ehrenfest method, the surface hopping technique involves an ensemble of trajectories. At any moment in time, the system is propagated on some pure adiabatic state i , which is selected according to its state population $|a_i|^2$. Changing adiabatic state occupations can thus result in nonadiabatic transitions between different adiabatic PESs (see fig. 4). The ensemble averaged number of trajectories evolving on adiabatic state i at any time is equal to its occupation number $|a_i|^2$.

In the original formulation of the surface hopping method by Tully and Preston,⁹ switches between adiabatic states were allowed only at certain locations defined prior to the simulation. Tully¹⁰ later generalized the method in such a way that nonadiabatic transitions can occur at any point in configuration space. At the same time, an algorithm — the so-called fewest switches criterion — was proposed which minimises the number of surface hops per trajectory whilst guaranteeing the correct ensemble averaged state populations at all times. The latter is important because excessive surface switching effectively results in weighted averaging over the adiabatic states much like in the case of the Ehrenfest method.

We shall now derive the fewest switches criterion. Out of a total of N trajectories, N_i will be in state i at time t ,

$$N_i(t) = \rho_{ii}(t)N \quad . \quad (40)$$

Here we have introduced the density matrix notation

$$\rho_{ij}(t) = a_i^*(t)a_j(t) \quad . \quad (41)$$

At a later time $t' = t + \delta t$ the new occupation numbers are

$$N_i(t') = \rho_{ii}(t')N \quad (42)$$

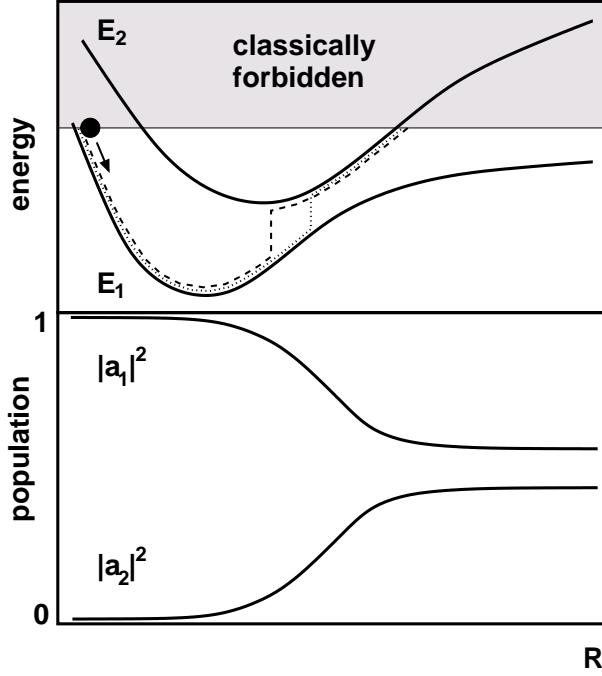


Figure 4. Top: avoided crossing between two adiabatic PES, E_1 and E_2 , and two typical forward surface hopping trajectories. Nonadiabatic transitions are most likely to occur in the coupling region. Bottom: corresponding adiabatic state populations $|a_1|^2$ and $|a_2|^2$. The system is prepared to be in state 1 initially with zero kinetic energy. Upon entering the coupling region state 2 is increasingly populated.

Let us suppose that $N_i(t') < N_i(t)$ or $\delta N = N_i(t) - N_i(t') > 0$. Then the minimum number of transitions required to go from $N_i(t)$ to $N_i(t')$ is δN hops from state i to any other state and zero hops from any other state to state i (see fig. 5). The probability $P_i(t, \delta t)$ for a transition out of state i to any other state during the time interval $[t, t + \delta t]$ is then given by

$$P_i(t, \delta t) = \frac{\delta N}{N} = \frac{\rho_{ii}(t) - \rho_{ii}(t')}{\rho_{ii}} \approx -\frac{\dot{\rho}_{ii}\delta t}{\rho_{ii}} , \quad (43)$$

where we have used

$$\dot{\rho}_{ii} \approx \frac{\rho_{ii}(t') - \rho_{ii}(t)}{\delta t} . \quad (44)$$

The lhs of eqn (44) can be written as

$$\dot{\rho}_{ii} = \frac{d}{dt}(a_i^* a_i) = \dot{a}_i^* a_i + a_i^* \dot{a}_i = (a_i^* a_i)^* + a_i^* \dot{a}_i = 2\Re(a_i^* \dot{a}_i) . \quad (45)$$

Inserting eqn (19) into eqn (45) we obtain

$$\dot{\rho}_{ii} = -2\Re \left(\sum_j \rho_{ij} C_{ij} e^{-\frac{i}{\hbar} \int (E_j - E_i) dt} \right) . \quad (46)$$

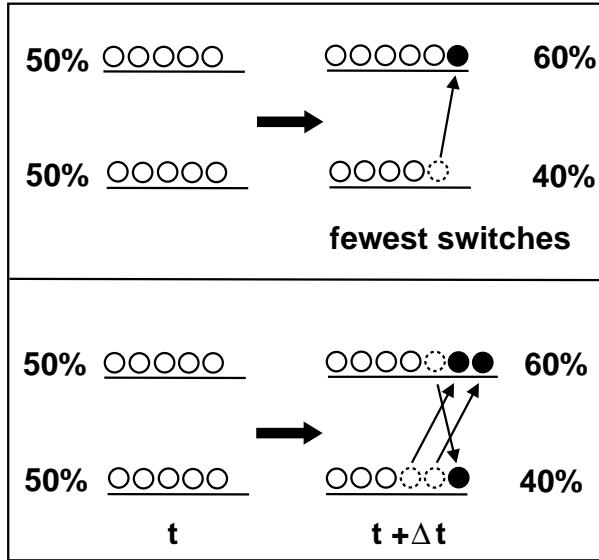


Figure 5. A two-state system with each state being equally (50%) populated at time t . At time $t + \Delta t$ the lower and the upper state are populated by 40 % and 60 % of ensemble members, respectively. The top panel shows how this distribution can be achieved with the minimum number of transitions, whereas the bottom panel shows one alternative route involving a larger number of transitions.

Substituting expression (46) into eqn (43) the probability P_i can be rewritten as follows

$$P_i(t, \delta t) = \frac{2\Re \left(\sum_j \rho_{ij} C_{ij} e^{-\frac{i}{\hbar} \int (E_j - E_i) dt} \right) \delta t}{\rho_{ii}} . \quad (47)$$

Since the probability, P_i , for a switch from state i to any other state must be the sum over all states of the probabilities, P_{ij} , for a transition from state i to a specific state j ,

$$P_i(t, \delta t) = \sum_j P_{ij}(t, \delta t) , \quad (48)$$

it follows from eqn (47) that

$$P_{ij}(t, \delta t) = \frac{2\Re \left(\rho_{ij} C_{ij} e^{-\frac{i}{\hbar} \int (E_j - E_i) dt} \right) \delta t}{\rho_{ii}} . \quad (49)$$

A transition from state i to state k is now invoked if

$$P_i^{(k)} < \zeta < P_i^{(k+1)} , \quad (50)$$

where ζ ($0 \leq \zeta \leq 1$) is a uniform random number and $P_i^{(k)}$ is the sum of the transition probabilities for the first k states,

$$P_i^{(k)} = \sum_j^k P_{ij} . \quad (51)$$

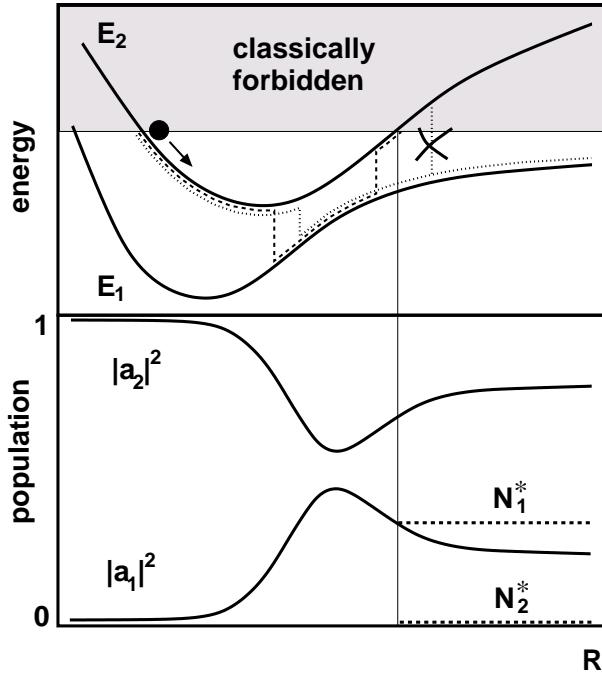


Figure 6. Top: avoided crossing between two adiabatic PES, E_1 and E_2 , and two typical forward surface hopping trajectories. Nonadiabatic transitions are most likely to occur in the coupling region. The cross indicates a classically forbidden transition; no switch is carried out in this case. Bottom: corresponding adiabatic state populations $|a_1|^2$ and $|a_2|^2$. The system is prepared in state 2 initially with zero kinetic energy. Upon entering the coupling region state 1 is increasingly populated. Upon exiting the coupling region, state population 1 decreases. For configurations \mathbf{R} for which E_2 is in the classically forbidden region, the percentages of trajectories in state i , N_i^* , are unequal to $|a_i|^2$; N_2^* is zero whereas N_1^* remains constant.

In order to conserve total energy after a surface hop has been carried out, the atomic velocities have to be rescaled. The usual procedure is to adjust only the velocity components in the direction of the nonadiabatic coupling vector $d_{ik}(\mathbf{R})$ (eqn (33)).¹⁰ We can qualitatively justify this practice by our earlier observation that the nonadiabatic contribution to the Ehrenfest forces also are in the direction of the nonadiabatic coupling vector $d_{ik}(\mathbf{R})$ (see eqn (39)). Certainly, such discontinuities in nuclear velocities must be regarded as a flaw of the surface hopping approach. In most physical scenarios, however, nonadiabatic surface switches take place only at relatively small potential energy separations so that the necessary adjustment to the nuclear velocities is reasonably small. Nevertheless, a severe limitation of the method is presented by its inability to properly deal with situations in which the amount of kinetic energy is insufficient to compensate for the difference in potential energy (so-called classically forbidden transitions). Tully's original suggestion not to carry out a surface hop while retaining the nuclear velocities in such cases has been demonstrated¹⁸ to be more accurate than later proposals to reverse the velocity components in the direction of the nonadiabatic coupling vector $d_{ik}(\mathbf{R})$.^{19,20} The example presented in Figure 6 illuminates how classically forbidden transitions cause

divergence between the target occupation numbers, $|a_i|^2$, and the actual percentages of trajectories evolving in state i , N_i^* .

It should be noted that surface hopping exhibits a large degree of electronic coherence through continuous integration of eqns (19) along the entire trajectory. On the one hand, this enables the method to reproduce quantum interference effects¹⁰ such as Stueckelberg oscillations.¹³ On the other hand, due to treating nuclei classically, dephasing of the electronic degrees of freedom may be too slow, a shortcoming shared by the surface hopping and the Ehrenfest method alike. A number of semiclassical approaches incorporating decoherence have, therefore, been proposed.^{21–27} Some of these alternative methods attempt to combine the advantages of surface hopping (mainly, pure adiabatic states in asymptotic regions) with those of the mean-field method (no discontinuities in potential energy, no disallowed transitions) by employing an effective potential whilst enforcing gradual demixing of the total wavefunction away from the coupling regions.^{25–27}

4.3 Car-Parrinello Surface Hopping

So far we have assumed that a number of adiabatic potential energy surfaces (at least two) have been obtained by solving the time-independent Schrödinger equation (10) in some unspecified manner. Instead of precalculating the entire PESs, it is advantageous to compute the electronic energies and nuclear gradients “on the fly” as the system is propagated along the trajectory. A popular method in this context has been the Diatomics-in-Molecules (DIM)^{28–40} method which cheaply provides the required electronic eigenvalues and atomic forces for a multitude of molecular valence states simultaneously through diagonalisation of the Hamiltonian matrix. However, although the DIM method works remarkably well for some simple systems such as cationic rare-gas clusters,^{41–44} it is not generally applicable to more complex systems.

For ground state calculations, density functional theory^{45–47} based *ab initio* MD in the spirit of Car and Parrinello⁴⁸ has become the method of choice to study large molecules and condensed phase systems. Recently, Car-Parrinello simulations have become possible also in the first excited singlet state using a restricted open-shell Kohn-Sham (ROKS) approach.⁴⁹ We now report here of a Tully-style¹⁰ trajectory surface hopping method coupling nonadiabatically the S_0 ground state and the S_1 excited state accessible within the Car-Parrinello framework.⁵⁰

4.3.1 Restricted Open-Shell Kohn-Sham Method

Let us first take a brief look at the ROKS method for the S_1 state. Starting from a closed-shell ground state, S_0 , consider an excitation of an electron out of the HOMO into the LUMO. The resulting two unpaired spins can be arranged in four different ways, as illustrated in fig. 7, parallel spins forming triplet determinants and antiparallel spins being equal mixtures of singlet and triplet determinants. The S_1 singlet wavefunction, ϕ_1 , is

constructed as

$$\begin{aligned}\phi_1 &= \frac{1}{\sqrt{2}} \{ |m_1\rangle + |m_2\rangle \} \\ &= \frac{1}{\sqrt{2}} \left\{ |\varphi_1^{(1)} \bar{\varphi}_1^{(1)} \varphi_2^{(1)} \bar{\varphi}_2^{(1)} \cdots \varphi_l^{(1)} \bar{\varphi}_{l+1}^{(1)}\rangle \right. \\ &\quad \left. + |\varphi_1^{(1)} \bar{\varphi}_1^{(1)} \varphi_2^{(1)} \bar{\varphi}_2^{(1)} \cdots \bar{\varphi}_l^{(1)} \varphi_{l+1}^{(1)}\rangle \right\} \quad ,\end{aligned}\quad (52)$$

where the “ket” notation signifies Slater determinants made up of Kohn-Sham orbitals, $\varphi_i^{(1)}$ (spin up) and $\bar{\varphi}_i^{(1)}$ (spin down); $l = \frac{n}{2}$ is half the number of electrons.

It has been shown by Ziegler et al.⁵¹ that the S_1 energy, $E(S_1)$, can be written as the difference between twice the energy of the mixed determinant, $E(m)$, and the energy of the triplet determinant, $E(t)$,

$$E(S_1) = 2E(m) - E(t) \quad . \quad (53)$$

Within the ROKS scheme, a *single* set of orbitals $\{\varphi_i^{(1)}\}$ is determined that minimises the energy functional,

$$E[\{\varphi_i^{(1)}\}] = 2\langle m | \mathcal{H}^{\text{KS}} | m \rangle - \langle t | \mathcal{H}^{\text{KS}} | t \rangle - \sum_{i,j=1}^{l+1} \lambda_{ij} \left\{ \langle \varphi_i^{(1)} | \varphi_j^{(1)} \rangle - \delta_{ij} \right\} \quad , \quad (54)$$

where \mathcal{H}^{KS} is the Kohn-Sham Hamiltonian⁴⁷ and the λ_{ij} are Lagrange multipliers taking care of the orthonormality of the orbitals.

Due to this optimisation the entire set of orbitals $\{\varphi_i^{(1)}\}$ will, in general, differ from the set of orbitals $\{\varphi_i^{(0)}\}$ that define the ground state wavefunction, ϕ_0 ,

$$\phi_0 = |\varphi_1^{(0)} \bar{\varphi}_1^{(0)} \varphi_2^{(0)} \bar{\varphi}_2^{(0)} \cdots \varphi_l^{(0)} \bar{\varphi}_l^{(0)}\rangle \quad . \quad (55)$$

As a consequence the two state functions, ϕ_0 and ϕ_1 , are nonorthogonal giving rise to the overlap matrix elements, S_{ij} ,

$$S_{01} = S_{10} \equiv S \quad , \quad S_{ii} = 1 \quad . \quad (56)$$

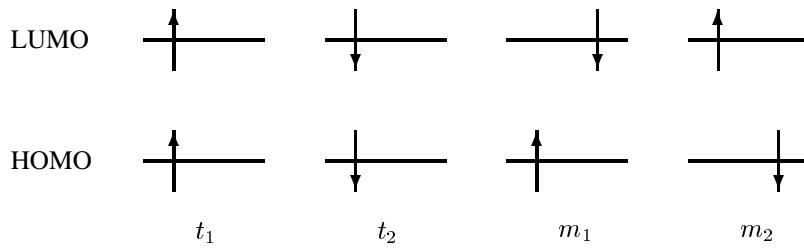


Figure 7. Four possible spin configurations upon excitation of one electron out of the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO). The two parallel spin configurations, t_1 and t_2 form triplet determinants, while the two antiparallel configurations, m_1 and m_2 form mixed determinants with equal singlet and triplet contributions.

4.3.2 S_0 - S_1 Surface Hopping

Inserting ansatz (18) using the above basis functions, ϕ_0 and ϕ_1 , into eqn (17) and replacing \mathcal{H}_{el} with \mathcal{H}^{KS} we obtain after integration over the electronic coordinates following multiplication by ϕ_i^* from the left

$$\sum_j a_j p_j (H_{ij} - E_j S_{ij}) = i\hbar \left\{ \sum_j \dot{a}_j p_j S_{ij} + \sum_j a_j p_j C_{ij} \right\} , \quad (57)$$

where the Hamiltonian matrix elements are given by

$$H_{ii} = \langle \Phi_i | \mathcal{H}^{\text{KS}} | \Phi_i \rangle = E_i , \quad (58)$$

$$H_{01} = H_{10} = E_0 S , \quad (59)$$

and the phase factor has been abbreviated as

$$p_j \equiv e^{-\frac{i}{\hbar} \int E_j dt} . \quad (60)$$

We should stress here that the discrepancy between eqns (57) and (19) arises purely because ϕ_1 is *not* an eigenfunction of \mathcal{H}^{KS} .

For $i = 0$, eq. (57) thus becomes

$$a_1 p_1 S (E_0 - E_1) = i\hbar \{ \dot{a}_0 p_0 + \dot{a}_1 p_1 S + a_1 p_1 C_{01} \} , \quad (61)$$

and for $i = 1$

$$0 = \dot{a}_0 p_0 S + \dot{a}_1 p_1 + a_0 p_0 C_{10} . \quad (62)$$

Solving equations (61) and (62) for \dot{a}_0 and \dot{a}_1 one finds

$$\dot{a}_0 = \frac{1}{S^2 - 1} \left[i a_1 \frac{p_1}{p_0} S (E_0 - E_1) + a_1 C_{01} \frac{p_1}{p_0} - a_0 C_{10} S \right] , \quad (63)$$

$$\dot{a}_1 = \frac{1}{S^2 - 1} \left[a_0 C_{10} \frac{p_0}{p_1} - a_1 C_{01} S - i a_1 S^2 (E_0 - E_1) \right] . \quad (64)$$

We integrate these two coupled differential equations numerically using a fourth order Runge-Kutta scheme.⁵² It is computationally attractive to work with the nonadiabatic coupling elements, C_{ij} (eqn (20)), instead of the nonadiabatic coupling vectors, \mathbf{d}_{ji} (eqn (33)), since the orbital velocities are readily available within the Car-Parrinello method.

If both electronic state functions were eigenfunctions of the Kohn-Sham Hamiltonian, $|a_0|^2$ and $|a_1|^2$ would be their respective occupation numbers. A look at the normalisation integral of the total wavefunction Φ ,

$$\langle \Phi | \Phi \rangle = |a_0|^2 + |a_1|^2 + 2S \Re \left(a_0^* a_1 \frac{p_1}{p_0} \right) \equiv 1 , \quad (65)$$

shows that the definition of state populations in this basis is ambiguous. We therefore expand the total wavefunction Φ in terms of an orthonormal set of auxiliary wavefunctions, ϕ'_0 and ϕ'_1 :

$$\Phi = d_0 \phi'_0 + d_1 \phi'_1 = b_0 \phi_0 + b_1 \phi_1 , \quad (66)$$

where

$$\langle \phi'_i | \phi'_j \rangle = \delta_{ij} \quad (67)$$

and

$$b_j = a_j p_j \quad . \quad (68)$$

Since Φ is normalised, the squares of our new expansion coefficients add up to unity and thus have the meaning of state populations in the orthogonal basis:

$$|d_0|^2 + |d_1|^2 = 1 \quad . \quad (69)$$

The orthonormal wavefunctions ϕ'_0 and ϕ'_1 can be expressed in terms of ϕ_0 and ϕ_1 as

$$\phi'_0 = c_{00}\phi_0 + c_{10}\phi_1 \quad , \quad (70)$$

$$\phi'_1 = c_{01}\phi_0 + c_{11}\phi_1 \quad , \quad (71)$$

$\mathbf{c}_0 = \begin{pmatrix} c_{00} \\ c_{10} \end{pmatrix}$ and $\mathbf{c}_1 = \begin{pmatrix} c_{01} \\ c_{11} \end{pmatrix}$ being solutions of the eigenvalue problem

$$\mathbf{H}\mathbf{C} = \mathbf{SCE} \quad . \quad (72)$$

Using the Hamiltonian matrix elements of eqns (58) and (59) and the overlap matrix of eq (56), one obtains the eigenvalues

$$e_0 = E_0 \quad (73)$$

and

$$e_1 = \frac{E_1 - S^2 E_0}{1 - S^2} \quad (> E_1, \text{ if } E_0 < E_1) \quad . \quad (74)$$

The corresponding eigenvectors are

$$\mathbf{c}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (75)$$

and

$$\mathbf{c}_1 = \begin{pmatrix} -S \\ 1 \end{pmatrix} \quad (76)$$

leading to the orthonormal wavefunctions

$$\phi'_0 = \phi_0 \quad , \quad (77)$$

$$\phi'_1 = \frac{1}{\sqrt{1 - S^2}} [-S\phi_0 + \phi_1] \quad . \quad (78)$$

Inserting (77) and (78) into (66) we determine the expansion coefficients to be

$$d_0 = b_0 + b_1 S \quad (79)$$

$$d_1 = b_1 \sqrt{1 - S^2} \quad . \quad (80)$$

The state occupation numbers are thus

$$|d_0|^2 = |b_0|^2 + S^2 |b_1|^2 + 2S \Re(b_0^* b_1) \quad , \quad (81)$$

$$|d_1|^2 = (1 - S^2)|b_1|^2 \quad (82)$$

or alternatively

$$|d_0|^2 = |a_0|^2 + S^2|a_1|^2 + 2S \Re \left(a_0^* a_1 \frac{p_1}{p_0} \right) \quad (83)$$

$$|d_1|^2 = (1 - S^2)|a_1|^2 \quad . \quad (84)$$

We are now in a position to apply Tully's fewest switches criterion (49) using the coefficients d_i to construct the density matrix (41).

4.3.3 Example: Photoisomerisation of Formaldimine

Figure 8 shows a schematic view of the photoreaction pathways of formaldimine. The reactant, R, is excited vertically from the ground state minimum into the S_1 state to form R^* . Subsequently, the system moves along the reaction coordinate, which predominantly involves an out-of-plane twist of the NH bond, into a conical intersection located at

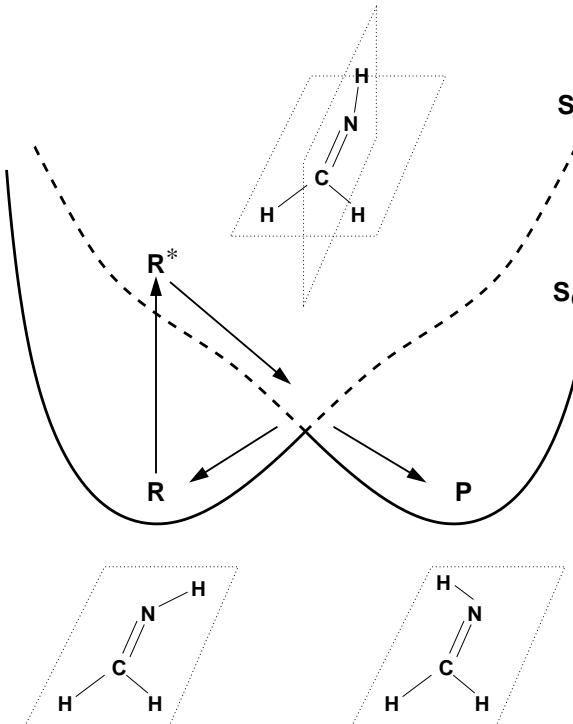


Figure 8. Schematic view of the photoreaction pathways of formaldimine. S_0 and S_1 energy curves are plotted against the reaction coordinate whose main contributor is the NH twist angle. The reactant R is vertically excited from the ground state into the S_1 state to form R^* . The system then falls into a conical intersection where relaxation to the ground state occurs. The reaction can proceed to either of the equivalent isomers, R and P. Formation of the photoproduct P corresponds to photoisomerisation.

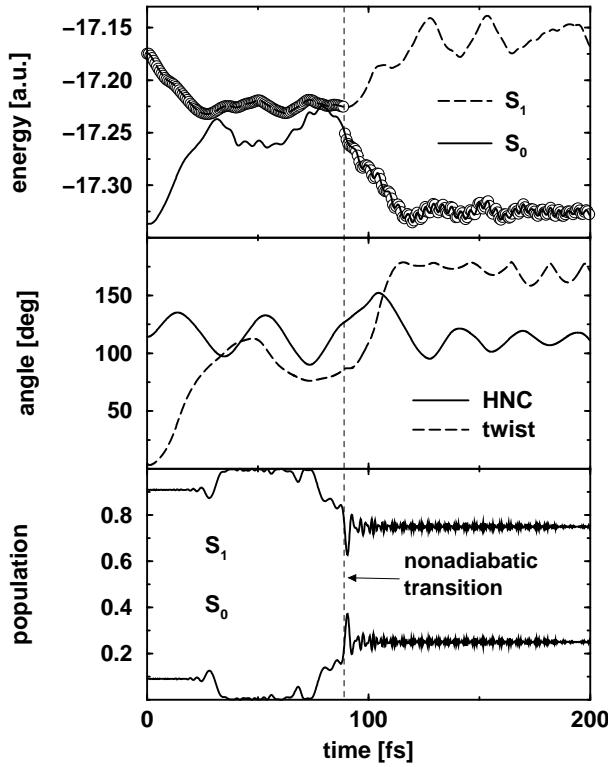


Figure 9. Top: Time evolution of S_0 and S_1 energies following photoexcitation in the case of a $R \rightarrow P$ reaction. The dashed line indicates the moment of the nonadiabatic transition to the ground state. Middle: Corresponding time evolution of the HNC and the HN twist angles. The HNC bond is seen to flip from 0° to 180° resulting in the photoproduct. For HNC angles around 106.5° at orthogonal twist geometry, the energy gap is seen to be minimal. Bottom: Corresponding adiabatic state populations, $|d_i(t)|^2$, of the orthogonal, auxiliary basis set (see eqn (66)).

orthogonal twist geometry. In this region of strong nonadiabatic coupling a transition to the ground state occurs leading either to the photoisomerisation product, P , or back to the reactant R .

We have picked 10 starting configurations at random from a ground state MD run at 300 K. For each of the two possible outcomes, i.e. $R \rightarrow P$ and $R \rightarrow R$, a typical trajectory is analysed in figs 9 and 10. The top panel of fig. 9 shows the evolution of the S_0 and S_1 energies as a function of simulation time for a trajectory leading to the photoproduct P . After vertical excitation of the molecule at $t = 0$, the system is seen to quickly move down into the S_1 potential well dramatically reducing the energy gap to the ground state. As illustrated by the middle panel of fig. 9, the main contribution to the S_1 energy reduction is due to NH twist angle changing from near planarity (0°) to orthogonality (90°). Near the minimum of the S_1 energy curve, where the nonadiabatic coupling is strongest, a nonadiabatic transition to the S_0 state occurs leading to rapid widening of the energy gap accompanied by a change in the twist angle from around 90° to near 180° . It is unclear

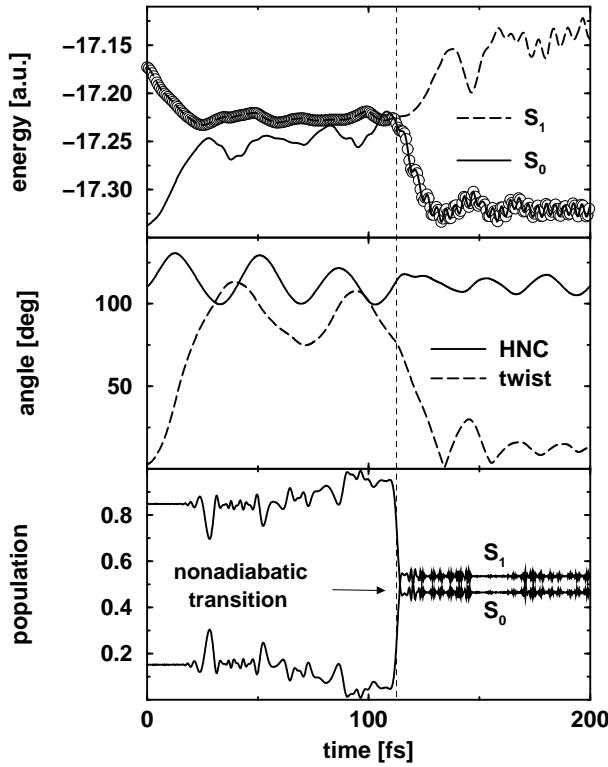


Figure 10. Top: Time evolution of S_0 and S_1 energies following photoexcitation in the case of a $R \rightarrow R$ event. The dashed line indicates the moment of the nonadiabatic transition to the ground state, which occurs shortly after the two states have actually crossed. Middle: Corresponding time evolution of the HNC and the HN twist angles. The HN bond is seen to orthogonalise initially and later flip back to 0° . Bottom: Corresponding adiabatic state populations, $|d_i(t)|^2$, of the orthogonal, auxiliary basis set (see eqn (66)).

what the role of the HNC angle is for this specific trajectory. From static calculations we know that small HNC angles energetically favour the S_1 state over the S_0 state at orthogonal twist geometry,^{50,53} the surface crossing being located at an HNC angle of roughly 106.5° . This observation is nicely confirmed by the fact that the energy gap in fig. 9 is minimal at simulation times when the molecule has approximately that geometry. As one would expect these locations also coincide with the peaks in the adiabatic state population curves, $|d_i(t)|^2$, shown in the bottom panel of fig. 9. After the surface switch has occurred the state occupations may be regarded as constant if one averages over high frequency oscillations. On the other hand, this noise can cause classically forbidden transitions resulting in discrepancy between the actual state distribution of trajectories and the semiclassical populations, $|d_i(t)|^2$.

In the case of the $R \rightarrow R$ event examined in fig 10, the situation is very similar with the exception of the fact that the nonadiabatic surface hop occurs one HNC vibrational period later. Furthermore, the NH twist angle relaxes back to near 0° after initial

orthogonalisation signifying an unsuccessful photoisomerisation attempt. It seems to be a common feature of all trajectories that after a surface hop has taken place, the state population, $|d_1(t)|^2$, of the S_1 state continues to fall sharply excluding the possibility of a transition back into the upper state according to eqn (43). By the time the S_1 occupation begins to increase again, the energy gap has grown significantly making a transition extremely unlikely before the system leaves the classically accessible region.

It is possible, in principle, to determine the quantum yield of photoisomerisation by averaging over an ensemble of surface hopping trajectories. Since this would be beyond the scope of this article, we can only state here our non-converged result of 70 %.

Acknowledgments

I should like to thank Prof. D. Marx for his support. Profs V. Staemmler and J. Hutter as well as Drs I. Frank and M. Odelius are kindly acknowledged for stimulating discussions on the Car-Parrinello surface hopping project. I am grateful to the many colleagues at Ruhr-Universität Bochum whose feedback was crucial in preparing this manuscript.

References

1. M. D. Hack and D. G. Truhlar. *J. Phys. Chem. A*, 104:7917, 2000.
2. M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Clarendon Press, Oxford, 1987.
3. D. Frenkel and B. Smit. *Understanding Molecular Simulation. From Algorithms to Applications*. Academic Press, Boston, 1996.
4. P. Ehrenfest. *Z. Phys.*, 45:455, 1927.
5. H.-D. Meyer and W. H. Miller. *J. Chem. Phys.*, 70:3214, 1979.
6. D. A. Micha. *J. Chem. Phys.*, 78:7138, 1983.
7. J. C. Tully. In B. J. Berne, G. Cicotti, and D. F. Coker, editors, *Classical and Quantum Dynamics in Condensed Phase Simulations*. World Scientific, Singapore, 1998.
8. J. C. Tully. In D. L. Thompson, editor, *Modern Methods for Multidimensional Dynamics Computations in Chemistry*. World Scientific, Singapore, 1998.
9. J. C. Tully and R. K. Preston. *J. Chem. Phys.*, 55:562, 1971.
10. J. C. Tully. *J. Chem. Phys.*, 93:1061, 1990.
11. N. C. Blair and D. G. Truhlar. *J. Chem. Phys.*, 79:1334, 1983.
12. P. J. Kuntz. *J. Chem. Phys.*, 95:141, 1991.
13. E. E. Nikitin. In H. Hartmann, editor, *Chemische Elementarprozesse*. Springer, Berlin, 1968.
14. E. E. Nikitin and L. Zülicke. *Theory of Chemical Elementary Processes*. Springer, Berlin, 1978.
15. L. Salem. *Electrons in Chemical Reactions: First Principles*. Wiley, New York, 1982.
16. L. Salem, C. Leforestier, G. Segal, and R. Wetmore. *J. Am. Chem. Soc.*, 97:479, 1975.
17. J. C. Tully, M. Gomez, and M. Head-Gordon. *J. Vac. Sci. Technol.*, A11:1914, 1993.

18. U. Müller and G. Stock. *J. Chem. Phys.*, 107:6230, 1997.
19. S. Hammes-Schiffer and J. C. Tully. *J. Chem. Phys.*, 101:4657, 1994.
20. D. F. Coker and L. Xiao. *J. Chem. Phys.*, 102:496, 1995.
21. F. Webster, P. J. Rossky, and R. A. Friesner. *Comp. Phys. Comm.*, 63:494, 1991.
22. F. J. Webster, J. Schnitker, M. S. Friedriches, R. A. Friesner, and P. J. Rossky. *Phys. Rev. Lett.*, 66:3172, 1991.
23. E. R. Bittner and P. J. Rossky. *J. Chem. Phys.*, 103:8130, 1995.
24. E. R. Bittner and P. J. Rossky. *J. Chem. Phys.*, 107:8611, 1997.
25. M. D. Hack and D. G. Truhlar. *J. Chem. Phys.*, 114:2894, 2001.
26. M. D. Hack and D. G. Truhlar. *J. Chem. Phys.*, 114:9305, 2001.
27. Y. L. Volobuev, M. D. Hack, M. S. Topaler, and D. G. Truhlar. *J. Chem. Phys.*, 112:9716, 2000.
28. F. O. Ellison. *J. Am. Chem. Soc.*, 85:3540, 1963.
29. J. C. Tully. *J. Chem. Phys.*, 59:5122, 1973.
30. J. C. Tully. *J. Chem. Phys.*, 58:1396, 1973.
31. J. C. Tully. *J. Chem. Phys.*, 64:3182, 1976.
32. J. C. Tully and C. M. Truesdale. *J. Chem. Phys.*, 65:1002, 1976.
33. P. J. Kuntz and A. C. Roach. *J. Chem. Soc. Faraday Trans.*, 68:259, 1971.
34. E. Steiner, P. R. Certain, and P. J. Kuntz. *J. Chem. Phys.*, 59:47, 1973.
35. P. J. Kuntz. *J. Phys. B*, 19:1731, 1986.
36. A. C. Roach and P. J. Kuntz. *J. Chem. Phys.*, 84:822, 1986.
37. J. C. Tully. In G. A. Segal, editor, *Modern Theoretical Chemistry*, Vol. 7A. Plenum Press, New York, 1977.
38. P. J. Kuntz. In R. B. Bernstein, editor, *Atom-Molecule Collision Theory*. Plenum Press, New York, 1979.
39. P. J. Kuntz. In M. Baer, editor, *Theory of Chemical Reaction Dynamics*, Vol. 1. Chemical Rubber, Boca Raton, 1985.
40. P. J. Kuntz. In Z. B. Maksic, editor, *Theoretical Models of Chemical Bonding*, Part 2. Springer, Berlin, 1990.
41. N. L. Doltsinis and P. J. Knowles. *Chem. Phys. Lett.*, 325:648, 2000.
42. N. L. Doltsinis. *Mol. Phys.*, 97, 1999.
43. N. L. Doltsinis and P. J. Knowles. *Chem. Phys. Lett.*, 301:241, 1999.
44. N. L. Doltsinis, P. J. Knowles, and F. Y. Naumkin. *Mol. Phys.*, 96:749, 1999.
45. R. G. Parr and W. Yang. *Density Functional Theory of Atoms and Molecules*. Oxford University Press, Oxford, 1989.
46. P. Hohenberg and W. Kohn. *Phys. Rev. B*, 136:864, 1964.
47. W. Kohn and L. J. Sham. *Phys. Rev. A*, 140:1133.
48. R. Car and M. Parrinello. *Phys. Rev. Lett.*, 55:2471, 1985.
49. I. Frank, J. Hutter, D. Marx, and M. Parrinello. *J. Chem. Phys.*, 108:4060, 1998.
50. N. L. Doltsinis and D. Marx. to be submitted.
51. T. Ziegler, A. Rauk, and E. J. Baerends. *Theor. Chim. Acta*, 43:261, 1977.
52. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flanner. *Numerical Recipes in Fortran 77*, volume 1. Cambridge University Press, 1999.
53. V. Bonačić-Koutecký and Josef Michl. *Theor. Chim. Acta*, 68:45, 1985.

Relieving the Fermionic and the Dynamical Sign Problem: Multilevel Blocking Monte Carlo Simulations

Reinhold Egger¹ and Chi H. Mak²

¹ Institut für Theoretische Physik, Heinrich-Heine-Universität
40225 Düsseldorf, Germany
E-mail: egger@thphy.uni-duesseldorf.de

² Department of Chemistry, University of Southern California
Los Angeles, CA 90089-0482, USA

This article gives an introduction to the multilevel blocking (MLB) approach to both the fermion and the dynamical sign problem in path-integral Monte Carlo simulations. MLB is able to substantially relieve the sign problem in many situations. Besides an exposition of the method, its accuracy and several potential pitfalls are discussed, providing guidelines for the proper choice of certain MLB parameters. Simulation results are shown for strongly interacting electrons in a 2D parabolic quantum dot, the real-time dynamics of several simple model systems, and the dissipative two-state dynamics (spin-boson problem).

1 Introduction: The Sign Problem

Quantum Monte Carlo (QMC) techniques are among the most powerful and versatile methods for the computer simulation of a large variety of interesting quantum systems encountered in physics, chemistry or biology.¹ In particular, QMC is capable of delivering numerically exact results. Despite the great potential of this method, there are several restrictions and handicaps inherent to all QMC techniques, the perhaps most pressing one being due to the *sign problem*.² There are various sign problems, namely the *fermionic sign problem* encountered in equilibrium (imaginary-time) simulations of strongly correlated many-fermion systems, and the *dynamical sign problem* in real-time (dynamical) simulations, which already shows up for a single particle. Unfortunately, apart from variational or approximate treatments (such as the fixed-node approximation), a completely general and totally satisfactory solution to the sign problem in QMC simulations is still lacking. Nevertheless, over the past few years considerable and substantial progress has been achieved along several different lines without introducing approximations into the QMC scheme, see, for instance, Refs.^{3–5}

In these notes focus is put on one specific class of QMC methods called *Path-integral Monte Carlo (PIMC)*. PIMC is based on a discretized path-integral representation of the quantities of interest. The sign problem then arises when different paths that contribute to averages carry different sign (or complex-valued phase). For instance, for the fermionic sign problem, as a consequence of exchange, one typically has to deal with determinants, so that non-positive-definite fermionic density matrix elements arise. The sign cancellations arising from sampling fermion paths then manifest themselves as a signal-to-noise ratio, $\eta \sim \exp(-N\beta E_0)$, that vanishes exponentially with both particle number N and inverse temperature $\beta = 1/k_B T$; here E_0 is a system-dependent energy scale. The small signal surviving the interference of many fermion paths is then inevitably lost in

the large background noise of the stochastic simulation. Similarly, when studying the real-time dynamics of even a single particle, the quantum-mechanical time evolution operator $\exp(-iHt/\hbar)$ attaches a complex-valued phase to each quantum path, which in turn gives rise to the dynamical sign problem. Again the signal-to-noise ratio will vanish exponentially, $\eta \sim \exp(-t^*/\tau_0)$, where t^* is the maximum real time under study and τ_0 a system- and implementation-specific characteristic time scale. The exponential scaling is typical for the “naive” approach, where one simply uses the absolute value of the complex-valued weight function for the MC sampling, and includes the phase information in the accumulation.

In Ref.³ a general scheme for tackling the sign problem in PIMC simulations was proposed. The method has been applied to interacting electrons in a quantum dot⁶ and to the real-time dynamics of simple few-degrees-of-freedom systems.⁷ The generalization of the algorithm to the case of effective actions – which are long-ranged along the Trotter or real-time direction – along with the application to the dissipative two-state (spin-boson) dynamics has been given in Refs.^{8,9} This *multilevel blocking* (MLB) approach represents the systematic implementation of a simple *blocking strategy*. The blocking strategy states that by sampling groups of paths (“blocks”) at the same time, the sign problem can always be reduced compared to sampling single paths as would be done normally; for a proof, see Sec. 2.1 below. By suitably bunching paths together into sufficiently small blocks, the sign cancellations among paths within the same block can be accounted for without the sign problem, simply because there is no sign problem for a sufficiently small system. The MLB approach is based on a recursive implementation of this idea, i.e. after building “elementary” blocks, new blocks are formed out of these, and the process is then iterated. This leads to a powerful implementation of the blocking strategy. This algorithm is able to turn the exponential severity of the sign problem into an “only” algebraic one. This is still difficult enough but in practice implies that significantly larger systems (lower temperature, longer real time) can be studied by MLB-PIMC than under the naive PIMC approach. Nevertheless, it should be stressed that MLB is definitely *not* a black-box scheme. There are several potential pitfalls related to incorrect or inadequate choices of certain MLB parameters, and one needs to be quite careful in applying this technique.¹⁰ Given some experience, however, it represents a powerful handle to relieve the sign problem, with potential for additional improvement.

The plan for the remainder of these notes is as follows. In Sec. 2 the MLB approach is discussed in detail, first in an intuitive way in Sec. 2.1, and subsequently on a more formal or technical level in Sec. 2.2 for fermions. The modifications for real-time simulations are summarized in Sec. 2.3. This is then followed by a discussion of the performance and the accuracy of this algorithm in Sec. 2.4. In Sec. 2.5, the generalization to effective-action problems is outlined. To demonstrate the power of this approach, MLB results are presented in Sec. 3. As an example for the fermionic sign problem, low-temperature simulation results for strongly correlated electrons in a 2D quantum dot are shown in Sec. 3.1. The remainder of that section is then concerned with the dynamical sign problem. After presenting results for several simple model systems in Sec. 3.2, the dynamics of the dissipative two-state system is discussed in Sec. 3.3. Finally, some conclusions can be found in Sec. 4.

2 Multilevel Blocking (MLB) Approach

Before diving into the details of the MLB approach, the underlying idea (“blocking strategy”) will be explained, focusing for simplicity on fermionic imaginary-time simulations. For those interested in working with this method, technical details and practical guidelines are provided in Secs. 2.2 to 2.4. In the last part the generalization of MLB to PIMC simulations of the effective-action type is described.

2.1 Blocking Strategy

Let us consider a many-fermion system whose state is described by a set of quantum numbers \vec{r} denoting, e.g. the positions and spins of *all* particles. These quantum numbers can correspond to electrons living on a lattice or in continuous space. For notational simplicity, we focus on calculating the equilibrium expectation value of a diagonal operator or correlation function (this can be easily generalized),

$$\langle A \rangle = \frac{\sum_{\vec{r}} A(\vec{r}) \rho(\vec{r}, \vec{r})}{\sum_{\vec{r}} \rho(\vec{r}, \vec{r})}, \quad (1)$$

where $\sum_{\vec{r}}$ represents either a summation for the case of a discrete system or an integration for a continuous system, and $\rho(\vec{r}, \vec{r}')$ denotes the (reduced) density matrix of the system. In PIMC applications, imaginary time is discretized into P slices of length $\epsilon = \beta/P$. Inserting complete sets at each slice $m = 1, \dots, P$, and denoting the corresponding configuration on slice m by \vec{r}_m , the diagonal elements of the density matrix at $\vec{r} = \vec{r}_P$ entering Eq. (1) are:

$$\rho(\vec{r}_P, \vec{r}_P) = \sum_{\vec{r}_1, \dots, \vec{r}_{P-1}} \prod_{m=1}^P \langle \vec{r}_{m+1} | e^{-\epsilon H} | \vec{r}_m \rangle. \quad (2)$$

Of course, periodic boundary conditions have to be enforced here, $\vec{r}_{P+1} = \vec{r}_1$. To proceed, one then has to construct accurate analytical approximations for the short-time propagator. This formulation of the problem excludes effective actions such as those arising from an integration over the fermions using the Hubbard-Stratonovich transformation,² since that generally leads to long-ranged imaginary-time interactions. The MLB approach suitable for such a situation⁸ is described below in Sec. 2.5.

Since we are dealing with a many-fermion system, the short-time propagators need to be antisymmetrized, leading to the appearance of determinants causing the sign problem. Strictly speaking, the antisymmetrization has to be done only on one time slice, but the intrinsic sign problem is much better behaved if one antisymmetrizes on all time slices. Choosing the absolute value of the product of the short-time propagators in Eq. (2) as the positive definite MC weight function $P[X]$, one has to keep the sign $\Phi[X]$ associated with a particular discretized path $X = (\vec{r}_1, \dots, \vec{r}_P)$ during the accumulation procedure,

$$\langle A \rangle = \frac{\sum_X P[X] \Phi[X] A[X]}{\sum_X P[X] \Phi[X]}. \quad (3)$$

Assuming that there are no further exclusivity problems in the numerator so that $A[X]$ is well-behaved, one can then analyze the sign problem in terms of the variance of the

denominator,

$$\sigma^2 \approx \frac{1}{N_s} (\langle \Phi^2 \rangle - \langle \Phi \rangle^2) , \quad (4)$$

where N_s is the number of MC samples taken and stochastic averages are calculated with P as the weight function. For the fermion sign problem, where $\Phi = \pm 1$ and hence $\langle \Phi^2 \rangle = 1$, the variance of the signal is controlled by the size of $|\langle \Phi \rangle|$.

Remarkably, one can achieve considerable progress by *blocking paths together*. Instead of sampling single paths along the MC trajectory, one can consider sampling sets of paths called blocks. Under such a blocking operation, the stochastic estimate for $\langle A \rangle$ takes the form

$$\langle A \rangle = \frac{\sum_B (\sum_{X \in B} P[X] \Phi[X] A[X])}{\sum_B (\sum_{X \in B} P[X] \Phi[X])} , \quad (5)$$

where one first sums over the configurations belonging to a block B in a way that is not affected by the sign problem, and then stochastically sums over the blocks. The summation within a block must therefore be done non-stochastically, or alternatively the block size must be chosen sufficiently small. Of course, there is considerable freedom in how to choose this blocking.

Let us analyze the variance σ'^2 of the denominator of Eq. (5). First define new sampling functions in terms of the blocks which are then sampled stochastically,

$$P'[B] = \left| \sum_{X \in B} P[X] \Phi[X] \right| , \quad \Phi'[B] = \text{sgn} \left(\sum_{X \in B} P[X] \Phi[X] \right) . \quad (6)$$

Rewriting the average sign in the new representation, i.e. using $P'[B]$ as the weight, then inserting the definition of P' and Φ' in the numerator,

$$\langle \Phi'[B] \rangle = \frac{\sum_B P'[B] \Phi'[B]}{\sum_B P'[B]} = \frac{\sum_X P[X] \Phi[X]}{\sum_B P'[B]} ,$$

and comparing to the average sign in the standard representation using $P[X]$ as the weight, one finds

$$\frac{|\langle \Phi' \rangle|}{|\langle \Phi \rangle|} = \frac{\sum_X P[X]}{\sum_B P'[B]} .$$

By virtue of the Schwarz inequality,

$$\sum_B P'[B] = \sum_B \left| \sum_{X \in B} P[X] \Phi[X] \right| \leq \sum_B \left| \sum_{X \in B} P[X] \right| = \sum_X P[X] ,$$

it follows that *for any kind of blocking*, the average sign improves, $|\langle \Phi' \rangle| \geq |\langle \Phi \rangle|$. Furthermore, since $\langle \Phi'^2 \rangle = \langle \Phi^2 \rangle = 1$, Eq. (4) implies that

$$\sigma'^2 \leq \sigma^2 , \quad (7)$$

and hence *the signal-to-noise ratio is always improved upon blocking configurations together*. Clearly, the worst blocking one could possibly choose would be to group the configurations into two separate blocks, one collecting all paths with positive sign and the other with negative sign. In this case, blocking yields no improvement whatsoever, and the

“ \leq ” becomes “ $=$ ” in Eq. (7). It is apparent from Eq. (7) that the blocking strategy provides a systematic handle to reduce the sign problem. In our realization of the blocking strategy, a block is defined by all paths that differ only at one time slice, i.e. only \vec{r}_m is updated with all other $\vec{r}_{n \neq m}$ being frozen.

A direct implementation of the blocking strategy does indeed improve the sign problem but will not remove its exponential severity. The reason is simply that for a sufficiently large system, there will be too many blocks, and once the signals coming from these blocks are allowed to interfere, one again runs into the sign problem, albeit with a smaller scale E_0 entering the signal-to-noise ratio. The resolution to this problem comes from the multilevel blocking (MLB) approach³ where one *applies the blocking strategy in a recursive manner to the blocks* again. In a sense, new blocks containing a sufficiently small number of elementary ones are formed, and this process is repeated until only one block is left. Each step of this hierarchy is called *level* in what follows. Blocks are then defined living on these *levels*, and after taking care of the sign cancellations within all blocks on a given (fine) level, the resulting sign information is transferred to the next (coarser) level. On each step, the blocking strategy ensures that no sign problem occurs *provided one has chosen sufficiently small block sizes*. By doing this recursively, the sign problem on all the coarser levels can be handled in the same manner. It is then possible to proceed without numerical instabilities from the bottom up to the top level. The cancellations arising at the top level will create a sign problem again, which is however strongly reduced. As is argued below, the resulting sign problem is characterized by an only algebraic severity.

In many ways, the MLB idea is related to the renormalization group approach. But instead of integrating out information on fine levels, sign cancellations are “synthesized” within a given level and subsequently their effects are transferred to coarser levels. While the renormalization group filters out information judged “relevant” and then ignores the “irrelevant” part, no such approximation is introduced here. Therefore our approach is actually closer in spirit to the multi-grid algorithm.¹¹ The technical implementation of MLB is discussed next following Ref.³

2.2 Systematic Implementation: MLB

To keep notation simple, the slice index m is used as a shorthand notation for the quantum numbers \mathbf{r}_m . From Eq. (2) the *level-0 bonds*, which are simply the short-time propagators, then follow in the form:

$$(m, m+1)_0 = \langle \mathbf{r}_{m+1} | e^{-\epsilon H} | \mathbf{r}_m \rangle . \quad (8)$$

Next the different levels $0 \leq \ell \leq L$, where L defines the Trotter number $P = 2^L$, have to be specified. Each slice m belongs to a unique level ℓ , such that $m = (2j+1)2^\ell$ and j is a nonnegative integer. For instance, the slices $m = 1, 3, 5, \dots, P-1$ belong to $\ell = 0$, $m = 2, 6, 10, \dots, P-2$ belong to $\ell = 1$, etc., such that there are $\mathcal{N}_\ell = 2^{L-\ell-1}$ (but $\mathcal{N}_L = 1$) different slices on level ℓ , see Fig. 1. An elementary blocking is achieved by grouping together configurations that differ only at slice m , so only \mathbf{r}_m varies in that block while all $\mathbf{r}_{m' \neq m}$ remain fixed. Sampling on level ℓ therefore extends over configurations $\{\mathbf{r}_m\}$ living on the \mathcal{N}_ℓ different slices. In the MLB scheme, one moves recursively from the finest ($\ell = 0$) up to the coarsest level ($\ell = L$), and the measurement of the diagonal operator is done only at the top level using the configuration \mathbf{r}_P .

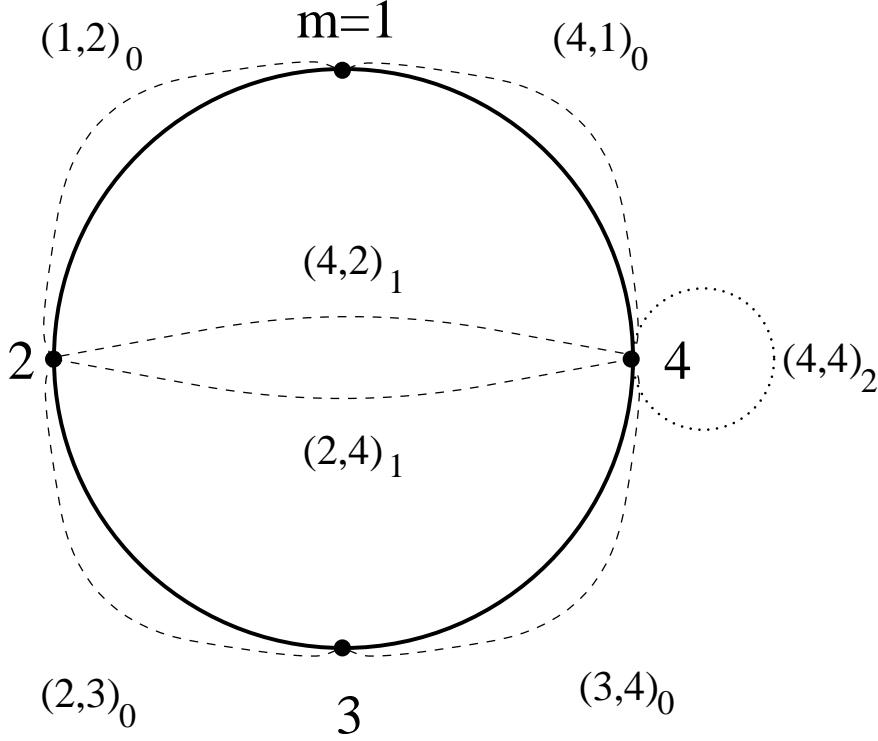


Figure 1. Levels for $L = 2$ ($P = 4$). Imaginary time flows along the circle (solid curve), and the slices $m = 1, 2, 3, 4$ are distributed among the three levels: The finest level $\ell = 0$ contains $m = 1, 3$, level $\ell = 1$ contains $m = 2$, and $\ell = 2$ contains $m = 4$. Level- ℓ bonds are indicated by dashed and dotted lines.

A MC sweep starts by changing only configurations associated with the slices on level $\ell = 0$ according to the standard weight

$$\mathcal{P}_0 = |(1, 2)_0 (2, 3)_0 \cdots (P, 1)_0|, \quad (9)$$

generating a MC trajectory containing K samples for each slice on level $\ell = 0$. These $\mathcal{N}_0 K$ samples are stored and they are used to generate additional coarser interactions among the higher-level slices,

$$\begin{aligned} (m, m+2)_1 &= \langle \text{sgn}[(m, m+1)_0 (m+1, m+2)_0] \rangle_{\mathcal{P}_0[m+1]} \\ &= K^{-1} \sum_{m+1} \text{sgn}[(m, m+1)_0 (m+1, m+2)_0], \end{aligned} \quad (10)$$

where the summation \sum_{m+1} extends over the K samples available for slice m . As will be discussed later on, the important MLB parameter K - subsequently called *sample number* - should be chosen as large as possible to ensure that the second equality in Eq. (10) is justified. The *level-1 bonds* (10) contain precious and crucial information about the sign

cancellations that occurred on the previous level $\ell = 0$. Using these bonds, the density matrix (2) is rewritten as

$$\rho(P, P) = \sum_{1,2,\dots,P-1} |(1,2)_0(2,3)_0 \cdots (P,1)_0|(2,4)_1 \cdots (P-2,P)_1(P,2)_1 . \quad (11)$$

Comparing this to Eq. (2), the entire sign problem has been transferred to the next coarser level by using the level-1 bonds.

In the next step, the sampling is carried out on level $\ell = 1$ in order to generate the next-level bonds, i.e. only slices $m = 2, 6, \dots, P - 2$ are updated, using the weight $\mathcal{P}_0 \mathcal{P}_1$ with

$$\mathcal{P}_1 = |(2,4)_1(4,6)_1 \cdots (P,2)_1| . \quad (12)$$

Moving the level-1 configurations modifies the level-0 bonds, which in turn requires that the level-1 bonds be updated. A direct re-calculation of these bonds according to Eq. (10) would be too costly. Instead, the stored configurations on level $\ell = 0$ are used to perform an importance sampling of the new level-1 bonds. Under the test move $m \rightarrow m'$, i.e. $r_m \rightarrow r'_m$, on level $\ell = 1$, the bond (10) can be obtained from

$$(m', m+2)_1 = \frac{\sum_{m+1} \frac{(m', m+1)_0(m+1, m+2)_0}{|(m, m+1)_0(m+1, m+2)_0|}}{\sum_{m+1} \frac{|(m', m+1)_0(m+1, m+2)_0|}{|(m, m+1)_0(m+1, m+2)_0|}} , \quad (13)$$

where \sum_{m+1} runs over the previously stored MC configurations r_{m+1} . Note that for small values of K , Eq. (13) is only approximative, and thus a sufficiently large value of K should be chosen. With the aid of Eq. (13), the updated level-1 bonds follow with only moderate computational effort. Generating a sequence of K samples for each slice on level $\ell = 1$ and storing them, *level-2 bonds* can be calculated in analogy to Eq. (10),

$$(m, m+4)_2 = \langle \text{sgn} [(m, m+2)_1(m+2, m+4)_1] \rangle_{\mathcal{P}_1 \mathcal{P}_0} . \quad (14)$$

Finally, the process is iterated up to the top level $\ell = L$ using the obvious recursive generalization of Eqs. (10) and (14) to define *level- ℓ bonds*.

Thereby the diagonal elements of the density matrix are obtained as

$$\begin{aligned} \rho(P, P) &= \sum_{1,2,\dots,P-1} |(1,2)_0(2,3)_0 \cdots (P,1)_0|(2,4)_1 \cdots (P-2,P)_1(P,2)_1 \\ &\cdots |(P/2,P)_{L-1}(P,P/2)_{L-1}|(P,P)_L . \end{aligned} \quad (15)$$

By virtue of this algorithm, the sign problem is transferred step by step up to the coarsest level. The expectation value (1) can thus be computed from

$$\langle A \rangle = \frac{\langle A(P) \text{sgn}(P, P)_L \rangle_{\mathcal{P}}}{\langle \text{sgn}(P, P)_L \rangle_{\mathcal{P}}} . \quad (16)$$

The manifestly positive definite MC weight \mathcal{P} used for the averaging in Eq. (16) can be read off from Eq. (15),

$$\begin{aligned} \mathcal{P} &= |(1,2)_0(2,3)_0 \cdots (P,1)_0|(2,4)_1 \cdots (P-2,P)_1(P,2)_1 \\ &\cdots |(P/2,P)_{L-1}(P,P/2)_{L-1}|(P,P)_L . \end{aligned} \quad (17)$$

The denominator in Eq. (16) gives the average sign and indicates to what extent the sign problem has been solved. For proper choice of MLB parameters, in particular the sample

number K , this method can solve the sign problem. The price to pay for the stability of the algorithm is the increased memory requirement $\sim K^2$ associated with having to store the sampled configurations.

2.3 Real-Time Simulations

The same method, described so far for fermions, can be applied with minor modifications to a computation of real-time correlation functions or occupation probabilities. For concreteness, let us focus on an equilibrium time-correlation function,

$$\langle A(0)B(t) \rangle = \frac{\text{Tr} \{ e^{-(\beta\hbar+it)H/\hbar} A e^{+itH/\hbar} B \}}{\text{Tr} \{ e^{-(\beta\hbar+it)H/\hbar} e^{+itH/\hbar} \}}. \quad (18)$$

Similarly other dynamical properties like the thermally symmetrized correlation function,¹²

$$C_s(t) = Z^{-1} \text{Tr} \{ e^{-(\beta\hbar/2+it)H/\hbar} A e^{-(\beta\hbar/2-it)H/\hbar} B \}, \quad (19)$$

with Z being the partition function, can be computed. In terms of path integrals, the traces in (18) involve two quantum paths, one propagated backward in time for the duration $-t$ and the other propagated in complex time for the duration $t - i\beta\hbar$. Discretizing each of the two paths into P slices, the entire cyclic path has a total of $2P$ slices. A slice on the first half of them has length $-t/P$ and on the second half $(t - i\beta\hbar)/P$. Denoting the quantum numbers (e.g. spin or position variables) at slice j by \mathbf{r}_j , the correlation function (18) reads

$$\frac{\int d\mathbf{r}_1 \cdots d\mathbf{r}_{2P} B(\mathbf{r}_{2P}) A(\mathbf{r}_P) \prod_{j=1}^{2P} (\mathbf{r}_j, \mathbf{r}_{j+1})_0}{\int d\mathbf{r}_1 \cdots d\mathbf{r}_{2P} \prod_{j=1}^{2P} (\mathbf{r}_j, \mathbf{r}_{j+1})_0}, \quad (20)$$

where the level-0 bond $(\mathbf{r}_j, \mathbf{r}_{j+1})_0$ is again the short-time propagator between slices j and $j + 1$, and $\mathbf{r}_{2P+1} = \mathbf{r}_1$. First assign all slices along the discretized path to different levels $\ell = 0, \dots, L$, where $P = 2^L$, in close analogy to the treatment for fermions, see Fig. 1. Each slice $j = 1, \dots, 2P$ belongs to a unique level ℓ , such that $j = (2k + 1)2^\ell$ and k is a nonnegative integer. For instance, slices $j = 1, 3, 5, \dots$ belong to level $\ell = 0$, slices $j = 2, 6, 10, \dots$ to $\ell = 1$, etc. The MLB algorithm starts by sampling only configurations which are allowed to vary on slices associated with the finest level $\ell = 0$, using the weight $\mathcal{P}_0 = |(\mathbf{r}_1, \mathbf{r}_2)_0 \cdots (\mathbf{r}_{2P}, \mathbf{r}_1)_0|$. The short-time level-0 bonds are then employed to synthesize longer-time level-1 bonds that connect the even- j slices. Subsequently the level-1 bonds are used to synthesize level-2 bonds, and so on. In this way the MLB algorithm moves recursively from the finest level ($\ell = 0$) up to increasingly coarser levels until $\ell = L$, where the measurement is done using \mathbf{r}_{2P} and \mathbf{r}_P .

Generating a MC trajectory containing K samples for each slice on level $\ell = 0$ and storing these samples, the level-1 bonds (10) can be computed, where the “sgn” has to be replaced by the complex-valued phase $\Phi[z] = e^{i\arg(z)}$. Their benefit becomes clear when rewriting the integrand of the denominator in (20) as

$$\mathcal{P}_0 \times (\mathbf{r}_2, \mathbf{r}_4)_1 \cdots (\mathbf{r}_{2P-2}, \mathbf{r}_{2P})_1 (\mathbf{r}_{2P}, \mathbf{r}_2)_1.$$

Comparing this to (20), the entire sign problem has been transferred to the next coarser level. In the next step, the sampling is carried out on level $\ell = 1$ in order to compute the

next-level bonds, using the weight $\mathcal{P}_0\mathcal{P}_1$ with $\mathcal{P}_1 = |(\mathbf{r}_2, \mathbf{r}_4)_1 \cdots (\mathbf{r}_{2P}, \mathbf{r}_2)_1|$. Generating a sequence of K samples for each slice on level $\ell = 1$, and storing these samples, level-2 bonds can be calculated,

$$(\mathbf{r}_j, \mathbf{r}_{j+4})_2 = \langle \Phi [(\mathbf{r}_j, \mathbf{r}_{j+2})_1 (\mathbf{r}_{j+2}, \mathbf{r}_{j+4})_1] \rangle_{\mathcal{P}_0\mathcal{P}_1}.$$

This process is then iterated up to the top level. Finally, the correlation function (18) can be computed from

$$\frac{\langle B(\mathbf{r}_{2P})A(\mathbf{r}_P)\Phi[(\mathbf{r}_P, \mathbf{r}_{2P})_L (\mathbf{r}_{2P}, \mathbf{r}_P)_L] \rangle_{\mathcal{P}}}{\langle \Phi[(\mathbf{r}_P, \mathbf{r}_{2P})_L (\mathbf{r}_{2P}, \mathbf{r}_P)_L] \rangle_{\mathcal{P}}}, \quad (21)$$

with the positive definite MC weight $\mathcal{P} = \mathcal{P}_0\mathcal{P}_1 \cdots \mathcal{P}_L$. The denominator in Eq. (21) gives the *average phase* and indicates to what extent the sign problem has been solved. With a suitable choice of MLB parameters, the average phase remains close to unity even for long times.

2.4 Accuracy and Pitfalls

Next questions concerning the *exactness* and *performance* of the MLB approach are addressed, in particular the dependence on the sample number K . Clearly, K needs to be sufficiently large to produce a reliable estimate for the level- ℓ bonds. If these bonds could be calculated exactly – corresponding to the limit $K \rightarrow \infty$ –, the manipulations leading to Eq. (15) yield the exact result. Hence for large enough K , the MLB algorithm must become exact and completely solve the sign problem. Since the level- ℓ bonds can however only be computed for finite K , the weight function \mathcal{P} amounts to using a noisy estimator, which in turn can introduce *bias* into the algorithm.¹³ In principle, this problem could be avoided by using a linear acceptance criterion instead of the algorithmically simpler Metropolis choice¹ which was employed in the applications reported here. But even with the Metropolis choice, the bias can be made arbitrarily small by increasing K . Therefore, with sufficient computer memory, the MLB approach can be made to give numerically exact results. One might then worry about the actual value of $K > K^*$ required to obtain stable and exact results. If the critical value K^* were to scale exponentially with β and/or system size, the sign problem would be present in disguise again.

Although a rigorous non-exponential bound on K^* has not yet been established, our experience with the MLB algorithm indicates that this scaling is at worst algebraic. This is corroborated by a recent careful study of this issue.¹⁰ Therefore the *exponential severity of the sign problem is replaced by an algebraic one under MLB*. A heuristic argument supporting this statement goes as follows. If one needs K samples for each slice on a given level in order to have satisfactory statistics despite of the sign problem, the total number of paths needed in the naive approach depends exponentially on P , namely $\sim K^P$. This is precisely the well-known exponential severity of the sign problem under the naive approach. However, with MLB the work on the last level, which is the only one affected by a sign problem provided K was chosen sufficiently large, scales only $\sim K^L$. Note that it does not scale $\sim K$ because one must update the level- ℓ bonds on all L finer levels as well. So in MLB, the work needed to sample the K^P paths with satisfactory statistical accuracy grows $\sim K^{\log_2 P} = P^{\log_2 K}$, i.e. algebraically with P . An important point to mention at this point concerns the high-temperature (or short-time) limit, where the direct PIMC

Table 1. MLB results for $N = 8$ and $\lambda = 2$ (see Sec. 3.1). N_s is the number of samples (in 10^4), t_{CPU} the total CPU time (in hours), MB the required memory (in mega-bytes), and $\langle \text{sgn} \rangle$ the average sign. Bracketed numbers are error estimates.

K	N_s	t_{CPU}	MB	$\langle \text{sgn} \rangle$	$E_N / \hbar\omega_0$
1	120	95	2	0.02	48.6(3)
100	7	33	14	0.48	48.43(8)
200	9	83	30	0.63	48.55(7)
400	8	174	64	0.73	48.53(9)
600	10	308	96	0.77	48.54(8)
800	9	429	150	0.81	48.59(8)

simulation does not face a significant sign problem. In this case, however, the above-mentioned bias problematics of the MLB-PIMC is quite serious and can give erroneous results. Fortunately, since that regime is of little interest to MLB, this is not a serious restriction. A more detailed discussion of this point can be found in Ref.¹⁰

To elucidate how the MLB algorithm works in practice, in Table 1 simulation results for $N = 8$ electrons in a quantum dot at various values of K are listed. For details, see Sec. 3.1. Compared to the naive approach where $K = 1$, using a moderate $K = 200$ already increases the average sign from 0.02 to 0.63, making it possible to obtain more accurate results from much fewer samples. The data in Table 1 also confirms that the bias can be systematically eliminated by increasing K , so that the energy found at $K \geq 200$ essentially represents the exact result (within error bars). The value $K^* \approx 200$ is quite typical for many applications. For a simple model system, a value $K^* \approx 50$ was found in Ref.¹⁰ Table 1 also shows that the CPU time per sample scales linearly with K , whereas memory requirements grow $\sim K^2$.

2.5 Effective Actions

So far the MLB algorithm was only discussed for the case of nearest-neighbor interactions along the Trotter/time direction. This situation is encountered under a direct Trotter-Suzuki breakup of the short-time propagator. In many cases, however, one has to deal with effective actions that may include *long-ranged interactions* along the (complex) time direction. Such effective actions arise from degrees of freedoms having been traced out, e.g. a harmonic heat bath,¹⁴ or through a Hubbard-Stratonovich transformation in auxiliary-field MC simulations of lattice fermions.¹ Remarkably, because such effective actions capture much of the physics such as symmetries or the dissipative influence of traced-out degrees of freedom, the corresponding path integral very often exhibits a significantly reduced intrinsic sign problem compared to the original (time-local) formulation. To be specific, let us focus on the dynamical sign problem arising in real-time PIMC computations here. The modifications required to implement the method for fermion simulations are then straightforward.

Let us consider a discretized path integral along a slightly different but fully equivalent contour in the complex-time plane compared to Sec. 2.3, namely a forward branch from $t = 0$ to $t = t^*$, where t^* is the maximum time studied in the simulation, followed by a branch going back to the origin, and then by an imaginary-time branch from $t = 0$

to $t = -i\hbar\beta$. Here a “factorized” initial preparation is studied, where the relevant degrees of freedom, $\mathbf{r}(t)$, are held fixed for $t < 0$.¹⁴ That implies that the imaginary-time dynamics must be frozen at the corresponding value, and one only needs to sample on the two real-time branches. Note that such a nonequilibrium calculation cannot proceed by first doing an imaginary-time QMC simulation followed by a generally troublesome analytic continuation of the numerical data.¹ Using time slices of length t^*/P , forward [$\mathbf{r}(t_m)$] and backward [$\mathbf{r}'(t_m)$] path configurations at time $t_m = mt^*/P$ are combined into the configuration s_m , where $m = 1, \dots, P$. The configuration at $t = 0$ is held fixed, and for $t = t^*$ one must be in a diagonal state, $\mathbf{r}(t^*) = \mathbf{r}'(t^*)$. For an efficient application of the current method, it is essential to combine several neighboring slices m into new blocks. For instance, think of $m = 1, \dots, 5$ as a new “slice” $\ell = 1$, $m = 6, \dots, 10$ as another slice $\ell = 2$, and so on. Combining q elementary slices into a block s_ℓ , instead of the original P slices one has $L = P/q$ blocks, where L is the number of MLB levels. Instead of the “circular” structure of the time contour inherent in the trace operation, it is actually more helpful to view the problem as a linear chain, where the MLB scheme proceeds from left to right. In actual applications, there is considerable freedom in how these blocks are defined, e.g. if there is hardly any intrinsic sign problem, or if there are only few variables in \mathbf{r} , one may choose larger values of q . Additional flexibility can be gained by choosing different q for different blocks.

Say one is interested in sampling the configurations s_L on the top level $\ell = L$ according to the appropriate matrix elements of the (reduced) density matrix,

$$\rho(s_L) = Z^{-1} \sum_{s_1, \dots, s_{L-1}} \exp\{-S[s_1, \dots, s_L]\}, \quad (22)$$

where S is the effective action under study and Z is a normalization constant so that $\sum_{s_L} \rho(s_L) = 1$. Due to the time-non-locality of this action, there will be interactions among all blocks s_ℓ . The sum in Eq. (22) denotes either an integration over continuous degrees of freedom or a discrete sum. In the case of interest here, the effective action is complex-valued and $e^{-S}/|e^{-S}|$ represents an oscillatory phase factor Φ . The “naive approach” to the sign problem is to sample configurations using the positive definite weight function $\mathcal{P} \sim |\exp\{-S\}|$, and to include the oscillatory phase in the accumulation procedure. Below it is assumed that one can decompose the effective action according to

$$S[s_1, \dots, s_L] = \sum_{\ell=1}^L W_\ell[s_\ell, \dots, s_L]. \quad (23)$$

All dependence on a configuration s_ℓ is then contained in the “partial actions” W_λ with $\lambda \leq \ell$. One could, of course, put all $W_{\ell>1} = 0$, but the approach becomes more powerful if a nontrivial decomposition is possible.

Let us now describe the algorithm in some detail, employing a somewhat different but equivalent notation than before. This may be helpful to some readers in order to better understand the MLB algorithm, see also Ref.¹⁰ for a formulation of Sec. 2.2 in this notation. The MC sampling starts on the finest level $\ell = 1$, where only the configuration $s_{\ell=1}$ containing the elementary slices $m = 1, \dots, q$ will be updated with all $s_{\ell>1}$ remaining fixed at their initial values s_ℓ^0 . Using the weight function

$$\mathcal{P}_0[s_1] = |\exp\{-W_1[s_1, s_2^0, \dots, s_L^0]\}|,$$

generate K samples $s_1^{(i)}$, where $i = 1, \dots, K$, and store them for later use. As usual, the sample number K should be chosen large enough. For $K = 1$, the algorithm will simply reproduce the naive approach. The stored samples are now employed to generate information about the sign cancellations. All knowledge about the interference that occurred at this level is encapsulated in the quantity

$$\begin{aligned} B_1 &= \left\langle \frac{\exp\{-W_1[s_1, \dots, s_L]\}}{|\exp\{-W_1[s_1, s_2^0, \dots, s_L^0]\}|} \right\rangle_{\mathcal{P}_0[s_1]} \\ &= C_0^{-1} \sum_{s_1} \exp\{-W_1[s_1, \dots, s_L]\} \\ &= K^{-1} \sum_{i=1}^K \frac{\exp\{-W_1[s_1^{(i)}, s_2, \dots, s_L]\}}{|\exp\{-W_1[s_1^{(i)}, s_2^0, \dots, s_L^0]\}|} = B_1[s_2, \dots, s_L], \end{aligned} \quad (24)$$

which are analogously called level-1 bonds, with the normalization constant $C_0 = \sum_{s_1} \mathcal{P}_0[s_1]$. Combining the second expression in Eq. (24) with Eq. (22), the density matrix reads

$$\rho(s_L) = Z^{-1} \sum_{s_2, \dots, s_{L-1}} \exp \left\{ - \sum_{\ell>1} W_\ell \right\} C_0 B_1 = Z^{-1} \sum_{s_1, \dots, s_{L-1}} \mathcal{P}_0 B_1 \prod_{\ell>1} e^{-W_\ell}. \quad (25)$$

When comparing Eq. (25) with Eq. (22), the sign problem has now been transferred to levels $\ell > 1$, since oscillatory phase factors only arise when sampling on these higher levels. Note that $B_1 = B_1[s_2, \dots, s_L]$ introduces couplings among *all* levels $\ell > 1$, in addition to the ones already contained in the effective action S .

Next proceed to the next level $\ell = 2$ and, according to Eq. (25), update configurations for $m = q + 1, \dots, 2q$ using the weight

$$\mathcal{P}_1[s_2] = |B_1[s_2, s_3^0, \dots, s_L^0] \exp\{-W_2[s_2, s_3^0, \dots, s_L^0]\}|. \quad (26)$$

Under the move $s_2 \rightarrow s'_2$, one should then resample and update the level-1 bonds, $B_1 \rightarrow B'_1$. Exploiting the fact that the stored K samples $s_1^{(i)}$ are correctly distributed for the original configuration s_2^0 , the updated bond can be computed according to

$$B'_1 = K^{-1} \sum_{i=1}^K \frac{\exp\{-W_1[s_1^{(i)}, s'_2, \dots, s_L]\}}{|\exp\{-W_1[s_1^{(i)}, s_2^0, \dots, s_L^0]\}|}. \quad (27)$$

Again, to obtain an accurate estimate for B'_1 , the number K should be sufficiently large. In the end, sampling under the weight \mathcal{P}_1 implies that the probability for accepting the move $s_2 \rightarrow s'_2$ under the Metropolis algorithm is

$$p = \frac{\left| \sum_i \frac{\exp\{-W_1[s_1^{(i)}, s'_2, s_3^0, \dots]\}}{|\exp\{-W_1[s_1^{(i)}, s_2^0, \dots]\}|} \right|}{\left| \sum_i \frac{\exp\{-W_1[s_1^{(i)}, s_2, s_3^0, \dots]\}}{|\exp\{-W_1[s_1^{(i)}, s_2^0, \dots]\}|} \right|} \times \frac{\left| \exp\{-W_2[s'_2, s_3^0, \dots]\} \right|}{\left| \exp\{-W_2[s_2, s_3^0, \dots]\} \right|}. \quad (28)$$

Using this method, one generates K samples $s_2^{(i)}$, stores them, and computes the level-2

bonds,

$$\begin{aligned}
B_2 &= \left\langle \frac{B_1[s_2, s_3, \dots] \exp\{-W_2[s_2, s_3, \dots]\}}{|B_1[s_2, s_3^0, \dots] \exp\{-W_2[s_2, s_3^0, \dots]\}|} \right\rangle_{\mathcal{P}_1[s_2]} \\
&= C_1^{-1} \sum_{s_2} B_1[s_2, \dots] \exp\{-W_2[s_2, \dots]\} \\
&= K^{-1} \sum_{i=1}^K \frac{B_1[s_2^{(i)}, s_3, \dots] \exp\{-W_2[s_2^{(i)}, s_3, \dots]\}}{|B_1[s_2^{(i)}, s_3^0, \dots] \exp\{-W_2[s_2^{(i)}, s_3^0, \dots]\}|} = B_2[s_3, \dots, s_L], \quad (29)
\end{aligned}$$

with $C_1 = \sum_{s_2} \mathcal{P}_1[s_2]$. Following above strategy, one then rewrites the reduced density matrix by combining Eq. (25) and the second expression in Eq. (29),

$$\rho(s_L) = Z^{-1} \sum_{s_3, \dots, s_{L-1}} \exp \left\{ - \sum_{\ell>2} W_\ell \right\} C_0 C_1 B_2 = Z^{-1} \sum_{s_1, \dots, s_{L-1}} \mathcal{P}_0 \mathcal{P}_1 B_2 \prod_{\ell>2} e^{-W_\ell}. \quad (30)$$

Clearly, the sign problem has been transferred one block further to the right along the chain. Note that the normalization constants C_0, C_1, \dots depend only on the initial configuration s_ℓ^0 so that their precise values need not be known. This procedure is then iterated in a recursive manner. Sampling on level ℓ using the weight function

$$\mathcal{P}_{\ell-1}[s_\ell] = |B_{\ell-1}[s_\ell, s_{\ell+1}^0, \dots] \exp\{-W_\ell[s_\ell, s_{\ell+1}^0, \dots]\}| \quad (31)$$

requires the recursive update of all bonds B_λ with $\lambda < \ell$. Starting with $B_1 \rightarrow B'_1$ and putting $B_0 = 1$, this recursive update is done according to

$$B'_\lambda = K^{-1} \sum_{i=1}^K \frac{B'_{\lambda-1}[s_\lambda^{(i)}, s_{\lambda+1}, \dots] \exp\{-W'_\lambda[s_\lambda^{(i)}, s_{\lambda+1}, \dots]\}}{|B_{\lambda-1}[s_\lambda^{(i)}, s_{\lambda+1}^0, \dots] \exp\{-W_\lambda[s_\lambda^{(i)}, s_{\lambda+1}^0, \dots]\}|}, \quad (32)$$

where the primed bonds or partial actions depend on s'_ℓ and the unprimed ones on s_ℓ^0 . Iterating this to get the updated bonds $B_{\ell-2}$ for all $s_{\ell-1}^{(i)}$, the test move $s_\ell \rightarrow s'_\ell$ is then accepted or rejected according to the probability

$$p = \left| \frac{B_{\ell-1}[s'_\ell, s_{\ell+1}^0, \dots] \exp\{-W_\ell[s'_\ell, s_{\ell+1}^0, \dots]\}}{B_{\ell-1}[s_\ell, s_{\ell+1}^0, \dots] \exp\{-W_\ell[s_\ell, s_{\ell+1}^0, \dots]\}} \right|. \quad (33)$$

On this level, one again generates K samples $s_\ell^{(i)}$, stores them and computes the level- ℓ bonds according to

$$B_\ell[s_{\ell+1}, \dots] = K^{-1} \sum_{i=1}^K \frac{B_{\ell-1}[s_\ell^{(i)}, s_{\ell+1}, \dots] \exp\{-W_\ell[s_\ell^{(i)}, s_{\ell+1}, \dots]\}}{|B_{\ell-1}[s_\ell^{(i)}, s_{\ell+1}^0, \dots] \exp\{-W_\ell[s_\ell^{(i)}, s_{\ell+1}^0, \dots]\}|}.$$

This process is iterated up to the top level, where the observables of interest may be computed. Since the sampling of B_ℓ requires the resampling of all lower-level bonds, the memory and CPU requirements of the algorithm laid out here are quite large. For $\lambda < \ell - 1$, one needs to update $B_\lambda \rightarrow B'_\lambda$ for all $s_{\ell'}^{(i)}$ with $\lambda < \ell' < \ell$, which implies a tremendous amount of computer memory and CPU time, scaling approximately $\sim K^L$ at the top level. Fortunately, an enormous simplification can often be achieved by exploiting the fact that the interactions among distant slices are usually weaker than between near-by slices. For

instance, when updating level $\ell = 3$, the correlations with the configurations $s_1^{(i)}$ may be very weak, and instead of summing over all K samples $s_1^{(i)}$ in the update of the bonds $B_{\lambda < \ell}$, one may select only a small subset. When invoking this argument, one should be careful to also check that the additional interactions coming from the level- λ bonds with $\lambda < \ell$ are sufficiently short-ranged. From the definition of these bonds, this is to be expected though.

3 Applications

In this section, several different applications of the MLB approach will be presented. The first will focus on the equilibrium behavior of interacting electrons in a parabolic quantum dot, a situation characterized by a fermionic sign problem. The two other subsections then deal with the dynamical sign problem.

3.1 Quantum Dots

Quantum dots are solid-state artificial atoms with tunable properties. Confining a small number of electrons N in a 2D electron gas in semiconductor heterostructures, novel effects due to the interplay between confinement and the Coulomb interaction have been observed experimentally.^{15–17} For small N , comparison of experiments to the generalized Kohn theorem indicates that the confinement potential is parabolic and hence quite shallow compared to conventional atoms. Employing the standard electron gas parameter r_s to quantify the correlation strength, only for small r_s , a Fermi-liquid picture is applicable. In the low-density (strong-interaction) limit of large r_s , classical considerations suggest a Wigner crystal-like phase with electrons spatially arranged in shells. We call this a *Wigner molecule* due to its finite extent. Of particular interest is the crossover regime between these two limits, where both single-particle and classical descriptions break down, and basically no other sufficiently accurate method besides QMC is available. Exact diagonalization is limited to very small N since one otherwise introduces a huge error due to the truncation of the Hilbert space. Hartree-Fock (and related) calculations become unreliable for large r_s and incorrectly favor spin-polarized states. Furthermore, density functional calculations can introduce uncontrolled approximations. Regarding QMC, to avoid the sign problem, the fixed-node approximation has been employed by Bolton¹⁸ and later by others.¹⁹ For $N > 5$, typical fixed-node errors in the total energy are found to be of the order of 10%.³ It is then clear that one should resort to exact methods, especially when looking at spin-dependent quantities, where often extremely small spin splittings are found. After our original studies,^{3,6} other studies using the naive PIMC approach were published.^{20,21} These studies are however concerned with the deep Wigner regime $r_s \gtrsim 20$,²¹ which is essentially a classical regime without sign problem not further discussed here, or employ a special virial estimator²⁰ that unfortunately appears to be incorrect except for fully spin-polarized states.²² A clean 2D parabolic quantum dot in zero magnetic field is described by

$$H = \sum_{j=1}^N \left(\frac{\vec{p}_j^2}{2m^*} + \frac{m^*\omega_0^2}{2} \vec{x}_j^2 \right) + \sum_{i < j=1}^N \frac{e^2}{\kappa |\vec{x}_i - \vec{x}_j|}. \quad (34)$$

The electron positions (momenta) are given by \vec{x}_j (\vec{p}_j), their effective mass is m^* , and the dielectric constant is κ . The MLB calculations are carried out at fixed N and fixed z -component of the total spin, $S = (N_\uparrow - N_\downarrow)/2$. As a check, the exact solution for $N = 2$ has been reproduced.

Here results for the energy, $E = \langle H \rangle$ (since H is a non-diagonal operator, two Trotter slices are kept at the top level), and the *spin sensitivity* $\xi_N(r_s)$ will be discussed. The latter quantity is useful to study the crossover from weak to strong correlations,

$$\xi_N(r_s) \propto \sum_{S,S'} \int_0^\infty dy y |g_S(y) - g_{S'}(y)| , \quad (35)$$

where the prefactor is chosen to give $\xi_N = 1$ for $r_s = 0$. This definition makes use of the spin-dependent two-particle correlation function

$$g_S(\vec{x}) = \frac{2\pi l_0^2}{N(N-1)} \left\langle \sum_{i \neq j=1}^N \delta(\vec{x} - \vec{x}_i + \vec{x}_j) \right\rangle , \quad (36)$$

which is isotropic. With $y = r/l_0$ prefactors are chosen such that $\int_0^\infty dy y g_S(y) = 1$. The confinement length scale $l_0 = \sqrt{\hbar/m^*\omega_0}$ allows the interaction strength to be parametrized by the dimensionless parameter $\lambda = l_0/a = e^2/\kappa\omega_0 l_0$, where a is the effective Bohr radius of the artificial atom. For any given N and λ , the density parameter $r_s = r^*/a$ with nearest-neighbor distance r^* follows by identifying r^* with the first maximum in $\sum_S g_S(r)$. The correlation function (36) is a very sensitive measure of Fermi statistics, in particular revealing the spin-dependent correlation hole. Because interactions tend to wash out the Fermi surface, the spin sensitivity (35) is largest for a Fermi gas, $r_s = 0$. Since for $r_s \rightarrow \infty$, one approaches the totally classical limit, where $g_S(r)$ is completely spin-independent, $\xi_N(r_s)$ decays from unity at $r_s = 0$ down to zero as $r_s \rightarrow \infty$. The functional dependence of this decay provides insight about the crossover phenomenon under study.

By computing charge densities, the PIMC simulations can directly reveal *shell formation* in real space.⁶ Such a spatial structure is clear evidence for near-classical Wigner molecule behavior. The classical shell filling sequence is as follows. For $N < 6$, the electrons arrange on a ring, but the sixth electron then goes into the center. Furthermore, electrons 7 and 8 enter the outer ring again. These predictions are in accordance with our PIMC data. Clear indications of a spatial shell structure at $N \geq 6$ can be observed already for $\lambda \approx 4$, albeit quantum fluctuations tend to wash them out somewhat. For $\lambda \gtrsim 4$, charge densities are basically insensitive to S . This is characteristic for a classical Wigner crystal, where the Pauli principle and spin-dependent properties are of little importance. Numerical results for the spin density in this regime simply follow the corresponding charge density according to $s_z(r) \simeq (S/N)\rho(r)$. A significant S -dependence of charge and spin densities is observed only for weak correlations. Figure 2 reveals that $\xi_N(r_s)$ is remarkably *universal*, i.e. it depends only very weakly on N . Its decay defines a crossover scale r_c , where an exponential fit for small r_s yields $r_c \approx 4$. For $r_s > 4$, the data can be well fitted by $\xi(r_s) \sim \exp(-\sqrt{r_s/r'_c})$, where $r'_c \approx 1.2$. Remarkably, this is precisely the behavior expected from a semiclassical WKB estimate for a Wigner molecule.⁶ The crossover value $r_c \approx 4$ is also consistent with the onset of spatial shell structures in the density, and with the spin-dependent ground state energies expected for a Wigner molecule. Therefore the

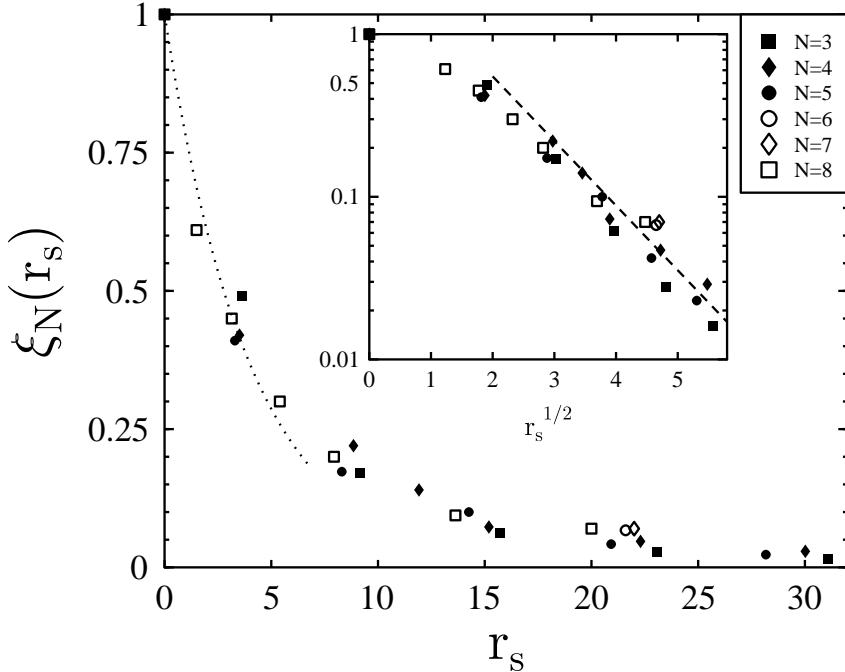


Figure 2. Numerical results for $\xi_N(r_s)$ at $k_B T / \hbar\omega_0 = 0.1$. Statistical errors are of the order of the symbol size. The dotted curve, given by $\exp(-r_s/r_c)$ with $r_c = 4$, is a guide to the eye only. The inset shows the same data on a semi-logarithmic scale as a function of $\sqrt{r_s}$. The dashed line is the WKB estimate (see text).

crossover from weak to strong correlations is characterized by the surprisingly small value $r_c \approx 4$, instead of $r_c \approx 37$ found for the bulk 2D electron gas.²³ This enormous stabilization of the Wigner molecule can be ascribed to the effects of the confinement potential. In the thermodynamic limit, $\omega_0 \rightarrow 0$ with r_s fixed, plasmons govern the low-energy physics, and hence the bulk value $r_c \approx 37$ becomes relevant for very large N . For GaAs based quantum dots, one can estimate⁶ that for $N \lesssim 10^4$, the value $r_c \approx 4$ is the relevant one. Remarkably, recent experiments on vertical quantum dots¹⁶ have found evidence for an even smaller crossover scale $r_c \approx 1.8$. The experimental study was carried out in a magnetic field, and the dot contained several impurities. Since both effects tend to stabilize a Wigner crystallized phase, our prediction and the experimental observation appear to be consistent.

MLB results for the energy at different parameter sets $\{N, S, \lambda\}$ are listed in Table 2. For given N and λ , if the ground state is (partially) spin-polarized with spin S , the simulations should yield the same energies for all $S' < S$. Within the accuracy of the calculation, this consistency check is indeed fulfilled. For strong correlations, $r_s > r_c$, the spin-dependent energy levels differ substantially from a single-particle orbital picture. In particular, the ground-state spin S can change and the relative energy of higher-spin states becomes much smaller than $\hbar\omega_0$. For $N = 3$ electrons, as r_s is increased, a transition occurs from $S = 1/2$ to $S = 3/2$ at an interaction strength $\lambda \approx 5$ corresponding to $r_s \approx 8$.

Table 2. MLB data for the energy for various $\{N, S, \lambda\}$ parameter sets at $k_B T/\hbar\omega_0 = 0.1$. Bracketed numbers denote statistical errors.

N	S	λ	$E/\hbar\omega_0$	N	S	λ	$E/\hbar\omega_0$
3	3/2	2	8.37(1)	5	5/2	8	42.86(4)
3	1/2	2	8.16(3)	5	3/2	8	42.82(2)
3	3/2	4	11.05(1)	5	1/2	8	42.77(4)
3	1/2	4	11.05(2)	5	5/2	10	48.79(2)
3	3/2	6	13.43(1)	5	3/2	10	48.78(3)
3	3/2	8	15.59(1)	5	1/2	10	48.76(2)
3	3/2	10	17.60(1)	6	3	8	60.42(2)
4	2	2	14.30(5)	6	1	8	60.37(2)
4	1	2	13.78(6)	7	7/2	8	80.59(4)
4	2	4	19.42(1)	7	5/2	8	80.45(4)
4	1	4	19.15(4)	8	4	2	48.3(2)
4	2	6	23.790(12)	8	3	2	47.4(3)
4	1	6	23.62(2)	8	2	2	46.9(3)
4	2	8	27.823(11)	8	1	2	46.5(2)
4	1	8	27.72(1)	8	4	4	69.2(1)
4	2	10	31.538(12)	8	3	4	68.5(2)
4	1	10	31.48(2)	8	2	4	68.3(2)
5	5/2	2	21.29(6)	8	4	6	86.92(6)
5	3/2	2	20.71(8)	8	3	6	86.82(5)
5	1/2	2	20.30(8)	8	2	6	86.74(4)
5	5/2	4	29.22(7)	8	4	8	103.26(5)
5	3/2	4	29.15(6)	8	3	8	103.19(4)
5	1/2	4	29.09(6)	8	2	8	103.08(4)
5	5/2	6	36.44(3)				
5	3/2	6	36.35(4)				
5	1/2	6	36.26(4)				

For $N = 4$, a Hund's rule case with a small- r_s ground state characterized by $S = 1$ is encountered. From our data, this standard Hund's rule covers the full range of r_s . A similar situation arises for $N = 5$, where the ground state is characterized by $S = 1/2$ for all r_s . Turning to $N = 6$, while one has filled orbitals and hence a zero-spin ground state for weak correlations, for $\lambda = 8$ a $S = 1$ ground state is found. A similar transition from a $S = 1/2$ state for weak correlations to a partially spin-polarized $S = 5/2$ state is found for $N = 7$. Finally, for $N = 8$, as expected from Hund's rule, a $S = 1$ ground state is observed for small r_s . However, for $\lambda \gtrsim 4$, corresponding to $r_s \gtrsim 10$, the ground state spin changes to $S = 2$, implying a different “strong-coupling” Hund's rule.

Let us finally address the issue of *magic numbers*. For small r_s , the simple picture of filling up single-particle orbitals predicts that certain N are exceptionally stable. Results for the energy per electron, E_N/N , in the spin-polarized state $S = N/2$, are shown in Figure 3. Notably, there are no obvious cusps or breaks in the N -dependence of the energy. The $\lambda = 2$ data in Fig. 3 suggest that an explanation of the experimentally observed magic numbers¹⁷ has to involve spin and/or magnetic field effects, since the single-particle picture breaks down so quickly. Remarkably, there are no pronounced cusps in E_N/N for strong correlations ($\lambda = 8$). Therefore magic numbers seem to play only a minor role in the Wigner molecule phase.

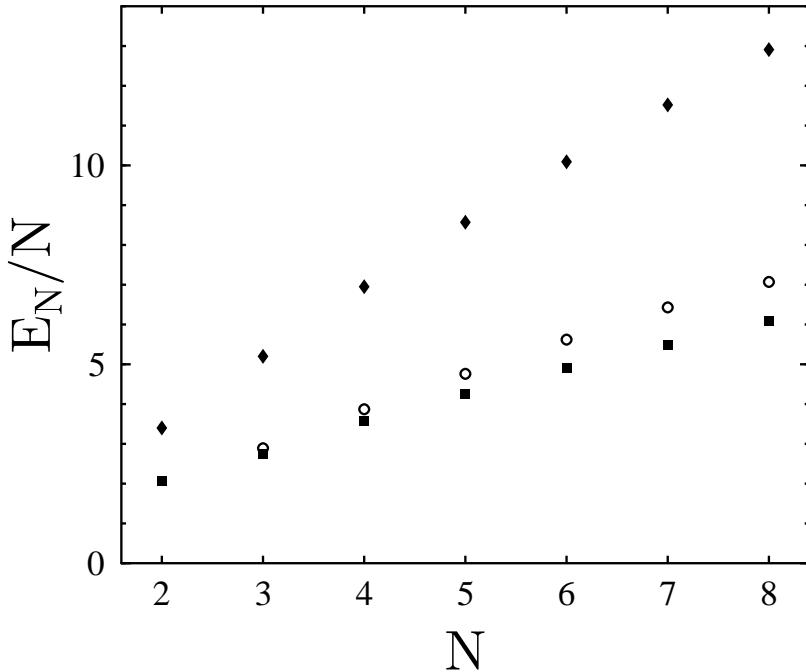


Figure 3. Energy per electron, E_N/N , for $S = N/2$ and $k_B T/\hbar\omega_0 = 1/6$, in units of $\hbar\omega_0$, for $\lambda = 2$ (squares) and $\lambda = 8$ (diamonds). Statistical errors are smaller than the symbol size. Open circles are $T = 0$ fixed-node QMC results¹⁸ for $\lambda = 2$.

3.2 Real-Time Simulations: Simple Model Systems

In each of the following examples, a time-correlation function was computed directly in real time for a simple model system, with increasing level of complexity. The average phase is larger than 0.6 for the presented data sets.

3.2.1 Harmonic Oscillator

For $H = p^2/2m + m\omega^2 x^2/2$, the real and imaginary parts of $\langle x(0)x(t) \rangle$ oscillate in time due to vibrational coherence. In dimensionless units $m = \omega = 1$, the oscillation period is 2π . Figure 4(a) shows MLB results for $C(t) = \text{Re} \langle x(0)x(t) \rangle$. With $P = 32$ for the maximum time $t = 26$, $K = 200$ samples were used for sampling the coarser bonds. Within error bars, the data coincide with the exact result and the algorithm produces stable results free of the sign problem. Without MLB, the signal-to-noise ratio was practically zero for $t > 2$.

3.2.2 Two-Level System

For a symmetric two-state system, $H = -\frac{1}{2}\Delta\sigma_x$, the dynamics is controlled by tunneling. The spin correlation function $\langle \sigma_z(0)\sigma_z(t) \rangle$ exhibits oscillations indicative of quantum coherence. Figure 4(b) shows MLB results for $C(t) = \text{Re} \langle \sigma_z(0)\sigma_z(t) \rangle$. Putting $\Delta = 1$, the

tunneling oscillations have a period of 2π . With $P = 64$ for the maximum time $t = 64$, only $K = 100$ samples were used for sampling the coarser bonds. The data agree well with the exact result. Again the simulation is stable and free of the sign problem. Without MLB, the simulation failed for $t > 4$.

3.2.3 Double-Well Potential

Next, let us examine a double-well system with the quartic potential $V(x) = -x^2 + \frac{1}{4}x^4$. At low temperatures, interwell transfer occurs through tunneling motions on top of intrawell vibrations. These two effects combine to produce nontrivial structures in the position correlation function. At high temperatures, interwell transfer can also occur by classical barrier crossings. Figure 4(c) shows MLB results for $C(t) = \text{Re} \langle x(0)x(t) \rangle$. The slow oscillation corresponds to interwell tunneling, with a period of approximately 16. The higher-frequency motions are characteristic of intrawell oscillations. In this simulation, $K = 300$ samples were used. The data reproduce the exact result well, capturing all the fine features of the oscillations. Again the calculation is stable and free of the sign problem, whereas a direct simulation failed for $t > 3$.

3.2.4 Multidimensional Tunneling System

As a final example, consider a problem with three degrees of freedom, in which a particle in a double-well potential is bilinearly coupled to two harmonic oscillators. The quartic potential in the last example is used for the double-well, and the harmonic potential in the first example is used for both oscillators. The coupling constant between each oscillator and the tunneling particle is $\alpha = 1/2$ in dimensionless units. For this example, the correlation function $C_s(t)$ in Eq. (19) has been computed for the position operator of the tunneling particle. Direct application of MC sampling to $C_s(t)$ has generally been found unstable for $t > \beta\hbar/2$.¹² In contrast, employing only moderate values of $K = 400$ to 900 allow to go up to $t = 10\beta\hbar$. Figure 4(d) shows MLB results for $C'_s(t) = \text{Re} C_s(t)$. For the coupled system, the position correlations have lost the coherent oscillations and instead decay monotonically with time. Coupling to the medium clearly damps the coherence and tends to localize the tunneling particle.

3.3 Spin-Boson Dynamics

Finally, to demonstrate the performance of the MLB approach for effective-action-type problems, the real-time dynamics of the celebrated spin-boson model¹⁴

$$H = -(\hbar\Delta/2)\sigma_x + (\hbar\epsilon/2)\sigma_z + \sum_{\alpha} \left[\frac{p_{\alpha}^2}{2m_{\alpha}} + \frac{1}{2}m_{\alpha}\omega_{\alpha}^2 \left(x_{\alpha} - \frac{c_{\alpha}}{m_{\alpha}\omega_{\alpha}^2}\sigma_z \right)^2 \right] \quad (37)$$

is discussed. This model has a number of important applications,¹⁴ e.g. the Kondo problem, interstitial tunneling in solids, quantum computing and electron transfer reactions, to mention only a few. The bare two-level system (TLS) has a tunneling matrix element Δ and the asymmetry (bias) ϵ between the two localized energy levels (σ_x and σ_z are

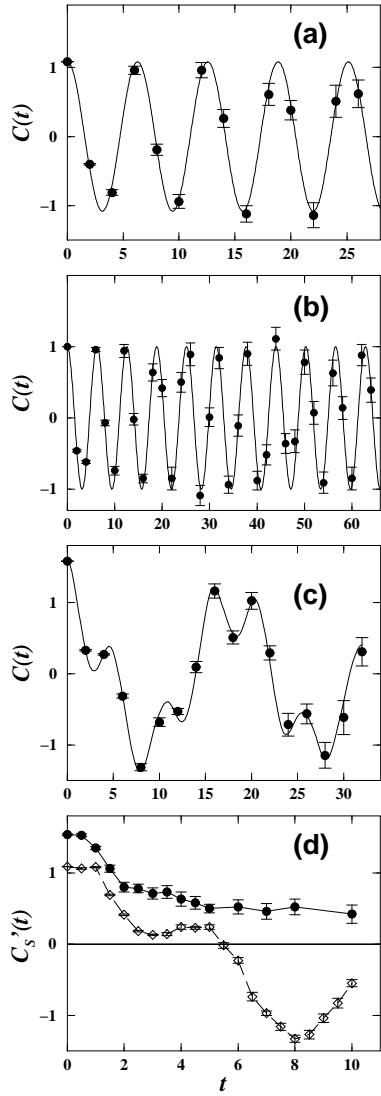


Figure 4. MLB results (closed circles) for various systems. Error bars indicate one standard deviation. (a) $C(t)$ for a harmonic oscillator at $\beta\hbar = 1$. The exact result is indicated by the solid curve. (b) Same as (a) for a two-level system at $\beta\hbar = 10$. (c) Same as (a) for a double-well system at $\beta\hbar = 1$. This temperature corresponds to the classical barrier energy. (d) $C_s'(t)$ for a double-well system coupled to two oscillators at $\beta\hbar = 1$. For comparison, open diamonds are for the uncoupled ($\alpha = 0$) system. Note that $C_s'(t)$ is similar but not identical to $C(t)$ shown in (c). Solid and dashed lines are guides to the eye only.

Pauli matrices). Dissipation is introduced via a linear heat bath, i.e. an arbitrary collection of harmonic oscillators $\{x_\alpha\}$ bilinearly coupled to σ_z . Concerning the TLS dynamics, all information about the coupling to the bath is contained in the spectral density

Table 3. MLB performance for $\alpha = 1/2$, $\omega_c/\Delta = 6$, $\Delta t^* = 10$, $P = 40$, and several L . q_ℓ denotes the number of slices for $\ell = 1, \dots, L$.

K	L	q_ℓ	$\langle \text{sgn} \rangle$
1	1	40	0.03
200	2	30 - 10	0.14
800	2	30 - 10	0.20
200	3	22 - 12 - 6	0.39
600	3	22 - 12 - 6	0.45

$J(\omega) = (\pi/2) \sum_\alpha (c_\alpha^2 / m_\alpha \omega_\alpha) \delta(\omega - \omega_\alpha)$, which has a quasi-continuous form in typical condensed-phase applications and dictates the form of the (twice-integrated) bath correlation function

$$Q(t) = \int_0^\infty \frac{d\omega}{\pi\hbar} \frac{J(\omega)}{\omega^2} \frac{\cosh[\omega\hbar\beta/2] - \cosh[\omega(\hbar\beta/2 - it)]}{\sinh[\omega\hbar\beta/2]} . \quad (38)$$

For the calculations here, an ohmic spectral density $J(\omega) = 2\pi\hbar\omega \exp(-\omega/\omega_c)$ has been assumed, for which $Q(t)$ can be found in closed form.¹⁴ Here ω_c is a cutoff frequency, and the damping strength is measured by the dimensionless Kondo parameter α . In the *scaling limit* $\Delta \ll \omega_c$ with $\alpha < 1$, all dependence on ω_c enters via a renormalized tunnel splitting¹⁴

$$\Delta_{\text{eff}} = [\cos(\pi\alpha)\Gamma(1-2\alpha)]^{1/2(1-\alpha)} (\Delta/\omega_c)^{\alpha/(1-\alpha)} \Delta , \quad (39)$$

and powerful analytical and alternative numerical methods are readily available.¹⁴ However, there are important applications (e.g. electron transfer reactions) that require to study the spin-boson problem away from the scaling limit. Here one generally has to resort to numerical methods. Basically all other available computational techniques can only deal with equilibrium quantities, or explicitly introduce approximations; for an overview and references, see Ref.¹⁴ The MLB approach is computationally more expensive than other methods but at the same time unique in yielding numerically exact results for the nonequilibrium spin-boson dynamics for arbitrary system parameters $\Delta, \epsilon, J(\omega)$ and $\beta = 1/k_B T$.

Below only results for the *occupation probability* $P(t) = \langle \sigma_z(t) \rangle$ under the nonequilibrium initial preparation $\sigma_z(t < 0) = +1$ are presented. $P(t)$ gives the time-dependent difference of the quantum-mechanical occupation probabilities of the left and right states, with the particle initially confined to the left state. To obtain $P(t)$ numerically, in a discretized path-integral representation one traces out the bath to get a long-ranged effective action, the *influence functional*.¹⁴ In discretized form the TLS path is represented by spins $\sigma_i, \sigma'_i = \pm 1$ on the forward- and backward-paths, respectively. The total action S consists of three terms. First, there is the “free” action S_0 determined by the bare TLS propagator U_0 ,

$$\exp(-S_0) = \prod_{i=0}^{P-1} U_0(\sigma_{i+1}, \sigma_i; t^*/P) U_0(\sigma'_{i+1}, \sigma'_i; -t^*/P) , \quad (40)$$

where t^* is the maximum time and P the Trotter number. Next there is the influence

functional, $S_I = S_I^{(1)} + S_I^{(2)}$, which contains the long-ranged interaction among the spins,

$$S_I^{(1)} = \sum_{j \geq m} (\sigma_j - \sigma'_j) \left\{ L'_{j-m} (\sigma_m - \sigma'_m) + i L''_{j-m} (\sigma_m + \sigma'_m) \right\},$$

where $L_j = L'_j + i L''_j$ is given by

$$L_j = [Q((j+1)t^*/P) + Q((j-1)t^*/P) - 2Q(jt^*/P)]/4 \quad (41)$$

for $j > 0$, and $L_0 = Q(t^*/P)/4$. In the scaling regime at $T = 0$, this effective action produces interactions $\sim \alpha/t^2$ between the spins (“inverse-square Ising model”). The contribution

$$S_I^{(2)} = i(t^*/P) \sum_m \gamma(m t^*/P) (\sigma_m - \sigma'_m)$$

describes the interactions with the imaginary-time branch where $\sigma_z = +1$, with the damping kernel

$$\gamma(t) = \frac{2}{\pi \hbar} \int_0^\infty d\omega \frac{J(\omega)}{\omega} \cos(\omega t).$$

The most difficult case for PIMC corresponds to an unbiased two-state system at zero temperature, $\epsilon = T = 0$. To check the code, the case $\alpha = 1/2$ was studied in some detail, where the exact solution¹⁴ is very simple, $P(t) = \exp(-\Delta_{\text{eff}} t)$. This exact solution only holds in the scaling limit, which is already reached for $\omega_c/\Delta = 6$ where the MLB-PIMC simulations yield precisely this result. Typical MLB parameters and the respective average sign are listed in Table 3. The first line in Table 3 corresponds to the naive approach. It is then clear that the average sign and hence the signal-to-noise ratio can be dramatically improved thus allowing for a study of long timescales t^* . For a fixed number of levels L , the average sign grows by increasing the parameter K . Alternatively, for fixed K , the average sign increases with L . Evidently, the latter procedure is more efficient in curing the sign problem, but at the same time computationally expensive. In practice, it is then necessary to find a suitable compromise.

Figure 5 shows scaling curves for $P(t)$ at $\alpha = 1/4$ for $\omega_c/\Delta = 6$ and $\omega_c/\Delta = 1$. The first value for ω_c/Δ is within the scaling regime. This is confirmed by a comparison to the noninteracting blip approximation (NIBA),¹⁴ which is known to be very accurate for $\alpha < 1$ in the scaling limit. However, for $\omega_c/\Delta = 1$, scaling concepts and also NIBA are expected to fail dramatically. This is seen in the simulations. MLB results show that away from the scaling limit, quantum coherence is able to persist for much longer, and both frequency and decay rate of the oscillations differ significantly from the predictions of NIBA. In electron transfer reactions in the adiabatic-to-nonadiabatic crossover regime, such coherence effects can then strongly influence the low-temperature dynamics. One obvious and important consequence of these coherence effects is the breakdown of a rate description, implying that theories based on an imaginary-time formalism might not be appropriate in this regime. A detailed analysis of this crossover regime using MLB is currently in progress.⁹

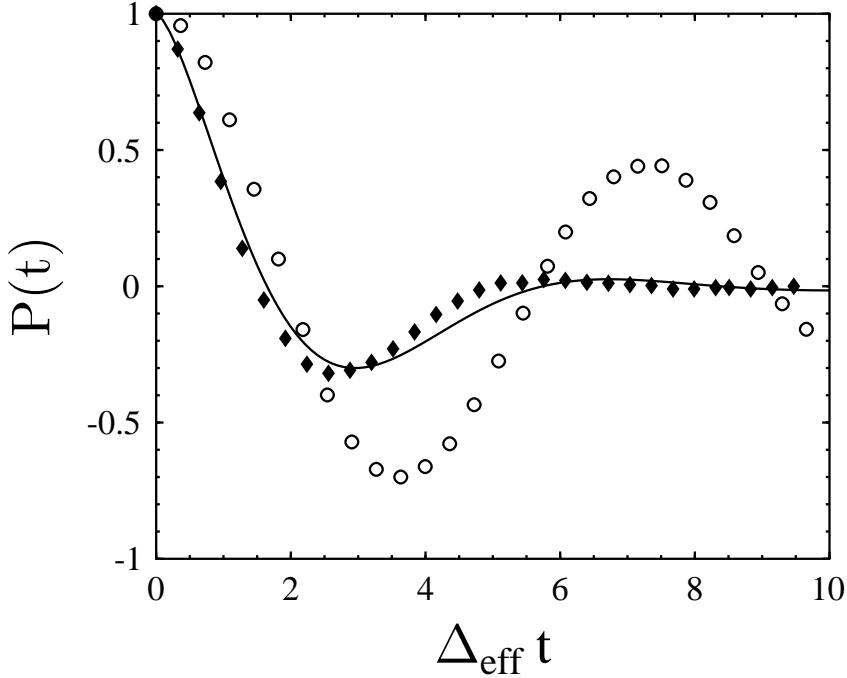


Figure 5. Scaling curves for $P(t)$ for $\alpha = 1/4$ with $\omega_c/\Delta = 6$ (closed diamonds) and $\omega_c/\Delta = 1$ (open circles). The solid curve is the NIBA prediction. Statistical errors are of the order of the symbol sizes.

4 Concluding Remarks

These notes summarize our previous activities using the multilevel blocking approach to the sign problem in path-integral Monte Carlo simulations. The approach holds substantial promise towards relieving (and eventually overcoming) the sign problem, but clearly there is still much room for improvement. The applications presented here demonstrate unambiguously that there are general and powerful handles to relieve the sign problem, even though a problem characterized by an intrinsic sign problem is still much harder than one without. We hope that especially young researchers will be attracted to work on this subject themselves.

Acknowledgments

Parts of this review are based on work with Lothar Mühlbacher. This research has been supported by the Volkswagen-Stiftung (I/73 259) and by the National Science Foundation (CHE-9970766).

References

1. See, e.g., *Quantum Monte Carlo Methods in Condensed Matter Physics*, ed. by M. Suzuki (World Scientific, Singapore, 1993), and references therein.
2. See, for instance, E.Y. Loh Jr., J. Gubernatis, R.T. Scalettar, S.R. White, D.J. Scalapino, and R.L. Sugar, Phys. Rev. B **41**, 9301 (1990).
3. C.H. Mak, R. Egger, and H. Weber-Gottschick, Phys. Rev. Lett. **81**, 4533 (1998).
4. S. Chandrasekharan and U.-J. Wiese, Phys. Rev. Lett. **83**, 3116 (1999).
5. M.H. Kalos and F. Pederiva, Phys. Rev. Lett. **85**, 3547 (2000).
6. R. Egger, W. Häusler, C.H. Mak, and H. Grabert, Phys. Rev. Lett. **82**, 3320 (1999); **83**, 462(E) (1999).
7. C.H. Mak and R. Egger, J. Chem. Phys. **110**, 12 (1999).
8. R. Egger, L. Mühlbacher, and C.H. Mak, Phys. Rev. E **61**, 5961 (2000).
9. L. Mühlbacher and R. Egger, in preparation.
10. M.V. Dikovsky and C.H. Mak, Phys. Rev. B **63**, 235105 (2001).
11. J. Goodman and A.D. Sokal, Phys. Rev. D **40**, 2035 (1989).
12. D. Thirumalai and B.J. Berne, J. Chem. Phys. **79**, 5029 (1984); *ibid.* **81**, 2512 (1984).
13. A.D. Kennedy and J. Kuti, Phys. Rev. Lett. **54**, 2473 (1985).
14. U. Weiss, *Quantum dissipative systems*, Vol.2 (World Scientific, Singapore, 1999).
15. R.C. Ashoori, Nature **379**, 413 (1996).
16. N.B. Zhitenev *et al.*, Science **285**, 715 (1999).
17. S. Tarucha *et al.*, Phys. Rev. Lett. **77**, 3613 (1996).
18. F. Bolton, Phys. Rev. Lett. **73**, 158 (1994).
19. F. Pederiva, C.J. Umrigar, and E. Lipparini, Phys. Rev. B **62**, 8120 (2000).
20. J. Harting, O. Mülken, and P. Borrman, Phys. Rev. B **62**, 10207 (2000).
21. A.V. Filinov, M. Bonitz, and Yu.E. Lozovik, Phys. Rev. Lett. **86**, 3851 (2001).
22. B. Reusch, private communication.
23. B. Tanatar and D.M. Ceperley, Phys. Rev. B **39**, 5005 (1989).

Statistical Analysis of Simulations: Data Correlations and Error Estimation

Wolfhard Janke

Institut für Theoretische Physik, Universität Leipzig
Augustusplatz 10/11, 04109 Leipzig, Germany
E-mail: wolfhard.janke@itp.uni-leipzig.de

This lecture gives an introduction to data correlations and error estimation in Monte Carlo simulations. The basic concepts are illustrated for studies of second-order phase transitions in classical spin systems such as the Ising model and for models exhibiting first-order phase transitions such as the q -state Potts model.

1 Introduction

The statistical mechanics of complex systems poses many hard problems which are often difficult to solve by analytical approaches. Numerical simulation techniques will therefore be indispensable tools on our way to a better understanding of such systems. Applications range from physics and chemistry to biological systems and even sociology and economy. Examples include (spin) glasses, disordered magnets, or biologically motivated studies of protein folding, to mention only a few important problems in classical physics. The broad class of quantum statistical problems in condensed matter and elementary particles physics as well as non-perturbative approaches to quantum gravity are further important examples.

Depending on the specific physical problem and the objectives at hand, the simulational approach is either based on molecular dynamics (MD) or Monte Carlo (MC) methods.^{1,2} Sometimes even a combination of both methods is used. For the purpose of this lecture I will focus in the following mainly on the MC approach. Thanks to advances in computer technology and significant algorithmic refinements in the past few years, MC studies have reached in many applications an accuracy that allows “quasi-exact” predictions. Since a MC simulation is a stochastic method, it is thus very important to supplement the data with carefully determined, reliable statistical error estimates. But also in other areas, where it is sometimes only feasible to obtain a first qualitative overview of the system behaviour from simulation studies, it is necessary to gain at least a rough idea of the data correlations involved and the order of magnitude of the statistical errors.

The required tools for the data analysis described below are designed for the general case and are perfectly suited for applications to very complex systems. Still, they can be illustrated and tested for very simple examples. In this lecture I will, therefore, mainly concentrate on Ising and Potts models, and sometimes even use synthetically generated stochastic data.

The rest of the paper is organized as follows. In the next section I first recall the definition of the example models and a few standard observables. Then some properties of phase transitions and related aspects of Monte Carlo simulations are briefly summarized. Here emphasis will be placed on those points which are necessary for an understanding of the tools of statistical data analysis which are described in Sec. 3 and illustrated with worked

out examples in Secs. 4 and 5. Section 6 is devoted to error propagation in multicanonical simulations. In Sec. 7, I conclude with a brief summary and a few general comments.

2 Model Systems and Phase Transitions

2.1 Models and Observables

When developing and testing advanced analysis tools for stochastically generated data it is not only convenient to work with very simple models but, in fact, even advantageous since usually at least a few exact analytical results are available for comparison. On the other hand, one should always be prepared that a really complex system may add further complications which are not present in the simple test cases. I nevertheless will follow the “bottom-up” approach, but at several places point out potential difficulties when more complex systems are considered.

The paradigm for a well-controlled second-order phase transition is the well-known Ising model³ whose partition function is defined as

$$Z_I(\beta) = \sum_{\{\sigma_i\}} \exp(-\beta H_I) , \quad H_I = - \sum_{\langle ij \rangle} \sigma_i \sigma_j , \quad \sigma_i = \pm 1 , \quad (1)$$

where $\beta = J/k_B T$ is the inverse temperature in natural units, the spins σ_i live on the sites i of a D-dimensional cubic lattice of volume $V = L^D$, and the symbol $\langle ij \rangle$ indicates that the lattice sum runs over all 2D nearest-neighbour pairs. We always assume periodic boundary conditions. In two dimensions (2D) and zero field this model has been solved exactly, even on finite lattices. For the three-dimensional (3D) model no exact solution is available, but there is an enormous amount of very precise data from MC simulations, high-temperature series expansions and, as far as critical exponents are concerned, also from field theoretical methods. The computer code for MC simulations of this model is easy to program and can be found, for instance, on the accompanying diskette to the article in Ref. 4, where Metropolis, heat-bath, Wolff single-cluster and Swendsen-Wang multiple-cluster update routines are described and implemented. In addition also programs for reweighting analyses and the exact solution of the 2D Ising model are provided in this reference.

Standard observables are the internal energy per site, $e = E/V$, with $E = -d \ln Z_I / d\beta \equiv \langle H_I \rangle$, the specific heat,

$$C/k_B = \frac{de}{d(k_B T)} = \beta^2 (\langle H_I^2 \rangle - \langle H_I \rangle^2) / V , \quad (2)$$

the magnetization

$$m = M/V = \langle |\mu| \rangle , \quad \mu = \sum_i \sigma_i / V , \quad (3)$$

and the susceptibility

$$\chi = \beta V (\langle \mu^2 \rangle - \langle |\mu| \rangle^2) . \quad (4)$$

In the high-temperature phase one often employs the fact that the magnetization vanishes in the infinite-volume limit and defines

$$\chi' = \beta V \langle \mu^2 \rangle . \quad (5)$$

Similarly, the spin-spin correlation function can then be taken as

$$G(\vec{x}_i - \vec{x}_j) = \langle \sigma_i \sigma_j \rangle . \quad (6)$$

At large distances, $G(\vec{x})$ decays exponentially, $G(\vec{x}) \sim \exp(-|\vec{x}|/\xi)$, and the spatial correlation length ξ can be defined as

$$\xi = - \lim_{|\vec{x}| \rightarrow \infty} (|\vec{x}| / \ln G(\vec{x})) . \quad (7)$$

The standard example exhibiting a first-order phase transition is the q -state Potts model⁵ defined by the Hamiltonian

$$H_P = - \sum_{\langle ij \rangle} \delta_{\sigma_i \sigma_j} , \quad \sigma_i \in 1, \dots, q , \quad (8)$$

where δ_{ij} is the Kronecker symbol. In 2D, this model is exactly known⁶ to exhibit a temperature-driven first-order transition at $\beta_t = \log(1 + \sqrt{q})$ for all $q \geq 5$. For $q \leq 4$ the transition is of second order, including the Ising case ($q = 2$) and the special percolation limit ($q = 1$). In 3D, there is strong numerical evidence that for all $q \geq 3$ the transition is of first order.⁷

2.2 Phase Transitions

The theory of phase transitions is a broad subject well covered by many textbooks. Here we shall confine ourselves to those properties that are important for understanding the requirements on the data analysis tools.

The characterising feature of second-order phase transitions is a *divergent* spatial correlation length ξ at the transition point β_c . This causes scale invariance, the “heart” of renormalization group treatments, and is the origin of the important concept of universality. The growth of spatial correlations in the vicinity of the critical point is illustrated in Fig. 1. At β_c one thus expects fluctuations on all length scales, implying power-law singularities in thermodynamic functions such as the correlation length,

$$\xi = \xi_0^{+, -} t^{-\nu} + \dots , \quad (9)$$

where $t \equiv |1 - T/T_c|$ and the \dots indicate subleading (analytical and confluent) corrections. This defines the (universal) critical exponent ν and the (non-universal) critical amplitudes $\xi_0^{+, -}$ on the high- and low-temperature side of the transition. Similar singularities of the specific heat, magnetization, and susceptibility as sketched in Fig. 2 define the critical exponents α , β , and γ , respectively, which are related with each other through scaling and hyper-scaling relations; only two of them (e.g. ν and γ) may be considered as independent.

When updating the spins with an importance sampling MC process,^{2,4,8} the information on the updated state of the spins has to propagate over the correlation volume before one obtains a new, statistically *independent* configuration. The number of update steps this takes is measured by the autocorrelation time τ (a formal definition is given below) which close to β_c scales according to

$$\tau \propto \xi^z \propto t^{-\nu z} . \quad (10)$$

Here we have introduced the independent *dynamical* critical exponent z , which depends on the employed update algorithm. For a *local* MC update procedure such as the Metropolis

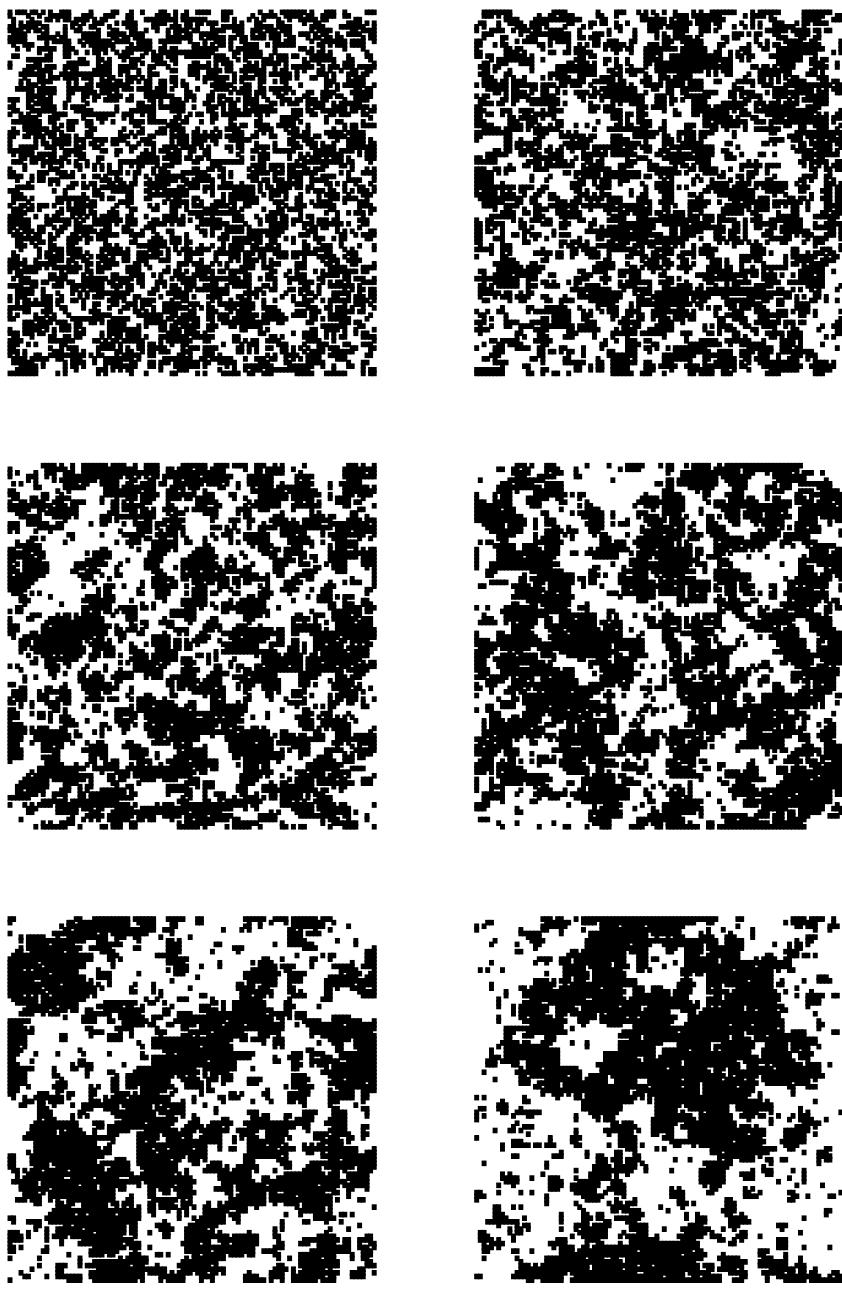


Figure 1. Approach of the critical region (lower right) starting from high temperatures (upper left) illustrating the development of large spatial correlations. Shown are 2D Ising model configurations for a 100×100 lattice with periodic boundary conditions at $\beta/\beta_c = 0.50, 0.70, 0.85, 0.90, 0.95$, and 0.98 .

or heat-bath algorithm, one expects that the updated information performs a random walk in configuration space, requiring on the average ξ^2 steps to propagate over a distance proportional to ξ . For local update algorithms one thus expects a dynamical critical exponent of $z \approx 2$. In fact, an exact lower bound is $z \geq \gamma/\nu = 2 - \eta$, and numerical estimates for the Ising model yield $z \approx 2.125$ in 2D and $z \approx 2.03$ in 3D.^{9,10} As we will see in the next section, this *critical slowing down* of the dynamics (based on local update rules) is responsible for the dramatic reduction of the statistical accuracy attainable close to a critical point in a given computer time allocation. This is the reason why cluster and multigrid update algorithms have become so important.⁸⁻¹⁰ Here the update rules are *non-local*, leading to a significant reduction of critical slowing down. The dynamical critical exponent z varies among the different non-local update schemes and depends on the model class considered. In most cases, however, one finds z smaller than unity, and when cluster algorithms are applied to the 2D Ising model it is even difficult to distinguish z from zero, i.e. a logarithmic divergence.

For systems of finite size, as in any numerical simulation, the correlation length can-

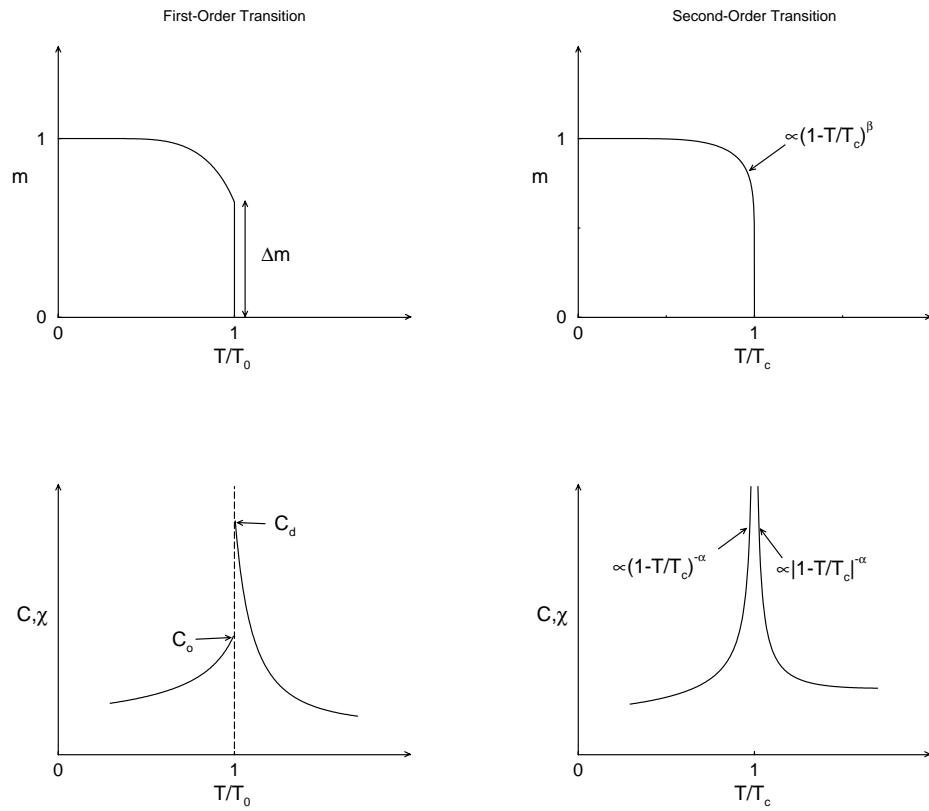


Figure 2. The characteristic behaviour of the magnetization, m , specific heat, C , and susceptibility, χ , at first- and second-order phase transitions.

not diverge, and also the divergences in all other quantities are rounded and shifted. For the specific heat of the 2D Ising model this is illustrated in Fig. 3. In the scaling formulas the role of ξ is then taken over by the linear size of the system, L . By writing $t \propto \xi^{-1/\nu} \rightarrow L^{-1/\nu}$, it becomes clear how thermodynamic scaling laws, e.g. $\chi \propto t^{-\gamma}$, should be replaced by *finite-size scaling* (FSS) Ansätze, e.g. $\chi \propto L^{\gamma/\nu}$, for finite geometries. In particular, by recalling (10) we obtain for the autocorrelation time a FSS of the form

$$\tau \propto L^z . \quad (11)$$

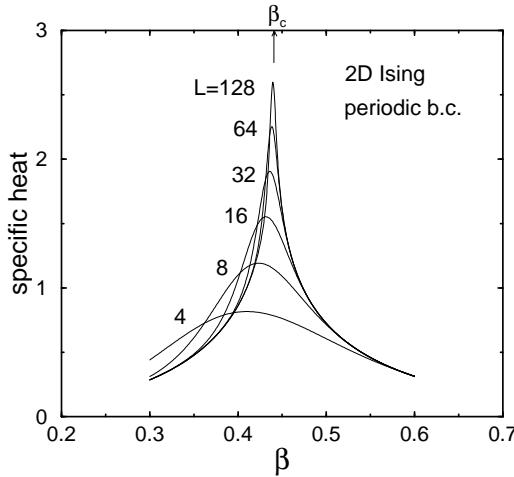


Figure 3. Finite-size scaling behaviour of the specific heat of the 2D Ising model on $L \times L$ lattices close to the infinite-volume critical point $\beta_c = \log(1 + \sqrt{2})/2 = 0.440\,686 \dots$

Most phase transitions in nature are of first order.^{11–14} An example experienced every day is ordinary melting. The characterising feature of first-order phase transitions are *discontinuities* in the order parameter (the jump Δm of the magnetization m in Fig. 2) or in the energy (the latent heat Δe), or both, at the transition point T_0 . This reflects the fact that at T_0 , two (or more) phases can coexist. In the example of the melting transition these are the solid (ordered) and liquid (disordered) phases. In contrast to a second-order transition, the correlation length in the coexisting pure phases is finite. Consequently the specific heat, the susceptibility and also the autocorrelation time do not diverge in the *pure* phases. There are, however, superimposed delta function like singularities associated with the jumps of e and m .

For finite systems the singularities are smeared out to narrow peaks with a height proportional to the volume and a width proportional to 1/volume. This signalizes that the system is now capable to flip from one phase into the other via mixed phase configurations. Mixed phase configurations are separated by interfaces which carry an extra free energy σL^{D-1} , where σ is the (reduced) interface tension and L^{D-1} is the projected area

of the interfaces. Since at T_0 the total bulk contribution to the free energy of the two co-existing phases equals that of the pure phases, the probability for the occurrence of mixed phase configurations is suppressed by the interfacial Boltzmann weight $\exp(-2\sigma L^{D-1})$, where the factor two accounts for the topological constraint that with periodic boundary conditions only an even number of interfaces can be present. This argument explains the typical double-peak structure of the energy and magnetization densities at a first-order phase transition. Due to the exponentially suppressed dip between the two peaks, for large system sizes it may take very long before the systems flips from one phase into the other. In fact, the autocorrelation time associated with this “flipping mode” is just the inverse of the suppression factor in the probability density,

$$\tau \propto \exp(2\sigma L^{D-1}) . \quad (12)$$

Since here the autocorrelations grow exponentially with the system size, this behaviour has been termed “supercritical slowing down” – even though nothing is “critical” at a first-order phase transition.

The standard acceleration methods such as cluster and multigrid algorithms cannot overcome this slowing down problem. The reason is that it is rooted in the shape of the probability density itself. In this situation a completely different strategy is necessary, namely generalized ensemble simulations. One of these techniques are multicanonical simulations^{15–17} whose statistical analysis requires special care.

In the multicanonical reweighting approach (for reviews see Refs. 18, 19 and 20) one rewrites the partition function by introducing an auxiliary function $f(E)$ as

$$Z = \sum_{\{s_i\}} e^{-\beta[H-f(E)]} e^{-\beta f(E)} , \quad (13)$$

and adjusts the reweighting factor $\exp(-\beta f(E))$ in such a way that the resulting histogram of E sampled according to the *multicanonical* probability distribution

$$\begin{aligned} p_{\text{muca}}(E) &\propto \exp[-\beta(H - f(E))] \\ &\equiv \exp[-\beta H_{\text{muca}}] \end{aligned} \quad (14)$$

is approximately flat between the two peaks. Here H_{muca} is the central object of a multicanonical simulation, and plays the same role in it as H does in a canonical simulation. Canonical observables $\langle \mathcal{O} \rangle_{\text{can}}$ can be recovered according to

$$\langle \mathcal{O} \rangle_{\text{can}} = \frac{\langle \mathcal{O} w \rangle}{\langle w \rangle} , \quad (15)$$

where $\langle \dots \rangle$ without subscript denote expectation values in the multicanonical distribution and

$$w \equiv \exp(\beta f(E)) \quad (16)$$

is the inverse reweighting factor. The multicanonical probability distribution p_{muca} may be updated using any legitimate MC algorithm, the simplest choice being a local Metropolis update. The fact that all canonical expectation values are expressed as ratios of multicanonical expectation values is the reason for the extra complications in the statistical analysis of multicanonical simulations.

3 Estimators, Autocorrelation Times, Bias and Resampling

About a decade ago most of the statistical analysis methods discussed in this section were still quite cumbersome since due to disk-space limitations they usually had to be applied “on the fly” during the simulation. In particular dynamical aspects of a given model are usually not easy to predict beforehand such that the guess of reasonable analysis parameters was quite difficult. The situation has changed dramatically when it became affordable to store hundreds of megabytes on hard-disks. Since then a simulation study can clearly be separated into “raw data generation” and “data analysis” parts. The interface between these two parts should consist of time series of measurements of the relevant physical observables taken during the actual simulations. In principle there are no limitations on the choice of observables \mathcal{O} which could be, for example, the energy H_I or the magnetization μ . Once the system is in equilibrium (which, in general, is non-trivial to assure), we simply save $\mathcal{O}_j \equiv \mathcal{O}[\{\sigma_i\}]_j$ where j labels the measurements. Given these data files one can perform detailed error analyses; in particular adapting parameters to a specific situation is now straightforward and very fast.

3.1 Estimators

If the time series data results from an importance sampling MC simulation, the expectation value $\langle \mathcal{O} \rangle$ can be estimated as a simple arithmetic mean over the Markov chain,

$$\overline{\mathcal{O}} = \frac{1}{N} \sum_{j=1}^N \mathcal{O}_j , \quad (17)$$

where we assume that the time series contains a total of N measurements. Conceptually it is important to distinguish between the expectation value $\langle \mathcal{O} \rangle$, which is an ordinary number, and the estimator $\overline{\mathcal{O}}$, which is a *random* number fluctuating around the theoretically expected value. Of course, in practice one does not probe the fluctuations of the mean value directly (which would require repeating the whole MC simulation many times), but rather estimates its variance,

$$\sigma_{\overline{\mathcal{O}}}^2 = \langle [\overline{\mathcal{O}} - \langle \overline{\mathcal{O}} \rangle]^2 \rangle = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2 , \quad (18)$$

from the distribution of the individual measurements \mathcal{O}_j . If the N subsequent measurements \mathcal{O}_j were all uncorrelated then the relation would simply be

$$\sigma_{\overline{\mathcal{O}}}^2 = \sigma_{\mathcal{O}_j}^2 / N , \quad (19)$$

where $\sigma_{\mathcal{O}_j}^2 = \langle \mathcal{O}_j^2 \rangle - \langle \mathcal{O}_j \rangle^2$ is the variance of the individual measurements. Here one assumes, of course, that the simulation is in equilibrium and uses time-translation invariance over the Markov chain. Equation (19) is true for any distribution $\mathcal{P}(\mathcal{O}_j)$ of the \mathcal{O}_j . For the energy or magnetization the latter distributions are often plotted as physically directly relevant histograms whose squared width ($= \sigma_{\mathcal{O}_j}^2$) is proportional to the specific heat or susceptibility, respectively.

Whatever form the distribution $\mathcal{P}(\mathcal{O}_j)$ assumes (which, in fact, is often close to Gaussian because the \mathcal{O}_j are usually already lattice averages over many degrees of freedom), by the central limit theorem the distribution of the mean value is Gaussian, at least for uncorrelated data in the asymptotic limit of large N . The variance of the mean, $\sigma_{\overline{\mathcal{O}}}^2$, is the squared

width of this (N dependent) distribution which is usually taken as the “one-sigma” squared error, $\epsilon_{\overline{\mathcal{O}}}^2 \equiv \sigma_{\overline{\mathcal{O}}}^2$, and quoted together with the mean value $\overline{\mathcal{O}}$. Under the assumption of a Gaussian distribution the interpretation is that about 68% of all simulations under the same conditions would yield a mean value in the range $[\overline{\mathcal{O}} - \sigma_{\overline{\mathcal{O}}}, \overline{\mathcal{O}} + \sigma_{\overline{\mathcal{O}}}]$. For a “two-sigma” interval which also is sometimes used, this percentage goes up to about 95.4%, and for a “three-sigma” interval the confidence level is higher than 99.7%.

3.2 Autocorrelation Times

Things become more involved for correlated measurements.^{21–23} Starting from the second identity in (18) and inserting (17), we obtain

$$\sigma_{\overline{\mathcal{O}}}^2 = \langle \overline{\mathcal{O}}^2 \rangle - \langle \overline{\mathcal{O}} \rangle^2 = \frac{1}{N^2} \sum_{i,j=1}^N \langle \mathcal{O}_i \mathcal{O}_j \rangle - \frac{1}{N^2} \sum_{i,j=1}^N \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle . \quad (20)$$

By collecting diagonal and off-diagonal terms we arrive at

$$\sigma_{\overline{\mathcal{O}}}^2 = \frac{1}{N^2} \sum_{i=1}^N (\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle^2) + \frac{1}{N^2} \sum_{i \neq j}^N (\langle \mathcal{O}_i \mathcal{O}_j \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_j \rangle) . \quad (21)$$

The first term is identified as the variance of the individual measurements times $1/N$. In the second sum we first use the symmetry $i \leftrightarrow j$ to reduce the summation to $\sum_{i \neq j}^N = 2 \sum_{i=1}^N \sum_{j=i+1}^N$. Then we reorder the summation and use time translation invariance to derive

$$\sigma_{\overline{\mathcal{O}}}^2 = \frac{1}{N} \left[\sigma_{\mathcal{O}_i}^2 + 2 \sum_{k=1}^N (\langle \mathcal{O}_1 \mathcal{O}_{1+k} \rangle - \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_{1+k} \rangle) \left(1 - \frac{k}{N} \right) \right] , \quad (22)$$

where, due to the last factor, the $k = N$ term may trivially be kept in the summation. Factoring out $\sigma_{\mathcal{O}_i}^2$, this can be written as

$$\epsilon_{\overline{\mathcal{O}}}^2 \equiv \sigma_{\overline{\mathcal{O}}}^2 = \frac{\sigma_{\mathcal{O}_i}^2}{N} 2\tau'_{\mathcal{O},\text{int}} . \quad (23)$$

Here we have introduced the so-called (proper) *integrated* autocorrelation time,

$$\tau'_{\mathcal{O},\text{int}} = \frac{1}{2} + \sum_{k=1}^N A(k) \left(1 - \frac{k}{N} \right) , \quad (24)$$

with

$$A(k) = \frac{\langle \mathcal{O}_i \mathcal{O}_{i+k} \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle}{\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_i \rangle} \quad (25)$$

denoting the normalized autocorrelation function ($A(0) = 1$).

For large time separations k the autocorrelation function decays exponentially,

$$A(k) \xrightarrow{k \rightarrow \infty} a e^{-k/\tau_{\mathcal{O},\text{exp}}} , \quad (26)$$

where $\tau_{\mathcal{O},\text{exp}}$ is the *exponential* autocorrelation time and a is some constant. Due to the exponential decay of $A(k)$ as $k \rightarrow \infty$, in any meaningful simulation with $N \gg \tau_{\mathcal{O},\text{exp}}$,

the correction term in parentheses in (24) can safely be neglected. The usually employed definition of the *integrated* autocorrelation time is thus

$$\tau_{\mathcal{O},\text{int}} = \frac{1}{2} + \sum_{k=1}^N A(k) . \quad (27)$$

Notice that, in general, $\tau_{\mathcal{O},\text{int}}$ (and also $\tau'_{\mathcal{O},\text{int}}$) is different from $\tau_{\mathcal{O},\text{exp}}$. In fact, one can show²⁴ that $\tau_{\mathcal{O},\text{int}} \leq \tau_{\mathcal{O},\text{exp}}$ in realistic models. Only if $A(k)$ is a pure exponential, the two autocorrelation times, $\tau_{\mathcal{O},\text{int}}$ and $\tau_{\mathcal{O},\text{exp}}$, coincide (up to minor corrections for small $\tau_{\mathcal{O},\text{int}}$, see Eq. (46) in Sec. 4 below).²³

The important point of Eq. (23) is that due to temporal correlations of the measurements the statistical error $\epsilon_{\overline{\mathcal{O}}} \equiv \sqrt{\sigma_{\overline{\mathcal{O}}}^2}$ on the MC estimator $\overline{\mathcal{O}}$ is enhanced by a factor of $\sqrt{2\tau_{\mathcal{O},\text{int}}}$. This can be rephrased by writing the statistical error similar to the uncorrelated case as $\epsilon_{\overline{\mathcal{O}}} = \sqrt{\sigma_{\mathcal{O}}^2/N_{\text{eff}}}$, but now with a parameter

$$N_{\text{eff}} = N/2\tau_{\mathcal{O},\text{int}} \leq N , \quad (28)$$

describing the *effective* statistics. This shows more clearly that only every $2\tau_{\mathcal{O},\text{int}}$ iterations the measurements are approximately uncorrelated and gives a better idea of the relevant effective size of the statistical sample. Since some quantities (e.g., the specific heat or susceptibility) can severely be underestimated if the effective statistics is too small,²⁵ any serious simulation should therefore provide at least a rough order-of-magnitude estimate of autocorrelation times.

3.3 Bias

For a better understanding of the latter point, let us consider as a specific example the specific heat, $C = \beta^2 V (\langle e^2 \rangle - \langle e \rangle^2) = \beta^2 V \sigma_{e_i}^2$. The standard estimator for the variance is

$$\hat{\sigma}_{e_i}^2 = \overline{e^2} - \overline{e}^2 = \overline{(e - \overline{e})^2} = \frac{1}{N} \sum_{i=1}^N (e_i - \overline{e})^2 . \quad (29)$$

What is the *expected* value of $\hat{\sigma}_{e_i}^2$? To answer this question, we subtract and add $\langle \overline{e} \rangle^2$,

$$\langle \hat{\sigma}_{e_i}^2 \rangle = \langle \overline{e^2} - \overline{e}^2 \rangle = \langle \overline{e^2} \rangle - \langle \overline{e} \rangle^2 - (\langle \overline{e^2} \rangle - \langle \overline{e} \rangle^2) , \quad (30)$$

and then use the previously derived result: The first two terms on the r.h.s. of (30) just give $\sigma_{e_i}^2$, and the second two terms in parentheses yield $\sigma_{\overline{e}}^2 = \sigma_{e_i}^2 2\tau_{e,\text{int}}/N$, as calculated in (23). Combining these two results we arrive at

$$\langle \hat{\sigma}_{e_i}^2 \rangle = \sigma_{e_i}^2 \left(1 - \frac{2\tau_{e,\text{int}}}{N} \right) = \sigma_{e_i}^2 \left(1 - \frac{1}{N_{\text{eff}}} \right) \neq \sigma_{e_i}^2 . \quad (31)$$

The estimator $\hat{\sigma}_{e_i}^2$ as defined in (29) thus systematically underestimates the true value by a term of the order of $\tau_{e,\text{int}}/N$. Such an estimator is called *weakly biased* (“weakly” because the statistical error $\propto 1/\sqrt{N}$ is asymptotically larger than the systematic bias; for medium or small N , however, also prefactors need to be carefully considered).

We thus see that for large autocorrelation times or equivalently small effective statistics N_{eff} , the bias may be quite large. Since $\tau_{e,\text{int}}$ scales quite strongly with the system size for local update algorithms, some care is necessary in choosing the run time N . Otherwise the FSS of the specific heat and thus the determination of the *static* critical exponent α/ν could be completely spoiled by the temporal correlations!

As a side remark we note that even in the completely uncorrelated case the estimator (29) is biased, $\langle \hat{\sigma}_{e_i}^2 \rangle = \sigma_{e_i}^2 (1 - 1/N)$, since with our conventions in this case $\tau_{e,\text{int}} = 1/2$ (some authors use a different convention in which τ more intuitively vanishes in the uncorrelated case; but this has certain disadvantages in other formulas). In this case one can (and usually does) define a bias-corrected estimator,

$$\hat{\sigma}_{e_i,\text{corr}}^2 = \frac{N}{N-1} \hat{\sigma}_{e_i}^2 = \frac{1}{N-1} \sum_{i=1}^N (e_i - \bar{e})^2 , \quad (32)$$

which obviously satisfies $\langle \hat{\sigma}_{e_i,\text{corr}}^2 \rangle = \sigma_{e_i}^2$. For the squared error on the mean value, this leads to the error formula $\epsilon_{\bar{e}}^2 = \hat{\sigma}_{\bar{e},\text{corr}}^2 = \hat{\sigma}_{e_i,\text{corr}}^2/N = \frac{1}{N(N-1)} \sum_{i=1}^N (e_i - \bar{e})^2$, i.e., to the celebrated replacement of one of the $1/N$ -factors by $1/(N-1)$ “due to one missing degree of freedom”.

3.4 Numerical Estimation of Autocorrelation Times

The above considerations show that not only for the error estimation but also for the computation of static quantities themselves it is important to have control over autocorrelations. Unfortunately, it is very difficult to give reliable a priori estimates, and an accurate numerical analysis is often too time consuming. As a rough estimate it is about ten times harder to get precise information on dynamic quantities than on static quantities like critical exponents. A (biased) estimator $\hat{A}(k)$ for the autocorrelation function is obtained by replacing in (25) the expectation values (ordinary numbers) by mean values (random variables), e.g., $\langle \mathcal{O}_i \mathcal{O}_{i+k} \rangle$ by $\overline{\mathcal{O}_i \mathcal{O}_{i+k}}$. With increasing k the relative variance of $\hat{A}(k)$ diverges rapidly. To get at least an idea of the order of magnitude of $\tau_{\mathcal{O},\text{int}}$ and thus the correct error estimate (23), it is useful to record the “running” autocorrelation time estimator

$$\hat{\tau}_{\mathcal{O},\text{int}}(k_{\max}) = \frac{1}{2} + \sum_{k=1}^{k_{\max}} \hat{A}(k) , \quad (33)$$

which approaches $\tau_{\mathcal{O},\text{int}}$ in the limit of large k_{\max} where, however, its statistical error increases rapidly. As a compromise between systematic and statistical errors, an often employed procedure is to determine the upper limit k_{\max} self-consistently by cutting off the summation once $k_{\max} \geq 6\hat{\tau}_{\mathcal{O},\text{int}}(k_{\max})$. In this case an a priori error estimate is available,^{9,10,23}

$$\epsilon_{\tau_{\mathcal{O},\text{int}}} = \tau_{\mathcal{O},\text{int}} \sqrt{\frac{2(2k_{\max}+1)}{N}} \approx \tau_{\mathcal{O},\text{int}} \sqrt{\frac{12}{N_{\text{eff}}}} . \quad (34)$$

For a 5% relative accuracy one thus needs at least $N_{\text{eff}} \approx 5000$ or $N \approx 10000 \tau_{\mathcal{O},\text{int}}$ measurements. As an order of magnitude estimate consider the 2D Ising model with $L = 100$ simulated with a local update algorithm. The integrated autocorrelation time for this example is of the order of $L^2 \approx 100^2$ (ignoring an a priori unknown prefactor of “order unity”

which depends on the considered quantity), thus implying $N \approx 10^8$. Since in each sweep L^2 spins have to be updated and assuming that each spin update takes about a μsec , we end up with a total time estimate of about 10^6 seconds ≈ 1 CPU-day to achieve this accuracy.

Another possibility is to approximate the tail end of $A(k)$ by a single exponential as in (26). Summing up the small k part exactly, one finds²⁶

$$\tau_{\mathcal{O},\text{int}}(k_{\max}) = \tau_{\mathcal{O},\text{int}} - ce^{-k_{\max}/\tau_{\mathcal{O},\text{exp}}} , \quad (35)$$

where c is a constant. The latter expression may be used for a numerical estimate of both the exponential and integrated autocorrelation times.²⁶

3.5 Binning Analysis

As the preceding discussions have shown, ignoring autocorrelations can lead to a severe underestimation of statistical errors. Invoking the full machinery of autocorrelation analysis, however, is often too cumbersome. On a day by day basis the following binning analysis is much more convenient (though somewhat less accurate). By grouping the original data into bins or blocks, one forms a new, shorter time series which is almost uncorrelated and can thus be analyzed by standard means. But even if the data are completely uncorrelated in time, one still has to handle the problem of error estimation for quantities that are not directly measured in the simulation but are computed as a non-linear combination of “basic” observables. This problem can either be solved by error propagation or by using the Jackknife method described in the next subsection.

Let us assume that the time series consists of N correlated measurements \mathcal{O}_i . One then forms N_B non-overlapping blocks of length k such that $N = N_B k$ (assuming that N was chosen cleverly; otherwise one has to discard some of the data and redefine N) and computes the block average $\mathcal{O}_{B,n}$ of the n -th block,

$$\mathcal{O}_{B,n} \equiv \frac{1}{k} \sum_{i=1}^k \mathcal{O}_{(n-1)k+i} , \quad n = 1, \dots, N_B . \quad (36)$$

The mean value over all block variables obviously satisfies $\overline{\mathcal{O}}_B = \overline{\mathcal{O}}$. If the block length k is large enough ($k \gg \tau$), the blocks are basically uncorrelated in time and their variance can be computed according to the unbiased estimator (32), leading to the squared statistical error of the mean value,

$$\epsilon_{\overline{\mathcal{O}}}^2 \equiv \sigma_{\overline{\mathcal{O}}}^2 = \sigma_B^2/N_B = \frac{1}{N_B(N_B-1)} \sum_{n=1}^{N_B} (\mathcal{O}_{B,n} - \overline{\mathcal{O}}_B)^2 . \quad (37)$$

By comparing with (23) we see that $\sigma_B^2/N_B = 2\tau_{\mathcal{O},\text{int}}\sigma_{\mathcal{O}_i}^2/N$, showing that one may also use

$$2\tau_{\mathcal{O},\text{int}} = k\sigma_B^2/\sigma_{\mathcal{O}_i}^2 \quad (38)$$

for the estimation of $\tau_{\mathcal{O},\text{int}}$. Estimates of $\tau_{\mathcal{O},\text{int}}$ obtained in this way are often referred to as “blocking τ ” or “binning τ ”.

3.6 Jackknife Analysis

Instead of considering rather small blocks of lengths k and their fluctuations as in the binning method, in a Jackknife analysis^{27,28} one forms N_B large Jackknife blocks $\mathcal{O}_{J,n}$ containing all data but one of the previous binning blocks,

$$\mathcal{O}_{J,n} = \frac{N\bar{\mathcal{O}} - k\mathcal{O}_{B,n}}{N-k} , \quad n = 1, \dots, N_B . \quad (39)$$

Each of the Jackknife blocks thus consists of $N - k$ data, i.e., it contains almost as many data as the original time series. When non-linear combinations of basic variables are estimated, the bias is hence comparable to that of the total data set (typically $1/(N - k)$ compared to $1/N$). The N_B Jackknife blocks are, of course, trivially correlated because one and the same original data enter in $N_B - 1$ different Jackknife blocks. This trivial correlation caused by re-using the original data over and over again has nothing to do with temporal correlations. As a consequence the Jackknife block variance σ_J^2 will be much smaller than the variance estimated in the binning method. Because of the trivial nature of the correlations, however, this reduction can be corrected by multiplying σ_J^2 with a factor $(N_B - 1)^2$, leading to

$$\epsilon_{\bar{\mathcal{O}}}^2 \equiv \sigma_{\bar{\mathcal{O}}}^2 = \frac{N_B - 1}{N_B} \sum_{n=1}^{N_B} (\mathcal{O}_{J,n} - \bar{\mathcal{O}}_J)^2 . \quad (40)$$

To summarize this section, any realization of a Markov chain, i.e., MC update algorithm, is characterised by autocorrelation times which enter directly in the statistical errors of MC estimates. Since temporal correlations always increase the statistical errors, it is a very important issue to develop MC update algorithms that keep autocorrelation times as small as possible. This is the reason why cluster and other non-local algorithms are so important.

4 A Simplified Model

It is always useful to have a few exact results available against which the numerical techniques can be checked. Of course, to continue analytically, we have to make some simplifying assumptions about the distribution $P(e)$ from which the e_i are drawn and about the temporal correlations of these measurements. In the following we shall hence assume that the e_i are Gaussian random variables. Furthermore, without loss of generality we simplify the notation and normalize the energies to have zero expectation, $\langle e_i \rangle = 0$, and unit variance, $\langle e_i^2 \rangle = 1$. This is convenient but inessential. Finally, the temporal correlations are modelled by a bivariate time series with correlation coefficient ρ ($0 \leq \rho < 1$),

$$e_0 = e'_0 , \\ e_i = \rho e_{i-1} + \sqrt{1 - \rho^2} e'_i , \quad i \geq 1 , \quad (41)$$

where the e'_i are *independent* Gaussian random variables satisfying $\langle e'_i \rangle = 0$ and $\langle e'_i e'_j \rangle = \delta_{ij}$. By iterating the recursion (41) it is then easy to see that

$$e_k = \rho e_{k-1} + \sqrt{1 - \rho^2} e'_k = \rho^k e_0 + \sqrt{1 - \rho^2} \sum_{l=1}^k \rho^{k-l} e'_l , \quad (42)$$

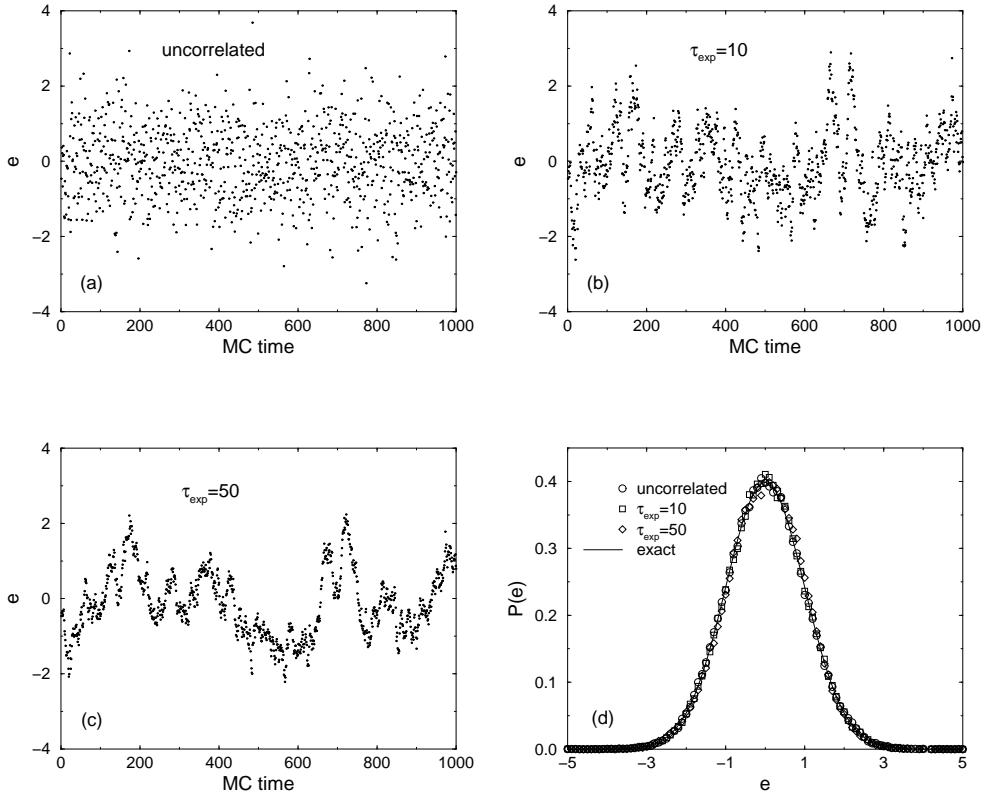


Figure 4. One percent of the “MC time” evolution according to the bivariate Gaussian process (41) in (a) the uncorrelated case respectively with (b) $\tau_{\text{exp}} = 10$ and (c) $\tau_{\text{exp}} = 50$. All three time evolutions with a total of 100 000 consecutive “measurements” lead to the same Gaussian histogram shown in (d).

and consequently

$$A(k) = \langle e_0 e_k \rangle = \rho^k \equiv e^{-k/\tau_{\text{exp}}} . \quad (43)$$

In this simplified model the autocorrelation function is thus a pure exponential with an exponential autocorrelation time given by

$$\tau_{\text{exp}} = -1 / \ln \rho . \quad (44)$$

It should be stressed that in realistic situations a purely exponential decay can only be expected asymptotically for large k where the slowest mode dominates. For smaller time separations usually also many other modes contribute whose correlation time is smaller. The visual appearance of uncorrelated and correlated data with $\tau_{\text{exp}} = 10$ and 50 is depicted in Figs. 4(a)-(c) where in each case one percent of the total “MC time” evolution consisting of 100 000 consecutive “measurements” is shown. Despite the quite distinct temporal evolutions, histogramming the time series leads to the same Gaussian distribution

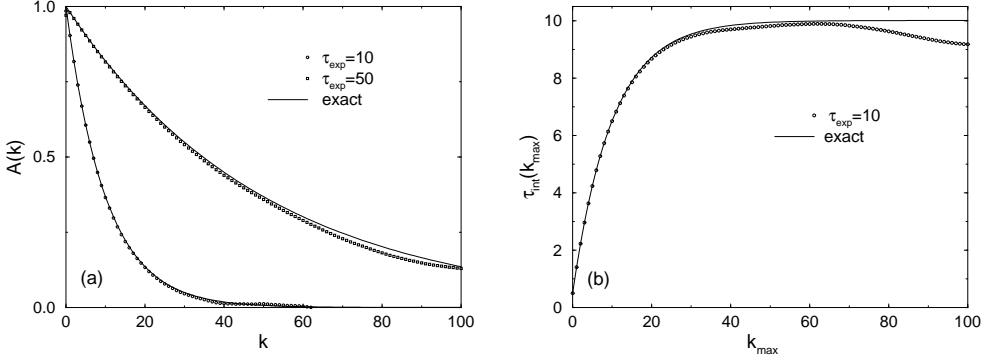


Figure 5. (a) Autocorrelation functions and (b) integrated autocorrelation time for $\tau_{\text{exp}} = 10$ on the basis of 100 000 “measurements” in comparison with exact results for the bivariate Gaussian model shown as the solid lines.

within error bars, as it should, cf. Fig. 4(d). The corresponding autocorrelation functions $A(k)$ are shown in Fig. 5(a).

The integrated autocorrelation time can also be calculated exactly,

$$\tau_{\text{int}} = \frac{1}{2} + \sum_{k=1}^{\infty} A(k) = \frac{1}{2} \frac{1+\rho}{1-\rho} = \frac{1}{2} \coth(1/2\tau_{\text{exp}}) . \quad (45)$$

For $\tau_{\text{exp}} \gg 1$ this can be approximated by

$$\tau_{\text{int}} = \tau_{\text{exp}} \left[1 + \frac{1}{12\tau_{\text{exp}}^2} + \mathcal{O}(1/\tau_{\text{exp}}^4) \right] , \quad (46)$$

i.e., for a purely exponential autocorrelation function we have, to a very good approximation, $\tau_{\text{int}} \approx \tau_{\text{exp}}$, which would immediately follow from $\tau_{\text{int}} \approx \int_0^\infty dk A(k) = \tau_{\text{exp}}$.

As explained in the last section, one usually truncates the summation in (45) self-consistently at about $k_{\max} = 6\tau_{\text{int}}$ ($\approx 6\tau_{\text{exp}}$) since $A(k)$ becomes very noisy for large time separations. Observing that (45) is nothing but a geometric series, also the resulting correction can be calculated exactly,

$$\tau_{\text{int}}(k_{\max}) \equiv \frac{1}{2} + \sum_{k=1}^{k_{\max}} A(k) = \frac{1}{2} \coth(1/2\tau_{\text{exp}}) \left[1 - \frac{2e^{-(k_{\max}+1)/\tau_{\text{exp}}}}{1 + e^{-1/\tau_{\text{exp}}}} \right] , \quad (47)$$

which simplifies in the case of large $\tau_{\text{exp}} \gg 1$ to

$$\tau_{\text{int}}(k_{\max}) = \tau_{\text{int}} \left[1 - \frac{2\tau_{\text{exp}}}{2\tau_{\text{exp}} + 1} e^{-k_{\max}/\tau_{\text{exp}}} \right] , \quad (48)$$

showing that with increasing k_{\max} the asymptotic value of $\tau_{\text{int}} \equiv \tau_{\text{int}}(\infty)$ is approached exponentially fast. This is illustrated in Fig. 5(b) for the bivariate Gaussian time series with $\tau_{\text{exp}} = 10$. Here we also see that for too large k_{\max} the estimate for $\tau_{\text{int}}(k_{\max})$ can deviate quite substantially from the exact value due to its divergent variance. The usually employed self-consistent cutoff would be around $6\tau_{\text{exp}} = 60$ where $\tau_{\text{int}}(k_{\max}) \approx 9.89$.

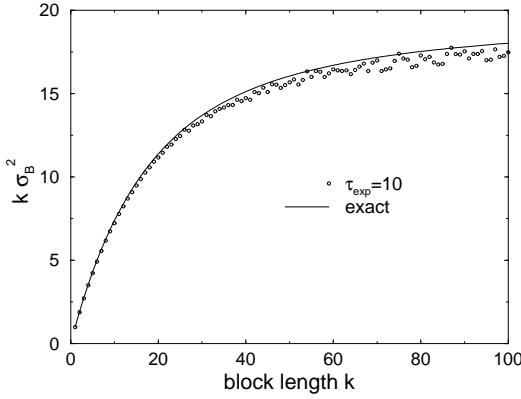


Figure 6. Binning analysis of 100 000 “measurements” in the bivariate Gaussian model. The solid line shows the exact result.

Let us now turn to the binning analysis, assuming that we decompose the total number of measurements N into N_B non-overlapping blocks of length k such that $N = N_B k$. The n th block of measurements then yields a block average of

$$e_{B,n} \equiv \frac{1}{k} \sum_{i=1}^k e_{(n-1)k+i} . \quad (49)$$

In our simple example the expected value is, of course, zero, $\langle e_{B,n} \rangle = \frac{1}{k} \sum_{i=1}^k \langle e_{(n-1)k+i} \rangle = 0$. Therefore, the variance of the block variables is just the expectation value of $e_{B,n}^2$,

$$\begin{aligned} \sigma_B^2 &= \langle e_{B,n}^2 \rangle = \frac{1}{k^2} \sum_{i,j=1}^k \rho^{|i-j|} \\ &= \frac{1}{k^2} \left[k + 2 \sum_{i=1}^k \sum_{j=1}^{i-1} \rho^{|i-j|} \right] \\ &= \frac{1}{k} \left[1 + \frac{2\rho}{1-\rho} - \frac{2\rho}{k} \frac{1-\rho^k}{(1-\rho)^2} \right] . \end{aligned} \quad (50)$$

Recalling (45) this can be rewritten as

$$k\sigma_B^2 = 2\tau_{\text{int}} \left[1 - \frac{\tau_{\text{int}}}{k} \left(1 - e^{-k/\tau_{\text{exp}}} \right) / \cosh^2(1/2\tau_{\text{exp}}) \right] , \quad (51)$$

and for $\tau_{\text{exp}} \gg 1$ to a very good approximation as

$$\begin{aligned} k\sigma_B^2 &\approx 2\tau_{\text{int}} \left[1 - \frac{\tau_{\text{int}}}{k} \left(1 - e^{-k/\tau_{\text{exp}}} \right) \right] \\ &\approx 2\tau_{\text{exp}} \left[1 - \frac{\tau_{\text{exp}}}{k} \left(1 - e^{-k/\tau_{\text{exp}}} \right) \right] , \end{aligned} \quad (52)$$

showing that with increasing block length k the asymptotic value $2\tau_{\text{int}}$ is approached according to a power law. For an illustration see Fig. 6.

5 A Realistic Example

In this section the autocorrelation and error analysis is illustrated for a realistic but still very simple model: The two-dimensional (2D) Ising model, simulated with the Metropolis algorithm at the infinite-volume critical point $\beta_c = \ln(1 + \sqrt{2})/2 \approx 0.440\,686\,793\,4\dots$ on a 16×16 square lattice with periodic boundary conditions. The raw data are the time series with 1 000 000 measurements of the energy and magnetization taken after each sweep over the lattice, after discarding the first 200 000 sweeps to equilibrate the system.

A small part of the time evolution of the energy and magnetization is shown in Fig. 7. Notice the very different time scales for the e and m plots. The energy plot should be compared with the Gaussian model time series in Figs. 4(b) and (c).

Next, using the complete time series the autocorrelation functions were computed according to (25). The only difference to the analysis of the simplified model is that instead of using the Gaussian data one now reads in the Ising model time series. The result for the energy autocorrelations is shown in Fig. 8. On the log-scale of Fig. 8(b) we clearly see the asymptotic linear behaviour of $\ln A(k)$. Apart from the noise for large k , which is also present in the simplified model for finite statistics, the main difference to the artificial data of the simplified model lies in the small k behaviour. For the Ising model we clearly notice an initial fast drop, corresponding to faster relaxing modes, before the asymptotic behaviour sets in. This is, in fact, the generic behaviour of autocorrelation functions in realistic models.

Once the autocorrelation function is known, it is straightforward to sum up the integrated autocorrelation time. The result for the energy is depicted in Fig. 9, yielding an estimate of $\tau_{e,\text{int}} \approx 27$. Also shown is the binning analysis which yields consistent results as it should (the horizontal line shows $2\tau_{e,\text{int}} \approx 54$).

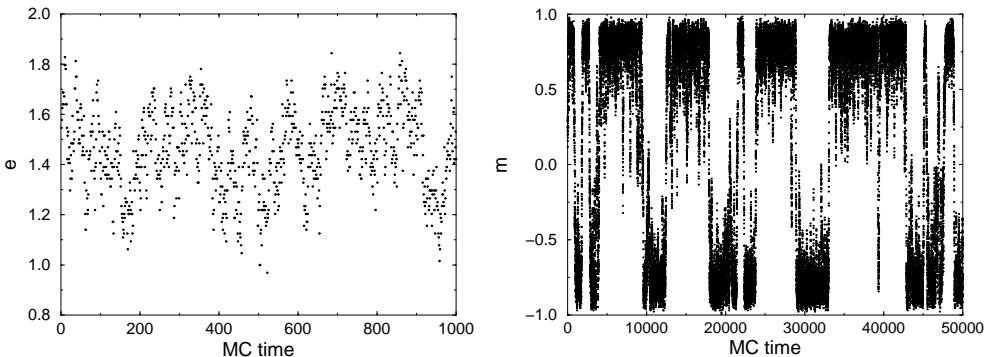


Figure 7. Part of the time evolution of the energy e and magnetization m for the 2D Ising model on a 16×16 lattice at β_c .

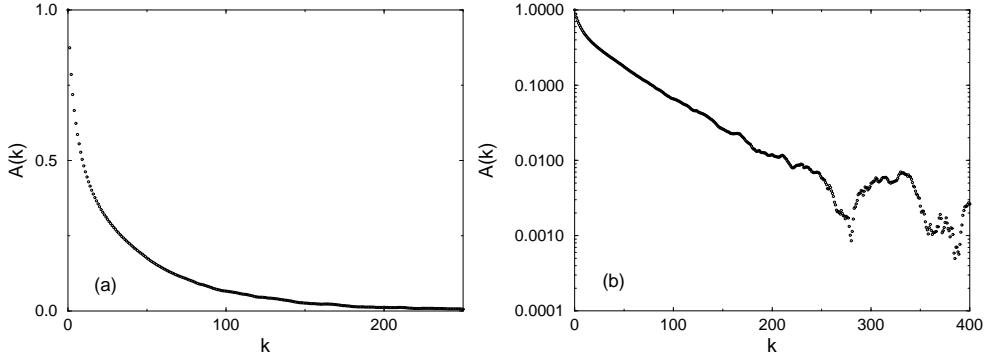


Figure 8. (a) Autocorrelation function of the energy for the 2D Ising model on a 16×16 lattice at β_c . (b) The same data as in (a) on a logarithmic scale, revealing the fast initial drop for small k and the noisy behaviour for large k .

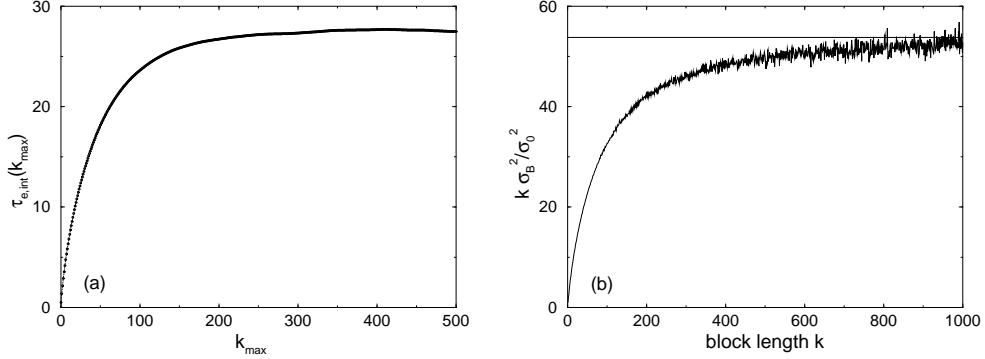


Figure 9. (a) Integrated autocorrelation time and (b) binning analysis for the energy of the 2D Ising model on a 16×16 lattice at β_c . The horizontal line in (b) shows $2\tau_{e,\text{int}} \approx 54$.

6 Error Propagation in Multicanonical Simulations

As a rather extreme example for dealing with non-linear combinations of basic observables we consider now the error analysis of multicanonical simulations.²⁶ As shown in (15), for any observable \mathcal{O} expectation values in the *canonical* ensemble, $\langle \mathcal{O} \rangle_{\text{can}}$, can be calculated as

$$\langle \mathcal{O} \rangle_{\text{can}} = \frac{\langle \mathcal{O} w \rangle}{\langle w \rangle} , \quad (53)$$

where $\langle \dots \rangle$ (without subscript) denote expectation values with respect to the *multicanonical* distribution and $w = \exp(\beta f)$ is the inverse reweighting factor. In a MC simulation with a total number N measurements these values are, as usual, estimated by the mean

values

$$\langle \mathcal{O}w \rangle \approx \overline{\mathcal{O}w} \equiv \frac{1}{N} \sum_{i=1}^N \mathcal{O}_i w_i , \quad (54)$$

$$\langle w \rangle \approx \overline{w} \equiv \frac{1}{N} \sum_{i=1}^N w_i , \quad (55)$$

where \mathcal{O}_i and w_i denote the measurements for the i -th configuration. Hence $\langle \mathcal{O} \rangle_{\text{can}}$ is estimated by

$$\langle \mathcal{O} \rangle_{\text{can}} \approx \hat{\mathcal{O}} \equiv \frac{\overline{\mathcal{O}w}}{\overline{w}} . \quad (56)$$

The estimator $\hat{\mathcal{O}}$ is biased,

$$\langle \hat{\mathcal{O}} \rangle = \langle \mathcal{O} \rangle_{\text{can}} \left[1 - \frac{\langle \overline{\mathcal{O}w}; \overline{w} \rangle}{\langle \overline{\mathcal{O}w} \rangle \langle \overline{w} \rangle} + \frac{\langle \overline{w}; \overline{w} \rangle}{\langle \overline{w} \rangle \langle \overline{w} \rangle} + \dots \right] , \quad (57)$$

and fluctuates around $\langle \hat{\mathcal{O}} \rangle$ with variance, i.e. squared statistical error

$$\epsilon_{\hat{\mathcal{O}}}^2 = \langle \mathcal{O} \rangle_{\text{can}}^2 \left[\frac{\langle \overline{\mathcal{O}w}; \overline{\mathcal{O}w} \rangle}{\langle \overline{\mathcal{O}w} \rangle^2} + \frac{\langle \overline{w}; \overline{w} \rangle}{\langle \overline{w} \rangle^2} - 2 \frac{\langle \overline{\mathcal{O}w}; \overline{w} \rangle}{\langle \overline{\mathcal{O}w} \rangle \langle \overline{w} \rangle} + \dots \right] . \quad (58)$$

Here we used the abbreviation

$$\langle \overline{\mathcal{O}w}; \overline{w} \rangle \equiv \langle \overline{\mathcal{O}w} \overline{w} \rangle - \langle \overline{\mathcal{O}w} \rangle \langle \overline{w} \rangle , \quad (59)$$

etc. to denote (connected) correlations of the mean values, which can be computed as

$$\langle \overline{\mathcal{O}w}; \overline{w} \rangle = \langle \mathcal{O}_i w_i; w_i \rangle \frac{2\tau_{\mathcal{O}w;w}^{\text{int}}}{N} , \quad (60)$$

where

$$\tau_{\mathcal{O}w;w}^{\text{int}} = \tau_{w;\mathcal{O}w}^{\text{int}} = \frac{1}{2} + \sum_{k=1}^N \frac{\langle \mathcal{O}_0 w_0; w_k \rangle}{\langle \mathcal{O}_0 w_0; w_0 \rangle} \left(1 - \frac{k}{N} \right) \quad (61)$$

is the associated integrated autocorrelation time of measurements in the multicanonical distribution. Hence the statistical error is given by

$$\begin{aligned} \epsilon_{\hat{\mathcal{O}}}^2 = \langle \mathcal{O} \rangle_{\text{can}}^2 & \left[\frac{\langle \mathcal{O}_i w_i; \mathcal{O}_i w_i \rangle}{\langle \mathcal{O}_i w_i \rangle^2} \frac{2\tau_{\mathcal{O}w;\mathcal{O}w}^{\text{int}}}{N} + \frac{\langle w_i; w_i \rangle}{\langle w_i \rangle^2} \frac{2\tau_{w;w}^{\text{int}}}{N} \right. \\ & \left. - 2 \frac{\langle \mathcal{O}_i w_i; w_i \rangle}{\langle \mathcal{O}_i w_i \rangle \langle w_i \rangle} \frac{2\tau_{\mathcal{O}w;w}^{\text{int}}}{N} \right] . \end{aligned} \quad (62)$$

Since for uncorrelated measurements $\tau_{\mathcal{O}w;\mathcal{O}w}^{\text{int}} = \tau_{\mathcal{O}w;w}^{\text{int}} = \tau_{w;w}^{\text{int}} = 1/2$, it is useful to define an *effective* multicanonical variance^a

$$\sigma_{\hat{\mathcal{O}}}^2 = \langle \mathcal{O} \rangle_{\text{can}}^2 \left[\frac{\langle \mathcal{O}_i w_i; \mathcal{O}_i w_i \rangle}{\langle \mathcal{O}_i w_i \rangle^2} + \frac{\langle w_i; w_i \rangle}{\langle w_i \rangle^2} - 2 \frac{\langle \mathcal{O}_i w_i; w_i \rangle}{\langle \mathcal{O}_i w_i \rangle \langle w_i \rangle} \right] , \quad (63)$$

^aIn the multicanonical distribution this is nothing but an *abbreviation* of the expression on the r.h.s. but *not* the variance in the multicanonical distribution.

3D 4-state Potts model, multibondic simulations							
L	$\tau_{E;E}^{\text{int}}$	$\tau_{Ew;Ew}^{\text{int}}$	$\tau_{w;w}^{\text{int}}$	$\tau_{Ew;w}^{\text{int}}$	τ_e^{eff}	τ_e^{jack}	τ_e^{flip}
8	71(5)	39(2)	9(1)	18(1)	77(5)	63	119(2)
10	95(10)	62(3)	38(2)	50(3)	107(8)	103	181(3)
12	148(12)	81(3)	53(2)	66(3)	189(34)	205	298(4)
14	229(24)	105(3)	74(2)	87(3)	316(21)	326	468(6)
16	303(27)	131(3)	100(4)	113(3)	488(65)	498	655(10)
20	584(41)	206(5)	166(4)	183(4)	940(58)	1009	1298(18)
24	1008(427)	280(10)	239(8)	256(11)	1607(473)	1471	2434(82)
30	2730(1293)	340(15)	334(13)	324(13)	3085(173)	4972	5429(238)
L	$\tau_{B;B}^{\text{int}}$	$\tau_{Bw;Bw}^{\text{int}}$	$\tau_{w;w}^{\text{int}}$	$\tau_{Bw;w}^{\text{int}}$	τ_b^{eff}	τ_b^{jack}	τ_b^{flip}
8	69(5)	40(2)	9(1)	18(1)	75(5)	62	106(2)
10	93(10)	62(3)	38(2)	50(3)	107(7)	102	177(3)
12	146(12)	81(3)	53(2)	66(3)	189(34)	204	286(4)
14	225(23)	104(3)	74(2)	87(3)	317(21)	327	451(6)
16	299(27)	131(3)	100(4)	113(3)	490(64)	498	649(10)
20	578(41)	206(5)	166(4)	183(4)	941(58)	1009	1281(18)
24	1005(427)	280(10)	239(8)	256(10)	1611(449)	1479	2418(83)
30	2723(1364)	340(15)	334(13)	324(13)	3081(172)	4978	5429(237)

Table 1. Autocorrelation times in multibondic simulations of the three-dimensional 4-state Potts model. Error estimates are obtained with the Jackknife method on the basis of 100 blocks for $L = 8 - 20$, 50 blocks for $L = 24$, and 40 blocks for $L = 30$.

such that the squared error (62) can be written in the usual form

$$\epsilon_{\mathcal{O}}^2 \equiv \sigma_{\mathcal{O}}^2 \frac{2\tau_{\mathcal{O}}}{N} , \quad (64)$$

with $\tau_{\mathcal{O}}$ collecting the various autocorrelation times in an averaged sense. For a comparison with canonical simulations we need one further step since

$$(\epsilon_{\mathcal{O}}^2)^{\text{can}} = \langle \overline{\mathcal{O}}; \overline{\mathcal{O}} \rangle_{\text{can}} = (\sigma_{\mathcal{O}_i}^2)^{\text{can}} \frac{2\tau_{\mathcal{O}}^{\text{can}}}{N} \quad (65)$$

but $\sigma_{\mathcal{O}}^2 \neq (\sigma_{\mathcal{O}_i}^2)^{\text{can}} = \langle \mathcal{O}_i; \mathcal{O}_i \rangle$. Hence we define an effective autocorrelation time $\tau_{\mathcal{O}}^{\text{eff}}$ through

$$\epsilon_{\mathcal{O}}^2 = (\sigma_{\mathcal{O}_i}^2)^{\text{can}} \frac{2\tau_{\mathcal{O}}^{\text{eff}}}{N} = (\epsilon_{\mathcal{O}}^2)^{\text{can}} \frac{\tau_{\mathcal{O}}^{\text{eff}}}{\tau_{\mathcal{O}}^{\text{can}}} , \quad (66)$$

i.e.,

$$\tau_{\mathcal{O}}^{\text{eff}} = \frac{\sigma_{\mathcal{O}}^2}{(\sigma_{\mathcal{O}_i}^2)^{\text{can}}} \tau_{\mathcal{O}} . \quad (67)$$

For symmetric distributions and *odd* observables we have $\langle \mathcal{O}_i w_i \rangle \equiv 0$ and this simplifies to

$$\epsilon_{\mathcal{O}}^2 = \frac{\langle \mathcal{O}_i w_i; \mathcal{O}_i w_i \rangle}{\langle w_i \rangle^2} 2\tau_{\mathcal{O}w;\mathcal{O}w}^{\text{int}} , \quad (68)$$

3D 4-state Potts model, multibondic simulations						
L	$\frac{\langle Ew; Ew \rangle}{\langle Ew \rangle^2}$	$\frac{\langle w; w \rangle}{\langle w \rangle^2}$	$\frac{\langle Ew; w \rangle}{\langle Ew \rangle \langle w \rangle}$	$\sigma_e^2 \times 10^2$	$\varepsilon_e \times 10^2$	$\varepsilon_e^{\text{jack}} \times 10^2$
8	0.11208(70)	0.05993(28)	0.07032(46)	9.751(67)	0.462(13)	0.417
10	0.2429(25)	0.15050(89)	0.1799(16)	8.363(70)	0.423(15)	0.415
12	0.4471(52)	0.3418(26)	0.3759(37)	7.731(71)	0.383(45)	0.398
14	0.6555(76)	0.5422(42)	0.5790(57)	7.331(83)	0.393(14)	0.399
16	1.047(14)	0.9210(87)	0.961(11)	6.988(92)	0.369(25)	0.373
20	1.561(22)	1.433(15)	1.468(18)	6.992(94)	0.363(13)	0.376
24	2.175(71)	2.038(56)	2.077(62)	6.15(22)	0.70(11)	0.673
30	2.555(91)	2.328(65)	2.381(73)	8.11(18)	1.119(32)	1.420

Table 2. Variances, covariances and expectation values of multibondic simulations of the three-dimensional 4-state Potts model, which enter the effective error estimate ε_e . σ_e^2 is the canonical variance of the energy. For the number of Jackknife blocks used in the computation of $\varepsilon_e^{\text{jack}}$ see Table 1.

such that

$$\tau_{\mathcal{O}} = \tau_{\mathcal{O}w; \mathcal{O}w}^{\text{int}} , \quad (69)$$

and

$$\tau_{\mathcal{O}}^{\text{eff}} = \frac{\sigma_{\mathcal{O}}^2}{(\sigma_{\mathcal{O}_i}^2)^{\text{can}}} \tau_{\mathcal{O}w; \mathcal{O}w}^{\text{int}} . \quad (70)$$

These formulas have been tested and compared with a standard Jackknife analysis in Refs. 26 and 29. As an example the various quantities computed are shown in Tables 1 and 2 for the case of the 3D 4-state Potts model simulated with the multibondic algorithm,^{30,31} a cluster algorithm variant of the multicanonical method. Here E denotes as usual the energy and B is the number of active bonds.

7 Summary

Thanks to the great improvements in Monte Carlo simulation methodology over the last decade, the results for at least certain model classes have reached such a high degree of precision that careful and reliable statistical error analyses become more and more important. The interplay of autocorrelations and correlations between basic observables in non-linear combined quantities requires rather involved analyses which are well understood in principle but still a bit cumbersome to implement in practice. Still, after spending months or even years of computer time for the generation of the raw data, it is certainly worth spending this comparatively little extra effort; since a Monte Carlo simulation is a stochastic method, trustworthy error estimates are an important part of the final results.

Acknowledgments

I would like to thank Bernd Berg, Stefan Kappler, Tilman Sauer and Martin Weigel for enjoyable collaborations and many useful discussions on various aspects of this lecture.

References

1. D. Frenkel and B. Smit, *Understanding Molecular Simulation – From Algorithms to Applications* (Academic Press, San Diego, 1996).
2. D.P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).
3. There is by now a huge amount of Ising model material available on the World Wide Web, including animations. For a list of useful links, see e.g. <http://oscar.cacr.caltech.edu/Hrothgar/Ising/references.html>.
4. W. Janke, *Monte Carlo Simulations of Spin Systems*, in: *Computational Physics: Selected Methods – Simple Exercises – Serious Applications*, eds. K.H. Hoffmann and M. Schreiber (Springer, Berlin, 1996), p. 10.
5. R.B. Potts, Proc. Camb. Phil. Soc. **48**, 106 (1952).
6. F.Y. Wu, Rev. Mod. Phys. **54**, 235 (1982); *ibid.* **55**, 315(E) (1983).
7. W. Janke and R. Villanova, Nucl. Phys. **B489**, 679 (1997).
8. W. Janke, *Nonlocal Monte Carlo Algorithms for Statistical Physics Applications*, Mathematics and Computers in Simulations **47**, 329 (1998).
9. A.D. Sokal, *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*, lecture notes, Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne, 1989.
10. A.D. Sokal, *Bosonic Algorithms*, in: *Quantum Fields on the Computer*, ed. M. Creutz (World Scientific, Singapore, 1992), p. 211.
11. J.D. Gunton, M.S. Miguel, and P.S. Sahni, in: *Phase Transitions and Critical Phenomena*, Vol. 8, eds. C. Domb and J.L. Lebowitz (Academic Press, New York, 1983), p. 269.
12. K. Binder, Rep. Prog. Phys. **50**, 783 (1987).
13. H.J. Herrmann, W. Janke, and F. Karsch (eds.), *Dynamics of First Order Phase Transitions* (World Scientific, Singapore, 1992).
14. W. Janke, *Recent Developments in Monte Carlo Simulations of First-Order Phase Transitions*, in: *Computer Simulations in Condensed Matter Physics VII*, eds. D.P. Landau, K.K. Mon, and H.-B. Schüttler (Springer, Berlin, 1994), p. 29.
15. B.A. Berg and T. Neuhaus, Phys. Lett. **B267**, 249 (1991).
16. B.A. Berg and T. Neuhaus, Phys. Rev. Lett. **69**, 9 (1992).
17. W. Janke, in Ref. 13, p. 365.
18. B.A. Berg, Fields Inst. Commun. **26**, 1 (2000).
19. W. Janke, *Monte Carlo Methods for Sampling of Rare Event States*, in: *Classical and Quantum Dynamics in Condensed Phase Simulations*, Proceedings of the International School of Physics “Computer Simulation of Rare Events and the Dynamics of Classical and Quantum Condensed-Phase Systems” and Euroconference on “Technical advances in Particle-based Computational Material Sciences”, Lerici, Italy, July 7 – 18, 1997, eds. B. Berne, G. Ciccotti, and D. Coker (World Scientific, Singapore, 1998), p. 195.
20. W. Janke, *Multicanonical Monte Carlo Simulations*, Physica **A254**, 164 (1998).
21. M.B. Priestley, *Spectral Analysis and Time Series*, 2 vols. (Academic, London, 1981), Chapters 5-7.
22. T.W. Anderson, *The Statistical Analysis of Time Series* (Wiley, New York, 1971).

23. N. Madras and A.D. Sokal, J. Stat. Phys. **50**, 109 (1988).
24. A.D. Sokal and L.E. Thomas, J. Stat. Phys. **54**, 797 (1989).
25. A.M. Ferrenberg, D.P. Landau, and K. Binder, J. Stat. Phys. **63**, 867 (1991).
26. W. Janke and T. Sauer, J. Stat. Phys. **78**, 759 (1995).
27. B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics [SIAM], Philadelphia, 1982).
28. R.G. Miller, Biometrika **61**, 1 (1974).
29. M.S. Carroll, W. Janke, and S. Kappler, J. Stat. Phys. **90**, 1277 (1998).
30. W. Janke and S. Kappler, Phys. Rev. Lett. **74**, 212 (1995).
31. W. Janke, *Multicanonical Multigrid and Multibondic Algorithms*, in: *Multiscale Phenomena and Their Simulation*, Proceedings of the International Conference, Bielefeld, September 30 – October 4, 1996, eds. F. Karsch, B. Monien, and H. Satz (World Scientific, Singapore, 1997), p. 147.
32. U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
33. G. C. Fox, M. A. Johnson, G. A. Lyzenga, S. W. Otto, J. K. Salmon, and D. W. Walker, *Solving Problems on Concurrent Processors: Volume 1* (Prentice Hall, Englewood Cliffs, NJ, 1988).
34. MD simulation of an FCC lattice with 1.213.857.792 atoms, see <http://www.itap.physik.uni-stuttgart.de/~joerg/imd.html>
35. MD simulation of a 1 μ s trajectory of a small protein in water: (a) Y. Duan, L. Wang and P. A. Kollman, Proc. Natl. Acad. Sci. USA **9**, 9897 (1998). (b) Y. Duan and P. A. Kollman, Science **282**, 740 (1998).
36. <http://www.cs.sandia.gov/~sjplimp/main.html>
37. <http://www.cs.sandia.gov/~sjplimp/crada.html>

Pseudo Random Numbers: Generation and Quality Checks

Wolfhard Janke

Institut für Theoretische Physik, Universität Leipzig
Augustusplatz 10/11, 04109 Leipzig, Germany
E-mail: wolfhard.janke@itp.uni-leipzig.de

Monte Carlo simulations rely on the quality of pseudo random numbers. Some of the most common algorithms for the generation of uniformly distributed pseudo random numbers and a few physically motivated quality checks are described. Non-uniform distributions of random numbers are also briefly discussed.

1 Introduction

Molecular dynamics and Monte Carlo simulations are important numerical techniques in classical and quantum statistical physics.^{1,2} Being a stochastic method, Monte Carlo simulations rely heavily on the use of random numbers. Other important areas making use of random numbers include stochastic optimization techniques and cryptography. In practice, random numbers are generated by deterministic recursive rules, formulated in terms of simple arithmetic operations. Obviously the emerging numbers can at best be pseudo random, and it is a great challenge to design random number generators that approximate “true randomness” as closely as possible. Besides this obvious requirement, pseudo random number generators should yield reproducible results, should be portable between different computer architectures, and should be as efficient as possible since in most applications many millions of random numbers are needed.

There is by now a huge literature on this topic,^{3–9} and a simple search in the World-Wide-Web yields hundreds of useful links. The purpose of these lecture notes is to give a brief introduction into the most commonly used pseudo random number generators and to describe a few of the quality checks performed on them which are particularly relevant for Monte Carlo simulation studies.

2 Pseudo Random Number Generators

2.1 Linear Congruential Generators

Among the simplest algorithms for pseudo random numbers are the linear congruential generators¹⁰ which are based on the integer recursion

$$X_{i+1} = (aX_i + c) \bmod m , \quad (1)$$

where the integers a , c and m are constants. These generators can be further classified into mixed ($c > 0$) and multiplicative ($c = 0$) types, usually denoted by LCG(a, c, m) and MLCG(a, m), respectively. A LCG generates a sequence of pseudo random integers X_1, X_2, \dots between 0 and $m - 1$; for a MLCG the lower bound is 1. Each X_i is then

scaled into the interval [0,1). If the multiplier a is a primitive root modulo m and m is prime, the period of this generator is $m - 1$.

A commonly used choice of parameters for the MLCG is the miracle number $a = 16\,807 = 7^5$ and $m = 2^{31} - 1$. This yields the GGL generator¹¹ (sometimes also denoted by CONG or RAN0¹²)

$$X_{i+1} = (16\,807 X_i) \bmod (2^{31} - 1) , \quad (2)$$

which has extensively been used on IBM computers. The period of $2^{31} - 2 \approx 2.15 \times 10^9$ is relatively short, however, and can easily be exhausted in present day simulations (100 000 Monte Carlo sweeps of a 100×100 lattice consist of 10^9 spin updates). Another known problem of this generator is that D -dimensional vectors (x_1, x_2, \dots, x_D) , $(x_{D+1}, x_{D+2}, \dots, x_{2D})$, \dots formed by consecutive normalized pseudo random numbers $x_i \in [0, 1]$ lie on a relatively small number of parallel hyperplanes. As will be shown in the next section, this is already clearly visible in the smallest non-trivial case $D = 2$.

Also the generator G05FAF of the NAG software package¹³ employs a multiplicative linear congruential algorithm MLCG(13^{13} , 2^{59}) or

$$X_{i+1} = (13^{13} X_i) \bmod 2^{59} , \quad (3)$$

which apart from the much longer period of $2^{59} - 1 \approx 5.76 \times 10^{17}$ has on vector computers the technical advantage that a vector of n pseudo random numbers agrees exactly with n successive calls of the G05FAF subroutine.

A frequently used MLCG is RANF,^{14,15} originally implemented on (64 bit) CDC computers and later taken as the standard random number generator on CRAY vector and T3D or T3E parallel computers. This generator uses two linear congruential recursions with modulus 2^{48} , i.e., it is a combination of MLCG($M_1, 2^{48}$) and MLCG($M_{64}, 2^{48}$) with

$$X_{i+1} = (M_1 X_i) \bmod 2^{48} , \quad (4)$$

$$X_{i+64} = (M_{64} X_i) \bmod 2^{48} , \quad (5)$$

where $M_1 = 44\,485\,709\,377\,909$ and $M_{64} = 247\,908\,122\,798\,849$. The period length of RANF is¹⁶ $2^{46} \approx 7.04 \times 10^{14}$ which is already long enough for most applications.

Another popular choice of parameters yields the generator RAND = LCG(69 069, 1, 2^{32}) with a rather short period of $2^{32} \approx 4.29 \times 10^9$, i.e., the recursion

$$X_{i+1} = (69\,069 X_i + 1) \bmod 2^{32} , \quad (6)$$

whose lattice structure is improved for small dimensions but also becomes poor for higher dimensions ($D \geq 6$). The multiplier 69 069 was strongly recommended by Marsaglia¹⁷ and is part of the so-called SUPER-DUPER generator¹⁴ which explains its popularity.

2.2 Lagged Fibonacci Generators

To increase the rather short period of linear congruential generators, it is natural to generalize them to the form

$$X_i = (a_1 X_{i-1} + \dots + a_r X_{i-r}) \bmod m , \quad (7)$$

with $r > 1$ and $a_r \neq 0$. The maximum period is then $m^r - 1$. The special choice $r = 2$, $a_1 = a_2 = 1$ leads to the Fibonacci generator

$$X_i = (X_{i-1} + X_{i-2}) \bmod m , \quad (8)$$

whose properties are, however, relatively poor. This has led to the introduction of lagged Fibonacci generators which are initialized with r integers X_1, X_2, \dots, X_r . Similar to (8) one then uses the recursion

$$X_i = (X_{i-r} \otimes X_{i-s}) \bmod m , \quad (9)$$

where $s < r$ and \otimes stands short for one of the binary operations $+$, $-$, \times , or the exclusive-or operation \oplus (XOR). These generators are denoted by $\text{LF}(r, s, m, \otimes)$. For the usually used addition or subtraction modulo 2^w (w is the word length in bits) the maximal period with suitable choices of r and s is $(2^r - 1)2^{w-1} \approx 2^{r+w-1}$.

An important example is $\text{RAN3} = \text{LF}(55, 24, m, -)$ or

$$X_i = (X_{i-55} - X_{i-24}) \bmod m , \quad (10)$$

where for instance in the Numerical Recipes implementation¹⁸ $m = 10^9$ is used. The period length of this specific generator is known¹⁹ to be $2^{55} - 1 \approx 3.60 \times 10^{16}$.

2.3 Shift Register Generators

Another important class of pseudo random number generators are provided by generalized feedback shift register algorithms²⁰ which are sometimes also called Tausworthe²¹ generators. They are based on the theory of primitive trinomials of the form $x^p + x^q + 1$, and are denoted by $\text{GFSR}(p, q, \oplus)$, where \oplus stands again for the exclusive-or operation XOR, or in formulas

$$X_i = X_{i-p} \oplus X_{i-q} . \quad (11)$$

The maximal possible period of $2^p - 1$ of this generator is achieved when the primitive trinomial $x^p + x^q + 1$ divides $x^n - 1$ for $n = 2^p - 1$, but for no smaller value of n . This condition can be met by choosing p to be a Mersenne prime, that is a prime number p for which $2^p - 1$ is also a prime.

A standard choice for the parameters is $p = 250$ and $q = 103$. This is the (in)famous (see below) R250 generator $\text{GFSR}(250, 103, \oplus)$ or

$$X_i = X_{i-250} \oplus X_{i-103} . \quad (12)$$

The period of R250 is $2^{250} - 1 \approx 1.81 \times 10^{75}$. In one of the earliest implementations²² the $\text{MLCG}(16\,807, 2^{31} - 1)$, i.e. the GGL recursion (2), was used to initialize the first 250 integers, but many different approaches for the initialization are possible and were indeed used in the literature (a fact which complicates comparisons).

2.4 Combined Algorithms

If done carefully, the combination of two different generators may improve the performance. An often employed generator based on this construction is the RANMAR generator.^{23,24} In the first step it employs a lagged Fibonacci generator,

$$X_i = \begin{cases} X_{i-97} - X_{i-33} , & \text{if } X_{i-97} \geq X_{i-33} , \\ X_{i-97} - X_{i-33} + 1 , & \text{otherwise .} \end{cases} \quad (13)$$

Only 24 most significant bits are used for single precision reals. The second part of the generator is a simple arithmetic sequence for the prime modulus $2^{24} - 3 = 16\,777\,213$,

$$Y_i = \begin{cases} Y_i - c & , \text{ if } Y_i \geq c \\ Y_i - c + d & , \text{ otherwise } \end{cases}, \quad (14)$$

where $c = 7\,654\,321/16\,777\,216$ and $d = 16\,777\,213/16\,777\,216$. The final random number Z_i is then produced by combining the obtained X_i and Y_i as

$$Z_i = \begin{cases} X_i - Y_i & , \text{ if } X_i \geq Y_i \\ X_i - Y_i + 1 & , \text{ otherwise } \end{cases}. \quad (15)$$

The total period of RANMAR is about²³ $2^{144} \approx 2.23 \times 10^{43}$. This generator has become very popular in high-statistics Monte Carlo simulations.

2.5 Marsaglia-Zaman Generator

The Marsaglia-Zaman generator is based on the so-called “subtract-and-borrow” algorithm.^{25,26} It is similar to the lagged Fibonacci generator but supplemented with an extra carry bit. If X_i and b are integers with

$$0 \leq X_i \leq b, \quad i < n, \quad (16)$$

then the recursion involving two lags p and q works according to the following prescription. For $n \geq q$ one first computes the difference

$$\Delta_n = X_{n-p} - X_{n-q} - c_{n-1}, \quad (17)$$

where $c_{n-1} = 0$ or 1 is the carry bit. One then determines X_n and c_n through

$$X_n = \begin{cases} \Delta_n & , c_n = 0 & , \text{ if } \Delta_n \geq 0 \\ \Delta_n + b & , c_n = 1 & , \text{ if } \Delta_n < 0 \end{cases}. \quad (18)$$

To start the recursion, the first q values X_0, X_1, \dots, X_{q-1} together with the carry bit c_{q-1} must be initialized. The generator with the particular choice of parameters $b = 2^{24}$, $p = 24$, and $q = 10$ is known²⁴ under the name RCARRY.^a It has the tremendously long period of^{24,28} $(2^{24})^{24}/48 \approx 2^{570}$ or about 5.15×10^{171} .

2.6 The “Luxury” RANLUX Generator

Also for the RCARRY generator some deficiencies in empirical tests of randomness were reported in the literature.⁵ By analysing Fibonacci generators from the viewpoint of multi-dimensional chaos, Lüscher²⁸ showed that there was slow and hence poor divergence from nearby initial points. Based on these results, James^{29,30} implemented the so-called “luxury” pseudo random number generator RANLUX, in which a certain number of points is discarded following each pass through the recirculation buffer. If the “luxury” level is LUX=0, no points are skipped (and RANLUX runs as RCARRY), if LUX=1, then after 24 values have been returned, 24 more values are discarded. Similarly, for LUX=2, 3,

^aNotice that the FORTRAN code for this algorithm printed in Ref. 24 contains a typo:²⁷ In the line UNI=SEEDS(I24)-SEEDS(J24)-CARRY, the indices I24 and J24 should be interchanged.

and 4, the routine discards 73, 199, and 365 values, respectively. In the high “luxury” levels this generator is rather slow, but the quality of the resulting pseudo random numbers is considered to be extremely good. RANLUX has therefore become very popular in the Monte Carlo community, in particular for computationally demanding problems where only a small portion of the total computing time is spent on the generation of pseudo random numbers (such as for instance lattice QCD with dynamical fermions) and for testing purposes. Recently it has been coded in PC assembler language³¹ and the FORTRAN90 versions have also been accelerated by conversion to integer arithmetic.³²

3 Quality Checks

The evaluation of the quality of pseudo random numbers is a difficult problem which has no unique solution. On the one hand there is no single practical test that can verify the realization of randomness in a given pseudo random number sequence. On the other hand, since all pseudo random number generators are based on deterministic rules, there exists always a test in which a given generator will fail. Given this situation one can only try to find some criteria which test at least the most fundamental properties of such “random” sequences. There is a well standardized set of statistical tests¹⁹ such as the uniformity test, the serial test, the gap test, the maximum t test, the collision test, the run test and the park test, bit level tests, the spectral test and visual tests which are well described in the comparative study of many random number generators by Vattulainen *et al.*^{5,6}

Among the most impressive visual tests is the search for lattice structures when consecutive pseudo random numbers are plotted as D -tupels. As can be inspected in Fig. 1, for a multiplicative linear congruential pseudo random number generator a lattice structure is already clearly visible in the smallest non-trivial case $D = 2$ when a small portion of, say, the x -axis is magnified. For this plot a sequence of 10^7 random numbers was generated with the GGL recursion (2), starting with the “pi”-seed $X_0 = 314159$. This gave a mean

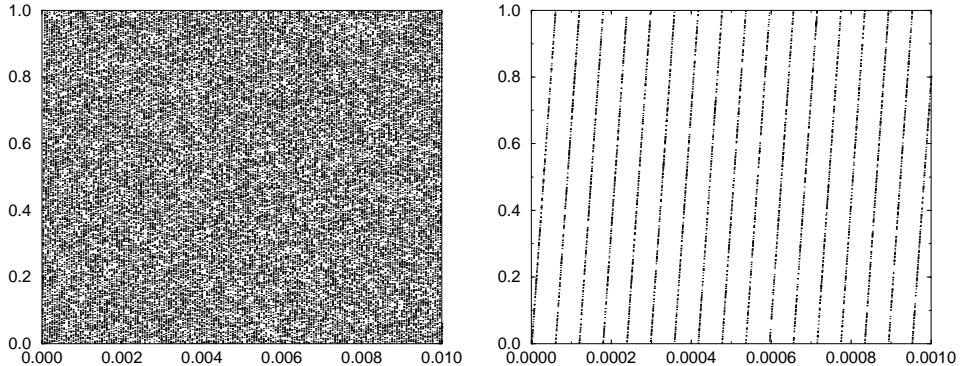


Figure 1. The two-dimensional distribution of 5×10^6 pseudo random number pairs (x_i, x_{i+1}) generated with the GGL recursion (2). While on the scale of the plot on the l.h.s. the distribution “looks” random, on the expanded scale of the plot on the r.h.s. an ordered line structure is clearly revealed.

value of $\bar{x} = 0.499\,795 \approx 1/2$ and a variance of $\sigma^2 = \overline{x^2} - \bar{x}^2 = 0.080\,08 \approx 1/12$. The failure of the MLCG can easily be highlighted by intentionally using a “poor” choice of parameters, for instance³³ $X_{i+1} = (5X_i) \bmod 2^7$, $X_0 = 1$, where the problem becomes immediately obvious.

When subjected to the various mathematical tests mentioned above, the R250 shift register generator turned out to be among the best generators. Consequently it has been used in many Monte Carlo studies. It, therefore, came as a surprise when Ferrenberg *et al.*³⁴ reported severe problems with this generator in applications to Monte Carlo simulations of the two-dimensional Ising model using the single-cluster Wolff update algorithm.³⁵ More precisely they performed simulations of a 16×16 square lattice with periodic boundary conditions at the exactly known infinite-volume transition point $\beta_c = \ln(1 + \sqrt{2})/2$. They generated 10 runs with 10^6 clusters each (which, on the average, cover at β_c about 55% of the lattice sites). As a result Ferrenberg *et al.* obtained for the energy a systematic deviation from the exact value³⁶ of about 42σ (at an accuracy level of 0.003%) and for the specific heat an even larger deviation of about -107σ (at an accuracy level of 0.03%). Further simulations with the same statistics but other pseudo random number generators behaved perfectly well.

Subsequently these results have been confirmed by many other authors,³⁷ and consensus has been reached that triplet correlations in $\langle x_n x_{n-k} x_{n-250} \rangle$ around $k = 147$ are the origin of the problem.³⁸ Notice that due to “time-reversal symmetry” this value is equivalent³⁹ to $k = 250 - 147 = 103$ – a value that just coincides with the second parameter $q = 103$ of the R250 generator! While the numerically determined correlator³⁸ reproduces the theoretically expected value of $(1/2)^3 = 0.125$ for almost all values of k , it drops down to about 0.107 at $k = 147$ or equivalently³⁹ at $k = 103$. This correlation does not only affect the cluster algorithm but also the Metropolis update which was shown³⁸ to fail in combination with R250 at the tricritical point of the Blume-Capel model for some “resonant” lattice sizes.

By analyzing the recursion of R250 analytically, Heuer *et al.*³⁹ succeeded to predict an anomalous triplet correlation of $3/28 \approx 0.107\,142\,857\,1\dots$ at the special value $k = 103$, in perfect agreement with the numerical observation (which they also reconfirmed). Equipped with this finding they were able to propose a modified version of R250 denoted as R250/521, which avoids these triplet correlations and indeed shows a much improved performance. For the two-dimensional Ising model on a 16×16 lattice at β_c they obtained with the R250/521 generator for the energy and specific heat systematic deviations of absolutely tolerable (and in fact, expected) 0.1σ and 1.5σ , respectively, for a set-up of the single-cluster Monte Carlo simulations which was otherwise equivalent to that of Ref. 34. This remarkable improvement indicates that the triplet correlations are very probably responsible for the systematic errors observed by Ferrenberg *et al.*³⁴

Around the same time, Shchur and Blöte⁴⁰ performed a systematic study of this problem by varying the size L of the square lattice as well as the “magic numbers” (p, q) , investigating the four pairs $(p, q) = (36, 11), (89, 38), (127, 64)$, and $(250, 103)$. For all pairs they found at β_c significant deviations from the exact result with a maximum for $L = 7, 12, 15$, and 22 , respectively. Since the average cluster size $\langle C \rangle$ of the single-cluster Wolff update algorithm is an (improved) estimator⁴¹ for the susceptibility of Ising models, it scales with lattice size according to $\langle C \rangle = aL^{\gamma/\nu}$, which in two dimensions specializes to $\langle C \rangle = aL^{7/4}$. The coefficient a is non-universal, and for the square geometry it takes the

generator	periodic b.c.		anti-periodic b.c.	
	e	c	e	c
R250	1.414 087(63) −2.0 σ	1.356 9(16) +19.4 σ	1.214 173(47) −1.6 σ	0.704 52(85) +7.5 σ
R250/521	1.414 206(59) −0.14 σ	1.325 5(16) −0.27 σ	1.214 277(47) +0.6 σ	0.698 17(86) +0.06 σ
RANLUX	1.414 253(60) +0.65 σ	1.326 9(13) +0.75 σ	1.214 215(48) −0.7 σ	0.698 08(82) −0.05 σ
exact/average	1.414 213.6	1.325 927.9	1.214 247(34)	0.698 08(82)

Table 1. Comparison of the energy e and specific heat c for a 10×192 Ising system as obtained in single-cluster Monte Carlo simulations using three different pseudo random number generators. The deviations from the exact (periodic b.c.) respectively average (anti-periodic b.c.) values are given in units of the standard deviation σ , indicated in parentheses behind the mean values.

numerically determined value of $a \approx 1.1$. Using this value in the above finite-size scaling formula one easily derives $\langle C \rangle \approx 33, 85, 126, 141 = 0.55 \times 256$, and 246 for $L = 7, 12, 15, 16$, and 22, respectively. By comparing with the parameter p one thus concludes that the largest deviations from the exact results happen when the average cluster size coincides with p .

Also for asymmetric, strip-like lattices of size 10×192 the failure of the R250 generator in combination with the single-cluster update algorithm was observed.⁴² Here periodic boundary conditions (b.c.) were applied in the long direction and both, periodic as well as anti-periodic b.c., in the short direction. The results shown in Table 1 are based on 2×10^6 measurements at β_c . Here the deviations of the energy and specific heat from the exact results are less pronounced than in Refs. 34 and 40. This is, however, consistent with the observation in Ref. 40, since the average cluster sizes, $\langle C \rangle \approx 159$ and $\langle C \rangle \approx 83$ for periodic and anti-periodic b.c., respectively, are relatively far away from the lag $p = 250$. The modified generator R250/521 as well as RANLUX (in “luxury” level LUX = 4), on the

generator	periodic b.c.		anti-periodic b.c.	
	ξ_e	ξ_σ	ξ_e	ξ_σ
R250	1.5796(46) −0.3 σ	12.7092(92) +3.3 σ	0.7803(84) −3.8 σ	4.2878(43) −0.7 σ
R250/521	1.5851(48) +0.8 σ	12.6787(74) −0.02 σ	0.8110(82) −0.1 σ	4.2925(39) +0.5 σ
RANLUX	1.5770(50) −0.8 σ	12.6789(85) +0.006 σ	0.8126(83) +0.1 σ	4.2887(40) −0.5 σ
exact/average	1.5812(35)	12.678 845	0.8118(58)	4.2906(28)

Table 2. Same comparison as in Table 1 for the correlation lengths ξ_e and ξ_σ of energy and magnetization densities, respectively, in the long direction of a 10×192 Ising lattice.

other hand, performed very well. When using the R250 generator, small but still significant deviations were also observed⁴² for the correlation lengths of the energy and spin densities (measured using the zero-momentum technique⁴³), cf. Table 2.

Another simulational test based on Schwinger-Dyson identities has been proposed by Ballesteros and Martín-Mayor.⁴⁴ Applications to the two- and three-dimensional Ising model confirmed the flaws in two dimensions reported earlier and showed that also in three dimensions the combination of R250 with the single-cluster update algorithm produces incorrect results.

4 Non-Uniform Pseudo Random Numbers

All basic pseudo random number generators discussed above are designed for uniformly distributed pseudo random numbers $x_i \in [0, 1]$. In many applications it is necessary, however, to be able to draw pseudo random numbers from non-uniform distributions.⁴⁵ One strategy is to divide this problem into two parts. First, one of the generators described above is used to generate uniformly distributed random numbers, which in a second step are appropriately transformed to follow the specific distribution at hand.

A standard procedure is the inversion method. For a given normalized probability density $f(x)$ one calculates the associated probability distribution (accumulated density in usual physics terms),

$$F(x) = \int_{x_{\min}}^x dx' f(x') . \quad (19)$$

Due to $F'(x) = f(x) \geq 0$ and the normalization condition, $F(x)$ grows monotonically from 0 to 1, such that the F values are uniformly distributed. Drawing a uniformly distributed pseudo random number R , equating $R = F(x)$ and, if the function inverse is known analytically, setting $x = F^{-1}(R)$ the problem is solved.

For the example of an exponential decay,

$$f(x) = \exp(-x) , \quad x \geq 0 , \quad (20)$$

one derives in this way $R = F(x) = 1 - \exp(-x)$ or

$$x = -\ln(1 - R) , \quad R \in [0, 1] . \quad (21)$$

Notice that since $R \in [0, 1]$, the formula should be programmed in the “complicated” way as shown here; rewriting it as $x = -\ln(R)$ one could occasionally hit $\ln 0$ which would cause a run-time error (with a reaction depending on the operating system used).

Another simple example is the Lorentzian density

$$f(x) = \frac{1}{\pi} \frac{\Gamma}{\Gamma^2 + x^2} , \quad (22)$$

where Γ parameterizes the width of the Lorentzian peak. Here one calculates $R = F(x) = \frac{1}{\pi} \int_{-\infty}^x dx' \frac{\Gamma}{\Gamma^2 + x'^2} = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x}{\Gamma}\right)$, which can be inverted to give

$$x = \Gamma \tan[\pi(R - 1/2)] , \quad R \in [0, 1] . \quad (23)$$

The final and most important example are Gaussian random numbers which follow the probability density (. . . no problem to remember if you saved a German 10 DM note . . .)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) , \quad (24)$$

where the parameter σ^2 is the squared width of the distribution. Here

$$R = F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x dx' \exp\left(-\frac{x'^2}{2\sigma^2}\right) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2\sigma^2}}\right) \right] , \quad (25)$$

with $\operatorname{erf}(\cdot)$ denoting the error function which cannot be inverted analytically. Either one could think of numerical inversion schemes (which indeed is the method of choice for really complicated probability distributions) – or one remembers the polar coordinates trick used to calculate the Gaussian integral and continues analytically: Considering the auxiliary two-dimensional product distribution $f_2(x, y) = f(x)f(y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$ and introducing polar coordinates $x = r \cos(\Theta)$, $y = r \sin(\Theta)$, one obtains

$$f_2(x, y) dx dy = \frac{1}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) r dr \frac{d\Theta}{2\pi} , \quad (26)$$

showing immediately that the angle Θ is uniformly distributed between 0 and 2π . Also for the radial coordinate r the inversion is now straightforward since

$$R_1 = F(r) = \frac{1}{\sigma^2} \int_0^r dr' r' \exp\left(-\frac{r'^2}{2\sigma^2}\right) = 1 - \exp\left(-\frac{r^2}{2\sigma^2}\right) . \quad (27)$$

We thus arrive at the so-called Box-Müller method: Draw two uniformly distributed random numbers R_1 and R_2 , and compute

$$r = \sqrt{-2\sigma^2 \ln(1 - R_1)} , \quad R_1 \in [0, 1) , \quad (28)$$

$$\Theta = 2\pi R_2 , \quad R_2 \in [0, 1) . \quad (29)$$

Then

$$x = r \cos(\Theta) , \quad (30)$$

$$y = r \sin(\Theta) , \quad (31)$$

is a pair of two independent Gaussian distributed pseudo random numbers.

Especially for Gaussian random numbers there is another procedure which directly makes use of the central limit theorem and the fact that averages of arbitrarily distributed random numbers (under certain rather mild conditions) tend asymptotically to a Gaussian distribution. In practice one uses, of course, again uniformly distributed pseudo random numbers x_i generated with one of the algorithms described in the previous section. Recalling the mean value $\bar{x} = 1/2$ and variance $\sigma^2 = \bar{x^2} - \bar{x}^2 = 1/12$ for uniform random numbers, it is straightforward to see that

$$X = \left(\sum_{i=1}^n x_i - \frac{n}{2} \right) \sqrt{\frac{12}{n}} \quad (32)$$

is (approximately) a Gaussian distributed random number around $X = 0$ with unit variance. Of course, since $X_{\max} = -X_{\min} = \sqrt{3n}$, this can be strictly true only asymptotically as $n \rightarrow \infty$. But even the convenient choice $n = 12$ leads already to a reasonable approximation⁴⁶ in the range $|X| < 2 = 2\sigma$ with errors less than 9×10^{-3} .

Another, physically motivated direct method based on simulating N molecules has recently been discussed by Fernández and Criado.⁴⁷

5 Summary

The generation of “good” pseudo random numbers is quite a delicate issue that requires some care and extensive quality tests. It is therefore highly recommended not to invent ones own “secret” recursion rules but to use one of the well-known generators which have been tested and applied by many other workers in the field. If such a well-accepted generator would turn out to be problematic in some specific application, one could at least be sure that the Monte Carlo community as a whole would work hard to track the origin of the problem – as it has happened with the (in)famous R250 generator. Being based on deterministic recursion rules, it is trivial that for any pseudo random number generator one can design a test where it would fail. Thanks to the by now available quite sophisticated mathematical and physically motivated empirical tests one can be very confident, however, that standard generators will yield sufficiently “random” numbers in most applications.

Acknowledgments

I would like to thank Tilman Sauer, Andreas Weber and Martin Weigel for pseudo random, but very useful discussions on randomly selected topics relevant for these lecture notes.

References

1. D. Frenkel and B. Smit, *Understanding Molecular Simulation – From Algorithms to Applications* (Academic Press, San Diego, 1996).
2. D.P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).
3. F. Gutbrod, in: *Annual Review of Computational Physics VI*, ed. D. Stauffer (World Scientific, Singapore, 1999), p. 203.
4. D. Stauffer, in: *Computational Physics: Selected Methods – Simple Exercises – Serious Applications*, eds. K.H. Hoffmann and M. Schreiber (Springer, Berlin, 1996), p. 1.
5. I. Vattulainen, K. Kankaala, J. Saarinen, and T. Ala-Nissila, Comp. Phys. Comm. **86**, 209 (1995).
6. I. Vattulainen, Licentiate in Technology thesis, University of Helsinki (1994) [cond-mat/9411062].
7. K. Kankaala, T. Ala-Nissila, and I. Vattulainen, Phys. Rev. **E48**, R4211 (1993).
8. I. Vattulainen, T. Ala-Nissila, and K. Kankaala, Phys. Rev. Lett. **73**, 2513 (1994).
9. I. Vattulainen, T. Ala-Nissila, and K. Kankaala, Phys. Rev. **E52**, 3205 (1995).
10. D.H. Lehmer, in: *Proc. 2nd Symp. on Large-Scale Digital Calculating Machinery* (Harvard University Press, Cambridge, 1951), p. 141.
11. S.K. Park and K.W. Miller, Comm. ACM **31**, 1192 (1988).
12. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in Fortran 77 – The Art of Scientific Computing*, second corrected edition (Cambridge University Press, Cambridge, 1996), pp. 269-270.

13. *NAG Fortran Library Manual, Mark 14, 7* (Numerical Algorithms Group Inc., 1990).
14. S.L. Anderson, SIAM Review **32**, 221 (1990).
15. G.S. Fishman, Math. Comp. **54**, 331 (1990).
16. A. De Matteis and S. Pagnutti, Parallel Computing **13**, 193 (1990).
17. G. Marsaglia, in: *Applications of Number Theory to Numerical Analysis*, ed. S.K. Zaremba (Academic Press, New York, 1972), p. 249.
18. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, in Ref. 12, p. 273.
19. D.E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, second edition (Addison-Wesley, Reading, Massachusetts, 1981).
20. T.G. Lewis and W.H. Payne, J. Assoc. Comput. Mach. **20**, 456 (1973).
21. R.C. Tausworthe, Math. Comp. **19**, 201 (1965).
22. S. Kirkpatrick and E.P. Stoll, J. Comp. Phys. **40**, 517 (1981).
23. G. Marsaglia and A. Zaman, Stat. & Prob. Lett. **8**, 329 (1990).
24. F. James, *A Review of Pseudorandom Number Generators*, Comp. Phys. Comm. **60**, 329 (1990).
25. G. Marsaglia, B. Narasimham, and A. Zaman, Comp. Phys. Comm. **60**, 345 (1990).
26. G. Marsaglia and A. Zaman, Ann. Appl. Prob. **1**, 462 (1991).
27. Private communication (1993) of F. James to M. Lüscher (Ref. 28).
28. M. Lüscher, Comp. Phys. Comm. **79**, 100 (1994).
29. F. James, Comp. Phys. Comm. **79**, 111 (1994).
30. F. James, Comp. Phys. Comm. **97**, 357 (1996).
31. K.G. Hamilton, Comp. Phys. Comm. **101**, 249 (1997).
32. K.G. Hamilton and F. James, Comp. Phys. Comm. **101**, 241 (1997).
33. P. Blaudeck, in: *Computational Physics: Selected Methods – Simple Exercises – Serious Applications*, eds. K.H. Hoffmann and M. Schreiber (Springer, Berlin, 1996), p. 9.
34. A.M. Ferrenberg, D. P. Landau, and Y.J. Wong, Phys. Rev. Lett. **69**, 3382 (1992).
35. U. Wolff, Phys. Rev. Lett. **62**, 361 (1989); Nucl. Phys. **B322**, 759 (1989).
36. B. Kaufman, Phys. Rev. **76**, 1232 (1949); A.E. Ferdinand and M.E. Fisher, Phys. Rev. **185**, 832 (1969). For a Fortran code, see W. Janke, in: *Computational Physics: Selected Methods – Simple Exercises – Serious Applications*, eds. K.H. Hoffmann and M. Schreiber (Springer, Berlin, 1996), p. 10, and the accompanying diskette.
37. P.D. Coddington, Int. J. Mod. Phys. **C5**, 547 (1994) [cond-mat/9309017].
38. F. Schmid and N.B. Wilding, Int. J. Mod. Phys. **C6**, 781 (1995) [cond-mat/9512135].
39. A. Heuer, B. Dünweg, and A.M. Ferrenberg, Comp. Phys. Comm. **103**, 1 (1997).
40. L.N. Shchur and H.W.J. Blöte, Phys. Rev. **E55**, R4905 (1997) [cond-mat/9703050].
41. U. Wolff, Nucl. Phys. **B334**, 581 (1990).
42. M. Weigel, Diploma thesis, Universität Mainz (1998), unpublished.
43. M. Weigel and W. Janke, Phys. Rev. Lett. **82**, 2318 (1999); Phys. Rev. **B62**, 6343 (2000).
44. H.G. Ballesteros and V. Martín-Mayor, Phys. Rev. **E58**, 6787 (1998).
45. L. Devroye, *Non-Uniform Random Variate Generation* (Springer, Berlin, 1986).
46. M. Abramowitz and I.A. Stegun (eds.), *Handbook of Mathematical Functions*, 9th printing (Dover, New York, 1972), p. 953.
47. J.F. Fernández and C. Criado, preprint cond-mat/9901202.
48. There is by now a huge amount of Ising model material available on the

World Wide Web, including animations. For a list of useful links, see
<http://oscar.cacr.caltech.edu/Hrothgar/Ising/references.html>.

49. S. J. Plimpton, *Fast parallel algorithms for short-range molecular dynamics*, J. Comp. Phys. **117**, 1–19 (1995).
50. G. C. Fox, M. A. Johnson, G. A. Lyzenga, S. W. Otto, J. K. Salmon, and D. W. Walker, *Solving Problems on Concurrent Processors: Volume 1* (Prentice Hall, Englewood Cliffs, NJ, 1988).
51. MD simulation of an FCC lattice with 1.213.857.792 atoms, see
<http://www.itap.physik.uni-stuttgart.de/~joerg/imd.html>
52. MD simulation of a 1 μ s trajectory of a small protein in water: (a) Y. Duan, L. Wang and P. A. Kollman, Proc. Natl. Acad. Sci. USA **9**, 9897 (1998). (b) Y. Duan and P. A. Kollman, Science **282**, 740 (1998).
53. <http://www.cs.sandia.gov/~sjplimp/main.html>
54. <http://www.cs.sandia.gov/~sjplimp/crada.html>

Integrators for Quantum Dynamics: A Numerical Analyst's Brief Review

Christian Lubich

Mathematisches Institut, Universität Tübingen
Auf der Morgenstelle 10, 72076 Tübingen, Germany
E-mail: lubich@na.uni-tuebingen.de
URL: <http://na.uni-tuebingen.de>

This note discusses and compares – in theoretical respects – various old and new approaches to numerical time integration for quantum dynamics: implicit vs. exponential midpoint rule; splitting, Chebyshev and Lanczos approximations to the exponential; Magnus integrators; integrators for almost-adiabatic quantum dynamics.

1 Introduction

This paper gives a concise review of numerical integrators for time-dependent Schrödinger equations

$$i\dot{\psi}(t) = H(t)\psi(t), \quad \psi(0) = \psi_0. \quad (1)$$

Numerical difficulties in the solution of such problems are due both to discretizing or modeling in space (which is not considered here) and to discretization in time, on which the focus is put in the present article. The computational Hamiltonian $H(t)$ is a space discretization or other finite-dimensional model of $H(t) = T + V(t)$ with a kinetic part such as $T = -(2m)^{-1}\Delta_x$, and with a potential $V(t)$ acting as a multiplication operator. Numerical problems are caused by the unbounded nature of the Hamiltonian and the resulting highly oscillatory behaviour of the wave function.

Several new and promising numerical methods have been devised in the last few years, and an improved understanding of well-established methods could be gained. In this review I will put a stronger emphasis on theoretical error bounds than is usual in computational physics articles. This is not done out of mathematical vanity, but because theoretical insight – together with numerical experiments – is important in identifying and comparing merits and flaws of different methods, and in guiding the way to improved methods. It is also useful to question the uncritical use of such universally accepted concepts as “second-order scheme”, which may be misleading for the problem at hand.

We start from the classical implicit midpoint rule and compare it with the exponential midpoint rule. The latter method requires computing the exponential of the Hamiltonian applied to a vector, for which we discuss three computational approaches: Splitting, Chebyshev and Lanczos approximations. As a way to enhance the accuracy of the exponential midpoint rule, Magnus integrators are then discussed. In a final section, we turn to novel integrators that are devised for treating almost-adiabatic quantum dynamics.

2 The Implicit Midpoint Rule

Contrary to most of the classical numerical integrators, such as explicit or implicit Runge-Kutta or multistep methods, the *implicit midpoint rule*

$$i \frac{\psi_{n+1} - \psi_n}{\Delta t} = H(t_{n+1/2}) \frac{1}{2}(\psi_{n+1} + \psi_n) \quad (2)$$

(with $t_{n+1/2} = \frac{1}{2}(t_{n+1} + t_n)$ and $t_n = n\Delta t$) has a *unitary propagator*:

$$\psi_{n+1} = r(-i\Delta t H(t_{n+1/2})) \psi_n \quad \text{with} \quad r(z) = \frac{1 + \frac{1}{2}z}{1 - \frac{1}{2}z}. \quad (3)$$

This is an important qualitative feature which the numerical method shares with the exact solution operator. It implies that the method preserves the L^2 norm (or the Euclidean norm in the spatially discretized situation):

$$\|\psi_{n+1}\| = \|\psi_n\|,$$

and hence is stable for arbitrary time steps Δt .

A further useful property is the *time reversibility* of the numerical scheme: exchanging $n \leftrightarrow n + 1$ and $\Delta t \leftrightarrow -\Delta t$ gives the same numerical method again. In terms of the propagator function $r(z)$, this is reflected by the property

$$r(-z) = r(z)^{-1},$$

which $r(z)$ shares with the exponential e^z .

But what can be said about the accuracy of the implicit midpoint rule? In the classical ODE setting of *bounded* and smooth $H(t)$, it is a well-known fact that the implicit midpoint rule is a *second-order* method, that is, the error satisfies

$$\|\psi_n - \psi(t_n)\| = \mathcal{O}(\Delta t^2) \quad (4)$$

uniformly for $n\Delta t$ in a bounded time interval $[t_0, t_{\text{end}}]$. Such statements on the order of a method should, however, be taken with caution in the present context: in our situation of a (spatially discretized) Schrödinger equation (1), the norm of $H(t)$ can be arbitrarily large or even unbounded, and hence the classical numerical ODE theory does not apply. Nevertheless, for the particular case of the implicit midpoint rule it can be shown that the constant hidden in the $\mathcal{O}(\Delta t^2)$ error bound is in fact independent of bounds of $H(t)$. It does depend, however, on bounds of \dot{H} and \ddot{H} , and on the maximum of the norm of the third time derivative of the solution ψ on the time interval under consideration. This latter dependence on solution derivatives is an unpleasant feature: unless we start from spatially very smooth initial data, the wave function $\psi(t)$ is highly oscillatory in time, and hence higher time derivatives can become large. Good accuracy can then be expected only for very small time steps, and this is indeed what happens here. We emphasize that accuracy, not stability, restricts the time step of the implicit midpoint rule applied to Schrödinger equations. High temporal smoothness, when available, leads to good accuracy also with larger time steps.

3 The Exponential Midpoint Rule

This method is obtained by formally replacing $r(z)$ by $\exp(z)$ in the formula (3) of the implicit midpoint rule:

$$\psi_{n+1} = \exp(-i\Delta t H(t_{n+1/2})) \psi_n . \quad (5)$$

Of course, instead of solving systems of linear equations, we now have to scope about how to compute the exponential of a large matrix. We will consider this aspect in the following sections and assume for the time being that the matrix exponential times a vector can be computed efficiently. We note that the above exponential midpoint rule again has a unitary propagator and it is time-reversible. It offers improved accuracy over the classical implicit midpoint rule: it satisfies a second-order error bound (4), but contrary to before, the constant hidden in the \mathcal{O} -notation is now independent of the time derivatives of the wave function $\psi(t)$.^{5,3} The result³ is

$$\|\psi_n - \psi(t_n)\| \leq C \Delta t^2 \max_{t \in [t_0, t_{\text{end}}]} \|D\psi(t)\|$$

under the assumption on the commutators ($[A, B] = AB - BA$)

$$\|[H(t), H(s)]\phi\| \leq c \|D\phi\| \quad \text{for all } t, s \text{ and } \phi.$$

Since the commutator of the Laplacian with a multiplication operator is a *first-order* differential operator, this condition holds with the gradient operator $D = \nabla_x$ in the spatially continuous case of $H(t) = (2m)^{-1}\Delta_x + V(t)$ with a smooth bounded potential $V(t)$, and with a discrete gradient in cases of spatially discretized problems.⁸

This theoretical fact explains – and numerical experiments confirm – that much larger time steps than with the classical implicit midpoint rule can be taken to achieve the same accuracy, in particular in cases of low regularity of the wave function.

4 Strang Splitting

A standard approach to computing the exponential of $H = T + V$ is to use the symmetric splitting (known as Strang splitting or Marchuk splitting or symmetric Trotter splitting in different communities)

$$\exp(-i\Delta t(T + V))\psi \approx \exp(-i\Delta t T/2) \exp(-i\Delta t V) \exp(-i\Delta t T/2)\psi . \quad (6)$$

The right-hand side is often much cheaper to compute. For instance, this is the case when T is a spectral discretization of the negative Laplacian $-(2m)^{-1}\Delta_x$, which is diagonalized by fast Fourier transforms, and V is represented by a diagonal matrix. Only the exponentials of diagonal matrices, which are trivially computed, are required in this situation.

The symmetric splitting is a *second-order* scheme:

$$\|\exp(-i\Delta t T/2) \exp(-i\Delta t V) \exp(-i\Delta t T/2)\psi - \exp(-i\Delta t(T + V))\psi\| = \mathcal{O}(\Delta t^2) .$$

Here again, such an order statement must be taken with caution. This error bound is easily obtained by using the series expansion of the exponential, but then the \mathcal{O} -term depends on the norms of T and V . Since T is typically a discretized Laplacian, such an estimate is of no use here. A second-order error bound that allows for an unbounded T has been derived

only recently.⁸ Under reasonable conditions on the commutators $[T, V]$ and $[T, [T, V]]$ and assuming bounds on V , it is shown that such an estimate holds with $\|T\psi\|$ appearing in the constant of the $\mathcal{O}(\Delta t^2)$ estimate. The spatial regularity of ψ thus enters the error. If only an energy bound $\psi^* H \psi \leq B$ is available, then the order of convergence of the splitting scheme may decrease to one. Numerical experiments confirm this theoretically predicted order reduction.

5 Chebyshev Approximation

When a computationally efficient splitting is not available, or when there is little spatial regularity in the problem, an alternative is to compute the exponential of H as a whole, using polynomial approximations to the exponential. In the Chebyshev approach, this requires bounds for the extreme eigenvalues E_{\min} and E_{\max} of H (which is here assumed to be given as a spatial discretization). One then uses a truncated Chebyshev expansion of $\exp(-ix)$ on the interval $[\Delta t E_{\min}, \Delta t E_{\max}]$:

$$\exp(-ix) \approx \sum_{n=0}^m c_n P_n(x), \quad \text{where } P_n(x) = T_n \left(\frac{2x - \Delta t E_{\max} - \Delta t E_{\min}}{\Delta t E_{\max} - \Delta t E_{\min}} \right)$$

with the usual Chebyshev polynomials $T_n(\xi)$ for the interval $[-1, 1]$. Then, one uses the approximation

$$\exp(-i\Delta t H)\psi \approx \sum_{n=0}^m c_n P_n(\Delta t H)\psi$$

which is computed efficiently using Clenshaw's algorithm. This requires m multiplications $H\phi$ of the Hamiltonian H with a vector. Concerning the quality of the approximation, there is nearly no error reduction for $m \leq \frac{1}{2}\Delta t(E_{\max} - E_{\min})$, but very rapid, superlinear error decay for m growing beyond that bound.¹⁵ The error is not influenced by regularity properties of ψ , as opposed to the situation of the Strang splitting. However, refining the space discretization increases E_{\max} and thus requires a higher degree m or a reduction of the time step Δt .

6 Lanczos Approximation

A different, and according to our numerical experience often more efficient approach to polynomial approximation of the product of the matrix exponential times a vector, is by using the *Lanczos process*.^{10,13} This approach to computing the exponential was proposed in the context of quantum dynamics,¹² and its convergence properties have meanwhile been analyzed.⁴ The symmetric Lanczos process generates recursively an orthonormal basis $V_m = [v_1 \cdots v_m]$ of the m th Krylov subspace $K_m(H, \psi) = \text{span}(\psi, H\psi, \dots, H^{m-1}\psi)$ such that

$$HV_m = V_m L_m + [0 \cdots 0 \beta_m v_{m+1}].$$

This requires m multiplications of H with a vector, where m is chosen much smaller than the dimension of the problem. The symmetric tridiagonal $m \times m$ matrix $L_m = V_m^T H V_m$ is

the orthogonal projection of H onto $K_m(H, \psi)$. This is used in the approximation^{2,4,12,14}

$$\exp(-i\Delta t H)\psi \approx V_m \exp(-i\Delta t L_m) V_m^T \psi$$

with $V_m^T \psi = [10 \cdots 0]^T$. The matrix exponential $\exp(-i\Delta t L_m)$ is computed cheaply from the eigendecomposition $L_m = Q_m D_m Q_m^T$, with diagonal D_m , via

$$\exp(-i\Delta t L_m) = Q_m \exp(-i\Delta t D_m) Q_m^T.$$

The Lanczos process is stopped if

$$\beta_m \left| \left(\exp(-i\Delta t L_m) \right)_{m,m} \right| < \text{tol},$$

where $(\cdot)_{m,m}$ denotes the (m, m) entry of the matrix, and tol is a user-specified tolerance. This stopping criterion is motivated⁶ by a generalization of a residual bound which is the most popular stopping criterion for solving linear systems. The convergence behaviour as a function of m is similar to Chebyshev approximation of the same degree in cases where the eigenvalues of H are densely distributed in the interval $[E_{\min}, E_{\max}]$, but convergence can be much more rapid when there are eigenvalue gaps within this interval.⁴ Moreover, this approach takes advantage of preferred eigendirections in the vector ψ . It does not require a priori estimates of the extreme eigenvalues. On the other hand, the Lanczos process needs the computation of scalar products of vectors that are not required in the Chebyshev approach.

7 Magnus Integrators

In the Magnus approach,¹¹ the solution of (1) is represented as

$$\psi(t_n + \Delta t) = \exp(\Omega_n)\psi(t_n), \quad (7)$$

where Ω_n is given as a series composed of integrals of commutators of $A(t) = -iH(t_n + t)$, the *Magnus series*

$$\begin{aligned} \Omega_n = & \int_0^{\Delta t} A(\tau) d\tau - \frac{1}{2} \int_0^{\Delta t} \left[\int_0^\tau A(\sigma) d\sigma, A(\tau) \right] d\tau \\ & + \frac{1}{4} \int_0^{\Delta t} \left[\int_0^\tau \left[\int_0^\sigma A(\mu) d\mu, A(\sigma) \right] d\sigma, A(\tau) \right] d\tau \\ & + \frac{1}{12} \int_0^{\Delta t} \left[\int_0^\tau A(\sigma) d\sigma, \left[\int_0^\tau A(\mu) d\mu, A(\tau) \right] \right] d\tau + \dots \end{aligned} \quad (8)$$

For smooth bounded matrices $A(t)$ the remainder in (8) is of size $\mathcal{O}(\Delta t^5)$, and hence the truncated series inserted into (7) gives a higher-order approximation to the solution value $\psi(t_n + \Delta t)$ for small Δt . A simpler expression that agrees with the truncated series up to terms of size $\mathcal{O}(\Delta t^5)$, is given in terms of the univariate integrals¹

$$B_k = \frac{1}{\Delta t^{k+1}} \int_{-\Delta t/2}^{\Delta t/2} t^k A(\frac{1}{2}\Delta t + t) dt$$

as $\Omega_n = \Delta t B_0 - \Delta t^2 [B_0, B_1] + \mathcal{O}(\Delta t^5)$. The integrals B_k can be replaced by suitable quadrature, e.g., by the fourth-order Gauss or Simpson rule. Using this approximation in (7) gives a time-reversible, unitary method of order 4:

$$\psi_{n+1} = \exp(\widehat{\Omega}_n)\psi_n \quad \text{with} \quad \widehat{\Omega}_n = \Delta t B_0 - \Delta t^2 [B_0, B_1]. \quad (9)$$

In Blanes et al.,¹ methods of order 6 requiring 4 commutators, and methods of order 8 requiring 10 commutators are also constructed. See Iserles et al.⁷ for a detailed review of Magnus integrators.

As with the previously considered methods, the order statements must be taken with caution in the case of unbounded operators $H(t)$. It turns out³ that the above method retains fourth order independently of the norm of $H(t)$ in the situation of $H(t) = T + V(t)$ with T a discretization of the negative Laplacian (with maximum eigenvalue $E_{\max} \sim \Delta x^{-2}$) and with a smooth potential $V(t)$, under a rather mild time step restriction

$$\Delta t \sqrt{E_{\max}} \leq \text{Const.}$$

We remark that this holds in spite of the fact that here the Magnus expansion is generally a divergent series. Convergence of the Magnus series would require a more stringent time step restriction $\Delta t E_{\max} \leq c (\approx 1)$.

8 Integrators for Almost-Adiabatic Quantum Dynamics

A different situation from that considered so far occurs in the treatment of problems of the type

$$i\varepsilon \dot{\psi} = H(t)\psi \quad (0 < \varepsilon \ll 1) \quad (10)$$

with a small parameter ε (which, in self-consistent field approaches, would correspond to the square root of the mass ratio of light and heavy particles, such as electrons and ions). Here it is assumed that $H(t)$ varies slowly compared to the fast time scale ε . All of the previously considered integrators require time steps $\Delta t \ll \varepsilon$. For $\varepsilon \rightarrow 0$ and $\Delta t > \varepsilon$, they do not approximate the adiabatic limit as given by the quantum-adiabatic theorem.

Numerical integrators that give good approximations to (10) with relatively large time steps $\Delta t > \varepsilon$, have recently been derived.⁹ These integrators are devised for situations where $H(t)$ is expensive to evaluate, but the substantially occupied eigenstates and eigenvalues of $H(t)$ can be obtained at comparatively small additional computational cost. This situation occurs in particular in reduced, relatively low-dimensional models, which are often appropriate for the description of near-adiabatic behaviour. Let $H(t)$ be diagonalized as

$$H(t) = Q(t)\Lambda(t)Q(t)^T, \quad \Lambda(t) = \text{diag}(\lambda_k(t))$$

with an orthogonal matrix $Q(t)$. (This can be extended to the situation where only a few of the lower eigenstates are computed.⁹) The numerical integrator is not applied directly to (10), but to an equivalent equation for the variable $\eta(t)$ defined by

$$Q(t)^T \psi(t) = \exp\left(-\frac{i}{\varepsilon}\Phi(t)\right) \eta(t) \quad \text{with} \quad \Phi(t) = \int_0^t \Lambda(\tau) d\tau.$$

Up to a rapidly rotating phase, η is the coefficient vector with respect to the eigenbasis representation of ψ . Then η solves the differential equation

$$\dot{\eta}(t) = \exp\left(\frac{i}{\epsilon}\Phi(t)\right) W(t) \exp\left(-\frac{i}{\epsilon}\Phi(t)\right) \eta(t) \quad (11)$$

with the skew-symmetric matrix $W = \dot{Q}^T Q$. The right-hand side of (11) is bounded (though highly oscillatory), and hence η is smoother than ψ . As long as the eigenvalues of $H(t)$ remain well-separated, $\eta(t)$ stays $\mathcal{O}(\epsilon)$ close to the initial value $\eta(0)$.

The simplest method is based on freezing the slow variables η , Λ and W over a time step and integrating analytically over the highly oscillatory exponentials. This gives the method⁹

$$\eta_{n+1} = \eta_{n-1} + 2h(S(t_n) \bullet E(\Phi_n) \bullet W_n) \eta_n, \quad (12)$$

where the bullets \bullet denote the entrywise product of matrices, $S(t)$ is the matrix with entries $\sin x_{kl}/x_{kl}$ with $x_{kl} = \Delta t(\lambda_k(t) - \lambda_l(t))/\epsilon$, and $E(\Phi)$ is the matrix with entries $\exp(\frac{i}{\epsilon}(\phi_k - \phi_l))$. W_n is a finite difference approximation to $W(t_n)$: $W_n = (2\Delta t)^{-1}(Q(t_{n+1}) - Q(t_{n-1}))^T Q(t_n)$, and Φ_n is the Simpson rule approximation to the integral $\Phi(t_n)$.

This method forms the basis for more accurate schemes also derived in the article⁹. That paper also gives an extension to adaptive time steps to treat avoided crossings of eigenvalues, where non-adiabatic behaviour with sudden energy redistributions occurs.

Acknowledgment

Much of the work in this area that I have been involved in was done together with Marlis Hochbruck or Tobias Jahnke.

References

1. S. Blanes, F. Casas, J. Ros, *Improved high order integrators based on the Magnus expansion*, BIT 40 (2000), 434–450.
2. V. L. Druskin, L. A. Knizhnerman, *Krylov subspace approximations of eigenpairs and matrix functions in exact and computer arithmetic*, Numer. Lin. Alg. Appl., 2 (1995), pp. 205–217.
3. M. Hochbruck, K. Lorenz, Ch. Lubich, *On Magnus integrators for time-dependent Schrödinger equations*, Report (in preparation), 2002.
4. M. Hochbruck, Ch. Lubich, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal. 34 (1997), 1911–1925.
5. M. Hochbruck, Ch. Lubich, *Exponential integrators for quantum-classical molecular dynamics*, BIT 39 (1999), 620–645.
6. M. Hochbruck, Ch. Lubich, H. Selhofer, *Exponential integrators for large systems of differential equations*, SIAM J. Sci. Comput. 19 (1998), 1552–1574.
7. A. Iserles, H.Z. Munthe-Kaas, S.P. Nørsett, A. Zanna, *Lie-group methods*, Acta Numerica 2000, 215–365.

8. T. Jahnke, Ch. Lubich, *Error bounds for exponential operator splittings*. BIT 40 (2000), 735–744.
9. T. Jahnke, Ch. Lubich, *Numerical integrators for quantum evolution close to the adiabatic limit*. Report, 2001.
10. C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bureau Standards 45 (1950), 255–281.
11. W. Magnus, *On the exponential solution of differential equations for a linear operator*, Comm. Pure Appl. Math. VII (1954), 649–673.
12. T. J. Park, J. C. Light, *Unitary quantum time evolution by iterative Lanczos reduction*, J. Chem. Phys. 85 (1986), 5870–5876.
13. B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
14. Y. Saad, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal. 19 (1992), 209–228.
15. H. Tal-Ezer, *Polynomial approximation of functions of matrices and applications*. J. Sci. Comput. 4 (1989), 25–60.

Long-Range Interactions in Many-Particle Simulation

Paul Gibbon and Godehard Sutmann

John von Neumann Institute for Computing
Central Institute for Applied Mathematics
Research Centre Jülich, 52425 Jülich, Germany
E-mail: {*p.gibbon, g.sutmann*}@fz-juelich.de

Numerical algorithms for accelerating the computation of N -body problems dominated by long-range inter-particle forces are reviewed. For periodic systems, the optimised Ewald lattice sum with an $O(N^{3/2})$ scaling forms a reference standard against which the newer, potentially faster Particle-Particle Particle-Mesh and Fast Multipole Methods can be measured. The more general N -body problem with arbitrary boundary conditions is also described, in which various multipole methods are now routinely used instead of direct summation for particle numbers in excess of 10^4 . These techniques are described in a tutorial fashion and rough comparisons given of their respective computational performance on scalar and parallel machine architectures.

1 Introduction

Until relatively recently, the general solution of the N-body problem – computing the trajectories of many mutually interacting particles – was considered intractable, except for small systems, or for particle assemblies in which the interaction potential is either physically or artificially truncated. Over the past half century, however, our definition of ‘small’ has stretched from a few dozen to several thousand bodies, thanks to advances in computing power. On the other hand, this increase in manageable system size is *not* as dramatic as one might expect from Moore’s ‘Law’, in which processing power has doubled every 18 months or so since the 1960s.

A brief inspection of the typical N-body force law reveals why this is so. Consider a classical system of N bodies with charges q_i and masses m_i interacting via a central potential V :

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = -q_i \nabla_i V \quad i = 1, 2, 3 \dots N, \quad (1)$$

where

$$V(\mathbf{r}_i) = \sum_{j \neq i}^N \frac{q_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (2)$$

To compute one iteration of the ensemble trajectory $\mathbf{r}_i(t)$ described by the equation of motion (1), we require $N(N - 1)$ operations. With the aid of Newton’s third law (action=reaction), we can exploit the symmetry of the potential to reduce the operation count by one half, but this still leaves us with an asymptotic scaling of $O(N^2)$ for large N . In other words, we need a 100-fold increase in computing power in order to increase the simulation size by an order of magnitude. This dispiriting fact of N -body life held up large-scale simulation of many-particle systems until the early 1980s, when a number of *algorithmic* advances reduced the computational effort to complexities ranging from $O(N^{3/2})$ down to $O(N)$, depending on the context of the problem.

In this article we present a tutorial survey of these techniques, which can be broadly classified into three categories: (i) Ewald summation, (ii) particle-mesh methods, and (iii) hierarchical or multipole methods. The Ewald method¹ is restricted to fully or partially periodic systems, but has been widely adopted for studies of condensed matter – ionic salts, molecules in solvent etc. – where it is important to eliminate surface effects which would arise in a small, isolated system. Particle-mesh codes^{2,3} are actually more widespread outside the MD community, especially in astrophysics, plasma physics and electrical engineering, but form a vital component of the so-called Particle-Mesh-Ewald (PME) method developed some 10 years ago by Darden.⁴ Multipole methods,⁵ which come in two main flavours – ‘Fast Multipole Methods’ and ‘Tree-Codes’ respectively – are based on the observation that distant charges (or masses, in the case of gravity) may be grouped together and substituted by a single multipole expansion, leading to a substantial saving in the number of interactions necessary to sum the potential or force.

All of these techniques for accelerating N -body force summation have recently been subjected to intense re-examination in order to produce codes suitable for parallel computer architectures. In the final section, some of the important design considerations for N -body simulation on parallel machines will be discussed, and an attempt is made to compare the relative performance of the most commonly used methods.

2 Ewald Summation

The technique of Ewald summation is hugely popular in contemporary molecular dynamics simulation, even though it applies to a special case: namely, periodic systems. By this we mean that the simulation region or ‘cell’ is effectively replicated in all spatial directions, so that particles leaving the cell reappear at the opposite boundary. For systems governed by a short-ranged potential – say Lennard-Jones or hard spheres – it is sufficient to take just the neighbouring simulation volumes into account, leading to the ‘minimum-image’ configuration shown in Fig. 1. The potential seen by the particle at r_i is summed over all other particles r_j , or their periodic images $(r_j \pm n)$, where $n = (i_x, i_y, i_z)L$, with $i_\alpha = 0, \pm 1, \pm 2 \dots \pm \infty$, whichever is closest. More typically, this list is further restricted to particles lying within a sphere centred on r_i .⁶ For long-range potentials, this arrangement is inadequate because the contributions from more distant images at $2L, 3L$ etc., are no longer negligible. One might argue that these contributions should more-or-less cancel, which they nearly do, but one has to take care in which order to do the sum: a simple example serves to illustrate the problem. Consider a system of two oppositely charged ions, periodically extended to form an infinite one-dimensional line of charges, each separated by a distance R – Fig. 2. The potential energy of the reference ion with charge $-q$ is:

$$\begin{aligned} U &= -2q^2 \left(\frac{1}{R} - \frac{1}{2R} + \frac{1}{3R} - \frac{1}{4R} \dots \right) \\ &= -\frac{2q^2}{R} \left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \dots \right) \\ &= -\frac{2q^2}{R} \log 2 \end{aligned} \quad (3)$$

The factor $2 \log 2$ is the Madelung constant, which is of central importance in the theory of ionic crystals.^{7,8} The series in (3) is actually *conditionally convergent*; the result depends

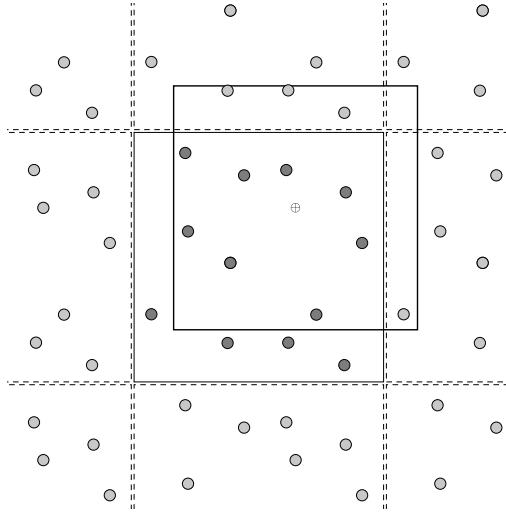


Figure 1. Periodic boundary conditions for simulation region (centre, dark-shaded particles at positions r_j), showing ‘minimum-image’ box for reference ion \oplus at position r_i containing nearest periodic images (light-shaded particles at positions $r_j \pm n$).

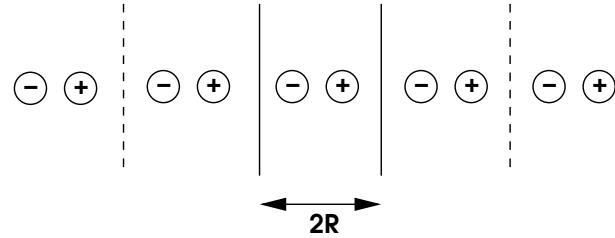


Figure 2. Infinite 1D lattice

on the summation order. To illustrate this, we can choose a different ordering, for example:

$$\begin{aligned}
 & 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \dots \\
 = & 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \frac{1}{9} + \dots \\
 & + \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \dots \\
 = & \sum_n \frac{(-1)^{n-1}}{n} + \frac{1}{2} \sum_n \frac{(-1)^{n-1}}{n} = \frac{3}{2} \log 2,
 \end{aligned}$$

giving us 50% more potential energy than we had before!

In three dimensions, determination of the Madelung constant – and hence the lattice potential energy – is non-trivial because successive terms in the series must be arranged so that positive and negative contributions nearly cancel. This is exactly the problem we are faced with when evaluating the potential on our reference ion in Fig. 1, albeit for an irregular assortment of charges: in what order should we sum over image boxes?

An intuitive and elegant way of doing this is to build up sets of images contained within successively larger spheres surrounding the simulation region²⁵ – Fig. 3. According to this

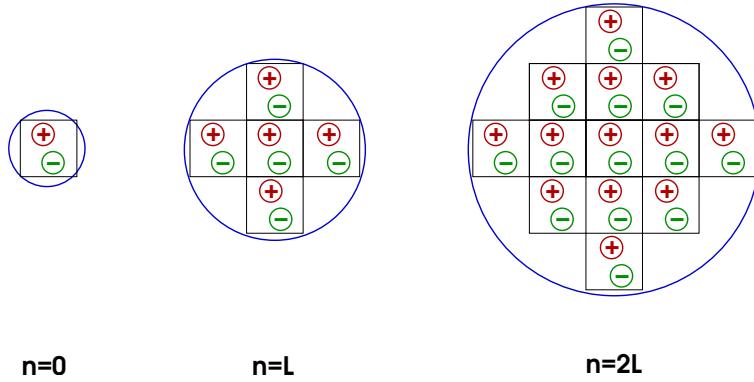


Figure 3. Constructing a convergent sum over periodic images. (Adapted from Allen & Tildesley)

scheme, the potential is expressed mathematically as:

$$V_s(r_i) = \sum_{\mathbf{n}}' \sum_{j=1}^N \frac{q_j}{|\mathbf{r}_{ij} + \mathbf{n}|}, \quad (4)$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{n} = (i_x, i_y, i_z)L$, with $i_\alpha = 0, \pm 1, \pm 2 \dots \pm \infty$ as before. The prime in the sum over \mathbf{n} indicates that the $j = i$ term is omitted for the primary cell $\mathbf{n} = 0$. Taking the image cells in the order prescribed by Fig. 3 does ensure that the sum (4) converges to the correct value, but only slowly. Strictly speaking, (4) is a conditionally convergent series; by contrast, potentials with a radial fall-off steeper than $\sim r^{-3}$ are *absolutely* convergent.⁹

2.1 Periodic Lattice Sum

As it stands, the summation over image boxes implied by (4) makes the prospects of speeding up our N -body problem look rather grim: we have turned an $O(N^2)$ problem into one requiring $N_{box} \times N^2$ operations! Ewald got around this by recasting the potential as the sum of two rapidly converging series: one in real space; the other in reciprocal, or k -space:

$$V_E(r_i) = \sum_{\mathbf{n}}' \sum_{j=1}^N q_j \frac{\operatorname{erfc}(\alpha |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} + \frac{4\pi}{L^3} \sum_{\mathbf{k} \neq 0} \sum_j q_j \exp\left(\frac{-|k|^2}{4\alpha^2}\right) \exp\{i\mathbf{k} \cdot (\mathbf{r}_j - \mathbf{r}_i)\} - \frac{2\alpha}{\pi^{1/2}} q_i \quad (5)$$

The term α is an arbitrary, but important parameter, which governs the relative convergence rates of the two main series. The last term is a ‘self-potential’ which cancels an equivalent contribution in the k -space sum. It is not immediately obvious why the double series (5) should be equivalent to (4). However, we can begin to make physical sense of it by noting that

$$\operatorname{erfc}(x) = 1 - \frac{2}{\pi^{1/2}} \int_0^x e^{-t^2} dt.$$

Thus, what we actually have is an expression of the form:

$$V_E(r_i) = V_s(r_i) - \sum_{\mathbf{n}} f(\mathbf{n}) + \sum_{\mathbf{k}} g(\mathbf{k}). \quad (6)$$

In other words, to get (5) from (4), we just use the trick of adding and subtracting an *additional* series, summed in real space and k -space respectively. In the Ewald sum, this new series is in fact the lattice sum for a Gaussian charge distribution

$$\rho(r) = A \exp(-\alpha^2 r^2). \quad (7)$$

The first two terms in (6) combine to give the rapidly converging real-space sum in (5) – as illustrated schematically in Fig. 4.

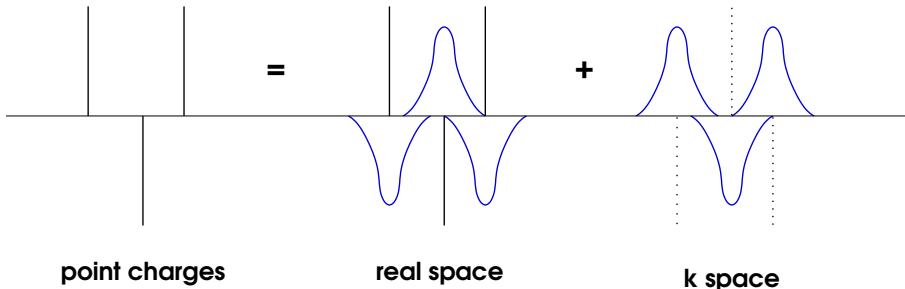


Figure 4. Splitting the sum for point charges into two rapidly convergent series for Gaussian-shaped charges.

The choice of kernel for the charge-smearing function is actually not too critical, and mainly influences the convergence characteristics of the final series. A comparison of several alternative functions was made some time ago by Heyes;¹⁰ for tutorial purposes here, however, we will stick to the simple Gaussian distribution originally used by Ewald himself:

$$\sigma(r) = \frac{\alpha^3}{\pi^{3/2}} \exp(-\alpha^2 r^2), \quad (8)$$

which is normalised such that

$$\int_0^\infty \sigma(r) dr = 1.$$

Note that α determines the width and height of the spreading function, and hence the effective size of the charges.

Let us begin with the real-space sum depicted in Fig. 4. To obtain this, we just subtract the lattice sum for the smeared-out charges from the original point-charge sum, thus:

$$\begin{aligned} V_r(r_i) &= \sum_{\mathbf{n}}' \sum_{j=1}^N \frac{q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \left[1 - \int_0^\infty \sigma(r - r_{ij}) d^3 r \right] \\ &= \sum_{\mathbf{n}}' \sum_j q_j \left[\frac{1}{|\mathbf{r}_{ij} + \mathbf{n}|} - \frac{4\alpha^3}{\pi^{1/2} |\mathbf{r}_{ij} + \mathbf{n}|} \int_0^{|\mathbf{r}_{ij} + \mathbf{n}|} r^2 \exp(-\alpha^2 r^2) dr \right. \\ &\quad \left. - \frac{4\alpha^3}{\pi^{1/2}} \int_{|\mathbf{r}_{ij} + \mathbf{n}|}^\infty r \exp(-\alpha^2 r^2) dr \right] \end{aligned}$$

The second term in the above expression can be integrated by parts to give an error function $\text{erf}(\alpha |\mathbf{r}_{ij} + \mathbf{n}|)$, plus a term which exactly cancels the third term. Carrying out this simplification, we finally get:

$$V_r(r_i) = \sum_{\mathbf{n}}' \sum_{j=1}^N q_j \frac{\text{erfc}(\alpha |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|}. \quad (9)$$

The reciprocal-space sum takes a little more work. First, we consider the charge density of the whole lattice at some arbitrary position r :

$$\rho(r) = \sum_j q_j \delta(r - r_j). \quad (10)$$

Since the lattice is periodic, we can express this equivalently as a Fourier sum:

$$\rho(r) = L^{-3} \sum_j \sum_{\mathbf{k}} f(\mathbf{k}) \exp(-i\mathbf{k} \cdot \mathbf{r}), \quad (11)$$

where $\mathbf{k} = 2\pi/L(i_{\hat{x}}, i_{\hat{y}}, i_{\hat{z}})$; $i_\alpha = 0, 1, 2, \dots$ etc., and

$$f(\mathbf{k}) = \int_{L^3} \rho(r) \exp(i\mathbf{k} \cdot \mathbf{r}) d^3 r, \quad (12)$$

where the integration is now restricted to the unit cell volume $V = L^3$. Substituting (10) into (12) and making use of a standard identity picks out modes corresponding to the point charges:

$$f(\mathbf{k}) = \sum_j q_j \exp(i\mathbf{k} \cdot \mathbf{r}_j) \quad (13)$$

Turning now to the *smeared* charge distribution:

$$\begin{aligned} \rho'(r) &= \sum_j q_j \sigma(r - r_j) \\ &= \int_V \rho(r - r') \sigma(r') d^3 r', \end{aligned} \quad (14)$$

we observe that this is just the convolution of $\rho(r)$ with $\sigma(r)$, which we know can be expressed in Fourier space as:¹¹

$$\rho'(\mathbf{r}) = \frac{1}{L^3} \sum_{\mathbf{k}} 'f(\mathbf{k})\phi(\mathbf{k}, \alpha) \exp(-i\mathbf{k} \cdot \mathbf{r}), \quad (15)$$

where $\phi(\mathbf{k}, \alpha)$ is the Fourier transform of the charge-smearing function $\sigma(r)$, i.e.:

$$\phi(\mathbf{k}, \alpha) = \exp\left(\frac{-|\mathbf{k}|^2}{4\alpha^2}\right). \quad (16)$$

We are now equipped to express the potential due to the smeared charges in k -space. At the reference position r_i , this is:

$$\begin{aligned} V_k(r_i) &= \int_0^\infty \frac{\rho'(r_i + r)}{r} d^3r \\ &= \frac{1}{L^3} \sum_{\mathbf{k}} 'f(\mathbf{k})\phi(\mathbf{k}, \alpha) \exp(-i\mathbf{k} \cdot \mathbf{r}_i) \int_0^\infty \frac{\exp(-i\mathbf{k} \cdot \mathbf{r})}{r} d^3r. \end{aligned}$$

The integral on the right of the above expression is a standard one,¹² and evaluates to $4\pi/k^2$. Combining this with the earlier results (13) and (16) for $f(\mathbf{k})$ and $\phi(\mathbf{k}, \alpha)$ respectively, we have finally:

$$V_k(r_i) = \frac{4\pi}{L^3} \sum_{\mathbf{k}} ' \sum_j q_j \exp\{i\mathbf{k} \cdot (\mathbf{r}_j - \mathbf{r}_i)\} \frac{\exp\left(\frac{-|\mathbf{k}|^2}{4\alpha^2}\right)}{|\mathbf{k}|^2}. \quad (17)$$

This potential includes an unphysical ‘self-term’ corresponding to a smeared out charge centered at r_i , which needs to be subtracted off:

$$\begin{aligned} V_s(r_i) &= q_i \int_0^\infty \sigma(r) d^3r \\ &= \frac{4\pi q_i \alpha^3}{\pi^{3/2}} i \int_0^\infty r \exp(-\alpha^2 r^2) \\ &= \frac{2\alpha}{\pi^{1/2}} q_i. \end{aligned} \quad (18)$$

Adding together our partial results (9), (17) and (18), we recover the Ewald sum quoted before in Equation 5. The equivalent expression for the force (or more correctly the electric field) can be found by direct differentiation with respect to the vector between the reference particle i and particle j :

$$\begin{aligned} \mathbf{E}(\mathbf{r}_i) &= -\frac{\partial V_E(\mathbf{r}_i)}{\partial \mathbf{r}_{ij}} \\ &= \sum_{\mathbf{n}} ' \sum_j \frac{q_j \mathbf{r}_{ij} \cdot \mathbf{n}}{r_{ij}^3 \cdot \mathbf{n}} \left[\text{erfc}(\alpha r_{ij} \cdot \mathbf{n}) + \frac{2\alpha r_{ij} \cdot \mathbf{n}}{\sqrt{\pi}} \exp(-\alpha^2 r_{ij}^2 \cdot \mathbf{n}) \right] \\ &\quad + \frac{4\pi}{L^3} \sum_{\mathbf{k} \neq 0} \sum_j q_j \frac{\mathbf{k}}{k^2} \exp\left(\frac{-k^2}{4\alpha^2}\right) \sin(\mathbf{k} \cdot \mathbf{r}_{ji}). \end{aligned} \quad (19)$$

In the above expression, we have used the shorthand notation $\mathbf{r}_{ij,n} \equiv \mathbf{r}_{ij} + \mathbf{n}$ and $\mathbf{r}_{ji} \equiv \mathbf{r}_j - \mathbf{r}_i$. More sophisticated derivations of lattice sums can be found in Leeuw *et al.*,⁹ who consider more general systems surrounded by a dielectric medium, by Heyes,¹⁰ who considers an arbitrary charge-spreading function, and by Perram *et al.*,¹³ who derive expressions for other types of potential (force-laws). A detailed analysis of the cutoff errors incurred by real-space and k -space sums has been made by Kolafa and Perram,¹⁴ for the special 2D case by Solvason *et al.*¹⁵

2.2 Scaling

In replacing (4) by (5) we immediately reap the benefits of rapid convergence. This can be seen more clearly when we make use of the previous results to compute the total potential energy of the system, summing over all charges, q_i :

$$\begin{aligned}\Phi_T &= \frac{1}{2} \sum_i q_i [V_r(r_i) + V_k(r_i) - V_s(r_i)] \\ &= \frac{1}{2} \sum_{\mathbf{n}}' \sum_{i=1}^N \sum_{j=1}^N q_i q_j \frac{\operatorname{erfc}(\alpha |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} \\ &= \frac{2\pi}{L^3} \sum_{\mathbf{k}}' \sum_{i=1}^N \sum_{j=1}^N q_i q_j \exp\{i\mathbf{k} \cdot (\mathbf{r}_j - \mathbf{r}_i)\} \frac{\exp\left(\frac{-|\mathbf{k}|^2}{4\alpha^2}\right)}{|\mathbf{k}|^2} - \frac{\alpha}{\pi^{1/2}} \sum_{i=1}^N q_i^2.\end{aligned}\quad (20)$$

Simple experimentation with the Ewald sums¹⁶ soon reveals a range of parameters in which one or both of the partial sums can be restricted. The example in Fig. 5, constructed for 40 randomly distributed positive and negative charges, shows that neither real-space nor k -space parts can be neglected in the range $\alpha L = 1-10$, even though large summation volumes were taken: $|\mathbf{n}|_{max} = 12$, $h^2 = (k_{max}/2\pi)^2 = 700$, on each side. Although we will come to the question of optimisation shortly, we can also get an idea of where to make cutoffs by successively truncating the two sums – Fig. 6. Inspection of these curves confirms the consensus choice found in the literature of $\alpha L \sim 2-5$, which allows the real-space sum to be restricted to a couple of box lengths ($|\mathbf{r}_j + \mathbf{n}| \leq 2L$), while maintaining reasonable accuracy. In fact, for the curve $n = 2$, $h^2 = 100$, the potential energy is accurate to better than 10^{-6} in the range $\alpha L = 2$ to $\alpha L = 9$.

The qualitative observations above still do not tell us how the overall computational effort scales with N , because the cutoff point in both sums may vary. Fixing $|\mathbf{r}_j + \mathbf{n}| \leq L$ – i.e., adopting the minimum image convention – would of course lead to an $O(N^2)$ scaling once more. The arbitrariness of the parameter α raises the question of whether one can choose the cutoffs n_{max} and k_{max} in either sum to reduce the overall effort. This seems a tall order, but just such a recipe was derived by Perram *et al.*,¹³ who showed that there does exist an optimal choice of parameters which reduces the scaling to $O(N^{3/2})$.

The trick is to weight the summation towards the k -space sum, thereby restricting the number of particle pairs which have to be considered in real space. Fincham¹⁷ gives an intuitive proof of Perram's $N^{3/2}$ scaling, which we reproduce here. First, we suppose that both sums are to converge to some small value, depending on the accuracy requirement of the application. We set this ‘small’ value equal to $\exp(-p)$. For the real-space sum (9),

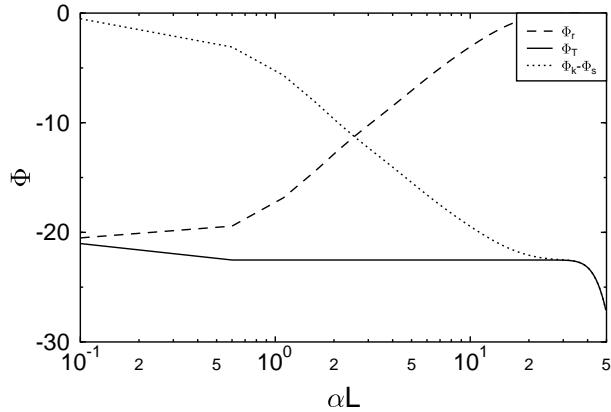


Figure 5. Convergence of real- (dashed) and k -space (dotted) Ewald potentials for different values of α .

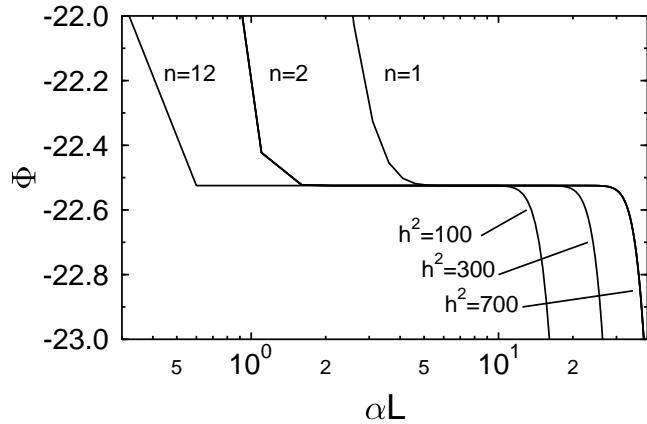


Figure 6. Ewald potential energy for different cutoffs in real- and k -space ($h^2 \equiv (k/2\pi)^2$).

this implies that at some cutoff radius R , we may write

$$\text{erfc}(\alpha r) |_{r=R} \sim \exp(-\alpha^2 R^2) = \exp(-p) .$$

From this we immediately obtain a constraint on α , namely:

$$\alpha = p^{1/2}/R. \quad (21)$$

Applying the same convergence criterion to the k -space sum, we have for some cutoff

wave-vector K :

$$\exp\left(-\frac{K^2}{4\alpha^2}\right) \sim \exp(-p),$$

or

$$p = \frac{K^2}{4\alpha^2}.$$

Thus, making use of our first constraint (21), we obtain

$$K = 2\alpha p^{1/2} = \frac{2p}{R}. \quad (22)$$

Once the accuracy (via p) and the cutoff radius R have been chosen, Equations 21 and 22 specify K and α . It remains for us to find a cutoff radius in the real-space sum that minimises the execution time. To estimate the latter, we assume that all N charges are distributed uniformly in a cubic box with side L , so that the number density $n = N/L^3$ remains constant as N is varied. The number of ions contained in a cutoff sphere is then:

$$N_c = \frac{4\pi}{3} R^3 n.$$

Hence, the execution time for the real-space sum can be approximated by:

$$T_r \simeq \frac{1}{2} N \frac{4\pi}{3} R^3 n t_r, \quad (23)$$

where t_r is the time needed to evaluate a single interaction pair. For the k -space sum, we have a total volume, using (22), of

$$\frac{4\pi}{3} K^3 = \frac{4\pi}{3} \frac{8p^3}{R^3},$$

whereby wave-vectors are chosen according to $\mathbf{k} = 2\pi(i_{\hat{x}}, i_{\hat{y}}, i_{\hat{z}})/L$ and $(i_{\hat{x}}, i_{\hat{y}}, i_{\hat{z}})$ is the usual integer triple. The volume per reciprocal point is just $(2\pi/L)^3$, so the number of points in the cutoff sphere is

$$N_k = \frac{4\pi}{3} \left(\frac{p}{\pi}\right)^3 \frac{N}{nR^3}, \quad (24)$$

and the overall execution time for the k -space sum is then

$$T_k = \frac{1}{2} \frac{4\pi}{3} \left(\frac{p}{\pi}\right)^3 \frac{N}{nR^3} t_k. \quad (25)$$

The total time for the Ewald summation is just the sum of (23) and (25):

$$T_{tot} = \frac{1}{2} \frac{4\pi}{3} \left[N n R^3 t_r + \left(\frac{p}{\pi}\right)^3 \frac{N}{nR^3} t_k \right].$$

If we fix the accuracy requirement p , the only free parameter remaining is R . The obvious thing to do is therefore to set $dT/dR = 0$, whereupon we find:

$$R_{opt} = \left(\frac{p}{\pi}\right)^{1/2} \left(\frac{t_k}{t_r}\right)^{1/6} \frac{N^{1/6}}{n^{1/3}}, \quad (26)$$

and

$$T_{opt} = 2T_r = 2T_k = \frac{4\pi}{3} N^{3/2} \left(\frac{p}{\pi}\right)^{3/2} (t_r t_k)^{1/2}. \quad (27)$$

The optimal configuration is thus equally divided (in terms of computation time) between real- and k -space sums. Assuming $t_r \sim t_k$, and stipulating a fairly conservative accuracy of $\exp(-p) \sim 5 \times 10^{-5}$, or $p = \pi^2$, we find from (26)

$$R_{opt} \simeq \pi^{\frac{1}{2}} L N^{-1/6},$$

with

$$\alpha L \simeq \frac{KL}{2\pi} = \pi^{\frac{1}{2}} N^{1/6}.$$

To illustrate this with a concrete example: for a system of 10000 particles, we would choose $R = 0.38L$ – less than the minimum-image box length – and $\alpha L = 8.2$.

For small systems, say $N < 10^4$, the conventional Ewald summation technique encapsulated by Equation 5 together with the simple optimisation recipe dictated by Equations 26, 21 and 22 is widely regarded as the standard method for periodic systems. The reciprocal-space sum itself can be optimised further by exploiting symmetries to reduce the number of lattice vectors needed.^{16,18} For larger systems, however, even the reduced $O(N^{3/2})$ cost-scaling eventually becomes prohibitive, forcing us to seek faster alternatives. Nevertheless, the direct Ewald method is still an important benchmark for assessing the performance of other, newer methods, some of which we will describe in the following sections.

3 Particle-Mesh Techniques

The deployment of a mesh or grid to speed up N -body calculations has long been standard practice in fields outside the traditional molecular dynamics arena, such as astrophysics and plasma physics, where the need to follow macroscopic trends over long timescales usually takes precedence over high accuracy or fine-grained spatial resolution. The term ‘macroscopic’ is loosely applied here to indicate that simulation charges (or masses) may represent a large number of physical entities. For example, a galaxy containing 10^{11} stars, each with mass M equal to one solar mass M_\odot , can be simulated by a system of 10^4 stars each weighing $10^7 M_\odot$. It turns out that this seemingly crude approximation works very well as long as one is interested in large-scale, collective effects, such as waves or instabilities with wavelengths on the same order as the system size itself. For *statistical* purposes, however, it is still highly desirable to use as many ‘particles’ as possible, which leaves one with the same type of computational challenge faced by someone wishing to follow the dynamics of a protein, say.

Since arbitrarily high accuracy was not (and is still not necessarily) a priority in many such applications, the $O(N)$ particle-mesh techniques, pioneered by Bunemann,¹⁹ Dawson,²⁰ Hockney³ and Birdsall² in the 1960s, quickly replaced direct summation as the workhorse computational tool in these fields.

3.1 Particle-in-Cell Codes

The subject of grid-based particle methods is too vast to do justice to within the confines of this review. For one thing, the representation of discrete charges on a grid introduces artificial structure factors, which in turn give rise to characteristic kinetic behaviour – such as modifications in the dielectric constant – storing up nasty surprises for the unwary. A quantitative understanding of such effects affords a certain amount of background theory, for which we heartily recommend the two definitive texts by Hockney & Eastwood³ and Birdsall & Langdon.² Nonetheless, it will prove instructive to review the basic concepts of the particle-mesh (PM) method. In doing so, we hope to clarify some of the ambiguous terminology which has crept into the MD literature on this subject.

Formally, the PM method can be derived by rigorous application of kinetic theory, simplifying the N -body problem with $6N$ degrees of freedom via the introduction of a smooth distribution function $f(\mathbf{r}, \mathbf{v}, t)$, obeying the kinetic Vlasov-Boltzmann equation:²¹

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} + q\mathbf{E} \cdot \frac{\partial f}{\partial \mathbf{v}} = \frac{\partial f}{\partial t} \Big|_c . \quad (28)$$

The term on the RHS is a collision term, describing the transfer of momentum between particles due to close encounters. Herein lies an important difference between PM and MD: in MD, collisions are treated automatically as part of the computation, whereas in a PM code, some approximate collision *model* must be introduced. Construction of a physically sensible and computationally stable model is fiendishly difficult, and luckily need not concern us here: for the time-being we will assume that our particle system is purely collisionless, and set $\partial f / \partial t|_c = 0$.

The distribution function $f(\mathbf{r}, \mathbf{v})$ is 6-dimensional, so the general solution of (28) is still intractable for most practical purposes. Even for problems reducible to a 1D geometry, one typically still needs to retain 2 or 3 velocity components in order to incorporate the appropriate electron motion and its coupling to Maxwell's equations, which effectively results in a 3- or 4-dimensional ‘Vlasov’-code. In the particle-mesh technique, the distribution function is represented instead by a large number of discrete ‘macro-particles’, each carrying a fixed charge q_i and mass m_i – Fig. 7, usually chosen so that the charge-to-mass ratio corresponds to a physical counterpart, for example e/m_e . This ensures that the particle trajectories in the simulation match those which would be followed by real electrons and ions.

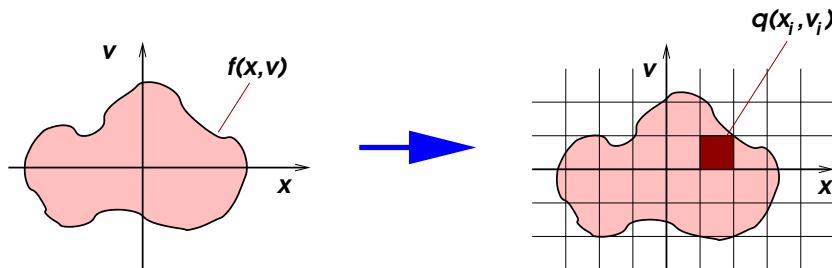


Figure 7. Correspondence between Vlasov and particle-in-cell representation of phase-space.

The particles are moved individually in Lagrangian fashion according to the equation of motion:

$$\frac{d}{dt}(\mathbf{v}_i) = \frac{q_i}{m_i} \mathbf{E} \quad i = 1, \dots, N \quad (29)$$

The density source needed to compute the electric field is obtained by mapping the local particle positions onto a grid via a weighting function W :

$$\rho(\mathbf{r}) = \sum_j q_j W(\mathbf{r}_j - \mathbf{r}), \quad j = 1, \dots, N_{\text{cell}} \quad (30)$$

where $W(\mathbf{r}_j - \mathbf{r})$ is a function describing the effective shape of the particles. Often it is sufficient to use a linear weighting for W – originally dubbed the ‘Cloud-in-Cell’ scheme by its inventors Birdsall & Fuss²² – although other more accurate methods are also possible. Once $\rho(\mathbf{r})$ is defined at the grid points, we can proceed to solve Poisson’s equation to obtain the new electric field. This is then interpolated back to the particle positions so that we can go back to the particle push step (29) and complete the cycle – Fig. 8.

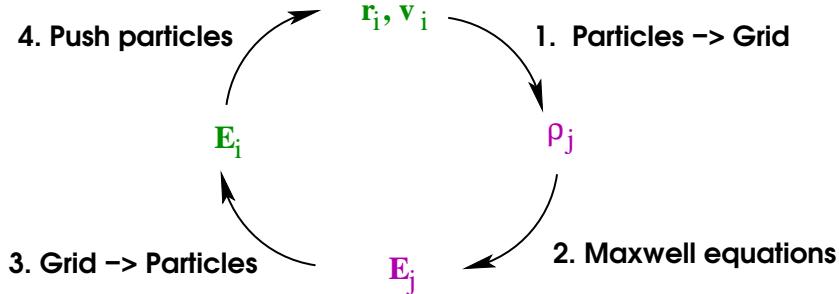


Figure 8. Schematic illustration of the particle-in-cell algorithm.

Because of its simplicity and ease of implementation, the PIC-scheme sketched above is currently one of the most widely used plasma simulation methods. It is particularly suitable for the study of *kinetic* or *non-Maxwellian* effects. The simplest variation of this technique is a ‘1D1V’-configuration: 1 space coordinate plus 1 velocity, the numerical behaviour of which was first examined by Dawson some forty years ago.²³

The heart of the code is based on the following difference equations:

$$\begin{aligned} \text{Particle pusher:} \quad v_i^{n+\frac{1}{2}} &= v_i^{n-\frac{1}{2}} + \frac{q_i}{m_i} E_i^n \Delta t, \\ x_i^{n+1} &= x_i^n + v_i^{n+\frac{1}{2}} \Delta t, \end{aligned} \quad (31)$$

$$\begin{aligned} \text{Density gather:} \quad \rho_j^{n+1} &= \sum_i q_i W(x_i - x_j), \\ S &= 1 - \frac{|x_i - x_j|}{\Delta x}, \end{aligned} \quad (32)$$

$$\text{Field integration:} \quad E_{j+\frac{1}{2}}^{n+1} = E_{j-\frac{1}{2}}^{n+1} + \rho_j^{n+1} \Delta x. \quad (33)$$

Notice that this scheme uses one of the simplest weighting schemes, namely linear interpolation, with the effect that the particles have an effective size of $2\Delta x$. Other weighting schemes are listed below in Fig. 9. The choice of scheme is, not surprisingly, a choice between speed and accuracy – see Hockney & Eastwood, Chapter 5 for a comprehensive discussion.³ We will return to this issue when we discuss the Particle-Mesh-Ewald method in Section 3.3. The subscript on W refers to the number of grid points along each axis

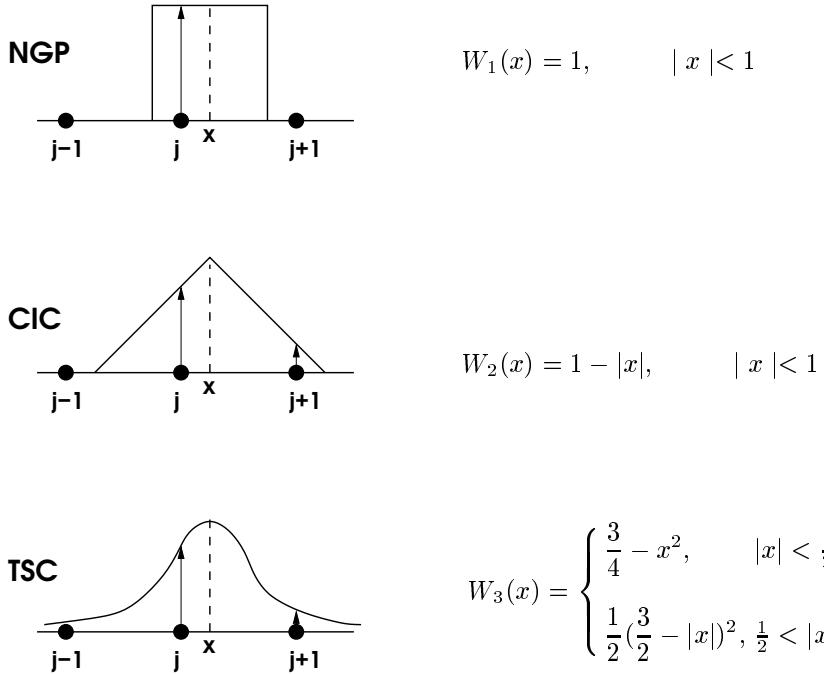


Figure 9. Charge assignment schemes: a) Nearest-grid-point, b) Cloud-in-cell and c) Triangular-shaped-cloud.

contributing to the charge-spreading. In all cases, $x \equiv (x_i - x_j)/\Delta x$ and $W(x) = 0$ outside the ranges indicated. The weighting function is extended into 2 and 3 dimensions by forming the product along each direction. For example, the popular CIC scheme in 3D takes the form of an 8-point weighting function:

$$W_2(\mathbf{r}) = (1 - |x|)(1 - |y|)(1 - |z|), \quad |x|, |y|, |z| < 1.$$

The difference scheme for the Poisson equation is also readily generalised to 3 dimensions:

$$6\phi_{j,k,l} - (\phi_{j+1,k,l} + \phi_{j-1,k,l} + \phi_{j,k+1,l} + \phi_{j,k-1,l} + \phi_{j,k,l+1} + \phi_{j,k,l-1}) = 4\pi\Delta^2\rho_{j,k,l}, \quad (34)$$

where the grid spacing is now assumed to be the same in all directions: $\Delta_x = \Delta_y = \Delta_z = \Delta$. Given that we can solve (34), the electric field is then obtained from:

$$E_{j,k,l} = -\frac{\phi_{j+1,k,l} - \phi_{j-1,k,l}}{2\Delta} - \frac{\phi_{j,k+1,l} - \phi_{j,k-1,l}}{2\Delta} - \frac{\phi_{j,k,l+1} - \phi_{j,k,l-1}}{2\Delta} \quad (35)$$

For periodic systems, Equation 34 can be solved using fast Fourier transforms (FFTs), by first computing $\tilde{\rho}(k)$ and then applying the inverse transform to $\tilde{\phi}(k) = \tilde{\rho}(k)/k^2$ to recover $\phi(r)$ in real space. This procedure results in $N_g \log N_g$ operations, where N_g is the total number of grid points.

Having obtained new field values on the mesh, these are then interpolated back to the particles using the *same* weighting scheme as for the charge assignment:

$$\mathbf{E}_i = \sum_{jkl} E_{j,k,l} W(r_i - r_{j,k,l}).$$

For example, in the CIC scheme in Fig. 9, a particle will receive field contributions from its nearest 8 grid points, weighted according to the volume-overlap between the mesh cell and a cube of side $\Delta/2$ centered on the particle.

Particle-mesh or particle-in-cell simulation typically uses many particles per cell, $N \gg N_g$, to keep field quantities as smooth as possible. The main computational cost is therefore not in the field solver, but in the integrator (or ‘particle pusher’, as it is commonly known), which is just $O(N)$. In electrodynamics, the pusher can be much more complicated than the simple linear acceleration implied by (31), often containing magnetic fields and relativistic factors. The drawback of mesh methods is that the spatial resolution is limited to distances $r \geq \Delta$, no matter how densely-packed the particles are, since these have an effective size $\sim \Delta$. In some sense, the particle-mesh technique is the ideal algorithm for long-range forces, because short-range interactions are automatically excluded! Of course this is not particularly helpful in the context of molecular dynamics, which is based on the ability to follow individual particle trajectories – including short-range encounters – explicitly and accurately.

3.2 P³M: Particle-Particle, Particle-Mesh

Ideally, one would like to have the best of both worlds offered by pure MD and PM respectively: high resolution of individual encounters, combined with a rapid mesh-based evaluation of long-range forces. This is precisely the philosophy behind the particle-particle, particle-mesh (P³M) method developed primarily by Eastwood in the 1970s.²⁴ The inter-particle force is initially split into two contributions:

$$\mathbf{F}_{ij} = \mathbf{F}_{ij}^{PP} + \mathbf{F}_{ij}^{PM}, \quad (36)$$

where the PP part is finite only over a few interparticle spacings, up to some cutoff radius r_c ; the PM part is assumed to be temporally and spatially smooth enough to be computed on a mesh – Fig. 10.

The question which immediately arises is: how do we implement this splitting in practice? Clearly, the short-range (PP) sphere should be as small as possible to minimise the number of direct PP interactions. If it is too small, however, spatial resolution will be lost which cannot be compensated for by the PM calculation. This is because PM-codes filter out all modes with $|\mathbf{k}| \geq \pi/\Delta$, where Δ is the mesh size.

The second issue concerns the matching of the force and potential contributions across the artificial boundary created by the cutoff sphere. We have already seen that the introduction of a grid causes the particles to acquire a finite form factor which depends on the details of the charge assignment scheme. We can illustrate how force splitting works by

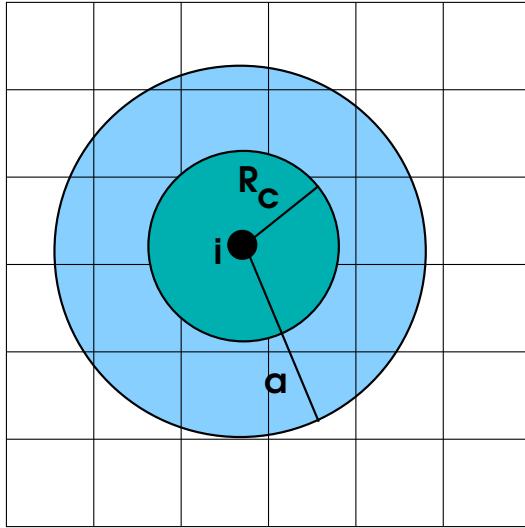


Figure 10. Force calculation using the P^3M splitting scheme.

taking the simplest of the schemes in Fig. 9, NGP, and working out its associated interparticle force. In three dimensions, the NGP scheme is equivalent to replacing point charges by spheres of radius $a/2$, with a uniform density given by:

$$\rho(r) = \begin{cases} \frac{6q}{\pi a^3}, & r < a/2 \\ 0, & r \geq a/2 \end{cases} \quad (37)$$

Elementary electrostatics shows that the force between two such spheres along the axis joining their centres is given by:

$$F_{\text{sphere}}(r) = \begin{cases} \frac{q^2}{a^2} \left(\frac{8r}{a} - \frac{9r^2}{a^2} + \frac{2r^4}{a^4} \right), & r < a \\ \frac{q^2}{r^2}, & r \geq a \end{cases} \quad (38)$$

The natural force-splitting choice in this case is to take the short-range cutoff $R_c = a$, and to set the PP force inside this sphere equal to the *difference* between the Coulomb force and this effective force contribution arising from the mesh points:

$$F^{PP}(r) = \begin{cases} F_c(r) - F_{\text{sphere}}(r), & r < a \\ 0, & r \geq a \end{cases}$$

This modified force-law is illustrated in Fig. 11. Note that we are not changing the physics here: the ultimate goal is still to match the exact Coulomb law for all r ; that is, when short- and long-range components are added together. Of course, we could guarantee

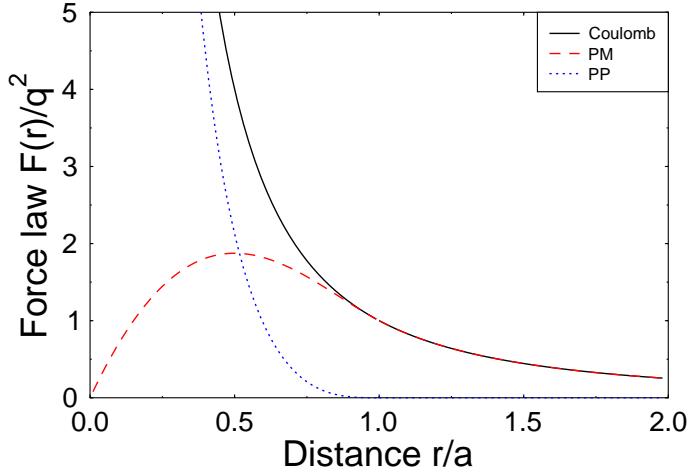


Figure 11. P^3M force-laws for short-range (PP, dotted line) and long-range (PM, dashed line) contributions.

this by summing over the *particles*:

$$F_i^{PM} = \sum_{j=1}^N F_{\text{sphere}}(r_i - r_j),$$

but then we would be back to square one, saddled with an $O(N^2)$ force-summation! The challenge is to arrange the fast, but approximate, particle-mesh calculation in such a way as to minimize the difference between the forces evaluated on the mesh and the ‘exact’ reference force given by (38).

Fortunately the particle-mesh algorithm outlined earlier (31)–(33) gives us plenty of leeway to achieve this goal. All four of the intermediate steps necessary to compute the ‘mesh’ forces – charge assignment, potential solver, differencing, and back-interpolation to the particles – present an opportunity to improve the matching of F_i^{PM} to the reference force. The details of this minimisation process are quite complex however, and the reader is referred to Hockney & Eastwood,³ Chapter 8. In the public version of their P^3M code,²⁴ charge assignment is performed using the TSC function, rather than NGP, which is found to give better overall performance. Once the PM parameters have been fixed, the calculation proceeds as in conventional MD: linked neighbour-lists are used to narrow down the search effort in the short-range calculation,^{25,6} and the velocities and positions are updated using a standard integrator.

The P^3M algorithm in its original form as advocated by Hockney & Eastwood has not been widely adopted for production MD simulation, despite its promising scaling characteristics for large particle number. This is perhaps due to the uncertainty created by the admittedly complicated force-splitting procedure, leaving the user with less than 100% confidence in its accuracy. Recent comparisons of P^3M with other techniques²⁶ has

demonstrated that these doubts are largely unfounded, however, so we can perhaps expect a rejuvenation of this method in other fields too.

A more transparent version of P^3M has been implemented for ‘dense plasma’ (neutral electron-ion) systems by Nishihara *et al.*²⁷ In their scheme, the short- and long-range contributions are organised by the mesh itself – Fig.12. In this example, the Coulomb forces on particle i are summed over the other particles within its own cell plus those (e.g.: j_1) in the neighbouring 26 cells (hatched region). The forces from particles outside (j_2) are computed from the mesh (shaded region), but *excluding* the charges inside the hatched region. The splitting is thus even more artificial than before, but actually easier to implement.

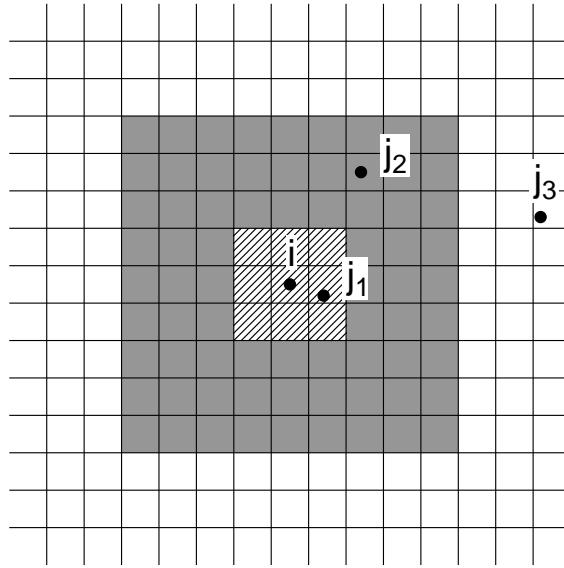


Figure 12. P^3M scheme using grid-based force-splitting.

3.3 Particle-Mesh-Ewald

By now the reader will probably have noticed the striking similarity between the classical Ewald method – summarised by Eq. 6 – and the force- or potential-splitting principle behind P^3M . This correspondence is perhaps surprising because in a sense, the two methods approach the N -body problem from different philosophical viewpoints: the Ewald method seeks to perform an exact, periodic lattice-sum by deliberately introducing a charge-spreading function to speed up convergence; P^3M is a means of rapidly evaluating the long-range force contribution through the use of a grid, due to which the charges automatically acquire a finite size.

In Section 2.2 we saw that the optimal configuration for the Ewald sum is achieved when the computational effort is equally shared between the real- and reciprocal-space

sums. In P³M, the bias is shifted even further towards the particle-mesh calculation (which is also usually performed in reciprocal space) to guarantee something like an $O(N)$ scaling. In 1993, Darden, York and Pedersen⁴ realised that the Ewald sum could be recast in the P³M form by using a large value for the convergence parameter α . This accelerates the convergence of the real-space sum, which can then be restricted to a smaller cutoff radius R_c , and ensures that the main contribution is computed in k -space. According to Eq. 24, this means summing over a large number of k -vectors, $N_k \sim N/nR_c^3$ for each particle.

At this point, Darden *et al.* took the analogy with P³M a step further, arguing that Gaussian-shaped charges could be equally well mapped onto a regular grid and the resulting potential computed by a FFT. Assuming the charges can be adequately represented on a $m \times m \times m$ mesh with a fixed number of particles per cell, i.e.: $m^3 \propto N$, then the scheme will scale as $N \log N$, the time it takes to perform the Fourier transform. Thus, instead of evaluating (17), we first compute the FFT of the gridded density, which by inspection of (15) is given by:

$$\tilde{\rho}_{jkl}(k) = \frac{1}{L^3} \exp\left(-\frac{k^2}{4\alpha^2}\right) \sum_i q_i \exp(i\mathbf{k} \cdot \mathbf{r}_i) \quad (39)$$

The potential in k -space is then just the density multiplied by an ‘influence’ function $\tilde{G}(k)$:

$$\tilde{\phi}_{jkl}(k) = \tilde{G}(k)\tilde{\rho}(k), \quad (40)$$

where

$$\tilde{G}(k, \alpha) = \frac{\exp(-k^2/4\alpha^2)}{k^2}, \quad (41)$$

and the fields are given simply by:

$$\begin{aligned} \tilde{E}_{jkl}(\mathbf{k}) &= -i\mathbf{k}\tilde{\phi}_{jkl}(k) \\ &= -i\mathbf{k}\tilde{G}(k, \alpha)\tilde{\rho}(k). \end{aligned} \quad (42)$$

The inverse FFT then yields the potential and fields at the grid points, which can then be interpolated back to the particle positions.²⁸

In its original form, the particle-mesh-Ewald (PME) algorithm of Darden *et al.* stuck with Gaussian charge shapes for consistency with the real-space part of the Ewald sum (which is evaluated conventionally according to Eq. 9). Unfortunately, this has the drawback that a large number of points – typically $8^3 \sim 100$ – are needed to map the charges onto the grid, creating a new numerical bottleneck. The advantage of the P³M scheme is that it restricts the charge distribution to 8 (CIC) or 27 (TSC) grid points respectively. A number of works have since appeared^{29,26} which fully adopt the P³M concept of a narrow, finite assignment function. Just as in P³M, the errors incurred by this procedure can be compensated by modifying the influence function $\tilde{G}(k)$ – (41) – to include the discreteness effects of the grid.³⁰ For consistency, the same charge shape should be used for the real-space sum, which, as shown by Heyes,¹⁰ will still guarantee rapid convergence. PME can thus be formally regarded as a special case of P³M, though we have deliberately treated the methods separately here because of their independent historical development.

Many implementations and applications of the PME method have already appeared,³¹ including a sophisticated variation based on B-spline interpolation,³² assessments of its performance relative to crude cutoff methods³³ as well as multipole methods.²⁸

4 Multipole Methods

So far in this review, we have dealt exclusively with periodic systems, which can be neatly handled by some form of Ewald summation. As we have seen in the previous two sections, there are essentially two choices in this case: the direct, optimised scheme of Perram *et al.*,¹³ or a grid-based P³M/PME scheme.^{4,29} There is, of course, a large number of N -body problems for which periodic boundaries are completely *inappropriate*, for example: galaxy dynamics, electron-beam transport, large proteins,³¹ and any number of problems with complex geometries. So how does one get round the N^2 -bottleneck if there is no symmetry to exploit?

Two new approaches to this problem were put forward in the mid-1980s, the first from Appel³⁴ and Barnes & Hut,³⁵ who proposed $O(N \log N)$ -schemes based on hierarchical grouping of distant particles; the second from Greengard & Rohklin,³⁶ who went one better, devising an $O(N)$ solution with rounding-error accuracy. These two methods – known today as the ‘hierarchical tree algorithm’ and the ‘fast multipole method’ respectively – have revolutionised N -body simulation in a much broader sense than the specialised periodic methods discussed earlier. They offer a generic means of accelerating the computation of many-particle systems governed by central, long-range potentials.

Although the FMM is currently more widespread in molecular dynamics than the Barnes-Hut (BH) tree algorithm, we nevertheless give both methods an equal airing here. For one thing, the methods are conceptually very similar (and are therefore related); secondly, both FMM and BH are still evolving, and it is likely that some hybrid, adaptive scheme may eventually prevail as a competitive alternative to, say, PME even for periodic simulations with moderate numbers of particles.

4.1 The Barnes-Hut Tree Algorithm

An inherent inefficiency in direct force-summation is that one does not distinguish near-neighbours from more distant particles; each pair evaluation requires the same computational effort, even though the individual contributions of distant particles may be negligibly small. Introduction of an artificial cutoff radius can separate out the important and less important partners, but this procedure only works well for short-range potentials; for Coulomb potentials, errors will accumulate as a result of abrupt truncation.

In 1986, Barnes and Hut³⁵ introduced a scheme in which the physical space is systematically divided up so as to establish and maintain a relationship between each particle and its neighbours. The resulting ‘tree’ structure can then be used to group distant clusters of particles into a single charge or mass, thereby reducing the number of interactions in the force/potential calculation. These codes are sometimes described as oct-tree codes to distinguish them from so-called binary tree codes,^{37,38} based on nearest neighbour *pairs*. Although binary trees might reflect the structure of the system more closely, the Barnes and Hut method is by far the most commonly used method due to its conceptual simplicity and easy, low-overhead tree construction.

There is no single correct way to go about the tree-building process, but one method which produces an identical structure to the original BH scheme is as follows.³⁹ First, a root cell is created containing all simulation particles – Fig.13a) This cell is then divided into eight equally sized subcells – Fig.13b). For each subcell, one asks whether it contains none, one, or more than one particle.

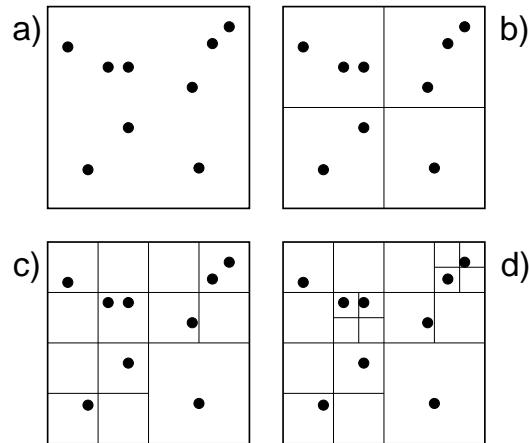


Figure 13. Step-by-step division of space for a simple 2-D particle distribution.

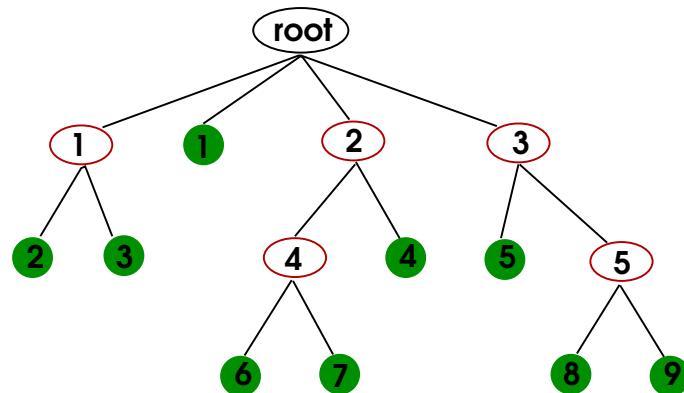


Figure 14. Tree data structure corresponding to Fig. 13d)

If the cell is empty, this cell is ignored; if there is one particle in the cell, this is stored as a ‘leaf’ node in the tree structure; if there are more particles in a cell, this cell is stored as a ‘twig’ node and subdivided further. The subdivision process continues until there are no cells with more than one particle left, which ultimately leads to Fig. 13d). The division of space just described is not used as a grid in the particle-mesh sense, but rather as a *bookkeeping* structure. At each division step, the tree data structure is augmented with the twig-nodes belonging to next level down in the hierarchy – Fig. 14. Each node in the tree is associated with a cubic volume of space containing a given number of particles; empty cells are not stored. Pointers to the *parents* of each leaf and twig node are also kept in the tree structure.

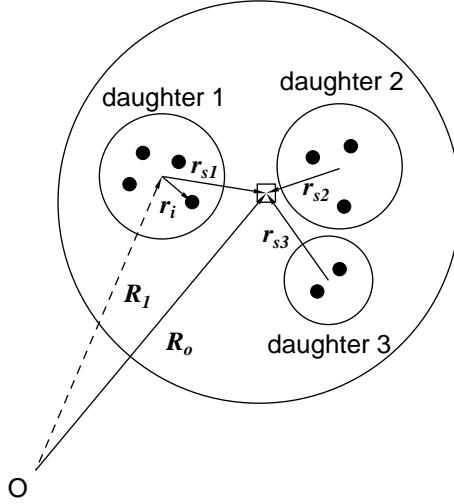


Figure 15. Origin shift for the multipole calculation: the circles symbolize the pseudoparticles (twig-nodes); \mathbf{r}_i is a vector from a constituent particle to the centre of charge of the daughter node and $\mathbf{r}_{s1}, \mathbf{r}_{s2}$, etc. are the shifting vectors to the new origin O' (\square), which is the centre of charge of the parent node.

Once the tree is in place, the twig nodes (represented by the 5 ellipses in Fig. 14) can be ‘loaded’ with information about their physical contents, like their centres of charge,

$$\mathbf{r}_{coc} = \frac{\sum_i |q_i| \mathbf{r}_i}{\sum_i |q_i|} \quad (43)$$

and their multipole moments:

$$M = \sum_i q_i \quad (44)$$

$$D_\alpha = \sum_i q_i r_{i\alpha} \quad (44)$$

$$Q_{\alpha\beta} = \sum_i q_i (3r_{i\alpha} r_{i\beta} - r_i^2 \delta_{\alpha\beta}) \quad (45)$$

This information will be needed later for the force calculation when the twigs are treated as pseudoparticles. This loading of twig-nodes, can be performed very rapidly by propagating information up the tree level-by-level, from individual particles (leaves), through the intermediate twigs until the root is reached.

To do this, we make use of the already-computed multipole moments of the daughter cell to calculate the moments of the parent cell. Each individual \mathbf{r}_i in the sum of a daughter moment is shifted by the same vector \mathbf{r}_{sd} . For example,

$$\begin{aligned} \sum_i q_i x_i &\rightarrow \sum_i q_i x_i - x_{sd} \sum_i q_i \\ \sum_i q_i x_i^2 &\rightarrow \sum_i q_i x_i^2 - 2x_{sd} \sum_i q_i x_i + x_{sd}^2 \sum_i q_i \end{aligned}$$

$$\sum_i q_i x_i y_i \rightarrow \sum_i q_i x_i y_i - x_{sd} \sum_i q_i y_i - y_{sd} \sum_i q_i x_i + x_{sd} y_{sd} \sum_i q_i$$

and so forth. These results are used later to calculate the contribution of selected twig-nodes to the total force or potential.

Obviously, the tree-building process incurs a certain overhead in an N -body code, so it is natural to ask how much. A rough estimate of how many divisions are needed to reach a typical cell, starting from the root, can be obtained from the average size of a cell containing one or more particles. The average volume of such a cell is the volume of the root cell V divided by the number of simulation particles N . Moreover, the average length of a cell is a power of $V^{1/3}/2$. Therefore,

$$\left(\frac{1}{N}\right)^{1/3} = \left(\frac{1}{2}\right)^x,$$

which means that the height x of the tree is of the order

$$\log_2 N^{1/3} = \frac{1}{3 \log 2} \log N \simeq \log N. \quad (47)$$

Starting from the root, an average of $\log N$ divisions are necessary to reach a given leaf. The tree contains N leaves, therefore the time required to construct the tree is $O(N \log N)$. In practice, tree-building actually comprises only 3–5% of the total force calculation, so it can be performed every timestep.

The tree structure provides the means to distinguish between close particles and distant particles without actually calculating the distance between every particle. The force between near-neighbours is calculated directly, whereas more distant particles are grouped together to pseudoparticles. An *interaction list* is thus built for each particle by traversing the tree from node-to-node and deciding whether to accept the node as-is, or subdivide further. There are actually a number of so-called ‘multipole acceptance criteria’ (MACs) of varying complexity,⁴⁰ the simplest of which is the original ‘ s/d ’ criterion introduced by Barnes and Hut.³⁵ Beginning at the root of the tree, the ‘size’ of the current node (or twig), s , is compared with its distance from the particle, d . If the ratio s/d is smaller than some preset value, θ , then the internal structure of the pseudoparticle is ignored and it is added to the interaction list for that particle. Otherwise, this node is resolved into its daughter nodes, each of which is recursively examined according to s/d and, if necessary, subdivided. Fig. 16 illustrates this comparison at two different stages of the tree-walk. This continues until all nodes have been examined, i.e.: when we have returned to the root.

The result of this procedure is an interaction list, the length of which depends both on N and θ , the multipole acceptance parameter – Fig. 17. The case $\theta = 0$ is equivalent to computing all particle-particle interactions; which is exact but rather pointless, because the operation count is again $O(N^2)$. This is in fact slower than direct PP because of the tree-building and traversal overheads. A practical choice using this MAC proves to be in the range $\theta = 0.3\text{--}1.0$, depending on the application.

The asymptotic scaling of this algorithm can be estimated by considering the average number of interactions n_{int} in a spherical, homogeneous distribution surrounding a test particle. For nonzero θ , it can be shown that:³⁹

$$n_{int} \sim \log N / \theta^2, \quad (48)$$

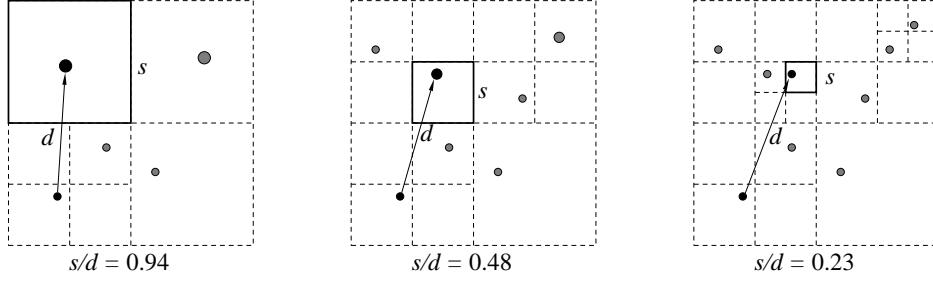


Figure 16. The relation s/d for different levels of the tree.

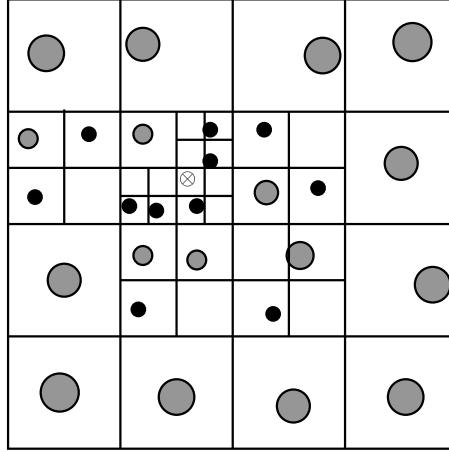


Figure 17. Interaction list generated for particle marked \otimes , using multipole acceptance parameter $\theta = 1.0$. The black circles are single charges; the shaded circles multipole expansions.

so the time required to calculate the force on a given particle is $O(\log N)$, which means the number of operations to compute the force on all N bodies will scale as $O(N \log N)$.

Having determined the interaction lists for each particle, all that remains is to compute the potentials and forces. However, we have already anticipated making use of multipole expansions to take account of the charge *distribution* inside the pseudoparticle terms. Referring to Fig. 18, the potential at particle P due to the pseudoparticle is the sum of the potentials Φ_i due to the particles contained in the cell,

$$\Phi(\mathbf{R}) = \sum_i \Phi_i(\mathbf{R} - \mathbf{r}_i),$$

where \mathbf{r}_i is the vector from the particle to the centre of mass and the origin is, for simplicity, the individual particle on which the force of the pseudoparticle is calculated. Here we

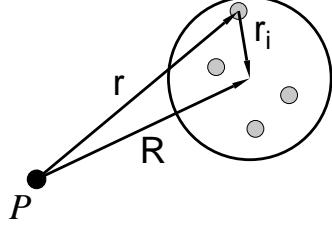


Figure 18. Geometry of multipole expansion.

consider a $1/r$ -potential, therefore

$$\begin{aligned}\Phi_i(\mathbf{R} - \mathbf{r}_i) &= \frac{q_i}{|\mathbf{R} - \mathbf{r}_i|} \\ &= \frac{q_i x_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}}.\end{aligned}$$

Expanding this potential about \mathbf{R} up to quadrupole order, gives:

$$\begin{aligned}\Phi(\mathbf{R}) &= \sum_i q_i \left[1 - \mathbf{r}_i \frac{\partial}{\partial \mathbf{r}} + \frac{1}{2} \mathbf{r} \cdot \mathbf{r} \frac{\partial}{\partial \mathbf{r}} \frac{\partial}{\partial \mathbf{r}} + \dots \right] \frac{1}{R} \\ &= \sum_i q_i \left[1 - x_i \frac{\partial}{\partial x} - y_i \frac{\partial}{\partial y} - z_i \frac{\partial}{\partial z} \right. \\ &\quad + \frac{1}{2} x_i^2 \frac{\partial}{\partial x} \frac{\partial}{\partial x} + \frac{1}{2} y_i^2 \frac{\partial}{\partial y} \frac{\partial}{\partial y} + \frac{1}{2} z_i^2 \frac{\partial}{\partial z} \frac{\partial}{\partial z} \\ &\quad + \frac{1}{2} x_i y_i \left(\frac{\partial}{\partial x} \frac{\partial}{\partial y} + \frac{\partial}{\partial y} \frac{\partial}{\partial x} \right) \\ &\quad + \frac{1}{2} y_i z_i \left(\frac{\partial}{\partial y} \frac{\partial}{\partial z} + \frac{\partial}{\partial z} \frac{\partial}{\partial y} \right) \\ &\quad \left. + \frac{1}{2} x_i z_i \left(\frac{\partial}{\partial x} \frac{\partial}{\partial z} + \frac{\partial}{\partial z} \frac{\partial}{\partial x} \right) \right] \frac{1}{R} \\ &= \sum_i q_i \left[\frac{1}{R} + x_i \frac{x}{R^3} + y_i \frac{y}{R^3} + z_i \frac{z}{R^3} \right. \\ &\quad + \frac{1}{2} x_i^2 \left(-\frac{1}{R^3} + \frac{3x^2}{R^5} \right) + \frac{1}{2} y_i^2 \left(-\frac{1}{R^3} + \frac{3y^2}{R^5} \right) \\ &\quad + \frac{1}{2} z_i^2 \left(-\frac{1}{R^3} + \frac{3z^2}{R^5} \right) \\ &\quad \left. + x_i y_i \left(\frac{3xy}{R^5} \right) + y_i z_i \left(\frac{3yz}{R^5} \right) + x_i z_i \left(\frac{3xz}{R^5} \right) \right]\end{aligned}$$

This can be rearranged to give:

$$\Phi(\mathbf{R}) = \frac{M}{R} + \sum_{\alpha} \frac{r_{\alpha} D_{\alpha}}{R^3} + \sum_{\alpha\beta} \frac{1}{2} Q_{\alpha\beta} \frac{r_{\alpha} r_{\beta}}{R^5}, \quad (50)$$

where M , D_{α} and $Q_{\alpha\beta}$ are the monopole, dipole and quadrupole moments of the pseudoparticles, defined previously in (44). The indices α and β refer to the 1 components x, y, z , so that $\mathbf{r}_{i\alpha} = (x_i, y_i, z_i)$ etc.

The corresponding electric field can be obtained directly from (50) by differentiating with respect to \mathbf{R} :

$$\mathbf{E}(\mathbf{R}) = -\frac{\partial}{\partial \mathbf{R}} \Phi(\mathbf{R}), \quad (51)$$

which gives for the each component γ :

$$E_{\gamma} = \frac{r_{\gamma}}{R^3} M + \sum_{\alpha} \left(\frac{3r_{\alpha} r_{\gamma}}{R^5} - \frac{r_{\alpha} \delta_{\alpha\gamma}}{R^3} \right) D_{\alpha} + \sum_{\alpha\beta} \frac{5r_{\alpha} r_{\beta} r_{\gamma}}{R^7} \cdot \frac{1}{2} Q_{\alpha\beta} - \sum_{\alpha} \frac{r_{\alpha}}{R^5} Q_{\alpha\gamma}.$$

Expanding the multipole moments and the summations, we arrive at a somewhat more convenient form for implementing in a code:⁵

$$\begin{aligned} E_x &= \frac{x}{R^3} \sum_i q_i \cdot \\ &- \left(\frac{1}{R^3} - \frac{3x^2}{R^5} \right) \cdot \sum_i q_i x_i + \frac{3xy}{R^5} \cdot \sum_i q_i y_i + \frac{3xz}{R^5} \cdot \sum_i q_i z_i. \\ &+ \left(\frac{15x^3}{R^7} - \frac{9x}{R^5} \right) \cdot \frac{1}{2} \sum_i q_i x_i^2 + \left(\frac{15xy^2}{R^7} - \frac{3x}{R^5} \right) \cdot \frac{1}{2} \sum_i q_i y_i^2 \\ &+ \left(\frac{15xz^2}{R^7} - \frac{3x}{R^5} \right) \cdot \frac{1}{2} \sum_i q_i z_i^2 + \left(\frac{15x^2y}{R^7} - \frac{3y}{R^5} \right) \cdot \sum_i q_i x_i y_i \\ &+ \left(\frac{15x^2z}{R^7} - \frac{3z}{R^5} \right) \cdot \sum_i q_i x_i z_i + \left(\frac{15xyz}{R^7} \right) \cdot \sum_i q_i y_i z_i. \end{aligned} \quad (53)$$

The y - and z -components can be found by cyclic rotation. Compared with the direct force calculation, Equation 53 above contains 8 additional multipole terms which must be evaluated for each twig-node in the interaction list. Clearly, this necessitates a certain overhead for the BH tree algorithm, so that it will only be more efficient above a certain particle number. Where this breakeven point is exactly, depends mainly on the accuracy desired, i.e. on the choice of θ . We defer comparison of multipole schemes until later (Fig.24), but for the moment, we note that the standard tree code relies very much on a trade-off between speed and accuracy – Fig. 19.

4.2 The Fast-Multipole Method (FMM)

The Fast-Multipole Method was developed by Greengard & Rohklin, shortly after the Barnes–Hut algorithm appeared.^{36,41,42} In some sense, therefore, the FMM can be thought

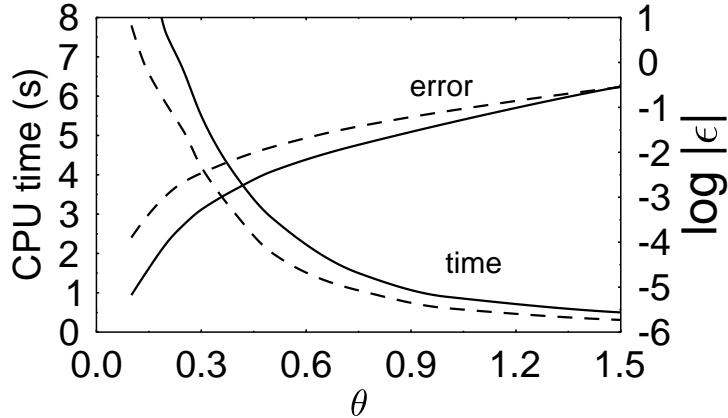


Figure 19. Trade-off between CPU time per integration step and average force error as a function of multipole acceptance parameter using monopole terms only (dashed lines) and quadrupole terms (solid lines).

of as an elegant refinement of the BH tree-code, but in fact it was developed independently. The FMM makes rigorous use of the fact that a multipole expansion to infinite order contains the total information of a particle distribution. As in the BH algorithm, the interaction between near-neighbours is calculated by direct particle–particle force summation, and more distant particles are treated separately. However, the distinction between these two contributions is obtained in a different way. In FMM the distant region is treated as a *single* ‘far-field’ contribution, which is calculated by a high-order multipole expansion.

By forming high-order multipole moments at the lowest level of the tree and carefully combining and shifting these centres up and down the tree, Greengard and Rokhlin showed that the N -body problem can in principle be reduced to an order $O(N)$ algorithm.³⁶ Arbitrary accuracy (e.g., within numerical rounding error) can be assured *a priori* by taking sufficient terms in the expansion. Because of this, it is essential to find a concise mathematical representation of the multipoles and their shifting theorems. The first FMM codes were two dimensional^{36,43} and exploited a convenient complex variable notation to represent the potentials and fields. The 3D formulation uses spherical harmonics instead and was also derived by Greengard,⁴¹ and later implemented by Schmidt and Lee⁴⁴ for periodic systems.

Like the tree algorithm, FMM starts with a box big enough to contain all the simulation particles, and this box is subsequently subdivided into boxes of length $d/2^r$ ($r = 0, 1, 2, \dots$) equivalent to 8^r equal sized subvolumes (4^r in two dimensions). In contrast to the tree method, however, this is done for *every box* up to a given maximum refinement level R , regardless of the number of particles it contains – Fig. 20. This maximum refinement level R is chosen so that the number of boxes is approximately equal to the number of the simulated particles N . This means that assuming the N particles are more or less homogeneously distributed, the maximum refinement level (in 3D) must be chosen as

$$R = \log_8 N.$$

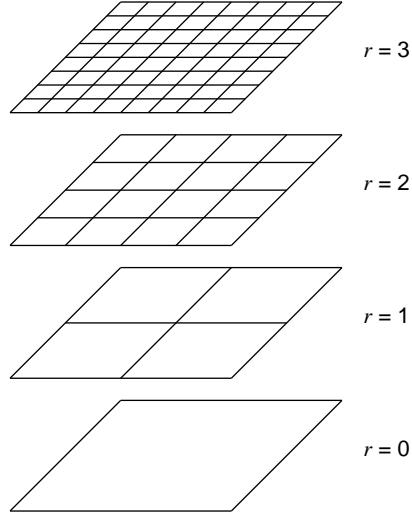


Figure 20. Division of the simulation box in a fast multipole code.

The FMM tree structure is used to build ‘near-neighbour’ lists of boxes at each refinement level r . Near-neighbours are defined as the box itself and any box at the same level with which it shares a boundary point. By contrast, a box on the same level which is *not* in a near-neighbour list is *well separated*: A local multipole expansion made about the centre of this box will then automatically be valid. Using the usual conventions⁴⁵ for spherical harmonics and units, the generalized version of (50) is:

$$\Phi(\mathbf{r}) = 4\pi \sum_{l,m} \frac{M_{lm} Y_{lm}(\theta, \phi)}{(2l+1)r^{l+1}} \quad (54)$$

in spherical coordinates (r, θ, ϕ) relative to an origin O , with multipole moments given by:

$$M_{lm} = \sum_i q_i r_i^l Y_{lm}^*(\theta_i, \phi_i), \quad (55)$$

where the charges now have the coordinates (r_i, θ_i, ϕ_i) . As in the Barnes-Hut algorithm, multipole moments are first computed for each box at the highest refinement level R , but this time relative to the *centre of the box* rather than the centre of charge of the particles. The maximum number of terms L in the multipole expansion is chosen such that:^{36,43}

$$\left(\sum_i |q_i| \right) 2^{-L} \leq \epsilon, \quad (56)$$

where ϵ is the desired precision.

Next, the multipole moments on the next coarsest refinement level $r = R - 1$ are calculated, and just as for the tree method, the shifted multipole moments of the daughter cells can be used to obtain the moments of the parent cell. Shifting the origin O to O' by

a translation vector \mathbf{r}_t , a transformation of the multipole moments is needed to obtain the expansion in terms of the new vector $\mathbf{r}' = \mathbf{r} - \mathbf{r}_t$ relative to O' . This transformation is given by:

$$M'_{l'm'} = \sum_{l,m} T_{l'm',lm}^{MM} M_{lm}, \quad (57)$$

with the transformation matrix

$$T_{l'm',lm}^{MM} = 4\pi \frac{(-r_t)^{l'-1} Y_{l'-l,m'-m}^*(\theta_t, \phi_t) a'_{l'-l,m'-m} a_{lm} (2l'+1)}{2(l+1)[2(l'-l)+1] a_{l'm'}},$$

and a_{lm} is defined as

$$a_{lm} = (-1)^{l+m} \frac{2(l+1)^{1/2}}{[4n(l+m)!(l-m)!]^{1/2}}.$$

This is just a generalization of the 1 shifting relations used for the (quadrupole order) tree

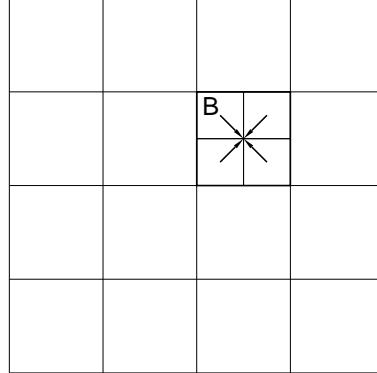


Figure 21. Shifting of the multipole expansion from the daughters on level $r = 3$ to the parents up $r = 2$ during the upward pass.

code, with the main difference that the moments are calculated relative to the centre of the cell (see Fig. 21). In this fashion, the moment expansions are carried up to to root level $r = 0$, which ultimately contains an L -term multipole expansion of the whole system.

So far, apart some hair-raising mathematics, there is little to choose between a standard tree code and the FMM. At this point, we could just use our L -term multipole expansions for the ‘box-to-particle’ strategy of the tree algorithm. In the FMM, however, we include the extra step of evaluating ‘box-box’ interactions. To do this, we perform a downward pass (from $r = 0$ to $r = R$), in which a careful distinction is made in the interaction lists for the boxes. There are actually *three* regions: the near field, the interactive field and the far field. The near field consists of the neighbouring cells; the far field is the entire simulation box *excluding* the cell in question and its neighbours. The interactive field is the part of the far field that is contained in the near field of this cell’s *parents* – Fig. 22.

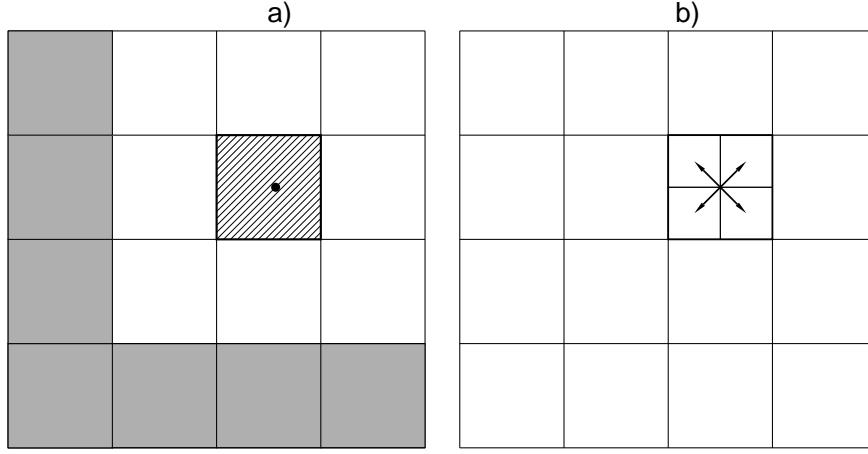


Figure 22. a) List of well-separated boxes (interactive field) which contribute to local expansion of the hatched box in b) at level $r = 2$.

In this downward pass, each multipole expansion is converted into a *local expansion* (i.e., a Taylor expansion about the centre of all well-separated boxes at each level). Using the notation above, this can be expressed thus:

$$\Psi(\mathbf{r}) = 4\pi \sum_{l,m} L_{lm} r^l Y_{lm}(\theta, \phi), \quad (60)$$

where L_{lm} are referred to as the local moments of the Taylor series expansion, obtained from the original multipole moments by the transformation:

$$L_{l'm'} = \sum_{l,m} T_{l'm',lm}^{LM} M_{lm}, \quad (61)$$

where the transformation matrix $T_{l'm',lm}^{LM}$ is now

$$T_{l'm',lm}^{LM} = 4\pi \frac{(-1)^{l+m} Y_{l'+l,m'-m}^*(\theta_t, \phi_t) a_{l,m} a_{l'm'}}{r_t^{l'+l+1} (2l+1)(2l'+l) a_{l'+l,m'-m}}. \quad (62)$$

The region of validity of the multipole expansions that contribute to the local expansion consists of all boxes at the same level which are not near neighbours. At levels $r = 0$ and $r = 1$, there are no boxes which fulfill this requirement, so we just set

$$\Psi_0 = \Psi_1 = 0$$

From $r = 2$ to $r = R$ the following operations can now be performed: For each box on level r , the local expansion of the parent box is shifted to the centre of each of its daughters as in Fig. 22b). In other words,

$$L'_{l'm'} = \sum_{l,m} T_{l'm',lm}^{LL} L_{lm},$$

with

$$T_{l'm',lm}^{LL} = 4\pi \frac{r_t^{l-l'} Y_{l'-l,m'-m}(\theta_t, \phi_t) a_{l'm'} a_{l-l',m-m'}}{(2l'+1)[2(l-l')+1] a_{lm}}.$$

In this local expansion of the daughter cell there are now boxes missing. These are the boxes that do not touch the current daughter cell, but do not contribute to the local expansion of the parent cell either – in other words, the boxes of the *interactive field* – Fig. 23a). Their contribution has to be added to the local expansion of the daughter cell. The result is the local expansion due to all particles in all boxes at the same level which are not near neighbours. However, the ultimate aim is to evaluate the potential or force not on the box, but rather on the particles inside the box. Therefore, once the highest refinement level is reached, the local expansions at the individual particle locations are evaluated.

Finally, the remaining interactions with the particles in neighbouring boxes and the box itself are added by direct summation of the particle–particle interactions – Fig. 23b).

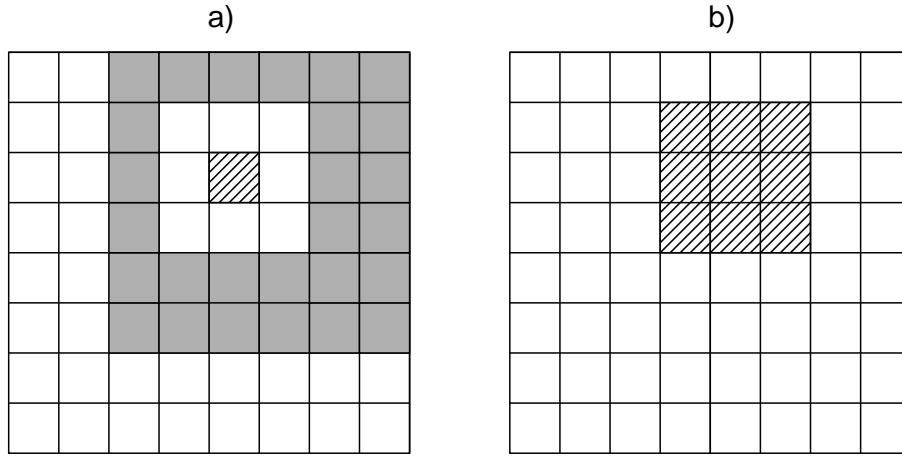


Figure 23. a) List of well-separated boxes (interactive field – grey) which contribute to the local expansion of the hatched box at level $r = 3$. b) Near-neighbour boxes (hatched region) for the direct particle–particle sum at the finest level (in this case $R = 3$)

Some recent improvements to these formulations – in which the central multipole transformations are optimised – have been proposed by Petersen et al.⁴⁶ and White and Head-Gordon.⁴⁷ A new implementation for high-precision quantum chemistry applications has been reported by Dachsel⁴⁸ at this Winter School.

At first sight it seems that FMM codes with a computation time proportional to N must be superior to tree codes – which scale as $N \log N$ – and particle–particle codes – which scale either as N^2 or $N^{3/2}$ depending on the boundary conditions. However, FMM codes have a large overhead due to the multipole and Taylor expansions, so one has to ask instead: at what point is it appropriate to use an FMM code instead of a PP or tree code? This is not a simple question to answer, because the computation time depends not only on the number of particles N , but also on how accurately the calculation should be performed.

For the tree code, accuracy is determined by the tolerance parameter θ , for the FMM code the number of multipole terms L and the refinement level R have the same function. Schmidt and Lee⁴⁴ showed that the overall polynomial dependence of the computation on N , L , and R is given by

$$P = aBL^2 + bNL^2 + cBL^4 + dB \left[g_1 + g_2 \left(\frac{N}{B} \right) + g_3 \left(\frac{N}{B} \right)^2 \right], \quad (65)$$

where a, b, c, g_1, g_2 , and g_3 are machine-dependent parameters and B is the number of boxes (8^R) on the highest refinement level. Interestingly, Eq. 65 shows that it is crucial that ratio N/B is kept small, because only then is the N^2 dependence actually removed, leaving a dominant $O(N)$ scaling. In practice, this means increasing the number of levels as N is increased, typically leading to a ratchet-like timing curve.

One of the assumptions of the FMM is that the distribution of the particles is more or less homogeneous. Nonuniform particle distributions would either require a high refinement level or a large number of terms in the multipole expansion, both resulting in higher overheads. Therefore, the original form of the fast multipole method is not very suitable for nonuniform distributions, or for dynamical systems which may develop large density contrasts over the course of time. To cope with this, adaptive fast multipole codes^{49–51} have been developed, which are basically hybrids between the BH and FM algorithms.

5 Performance and Parallelism

Comparison of N -body algorithms is a dangerous business because unless one is completely impartial, there is a tendency to neglect one's least favourite scheme. A classic example is the direct N -particle summation, which can be performed in a time $N(N - 1)/2$ by exploiting the ‘action-reaction’ symmetry of the Coulomb force-law. Overlooking this fact immediately makes the naive PP scheme a factor of 2 slower than it needs to be. Less trivial optimisations are also often neglected in the Ewald method for periodic systems, leading to wildly differing conclusions for the ‘crossover’ points at which alternative schemes – PME, FMM etc. – are faster.

5.1 Open Systems

Because periodicity necessitates additional complexity and overheads which differ for the Ewald and multipole schemes, we first consider the simpler case of open boundaries, where the system is surrounded by an infinite vacuum with no external forces present. Here we will compare the pure PP algorithm against the Barnes-Hut³⁵ tree algorithm (BH) and the standard Greengard FMM.³⁶ To make the comparison meaningful we perform tests for fixed accuracy, defined by the relative RMS particle-particle force error, estimated as follows:

$$\varepsilon_\alpha = \left\{ \frac{\sum_{i=1}^{N_{test}} (f_{\alpha i}^a - f_{\alpha i}^r)^2}{\sum_{i=1}^{N_{test}} (f_{\alpha i}^r)^2} \right\}^{\frac{1}{2}},$$

where $f_{\alpha i}^a$ and $f_{\alpha i}^r$ are the components of the forces on particle i , evaluated using the approximate and direct (reference) methods respectively. For homogeneous, symmetric

density distributions, we can average over the three force components:

$$\varepsilon_f = \frac{1}{3} \sum_{\alpha} \varepsilon_{\alpha}.$$

Note that this is a stronger measure than, say the relative error in potential energy

$$\varepsilon_{pot} = \left| \frac{\Phi^a - \Phi^r}{\Phi^r} \right|,$$

where individual errors can sometimes cancel each other. Generally, error estimates based on ε_f are larger than those found with ε_{pot} , and represent a more reliable and stringent measure for dynamical applications.^{52,53}

Benchmarks for these three algorithms performed on a SunBlade Sparc machine are shown in Fig. 24. The initial distribution used was a neutral, homogeneous sphere of randomly placed positive and negative charges. With the BH algorithm, setting the multipole acceptance parameter θ to 0.2 and 0.5 resulted in average force errors of 0.1% and 1% respectively, almost independent of N – see Fig. 19. To achieve equivalent precision with the FMM is not just a matter of choosing the appropriate number of multipoles, because the refinement level needs to be adjusted too as N is increased. Esselink⁵³ investigated this effect in some detail, and we quote adjusted timings from his FMM code for 7- and 4-term expansions. The above curves reveal two noticeable features. First, the Barnes-Hut

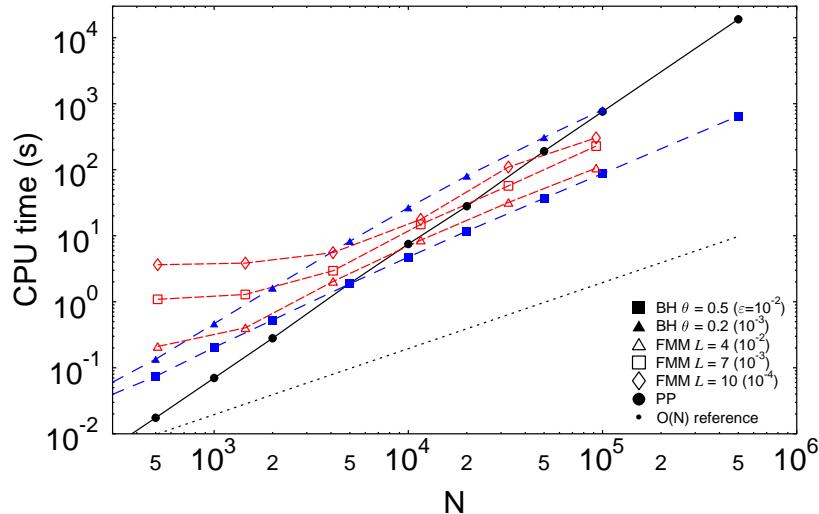


Figure 24. Comparison of computation time as a function of the number of simulation particles N between particle-particle, hierarchical tree, and the fast multipole code for a 3-dimensional open system. The FMM timings are taken (and adjusted) from Esselink (1995)

algorithm is clearly hard to beat for ‘low-precision’ applications (where a force error of 1%

can be tolerated), displaying a breakeven point over PP of around 5000 for the BH code compared to about 10000 for FMM. If high precision is needed however (for, say Monte-Carlo configurational computations), then FMM comes into its own for particle numbers in excess of 10^5 .

5.2 Periodic Systems

As we hinted at before, timing comparisons for the classic, periodic lattice of charges with which we began this article have proved to be the most controversial in the past, with claimed breakeven points ranging from 300 to 10^5 ! These discrepancies can usually be traced to differing levels of optimisation for both Ewald and FMM algorithms. Rather than attempt to implement a ‘perfect’ Ewald sum ourselves, we take timings from Esselink’s paper,⁵³ which describes several optimisations in some detail. The FMM timings, on the other hand, we draw from Schmidt & Lee’s paper,⁴⁴ where they also describe and implement a fully periodic version. The overhead caused by periodic boundaries is actually just a few percent if one uses the multipole expansion for the whole system for the periodic images. Figure 25 shows a comparison of the computation time required by the Ewald method, tree code, and fast multipole codes for a 3-dimensional, fully periodic system of charges. Extrapolating the FMM curves for these error levels (around 10^{-4} and 10^{-3} for the upper and lower curve respectively), one can estimate the breakeven point somewhere between 5×10^4 and 10^5 - a little higher than for open systems. This value seems to be consistent with a more rigorous comparison by Solvason *et al.* for 2D periodic systems.¹⁵

The other two curves shown are taken from a periodic tree code (Barnes-Hut-Ewald – BHE), developed by Pfalzner & Gibbon⁵⁴ for plasma physics applications. These appear to be comparable with timings reported for PME/P³M codes,^{30,28,26} though the tree code is perhaps less accurate in the form used. Another interesting periodic BH derivation has been presented by Duan & Krasny,⁵⁵ who also deduce a breakeven point of about 10^4 particles.

5.3 Parallelisation Strategies

The widespread availability of parallel supercomputers has made N -body calculation very attractive as a simulation tool, even bringing direct $O(N^2)$ force summation into the realms of feasibility for a restricted set of problems.⁵⁶ The direct N^2 algorithm contains a natural parallelism, requiring a simple sharing of particles between the processors. Communication overheads can be minimised by passing partial results between processors in a ring (systolic loop), as described elsewhere in these proceedings.⁶

The methods described in this review are all designed to yield a better *algorithmic scaling* than the brute force option, i.e. a speed-up relative to direct summation regardless of machine architecture. Having established a breakeven point for a given accuracy, it is important to know whether this will still hold on a massively parallel computer. In practice, we find that N -body algorithms have widely differing efficiencies when implemented in parallel, so we now briefly outline the main parallelisation strategies used to date.

The conventional Ewald method which we started with in Section 2 is quite easy to implement in parallel. Even with a modest number of processors, there is an immediate

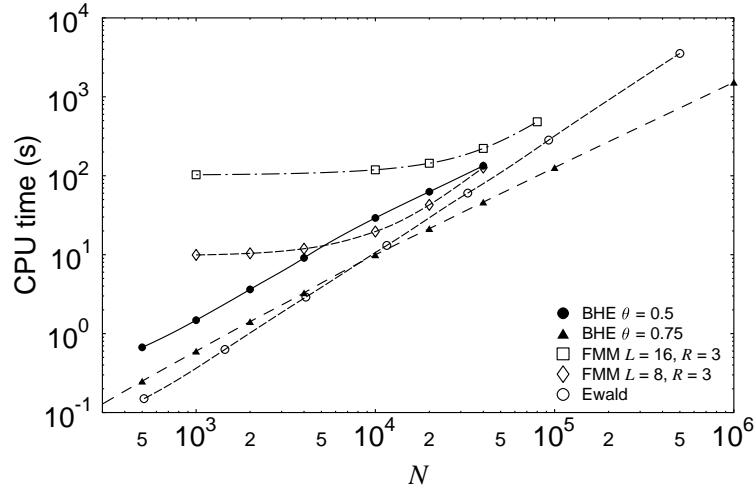


Figure 25. Comparison of computation time as a function of the number of simulation particles N between particle-particle, hierarchical tree (BHE), and the fast multipole code for a neutral system of charges with periodic boundary conditions. The FMM timings are derived from Schmidt and Lee (1991).

opportunity for task sharing provided by the splitting of the sum into real- and reciprocal-space, which yields two more-or-less equally time-consuming contributions. The real-space part can be split further according to the standard PP recipe; the k -space sum can be reformulated as a particle sum multiplied by structure factors which can be computed independently on each processor.⁵⁷

Unfortunately, parallelism is not so clear-cut for the PME or P³M methods: if anything it becomes less efficient, because the Ewald sum gets heavily and deliberately biased towards the now gridded Fourier component. On a scalar machine, the latter can be rapidly evaluated by FFTs, leading to the desired $O(N \log N)$ -scaling. However, the binary, recursive nature of the FFT is not well suited to distributed-memory parallelism, and ultimately results in poor scalability.⁵⁸ An alternative P³M scheme with very promising scaling characteristics has recently been suggested by Beckers *et al.*, in which the FFT is abandoned in favour of an iterative Poisson solver.⁵⁹ Parallel implementation of this scheme involves a standard spatial decomposition commonly found in fluid dynamics applications, in which a ‘halo’ of ghost cells a few grid points wide is placed around each processor domain. The halos act as communication buffers for grid information held by neighbouring processors needed for the local integration of Poisson’s equation.

Multipole-based N -body methods require somewhat more effort to implement on parallel architectures, but have nonetheless attracted considerable attention because of the potentially rewarding prospect of Giga-particle simulation which modern Teraflop machines would offer. A parallel version of the original, non-adaptive 2D FMM was proposed by Greengard himself as early as 1990.⁶⁰ This scheme, based on task-sharing in each of the separate near- and far-field stages of the FMM, works well on shared memory machines,

but less efficiently on distributed memory systems. Nonetheless, these ideas inspired the first parallel FMM for MD simulation of macromolecules, developed and demonstrated by the Duke University group in the early 1990s.^{50,61}

Of all the algorithms described here, the Barnes-Hut tree code probably presents the biggest challenge for parallel implementation. At first sight, the hierarchical data structure would seem to rule out parallelism altogether, but it was soon realised that both tree construction and interaction lists could be at least vectorised on a level-by-level basis,^{62–64} leaving a straightforward $N \times N_{list}$ force summation to contend with. However, all this only works with shared memory. On distributed memory machines, the tree structure either has to be global to all processors – a very expensive and wasteful option – or somehow divided up equally among them. The problem is that the access to the *whole* tree is in principle needed to build an interaction list. Searching for nonlocal nodes on remote processors using the pointer addressing schemes typical of scalar tree codes would entail a huge communication overhead. This problem was recognised by Salmon and Warren,^{65,66} who practically reinvented the BH algorithm by scrapping pointers in favour of a set of unique binary keys to represent particle and node coordinates. Domain decomposition is reduced to cutting a list of keys sorted into an appropriate order. The main drawback of this scheme is that memory locations are accessed by mapping the large number of possible keys onto a hash table; a process that risks ‘collisions’ – two or more keys yielding the same address. Various ways have been proposed to minimise this effect, including sorting addresses according to access frequency, and distributing ‘work’ rather than tree-nodes to remote processors.⁶⁷

The trend in machine architecture is towards ever larger processor arrays – currently approaching the 10000 mark,⁶⁸ which puts a premium on algorithm efficiency. Even a 1% inefficiency due to a non-parallel component or communication overhead can lead to severe performance deterioration for 1000 processors or more.⁶ Which of the above algorithms is best suited to large-scale computations on large machines is really an open question. The mesh- and multipole-based methods may have a mathematically superior scaling, but are harder to implement efficiently in parallel, and will thus continue to present a challenge as computers increase in size.

6 Summary

In this article we have attempted to provide a guide to the alternative methods available for accelerating the force or potential calculation for long-range N -body problems. Although it is impossible to give any strict recommendations, we can draw up some general rules-of-thumb as to which scheme to choose. For the special but important case of periodic systems, some form of P³M is widely acknowledged as being faster than the classical Ewald sum for reasonable accuracy (10^{-4} relative force error, say) – in the range $N = 10^4$ – 10^5 . Thereafter, it may be worth investing in FMM, particularly if very high precision is desired for one-shot configurational calculations, for example. The lesser known Barnes-Hut-Ewald schemes may also be competitive with either of these methods for all N , though this remains to be demonstrated conclusively. For open systems, multipole methods are the only alternative, and have become routine in astrophysical and plasma physics applications. For static, high-precision problems where $N > 10^5$, FMM is again hard to beat. On the other hand, dynamical applications (MD) call for force-accuracy comparable with the in-

tegration scheme (typically 0.1–1%), which favours the much simpler BH algorithm. This segregation of requirements does not rule out future hybrid, adaptive multipole schemes suitable for implementing on massively parallel architectures.

Acknowledgments

One of us (PG) acknowledges discussions on the intricacies of optimised FMM with H. Dachsel.

References

1. P. P. Ewald, *Die Berechnung optischer und elektrostatischer Gitterpotentiale*, Ann. Phys. **64**, 253 (1921).
2. C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulation*, (McGraw-Hill, New York, 1985).
3. R. L. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, (McGraw-Hill, New York, 1981).
4. T. Darden, D. York, and L. Pedersen, *Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems*, J. Chem. Phys. **98**, 10089–10092 (1993).
5. S. Pfalzner and P. Gibbon, *Many Body Tree Methods in Physics*, (Cambridge University Press, New York, 1996).
6. G. Sutmann, “Classical molecular dynamics”, in *Quantum Simulations of Complex Many-Body Systems: From Theory to Applications*, J. Grotendorst, Ed. (NIC, Jülich, 2002).
7. E. Madelung, *Das elektrische Feld in Systemen von regelmässig angeordneten Punktladungen*, Phys. Z. **19**, 524–532 (1918).
8. C. Kittel, *Introduction to Solid State Physics*, (Wiley & Sons, New York, 5th edition, 1976).
9. S. W. De Leeuw, J. W. Perram, and E. R. Smith, *Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants*, Proc. Roy. Soc. Lon. **373**, 27–56 (1980).
10. D. M. Heyes, *Electrostatic potentials and fields in infinite point charge lattices*, J. Chem. Phys. **74**, 1924–1929 (1981).
11. M. L. Boas, *Mathematical Methods in the Physical Sciences*, (Wiley & Sons, New York, 1st edition 1966).
12. R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* vol. 1, (Addison-Wesley, Reading, Mass., 1963).
13. J. W. Perram, H. G. Petersen, and S. W. D. Leeuw, *An algorithm for the simulation of condensed matter which grows as the $N^{3/2}$ power of the number of particles*, Mol. Phys. **65**, 875–893 (1988).
14. J. Kolafa and J. W. Perram, *Cutoff errors in the Ewald summation formulae for point charge systems*, Mol. Sim. **9**, 351–368 (1992).
15. D. Solvason, J. Kolafa, H. G. Petersen, and J. W. Perram, *A rigorous comparison of the Ewald method and the fast multipole method in two dimensions*, Comp. Phys. Commun. **87**, 307–318 (1995).

16. M. J. Sangster and M. Dixon, *Interionic potentials in alkali halides and their use in simulations of the molten salts*, Adv. in Phys. **25**, 247–342 (1976).
17. D. Fincham, *Optimisation of the Ewald sum for large systems*, Mol. Sim. **13**, 1–9 (1994).
18. G. Rajagopal and R. J. Needs, *An optimized Ewald method for long-ranged potentials*, J. Comp. Phys. **115**, 399–405 (1994).
19. O. Bunemann, *Dissipation of currents in ionized media*, Phys. Rev. **115**, 503–517 (1959).
20. J. M. Dawson, *Particle simulation of plasmas*, Rev. Mod. Phys. **55**, 403–447 (1983).
21. R. Dendy, Ed., *Plasma Physics: An Introductory Course*, (Cambridge University Press, Cambridge, 1993).
22. C. K. Birdsall and D. Fuss, *Clouds-in-clouds, clouds-in-cells physics for many-body plasma simulation*, J. Comp. Phys. **3**, 494–511 (1969).
23. J. M. Dawson, *One-dimensional plasma model*, Phys. Fluids **5**, 445–459 (1962).
24. J. Eastwood, R. W. Hockney, and D. N. Lawrence, *P3M3DP: The three-dimensional periodic particle-particle / particle-mesh program*, Comp. Phys. Commun. **19**, 215–261 (1980).
25. M. P. Allen and D. J. Tildesley, *Computer simulations of liquids*, (Oxford University Press, Oxford 1987).
26. E. L. Pollock and J. Glosli, *Comments on P^3M , FMM, and the Ewald method for large periodic Coulombic systems*, Comp. Phys. Commun. **95**, 93–110 (1996).
27. H. Furukawa and K. Nishihara, *Reduction in bremsstrahlung emission from host, dense binary-ionic-mixture plasmas*, Phys. Rev. A **42**, 3532–3543 (1990).
28. H. G. Petersen, *Accuracy and efficiency of the particle mesh Ewald method*, J. Chem. Phys. **103**, 3668–3679 (1995).
29. B. A. Luty, I. G. Tironi, and W. F. van Gunsteren, *Lattice-sum methods for calculating electrostatic interactions in molecular simulations*, J. Chem. Phys. **103**, 3014–3021 (1995).
30. B. A. Luty, M. E. Davis, I. G. Tironi, and W. F. van Gunsteren, *A comparison of Particle-Particle Particle-Mesh and Ewald methods for calculating electrostatic interactions in periodic molecular systems*, Mol. Sim. **14**, 11–20 (1994).
31. T. Schlick, R. D. Skeel, A. T. Brunger, L. V. Kale, J. A. Board Jr., J. Hermans, and K. Schulten, *Algorithmic challenges in computational molecular biophysics*, J. Comp. Phys. **151**, 9–48 (1999).
32. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *A smooth particle mesh Ewald method*, J. Chem. Phys. **103**, 8577–8593 (1995).
33. D. M. York, T. A. Darden, and L. G. Pedersen, *The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the Ewald and truncated list methods*, J. Chem. Phys. **99**(10), 8345–8348 (1993).
34. A. Appel, *An efficient program for many-body simulation*, SIAM J. Sci. Statist. Comput. **6**, 85 (1985).
35. J. Barnes and P. Hut, *A hierarchical $O(N \log N)$ force-calculation algorithm*, Nature **324**, 446–449 (1986).
36. L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*, J. Comp. Phys. **73**, 325–348 (1987).
37. W. H. Press, ”, in *The use of supercomputers in stellar dynamics*, P. Hut and S. L. W.

McMillan, Eds. (Springer, New York 1986), p. 184.

38. W. Benz, *Applications of smooth particle hydrodynamics (SPH) to astrophysical problems*, Comp. Phys. Commun. **48**, 97–105 (1988).
39. L. Hernquist, *Hierarchical N-body methods*, Comp. Phys. Commun. **48**, 107–115 (1988).
40. J. K. Salmon and M. S. Warren, *Skeletons from the treecode closet*, J. Comp. Phys. **111**, 136–155 (1994).
41. L. Greengard, *The rapid evaluation of potential fields in particle systems*, MIT Press Cambridge, Mass. (1988).
42. L. Greengard, *The numerical solution of the N-body problem*, Computers in Physics pp. 142–152 Mar./Apr. (1990).
43. J. Ambrosiano, L. Greengard, and V. Rokhlin, *The fast multipole method for gridless particle simulation*, Comp. Phys. Commun. **48**, 117–125 (1988).
44. K. E. Schmidt and M. A. Lee, *Implementing the fast multipole method in three dimensions*, J. Stat. Phys. **63**, 1223–1235 (1991).
45. J. D. Jackson, *Classical Electrodynamics*, (Wiley, New York, 2nd edition, 1975).
46. H. G. Peterson, D. Soelvason, J. W. Perram, and E. R. Smith, *The very fast multipole method*, J. Chem. Phys. **101**, 8870–8876 (1994).
47. C. A. White and M. Head-Gordon, *Derivation and efficient implementation of the fast multipole method*, J. Chem. Phys. **101**, 6593–6605 (1994).
48. H. Dachsel, private communication. See also Poster by same author at the NIC Winter School (2002).
49. J. Carrier, L. Greengard, and V. Rokhlin, *A fast adaptive multipole algorithm for particle simulations*, SIAM J. Sci. Stat. Comput. **9**, 669–686 (1988).
50. J. A. Board Jr., J. W. Causey, J. F. Leathrum Jr., A. Windemuth, and K. Schulten, *Accelerated molecular dynamics simulation with the parallel fast multipole algorithm*, Chem. Phys. Lett. **198**, 89–94 (1992).
51. H. Cheng, L. Greengard, and V. Rokhlin, *A fast adaptive multipole algorithm in three dimensions*, J. Comp. Phys. **155**, 468–498 (1999).
52. L. Hernquist, *Performance characteristics of tree codes*, Astrophys. J. Supp. **64**, 715–734 (1987).
53. K. Esselink, *A comparison of algorithms for long-range interactions*, Comp. Phys. Commun. **87**, 375–395 (1995).
54. S. Pfalzner and P. Gibbon, *A hierarchical tree code for dense plasma simulation*, Comp. Phys. Commun. **79**, 24–38 (1994).
55. Z.-H. Duan and R. Krasny, *An Ewald summation based multipole method*, J. Chem. Phys. **113**, 3492–3495 (2000).
56. R. Spurzem, *Direct N-body simulations*, J. Comp. Appl. Math. **109**, 407–432 (1999).
57. R. K. Kalia, S. de Leeuw, A. Nakano and P. Vashishta, *Molecular dynamics simulations of Coulombic systems on distributed-memory MIMD machines*, Comp. Phys. Commun. **74**, 316–326 (1993).
58. A. Gupta and V. Kumar, *The scalability of FFT on parallel computers*, IEEE Trans. Parallel Dist. Systems **4**, 922–932 (1993).
59. J. V. L. Beckers, C. P. Lowe, and S. W. de Leeuw, *An iterative PPPM method for simulating Coulombic systems on distributed memory parallel computers*, Mol. Sim. **20**, 369–383 (1998).

60. L. Greengard and W. D. Groop, *A parallel version of the fast multipole method*, Comput. Math. Applic. **20**, 63–71 (1990).
61. “Distributed parallel multipole tree algorithm”, Duke University Technical Report (2000), <http://www.ee.duke.edu/research/SciComp/Docs/>
62. J. E. Barnes, *A modified tree code: don’t laugh; it runs*, J. Comp. Phys. **87**, 161–170 (1990).
63. J. Makino, *Vectorization of a treecode*, J. Comp. Phys. **87**, 148–160 (1990).
64. L. Hernquist, *Vectorization of tree traversals*, J. Comp. Phys. **87**, 137–147 (1990).
65. M. S. Warren and J. K. Salmon, “A parallel hashed oct-tree n -body algorithm”, in *Supercomputing ’93 Los Alamitos* (1993) IEEE Comp. Soc. pp. 12–21.
66. M. S. Warren and J. K. Salmon, *A portable parallel particle program*, Comp. Phys. Commun. **87**(266–290) (1995).
67. A. Gramma, V. Kumar, and A. Sameh, *Scalable parallel formulations of the Barnes-Hut method for N-body simulations*, Parallel Comp. **24**, 797–822 (1998).
68. Top 500 computer sites, <http://www.top500.org/>

Parallel Programming Models, Tools and Performance Analysis

Bernd Mohr¹ and Michael Gerndt²

¹ John von Neumann Institute for Computing
Central Institute for Applied Mathematics
Research Centre Jülich, 52425 Jülich, Germany
E-mail: b.mohr@fz-juelich.de

² Technische Universität München
Institut für Informatik, LRR
80290 München, Germany
E-mail: gerndt@in.tum.de

The major parallel programming models for scalable parallel architectures are the message passing model and the shared memory model. This article outlines the main concepts of these models as well as the industry standard programming interfaces MPI and OpenMP. To exploit the potential performance of parallel computers, programs need to be carefully designed and tuned. We will discuss design decisions for good performance as well as programming tools that help the programmer in program tuning.

1 Introduction

Although the performance of sequential computers increases incredibly fast, it is insufficient for a large number of challenging applications. Applications requiring much more performance are numerical simulations in industry and research as well as commercial applications such as query processing, data mining, and multi-media applications. Architectures offering high performance do not only exploit parallelism on a very fine grain within a single processor but apply a medium to large number of processors concurrently to a single application. High-end parallel computers deliver up to 30 Teraflop/s (10^{12} floating point operations per second) and are developed and exploited within the ASCI (Accelerated Strategic Computing Initiative) program of the Department of Energy in the USA.

This article concentrates on programming numerical applications on distributed memory computers introduced in Sec. 1.1. Parallelization of those applications centers around selecting a decomposition of the data domain onto the processors such that the workload is well balanced and the communication between processors is reduced (Sec. 1.2).⁷

The parallel implementation is then based on either the message passing or the shared memory model (Sec. 2). The standard programming interface for the message passing model is MPI (Message Passing Interface),^{14,11} offering a complete set of communication routines (Sec. 2.1). OpenMP^{4,13} is the standard for directive-based shared memory programming and will be introduced in Sec. 2.2.

Since parallel programs exploit multiple threads of control, debugging is even more complicated than for sequential programs. Sec. 3 outlines the main concepts of parallel debuggers and presents TotalView,¹⁵ the most widely available debugger for parallel programs.

Although the domain decomposition is key to good performance on parallel architectures, program efficiency also heavily depends on the implementation of the communication and synchronization required by the parallel algorithm and the implementation techniques chosen for sequential kernels. Optimizing those aspects is very system dependent and thus, an interactive tuning process consisting of measuring performance data and applying optimizations follows the initial coding of the application. The tuning process is supported by programming model specific performance analysis tools. Sec. 4 presents basic performance analysis techniques and introduces the widely available performance analysis tools VAMPIR¹⁶ (for MPI programs) and GuideView⁹ (for OpenMP).

1.1 Parallel Architectures

Parallel computers that scale beyond a small number of processors circumvent the main memory bottleneck by distributing the memory among the processors. Current architectures³ are composed of single-processor nodes with local memory or of multiprocessor nodes where each node's main memory is shared among its processors. The latter are often called SMP (Symmetrical Multi Processor) nodes.

The most important characteristic of this *distributed memory architecture* is that access to the local memory is faster than to remote memory. It is the challenge for the programmer to assign data to the processors such that most of the data accessed during the computation are already in the node's local memory.

Three major classes of distributed memory computers can be distinguished:

No Remote Memory Access (NORMA) computers do not have any special hardware support to access another node's local memory. Processors obtain data from remote memory only by exchanging messages between processes on the requesting and the supplying node.

Remote Memory Access (RMA) computers allow to access remote memory via specialized operations implemented by hardware. The accessed memory location is not determined via an address in a shared linear address space but via a tuple consisting of the processor number and the local address in the target processor's address space.

Cache-Coherent Non Uniform Memory Access (ccNUMA) computers do have a shared physical address space. All memory locations can be accessed via usual load and store operations. Access to a remote location results in a copy of the appropriate cache line in the processor's cache. Coherence algorithms ensure that multiple copies of a cache line are kept coherent, i.e., the copies do have the same value.

While most of the early parallel computers were NORMA systems, today's systems are either RMA or ccNUMA computers. This is because remote memory access is a light-weight communication protocol that is more efficient than standard message passing since data copying and process synchronization are eliminated. In addition, ccNUMA systems offer the abstraction of a shared linear address space resembling physically shared memory systems. This abstraction simplifies the task of program development but does not necessarily facilitate program tuning.

Typical examples of the three classes are clusters of workstations (NORMA), CRAY T3E (RMA), and SGI Origin 3000 (ccNUMA).

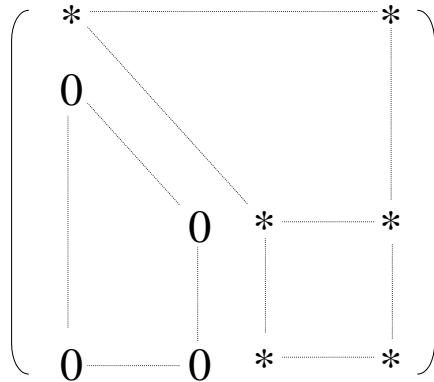


Figure 1. Structure of the matrix during Gaussian elimination.

1.2 Data Parallel Programming

Applications that scale to a large number of processors usually perform computations on large data domains. For example, crash simulations are based on partial differential equations that are solved on a large finite element grid and molecular dynamics applications simulate the behavior of a large number of particles. Other parallel applications apply linear algebra operations to large vectors and matrices. The elemental operations on each object in the data domain can be executed in parallel by the available processors.

The scheduling of operations to processors is determined according to a selected *domain decomposition*.⁸ Processors execute those operations that determine new values for local elements (owner-computes rule). While processors execute an operation, they may need values from other processors. The domain decomposition has thus to be chosen so that the distribution of operations is balanced and the communication is minimized. The third goal is to optimize single node computation, i.e., to be able to exploit the processor's pipelines and the processor's caches efficiently.

A good example for the design decisions taken when selecting a domain decomposition is Gaussian elimination.² The main structure of the matrix during the iterations of the algorithm is outlined in Fig. 1.

The goal of this algorithm is to eliminate all entries in the matrix below the main diagonal. It starts at the top diagonal element and subtracts multiples of the first row from the second and subsequent rows to end up with zeros in the first column. This operation is repeated for all the rows. In later stages of the algorithm the actual computations have to be done on rectangular sections of decreasing size. If the main diagonal element of the current row is zero, a pivot operation has to be performed. The subsequent row with the maximum value in this column is selected and exchanged with the current row.

A possible distribution of the matrix is to decompose its columns into blocks, one block for each processor. The elimination of the entries in the lower triangle can then be performed in parallel where each processor computes new values for its columns only. The main disadvantage of this distribution is that in later computations of the algorithms only a subgroup of the processors is actually doing any useful work since the computed rectangle is getting smaller.

To improve load balancing, a cyclic column distribution can be applied. The computations in each step of the algorithm executed by the processors differ only in one column.

In addition to load balancing also communication needs to be minimized. Communication occurs in this algorithm for broadcasting the current column to all the processors since it is needed to compute the multiplication factor for the row. If the domain decomposition is a row distribution, which eliminates the need to communicate the current column, the current row needs to be broadcast to the other processors.

If we consider also the pivot operation, communication is necessary to select the best row when a row-wise distribution is applied since the computation of the global maximum in that column requires a comparison of all values.

Selecting the best domain decomposition is further complicated due to optimizing single node performance. In this example, it is advantageous to apply BLAS3 operations for the local computations. These operations make use of blocks of rows to improve cache utilization. Blocks of rows can only be obtained if a block-cyclic distribution is applied, i.e., columns are not distributed individually but blocks of columns are cyclically distributed.

This discussion makes clear, that choosing a domain decomposition is a very complicated step in program development. It requires deep knowledge of the algorithm's data access patterns as well as the ability to predict the resulting communication.

2 Programming Models

The two main programming models, *message passing* and *shared memory*, offer different features for implementing applications parallelized by domain decomposition.

The message passing model is based on a set of processes with private data structures. Processes communicate by exchanging messages with special send and receive operations. The domain decomposition is implemented by developing a code describing the local computations and local data structures of a single process. Thus, global arrays have to be split up and only the local part has to be allocated in a process. This handling of global data structures is called *data distribution*. Computations on the global arrays also have to be transformed, e.g., by adapting the loop bounds, to ensure that only local array elements are computed. Access to remote elements have to be implemented via explicit communication, temporary variables have to be allocated, messages have to be constructed and transmitted to the target process.

The shared memory model is based on a set of threads that are created when parallel operations are executed. This type of computation is also called *fork-join parallelism*. Threads share a global address space and thus access array elements via a global index.

The main parallel operations are *parallel loops* and *parallel sections*. Parallel loops are executed by a set of threads also called a *team*. The iterations are distributed among the threads according to a predefined strategy. This scheduling strategy implements the chosen domain decomposition. Parallel sections are also executed by a team of threads but the tasks assigned to the threads implement different operations. This feature can for example be applied if domain decomposition itself does not generate enough parallelism and whole operations can be executed in parallel since they access different data structures.

In the shared memory model, the distribution of data structures onto the node memories is not enforced by decomposing global arrays into local arrays, but the global address space is distributed onto the memories by the operating system. For example, the pages

of the virtual address space can be distributed cyclically or can be assigned at first touch. The chosen domain decomposition thus has to take into account the granularity of the distribution, i.e., the size of pages, as well as the system-dependent allocation strategy.

While the domain decomposition has to be hard-coded into the message passing program, it can easily be changed in a shared memory program by selecting a different scheduling strategy for parallel loops.

Another advantage of the shared memory model is that automatic and incremental parallelization is supported. While automatic parallelization leads to a first working parallel program, its efficiency typically needs to be improved. The reason for this is that parallelization techniques work on a loop-by-loop basis and do not globally optimize the parallel code via a domain decomposition. In addition, dependence analysis, the prerequisite for automatic parallelization, is limited to access patterns known at compile time.

In the shared memory model, a first parallel version is relatively easy to implement and can be incrementally tuned. In the message passing model instead, the program can be tested only after finishing the full implementation. Subsequent tuning by adapting the domain decomposition is usually time consuming.

2.1 MPI

The Message Passing Interface (MPI)^{14,11} was developed between 1993 and 1997. It includes routines for point-to-point communication, collective communication, one-sided communication, and parallel IO. While the basic communication primitives have already been defined since 1994 and implemented on almost all parallel computers, remote memory access and parallel IO routines are part of MPI 2.0 and are only available on few machines.

2.1.1 MPI Basic Routines

MPI consists of more than 120 functions. But realistic programs can already be developed based on no more than six functions:

MPI_Init initializes the library. It has to be called at the beginning of a parallel operation before any other MPI routines are executed.

MPI_Finalize frees any resources used by the library and has to be called at the end of the program.

MPI_Comm_size determines the number of processors executing the parallel program.

MPI_Comm_rank returns the unique process identifier.

MPI_Send transfers a message to a target process. This operation is a blocking send operation, i.e., it terminates when the message buffer can be reused either because the message was copied to a system buffer by the library or because the message was delivered to the target process.

MPI_Recv receives a message. This routine terminates if a message was copied into the receive buffer.

2.1.2 MPI Communicator

All communication routines depend on the concept of a *communicator*. A communicator consists of a process group and a communication context. The processes in the process

group are numbered from zero to process count - 1. The process number returned by `MPI_Comm_rank` is the identification in the process group of the communicator which is passed as a parameter to this routine.

The communication context of the communicator is important in identifying messages. Each message has an integer number called a *tag* which has to match a given selector in the corresponding receive operation. The selector depends on the communicator and thus on the communication context. It selects only messages with a fitting tag and having been sent relative to the same communicator. This feature is very useful in building parallel libraries since messages sent inside the library will not interfere with messages outside if a special communicator is used in the library. The default communicator that includes all processes of the application is `MPI_COMM_WORLD`.

2.1.3 MPI Collective Operations

Another important class of operations are *collective operations*. Collective operations are executed by a process group identified via a communicator. All the processes in the group have to perform the same operation. Typical examples for such operations are:

`MPI_Barrier` synchronizes all processes. None of the processes can proceed beyond the barrier until all the processes started execution of that routine.

`MPI_Bcast` allows to distribute the same data from one process, the so-called *root* process, to all other processes in the process group.

`MPI_Scatter` also distributes data from a root process to a whole process group, but each receiving process gets different data.

`MPI_Gather` collects data from a group of processes at a root process.

`MPI_Reduce` performs a global operation on the data of each process in the process group. For example, the sum of all values of a distributed array can be computed by first summing up all local values in each process and then summing up the local sums to get a global sum. The latter step can be performed by the reduction operation with the parameter `MPI_SUM`. The result is delivered to a single target processor.

2.1.4 MPI IO

Data parallel applications make use of the IO subsystem to read and write big data sets. These data sets result from replicated or distributed arrays. The reasons for IO are to read input data, to pass information to other programs, e.g., for visualization, or to store the state of the computation to be able to restart the computation in case of a system failure or if the computation has to be split into multiple runs due to its resource requirements.

IO can be implemented in three ways:

1. Sequential IO

A single node is responsible to perform the IO. It gathers information from the other nodes and writes it to disk or reads information from disk and scatters it to the appropriate nodes. While the IO is sequential and thus need not be parallelized, the full performance of the IO subsystem might not be utilized. Modern systems provide high performance IO subsystems that are fast enough to support multiple IO requests from different nodes in parallel.

2. Private IO

Each node accesses its own files. The big advantage of this implementation is that no synchronization among the nodes is required and very high performance can be obtained. The major disadvantage is that the user has to handle a large number of files. For input the original data set has to be splitted according to the distribution of the data structure and for output the process-specific files have to be merged into a global file for postprocessing.

3. Parallel IO

In this implementation all the processes access the same file. They read and write only those parts of the file with relevant data. The main advantages are that no individual files need to be handled and that reasonable performance can be reached. The disadvantage is that it is difficult to reach the same performance as with private IO. The parallel IO interface of MPI provides flexible and high-level means to implement applications with parallel IO.

Files accessed via MPI IO routines have to be opened and closed by collective operations. The open routine allows to specify hints to optimize the performance such as whether the application might profit from combining small IO requests from different nodes, what size is recommended for the combined request, and how many nodes should be engaged in merging the requests.

The central concept in accessing the files is the *view*. A view is defined for each process and specifies a sequence of data elements to be ignored and data elements to be read or written by the process. When reading or writing a distributed array the local information can be described easily as such a repeating pattern. The IO operations read and write a number of data elements on the basis of the defined view, i.e., they access the local information only. Since the views are defined via runtime routines prior to the access, the information can be exploited in the library to optimize IO.

MPI IO provides blocking as well as nonblocking operations. In contrast to blocking operations, the nonblocking ones only start IO and terminate immediately. If the program depends on the successful completion of the IO it has to check it via a test function. Besides the collective IO routines which allow to combine individual requests, also non-collective routines are available to access shared files.

2.1.5 MPI Remote Memory Access

Remote memory access (RMA) operations (also called *1-sided communication*) allow to access the address space of other processes without participation of the other process. The implementation of this concept can either be in hardware, such as in the CRAY T3E, or in software via additional threads waiting for requests. The advantages of these operations are that the protocol overhead is much lower than for normal send and receive operations and that no polling or global communication is required for setting up communication.

In contrast to explicit message passing where synchronization happens implicitly, accesses via RMA operations need to be protected by explicit synchronization operations.

RMA communication in MPI is based on the *window concept*. Each process has to execute a collective routine that defines a window, i.e., the part of its address space that can be accessed by other processes.

The actual access is performed via *put* and *get operations*. The address is defined by the target process number and the displacement relative to the starting address of the window for that process.

MPI also provides special synchronization operations relative to a window. The MPI_Win_fence operation synchronizes all processes that make some address ranges accessible to other processes. It is a collective operation that ensures that all RMA operations started before the fence operation terminate before the target process executes the fence operation and that all RMA operations of a process executed after the fence operation are executed after the target process executed the fence operation.

2.2 OpenMP

OpenMP^{4,13} is a directive-based programming interface for the shared memory programming model. It consists of a set of directives and runtime routines for Fortran 77 (published 1997), for Fortran 90 (2000), and a corresponding set of pragmas for C and C++ (1998).

Directives are special comments that are interpreted by the compiler. Directives have the advantage that the code is still a sequential code that can be executed on sequential machines and thus no two versions, a sequential and a parallel version, need to be maintained.

Directives start and terminate parallel regions. When the master thread hits a parallel region a team of threads is created or activated. The threads execute the code in parallel and are synchronized at the beginning and the end of the computation. After the final synchronization the master thread continues sequential execution after the parallel region. The main directives are:

!\$OMP PARALLEL DO specifies a loop that can be executed in parallel. The DO loop's iterations can be distributed in various ways including STATIC(CHUNK), DYNAMIC(CHUNK), and GUIDED(CHUNK) among the set of threads. STATIC(CHUNK) distribution means that the set of iterations are consecutively distributed among the threads in blocks of CHUNK size (resulting in block and cyclic distributions). DYNAMIC(CHUNK) distribution implies that iterations are distributed in blocks of CHUNK size to threads on a first-come-first-served basis. GUIDED (CHUNK) means that blocks of exponentially decreasing size are assigned on a first-come-first-served basis. The size of the smallest block is determined by CHUNK size.

!\$OMP PARALLEL SECTIONS starts a set of sections that are executed in parallel by a team of threads.

!\$OMP PARALLEL introduces a code region that is executed redundantly by the threads. It has to be used very carefully since assignments to global variables will lead to conflicts among the threads and possibly to nondeterministic behavior.

!\$OMP DO is a work sharing construct and may be used within a parallel region. All the threads executing the parallel region have to cooperate in the execution of the parallel loop. There is no implicit synchronization at the beginning of the loop but a synchronization at the end. After the final synchronization all threads continue after the loop in the replicated execution of the program code.

The main advantage of this approach is that the overhead for starting up the threads is eliminated. The team of threads exists during the execution of the parallel region and need not be built before each parallel loop.

!\$OMP SECTIONS is also a work sharing construct that allows the current team of threads executing the surrounding parallel region to cooperate in the execution of the parallel sections.

Program data can either be shared or private. While threads do have their own copy of private data, only one copy exists of shared data. This copy can be accessed by all threads. To ensure program correctness, OpenMP provides special synchronization constructs. The main constructs are *barrier synchronization* enforcing that all threads have reached this synchronization operation before execution continues and *critical sections*. Critical sections ensure that only a single thread can enter the section and thus, data accesses in such a section are protected from race conditions. For example, a common situation for a critical section is the accumulation of values. Since an accumulation consists of a read and a write operation unexpected results can occur if both operations are not surrounded by a critical section.

3 Parallel Debugging

Debugging parallel programs is more difficult than debugging sequential programs not only since multiple processes or threads need to be taken into account but also because program behavior might not be deterministic and might not be reproducible. These problems are not solved by current state-of-the-art commercial parallel debuggers. They deal only with the first problem by providing menus, displays, and commands that allow to inspect individual processes and execute commands on individual or all processes.

The widely used debugger is TotalView from Etnus Inc.¹⁵ It provides breakpoint definition, single stepping, and variable inspection via an interactive interface. The programmer can execute those operations for individual processes and groups of processes. TotalView also provides some means to summarize information such that equal information from multiple processes is combined into a single information and not repeated redundantly. It also supports MPI and OpenMP programs on many platforms.

4 Parallel Performance Analysis

Performance analysis is an iterative subtask during program development. The goal is to identify program regions that do not perform well. Performance analysis is structured into four phases:

1. Measurement

Performance analysis is done based on information on runtime events gathered during program execution. The basic events are, for example, cache misses, termination of a floating point operation, start and stop of a subroutine or message passing operation. The information on individual events can be summarized during program execution or individual trace records can be collected for each event.

Summary information has the advantage to be of moderate size while trace information tends to be very large. The disadvantage is that it is not fine grained; the behavior of individual instances of subroutines can for example not be investigated since all the information has been summed up.

2. Analysis

During analysis the collected runtime data are inspected to detect *performance problems*. Performance problems are based on *performance properties*, such as the existence of message passing in a program region, which have a condition for identifying it and a severity function that specifies its importance for program performance.

Current tools support the user in checking the conditions and the severity by visualizing program behavior. Future tools might be able to automatically detect performance properties based on a specification of possible properties. During analysis the programmer applies a threshold. Only performance properties whose severity exceeds this threshold are considered to be performance problems.

3. Ranking

During program analysis the severest performance problems need to be identified. This means that the problems need to be ranked according to the severity. The most severe problem is called the *program bottleneck*. This is the problem the programmer tries to resolve by applying appropriate program transformations.

4. Refinement

The performance problems detected in the previous phases might not be precise enough to allow the user to start optimization. At the beginning of performance analysis, summary data can be used to identify critical regions. The summary data might not be sufficient to identify why, for example, a region has high message passing overhead. The reason, e.g., very big messages or load imbalance, can be identified only with more detailed information. Therefore the performance problem should be refined into hypotheses about the real reason and additional information be collected in the next performance analysis cycle.

Current techniques for performance data collection are *profiling* and *tracing*. Profiling collects summary data only. This can be done via *sampling*. The program is regularly interrupted, e.g., every 10 ms, and the information is added up for the source code location which was executed in this moment. For example, the UNIX profiling tool *prof* applies this technique to determine the fraction of the execution time spent in individual subroutines.

A more precise profiling technique is based on *instrumentation*, i.e., special calls to a *monitoring library* are inserted into the program. This can either be done in the source code by the compiler or specialized tools, or can be done in the object code. While the first approach allows to instrument more types of regions, for example, loops and vector statements, the latter allows to measure data for programs where no source code is available. The monitoring library collects the information and adds it to special counters for the specific region.

Tracing is a technique that collects information for each event. This results, for example, in very detailed information for each instance of a subroutine and for each message sent to another process. The information is stored in specialized trace records for each event type. For example, for each start of a send operation, the time stamp, the message size and the target process can be recorded, while for the end of the operation, the time stamp and bandwidth are stored.

The trace records are stored in the memory of each process and are written to a trace file either when the buffer is filled up or when the program terminates. The individual trace files of the processes are merged together into one trace file ordered according to the time stamps of the events.

The following sections describe two widely available performance analysis tools for MPI programs (VAMPIR) and OpenMP applications (GuideView).

4.1 VAMPIR

VAMPIR (Visualization and Analysis of MPI Resources) is an event trace analysis tool^{12, 16} initially developed by the Central Institute for Applied Mathematics of the Research Centre Jülich and now is commercially distributed by the German company PALLAS. VAMPIR has three components:

- The VAMPIR tool itself is a graphical event trace browser implemented for the X11 Window system using the Motif toolkit. It is available for all major UNIX platforms.
- The VAMPIR runtime library (VampirTrace) provides an API for collecting, buffering, and generating event traces as well as a set of wrapper routines for MPI and shmem communication routines which record message traffic in the event trace.
- In order to observe functions or subroutines in the user program, their entry and exit has to be instrumented by inserting calls to the VAMPIR runtime library. Observing message passing functions is handled by linking the program with the VAMPIR wrapper function library.

VAMPIR comes with a source instrumenter for ANSI Fortran 77. Programs written in other programming languages (e.g., C or C++) have to be instrumented manually.

During the execution of the instrumented user program, the VAMPIR runtime library records entry and exits to instrumented user and message passing functions and the sending and receiving of messages. For each message, its tag, communicator, and length is recorded. Through the use of a configuration file, it is possible to switch the runtime observation of specific functions on and off. This way, the program doesn't have to be re-instrumented and re-compiled for every change in the instrumentation.

Large parallel programs consist of dozens or even hundreds of functions. To ease the analysis of such complex programs, VAMPIR arranges the functions into groups, e.g., user functions, MPI routines, I/O routines, and so on. The user can control/change the assignment of functions to groups and can also define new groups.

VAMPIR provides a wide variety of graphical displays to analyze the recorded event traces:

- The dynamic behavior of the program can be analyzed by timeline diagrams for either the whole program or a selected set of nodes. By default, the displays show the whole event trace, but the user can zoom-in to any arbitrary region of the trace. Also, the user can change the display style of the lines representing messages based on their tag/communicator or the length. This way, message traffic of different modules or libraries can easily be visually separated.

- The parallelism display shows the number of nodes in each function group over time. This allows to easily locate specific parts of the program, e.g., parts with heavy message traffic or IO.
- VAMPIR also provides a large number of statistical displays. It calculates how often each function or group of functions was called and the time spent in them. Message statistics show the number of messages sent, and the minimum, maximum, sum, and average length or transfer rate between any two nodes. The statistics can be displayed as barcharts, histograms, or textual tables.

A very useful feature of VAMPIR is that the statistic displays can be linked to the timeline diagrams. By this, statistics can be calculated for any arbitrary, user selectable part of the program execution.

- If the instrumenter/runtime library provides the necessary information in the event trace header, the information provided by VAMPIR can be related back to the source code. VAMPIR provides a source code and a call graph display to show selected functions or the location of the send and the receive of a selected message.

In summary, VAMPIR is a very powerful and highly configurable event trace browser. It displays trace files in a variety of graphical views, and provides flexible filter and statistical operations that condense the displayed information to a manageable amount. Rapid zooming and instantaneous redraw allow to identify and focus on the time interval of interest.

4.2 GuideView

GuideView⁹ is the integrated profiling performance analysis component of the OpenMP Compilation Environment KAP/Pro of KAI. It can be used to look for typical OpenMP performance problems like load imbalance, false sharing, or excessive synchronization.

The necessary instrumentation for performance data collection is automatically inserted on user request by the Guide OpenMP compiler. During program execution, the Guide runtime system collects execution statistics for each OpenMP construct in each thread. Execution time is measured and categorized into user code execution in sequential and parallel mode, sequential and parallel overhead, time spent in lock functions and barriers, as well as load imbalance at barriers. Afterwards, the collected performance data can be analyzed with the GuideView tool. It provides performance visualizations for the whole program, on a per thread basis, and on a per OpenMP region basis.

In addition, performance data files from different program runs can be loaded and analyzed simultaneously. This allows to compare the program performance based on different input datasets and/or thread numbers.

5 Summary

This article gave an overview of parallel programming models as well as programming tools. Parallel programming will always be a challenge for programmers. Higher-level programming models and appropriate programming tools only facilitate the process but do not make it a simple task.

While programming in MPI offers the greatest potential performance, shared memory programming with OpenMP is much more comfortable due to the global style of the resulting program. The sequential control flow among the parallel loops and regions matches much better with the sequential programming model all the programmers are trained for.

Although program tools were developed over years, the current situation seems not to be very satisfying. Program debugging is done per thread, a technique that does not scale to larger numbers of processors. Performance analysis tools do also suffer scalability limitations and, in addition, the tools are complicated to use. The programmers have to be experts for performance analysis to understand potential performance problems, their proof conditions, and their severity. In addition they have to be experts for powerful but also complex user interfaces.

Future research in this area has to try to automate performance analysis tools, such that frequently occurring performance problems can be identified automatically. It is the goal of the IST working group APART on *Automatic Performance Analysis: Resources and Tools* to investigate base technologies for future more intelligent tools.¹ An important result of this work is a collection of performance problems for parallel programs that have been formalized with the ASL, the *APART Specification Language*.⁶ This approach will lead to a formal representation of the knowledge applied in the manually executed performance analysis process and thus will make this knowledge accessible for automatic processing. First automatic tools are already available: ParaDyn¹⁰ from the University of Wisconsin-Madison, Kappa-PI⁵ from the Universitat Autònoma de Barcelona, and EXPERT^{17,18} from the Research Centre Jülich.

A second important trend that will effect parallel programming in the future is the move towards clustered shared memory systems. Clearly, a hybrid programming approach will be applied on those systems for best performance, combining message passing between the individual SMP nodes and shared memory programming in a node. This programming model will lead to even more complex programs and program development tools have to be enhanced to be able to help the user in developing these codes.

References

1. APART: *IST Working Group on Automatic Performance Analysis Resources and Tools*, <http://www.fz-juelich.de/apart/>, 2001.
2. D. P. Bertsekas, J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, ISBN 0-13-648759-9, 1989.
3. D. E. Culler, J. P. Singh, A. Gupta, *Parallel Computer Architecture - A Hardware/Software Approach*, Morgan Kaufmann Publishers, ISBN 1-55860-343-3, 1999.
4. L. Dagum, R. Menon, *OpenMP: An Industry-Standard API for Shared-memory Programming*, IEEE Computational Science & Engineering, Vol. 5, No. 1, 46–55, 1998.
5. A. Espinosa, *Automatic Performance Analysis of Parallel Programs*, PhD thesis, Universitat Autònoma de Barcelona, 2000.
6. Th. Fahringer, M. Gerndt, B. Mohr, F. Wolf, G. Riley, J. Träff, *Knowledge Specification for Automatic Performance Analysis*, APART Technical Report, Research Centre Juelich FZJ-ZAM-IB-2001-08, 2001.
7. I. Foster, *Designing and Building Parallel Programs*, Addison Wesley, ISBN 0-201-57594-9, 1994.

8. G. Fox, *Domain Decomposition in Distributed and Shared Memory Environments*, International Conference on Supercomputing June 8-12, 1987, Athens, Greece, Lecture Notes in Computer Science 297, edited by C. Polychronopoulos, 1987.
9. KAI: *GuideView*, <http://www.kai.com/parallel/openmp.html>, 2001.
10. B. P. Miller, M. D. Callaghan, J. M. Cargille, J. K. Hollingsworth, R. B. Irvine, K. L. Karavanic, K. Kunchithapadam, and T. Newhall, The Paradyn Parallel Performance Measurement Tool, *IEEE Computer*, Vol. 28, No. 11, 37–46, 1995.
11. MPI Forum: *Message Passing Interface*, <http://www mpi-forum.org>, 2001.
12. W. E. Nagel, A. Arnold, M. Weber, H.C. Hoppe, K. Solchenbach, *VAMPIR: Visualization and Analysis of MPI Resources*, *Supercomputer* 63, Vol. 12, No. 1, 69–80, 1996.
13. OpenMP Forum: *OpenMP Standard*, <http://www.openmp.org>, 2001.
14. M. Snir, St. Otto, St. Huss-Lederman, D. Walker, J. Dongarra, *MPI - The Complete Reference*, MIT Press, ISBN 0-262-69216-3, 1998.
15. Etnus Inc.: *Totalview*, <http://www.etnus.com/Products/Totalview/>, 2001.
16. Pallas GmbH: *VAMPIR*, <http://www.pallas.de/pages/vampir.htm>, 2001.
17. F. Wolf, B. Mohr, Automatic Performance Analysis of MPI Applications Based on Event Traces, In *Proc. of the European Conference on Parallel Computing (Euro-Par)*, 123–132, Munich (Germany), 2000.
18. F. Wolf, B. Mohr, *Automatic Performance Analysis of SMP Cluster Applications*, Technical Report, Research Centre Juelich FZJ-ZAM-IB-2001-05, 2001.

Iteratively Solving Large Sparse Linear Systems on Parallel Computers

H. Martin Bücker

Institute for Scientific Computing
Aachen University of Technology, 52056 Aachen, Germany
E-mail: buecker@sc.rwth-aachen.de

Large systems of linear equations arise frequently as building blocks in many areas of scientific computing. Often, these linear systems are somehow structured or sparse, i.e., only a small number of the entries of the coefficient matrix are nonzero. In this note, numerical techniques for the solution of such linear systems are surveyed starting with a description of direct methods. When direct methods lead to excessive fill-in or when the coefficient matrix is not explicitly available, iterative methods enter the picture. These methods involve the coefficient matrix solely in the form of matrix-vector multiplications eliminating the problems of direct methods. After a summary of classical iterative methods based on relaxation of coordinates, the focus is on modern iterative methods making use of projection techniques. In particular, Krylov subspace methods are explained with an emphasis on their underlying structure rather than on their implementation details. Additional topics that are indispensable in the context of parallel computing such as reducing synchronization overhead and graph partitioning are also covered.

1 An Algorithmic Shift in Large-Scale Computations

Why would you want more than Gaussian elimination for the solution of systems of linear equations? The answer is that, sometimes, you *have* to use different techniques – simply to get a solution. In situations where the coefficient matrix is large and sparse, Gaussian elimination is often not applicable because of its excessive storage requirements. Of course, what is considered to be “large” varies with time. The meaning of “sparse” is somewhat vague too. A common definition is due to Wilkinson who called a matrix “sparse” whenever it is possible to take advantage of the number and location of its nonzero entries. In this survey, we assume that “sparse” means the usage of an appropriate storage scheme such that, given an $N \times N$ matrix A and some N -dimensional vector \mathbf{x} , then the number of arithmetic operations needed to compute the matrix-vector multiplication $A\mathbf{x}$ is small, say N or $N \log N$, compared to the N^2 operations of the conventional matrix-vector multiplication. Note that there are dense, but somehow structured matrices, for instance Toeplitz matrices, for which a matrix-vector multiplication can be carried out in $N \log N$ time or even better.

Under the assumption of being capable of efficiently computing a matrix-vector multiplication, we will survey numerical techniques for the solution of systems of linear equations

$$A\mathbf{x} = \mathbf{b}, \quad (1)$$

where the coefficient matrix A is nonsingular. We will concentrate on nonsymmetric matrices and refer the reader to the book by Fischer¹ for the symmetric case where, among others, the well-known conjugate gradient (CG) method for symmetric positive definite systems is described.

In Sec. 2, we will show the reason why, in large-scale computations, there is a shift from direct methods whose most prominent representative is Gaussian elimination to iterative

methods. The discussion of iterative methods begins with classical techniques summarized in Sec. 3. We will then lead over to the general framework of projection methods to start the survey of modern iterations in Sec. 4. Krylov subspace methods described in Sec. 5 fall under the class of projection methods. These methods are commonly considered to be among the most powerful iterative methods when combined with preconditioning techniques briefly mentioned in Sec. 6. On parallel computers, a number of additional issues are raised including the reduction of synchronization cost, explained in Sec. 7, and graph partitioning to efficiently compute a matrix-vector multiplication outlined in Sec. 8.

2 Difficulties with Direct Methods

Direct methods constitute one of the two classes of techniques for the solution of linear systems of type (1). In these methods, the exact solution $\mathbf{x}_* = A^{-1}\mathbf{b}$ is obtained after an a priori known, definite number of successive transformations. The storage for the coefficient matrix is usually overwritten during the course of the process by explicitly manipulating rows and columns of the matrix. Prominent examples of direct methods for nonsymmetric and symmetric positive definite systems are Gaussian elimination and Cholesky factorization, respectively.

2.1 Gaussian Elimination

Gaussian elimination is a typical direct approach for non-Hermitian linear systems. In the first phase of solving linear systems by Gaussian elimination, a decomposition or factorization of the form

$$A = PLU \quad (2)$$

where P is an $N \times N$ permutation matrix is computed. Furthermore, the factor

$$L = \begin{bmatrix} 1 & & & \\ * & 1 & & \\ \vdots & \ddots & \ddots & \\ * & \dots & * & 1 \end{bmatrix}$$

is a Lower triangular $N \times N$ matrix with unit diagonal entries and the factor

$$U = \begin{bmatrix} * & \dots & \dots & * \\ & * & & \vdots \\ & & \ddots & \vdots \\ & & & * \end{bmatrix}$$

is an *U*pper triangular matrix of the same size.

In the second phase of solving a linear system by means of Gaussian elimination, the original problem (1) is reformulated in terms of the factors L and U by using (2). Since permutation matrices are orthogonal, i.e., they satisfy $P^{-1} = P^T$, the result of the reformulation is given by the two linear systems

$$Ly = P^T\mathbf{b} \quad \text{and} \quad Ux = \mathbf{y}.$$

This reformulation, at first, appears unreasonable because a single linear system is replaced by two linear systems of the same size. The key idea behind the approach is that the two resulting systems are “extremely easy” to solve due to their tridiagonal structure.

Time complexity of Gaussian elimination is given by $2N^3/3 + \Theta(N^2)$ arithmetic operations. The computational dominant part of Gaussian elimination is the computation of the factorization (2); that is, the computation of L and U . Gaussian elimination can be implemented in-place meaning that the entries of A are overwritten by the entries of L and U during the course of the process. Thus, the storage requirement of Gaussian elimination is $N^2 + \Theta(N)$.

2.2 Cholesky Factorization

For Hermitian positive definite systems, the Gaussian elimination simplifies to a method known as Cholesky factorization. In the first phase of a Cholesky factorization, a decomposition

$$A = LL^H \quad (3)$$

is computed where, as before, L is a Lower triangular matrix

$$L = \begin{bmatrix} * & & & \\ \vdots & * & & \\ \vdots & & \ddots & \\ * & \dots & \dots & * \end{bmatrix}$$

but now with general, real diagonal entries $l_{ii} > 0$. When there is ambiguity, the so-called Cholesky triangle is indexed by the matrix to be decomposed; that is, the symbol L_A is used to denote the Cholesky triangle of a matrix A satisfying (3).

In the second phase of a Cholesky factorization, the original problem (1) is solved in terms of the Cholesky triangle:

$$Ly = \mathbf{b} \quad \text{and} \quad L^H \mathbf{x} = \mathbf{y}.$$

As in Gaussian elimination, the dominant part of the Cholesky factorization is the computation of the decomposition (3); that is, the computation of the Cholesky triangle L . The overall time complexity is $N^3/3 + \Theta(N^2)$ arithmetic operations. It is possible to arrange the Cholesky factorization so that L overwrites the lower triangle of A . Thus, the storage requirement of the Cholesky factorization is $N^2/2 + \Theta(N)$. Note that, compared to Gaussian elimination for general matrices, the factorization for Hermitian positive definite matrices needs approximately half as much operations as well as half of the storage.

2.3 Additional Remarks

The situation in Gaussian elimination and Cholesky factorization is rather typical for direct methods in the following sense:

- Direct methods commonly proceed in two phases. During the first, computationally intensive phase, a decomposition of the coefficient matrix into factors is computed, either implicitly or explicitly. In the second phase, the original problem is reformulated and finally solved in terms of these factors.

- The number of arithmetic operations and the storage requirement of direct methods is often cubic and square, respectively, in the order of the coefficient matrix. Both properties are likely to be unacceptable in large-scale computations where the order of the matrix is rapidly increasing with time.

The discussion given in the preceding two subsections presents the direct solution of systems of linear equations in terms of matrices. The traditional implementations on a scalar level vary significantly and their performance on today's computers with deep memory hierarchies differ dramatically. Certain reorganizations of the algorithms in terms of matrix-vector and matrix-matrix operations rather than on scalar operations are known to substantially increase the performance as discussed by Dongarra et al.² Rather than concentrating on these techniques of tuning performance, the focus here is on two serious and inherent weaknesses of direct methods when applied to large and sparse—as opposed to small and dense—systems.

2.4 The Problem of Fill-in

The following well-known example lucidly explains the difficulties of direct methods in the context of sparsity. Suppose that the task is to solve a symmetric positive definite system of order N with a sparse coefficient matrix

$$A = \begin{bmatrix} * & * & * & * & * & \cdots & * & * & * \\ * & * & & & & & & & \\ * & & * & & & & & & \\ * & & & * & & & & & \\ \vdots & & & & * & & \ddots & & \\ * & & & & & & & * & \\ * & & & & & & & & * \end{bmatrix} \quad (4)$$

whose sparsity pattern is given in the form of an arrow. Since A is symmetric a sparse storage scheme needs only to store $2N - 1$ nonzero elements of A for its complete representation. The corresponding Cholesky triangle is given by

$$L_A = \begin{bmatrix} * & & & & & & \\ * & * & & & & & \\ * & * & * & & & & \\ * & * & * & * & & & \\ * & * & * & * & * & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \\ * & * & * & * & * & \cdots & * \\ * & * & * & * & * & \cdots & * \\ * & * & * & * & * & \cdots & * \\ * & * & * & * & * & \cdots & * \end{bmatrix} \quad (5)$$

and consists of $\Theta(N^2)$ nonzero entries. By comparing the sparsity of the lower triangular part of A and L_A , we find that L_A has a lot more nonzero entries than A . The phenomenon of turning a zero element of a sparse matrix into a nonzero element during a factorization is called fill-in. This kind of behavior is by no means restricted to the Cholesky factorization but applies to different factorization schemes as well. Fill-in is a general phenomenon of direct methods and may lead to severe storage problems in the context of high-performance computing. If A is large it may already be hard to keep its sparse representation within the limits of available storage capacity. Thus, fill-in is a measure of memory needed in addition to the extreme amount of storage used in high-performance applications anyway.

It is straight forward to ask whether the sparsity pattern of A has an influence on the amount of fill-in produced during a factorization. A different sparsity pattern results from renaming of the unknowns and reordering of the equations which can be represented by

a particular kind of permutation defined as follows. Given a matrix A as well as a permutation P , the matrix $P^T AP$ is said to be a symmetric permutation of A . A symmetric permutation of the matrix in (4) is given by

$$P^T AP = \begin{bmatrix} * & & & & * \\ * & * & & & * \\ * & * & * & & * \\ * & * & * & \ddots & * \\ \vdots & & & & \vdots \\ * & * & * & * & * \end{bmatrix} \quad (6)$$

where only the first and last components and the two corresponding equations are permuted. For symmetric positive definite matrices A , it is possible to show that any symmetric permutation $P^T AP$ is also symmetric positive definite. In other words, symmetric permutations preserve the property of symmetric positive definiteness. So, the Cholesky factorization can be applied to the matrix $P^T AP$ in (6) leading to a Cholesky triangle

$$L_{P^T AP} = \begin{bmatrix} * & & & & * \\ * & * & & & * \\ * & * & * & & * \\ * & * & * & \ddots & * \\ \vdots & & & & \vdots \\ * & * & * & * & * \end{bmatrix}. \quad (7)$$

Comparing the Cholesky triangles L_A in (5) and $L_{P^T AP}$ in (7), we observe that the sparsity pattern of the matrix to which a Cholesky factorization is applied does have a significant effect on the sparsity pattern of the matrix generated by the Cholesky factorization. The number of nonzero elements of L_A is $\Theta(N^2)$ whereas there are only $\Theta(N)$ nonzero elements in $L_{P^T AP}$. Unfortunately, the computation of a symmetric permutation leading to the minimum number of fill-in turns out to be a hard combinatorial optimization problem. More precisely, the so-called minimum fill-in problem is NP-complete meaning that, currently, there is no deterministic algorithm for its solution where the number of arithmetic operations scales polynomially with the order of the matrix. Moreover, from the point of view of theoretical computer science, it is very unlikely that one will ever find such an algorithm; see the book by Garey and Johnson³ for more information on NP-completeness.

In summary, direct methods applied to sparse linear systems may lead to a dramatically high amount of fill-in prohibiting their use in large-scale applications. Furthermore, there is currently no computationally efficient technique to compute the minimum fill-in. Recently, an approximation algorithm for the minimum fill-in problem was developed⁴ but its suitability for practical use is still open.

2.5 The Problem of Needing Explicit Access to the Matrix

Direct methods manipulate rows or columns of the coefficient matrix. Therefore, there is need to explicitly access entries of the coefficient matrix. In some applications, however, the matrix is not explicitly given. That is, the computation of the complete matrix is extremely expensive in terms of arithmetic operations whereas there is a computationally efficient procedure to compute the product of the coefficient matrix and some given vector.

An example of such a situation occurs in the solution of nonlinear systems of equations by Newton-type methods. Here, a subtask is to repeatedly solve linear systems of the form $J\mathbf{x} = \mathbf{b}$, where J is the Jacobian matrix of some function f . If there is a program in

a high-level programming language like Fortran or C evaluating f for a given set of inputs, a technique called automatic differentiation is applicable to provide efficient code for computing Jacobian-vector multiplications. In contrast to numerical differentiation based on divided differencing delivering approximations to derivatives, automatic differentiation produces derivatives accurate up to machine precision. See the book by Griewank⁵ or the web portal <http://www.autodiff.org> for an introduction to and more details on automatic differentiation.

The computation of all entries of J by automatic differentiation is more efficient than numerical differentiation under a wide range of circumstances.^{6–8} The crucial point in the context of this note is the fact that, using automatic differentiation, Jacobian-vector multiplications are computationally even N times more efficient than computing all entries of J . From a conceptual point of view, one may explain the factor N by comparing a single matrix-vector multiplication and a sequence of N matrix-vector multiplications $J\mathbf{e}_1, J\mathbf{e}_2, \dots, J\mathbf{e}_N$, where \mathbf{e}_i is the i th Cartesian unit vector, to compute all columns of J .

Since iterative methods make use of the coefficient matrix in the form of matrix-vector multiplications they do not suffer from neither the problem of fill-in nor from the problem of needing explicit access to the coefficient matrix.

3 Classical Iterations

Iterative methods enter the picture when direct methods produce excessive fill-in or the coefficient matrix is not explicitly available. By using the coefficient matrix in the form of matrix-vector multiplications, iterative methods are capable of handling these situations.

In their n th step, iterative methods compute approximations \mathbf{x}_n to the exact solution $\mathbf{x}_* = A^{-1}\mathbf{b}$ of the linear system (1). The corresponding residual vector is defined by

$$\mathbf{r}_n = \mathbf{b} - A\mathbf{x}_n \quad (8)$$

and determines how far the approximation \mathbf{x}_n is from the right hand side \mathbf{b} . The goal of any iterative method is to drive the residual vector to the zero vector because in this case $\mathbf{b} = A\mathbf{x}_n$ and thus the approximation \mathbf{x}_n equals the exact solution \mathbf{x}_* . As a matter of fact, it is indispensable that, in order to beat the $\Theta(N^3)$ time complexity of direct methods, the approximations \mathbf{x}_n should converge fast to the exact solution or a sufficiently accurate solution. To indicate this, the notation (big) N is used to denote the order of a (large) matrix and the symbol (little) n is used for a (small) iteration index expressing the desirable relation $n \ll N$; that is, any viable iterative method should converge in a number of steps that is significantly smaller than the order of the matrix.

Classical iterative methods for the solution of linear systems date back at least to the 19th century. They are characterized by defining the approximations by a sequence of the form

$$M\mathbf{x}_n = \mathbf{b} + S\mathbf{x}_{n-1} \quad (9)$$

where

$$A = M - S \quad (10)$$

$[\mathbf{x}_n, \mathbf{r}_n] = \text{RELAXATION}(A, \mathbf{b}, \mathbf{x}_0)$ If $A \in \mathbb{C}^{N \times N}$ with splitting $A = M - S$ with non-singular M , this algorithm computes approximations \mathbf{x}_n (with corresponding residuals \mathbf{r}_n) to the solution of the linear system $A\mathbf{x} = \mathbf{b}$ for any starting vector \mathbf{x}_0 .

- 1: Choose $\mathbf{x}_0 \in \mathbb{C}^N$, set $\mathbf{r}_0 \leftarrow \mathbf{b} - A\mathbf{x}_0$, and solve $M\mathbf{z}_0 = \mathbf{r}_0$
 - 2: **for** $n = 1, 2, 3, \dots$ **do** {until convergence}
 - 3: $\mathbf{x}_n \leftarrow \mathbf{z}_{n-1} + \mathbf{x}_{n-1}$
 - 4: $\mathbf{r}_n \leftarrow \mathbf{b} - A\mathbf{x}_n$
 - 5: Solve $M\mathbf{z}_n = \mathbf{r}_n$
 - 6: **end for**
-

Figure 1. General form of a classical iterative method based on relaxation of coordinates.

is a general matrix splitting. Given any starting vector \mathbf{x}_0 , these iterative methods obtain the next approximation by modifying one or a few components of the current approximation. This class of methods is said to be based on relaxation of coordinates.

To derive a basic formulation of relaxation methods, observe that inserting S from (10) into (9) yields

$$M\mathbf{x}_n = \mathbf{b} - A\mathbf{x}_{n-1} + M\mathbf{x}_{n-1}.$$

If M is nonsingular an equivalent form is given by

$$\mathbf{x}_n = M^{-1}\mathbf{r}_{n-1} + \mathbf{x}_{n-1}.$$

The resulting process is depicted in Fig. 1. It is important to remark that linear systems with coefficient matrices M should be “easy” to solve because these systems are to be solved in each step of the iteration.

The Jacobi iteration and the Gauss–Seidel iteration are popular examples of these classical iterative methods. If D , L , and U denote the diagonal, the strict lower triangle, and the strict upper triangle, respectively, then the matrix splittings of the Jacobi and Gauss–Seidel iterations are given by

$$M_{\text{Jac}} = D, \quad S_{\text{Jac}} = -(L + U), \quad (11)$$

$$\text{and} \quad M_{\text{GS}} = D + L, \quad S_{\text{GS}} = -U, \quad (12)$$

where the subscripts are used to identify the two methods.

A discussion of the convergence behavior of relaxation methods can be found in almost any introductory textbook on iterative methods. Typically, the analysis is formulated in terms of the spectral radius of the so-called iteration matrix, $M^{-1}S$. The spectral radius of a matrix B is defined by

$$\rho(B) := \max_{\lambda \text{ is eigenvalue of } B} |\lambda|.$$

The following theorem whose proof can be found in the book by Golub and van Loan⁹ summarizes an important result.

Theorem 3.1. *An iterative scheme of the form*

$$M\mathbf{x}_n = S\mathbf{x}_{n-1} + \mathbf{b}$$

for the solution of $A\mathbf{x} = \mathbf{b}$ with nonsingular coefficient matrix $A = M - S$ converges to the exact solution $\mathbf{x}_ = A^{-1}\mathbf{b}$ for any starting vector \mathbf{x}_0 , if M is nonsingular and*

$$\rho(M^{-1}S) < 1.$$

Relaxation methods are not considered to be really efficient for solving large-scale problems. However, they are still in use as building blocks of multigrid methods or to construct preconditioners.

4 Projection Methods

A general framework to discuss iterative techniques for the solution of linear systems is a projection process. The idea of a projection method is to extract the next approximations \mathbf{x}_n from a search subspace \mathcal{K} . If the dimension of \mathcal{K} is given by m , then, in general, m restrictions or constraints are necessary to be able to extract \mathbf{x}_n from \mathcal{K} . Typically, the constraints are imposed by orthogonalizing the residual vector \mathbf{r}_n with respect to a subspace of constraints \mathcal{L} .

To illustrate the situation, let there be two subspaces, \mathcal{K} and \mathcal{L} , of dimension m with two sets of basis vectors,

$$\mathcal{K} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\} \quad \text{and} \quad \mathcal{L} = \text{span}\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}.$$

Given some current approximation \mathbf{x}_{n-1} , the next approximation \mathbf{x}_n is constructed in the search subspace, i.e.,

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \mathcal{K} \tag{13}$$

subject to

$$\mathbf{r}_n \perp \mathcal{L}. \tag{14}$$

To proceed with the discussion we introduce two $N \times m$ matrices whose columns are given by the basis vectors of \mathcal{K} and \mathcal{L} ,

$$V_m = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_m] \quad \text{and} \quad W_m = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_m],$$

respectively. Then, an equivalent form of the next approximation is given by

$$\mathbf{x}_n = \mathbf{x}_{n-1} + V_m \mathbf{y}_m, \tag{15}$$

where $\mathbf{y}_m \in \mathbb{C}^m$ is determined by the constraint (14). More precisely, the governing equation of \mathbf{y}_m follows from the reformulation of (14) in the form of

$$W_m^H \mathbf{r}_n = \mathbf{0}.$$

By inserting the residual vector corresponding to (15), an equivalent form is given by

$$W_m^H \mathbf{r}_{n-1} = W_m^H A V_m \mathbf{y}_m$$

$\mathbf{x}_n = \text{PROJECTION}(A, \mathbf{b}, \mathbf{x}_0)$ If $A \in \mathbb{C}^{N \times N}$ and \mathbf{x}_0 is a starting vector, this algorithm computes approximations \mathbf{x}_n to the solution of the linear system $A\mathbf{x} = \mathbf{b}$.

- 1: Choose $\mathbf{x}_0 \in \mathbb{C}^N$
 - 2: **for** $n = 1, 2, 3, \dots$ **do** {until convergence}
 - 3: Choose subspaces \mathcal{K} and \mathcal{L}
 - 4: Choose basis $V_m = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_m]$ of \mathcal{K} and
choose basis $W_m = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_m]$ of \mathcal{L}
 - 5: $\mathbf{r}_{n-1} \leftarrow \mathbf{b} - A\mathbf{x}_{n-1}$
 - 6: $\mathbf{y}_m \leftarrow (W_m^H A V_m)^{-1} W_m^H \mathbf{r}_{n-1}$
 - 7: $\mathbf{x}_n \leftarrow \mathbf{x}_{n-1} + V_m \mathbf{y}_m$
 - 8: **end for**
-

Figure 2. General form of a projection method.

which finally leads to

$$\mathbf{y}_m = (W_m^H A V_m)^{-1} W_m^H \mathbf{r}_{n-1}.$$

The preceding derivation leads to the general form of a projection method depicted in Fig. 2. Note that the algorithm still depends on the choice of \mathcal{K} , \mathcal{L} , and their bases.

In general, a projection method onto the search subspace \mathcal{K} orthogonal to the subspace of constraints \mathcal{L} is characterized by (13) and (14). If the search subspace \mathcal{K} is the same as the subspace of constraints \mathcal{L} , the process is called an orthogonal projection method. In an oblique projection method, \mathcal{K} is different from \mathcal{L} .

Several theoretical results are known under the general scenario of projection methods. For instance, without being specific about the subspaces \mathcal{K} and \mathcal{L} , the following result holds if the subspace of constraints is chosen to satisfy $\mathcal{L} = A\mathcal{K}$.

Theorem 4.1 (Optimality of projection method with $\mathcal{L} = A\mathcal{K}$). *A vector \mathbf{x}_n is the next approximation of a projection method onto the search subspace \mathcal{K} along the subspace of constraints $\mathcal{L} = A\mathcal{K}$ if and only if*

$$\|\mathbf{b} - A\mathbf{x}_n\| = \min_{\mathbf{x} \in \mathbf{x}_{n-1} + \mathcal{K}} \|\mathbf{b} - A\mathbf{x}\|. \quad (16)$$

Here and in the sequel, the notation $\|\cdot\|$ is used to denote the Euclidean norm. The proof of the preceding theorem is given in the book by Saad¹⁰ which also contains additional material on general optimality results.

5 Krylov Subspace Methods

In the canonical form of projection methods introduced in the preceding section, the particular choice of a search subspace \mathcal{K} as well as of a subspace of constraints \mathcal{L} is not specified. In this section, we will explicitly describe candidates for both, \mathcal{K} and \mathcal{L} .

5.1 The Search Subspace

Krylov subspace methods are currently considered to be among the most powerful iterative methods for the solution of large sparse linear systems. A projection method onto the search subspace

$$\mathcal{K}_n(A, \mathbf{r}_0) := \text{span}\{\mathbf{r}_0, A\mathbf{r}_0, A^2\mathbf{r}_0, \dots, A^{n-1}\mathbf{r}_0\}$$

is called a Krylov subspace method. The subspace $\mathcal{K}_n(A, \mathbf{r}_0)$ is referred to as the n th Krylov subspace generated by the matrix A and the vector \mathbf{r}_0 . Any Krylov subspace method for the solution of $A\mathbf{x} = \mathbf{b}$ is characterized by constructing the vector $\mathbf{x}_n - \mathbf{x}_0$ in subspaces of the specific form $\mathcal{K}_n(A, \mathbf{r}_0)$ where $\mathbf{r}_0 := \mathbf{b} - A\mathbf{x}_0$ is the initial residual vector associated with the initial guess \mathbf{x}_0 . A key feature of any Krylov subspace method is to find accurate approximations $\mathbf{x}_n \in \mathbf{x}_0 + \mathcal{K}_n(A, \mathbf{r}_0)$ when $n \ll N$.

The straightforward approach to construct a basis of $\mathcal{K}_n(A, \mathbf{r}_0)$ is to repeatedly multiply the starting vector \mathbf{r}_0 by the matrix A . The resulting algorithm given in Fig. 3 is known as the power method but is a numerically useless process to span a basis of $\mathcal{K}_n(A, \mathbf{r}_0)$. The reason is that the power method converges to the largest eigenvalue (in absolute value) and, therefore, the vectors will soon become linearly dependent in finite-precision arithmetic.¹¹ To emphasize the importance of a numerically stable process for the generation of the Krylov subspace, we abstract from the particular “application” of solving linear systems and use a general starting vector \mathbf{v}_1 in the following discussion.

$V_n = \text{POWERMETHOD}(A, \mathbf{v}_1)$ If $A \in \mathbb{C}^{N \times N}$ and \mathbf{v}_1 is a suitable starting vector, this algorithm computes a (numerically useless) basis $V_n = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n] \in \mathbb{C}^{N \times n}$ of $\mathcal{K}_n(A, \mathbf{v}_1)$.

- 1: Choose $\mathbf{v}_1 \in \mathbb{C}^N$ such that $\|\mathbf{v}_1\| = 1$
 - 2: **for** $n = 1, 2, 3, \dots$ **do** {until invariance}
 - 3: $\tilde{\mathbf{v}}_{n+1} \leftarrow A\mathbf{v}_n$
 - 4: $\alpha_{n+1} \leftarrow \|\tilde{\mathbf{v}}_{n+1}\|$
 - 5: $\mathbf{v}_{n+1} \leftarrow \frac{1}{\alpha_{n+1}}\tilde{\mathbf{v}}_{n+1}$
 - 6: **end for**
-

Figure 3. Power method.

Arnoldi Algorithm

A more promising approach for the generation of a basis of $\mathcal{K}_n(A, \mathbf{v}_1)$ is the Arnoldi process¹² which computes an orthonormal basis of $\mathcal{K}_n(A, \mathbf{v}_1)$. The process of orthogonalization works as follows. If $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ already is an orthonormal set of basis vectors, then the vector

$$\mathbf{z} = \tilde{\mathbf{v}}_{n+1} - (\mathbf{v}_1^H \tilde{\mathbf{v}}_{n+1})\mathbf{v}_1 - (\mathbf{v}_2^H \tilde{\mathbf{v}}_{n+1})\mathbf{v}_2 - \cdots - (\mathbf{v}_n^H \tilde{\mathbf{v}}_{n+1})\mathbf{v}_n$$

is orthogonal to $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$. This can be derived from multiplication by \mathbf{v}_i^H where $i = 1, 2, \dots, n$ resulting in

$$\mathbf{v}_i^H \mathbf{z} = \mathbf{v}_i^H \tilde{\mathbf{v}}_{n+1} - (\mathbf{v}_i^H \tilde{\mathbf{v}}_{n+1}) \mathbf{v}_i^H \mathbf{v}_i = 0$$

due to the orthonormality of the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$. A new set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n, \mathbf{v}_{n+1}\}$ of orthonormal basis vectors is easily obtained by adding \mathbf{v}_{n+1} as a scaled version of \mathbf{z} .

The resulting Arnoldi method is shown in Fig. 4 where the generated vectors \mathbf{v}_i are called Arnoldi vectors. Step n of this algorithm computes the product of the previous Arnoldi vector \mathbf{v}_n and A , i.e., $\tilde{\mathbf{v}}_{n+1} = A\mathbf{v}_n$, and orthogonalizes $\tilde{\mathbf{v}}_{n+1}$ towards all previous Arnoldi vectors \mathbf{v}_i with $i = 1, 2, \dots, n$ by the procedure explained above. Finally, $\tilde{\mathbf{v}}_{n+1}$ is scaled to unity in the Euclidean norm. This kind of orthonormalization is called the standard Gram–Schmidt process.

The complete Arnoldi process can be written in matrix notation by first combining lines 3, 7 and 9 in the following vector equation

$$A\mathbf{v}_n = h_{n+1,n}\mathbf{v}_{n+1} + \sum_{i=1}^n h_{in}\mathbf{v}_i$$

representing an $(n+1)$ -term recurrence for the computation of the Arnoldi vectors. Collecting the recurrence coefficients in an upper Hessenberg matrix

$$H_n = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \ddots & \ddots & \ddots & \vdots \\ h_{n,n-1} & h_{nn} \end{bmatrix} \in \mathbb{C}^{n \times n},$$

a matrix with zeros below the first subdiagonal, leads to the matrix form summarized in the following result.

$V_n = \text{BASICARNOLDI}(A, \mathbf{v}_1)$ If $A \in \mathbb{C}^{N \times N}$ and \mathbf{v}_1 is a suitable starting vector, this algorithm computes an orthonormal basis $V_n = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n] \in \mathbb{C}^{N \times n}$ of $\mathcal{K}_n(A, \mathbf{v}_1)$ via the standard Gram–Schmidt process.

- 1: Choose $\mathbf{v}_1 \in \mathbb{C}^N$ such that $\|\mathbf{v}_1\| = 1$
 - 2: **for** $n = 1, 2, 3, \dots$ **do** {until invariance}
 - 3: $\tilde{\mathbf{v}}_{n+1} \leftarrow A\mathbf{v}_n$
 - 4: **for** $i = 1, 2, \dots, n$ **do**
 - 5: $h_{in} \leftarrow \mathbf{v}_i^H \tilde{\mathbf{v}}_{n+1}$
 - 6: **end for**
 - 7: $\tilde{\mathbf{v}}_{n+1} \leftarrow \tilde{\mathbf{v}}_{n+1} - \sum_{i=1}^n h_{in}\mathbf{v}_i$
 - 8: $h_{n+1,n} \leftarrow \|\tilde{\mathbf{v}}_{n+1}\|$
 - 9: $\mathbf{v}_{n+1} \leftarrow \frac{1}{h_{n+1,n}}\tilde{\mathbf{v}}_{n+1}$
 - 10: **end for**
-

Figure 4. Arnoldi method via standard Gram–Schmidt orthogonalization.

Theorem 5.1 (Arnoldi). *In exact arithmetic the Arnoldi vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ generated during the course of Fig. 4 form an orthonormal basis*

$$V_n^H V_n = I_n \quad (17)$$

of the Krylov subspace

$$\mathcal{K}_n(A, \mathbf{v}_1) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}. \quad (18)$$

Successive Arnoldi vectors are related by

$$AV_n = V_n H_n + h_{n+1,n} \mathbf{v}_{n+1} \mathbf{e}_n^T, \quad (19)$$

and the matrix A is reduced to (full) upper Hessenberg form

$$V_n^H A V_n = H_n \quad (20)$$

by means of unitary transformations V_n .

The very best of the Arnoldi method is the orthogonality of its basis. Orthogonality is a highly-desired feature from the point of view of numerical stability. Moreover, it can be exploited to the advantage of minimizing the Euclidean norm of the residual in an iterative method referred to as GMRES as is shown in the next subsection. The main disadvantage of the Arnoldi method is that its computation is expensive in terms of both arithmetic operations and storage requirement. The Arnoldi process is based on $(n+1)$ -term recurrences as is reflected in line 7 of Fig. 4 where all previous Arnoldi vectors are involved or, equivalently, by the fact that the upper Hessenberg matrix H_n is full. Due to this property the Arnoldi method is said to be based on long recurrences. The n th iteration of the Arnoldi method requires $\Theta(n \cdot N)$ arithmetic operations as well as $\Theta(n \cdot N)$ storage. The most unpleasant feature in practical applications when large sparse matrices are involved is the storage requirement of the Arnoldi process that grows linearly with the iteration number. The use of the Arnoldi process may therefore sometimes be prohibited by its long recurrences.

A more stable formulation of the Arnoldi method results from replacing the standard Gram–Schmidt process by a modified Gram–Schmidt orthogonalization; see Saad¹⁰ for details.

Lanczos Algorithm

Another process for the generation of a basis of $\mathcal{K}_n(A, \mathbf{v}_1)$ is the Lanczos algorithm.¹³ In contrast to the long recurrences of the Arnoldi method, the Lanczos algorithm is based on three-term recurrences. However, the basis is no longer unitary. Furthermore, the Lanczos algorithm not only computes a basis of $\mathcal{K}_n(A, \mathbf{v}_1)$ for some starting vector \mathbf{v}_1 but also an additional basis of $\mathcal{K}_n(A^H, \mathbf{w}_1)$ for a second starting vector \mathbf{w}_1 . The process is depicted in Fig. 5 and summarized in the following theorem.

Theorem 5.2 (Lanczos). *In exact arithmetic the Lanczos vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ and $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ generated during the course of Fig. 5 are scaled to unity in the Euclidean norm and form a biorthogonal system, i.e.,*

$$W_n^H V_n = D_n := \text{diag}(\delta_1, \delta_2, \dots, \delta_n), \quad (21)$$

$[V_n, W_n] = \text{BIOLANZOS}(A, \mathbf{v}_1, \mathbf{w}_1)$ If $A \in \mathbb{C}^{N \times N}$ and $\mathbf{v}_1, \mathbf{w}_1$ are suitable starting vectors, this algorithm computes biorthogonal bases $V_n = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n] \in \mathbb{C}^{N \times n}$ and $W_n = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_n] \in \mathbb{C}^{N \times n}$ of $\mathcal{K}_n(A, \mathbf{v}_1)$ and $\mathcal{K}_n(A^H, \mathbf{w}_1)$, respectively.

- 1: Choose $\mathbf{v}_1, \mathbf{w}_1 \in \mathbb{C}^N$ such that $\|\mathbf{v}_1\| = \|\mathbf{w}_1\| \leftarrow 1$ and $\delta_1 \leftarrow \mathbf{w}_1^H \mathbf{v}_1 \neq 0$
 - 2: Set $\mathbf{v}_0 = \mathbf{w}_0 \leftarrow \mathbf{0}$ and $\gamma_1 = \rho_1 \leftarrow 0, \xi_1 \neq 0$
 - 3: **for** $n = 1, 2, 3, \dots$ **do** {until invariance}
 - 4: $\alpha_n \leftarrow \mathbf{w}_n^H A \mathbf{v}_n / \delta_n$
 - 5: $\tilde{\mathbf{v}}_{n+1} \leftarrow A \mathbf{v}_n - \alpha_n \mathbf{v}_n - \gamma_n \mathbf{v}_{n-1}$
 - 6: $\tilde{\mathbf{w}}_{n+1} \leftarrow A^H \mathbf{w}_n - \overline{\alpha_n} \mathbf{w}_n - \frac{\overline{\gamma_n} \overline{\rho_n}}{\xi_n} \mathbf{w}_{n-1}$
 - 7: $\rho_{n+1} \leftarrow \|\tilde{\mathbf{v}}_{n+1}\|$
 - 8: $\xi_{n+1} \leftarrow \|\tilde{\mathbf{w}}_{n+1}\|$
 - 9: $\mathbf{v}_{n+1} \leftarrow \frac{1}{\rho_{n+1}} \tilde{\mathbf{v}}_{n+1}$
 - 10: $\mathbf{w}_{n+1} \leftarrow \frac{1}{\xi_{n+1}} \tilde{\mathbf{w}}_{n+1}$
 - 11: $\delta_{n+1} \leftarrow \mathbf{w}_{n+1}^H \mathbf{v}_{n+1}$
 - 12: $\gamma_{n+1} \leftarrow \frac{1}{\xi_{n+1}} \delta_{n+1} / \delta_n$
 - 13: **end for**
-

Figure 5. Biorthogonal Lanczos method.

as well as a pair of bases of the Krylov subspaces

$$\mathcal{K}_n(A, \mathbf{v}_1) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}, \quad (22)$$

$$\mathcal{K}_n(A^H, \mathbf{w}_1) = \text{span}\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}. \quad (23)$$

Successive Lanczos vectors are related by

$$AV_n = V_n T_n + \rho_{n+1} \mathbf{v}_{n+1} \mathbf{e}_n^T, \quad (24)$$

$$A^H W_n = W_n D_n^{-H} T_n^H D_n^H + \xi_{n+1} \mathbf{w}_{n+1} \mathbf{e}_n^T, \quad (25)$$

and the matrix A is reduced to tridiagonal form

$$W_n^H A V_n = D_n T_n \quad (26)$$

by means of similarity transformations V_n .

The Lanczos algorithm computes the Lanczos vectors \mathbf{v}_i and \mathbf{w}_i by means of three-term recurrences as is reflected in a tridiagonal matrix T_n rather than a full upper Hessenberg matrix as in the Arnoldi process. Compared to the Arnoldi algorithm, the price to be paid for these short recurrences is the computation of a second sequence \mathbf{w}_i in addition to the sequence \mathbf{v}_i . The sequence \mathbf{w}_i is based on repeatedly multiplying vectors by A^H rather than by A as for the sequence \mathbf{v}_i . So, each iteration of the Lanczos algorithm needs two matrix-by-vector products and thus is computationally twice as expensive as the Arnoldi method.

A professional implementation of the Lanczos algorithm is based on look-ahead techniques; see Gutknecht¹⁴ and the references given therein.

Additional Remark

While the Arnoldi process corresponds to a Hessenberg orthogonalization, the Lanczos algorithm is summarized as a tridiagonal biorthogonalization. Unfortunately, it is not possible to define algorithms making use of a tridiagonal orthogonalization, i.e., combining optimal projection techniques and short recurrences.¹⁵

5.2 The Subspace of Constraints

A Krylov subspace method is a projection method onto the Krylov subspace $\mathcal{K}_n(A, \mathbf{r}_0)$ along a subspace of constraints \mathcal{L}_n . More formally, the iterates of a Krylov subspace method are of the form

$$\mathbf{x}_n \in \mathbf{x}_0 + \mathcal{K}_n(A, \mathbf{r}_0) \quad (27)$$

subject to

$$\mathbf{r}_n \perp \mathcal{L}_n. \quad (28)$$

That is, the iterates are constructed in the particular search space $\mathcal{K}_n(A, \mathbf{r}_0)$ while their actual definition is based on the restriction that the associated residual vector

$$\mathbf{r}_n = \mathbf{b} - A\mathbf{x}_n \quad (29)$$

is orthogonal to a subspace \mathcal{L}_n . In this subsection, we will discuss different candidates for this subspace of constraints. The Arnoldi and Lanczos methods started with

$$\mathbf{v}_1 = \frac{\mathbf{r}_0}{\|\mathbf{r}_0\|} \quad (30)$$

will then be used as underlying processes of different Krylov subspace methods derived by applying these different approaches for the choice of \mathcal{L}_n .

Using (15), an equivalent representation of the iterates (27) is given by

$$\mathbf{x}_n = \mathbf{x}_0 + V_n \mathbf{y} \quad \text{for some } \mathbf{y} \in \mathbb{C}^n. \quad (31)$$

Here, \mathbf{y} is a free parameter vector that is fixed by imposing the condition (28). Note from (29) that, since the coefficient matrix is nonsingular, there is a bijection between \mathbf{x}_n and \mathbf{r}_n . Thus, imposing a condition on \mathbf{r}_n corresponds to fixing the free parameter vector \mathbf{y} in the representation (31) of \mathbf{x}_n . Let $\mathbf{y}_n \in \mathbb{C}^n$ denote the corresponding vector determining the actual definition of the iterates, i.e.,

$$\mathbf{x}_n = \mathbf{x}_0 + V_n \mathbf{y}_n. \quad (32)$$

Then, the resulting Krylov subspace methods differ in

- the choice of the basis V_n of the search subspace $\mathcal{K}_n(A, \mathbf{r}_0)$ and
- the definition of \mathbf{y}_n by the choice of the subspace of constraints \mathcal{L}_n .

An interpretation is as follows. A designer of a Krylov subspace method has two degrees of freedom. Firstly, there is the choice of how a basis of the underlying Krylov subspaces is generated in a numerically stable way, for instance, the Arnoldi and Lanczos methods may be used. Secondly, there is the option to define the actual iterates by different choices of \mathbf{y}_n four of which are described in this subsection.

Fixing the free parameter vector \mathbf{y} in (31) by imposing the condition (28) relates the vector \mathbf{y}_n to a basis

$$W_n = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_n]$$

of the subspace \mathcal{L}_n as follows. Rewriting (28) in terms of the basis W_n of \mathcal{L}_n yields

$$W_n^H \mathbf{r}_n = \mathbf{0}.$$

Inserting the residual vector associated to the iterate (32) which is given by

$$\mathbf{r}_n = \mathbf{r}_0 - AV_n \mathbf{y}_n$$

results in

$$W_n^H AV_n \mathbf{y}_n = W_n^H \mathbf{r}_0. \quad (33)$$

In the following paragraphs, different bases W_n of \mathcal{L}_n are used in this equation leading to different vectors \mathbf{y}_n .

The Ritz–Galerkin Approach

There are two broad classes of projection methods. The classification is based on whether or not the search subspace \mathcal{K}_n is the same as the subspace of constraints \mathcal{L}_n . Recall that, in an orthogonal projection method, both subspaces are equal while they are different in an oblique projection method. An orthogonal Krylov subspace method takes

$$\mathcal{K}_n = \mathcal{L}_n = \mathcal{K}_n(A, \mathbf{r}_0).$$

This approach is called *Ritz–Galerkin approach*. To derive an orthogonal Krylov subspace method this approach is applied to an underlying process for the generation of $\mathcal{K}_n(A, \mathbf{r}_0)$. When applying the Ritz–Galerkin approach to a process like the Arnoldi method there is no need for the generation of a second basis W_n because the two subspaces \mathcal{K}_n and \mathcal{L}_n are the same and the underlying process already generates a basis V_n of the Krylov subspace $\mathcal{K}_n(A, \mathbf{r}_0)$. Thus, inserting $W_n = V_n$ in (33) leads to

$$V_n^H AV_n \mathbf{y}_n = V_n^H \mathbf{r}_0.$$

Assume that the Ritz–Galerkin approach is applied to the Arnoldi method started with (30). Then, using Theorem 5.1 results in

$$H_n \mathbf{y}_n = \|\mathbf{r}_0\| \mathbf{e}_1, \quad (34)$$

where $\mathbf{e}_1 = (1, 0, 0, \dots, 0)^T$ is the first Cartesian unit vector of appropriate dimension. That is, from a conceptual point of view, the vector \mathbf{y}_n defining the iterate \mathbf{x}_n is available by solving a linear system whose coefficient matrix is given by the upper Hessenberg matrix H_n generated by the Arnoldi process. Note that a Hessenberg system of type (34) is to be solved in each iteration, n , of the resulting process for the solution of the original system $A\mathbf{x} = \mathbf{b}$. Fortunately, the system (34) is a “small” $n \times n$ system as opposed to the original large $N \times N$ system. The actual implementation of the resulting method called Full Orthogonalization Method (FOM) is beyond the scope of this note. A highlevel description of FOM is given in Fig. 6. There are many possible variations including those dealing with practical issues such as reducing the high computational and memory cost incurred by the underlying long recurrences of the Arnoldi process; see Saad¹⁰ for more details on restarting FOM and on truncation of the underlying orthogonalization process.

$\mathbf{x}_n = \text{FOM}(A, \mathbf{b}, \mathbf{x}_0)$ If $A \in \mathbb{C}^{N \times N}$, this algorithm computes approximations \mathbf{x}_n to the solution of the linear system $A\mathbf{x} = \mathbf{b}$ for any starting vector \mathbf{x}_0 .

```

1:  $\mathbf{r}_0 \leftarrow \mathbf{b} - A\mathbf{x}_0$ 
2:  $\mathbf{v}_1 \leftarrow \mathbf{r}_0 / \|\mathbf{r}_0\|$ 
3: for  $n = 1, 2, 3, \dots$  do {until convergence}
4:   Step  $n$  of Arnoldi method producing  $H_n$  and  $V_n$ 
5:    $\mathbf{y}_n \leftarrow H_n^{-1} \|\mathbf{r}_0\| \mathbf{e}_1$ 
6:    $\mathbf{x}_n \leftarrow \mathbf{x}_0 + V_n \mathbf{y}_n$ 
7: end for

```

Figure 6. Highlevel Full Orthogonalization Method (FOM).

The Petrov–Galerkin Approach

In contrast to an orthogonal projection method where the search subspace \mathcal{K}_n and the subspace of constraints \mathcal{L}_n are the same, \mathcal{K}_n and \mathcal{L}_n are different in an oblique projection method. An oblique Krylov subspace method can be derived by taking

$$\mathcal{L}_n = \mathcal{K}_n(A^H, \mathbf{w}_1),$$

where $\mathbf{w}_1 \in \mathbb{C}^N$ is a nonzero starting vector for the generation of a basis of \mathcal{L}_n . An approach where \mathcal{K}_n and \mathcal{L}_n are different is called *Petrov–Galerkin approach*. Recall that the Lanczos algorithm not only generates a basis \mathbf{V}_n of $\mathcal{K}_n(A, \mathbf{r}_0)$ but also computes another basis \mathbf{W}_n of $\mathcal{K}_n(A^H, \mathbf{w}_1)$. Thus, a basis of \mathcal{L}_n is available by means of the Lanczos process suggesting the application of the Petrov–Galerkin approach to the Lanczos algorithm. To this end, start the Lanczos algorithm with two starting vectors \mathbf{v}_1 as given by (30) and \mathbf{w}_1 satisfying

$$\|\mathbf{w}_1\| = 1 \quad \text{and} \quad \mathbf{v}_1^H \mathbf{w}_1 \neq 0. \quad (35)$$

Then, (33) and Theorem 5.2 lead to

$$D_n T_n \mathbf{y}_n = \delta_1 \|\mathbf{r}_0\| \mathbf{e}_1.$$

Since D_n is a nonsingular diagonal matrix an equivalent form is given by

$$T_n \mathbf{y}_n = \|\mathbf{r}_0\| \mathbf{e}_1. \quad (36)$$

This situation is quite similar to the Ritz–Galerkin approach described in the previous section in the sense that, in every iteration n , a small $n \times n$ systems has to be solved in order to obtain the vector \mathbf{y}_n . Here, however, the coefficient matrix T_n generated by the Lanczos process is tridiagonal whereas the coefficient matrix in (34) is a (full) upper Hessenberg matrix. The resulting method is called Bi-Lanczos method¹⁶ and its highlevel description is depicted in Fig. 7.

$\mathbf{x}_n = \text{BI-LANZOS}(A, \mathbf{b}, \mathbf{x}_0)$ If $A \in \mathbb{C}^{N \times N}$, this algorithm computes approximations \mathbf{x}_n to the solution of the linear system $A\mathbf{x} = \mathbf{b}$ for any starting vector \mathbf{x}_0 .

```

1:  $\mathbf{r}_0 \leftarrow \mathbf{b} - A\mathbf{x}_0$ 
2:  $\mathbf{v}_1 \leftarrow \mathbf{r}_0 / \|\mathbf{r}_0\|$ 
3: Choose  $\mathbf{w}_1$  such that  $\|\mathbf{w}_1\| \leftarrow 1$  and  $\mathbf{v}_1^H \mathbf{w}_1 \neq 0$ 
4: for  $n = 1, 2, 3, \dots$  do {until convergence}
5:   Step  $n$  of biorthogonal Lanczos algorithm producing  $T_n$  and  $V_n$ 
6:    $\mathbf{y}_n \leftarrow T_n^{-1} \|\mathbf{r}_0\| \mathbf{e}_1$ 
7:    $\mathbf{x}_n \leftarrow \mathbf{x}_0 + V_n \mathbf{y}_n$ 
8: end for

```

Figure 7. Highlevel Bi-Lanczos Method.

The Minimum Residual Approach

In Theorem 4.1, an optimality result for oblique projection methods is given under quite general assumptions. In particular, without being specific about the subspaces \mathcal{K}_n and \mathcal{L}_n , an oblique projection method minimizes the Euclidean norm of the residual vector over the space $\mathbf{x}_0 + \mathcal{K}_n$ if and only if the subspace of constraints is defined by $\mathcal{L}_n = A\mathcal{K}_n$. This result is applied here to Krylov subspace methods where $\mathcal{K}_n = \mathcal{K}_n(A, \mathbf{r}_0)$.

Suppose that, in a Krylov subspace method, the subspace of constraints is chosen as

$$\mathcal{L}_n = A\mathcal{K}_n(A, \mathbf{r}_0).$$

Then according to Theorem 4.1, the iterate \mathbf{x}_n is given by the vector whose associated residual is minimal in the Euclidean norm over all vectors $\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_n(A, \mathbf{r}_0)$, i.e.,

$$\|\mathbf{b} - A\mathbf{x}_n\| = \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_n(A, \mathbf{r}_0)} \|\mathbf{b} - A\mathbf{x}\|. \quad (37)$$

The approach is known as the *minimum residual approach* and can be applied to the Arnoldi process as follows. Recall that every vector $\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_n(A, \mathbf{r}_0)$ can be represented in the form (31) involving the free parameter vector \mathbf{y} . Therefore, rather than using \mathbf{x} , the minimization problem (37) can be reformulated in terms of the free parameter vector \mathbf{y} . That is, solving the minimization problem (37) in iteration n implies fixing the vector \mathbf{y} to a particular vector \mathbf{y}_n or, equivalently, choosing the subspace \mathcal{L}_n . More precisely, if the Arnoldi process is started with \mathbf{v}_1 as in (30), the residual vector associated to (31) is given by

$$\mathbf{r}_n = \|\mathbf{r}_0\| \mathbf{v}_1 - A V_n \mathbf{y} \quad \text{for some } \mathbf{y} \in \mathbb{C}^n.$$

Using the fact that \mathbf{v}_1 is the first column of V_{n+1} and making use of Theorem 5.1 yields

$$\mathbf{r}_n = V_{n+1} (\|\mathbf{r}_0\| \mathbf{e}_1 - \underline{H}_{n+1} \mathbf{y}) \quad \text{for some } \mathbf{y} \in \mathbb{C}^n, \quad (38)$$

where the symbol

$$\underline{H}_{n+1} := \begin{bmatrix} H_n \\ 0 & 0 & \dots & 0 & h_{n+1,n} \end{bmatrix} \in \mathbb{C}^{(n+1) \times n}$$

denotes the matrix obtained from H_n in (19) by adding another row at the bottom. This representation is used to rewrite (37) in the form

$$\|V_{n+1}(\|\mathbf{r}_0\|\mathbf{e}_1 - \underline{H}_{n+1}\mathbf{y}_n)\| = \min_{\mathbf{y} \in \mathbb{C}^n} \|V_{n+1}(\|\mathbf{r}_0\|\mathbf{e}_1 - \underline{H}_{n+1}\mathbf{y})\|.$$

Since the Arnoldi process generates a unitary matrix V_{n+1} and the Euclidean norm is invariant under unitary transformations, an equivalent form is given by

$$\|V_{n+1}(\|\mathbf{r}_0\|\mathbf{e}_1 - \underline{H}_{n+1}\mathbf{y}_n)\| = \min_{\mathbf{y} \in \mathbb{C}^n} \|\|\mathbf{r}_0\|\mathbf{e}_1 - \underline{H}_{n+1}\mathbf{y}\|. \quad (39)$$

Thus, in each iteration n of the resulting method, an $(n+1) \times n$ least-squares problem of type (39) is to be solved. Since the Hessenberg matrix \underline{H}_{n+1} has full rank, there is a unique solution \mathbf{y}_n . This method is known as the Generalized Minimum RESidual method (GMRES)¹⁷ whose highlevel description is given in Fig. 8. Finally, note that (39) is actually not a single least-squares problem, but defines a sequence of least-squares problems where in each step a row and a column in \underline{H}_{n+1} are appended. Moreover, its Hessenberg structure can be exploited to efficiently solve the sequence of least-squares problems with less computational cost than solving a new least-squares problem in every iteration from scratch. The corresponding technique is detailed in Saad.¹⁰

x_n = GMRES(A, b, x₀) If $A \in \mathbb{C}^{N \times N}$, this algorithm computes approximations \mathbf{x}_n to the solution of the linear system $A\mathbf{x} = \mathbf{b}$ for any starting vector \mathbf{x}_0 .

```

1:  $\mathbf{r}_0 \leftarrow \mathbf{b} - A\mathbf{x}_0$ 
2:  $\mathbf{v}_1 \leftarrow \mathbf{r}_0 / \|\mathbf{r}_0\|$ 
3: for  $n = 1, 2, 3, \dots$  do {until convergence}
4:   Step  $n$  of Arnoldi method producing  $\underline{H}_{n+1}$  and  $V_n$ 
5:   Compute minimizer  $\mathbf{y}_n$  of  $\|\|\mathbf{r}_0\|\mathbf{e}_1 - \underline{H}_{n+1}\mathbf{y}\|$ 
6:    $\mathbf{x}_n \leftarrow \mathbf{x}_0 + V_n\mathbf{y}_n$ 
7: end for

```

Figure 8. Highlevel Generalized Minimum Residual Method (GMRES).

The Quasi-Minimal Residual Approach

Suppose the Lanczos algorithm is started with two nonzero starting vectors \mathbf{v}_1 and \mathbf{w}_1 satisfying (30) and (35), respectively. Then, similar to the derivation of (38), the Lanczos algorithm summarized by Theorem 5.2 leads to

$$\mathbf{r}_n = V_{n+1}(\|\mathbf{r}_0\|\mathbf{e}_1 - \underline{T}_{n+1}\mathbf{y}) \quad \text{for some } \mathbf{y} \in \mathbb{C}^n, \quad (40)$$

where

$$\underline{T}_{n+1} := \begin{bmatrix} T_n \\ 0 & 0 & \dots & 0 & \rho_{n+1} \end{bmatrix} \in \mathbb{C}^{(n+1) \times n}.$$

Since the matrix V_{n+1} is no longer unitary in the Lanczos process, it is not possible to minimize $\|\mathbf{r}_n\|$ by an equivalent “small” problem of type (39) in a similar way as in the minimum residual approach. Rather than minimizing $\|\mathbf{r}_n\|$ which would be desirable but computationally expensive, the *quasi-minimal residual approach* minimizes a factor of the representation (40) of the residual. More precisely, the free parameter vector \mathbf{y} is fixed by

$$\|\mathbf{r}_0\|\mathbf{e}_1 - \underline{T}_{n+1}\mathbf{y}_n\| = \min_{\mathbf{y} \in \mathbb{C}^n} \|\mathbf{r}_0\|\mathbf{e}_1 - \underline{T}_{n+1}\mathbf{y}\|.$$

So, instead of $\|\mathbf{r}_n\|$, only the norm of the factor of (40) given in parentheses is minimized. The complete highlevel description of the resulting Quasi-Minimal Residual method (QMR)¹⁸ is given in Fig. 9.

x_n = QMR(A, b, x₀) If $A \in \mathbb{C}^{N \times N}$, this algorithm computes approximations \mathbf{x}_n to the solution of the linear system $A\mathbf{x} = \mathbf{b}$ for any starting vector \mathbf{x}_0 .

```

1:  $\mathbf{r}_0 \leftarrow \mathbf{b} - A\mathbf{x}_0$ 
2:  $\mathbf{v}_1 \leftarrow \mathbf{r}_0 / \|\mathbf{r}_0\|$ 
3: Choose  $\mathbf{w}_1$  such that  $\|\mathbf{w}_1\| \leftarrow 1$  and  $\mathbf{v}_1^H \mathbf{w}_1 \neq 0$ 
4: for  $n = 1, 2, 3, \dots$  do {until convergence}
5:   Step  $n$  of biorthogonal Lanczos algorithm producing  $\underline{T}_{n+1}$  and  $V_n$ 
6:   Compute minimizer  $\mathbf{y}_n$  of  $\|\mathbf{r}_0\|\mathbf{e}_1 - \underline{T}_{n+1}\mathbf{y}\|$ 
7:    $\mathbf{x}_n \leftarrow \mathbf{x}_0 + V_n\mathbf{y}_n$ 
8: end for
```

Figure 9. Highlevel Quasi-Minimal Residual Method (QMR).

Additional Remark

The presentations of the algorithms given in this survey concentrate on the underlying principles of Krylov subspace methods. These principles are useful in understanding most of the other Krylov subspace methods. The highlevel presentations are not meant to replace the study of the original papers such as the ones by Hestenes and Stiefel¹⁹ for CG, Saad and Schulz¹⁷ for GMRES, Freund and Nachtigal¹⁸ for QMR, to name just a few. Important implementation details given in these articles are vast and indispensable for efficient and professional software. In particular, a statement of the form $\mathbf{x}_n \leftarrow \mathbf{x}_0 + V_n\mathbf{y}_n$ in the highlevel presentations does *not* necessarily mean to store the complete matrix V_n , i.e., all vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.

6 Preconditioning

The convergence of an iterative method applied to a linear system depends on the properties of the coefficient matrix. With the exception of GMRES, little is known about the details of the convergence behavior for general linear systems. To achieve or accelerate the

convergence of an iterative method, a given linear system $A\mathbf{x} = \mathbf{b}$ is often transformed into an equivalent system called preconditioned system. There are right preconditioning techniques of the form

$$AM\mathbf{y} = \mathbf{b} \quad \text{and} \quad M\mathbf{y} = \mathbf{x}$$

and left preconditioners of type

$$M A\mathbf{x} = M\mathbf{b}.$$

Convergence will be fast if AM or MA are, in some sense, “close to” the identity I for right and left preconditioning, respectively. Preconditioning involves the additional work of computing the preconditioner M and the repeated solution of linear systems with M as the coefficient matrix. The requirements for an effective preconditioner are as follows:

- Linear systems with coefficient matrix M should be easy to solve. An extreme case is when $M = I$, but then convergence is not accelerated at all; essentially, M is no proper preconditioner.
- The acceleration of the convergence should be fast. The extreme case here is when $M = A^{-1}$ in which the process converges in a single step but the construction of M is as hard as solving the original system.

Effective preconditioners lie between these two extremes. Some examples of preconditioning techniques are as follows.

Diagonal scaling is a simple preconditioner where $M = \text{diag}^{-1}(A)$. Another popular preconditioning class is based on incomplete Cholesky or LU factorizations. Numerous variants of ILU preconditioners are in use. In problems arising from partial differential equations, preconditioners are constructed from coarse-grid approximations. One of the key ideas of multigrid methods is to use, as a preconditioner, one or more steps of a classical iteration such as Jacobi or Gauss–Seidel. Another well-known strategy is based on domain decomposition techniques where the idea is to have solvers for certain local sub-domains used to form a preconditioner for the overall global problem. Here, the local subdomains can be handled in parallel. Polynomial preconditioners are also interesting with respect to parallelism. Approximate inverse preconditioners may also offer a high degree of parallelism. Preconditioning is described in the books by Saad,¹⁰ Greenbaum,²⁰ Axelsson,²¹ and Meurant.²²

7 Reducing Synchronization

The parallelization of Krylov subspace methods on distributed memory processors is straightforward²³ and consists in parallelizing the three kinds of operations: vector updates, matrix-vector products, and inner products. Vector updates are perfectly parallelizable and, for large sparse matrices, matrix-vector products can be implemented with communication between only nearby processors. The bottleneck is usually due to inner products enforcing global communication, i.e., communication of all processors at the same time.

There are two strategies to remedy the performance degradation which, of course, can be combined. The first is to restructure the code such that most communication is overlapped with useful computation. The second is to eliminate data dependencies such that

several inner products can be computed simultaneously. In this section an example of the latter strategy is given where the number of global synchronization points is reduced. A global synchronization point is defined as the locus of an algorithm at which all local information has to be globally available in order to continue the computation.

Consider once more the Lanczos process depicted in Fig. 5. In a parallel implementation of the main loop, global communication is necessary for the inner products in lines 4 and 11 as well as for the norms in lines 7 and 8. Because the computation of the norms can be computed simultaneously, there are three global synchronization points per iteration.

For this algorithm, a simple reorganization of the statements is used to eliminate two of these global synchronization points. The idea is to delay the computation of the vectors \mathbf{v}_{n+1} and \mathbf{w}_{n+1} . This is easily accomplished by observing that lines 9 and 10 lead to an equivalent form of line 11 given by

$$\delta_{n+1} = \mathbf{w}_{n+1}^H \mathbf{v}_{n+1} = \frac{\tilde{\mathbf{w}}_{n+1}^H \tilde{\mathbf{v}}_{n+1}}{\xi_{n+1} \rho_{n+1}}. \quad (41)$$

Similarly, line 4 is reformulated as

$$\alpha_n = \frac{\tilde{\mathbf{w}}_n^H A \tilde{\mathbf{v}}_n}{\xi_n \rho_n \delta_n}.$$

Then, a new variant of the algorithm is given by replacing all δ_n 's by a new quantity

$$\tilde{\delta}_{n+1} := \tilde{\mathbf{w}}_{n+1}^H \tilde{\mathbf{v}}_{n+1} = \delta_{n+1} \xi_{n+1} \rho_{n+1}$$

where the last equation results from (41). This parallel variant is depicted in Fig. 10. It is more scalable than the original algorithm of Fig. 5 because the computations of the two inner products in lines 4 and 12 and the two norms in lines 10 and 11 are all independent of each other and can be computed simultaneously. Thus, in this variant of the algorithm, there is only a single global synchronization point per iteration.

Reducing synchronization cost by a simple rearrangement of the statements is possible for the Lanczos algorithm based on three-term recurrences with the option to scale both Lanczos vectors, i.e., the algorithm given in Fig. 5. However, there is a corresponding coupled two-term formulation²⁴ of the Lanczos algorithm that has a better reputation with respect to numerical stability where such a rearrangement is not immediately available. Here, algorithms have to be redesigned with parallelism in mind.^{25,26} Synchronization is also reduced in algorithms different from the Lanczos algorithm.^{27–29} It is possible to combine global synchronization points not only within a single iteration but also within multiple iterations.^{30,31} The performance of Krylov subspace methods on parallel computers is modeled by de Sturler,³² Gupta et al.,³³ and Bücker.³⁴

8 Matrix-Vector Multiplications and Graph Partitioning

Among the basic computational kernels of a Krylov subspace method, the most computationally expensive operation is typically the matrix-vector multiplication. A careful implementation of this operation is therefore important. On a parallel computer with distributed memory, a distribution of the data to processors has to be carried out where it is desirable to balance the computational load on each processor while minimizing the interprocessor communication. This can be modeled by a graph partitioning problem as follows.

$[V_n, W_n] = \text{SCABIOLANCZOS}(A, \tilde{\mathbf{v}}_1, \tilde{\mathbf{w}}_1)$ If $A \in \mathbb{C}^{N \times N}$ and $\tilde{\mathbf{v}}_1, \tilde{\mathbf{w}}_1$ are suitable starting vectors, this algorithm computes biorthogonal bases $V_n = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n] \in \mathbb{C}^{N \times n}$ and $W_n = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_n] \in \mathbb{C}^{N \times n}$ of $\mathcal{K}_n(A, \mathbf{v}_1)$ and $\mathcal{K}_n(A^H, \mathbf{w}_1)$, respectively. A single synchronization point is used.

- 1: Choose $\tilde{\mathbf{v}}_1, \tilde{\mathbf{w}}_1 \in \mathbb{C}^N$ such that $\tilde{\delta}_1 \leftarrow \tilde{\mathbf{w}}_1^H \tilde{\mathbf{v}}_1 \neq 0$
 - 2: Set $\mathbf{v}_0 = \mathbf{w}_0 \leftarrow \mathbf{0}$ and $\rho_0 = \xi_0 \leftarrow 0$, $\rho_1 \leftarrow \|\tilde{\mathbf{v}}_1\|$, $\xi_1 \leftarrow \|\tilde{\mathbf{w}}_1\|$, $\tilde{\delta}_0 \neq 0$
 - 3: **for** $n = 1, 2, 3, \dots$ **do** {until invariance}
 - 4: $\alpha_n \leftarrow \tilde{\mathbf{w}}_n^H A \tilde{\mathbf{v}}_n / \tilde{\delta}_n$
 - 5: $\gamma_n \leftarrow \xi_n \xi_{n-1} \rho_{n-1} \tilde{\delta}_n / (\xi_n \rho_n \tilde{\delta}_{n-1})$
 - 6: $\mathbf{v}_n \leftarrow \frac{1}{\rho_n} \tilde{\mathbf{v}}_n$
 - 7: $\mathbf{w}_n \leftarrow \frac{1}{\xi_n} \tilde{\mathbf{w}}_n$
 - 8: $\tilde{\mathbf{v}}_{n+1} \leftarrow \frac{1}{\rho_n} A \tilde{\mathbf{v}}_n - \alpha_n \mathbf{v}_n - \gamma_n \mathbf{v}_{n-1}$
 - 9: $\tilde{\mathbf{w}}_{n+1} \leftarrow A^H \mathbf{w}_n - \bar{\alpha}_n \mathbf{w}_n - \frac{\bar{\gamma}_n \bar{\rho}_n}{\xi_n} \mathbf{w}_{n-1}$
 - 10: $\rho_{n+1} \leftarrow \|\tilde{\mathbf{v}}_{n+1}\|$
 - 11: $\xi_{n+1} \leftarrow \|\tilde{\mathbf{w}}_{n+1}\|$
 - 12: $\tilde{\delta}_{n+1} \leftarrow \tilde{\mathbf{w}}_{n+1}^H \tilde{\mathbf{v}}_{n+1}$
 - 13: **end for**
-

Figure 10. A scalable variant of the Bi-Lanczos method.

Consider a matrix-vector multiplication of the form $\mathbf{y} = A\mathbf{x}$ where the N -dimensional vector \mathbf{y} is the result of applying the $N \times N$ coefficient matrix A to some given N -dimensional vector \mathbf{x} . Assume that the sparsity is exploited by computing the i th entry of \mathbf{y} via

$$y_i = \sum_{j \text{ with } A(i,j) \neq 0} A(i,j) \cdot x_j, \quad (42)$$

where the summation is over the nonzero elements of the i th row of A .

A graph representation of a nonsymmetric matrix with symmetric nonzero pattern is given by a set of nodes $V = \{1, 2, \dots, N\}$ where a node is associated to every row of A and a set of edges

$$E = \{(i, j) \mid A(i, j) \neq 0 \text{ for } i \neq j\}$$

used to describe the nonzero entries. An example of a matrix and its associated graph is given in Fig. 11 for $N = 10$. A data distribution on p processors, where x_i , y_i , and the i th row of A are stored on the same processor for all $1 \leq i \leq N$, may be expressed by a partition $P : V \rightarrow \{1, 2, \dots, p\}$ which decomposes the set of nodes into p subsets $V = V_1 \cup V_2 \cup \dots \cup V_p$ with $V_i \cap V_j = \emptyset$ for $i \neq j$. In Fig. 11, the partition of the graph on $p = 3$ processors is shown by three dashed lines. Here, the subset describing the data stored on processor 1 is given by $V_1 = \{1, 2, 3, 4\}$. The data distribution of the remaining processors is represented by $V_2 = \{5, 6, 7\}$ and $V_3 = \{8, 9, 10\}$.

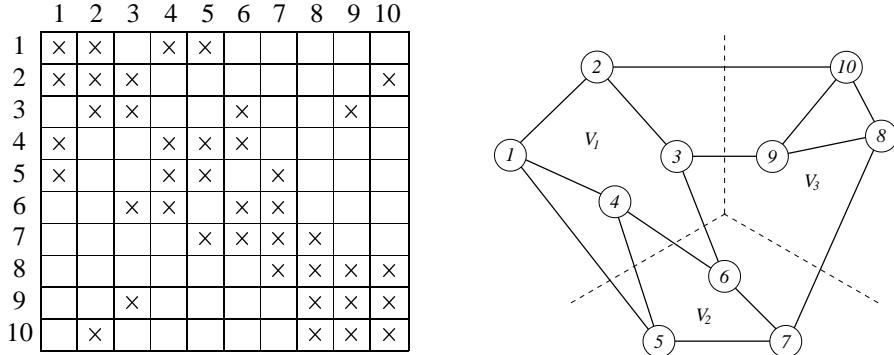


Figure 11. A nonsymmetric 10×10 matrix with a symmetric nonzero pattern (left) and its associated graph with partitions V_1 , V_2 and V_3 (right).

A reformulation of (42) in terms of graph terminology is

$$y_i = a_{ii}x_i + \sum_{\substack{(i,j) \in E \\ P(i)=P(j)}} A(i,j) \cdot x_j + \sum_{\substack{(i,j) \in E \\ P(i) \neq P(j)}} A(i,j) \cdot x_j.$$

Here, the first two terms can be computed on processor i without communication to any other processor. The condition $P(i) \neq P(j)$ in the last term shows that its computation requires communication between processor i and processor j .

Suppose that the number of nonzeros is roughly the same for each row of A . Then, the number of arithmetic operations of a matrix-vector multiplication Ax is well-balanced among the p processors if the partition P satisfies

$$|V_1| \approx |V_2| \approx \dots \approx |V_p|.$$

Moreover, a rough measure of minimizing interprocessor communication is to find a partition P that minimizes the number of edges (i, j) with $P(i) \neq P(j)$, i.e., those edges connecting nodes in different subsets of the partitions. The number of edges whose end nodes have been mapped to different processors is called the cut size.

Finding a partition balancing the number of nodes while minimizing its cut size is known as the graph partitioning problem. This problem is a hard combinatorial problem known as NP-hard.³ Thus, it is unlikely that there is a deterministic polynomial-time algorithm that always finds an optimal partition. Since the graph partitioning problem—or more general formulations involving node and edge weights—is of interest in a variety of areas, a large number of heuristics have been developed; see the surveys by Fjällström,³⁵ Schloegel et al.,³⁶ and Hendrickson and Kolda.³⁷

9 Concluding Remarks

Large sparse linear systems arise frequently in different areas of scientific computing. Due to their sheer size, parallel computing is often mandatory. With the ever-increasing computational performance and storage capacity of the computer technology, the order of the linear systems also increase at a noticeable speed: What today seems a large system is likely to be considered to be small in a few years. Therefore, the capability of exploiting structure and/or sparsity of the coefficient matrix is a crucial ingredient to any computational technique for the solution of linear systems from real-world applications.

Direct methods such as Gaussian elimination or Cholesky factorization may lead to excessive fill-in for large sparse matrices. By making use of the coefficient matrix solely in the form of matrix-vector multiplication, iterative methods do not suffer from the fill-in problem. Classical iterative methods such as Jacobi or Gauss–Seidel iteration typically do not converge fast enough but are useful as building blocks in more efficient techniques. Krylov subspace methods are currently considered to be among the most powerful iterative techniques. Prominent examples include the conjugate gradient (CG) method and the generalized minimum residual (GMRES) method. Preconditioning is an important mechanism to accelerate the convergence of Krylov subspace methods.

When large sparse systems are iteratively solved on parallel computers, a number of additional issues arise. If the number of processors is large, performance is usually decreased by synchronization involved in the computation of inner product-like operations. The cost of synchronization may sometimes be significantly reduced by small rearrangements of some given parallel implementation of a serial algorithm. However, there is more improvement to be expected if a new algorithm is designed from scratch with parallelism already in mind. Graph partitioning can be used to efficiently perform a matrix-vector multiplication in parallel. Here, the idea behind graph partitioning is to balance the computational work while minimizing interprocessor communication.

10 Bibliographic Comments

Thousands of papers have been written on direct methods for the solution of linear systems. The classic book by Wilkinson³⁸ is an early reference including a careful study of Gaussian elimination with respect to rounding errors. The more general field of matrix computations is treated in a standard textbook by Golub and van Loan.⁹ Direct methods exploiting sparsity are described in the books by George and Liu,³⁹ Duff et al.,⁴⁰ Osterby and Zlatev,⁴¹ and Pissanetzky.⁴² The books by Varga⁴³ and Young⁴⁴ started the study of classical iterative methods. Modern iterative methods including preconditioning are described in the books by Fischer,¹ Greenbaum,²⁰ Saad,¹⁰ Axelsson,²¹ and Meurant.²² Iterative methods are surveyed in the papers by Freund et al.⁴⁵ and by Gutknecht.¹⁴

Multigrid methods are an important class of modern techniques for the solution of linear systems. They are omitted in this survey simply to limit the discussion. Multigrid methods can be viewed as a combination of an iterative scheme and a preconditioner. A seminal paper in the area of multigrid methods is the one by Brandt.⁴⁶ A short introduction to multigrid methods is given in the book by Briggs.⁴⁷ Additional material is described in the books by Hackbusch,⁴⁸ Hackbusch and Trottenberg,⁴⁹ and Briggs et al.⁵⁰

An excellent starting point to parallel computing in general is the book by Kumar et al.⁵¹ The two articles by Demmel⁵² and Demmel et al.⁵³ give a survey on parallel numerical algorithms. A more recent survey on parallel techniques for the solution of linear systems, both direct and iterative, is given by Duff and van der Vorst.⁵⁴ The paper by Saad⁵⁵ concentrates on parallel iterative methods.

Acknowledgments

The author would like to thank the organizing team at Forschungszentrum Jülich as well as the Scientific Programme Committee for carrying out a successful Winter School at Rolduc. I am also indebted to F. Hoßfeld and the whole team at the Central Institute for Applied Mathematics, Forschungszentrum Jülich, for making available a professional environment for parts of my research.

References

1. B. Fischer. *Polynomial Based Iteration Methods for Symmetric Linear Systems*. Advances in Numerical Mathematics. Wiley and Teubner, Chichester, 1996.
2. J. J. Dongarra, I. S. Duff, and H. A. van der Vorst. *Numerical Linear Algebra for High-Performance Computers*. SIAM, Philadelphia, 1998.
3. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, San Francisco, 1979.
4. A. Natale, R. Shamir, and R. Sharan. A polynomial approximation algorithm for the minimum fill-in problem. *SIAM Journal on Computing*, 30(4):1067–1079, 2000.
5. Andreas Griewank. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, 2000.
6. M. Berz, C. Bischof, G. Corliss, and A. Griewank. *Computational Differentiation: Techniques, Applications, and Tools*. SIAM, Philadelphia, 1996.
7. George Corliss, Christèle Faure, Andreas Griewank, Laurent Hascoët, and Uwe Naumann, editors. *Automatic Differentiation of Algorithms: From Simulation to Optimization*. Springer, 2002. (to appear).
8. A. Griewank and G. Corliss. *Automatic Differentiation of Algorithms*. SIAM, Philadelphia, 1991.
9. G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, third edition, 1996.
10. Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston, 1996.
11. N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM Publications, Philadelphia, PA, 1996.
12. W. E. Arnoldi. The Principle of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem. *Quarterly of Applied Mathematics*, 9:17–29, 1951.
13. C. Lanczos. An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators. *Journal of Research of the National Bureau of Standards*, 45(4):255–282, 1950.

14. M. H. Gutknecht. Lanczos-Type Solvers for Nonsymmetric Linear Systems of Equations. In *Acta Numerica 1997*, pages 271–397. Cambridge University Press, Cambridge, 1997.
15. V. Faber and T. Manteuffel. Necessary and Sufficient Conditions for the Existence of a Conjugate Gradient Method. *SIAM Journal on Numerical Analysis*, 21(2):352–362, 1984.
16. C. Lanczos. Solutions of Systems of Linear Equations by Minimized Iterations. *Journal of Research of the National Bureau of Standards*, 49(1):33–53, 1952.
17. Y. Saad and M. H. Schulz. GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986.
18. R. W. Freund and N. M. Nachtigal. QMR: A Quasi-Minimal Residual Method for Non-Hermitian Linear Systems. *Numerische Mathematik*, 60(3):315–339, 1991.
19. M. Hestenes and E. Stiefel. Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, 49:409–436, 1952.
20. A. Greenbaum. *Iterative Methods for Solving Linear Systems*. SIAM, Philadelphia, PA, 1997.
21. O. Axelsson. *Iterative Solution Methods*. Cambridge University Press, Cambridge, 1994.
22. G. Meurant. *Computer Solution of Large Linear Systems*, volume 28 of *Studies in Mathematics and Its Applications*. North-Holland, Amsterdam, 1999.
23. Y. Saad. Krylov Subspace Methods on Supercomputers. *SIAM Journal on Scientific and Statistical Computing*, 10(6):1200–1232, 1989.
24. R. W. Freund and N. M. Nachtigal. An Implementation of the QMR Method Based on Coupled Two-Term Recurrences. *SIAM Journal on Scientific Computing*, 15(2):313–337, 1994.
25. H. M. Bücker and M. Sauren. A Parallel Version of the Quasi-Minimal Residual Method Based on Coupled Two-Term Recurrences. In J. Waśniewski, J. Dongarra, K. Madsen, and D. Olesen, editors, *Applied Parallel Computing: Industrial Computation and Optimization, Proceedings of the Third International Workshop, PARA '96, Lyngby, Denmark, August 18–21, 1996*, volume 1184 of *Lecture Notes in Computer Science*, pages 157–165, Berlin, 1996. Springer.
26. H. M. Bücker and M. Sauren. A Variant of the Biconjugate Gradient Method Suitable for Massively Parallel Computing. In G. Bilardi, A. Ferreira, R. Lüling, and J. Rolim, editors, *Solving Irregularly Structured Problems in Parallel, Proceedings of the Fourth International Symposium, IRREGULAR'97, Paderborn, Germany, June 12–13, 1997*, volume 1253 of *Lecture Notes in Computer Science*, pages 72–79, Berlin, 1997. Springer.
27. E.F. D'Azevedo, V.L. Eijkhout, and C.H. Romine. Lapack Working Note 56: Reducing Communication Costs in the Conjugate Gradient Algorithm on Distributed Memory Multiprocessors. Technical Report CS-93-185, University of Tennessee, Knoxville, 1993.
28. G. Meurant. Multitasking the Conjugate Gradient Method on the CRAY X-MP/48. *Parallel Computing*, 5:267–280, 1987.
29. Y. Saad. Practical Use of Polynomial Preconditionings for the Conjugate Gradient

- Method. *SIAM Journal on Scientific and Statistical Computing*, 6(4):865–881, 1985.
30. A. T. Chronopoulos and C. W. Gear. *s*-Step Iterative Methods for Symmetric Linear Systems. *Journal of Computational and Applied Mathematics*, 25:153–168, 1989.
 31. A. T. Chronopoulos and C. D. Swanson. Parallel Iterative S-step Methods for Unsymmetric Linear Systems. *Parallel Computing*, 22(5):623–641, 1997.
 32. E. de Sturler. A Performance Model for Krylov Subspace Methods on Mesh-Based Parallel Computers. *Parallel Computing*, 22:57–74, 1996.
 33. A. Gupta, V. Kumar, and A. Sameh. Performance and Scalability of Preconditioned Conjugate Gradient Methods on Parallel Computers. Technical Report TR 92–64, Department of Computer Science, University of Minnesota, Minneapolis, MN – 55455, November 1992. Revised April 1994.
 34. H. M. Bücker. Using the Isoefficiency Concept for the Design of Krylov Subspace Methods. In M. H. Hamza, editor, *Proceedings of the IASTED International Conference on Applied Simulation and Modelling, Marbella, Spain, September 4–7, 2001*, pages 404–411, Anaheim, CA, USA, 2001. ACTA Press.
 35. P.-O. Fjällström. Algorithms for graph partitioning: a survey. *Linköping Electronic Articles in Computer and Information Science*, 3(10), 1998.
 36. K. Schloegel, G. Karypis, and V. Kumar. Graph partitioning for high-performance scientific simulations. In J. Dongarra, I. Foster, G. Fox, K. Kennedy, L. Torczon, and A. White, editors, *CRPC Parallel Computing Handbook*. Morgan Kaufmann, to appear.
 37. B. Hendrickson and T. G. Kolda. Graph partitioning models for parallel computing. *Parallel Computing*, 26(2):1519–1534, 2000.
 38. J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, 1965.
 39. A. George and J. W. H. Liu. *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1981.
 40. I. S. Duff, A. M. Erisman, and J. K. Reid. *Direct Methods for Sparse Matrices*. Clarendon Press, Oxford, 1986.
 41. O. Osterby and Z. Zlatev. *Direct Methods for Sparse Matrices*. Springer, New York, 1983.
 42. S. Pissanetzky. *Sparse Matrix Technology*. Academic Press, New York, 1984.
 43. R. S. Varga. *Matrix Iterative Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1962.
 44. D. M. Young. *Iterative Solution of Large Linear Systems*. Academic Press, New York, 1971.
 45. R. W. Freund, G. H. Golub, and N. M. Nachtigal. Iterative Solution of Linear Systems. In *Acta Numerica 1992*, pages 1–44. Cambridge University Press, Cambridge, 1992.
 46. A. Brandt. Multilevel adaptive solutions to boundary value problems. *Mathematics of Computation*, 31:333–390, 1977.
 47. W. L. Briggs. *A Multigrid Tutorial*. SIAM, Philadelphia, 1987.
 48. W. Hackbusch. *Multigrid Methods and Applications*. Springer, Berlin, 1985.
 49. W. Hackbusch and U. Trottenberg. *Multigrid Methods*. Springer, Berlin, 1982.
 50. W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*. SIAM, Philadelphia, second edition, 2000.
 51. V. Kumar, A. Grama, A. Gupta, and G. Karypis. *Introduction to Parallel Computing: Design and Analysis of Algorithms*. Benjamin/Cummings, Redwood City, 1994.

52. J. W. Demmel. Trading Off Parallelism and Numerical Stability. In M. S. Moonen, G. H. Golub, and B. L. R. De Moor, editors, *Linear Algebra for Large Scale and Real-Time Applications*, volume 232 of *NATO ASI Series E: Applied Sciences*, pages 49–68. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993. Proceedings of the NATO Advanced Study Institute on Linear Algebra for Large Scale and Real-Time Applications, Leuven, Belgium, August 1992.
53. J. W. Demmel, M. T. Heath, and H. A. van der Vorst. Parallel Numerical Linear Algebra. In *Acta Numerica 1993*, pages 111–197. Cambridge University Press, Cambridge, 1993.
54. I. S. Duff and H. A. van der Vorst. Developments and Trends in the Parallel Solution of Linear Systems. *Parallel Computing*, 25(13–14):1931–1970, 1999.
55. Y. Saad. Parallel Iterative Methods for Sparse Linear Systems. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, volume 8 of *Studies in Computational Mathematics*, pages 423–440. Elsevier Science, Amsterdam, The Netherlands, 2001.

Already published:

Modern Methods and Algorithms of Quantum Chemistry - Proceedings

Johannes Grotendorst (Editor)

NIC Series Volume 1

Winterschool, 21 - 25 February 2000, Forschungszentrum Jülich

ISBN 3-00-005618-1, February 2000, 562 pages

**Modern Methods and Algorithms of Quantum Chemistry -
Poster Presentations**

Johannes Grotendorst (Editor)

NIC Series Volume 2

Winterschool, 21 - 25 February 2000, Forschungszentrum Jülich

ISBN 3-00-005746-3, February 2000, 77 pages

**Modern Methods and Algorithms of Quantum Chemistry -
Proceedings, Second Edition**

Johannes Grotendorst (Editor)

NIC Series Volume 3

Winterschool, 21 - 25 February 2000, Forschungszentrum Jülich

ISBN 3-00-005834-6, December 2000, 638 pages

**Nichtlineare Analyse raum-zeitlicher Aspekte der
hirnelektrischen Aktivität von Epilepsiepatienten**

Jochen Arnold

NIC Series Volume 4

ISBN 3-00-006221-1, September 2000, 120 pages

**Elektron-Elektron-Wechselwirkung in Halbleitern:
Von hochkorrelierten kohärenten Anfangszuständen
zu inkohärentem Transport**

Reinhold Lövenich

NIC Series Volume 5

ISBN 3-00-006329-3, August 2000, 145 pages

**Erkennung von Nichtlinearitäten und
wechselseitigen Abhängigkeiten in Zeitreihen**

Andreas Schmitz

NIC Series Volume 6

ISBN 3-00-007871-1, May 2001, 142 pages

Multiparadigm Programming with Object-Oriented Languages

Kei Davis, Yannis Smaragdakis, Jörg Striegnitz (Editors)

NIC Series Volume 7

Proceedings, Workshop MPOOL, 18 May 2001, Budapest

ISBN 3-00-007968-8, June 2001, 160 pages

Europhysics Conference on Computational Physics

Friedel Hossfeld, Kurt Binder (Editors)

NIC Series Volume 8

Book of Abstracts, 5 - 8 September 2001, Aachen

ISBN 3-00-008236-0, September 2001, 500 pages

NIC Symposium 2001 (in preparation)

Horst Rollnik, Dietrich Wolf (Editors)

NIC Series Volume 9

Proceedings, 5 - 6 December 2001, Jülich, Germany

ISBN 3-00-009055-X

All volumes are available online at <http://www.fz-juelich.de/nic-series/>.