

Springer
Proceedings in Physics 45



Springer Proceedings in Physics

Managing Editor: H.K. V. Lotsch

- Volume 30 *Short-Wavelength Lasers and Their Applications*
Editor: C. Yamanaka
- Volume 31 *Quantum String Theory* Editors: N. Kawamoto and T. Kugo
- Volume 32 *Universalities in Condensed Matter*
Editors: R. Jullien, L. Peliti, R. Rammal, and N. Boccara
- Volume 33 *Computer Simulation Studies in Condensed Matter Physics: Recent Developments*
Editors: D. P. Landau, K. K. Mon, and H.-B. Schüttler
- Volume 34 *Amorphous and Crystalline Silicon Carbide and Related Materials*
Editors: G. L. Harris and C. Y.-W. Yang
- Volume 35 *Polycrystalline Semiconductors: Grain Boundaries and Interfaces*
Editors: H. J. Möller, H. P. Strunk, and J. H. Werner
- Volume 36 *Nonlinear Optics of Organics and Semiconductors*
Editor: T. Kobayashi
- Volume 37 *Dynamics of Disordered Materials*
Editors: D. Richter, A. J. Dianoux, W. Petry, and J. Teixeira
- Volume 38 *Electroluminescence* Editors: S. Shionoya and H. Kobayashi
- Volume 39 *Disorder and Nonlinearity*
Editors: A. R. Bishop, D. K. Campbell, and S. Pnevmatikos
- Volume 40 *Static and Dynamic Properties of Liquids*
Editors: M. Davidović and A. K. Soper
- Volume 41 *Quantum Optics V* Editors: J. D. Harvey and D. F. Walls
- Volume 42 *Molecular Basis of Polymer Networks*
Editors: A. Baumgärtner and C. E. Picot
- Volume 43 *Amorphous and Crystalline Silicon Carbide II: Recent Developments*
Editors: M. M. Rahman, C. Y.-W. Yang, and G. L. Harris
- Volume 44 *Optical Fiber Sensors*
Editors: H. J. Arditty, J. P. Dakin, and R. Th. Kersten
- Volume 45 *Computer Simulation Studies in Condensed Matter Physics II: New Directions*
Editors: D. P. Landau, K. K. Mon, and H.-B. Schüttler
- Volume 46 *Cellular Automata and Modeling of Complex Physical Systems*
Editors: P. Manneville, N. Boccara, G. Y. Vichniac, and R. Bidaux

Volumes 1 – 29 are listed on the back inside cover

Computer Simulation Studies in Condensed Matter Physics II

New Directions

**Proceedings of the Second Workshop,
Athens, GA, USA, February 20–24, 1989**

**Editors: D.P. Landau, K.K. Mon,
and H.-B. Schüttler**

With 88 Figures

**Springer-Verlag Berlin Heidelberg New York
London Paris Tokyo Hong Kong**

Professor David P. Landau, Ph.D.

Professor K.K. Mon, Ph.D.

Professor Heinz-Bernd Schüttler, Ph.D.

Center for Simulational Physics, The University of Georgia, Athens, GA 30602, USA

ISBN-13:987-3-642-75236-0 e-ISBN-13:978-3-642-75234-6

DOI: 10.1007/978-3-642-75234-6

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. Duplication of this publication or parts thereof is only permitted under the provisions of the German Copyright Law of September 9, 1965, in its version of June 24, 1985, and a copyright fee must always be paid. Violations fall under the prosecution act of the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1990

Softcover reprint of the hardcover 1st edition 1990

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

2154/3150-543210 – Printed on acid-free paper

Preface

Computer simulations are becoming increasingly important in condensed matter physics, making significant contributions in new approaches to and in new results for important physical problems. The workshop on “Recent Developments in Computer Simulation Studies in Condensed Matter Physics”, held at the Center for Simulational Physics at the University of Georgia, February 20–24, 1989, brought together experienced practitioners in this field, as well as students and faculty members who are new to simulations, and provided a forum for the presentation and exchange of new ideas and recent developments. These proceedings are a record of the workshop and are published with the goal of timely dissemination of the papers to a wider audience.

Although by their very nature workshops are limited in their scope of coverage, a broad range of current topics is discussed. The papers present new techniques and results on both static and dynamical phenomena of classical and quantum mechanical systems. The reader will find descriptions of studies of static properties as well as time-dependent ones where time may be real time, stochastic time or imaginary (path integral) time.

A substantial portion of the material, the first part of the proceedings, deals with simulations of classical systems. Sophisticated implementation of vector algorithms is reviewed and new approaches using parallel systems are discussed. New methods for analyzing Monte Carlo data are presented as well as new applications of the Monte Carlo method itself, e.g., damage spreading. Both standard and novel Monte Carlo methods have been used to study static properties at phase transitions and new material is presented for the formation of aggregates. Molecular dynamics and/or Monte Carlo approaches have been developed for studying fluid flow as well as for polymer systems.

The second part of the proceedings is devoted to the simulation of quantum systems. New algorithms and results for interacting fermion systems are presented with emphasis on models which are relevant to superconductivity and magnetism. The workshop also included several presentations about emerging computer technologies and programming methods but for the most part no written record is available.

We hope that the readers will benefit from papers in their own particular field as well as gain new ideas from other related specialities.

This workshop was made possible through the generous support of the Center for Simulational Physics and the Vice President for Research at the University of Georgia.

Athens, GA, April 1989

D.P. Landau · K.K. Mon · H.-B. Schüttler

Contents

New Directions: An Introduction By D.P. Landau, K.K. Mon, and H.-B. Schüttler	1
--	---

Part I Classical Systems

Thoughts on Vectorized Algorithms for Molecular Dynamics By D.C. Rapaport	4
Parallelization of Computational Physics Algorithms By D.W. Heermann and A.N. Burkitt (With 6 Figures)	16
Classification and Exponents of Cellular Automata By D. Stauffer (With 3 Figures)	26
Mass Multifractality in Cluster Growth Models By T. Vicsek (With 4 Figures)	34
Noise Reduction in DLA By M.A. Novotny (With 8 Figures)	45
Damage Spreading By H.J. Herrmann (With 10 Figures)	56
Molecular Dynamics Simulations of Proteins and Protein-Protein Complexes By J.E. Wampler, D.E. Stewart, and S.L. Gallion (With 5 Figures)	68
Computer Simulation of Rotating Molecules in Condensed Phases By R.W. Gerling (With 12 Figures)	85
Novel Contexts of Computer Simulation By H.L. Frisch (With 9 Figures)	99

Part II Quantum Systems

Monte Carlo Studies of Many-Electron Systems By R.L. Sugar (With 11 Figures)	116
Universal Reduction of T_c in Strong Coupling Superconductors due to Anderson/Kondo Impurities By M. Jarrel (With 4 Figures)	137

Part III Contributed Papers

New Interfacial Phenomena in a Driven Diffusive System By J.L. Vallés, K.-t. Leung, and R.K.P. Zia (With 3 Figures)	152
Spin Dynamics of Cuprate Superconductors: Exact Results from Numerical Continued Fraction Expansions By C.-X. Chen and H.-B. Schüttler (With 3 Figures)	157
Statistical Mechanics in Optimization Problems By Pik-Yin Lai	163
Structural Anisotropy of Aggregation Clusters By H.M. Lindsay, R. Klein, D.A. Weitz, M.Y. Lin, and P. Meakin (With 2 Figures)	169
Monte Carlo Studies of Spinodal Decomposition in the 2d Kinetic Ising Model By J.G. Amar (With 5 Figures)	173
Histogram Methods for Monte Carlo Data Analysis By R.H. Swendsen and A.M. Ferrenberg	179
Theoretical Models for Adsorption and Diffusion of In on Cu(100) By M.S.S. Challa, Yi Li, M.R. Press, S.N. Khanna, P. Jena, and M. Yousouff (With 3 Figures)	184
Index of Contributors	189

New Directions: An Introduction

D.P. Landau, K.K. Mon, and H.-B. Schüttler

Center for Simulational Physics, University of Georgia, Athens, GA 30602, USA

This year's workshop includes papers which deal with a number of different rapidly developing aspects of computer simulations in condensed matter physics. The invited talks include rather long papers with some pedagogical content and at the end of this volume we present a few shorter, contributed papers which represent very recent research results.

The first two "long" papers deal primarily with techniques for implementing simulations on modern high performance machines. Rapaport first deals with implementing algorithms on vector machines. The problem of molecular dynamics simulations is approached from different directions and the relative advantages and disadvantages are discussed. Heermann and Burkitt show instead how to best prepare a simulation for a parallel processing and present examples, with performance tests, for simple Ising models and for cluster counting, on an array of transputers.

Stauffer considers all possible simple two dimensional cellular automata, and by simulating them using many different starting configurations for each rule determines what fraction of them may lead to chaotic behavior.

Two papers are related to growth and aggregation. Vicsek discusses multifractality in cluster growth models and presents numerical results for generalized dimensions. Novotny reviews recent applications of noise reduction in Diffusion Limited Aggregation (DLA) models and shows that in some cases noise reduction can change the asymptotic behavior of DLA-type aggregates.

In the next paper Herrmann discusses the dynamics of Monte Carlo simulations in terms of the concept of damage spreading. He shows that there are quite unexpected differences in the dynamical behavior of the damage spreading, arising in apparently very similar updating methods such as Glauber and heat bath updating in the two dimensional Ising model, for example. He discusses novel dynamical phase transitions which can be identified in terms of the damage spreading concept. An application of these ideas in the context of spin glass simulations is also discussed.

The paper by Frisch reviews applications of simulational methods in two novel contexts. The first considers extraction of a dynamical relaxation time from a biased sampling Monte Carlo computation on equilibrium data. The second deals with the approximate representation of continuous functions of many variables of physical interest in terms of continuous functions of one variable.

The paper by Gerling reports on applications of the Langevin formalism to simulate nonlinear dynamical phenomena in solids. He has studied the rotational motion of the NH₄-molecule in crystalline NH₄Cl using realistic crystalline model potentials. The simulations give detailed insight into the relaxation behavior of this rather complicated dynamical system and help to explain the anomalous relaxation times observed in this material. Simulations on a simple model for the coupling between rotational motion and lattice vibrations are also discussed.

The paper by Wampler *et al.* provides an introduction and discussion of recent advances in the application of molecular dynamics techniques to the study of protein complexes considered at the atomic level. The essential role of supercomputing and the success of the simulational methods used are discussed.

During the past twelve months significant advances have been made in the development of Monte Carlo techniques for strongly correlated lattice fermion systems at low temperature. Sugar reports on the recent applications of the new methods to the two dimensional Hubbard model. Of particular interest in the context of high-temperature superconductivity are the results for magnetic correlations, single particle momentum distributions, and superconducting pairing correlations. Jarrell discusses an interesting application of combined Monte Carlo simulation and diagrammatic techniques to the study of superconducting T_c suppression by

magnetic impurities. His results appear to resolve longstanding discrepancies in the values of the Kondo temperature of magnetic impurity doped superconductors as obtained by resistivity and from T_c suppression data.

A wide variety of topics is addressed in a number of contributed papers. Valles *et al.* reported novel interfacial phenomena in a driven diffusive system with shifter periodic boundary conditions. A critical angle for the boundary condition shift is found beyond which the system splits into a multistrip configuration. Lai considers using concepts and techniques from statistical mechanics in treating combinatorial optimization problems. The NP-complete problem of graph coloring is used as an illustration. Lindsay *et al.* report on structural anisotropy of aggregation clusters and argue that the fractal description of the structure of aggregation clusters is incomplete because it does not encompass the anisotropy of the clusters. Ferrenberg and Swendsen discuss new methods for optimizing the analysis of data from Monte Carlo simulations and demonstrate their utility with several examples. Amar discusses Monte Carlo studies of spinodal decomposition in the two dimensional Ising model, and Challa *et al.* present simulations of the kinetics of adsorbed In atoms on Cu(100) surfaces. Chen and Schüttler report exact numerical results for the spin dynamics of antiferromagnetic quantum Heisenberg and Hubbard model systems.

Part I

Classical Systems

Thoughts on Vectorized Algorithms for Molecular Dynamics

D.C. Rapaport

Physics Department, Bar-Ilan University, Ramat-Gan, Israel

The nature of the algorithms used in molecular dynamics simulation as well as the benefits and limitations of vector processing are reviewed, and methods for efficiently matching the computations to the computers are discussed.

1. Introduction

Molecular dynamics simulation addresses the problem of understanding the equilibrium and non-equilibrium properties of substances at the most detailed microscopic level. The inability of theoretical physics to deal with interacting many-particle systems without resorting to approximation methods, the range of validity of which is very often limited, and with implications that are in many instances difficult to assess, has resulted in a situation where simulation methods -- the Monte Carlo approach for equilibrium properties and the Molecular Dynamics approach for both equilibrium and non-equilibrium properties -- are often the only ways of studying the detailed behavior of such systems. Of these two approaches, both of which are in widespread use (in either the basic form or in a variety of extensions of the basic approach), it is the molecular dynamics method that offers the greater scope for more thorough exploration; this is due to the fact that the aim of the Monte Carlo method is to sample phase space without addressing the dynamical aspects of the behavior. However, in its present form, the molecular dynamics approach is limited to classical systems, and an efficient way of dealing with quantum many-body systems has yet to be proposed. The basic methodology and selected applications of the various approaches to simulation are described in refs. /1/ and /2/.

The subject of this lecture is the design considerations involved in implementing the different kinds of molecular dynamics algorithms on modern vector supercomputers. While the emphasis will be on the molecular dynamics approach, some of the underlying ideas may well prove to be applicable to Monte Carlo simulation as well.

It is well known among those who have had even the least exposure to developing programs for vector computers, that algorithms designed to operate efficiently on more conventional scalar (non-vector) computers may prove to be significantly less efficient on vector machines -- in the sense that the potential performance improvement that is claimed by the manufacturer for certain kinds of computations is far from realized in others. The hardware design of vector computers /3/

requires that data be organized in specific ways in order to permit the processor to operate anywhere near the peak design speed; failing this, the performance can be seriously degraded and can indeed be little better than a scalar computer costing a great deal less.

The performance gain arising from the ability to use the vector hardware efficiently can on occasion exceed an order of magnitude. The difficulty faced by supercomputer users at large is that there is no standardized architecture for vector computers, and the scheme developed to produce optimal performance on one machine may prove less effective, or even completely unfeasible, on another kind of machine. Even more subtle is the case of processors ostensibly having the same architecture (a term such as software-compatible is frequently used to characterize this similarity), but with different numbers of independent functional units that are capable of operating in parallel; while the internal structure of the computer is often hidden from the user, for some kinds of computations this too can have a significant effect on performance and thus also influence the choice of an optimal algorithm.

In this lecture we will discuss some of the issues involved in trying to tailor molecular dynamics algorithms to various kinds of vector supercomputer. Only a subset of the possible design options (those that are applicable to ETA-10 processors and their predecessor the Cyber 205, and to the Cray line of computers) will be considered; the reason for this limitation is obvious, since investigating the possible options for each computer is a non-trivial exercise, requiring a certain degree of familiarity with the capabilities of individual machines. There is of course no guarantee that even after a certain level of efficiency has been achieved, additional information about lesser known features of the machine could lead to further improvements in performance.

Another approach to performance enhancement in the supercomputer field is the use of parallelism -- many processors pooling their efforts to solve a single problem; while there are some aspects that are common to both vector and parallel processing, the latter introduces a whole new set of considerations, and so the subject will not be covered here. Yet another possibility, which will also not be addressed, is the design of a special-purpose processor whose capabilities are matched to the particular computation at hand; this is a comparatively new approach whose potential for cost-effective simulation is considerable.

2. Classes of problem studied by molecular dynamics

The kinds of system that are modeled by molecular dynamics simulation can be classified on the basis of several essentially independent characteristics, each of which influences both the complexity of the computational algorithm as well as the nature and extent of the computations required. These characteristics include whether the substance being modeled is in the solid or fluid state; whether the interaction range extends over minimal distance (for example, only as far as the

mean nearest neighbor separation), or reaches to moderately, or even widely separated molecules; whether the interactions are easily evaluated (soft spheres are one example of a system with simple interactions) or, at the other extreme, consist of many contributions stemming from several sets of pairs of sites on either molecule; whether the system contains only a relatively small number of molecules, or is substantially larger (the notion of what constitutes a 'large' system changes with epoch). The principal effects of each of these characteristics will be addressed in the paragraphs that follow.

There are a number of other features that are also relevant to the specification of the system and the details of how the simulation is implemented. Examples include whether the dynamical equations of motion are Newtonian, or are based on more general methods (such as isothermal or otherwise constrained dynamics); whether, in the case of a polyatomic molecular system, the molecules maintain their individual internal structures by means of relatively stiff interatomic interactions, or are modeled by using rigid geometric constraints. These features are of course central to the simulation, as is, for example, the nature of the numerical technique used for integrating the equations of motion; however, these features generally have only minor impact on the way the computational algorithm would be designed to accommodate the demands of a particular vector computer.

The distinction between simulations involving substances in the solid and fluid states is the essentially fixed neighbor relationships of the particles (which could equally well be atoms, ions or molecules) in the former versus the continually changing local spatial organization of the latter. In the solid state, the identities of the interacting neighbor pairs need be determined just once, and this determination remains valid for the duration of the simulation. In the liquid state the situation is very different, since the identities of the neighbors of any given molecule vary in an unsystematic way, and the problem of identifying all pairs of molecules that lie within interaction range can involve a significant amount of computational effort.

Depending on the nature of the system being modeled, the interaction range can vary from practically zero in the case of a fluid of 'billiard balls', to an effectively infinite value for a system of charged particles or gravitating masses. In the case of long-range interactions, all pairs of particles (be they atomic or stellar) interact, so the need for determining neighbors does not arise if all interactions are treated on an equal footing; there do exist techniques for evaluating the interactions between distant particles in a less precise fashion (thereby reducing the amount of work required) while correctly dealing with nearby particles /4/ -- then the problem of identifying nearby particles becomes important just as in the case of a fluid with short-range forces. With long-range forces there is, however, another kind of problem that concerns the handling of periodic boundaries (used to reduce finite-size effects whenever the system simulated is regarded as representing just a small, but typical part of a much larger entity); the resolution of this problem usually involves modifying the interactions (based on the computation of Ewald sums) to account for the periodic images of the system /5/. At the other extreme case of short-range interactions, the determination of which pairs of molecules interact

can, if the interactions are comparatively simple in form, require even more computation than the actual evaluation of the forces on the molecules. The scope for innovation is greatest for the short-range case, because it is here that the greatest effort is required to achieve effective vector performance /6/. With fluids in which the interaction range is intermediate -- typically a few times the mean particle separation -- both the determination of the interacting pairs and the interaction computations proper can contribute significantly to the overall effort.

The complexity of the interactions between molecules determines the amount of effort that must be invested in computing the forces, as a fraction of the total computation. Any model of a molecule that is more complex than monatomic involves several spatially separated interaction sites; the total interaction (which, for use in the equations of motion, is eventually expressed as a combination of a force acting at the center of mass and a torque) between two molecules is the sum over all pairs of sites on the two molecules, each term of which requires the evaluation of a function whose argument is the separation between the sites. In such a situation it is generally true (at least for the system sizes considered to date) that the interaction computations dominate. If, on the other hand, the interactions are between structureless particles, and, in addition to having a simple functional form, they are only short range in nature, then other parts of the computation, in particular the identification of interacting pairs, can contribute substantially to the overall work required by the simulation.

The number of particles in the system is also a factor in the algorithm design. For the more complex kinds of interactions, as well as for long-range interactions, the limits on presently available computer power dictate that the size of system be kept comparatively small; current levels may reach several hundred particles, although for long-range forces it is possible to tackle significantly larger systems (tens or hundreds of thousands of particles) if, as suggested earlier, approximate methods of representing interactions between distant particles are used (such as computing the field corresponding to a particular configuration only at a discrete set of grid points and then using interpolation, instead of directly computing the forces at the actual particle locations). For simple short-range forces, much larger systems ranging in size from one to a hundred thousand particles (and even more) can be handled.

The data organizational requirements affect each of these types of problems in different ways. As will become clear in the course of this lecture, efficient vectorization can require modification of each of the algorithms along very different lines; some of these modifications prove to be relatively simple, while others are more intricate. For very small calculations, none of these considerations may be especially relevant, but for larger problems, the possible order-of-magnitude gain in computation speed may spell the difference between a reasonable computation and one that is no longer feasible. As mentioned earlier, the notion of what is feasible is strongly epoch-dependent, and the largest of today's simulations will at some time in the future be carried out in real time on a desk-top computer; this optimistic prediction is of little consolation when one is faced with a problem demanding immediate solution.

3. The nature of vector computation

The vector supercomputer represents a compromise between the computer designer's desire to achieve maximal computation rate (within specified cost constraints) and the user's demand for the fastest possible computations over a broad range of problems. While this compromise has proved to have considerable benefit to both parties in a great many kinds of computation in science and engineering, there is no shortage of situations where the performance potential of the supercomputer is far from realized. What distinguishes algorithms that are effectively mapped onto a vector computer is the manner in which data is accessed and the nature of the processing carried out. Peak gains are achieved when the data is retrieved sequentially from storage, when only certain combinations of the basic arithmetic operations (preferably excluding division) are carried out, and when the results are returned sequentially to storage. Any deviation from a general operational pattern of this kind results in sub-optimal performance. However, with the exception of certain kinds of matrix and vector computation which manage to follow this prescription precisely, this state of perfection is unattainable. The issue then is how to achieve the best performance given the preferred manner of operation of the hardware.

In addition to the requirement that the data be sequentially ordered for vector processing, the processor design imposes a fixed startup overhead associated with each vectorized operation; this overhead is independent of the number of data items processed in the course of the operation. An immediate consequence of this overhead is that, if the vectors are too short, the somewhat paradoxical situation where vector processing is actually slower than the corresponding set of scalar operations (on the same computer) can be achieved; this is certainly something to be avoided. The minimal vector length requirements vary, depending on both the type of operation and the machine itself. But it is clear that in addition to rearranging the data to oblige the processor, it will also be necessary to ensure that the resulting data are organized into vectors that are of adequate length to ensure that the fixed overheads do not nullify the expected performance gains. There is no guarantee that this can be achieved in all cases; there are indeed calculations for which effective vectorization is not possible.

To help the user tailor the computations for the vector processor the machine instruction set generally features the capability for reorganizing data at a fairly rapid rate, one that tends to be intermediate between vector and scalar processing speeds. There are different approaches to data rearrangement and selection, and not all approaches are to be found on all machines. Even when a particular scheme for rearranging data is implemented, the question of how fast such operations are carried out in comparison with the peak (vector) computation speed is something which must be taken into account.

The two principal schemes for selecting and reordering data are known as gather-scatter and compress-expand. The act of gathering data involves the use of a vector of indices to access some or all of the elements of a set of data items -- in

no particular order as far as the computer is concerned -- and store them sequentially in another vector; the scatter operation is the converse, in which the index vector is used to store a sequential set of data items in some alternative order in another (possibly longer) vector (where some elements might even be left unmodified). Compression involves selecting an ordered subset of data items from a vector and storing them as adjacent elements in another vector; expansion is just the reverse of this operation. Because the order of the data is preserved under the compress/expand operations, information concerning which of the elements are to be compressed, or where the elements resulting from an expansion are to be placed, can be represented in terms of a bit vector (where the ones and zeroes represent, respectively, the elements that are or are not involved), an extremely compact alternative to the more general vector of indices utilized by gather/scatter. The fraction of the total elements needed in operations that require ordered subsets of elements can determine whether gathering or compression is the more effective operation -- provided the computer offers a choice; not all machines do.

Even if the computer has the flexibility for dealing with data that are arranged in a less than optimal manner, there is the issue of whether the compiler that translates the high-level source language program into the actual machine instructions is capable of recognizing the kinds of operations needed. Judging by the achievements to date, compiler efficiency is an even more complex issue than hardware efficiency, and the performance of different compilers exhibits a wide variation in this respect. Even if the compiler is very perceptive and competent at mapping the program onto the hardware, there are situations in which certain language constructs (for example the case of an algebraic statement which implies a recursive dependence on something that has only just been computed, which might be unvectorizable because of the manner in which pipeline processing restricts dependencies between individual data items involved) may prevent the compiler from dealing effectively with that segment of the program; in such situations further information -- expressed in the form of directives that lie outside the programming language -- may help the compiler perform its task (this might include information indicating that even though a recursive dependency has been detected it can still be safely vectorized). The capacity for aiding the compiler in this way also varies from one brand of computer to another.

The alternative to total dependence on the capability of the compiler is to directly invoke the machine instructions. This can be done by programming in assembly language, but is best avoided in favor of a sometimes available alternative which facilitates accessing machine instructions by subroutine calls from the higher-level language. The advantage of the latter approach is that it need only be resorted to when the compiler cannot deal with the problem, and it leaves the text of the program in more intelligible form; for most of the program the compiler should prove up to the task, and only certain critical sections may need to be handled in this way. This feature too is not necessarily provided on all computers.

4. Vectorized molecular dynamics

In the preceding sections we discussed, in very general terms, the kinds of molecular dynamics simulations that are encountered, as well as the issues that must be addressed in trying to reconfigure the computational algorithms to match the needs of the vector processor. In this section we deal with the question of reorganizing the molecular dynamics calculation into a form suitable for efficient vector processing.

The most difficult instance of algorithm conversion is the one that deals with the case of a very large fluid system with only short-range interactions between particles. The discussion of this case will be deferred until last; the vectorization of the remaining classes of simulation are much more readily described and carried out. We note that it is the details of the interaction computations on which the discussion is focused; the actual integration of the coupled equations of motion, as well as other relatively minor -- from the point of view of computational effort -- aspects of the simulations (such as establishing the initial conditions, or modifying particle coordinates to satisfy periodic boundary conditions) present no significant problems since these parts of the simulation constitute a relatively minor portion of the calculation that is readily (normally in a completely automatic manner) vectorized.

For relatively small systems, typically up to a few hundred particles, the quickest method for a vector processor is generally the simplest one, namely to consider all pairs of particles, even if they lie beyond the interaction range. The interaction computation can easily be altered to return a zero value if the cutoff separation is exceeded. If the pairing is carried out in a suitable manner that ensures that interacting pairs are organized as vectors associated with the particles concerned in an equitable manner /7/ (since the interactions between pairs of particles can be regarded as a triangular matrix, the simplest scheme of processing a row or column at a time is not the most efficient way to approach the problem) the data to be processed can be organized into vectors of adequate length to avoid serious problems with the fixed startup overheads. The additional computations required to evaluate interactions between particles lying outside the interaction range are, in this case, not of sufficient quantity that the alternative techniques described below would yield any improvement in performance.

If the molecules of the fluid are complex, in the sense that there are several interaction sites at fixed relative positions within the molecule, then the amount of computation required normally precludes consideration of systems larger than about a thousand particles (at the time of writing). In order to vectorize such a computation it is again worth considering the possibility of treating all pairs of molecules in the system (and subsequently discarding those pairs separated by more than the cutoff range -- where the separation now refers to the centers of mass), but this time the calculation is broken into several sections (at the conceptual level, although the implementation itself may involve treating all these sections in a uniform manner as parts of a single 'outer' loop) each of which deals with a distinct pair of interaction sites in all the molecule pairs; for each pair of site types

the innermost loop of the program evaluates the contributions from just those site pairs for all molecules of the system. This is the opposite of the usual approach in which all site pairs belonging to a particular pair of molecules are treated together; it also requires additional storage to hold interim data for the coordinates of the interaction sites on all of the molecules as well as the forces acting at each of the sites, but the rearrangement of the order of the computational loops results in an algorithm which vectorizes with no difficulty.

The computations associated with two other classes of system also readily vectorize. If the interaction is long-ranged, all pairs of particles interact and there is little opportunity for any shortcuts; however the necessarily heavy computation will be carried out in fully vectorized fashion. Simulation of solids also presents an ideal situation for vector processing. The fact that the interacting pairs remain the same throughout the computation means that there is little problem in rearranging the computations in an appropriate way; this might be done by ordering the processing of interacting pairs according to the direction and range by which they are offset in the lattice.

This leaves the problem of a fluid with short-range interactions, where 'short' is defined relative to the length of the region occupied by the fluid. The limiting case of short-range interactions is represented by a fluid of hard elastic spheres or disks, in which the interactions are essentially step functions /8/; an entirely different approach is called for in this case /9/ which, because of the complex data organization needed for efficient performance, does not permit vectorization. On a scalar computer there are two techniques in use, preferably both together, which reduce the size-dependence of the computation from quadratic to linear.

The first technique used to improve computational efficiency is based on the introduction of a fictitious space-filling array of cells to which the particles are assigned. There is then no difficulty in determining which particle pairs interact using an amount of computation proportional to system size. If the linear dimensions of the cell are chosen so that they exceed the interaction range then only particle pairs that lie in the same or in neighboring cells are possible interaction candidates.

The second effort-reducing technique is to construct a list containing those neighboring pairs of particles that actually interact /10/. Since the earlier subdivision into cells overestimates the number of potentially interacting pairs that must still be examined to determine if they really do interact (by factors of approximately three and seven in two and three dimensions respectively), there is clearly something to be gained from this approach, although there is a substantial penalty to be paid in terms of storage requirements for the extra information generated. Since rearrangement of particles in the fluid is a gradual process it may not be necessary to update this list at every time step; the frequency of updates can be reduced if the list includes particle pairs whose separation exceeds the maximum range of interaction by a prescribed small amount -- there is a tradeoff involved between the work involved in generating the neighbor list and the increased size of list if addi-

tional non-interacting pairs are included. By monitoring the maximum particle displacements at each time step it is possible to ensure that the neighbor list is recomputed as infrequently as possible, while at the same time guaranteeing that no interactions are missed. Choice of a suitable separation criterion for inclusion in the list is a matter for experimentation, but the typical value might be one that exceeds the interaction range by an amount equal to the maximum expected particle displacement over perhaps twenty time steps (monitoring the displacements ensures that the list will be reconstructed as often as is actually necessary).

This represents the most effective approach to dealing with this kind of problem on a scalar computer; the question is how to achieve a similar reduction in computational effort in a vector environment. Quite obviously one would not want to completely abandon the approach which leads to a linear rather than quadratic size dependence since there is no way the gains resulting from vectorization can compensate for this loss. But to achieve this dependence requires representing the information embodied in the cell subdivision, and possibly the neighbor list as well, in a form that can be handled in vectorized fashion. As will become apparent, the solution is basically a simple one, although the implementation can become somewhat awkward.

On a scalar computer, the information concerning which particles belong to which cells is stored in set of linked lists, one such list per cell. A linked list /11/ is a way of structuring data that is not intended to be accessed sequentially; with each data item there is associated a pointer to the next data item (if any -- otherwise the pointer is given a special 'null' value). Starting from the head of the list, which is merely a pointer associated with the cell itself, the contents of the cell can be determined by following the chain of pointers. The reason for using a linked list is the fact that no a priori limits are imposed on cell occupancy, as would be the case if a fixed amount of storage were reserved for each cell; the result is a considerable saving in storage -- often as important a factor in software implementation as the speed of the computation. Unfortunately the linked list concept is totally incompatible with the vector approach and so must be sacrificed in the interest of overall performance, even if the storage requirements increase as a consequence. The question is the choice of an effective alternative.

The technique actually used /6/ requires that the identity of the cell occupied by each particle be determined as a first step. Then the data reorganization is carried out by placing the particle serial numbers in one of a set of 'layers'. There are enough such layers to accommodate the maximum possible cell occupancy, and each layer contains a single storage element for each cell. The first particle in a particular cell is assigned to the first layer (the order of particles within cells is of course arbitrary), the second particle in the cell -- if present -- to the second layer, and so on. Assigning a particle to a layer requires knowing how many particles in that cell have already been assigned. A simple scheme which scans the complete set of particles just once is obviously not vectorizable, since the number of layers that must be scanned in order to assign each particle varies, depending on cell occupancy. The alternative, vectorizable scheme, involves filling the layers one at

a time, even though it implies scanning the entire set of particles several times. This technique is not as bad as it might seem however, because once a particle has been assigned to a layer it can be eliminated from the set of particles yet to be assigned. The processing involved in the construction of each such layer is fully vectorized, and all but the final layers corresponding to the tail of the cell-occupancy distribution provide sufficiently long vectors to yield effective vector-processing performance (assuming a large enough system).

Once a layer has been populated, the data it contains can be compacted if storage is at a premium. While linked lists can be used (for those layers that are sufficiently sparsely populated so that a linked list requires less storage than the original data), a particularly effective method for those machines which support the operations of compression and expansion is to pack the layer data in an ordered manner and use a bit vector to indicate which of the cells are actually occupied. When the data is later required by the interaction computations the layer contents can be re-expanded. A further extension of this idea, that plays an important role in the computations for extremely large systems (a hundred thousand particles or more -- by current standards), is that the region of space occupied by the system can be divided into a number of subregions, and only the particular portion of the layer associated with the part of the system actually being considered need be expanded. In the case of a subdivided system, additional effort is required to ensure that interacting particles located in distinct subregions of the system are dealt with properly, but the extra bookkeeping is not unmanageable and does not add significantly to the workload.

The interaction computations that follow the layer assignment involve the treatment of all pairs of occupied layers; for each pair of layers the pairs of particles at locations in the layers corresponding to the cells for which interactions are possible are considered -- but only if the cell locations in both layers are occupied by particles; a particular case that requires special attention is when the layer is paired with itself -- each pair of particles appears twice and so only half the pairs should be treated. This calculation also vectorizes readily, although effort is wasted when sparsely occupied layers are paired; it is quite possible to have no interaction terms emerging from the processing of such layer pairs. Despite the wasted effort, this scheme is probably the most efficient on machines that support the compress/expand operations at the hardware level.

An alternative approach is to use the layers to construct neighbor lists. The neighbor lists produced in this manner will have the property that in the list segment corresponding to a given layer no particle can appear more than once. Thus the interaction computations can again be vectorized, and the additional saving is that the neighbor list construction, together with the prerequisite layer assignment, need not be carried out at every time step. There is however a heavy storage penalty associated with neighbor lists, especially if the interaction range allows many interacting neighbors per particle; it is unlikely that the approach would be useful for the extremely large systems that are now being studied -- an example of the sacrifice in speed to economize on storage.

More of the bookkeeping mentioned earlier is needed to take periodic boundaries into account, as well as to deal with the interactions if the subdivision technique is employed. In particular, the conditional tests associated with periodic boundaries when computing interactions (the interparticle forces wrap around the ‘edges’ of the system, so that particles at opposite ends can in fact interact if both are sufficiently close to their respective boundaries) can be avoided by replicating the particles near the boundaries; these dummy particles are used only for computing the interactions and are then discarded.

Since layer assignment requires significantly less work than the interaction computations themselves, the assignment can be carried out afresh at each time step; thus there is no need, for example, to monitor particle displacements (as was done when using neighbor lists) to avoid the possibility that cell occupancies may no longer be correct. Finally, another shortcut that can be used if an efficient gather operation is available is to employ tabulated interactions and to perform a table lookup (possibly supplemented by interpolation) rather than evaluate the interactions from scratch; again the decision as to which technique is preferable depends on the machine.

5. Summary

It should be clear from the foregoing that there is no single technique that can be used for implementing all molecular dynamics simulations on all vector supercomputers. There are a broad variety of problem types, and an almost equally broad range of supercomputer architectures. A particular kind of computation which runs efficiently on one brand, or even model, of machine may not perform as well on another. The storage requirements can also differ significantly between machines, and there is often a need to determine an optimum balance between speed and memory needs. This is not the place to go into further details of the techniques which have and are being published elsewhere; suffice it to say that the methods have made possible production type simulations of 200,000 particle systems over a similar number of time steps, and are capable of dealing with even larger systems as suitable computers become available (tests on systems containing as many as 500,000 particles have been carried out). Such enormous (by usual standards) systems appear to be necessary for the study of classes of hydrodynamic instability at the microscopic level /12/, and could well turn out to be necessary in other contexts as well.

Acknowledgements

Many of the developments described here were carried out during visits to the Center for Simulational Physics at the University of Georgia, and the Department of Physics at Mainz University; David Landau and Kurt Binder are thanked for their hospitality and support.

References

1. *Molecular Dynamics Simulation of Statistical Mechanical Systems*, Proceedings of the International School of Physics "Enrico Fermi", Course XCVII, Varenna, 1985, eds. G. Ciccotti and W.G. Hoover (North-Holland, Amsterdam, 1986).
2. M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
3. R.W. Hockney and C.R. Jesshope, *Parallel Computers*, 2nd edn. (Adam Hilger, Bristol, 1988).
4. R.W. Hockney and J.W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1981).
5. J.-P. Hansen, in ref. /1/, p.89.
6. D.C. Rapaport, Computer Phys. Repts. **9**, 1 (1988).
7. S. Brode and R. Alrichs, Computer Phys. Comm. **42**, 51 (1986).
8. B.J. Alder and T.E. Wainwright, J. Chem. Phys. **31**, 459 (1959).
9. D.C. Rapaport, J. Comput. Phys. **34**, 184 (1980).
10. L. Verlet, Phys. Rev. **159**, 98 (1967).
11. D.E. Knuth, *Fundamental Algorithms* (Addison-Wesley, Reading, 1968).
12. D.C. Rapaport, Phys. Rev. A **36**, 3288 (1987).

Parallelization of Computational Physics Algorithms

D.W. Heermann* and A.N. Burkitt

Fachbereich 8: Physik, Bergische Universität, Gauß Str.20,
D-5600 Wuppertal, Fed. Rep. of Germany

*Also at: Institut für Physik, Staudinger Weg 7, Universität Mainz,
D-6500 Mainz, Fed. Rep. of Germany

We report on the parallelization of two widely used algorithms in computational physics: The local-update Metropolis Monte Carlo simulation of the Ising model and a cluster identification algorithm which is used for percolation or percolation-like problems. The algorithm for the identification of clusters for the percolation-like problems was tested on the Swendsen-Wang method for the simulation of the Ising model. Simulation results are presented for a quench from a disordered state to a state below the coexistence curve. We show that the resulting domain growth has an exponential instead of a power law. The simulations were carried out on a parallel computer based on the transputer concept with up to 128 processors. The performance data show that the algorithms can perform with a linear speed-up. A scaling law for the performance of geometric parallel algorithms is proposed and tested.

1 Introduction

Almost all algorithms in prevalent use today in computational science are time-driven [1-5]. In general one wants to calculate properties of a model characterized by the Hamiltonian \mathcal{H} , describing the interaction between the entities of the model. A simulation algorithm for such a model generates states $s \in \Omega$ starting from some initial state s_0 . The structure, which holds for deterministic as well as for stochastic methods, is

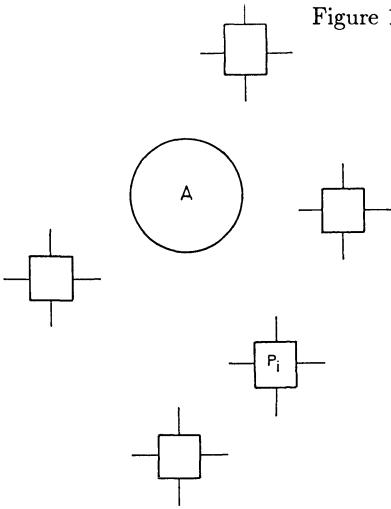
```
A:   s := s0 ∈ Ω
      LOOP
        change state s → s' ∈ Ω
      ENDLOOP
```

This algorithm is run with different initial conditions s_0 to generate trajectories through the state space of the system. Also there are other parameters, which we name generically by \underline{x} , that one wants to vary. For the calculation of the phase diagram of a ferromagnet the temperature and the magnetic field need to be varied; the number of particles or the volume is often changed to study the influence of effects of finite size. We can write in a formal way for the general simulation problem

```
P:   some s0 some  $\underline{x}$ : A(s0,  $\underline{x}$ )
```

i.e., run the algorithm for some initial conditions and for some parameters. From this formal structure it is apparent that one strategy for the parallelization of simulation algorithms in computational science is *replication* [6,7] (see Figure 1). The basic algorithm \mathcal{A} is replicated identically on the available processors. Each processor is assigned the algorithm \mathcal{A} with a specific initial condition s_0 and particular set of parameters \underline{x} .

Figure 1: Shown is a graphic representation of the replication algorithm idea.



This homogeneous parallelization is certainly the simplest form of parallelization, though a very effective one. This simple scheme is only applicable if the single processor is powerful enough to manage the single problem $\mathcal{A}(s_0, \underline{x})$ in a reasonable time. If the number of particles in the simulated system is large, or the interaction between them too complicated, then one quickly reaches the capacity of a single processor.

The formal structure of the general time-driven simulation algorithm discussed above neglects completely the structure implied by the underlying physics problem. Causality plays an important role and can be exploited for the parallelization. If the interaction between the particles is short ranged, then parts of the system can be handled simultaneously. More generally, if the underlying rule for a state change is *local*, then the system can always be partitioned into quasi-independent parts.

The locality gives a sufficient condition for the *geometric* decomposition of the problem. In the geometric decomposition the state change from s to $s' \in \Omega$ is split into parts. The parts are not independent. Usually a global constraint must be fulfilled. The global constraint can be, for example, a condition of detailed balance [1-4,6]. Such a constraint brings with it, by necessity, that the parts must interact at certain stages, and hence the parts are quasi-independent.

2 Geometric Parallelization of the Ising Model

Let us consider as a first example the Ising model:

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} s_i s_j \quad s_i = \pm 1, \quad J > 0 \quad (1)$$

The point to note is the locality of an attempted change in the configuration if one uses the Metropolis procedure [1-4] for a site update. A single site is chosen for a possible change. The change can be determined from the local neighbourhood alone. In the Hamiltonian \mathcal{H} only nearest neighbour interactions appear. In two dimensions this means only the four nearest neighbours. In Figure 2 we have indicated by the black and white checkerboard pattern which spins could in principle be simultaneously updated.

The Ising model has an inherent speed-up [8], defined by the complexity of the algorithm on a single processor divided by the unlimited processor complexity of $L^d/2$. In practice this

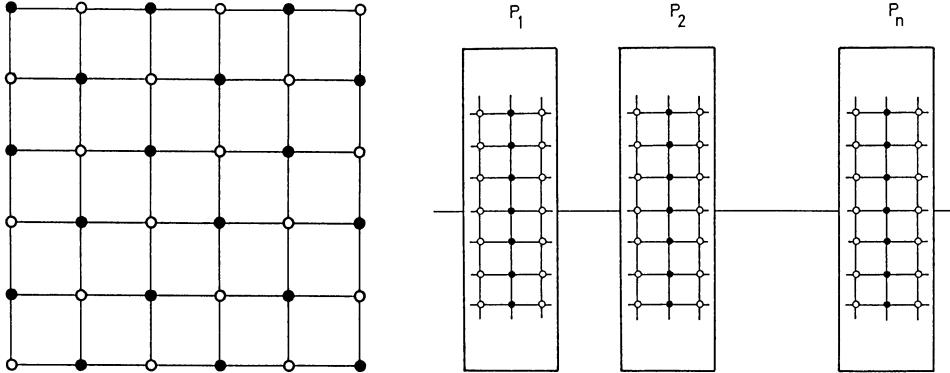


Figure 2: Shown is the decomposition of a simple square lattice into two inter-penetrating sub-lattices as used for pipeline machines. The figure on the right shows the decomposition into sublattices as used for parallel processors with local memory and message passing.

would require a special purpose machine, which indeed has been built [9,10], or a machine which closely fits the model, such as the DAP [11,12,13], to obtain at least approximately this order of speed-up.

The more realistic approach is to coarse-grain the lattice into blocks. i.e., the lattice is partitioned into b blocks. In principle we would get a speed-up of b if we have b processors available. In Figure 2 is shown how we partitioned the lattice into blocks. The lattice is divided into strips. Within each strip the lattice is partitioned into virtual blocks. This is necessary to ensure the global condition of detailed balance [1-4]. Only certain blocks are active at a particular time. The necessary information for the updates of spins is either within the processor (periodic boundary conditions in one dimension is easily implemented, since all information is already within the memory of a processor) or must be gathered from the two nearest neighbour processors. The gathering is done in an overlapped fashion. The information for the block next in line for execution is gathered while the current block is processed.

For simplicity we will be using only a two dimensional lattice for the timing of the algorithms. The same principle can of course be applied to higher dimensional lattices.

3 Geometric Parallelization of Cluster Algorithms

An algorithm for updating spin configurations which is based upon the correspondence between the Ising model and percolation [14] has been developed by Swendsen and Wang [15]. This algorithm reduces substantially the effect of critical slowing down for temperatures in the neighbourhood of the critical point.

In this algorithm there are, in addition to the spins, also bonds between spins. These bonds are broken between spins that are anti-parallel and they connect spins, with a probability that is temperature dependent, that are parallel. All spins connected by bonds then belong to the same cluster and there can clearly be a great many clusters associated with any particular spin configuration, some of which may consist of only one spin and others that may be very large. In the Swendsen-Wang algorithm all spins in a cluster may be simultaneously flipped in a single Monte Carlo step. The algorithm is thus very non-local since spins far apart may belong to the same cluster. For a detailed exposition of the algorithm the reader is referred to ref [15].

The computationally difficult part of this problem is to identify clusters in a spin system that is distributed over a number of processors. This cluster identification algorithm which can

be used for percolation problems [16] is based on the labelling technique developed by Hoshen and Kopelman [17] and extended by Kertesz [18].

We start by assigning a cluster label 1 to the spin in the “upper left corner” of the lattice. We also use a “stack vector” (or “cluster permutation vector”), whose first element is now assigned a value -1 or -3 (for spins +1 and -1 resp.). Each spin in the lattice is sequentially visited and if it is not connected with any labelled cluster it is assigned a new label, and the corresponding spin-value entry is placed in the stack vector. Any spin that is connected to only one labelled cluster is assigned that particular cluster label. In the case where a spin is connected to more than one labelled cluster it is assigned the lowest of these cluster-labels, and the other cluster(s) which become connected by this spin are joined by overwriting their (negative) spin values in the stack vector with this (positive) lowest cluster label (for further details, see ref [17]).

In this cluster labelling procedure it is thus possible for two clusters, which were originally widely separated, to be part of a single larger cluster. This feature needs very careful attention when the lattice is divided into strips that are distributed over a number of processors. The algorithm is now divided into a number of steps. We proceed in the first stage by treating each strip of the lattice separately; the bonds along the two boundaries are considered to consist entirely of broken bonds and a “local cluster label” is assigned to each cluster in the strip. Clusters within the strip are built using a local stack vector in the way described above. Simultaneously the spins on the boundary are communicated with the neighbouring strip and the actual bonds on the boundary are generated. The next stage is to create a “join vector” at each boundary that contains the global cluster labels of cluster pairs that are connected across a boundary by a bond. (The global cluster label on the r^{th} processor in a series of p processors can be defined as $(r - 1) * (L/p) * L$ plus the local cluster label, where each processor contains a strip of $(L/p) * L$ spins ; this ensures an overall unambiguous ordering of the clusters). All the clusters that enter into the join vectors are collected into a global stack vector. The join vectors are communicated cyclically through all the processing elements and are used to connect clusters in the global stack vector (in a similar fashion to that of connecting clusters in the local stack vector).

After a full cycle of the join vectors through all the processors the new spin values on each processor can be read from the global stack vector (for clusters that were connected across a boundary) or the local stack vector (for spins in clusters lying entirely within a strip), and in this way a new spin configuration is generated. For a more thorough description of this algorithm on many processors, see ref [19].

The non-local nature of the problem means that information must be passed from every processor to every other processor involved in the lattice simulation. This is unlike the usual local-update Monte Carlo algorithm described in the previous section, where it is only necessary for information to be exchanged between immediately adjacent lattice strips. This feature is the determining factor in the efficiency of the parallel algorithm, because the time spent in communication between processors increases not only with lattice size, but also with the number of processors over which the lattice is distributed. This is the source of the very different performance characteristics of the local and non-local algorithms, which will be discussed in the following section.

Before going into the discussion on the performance of the presented algorithms we want to discuss some consequences for the physics which result from the non-local nature of the Ising model with the Swendsen-Wang dynamics.

4 Domain Growth for the Non-Local Ising Model

Algebraic relaxation out of a non-equilibrium situation is what is generally observed. For many models with conserved and non-conserved order parameter a power law behaviour has been demonstrated by simulation and suggested by analytical considerations [20-27]. For all these

models the dynamics is based on *local changes* to the system. The question which we address now is in how far the relaxation out of a non-equilibrium state, produced by a quench, is influenced by non-local changes.

Let $l(t)$ denote the average size of a domain after a quench. Typically one considers a temperature quench. The system is prepared in a high temperature disordered state. For convenience we discuss only models with a non-conserved order-parameter. After the system has been set up, the control parameter, here temperature, is changed such that the system is now in a non-equilibrium state below the coexistence curve. The relaxation out of this state into a stable equilibrium state is characterized in most models by the growth law for the average domain size [20,21,27]

$$l(t) \propto t^n \quad (2)$$

The reason that all models simulated so far with the Glauber/Metropolis or stochastic dynamics [28,29,30] give essentially the same behaviour is that they are based on local rules. Only local spin flips or site updates in the different schemes are allowed. The initial domains, separated by walls, grow according to the local rule. A domain can only grow by moving the domain wall in small steps.

The relaxation dynamics is completely different once the rule for a change is non-local. One way of implementing a non-local rule is provided by the Swendsen-Wang algorithm [15].

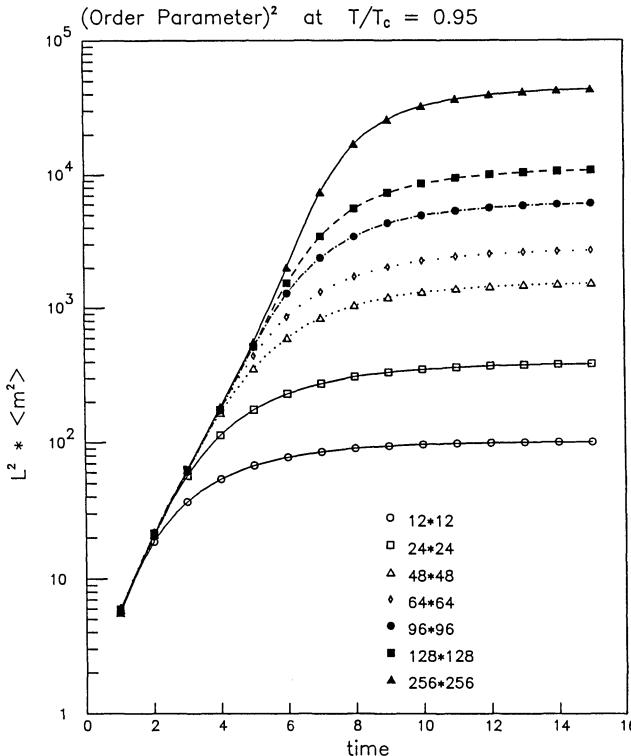


Figure 3: Shown is a semi-log plot of the results of the average domain size R and time t as measured in Monte Carlo sweeps. The data presented here are for a temperature $T/T_c = 0.95$. Different symbols denote different system sizes as indicated in the legend. The statistical errors are smaller than the symbol sizes.

The simulations of the quench from a high temperature state to a state below the coexistence curve were carried out for the two-dimensional Ising model. We have simulated a variety of system sizes ($N = L^2$, $L = 12, 24, 48, 96, 128, 256$) and temperatures ($T/T_c = 0.9, 0.95, 0.98$). All of these simulations were done on the parallel computer at the Gessellschaft für Mathematik und Datenverarbeitung. This machine consists of up to 128 transputers. Further simulations are being carried out on the Computing Surface at the University of Mainz.

The order parameter, i.e. the average domain size, is defined as

$$l(t)^2 \equiv \frac{L^2 \langle m^2 \rangle_t}{\langle m \rangle_{eq}^2}, \quad (3)$$

where

$$\langle m^2 \rangle = \langle (\sum_{i=1}^N s_i/L)^2 \rangle, \quad s_i = \pm 1 \quad (4)$$

and the brackets are meant as averages over independent quenches.

Contrary to the results from the simulations with a local rule, the data from our simulation (cf. Fig. 3) clearly show an exponential instead of an algebraic relaxation towards the stable equilibrium state

$$L^2 \langle m^2 \rangle \propto \exp(ct) \quad (5)$$

Shown are the data for the temperature $T/T_c = 0.95$.

5 Performance Data

The notion of speed-up [8] which we will be using below is

$$S(p) = \frac{\text{time complexity on one processor}}{\text{time complexity on } p \text{ processors}} \quad (6)$$

In the ideal case we expect to gain a linear speed-up, i.e., if we cut the two dimensional lattice into p strips with each strip assigned to one of the p processors. Hence each processor is responsible for $(L/p) * L$ data elements. Of course, we have assumed that the input/output operations of the basic algorithms on one processor do not degrade the performance of others.

In Figure 4 are shown the results for the above outlined algorithm for the Ising model with a local update algorithm. The data shows that initially the speed-up we gain is indeed linear. Eventually, though, the problem for a single processor is too small and too many messages have to be communicated. In the case of the 128^2 lattice a single processor in fact handles only a single spin for $p = 128$ before communication is necessary. The problem is too atomized for the specific processor, i.e., the coarse graining is too fine. We would do much better running two simulations with one on each half of the system. This can be seen by looking at the utilization or efficiency defined by

$$\eta = \frac{S(p)}{p} \quad (7)$$

which drops to 50 percent utilization for about 32 processors for a lattice of size $L = 128$. Hence a mixture of the replication algorithm, which was discussed in the beginning of the paper, and the geometric algorithm would give the best yield.

Note that the results for the 1024^2 system show a strong deviation form a linear speed-up already for $p = 128$. This strong deviation is due to some link problems between two $p = 64$ clusters. It should be emphasized that this problem can be rectified. Hence all data for $p = 128$ should be corrected upwards.

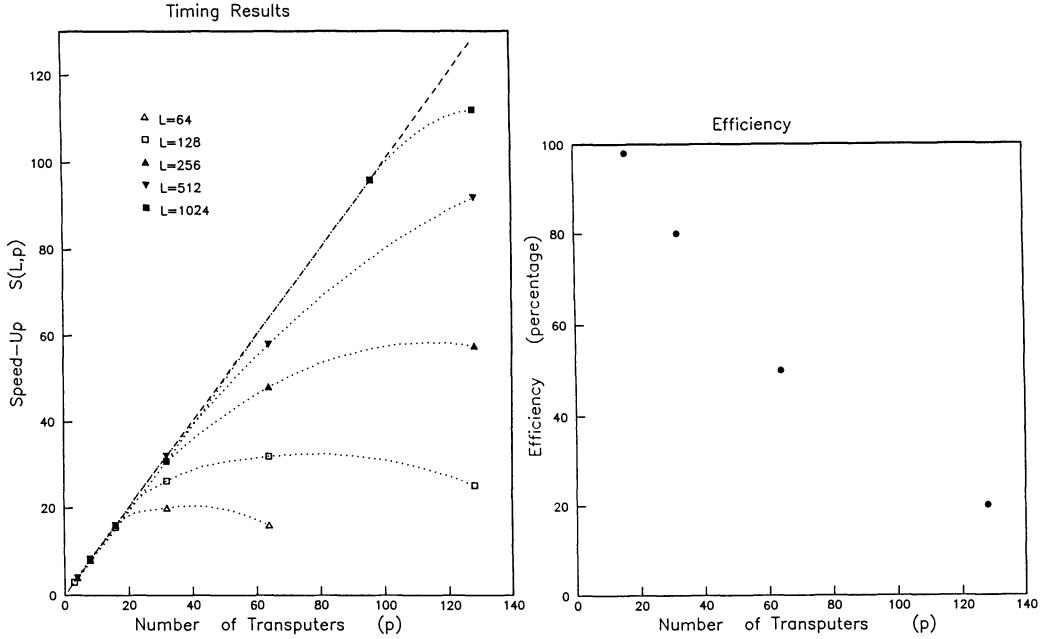


Figure 4: Speed-up of the geometric algorithm as a function of the number of processors p for the local-update Ising problem (left). Different curves denote different lattice sizes as indicated in the legend. The right figure shows the processor utilization for the case of a 128^2 lattice

The picture is essentially the same as for the Swendsen-Wang cluster algorithm for the updating (c.f. Figure 5). What we have to consider here is the by far larger communication overhead which is a result of the constraint of global connectivity. Initially the speed-up is linear as was the case for the Ising algorithm. The saturation sets in much earlier than for the Ising algorithm. This is due to the longer communication paths and also larger amounts of data which have to be passed.

It is interesting to note that the data obeys a scaling law. The speed-up S , in general, is a function of the problem size L and of the number of processors p . The observation is that the speed-up depends only upon the ratio of the problem size and the number of processors. This relation cannot be linear. If so then there would be no saturation and the observation would be trivial. The point is that we need to generalize this argument to propose a scaling of the form

$$S(L, p) = L^x \tilde{S}(p/L^x) \quad (8)$$

We expect that for large lattices and for a small number of processors the speed-up will be linear. In this limit the dependence on L should cancel out of the speed-up formula. On the other hand, for a large number of processors, communication will dominate so that the relation is non linear. The exponent x can be interpreted as giving an effective lattice size for the given algorithm. If no communication overhead occurs for all p and L then x would be one. A small x says that the effective lattice is much smaller than the original.

The scaling for the two problems is shown in Figure 6. The family of speed-up curves, both for the Ising as well as the cluster algorithm, can be collapsed onto a single scaling curve. The scaling functions themselves are not universal. Each algorithm gives a different scaling curve, but the scaling itself holds for both.

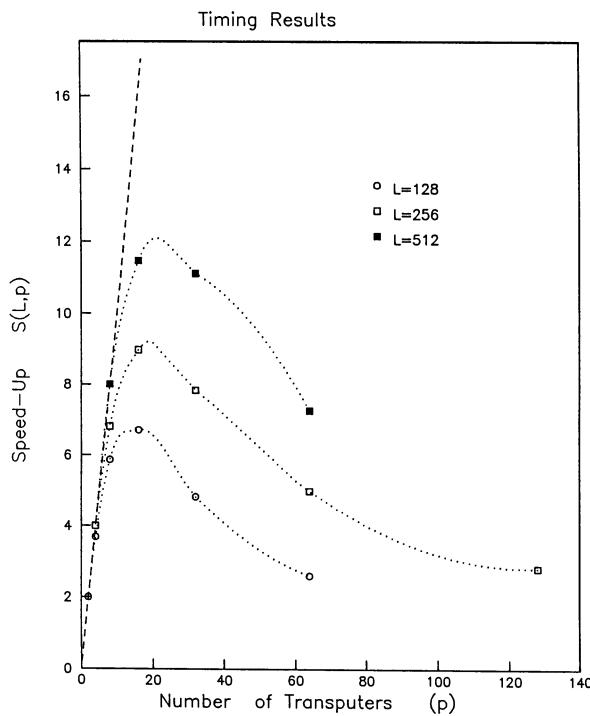


Figure 5: Speed-up of the geometric algorithm as a function of the number of processors p for the cluster problem. Different curves denote different lattice sizes as indicated in the legend.

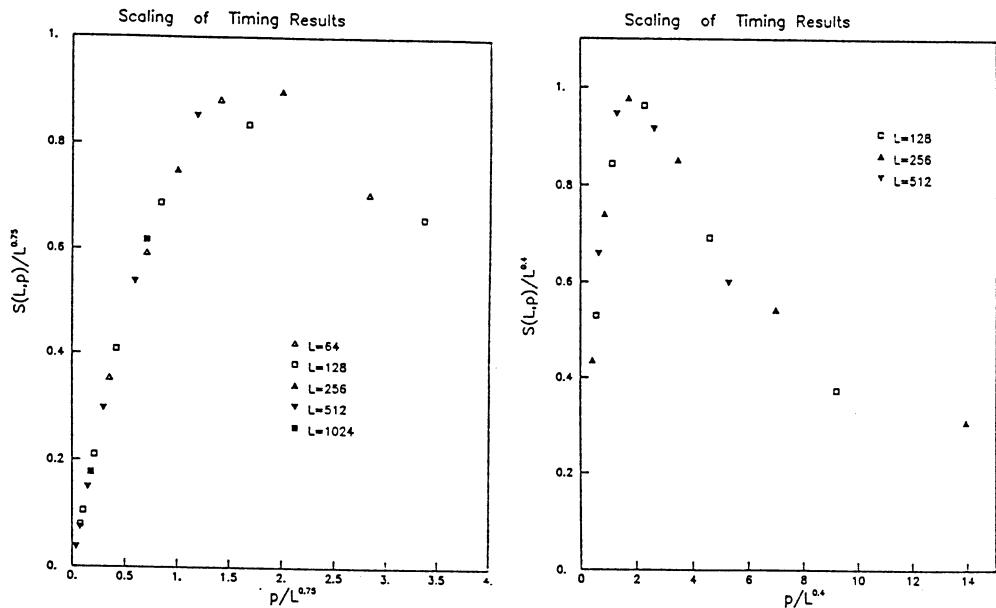


Figure 6: Shown is the scaling property of the geometric algorithms which are proposed in this text. The left figure shows the result for the local-update Ising problem and the right figure those of the cluster algorithm.

The exponents for the local-update Ising and the cluster algorithm are different. The exponent for the local Ising algorithm is

$$x_{Ising} \approx 0.75$$

while the exponent for the cluster algorithm is

$$x_{SW} \approx 0.4$$

Since this scaling holds for both problems we conjecture that such a scaling holds for all geometric algorithms. Of course, the scaling function \tilde{S} , i.e., the non-universal part, will depend on the processor speed, the detailed communication and the specific problem. It is also not clear in how far a general statement on the exponents can be made.

6 Discussion

We have shown that algorithms which are widely used in computational physics can be parallelized with a significant gain in speed-up. However, our results also show that the problem size must be appropriate for the number of processors. A problem which is too fine grained cannot run efficiently on a many processor system. The finding of scaling for the speed-up for different lattice sizes and processor number needs further investigation. What has to be solved is in how far one can attach at all a meaning to the exponents, ie., are they universal or not. Or, to rephrase the question, do local and non-local type algorithms give different exponents and is the range continuous?

Acknowledgments We would like thank H. Mühlenbein of the Gesellschaft für Mathematik und Datenverarbeitung for generously providing us with time on the GMD machine. The support from the Sonderforschungsbereich 262, Materialwissenschaftlicher Forschungszentrum at the University of Mainz and the Schwerpunkt Schi 257/2-2 is gratefully acknowledged.

References

- [1] K. Binder, editor. *Monte Carlo Methods*. Springer Verlag, Heidelberg, 1986.
- [2] M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods*. Volume 1, Wiley, New York, 1986.
- [3] D. W. Heermann. *Computer Simulation Methods in Theoretical Physics*. Springer Verlag, Heidelberg, 1986.
- [4] K. Binder and D. W. Heermann. *Monte Carlo Simulation in Statistical Physics: An Introduction*. Springer Series in Solid-State Sciences 80, Springer Verlag, Heidelberg, 1988.
- [5] W. G. Hoover. *Molecular Dynamics*. Volume 258 of *Lect. Notes. Phys.*, Springer Verlag, Heidelberg, 1986.
- [6] D. W. Heermann and A. N. Burkitt. *Parallel Algorithms and Computational Physics*. Springer Verlag, Heidelberg, 1989 to appear.
- [7] D. W. Heermann and R. C. Desai. *Comp. Phys. Commun.* 50, 297 (1988).

- [8] A. Samal and T. Henderson. *International Journal of Parallel Programming* **16**, 341 (1987).
- [9] M. N. Barber, R. B. Pearson, J. L. Richardson, and D. Toussaint. *Phys. Rev. B* **32**, 1720 (1985).
- [10] A. Hoogland, A. Campagner, and H. W. J. Blöte. *Physica* **132A**, 457 (1985).
- [11] R. W. Hockney and C. R. Jesshope. *Parallel Computers*. Adam Hilger, Bristol, 1981.
- [12] R. W. Gostick. *ICL Tech. J.* **1**, 116 (1979).
- [13] G. S. Pawley and G. W. Thomas. *J. Comp. Phys.* **47**, 165 (1982).
- [14] P. W. Fortuin and P. W. Kasteleyn. *Physica (Utrecht)* **57**, 536 (1972).
- [15] R. H. Swendsen and J. -S. Wang. *Phys. Rev. Lett.* **58**, 86 (1987).
- [16] D. Stauffer. *Introduction to Percolation Theory*. Taylor and Francis, London, 1985.
- [17] J. Hoshen and R. Kopelman. *Phys. Rev. B* **14**, 3428 (1976).
- [18] J. Kertesz. private communication.
- [19] A. N. Burkitt and D. W. Heermann. Parallelization of a cluster algorithm. November 1988. Wuppertal Preprint (to appear in *Comp. Phys. Commun.*).
- [20] J. D. Gunton, M. san Miguel, and P. S. Sahni. In C. Domb and J. L. Lebowitz, editors, *Phase Transitions and Critical Phenomena*, Academic Press, New York, Vol. 8 1983.
- [21] K. Binder and D. W. Heermann. In R. Pynn and T. Skjeltrop, editors, *Scaling Phenomena in Disordered Systems*, Plenum Press, New York, 1985.
- [22] S. M. Allen and J. W. Cahn. *Acta Metall* **27**, 1085 (1979).
- [23] E. T. Gawlinski, M. Grant, J. D. Gunton, and K. Kaski. *Phys. Rev. B* **31**, 281 (1985).
- [24] O. G. Mouritsen. *Phys. Rev. B* **32**, 1632 (1985).
- [25] G. S. Grest, D. J. Srolovitz, and M. P. Anderson. *Phys. Rev. Lett.* **52**, 1321 (1984).
- [26] A. Milchev, K. Binder, and D. W. Heermann. *Z. Phys. B* **63** - *Condensed Matter*, 521 (1986).
- [27] A. Sadiq and K. Binder. *J. Stat. Phys.* **35**, 517 (1984).
- [28] R. Toral, A. Chakrabarti, and J. D. Gunton. *Phys. Rev. Lett.* **60**, 2311 (1988).
- [29] T. M. Rogers, K. R. Elder, and R. C. Desai. *Phys. Rev. B*, (to be published).
- [30] G. F. Mazenko and O. T. Valls. *Phys. Rev. Lett.* **59**, 680 (1987).

Classification and Exponents of Cellular Automata

D. Stauffer

HLRZ, c/o KFA Jülich, D-5170 Jülich, Fed. Rep. of Germany

All 65536 cellular automata on the nearest neighbor square lattice (no memory effects) were simulated numerically and classified according to simple Wolfram-type criteria. Only about one out of ten rules leads to a fixed point or an oscillation of period two. More useful seems the classification according to damage spreading: seven percent of the rules let an initial disturbance die out, in 30 percent the damage remains localized, and in the other cases it spreads over the whole lattice: Chaos.

Critical exponents were determined for a random mixture of all these automata, i.e. for the square lattice Kauffman model. The amount of damage created by isolated initial disturbances vanishes, as a function of the distance from the critical point, with an exponent $\beta = 0.34$ on the chaotic side, and with an exponent $\gamma = 2.5$ on the non-chaotic side of the phase transition.

This paper consists of two parts: The original work in the second part on critical exponents of the Kauffman model¹⁻⁵ was not reported at the conference because of time limitations, and I gave there a summary of a general classification⁶ of cellular automata in the style of Wolfram⁷ (my first part here).

GENERAL CLASSIFICATION

In cellular automata, each site of a large lattice carries a spin which is either up or down. The orientation of each spin at time $t+1$ is fixed completely by the orientation of its neighbor spins at time t . These are the simple automata we will deal here with; one can generalize the above definitions to include spins with more than two states, interactions to more than just nearest neighbors, memory effects due to the orientation of the center spin itself, and probabilistic cellular automata like the Ising model which determine the spin orientation only with a certain probability.

The easiest case is one dimension, where each spin depends only on its left and its right neighbor. The logical AND, for example, gives an up spin at time $t+1$ only if both neighbors are up at time t ; otherwise the center spin is down at time $t+1$. Since each of the $K=2$ neighbors has two orientations, we have in total $2^K = N = 4$ neighbor configurations. For each neighbor configuration we have a rule which gives an up spin, and another rule which gives a down

spin. This amounts to $2^N = 16$ possible rules: and, or, nand, nor, equivalence, exclusive-or, tautology (always up), contradiction (always down), up iff left up and right down, down iff left up and right down, up iff right up and left down, down iff right up and left down, up iff left up, up iff right up, up iff left down, up iff right down. If the spin depends also on its own orientation the number of neighbors is increased from $K=2$ to $K=3$, the number $N = 2^K$ of neighbor configurations from $N=4$ to $N=8$, and thus the number of possible rules 2^N from 16 to 256. They were investigated systematically by Wolfram⁷ and classified into four groups, depending on what a random initial distribution of spins leads to eventually:

- I. Fixed points with all spins parallel
- II. Fixed points with finite clusters, oscillations with short periods
- III. Chaos (neither finite clusters nor small periods)
- IV. Propagating structures

These four Wolfram classes are not mathematically rigorously defined but seem to me a useful first grouping, somewhat similar to the Ehrenfest classification of phase transitions (first order, second order, third order etc): We know today that this classification is incomplete and partly misleading and that the universality classes of critical phenomena are determined instead by the values of non-integral critical exponents. Nevertheless these modern scaling and universality theories perhaps would not have been found without the earlier Ehrenfest classification. Similarly I believe the Wolfram classification will have to be replaced later by universality classes based on critical exponents for phase transitions between Wolfram classes II and III, as we will discuss for the Kauffman model. Nevertheless as a starting point the above classification could be useful also for later critical exponent studies.

However, for that purpose we need a precise definition easy to handle by computer simulation. Whether a periodic motion has a short or long period is a matter of taste. A finite cellular automat with L sites, in one dimension, has at most a period of 2^L , which is large but finite. A more precise distinction would be whether the period increases exponentially with system size, or according to a power law. Such a distinction worked reasonably for the Kauffman model (Stauffer²) but requires a lot of computer memory and computer runs; it is hardly suited for a simple overview of numerous rules according to some automatic computer evaluation. Thus instead we simply asked if the oscillation had a period of two, or if there was no oscillation at all. These questions can be answered with little extra memory and effort.

Also, what does chaos mean in a system where the local variables are up and down only ? We follow Derrida² and call the system chaotic if the final configuration depends on the initial configuration. More precisely, we simulate two replicas of the same lattice. Both obey the same rule, and also the initial spin configuration is nearly the same. Only at one site, or at another well-localized set of points, one replica has the spin opposite to the spin of the other

replica. Now the simulation starts, and this localized difference (called damage or difference picture or Hamming distance) may spread over the whole lattice, may remain localized near its origin, or may die out fully. Once the damage is zero, it remains zero forever in our deterministic systems. We call the system chaotic if (at least with nonzero probability) the damage can spread from the lattice center to the boundary of the system. This corresponds to a stability analysis similar to classical mechanics, where a small perturbation in an energy minimum for a single particle leads to a localized excitation whereas in an energy maximum it can change the position completely.

In one dimension, chaos in the above sense is possible since we can use the rule where each spin takes the orientation of its left neighbor. The initial damage then travels unchanged from left to right. However, phase transitions between chaotic and non-chaotic behavior, or other types of phase transitions, as a function of the initial concentration of up spins seem impossible. Also the honeycomb lattice ($K=3$) in two dimensions does not always have the desired phase transitions (Stauffer ²). To look for phase transitions, the simplest case thus seems to be the square lattice with $K = 4$. We thus have $N = 2^4 = 16$ neighbor configurations and $2^{16} = 65536$ different rules.

We cannot write 65536 different computer programs reliably and thus have to use one general algorithm which treats all of them in one big loop. We also have to simulate the lattices as efficiently as possible if we want to clarify in one hour a question relating to all 65536 rules. The fourfold rotational symmetry of the square lattice reduces our effort for most cases by a factor 4, and the up-down symmetry by a factor 2; only 64 rules are rotationally symmetric. Thus we looked only at 8368 rules. Previous work of Packard and Wolfram ⁸ was restricted to totalitarian rules depending only on the total number of up neighbors and thus in our case to 32 rules. Such rules could not be used to describe anisotropic problems like flow.

Having selected the 8368 really different rules, we simulate them by the efficient algorithm of da Silva and Herrmann⁵ which was improved during the Workshop by P. de Forcrand (University of Minnesota) to nearly 150 steps per microsecond and per Cray processor (see appendix). It stores 64 sites in one computer word, finds out with a combination of logical statements which neighbor configuration is realized, and then updates each site according to the rule value for that neighbor configuration.

For details of the results we refer to ref.6. We simulated up 8832*8832 lattice sites and up to 31623 updates per site. Nearly three percent of all 65536 rules ended up with all spins down, the same amount with all spins up, 2 percent oscillated with period two with not all spins parallel, 0.7 percent oscillated between all spins up and down, and 0.1 percent ended up in fixed points with not all spins parallel. Thus nearly 92 percent of all rules have periods larger than two.

More interesting perhaps is then the classification according to chaos. We damaged initially the center line of our square lattice, and called the system chaotic if the damage later reached a boundary line parallel to the initially damaged line. Results depended strongly on when the initial damage was introduced; it seems better to wait for a stationary state of the lattice before introducing this initial damage. In this case, we found in 4 percent of the rules the damage to heal out completely, in 37 percent of the rules it remained localized, and for the remaining 59 percent it touched the boundary, indicating chaos.

In all this work, initially half of the spins were up, and the other half down, distributed randomly. If we changed the initial distribution to a 1:9 or 9:1 ratio, hundreds of transitions between the different classes mentioned above were found. These transitions will give a wide field of research for the future; it particular one will have to check which of them are just finite-size effects like in bootstrap percolation.

What use is all this? We have not pointed out here a single physics problem and then found a cellular automata approximation for it. Instead we looked at all automata of a certain large class to classify their behavior. Further research in this direction should finally give indications which automata can give which desired results of more complex physics problems, and then one could use these automata as approximants for the real problem. Science not only progresses by looking at fixed applications, but also by systematically investigating new fields on their own merit.

KAUFFMAN MODEL

The Kauffman model¹ on regular lattices has been studied² for some time as an example of an extremely disordered cellular automaton, even though its biological significance seems limited³. Each site of a large lattice obeys its own rule, randomly selected from all possible $2^N, N = 2^K$ rules in a system of K neighbors. With probability p a rule is selected which for a given neighbor configuration gives the result spin up for the spin = ± 1 at this site for the next time step. A phase transition to chaos was observed by Derrida et al.^{2,4} both as a function of the number K of neighbors and of the probability p. In the chaotic phase, e.g. for $p > p_c = 0.3$ on the square lattice, K=4, a single spin flip can affect a large fraction of the lattice in its later time development. Thus¹ one simulates two replicas of the lattice, both with the same distribution of rules and nearly the same spin configuration. Only in the center of the lattice is one spin antiparallel in the two replicas. This damage then in the following simulation may spread over the whole lattice (chaos), remain localized, or finally die out. The fraction of sites damaged later by this single spin flip, i.e. the fraction of spins which differ in a site-by-site comparison of the two replicas at the same time, serves as an order parameter² for the chaotic phase, similar to the

(normalized) magnetization of ferromagnetism. Thus for chaos this fraction is nonzero, while in the nonchaotic phase, $p < p_c$, it is zero in the thermodynamic limit.

The word damage makes sense if initially only one spin, or a localized group of spins, is different in the two replicas. If instead a finite fraction of sites, distributed randomly over the whole lattice, has different spins, then the role of this fraction has been compared² with that of a magnetic field in a ferromagnet. Thus we denote as the Hamming distance $H(t)$ the number of spins differing in the two configurations after t iterations of the whole lattice, normalized by the total number of lattice sites. The ratio $H(t)/H(0)$ for small initial Hamming distance $H(0)$ in the nonchaotic phase, or more generally the derivative of $H(t)$ with respect to $H(0)$, is called the susceptibility χ . Close to the transition between chaotic and non-chaotic behavior one defines the critical exponents β and γ through

$$\chi(t) \propto (p_c - p)^{-\gamma},$$

$$H(t) \propto (p - p_c)^\beta,$$

for $t \rightarrow \infty$, $H(0) \rightarrow 0$, in analogy with magnetic susceptibility and spontaneous magnetization.

Past attempts to determine γ and β directly were unsuccessful. Now the better algorithm of da Silva and Herrmann⁵, their better estimate $p_c = 0.298$, and a better computer Cray-XMP are available, and another attempt was made. In practice 16 sites had initially opposite spins, and the linear dimension L of the square lattice was varied up to 5056 (25 million sites) to reduce finite size effects like $H(0)$. If initially only one spin was flipped we doubled $H(t \rightarrow \infty)$ since then at each time only one of two sublattices carries different spins. For the same reason we ensured that the 16 equally spaced sites damaged at the beginning were half on one sublattice and half on the other. Periodic boundary conditions, with L a multiple of 64, were employed throughout. The speed was up to 130 steps per microsecond and processor.

Fig.1 shows the susceptibility χ below p_c and the order parameter above p_c , as a function of $|p - p_c|$. The slopes in these log-log plots give

$$\gamma = 2.5, \beta = 0.34.$$

Actually this estimate for β does not come from the overall slope (about 0.31) since the curve is slightly s-shaped and its maximum slope in the center (away from finite-size corrections and away from corrections to scaling) seems a better estimate.

If we believe in (hyper-)scaling, $d\nu = \gamma + 2\beta$, then the correlation exponent ν for $d=2$ would be 1.6, and the fractal dimension of the Hamming distance would be $D = d - \beta/\nu = 1.78$. The latter exponent is compatible with the new (Stauffer²) direct estimate $1.8 < D < 1.9$ via damage spreading². (Stauffer²

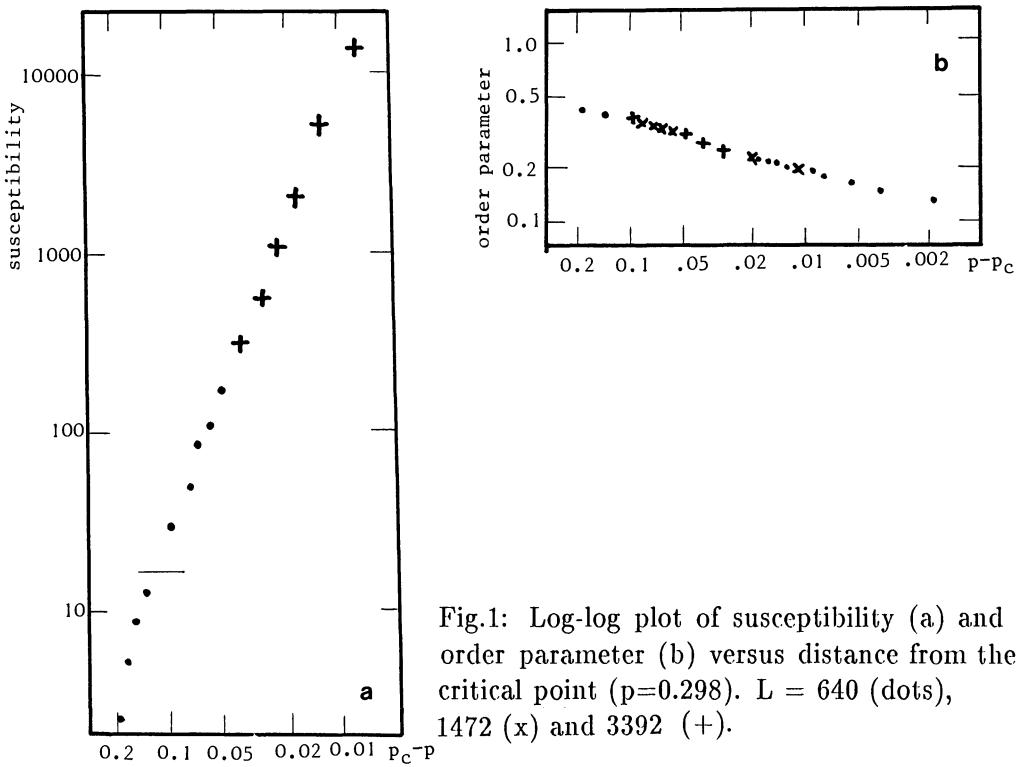


Fig.1: Log-log plot of susceptibility (a) and order parameter (b) versus distance from the critical point ($p=0.298$). $L = 640$ (dots), 1472 (x) and 3392 (+).

used a higher threshold estimate above 0.3, which would slightly increase γ and decrease β .) Ref.5 found $x = 0.06 \pm 0.05$ and $y = 0.45 \pm 0.10$ for two exponents x and y , which according to usual finite-size scaling could be identified with $y\beta$ and $1/(1+\nu)$, respectively. Our estimates then would mean $x = 0.13 \pm 0.01$ and $y = 0.39 \pm 0.01$, which only for y agrees well with ref.5.

Fig.2 shows right at the critical point, $p = 0.298$, the variation of the Hamming distance with time if we start with 16 sites damaged in a large lattice ($L = 3392$ and 5056). The slope in this log-log plot is about 1.2 (smaller if p is smaller) roughly compatible with the ratio of fractal dimensions for mass and time (which were both about 1.6 according to Stauffer²). For larger times, saturation of the Hamming distance sets in and the slope is thus diminished due to finite-size effects.

For Ising magnets, also on the ferromagnetic side of the phase transition a susceptibility is defined as the initial slope of the magnetization versus field curve, and analogously for percolation one can measure the mean cluster size of finite clusters even if an infinite cluster is present. Surprisingly, for the present Kauffman model this slope $dH(\infty)/dH(0)$ seems to be infinite for vanishingly small initial fraction $H(0)$ of flipped spins: Fig.3 shows a logarithmic variation of the final Hamming distance with the initial Hamming distance in the chaotic phase. This logarithmic variation, as well as the logarithmic time dependence

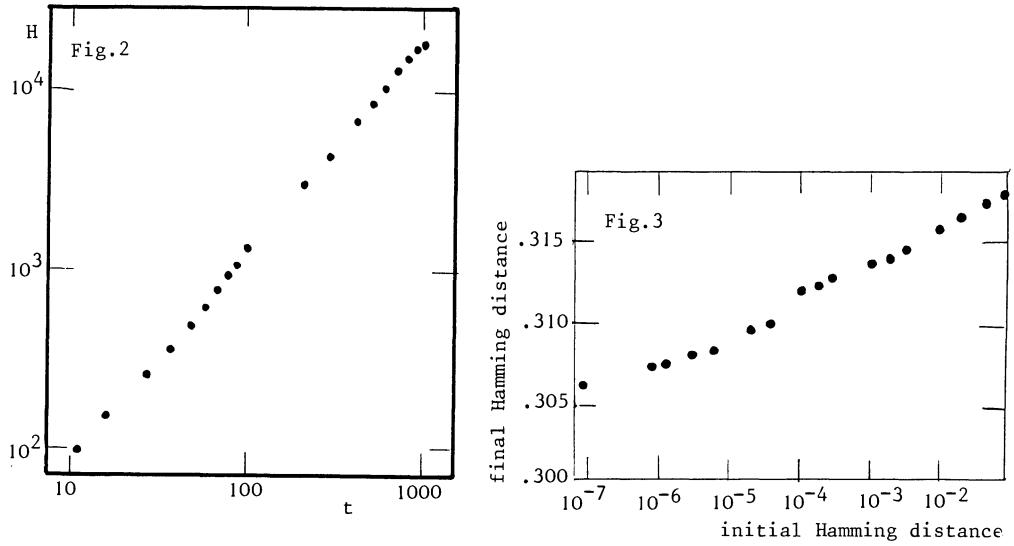


Fig.2: Log-log plot of Hamming distance H versus time t at the critical point, $p = 0.298$. Average of ten runs for $L = 5056$.

Fig.3: Semilogarithmic plot of final Hamming distance versus initial Hamming distance in the chaotic regime, $p=0.35$ ($L=3392$ and $L=5056$).

of the fraction of stable and unstable sites (Hansen²), remains to be explained.

In summary, we found $\gamma = 2.5$ and $\beta = 0.34$. Right at $p = p_c$ the Hamming distance increases with a power law in time, with an exponent near 1.2. The susceptibility in the chaotic phase shows an unexpected logarithmic variation.

1. S.A.Kauffman, J.Theor.Biol. 22, 437 (1969).
2. B.Derrida and D.Stauffer, Europhys. Lett. 2, 739 (1986); D.Stauffer, Phil.Mag.B 56, 510 (1987) and page 246 in: *Universalities in Condensed Matter*, ed. by R.Jullien, L.Peliti, R.Rammal and N.Boccara, Springer Verlag, Heidelberg 1988; L. de Arcangelis, J.Phys. A 20, L 369 (1987); A.Hansen, J.Phys. A 21, 2481 (1988); D.Stauffer, Physica D , Oct. 1989.
3. U.Keller, B.Thomas, and H.J.Pohley, J.Stat.Phys. 52, 1129 (1988).
4. B.Derrida and Y.Pomeau, Europhys. Lett. 1, 59 (1987); B.Derrida and G.Weisbuch, J.Physique 47, 1297 (1987); for analogies with neural networks see K.E.Kurten, J.Phys. A21, L 615 (1988).
5. L.R. da Silva and H.J.Herrmann, J.Stat.Phys. 52, 463 (1988).
6. D.Stauffer, Physics A 157, 645 (1989).
7. S.Wolfram, Rev.Mod.Phys. 55, 601 (1983) and *Theory and Applications of Cellular Automata*, World Scientific, Singapore 1986.
8. N.Packard and S.Wolfram, J.Stat.Phys. 38, 901 (1985)

APPENDIX: General Automata Program

In this central part of the da Silva – Herrmann – de Forcrand algorithm, NR gives the 16 values (0 or 1) of the rule, N(i,j) stores the old spins, and IA(i,j) the new spins, with 64 sites in one N or IA word. The lattice size is L*(L+2) with L/64 = MM; LP1=L+1; M1 = MM+1.

```

DATA INDH/1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4/,
1      INDV/1,2,3,4,1,2,3,4,1,2,3,4,1,2,3,4/
...
32    DO 32 J=1,MM
      N(1,J)=N(LP1,J)
      N(LP2,J)=N(2,J)
      DO 31 I=2,LP1
      N(I,0)=SHIFT(N(I,MM),63)
      31   N(I,M1)=SHIFT(N(I,1), 1)
      DO 5 J=1,MM
      DO 41 I=2,LP1
      NH(I,1)=AND(      N(I,J+1),      N(I,J-1))
      NH(I,2)=AND(      N(I,J+1) ,COMPL(N(I,J-1)))
      NH(I,3)=AND(COMPL(N(I,J+1)),      N(I,J-1))
      NH(I,4)=AND(COMPL(N(I,J+1)),COMPL(N(I,J-1)))
      NV(I,1)=AND(      N(I+1,J),      N(I-1,J))
      NV(I,2)=AND(      N(I+1,J) ,COMPL(N(I-1,J)))
      NV(I,3)=AND(COMPL(N(I+1,J)),      N(I-1,J))
      NV(I,4)=AND(COMPL(N(I+1,J)),COMPL(N(I-1,J)))
      41   DO 42 I=2,LP1
      42   IA(I,J)=0
      DO 43 K=1,16
      IF(NS(K).EQ.0) GOTO 43
      DO 44 I=2,LP1
      44   IA(I,J)=IA(I,J)+AND(NH(I,INDH(K)),NV(I,INDV(K)))
      43   CONTINUE
      5    CONTINUE

```

Mass Multifractality in Cluster Growth Models

T. Vicsek*

Department of Physics, Emory University, Atlanta, GA 30322, USA

*On leave from: Institute for Technical Physics, Budapest, P.O. Box 76, 1325 Hungary

Abstract It is demonstrated that clusters constructed by asymmetric recursion rules can be described in terms of mass multifractality. The set of generalized dimensions D_q associated with the geometry of such objects is determined numerically for the growing asymmetric Cantor set using the box counting and the sand box methods. These approaches are not found equally efficient in evaluating the D_q values. The direct determination of the $f(\alpha)$ spectrum corresponding to the singular distribution of mass in very large off-lattice diffusion-limited aggregates indicates the mass multifractality of these clusters as well.

1. Introduction

The growth of objects with fractal geometry [1-4] is a common process which can be observed in a wide range of phenomena in physics, chemistry, biology and technology. In many cases the reason for the fractal properties of the structures is the instability of the growth due to the presence of a non-local field satisfying the Laplace equation. Such diffusion-limited (or Laplacian) growth processes typically lead to branching fractals with deep fjords between the branches. The most studied examples include electrodeposition [5,6], crystallization [7,8], dielectric breakdown [9] and viscous fingering [10,11]. Although considerable amount of information has accumulated about these processes [4,12-15], the basic phenomenon still lacks theoretical description based on first principles. In addition, some of the questions related to the phenomenology of growing fractal objects are also open, including such problems as the role of noise in the asymptotic behaviour of patterns with no anisotropy or the multifractal nature of the mass distribution in these structures.

The velocity of the interface of Laplacian patterns is determined by the gradient of the field at the surface. From the related simulations and experimental studies it is known [16-20] that this gradient satisfies multifractal scaling [21-25]. As an example let us consider an isolated charged conductor. The electric field at the sharp tips of such an object becomes very large in accord with the behaviour of the Laplace equation for the electric potential. One expects that the electric field follows a multifractal distribution if the growing structure in the asymptotic limit has infinitely many tips of infinitely many types (the tips are surrounded by different local configurations determining the singularity of the gradient at the tip). It is rather natural to ask: Is the structure emerging as a result of such a complex growth velocity distribution a monofractal (i.e. an object which can be described in terms of a single exponent characterizing the distribution of mass in it) or does it become a (mass or geometrical) multifractal as well?

There are a few very recent results related to the above question. The evaluation of the generalized dimensions for relatively small diffusion-limited aggregates [26] (DLA) and electrodeposits led to the conclusion [27] that these objects can be regarded as monofractals. On the other hand, a small cell renormalization group calculation of DLA [28] and a study of sedimentary rock samples [29] suggested that the mass distribution in these systems can be described in terms of fractal measures. The latest results (Ref. 30 and this paper) about the structure of extremely large DLA clusters, however, indicate that the sizes of the objects considered in the above referenced investigations were too small to see the true asymptotic behaviour.

To attempt the clarification of some of the controversial points concerning mass multifractality we shall chose the following method. First, two types of growth processes will be considered: we shall determine the scaling properties of the mass distribution in both i) deterministic growth models and ii) off-lattice diffusion-limited aggregates. Second, several approaches will be used to calculate the relevant quantities in order to get a better insight into the numerical aspects of the problem through comparing the results obtained from the different techniques.

2. Mass multifractality

Before describing the results let us first introduce and demonstrate the concept of mass multifractality for cluster models according to Ref 31. We shall assume that the structure is defined on a lattice and its linear size and mass are L and M_0 respectively. Furthermore, a denotes the unit of the lattice on which the cluster is growing (it is also equal to the size of the particles which are represented by occupied lattice sites) and l is the lattice spacing (box size) of the grid which is put onto the cluster to calculate its fractal dimension. Then one can define M_i as the mass (the number of particles) of the i th box ($i = 1, 2, \dots$). Knowing the set of M_i values one can determine the quantity $N(M)$ which is the number of boxes with mass M . Assume that we plot $\ln N(M)$ versus $\ln M/M_0$ for various l . If these histograms fall onto the same universal (l independent) curve after rescaling both coordinates by a factor $\ln(l/L)$ [25], the structure is a geometrical multifractal [31]. From such a behaviour it follows that the distribution of mass has to satisfy the relations

$$M \sim M_0 \left(\frac{l}{L} \right)^\alpha, \quad (1)$$

and

$$N(\alpha) \sim \left(\frac{l}{L} \right)^{-f(\alpha)}, \quad (2)$$

where α is the mass index

$$\alpha = \frac{\ln M/M_0}{\ln l/L}, \quad (3)$$

$N(\alpha)$ is the number of boxes with mass index α , and $(l/L) \rightarrow 0$. It is important to point out that the condition

$$a \ll l \ll L \quad (4)$$

has to be satisfied when these definitions are considered. One can also introduce the so called generalized dimensions D_q which are defined through the moments of the box masses

$$\chi_q(M, l/L) \equiv \sum_i M_i^q \sim M_0^q \left(\frac{l}{L}\right)^{(q-1)D_q} \quad (5)$$

in the limit $\frac{l}{L} \rightarrow 0$. The generalized dimensions are related to the multifractal spectrum $f(\alpha)$ via the equation

$$(q-1)D_q = q\alpha(q) - f(\alpha(q)) \quad (6)$$

with $f'(\alpha(q)) = q$.

Next we use a deterministic cluster model to demonstrate how the above formalism describes mass multifractality. Deterministic models based on constructing a fractal using recursion represent a useful tool in studying fractal growth because they allow for exact treatment of a number of properties [31,32]. For example the scaling of the cluster size distribution in diffusion-limited deposits observed in the simulations holds exactly in a deterministic model for DLA [32]. Recently branching Julia sets were proposed as deterministic structures with a harmonic measure on them analogous to the growth probability distribution of Laplacian patterns [33]. Numerical integration of the Laplace equation represents a further method to the description of fractal growth by deterministic approaches [34,35].

Let us consider the following example for the construction of an aggregate type mass multifractal in two dimensions (Fig. 1a). The rule is quite simple: In the k th step the twice enlarged version of the configuration corresponding to the $k-1$ th stage of growth is added to the four corners of the already existing cluster. Another example is the growing asymmetric Cantor set which is shown in Fig. 1b. Its construction is analogous to that of the example displayed in Fig 1a. To calculate the generalized dimensions defined by (5) for such fractals one can use similarity arguments. Let us first consider a more general case and assume that our mass multifractal is constructed by the folowing procedure: in the first step ($k=1$) the i th part s obtained from the starting object by reduction using a rescaling factor $r_j < 1$ and is given a weight (mass) $m_j = r_j^d / \sum_j^n r_j^d$ (so that $\sum_j^n m_j = 1$), where n is the number of pieces the starting configuration is substituted with. The choice for m_j assures that the weight of a part is proportional to its volume, and because of this it can be considered as mass. At the next stage ($k=2$) each part is devided into n parts with weights reduced further by m_j and size rescaled by r_j . The multifractal is obtained in the $k \rightarrow \infty$ limit. Note that in this way one generates a fractal which instead of growing, remains finite, but becomes more and more ramified. However, for every finite k the two kinds of fractals (the growing and the finite) can be trivially rescaled into each other.

As a result of the above procedure, for $k \rightarrow \infty$ we obtain an object which can also be considered as having n fractal parts, each being a rescaled version of the whole fractal. The total mass associated with the j th part is m_j . Therefore,

$$\chi_{q,j}(\epsilon) = \sum_i m_j^q \chi_q(\epsilon/r_j), \quad (7)$$

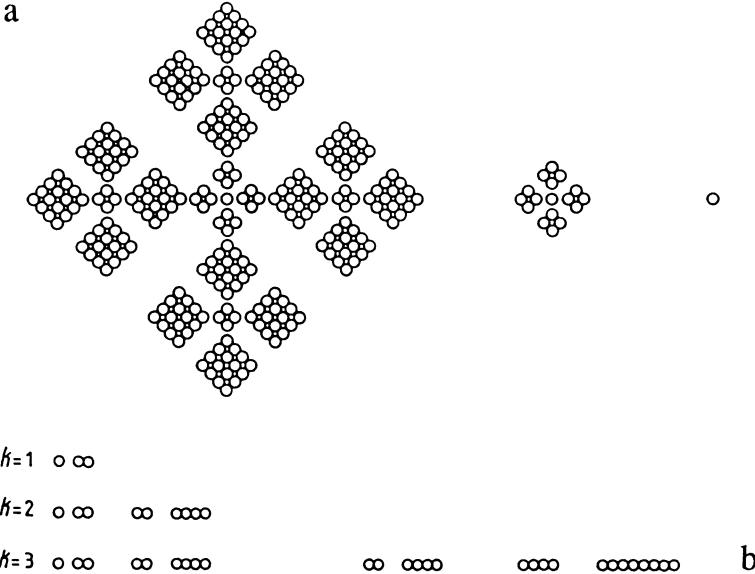


Figure 1. The construction of growing mass multifractal clusters embedded into $d = 2$ (Fig. 1a) and $d = 1$ (Fig 1b) dimensions.

where $\chi_{q,j}(\epsilon)$ is the quantity defined by (5) evaluated for the j th part, and $\epsilon = l/L$. Obviously,

$$\chi_q(\epsilon) = \sum_j^n \chi_{q,j}(\epsilon). \quad (8)$$

Using (5) and (7) in (8) we get the implicit equation for D_q

$$\sum_j^n m_j^q r_j^{(1-q)D_q} = 1 \quad (9)$$

which demonstrates that recursive fractal models with uniform distribution on them have a non-trivial D_q spectrum.

One of the goals of this paper is to demonstrate that the various methods which can be used to determine D_q -s are not equally efficient. As a geometrical multifractal to work with we shall chose the growing asymmetric Cantor set since it is simple enough to allow for extensive calculations with moderate computing power [36]. Before we discuss our results concerning the generalized dimensions of the growing asymmetric Cantor set we briefly review the definitions of the two methods which will be investigated.

Box counting: This is the more standard method, which was already implicitly described when the definition of $f(\alpha)$ was given. The object is covered with a lattice of unit size l and the number of non-empty boxes $N(l)$ is determined. Let us define the generalized box counting dimensions

$$D_q^{bc}(\epsilon) = \frac{\ln(\sum(M_i/M_0)^q)}{\ln \epsilon} \frac{1}{q-1}. \quad (10)$$

for arbitrary $\epsilon < 1$. Then we can investigate the conditions under which D_q^{bc} approximates the true D_q well enough. The definition (5) is recovered in the $\epsilon \rightarrow 0$ limit.

Sand box method: The main quantity used in this approach is the mass $M(R)$ (the number of particles) within a region of radius R centered on the fractal (typically on its origin). It is widely accepted that for growing fractals in the $R \rightarrow \infty$ limit the expression $M(R) \sim R^{D_0}$ determines the fractal dimension, where R is the actual length of the fractal, $R = L$, $M(L) = M_0$. However, as we shall show on a simple example, this statement does not hold, at least, for some geometrical multifractals. In fact, the generalized dimension defined by $M(L) \sim L^D$ exists, but its value is equal to $D_{-\infty}$, different from D_0 if the centre is the origin of the fractal.

If we chose an arbitrary point belonging to the fractal as the centre of the sandbox (instead of the origin), because of Eq. (3) the quantity $\ln \frac{M}{M(R)} / \ln \frac{L}{R}$ for $R \ll L$ approaches the local mass index α characteristic for the position of the given centre. Both α and the actual form of the function $\ln M(R)$ versus $\ln R$ depends on the choice for the centre. In order to obtain well defined dimensions independent of the local behaviour, we shall study the *average* value of the masses $M(R)$ and their powers over *randomly distributed* centres on the fractal. Thus, we are interested in the question how $\langle M^q(R) \rangle$ scales with increasing R . Rewriting (5) we find

$$\sum_i \left(\frac{M_i}{M_0} \right)^{q-1} \frac{M_i}{M_0} \sim \left(\frac{l}{L} \right)^{(q-1)D_q}. \quad (11)$$

On the basis of the above expression we define the sand box dimension of order q as

$$D_q^{sb}(R/L) = \frac{\ln \langle [M(R)/M_0]^{q-1} \rangle}{\ln(R/L)} \frac{1}{q-1}, \quad (12)$$

for any R between a and L , where the averaging is taken over centers randomly distributed within the fractal. In the following we shall investigate how D_q^{sb} is related to D_q and D_q^{bc} .

3. Results for the growing asymmetric Cantor set

Let us define the growing asymmetric Cantor set as shown in Fig. 1b. In the first step the (seed) structure is made of three particles placed at the first, third and the fourth sites, respectively. In the next step the twice enlarged version of the first configuration is added to the seed between the 9th and 16th sites. After the n th step of the construction the linear size of the structure is 4^n , and it is made of 3^n units. The fractal is obtained in the $n \rightarrow \infty$ limit.

The D_q spectrum can be determined for this fractal exactly. The general equation given in Ref. 31 for our case can be solved explicitly leading to

$$D_0 = \frac{\ln(\sqrt{5} + 1)}{\ln 2} - 1 \simeq 0.6942, \quad D_\infty = \frac{\ln(2/3)}{\ln(1/2)} \simeq 0.5849,$$

and $D_{-\infty} = \frac{\ln(1/3)}{\ln(1/4)} \simeq 0.7924.$ (13)

It is easy to check that M_0 scales with L according to $D_{-\infty}$ instead of D_0 !

During the application of the *box counting method* to the growing asymmetric Cantor set we put a grid on the fractal and gradually decreased the grid size from the largest possible size L to that of the particles [36]. We determined the number of particles (M_i) in each box by an exact recursion relation. In this way it was possible to study relatively large systems (corresponding to $L = 4^{50}$). In Figures 2a and 2b the generalized dimensions

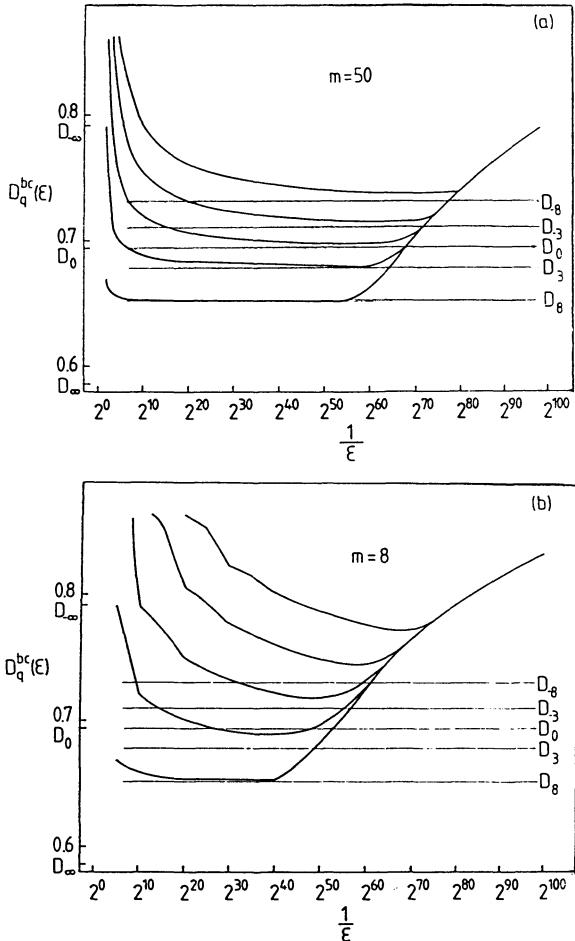


Figure 2. The box size dependent generalized dimensions D_q^{bc} determined from the box counting method versus $\ln \epsilon$. Fig. 2a corresponds to the case $n = 50$ (i. e. to $L = 2^{100}$), while Fig. 2b shows the analogous results obtained for $n = 8$ ($L = 2^{16}$).

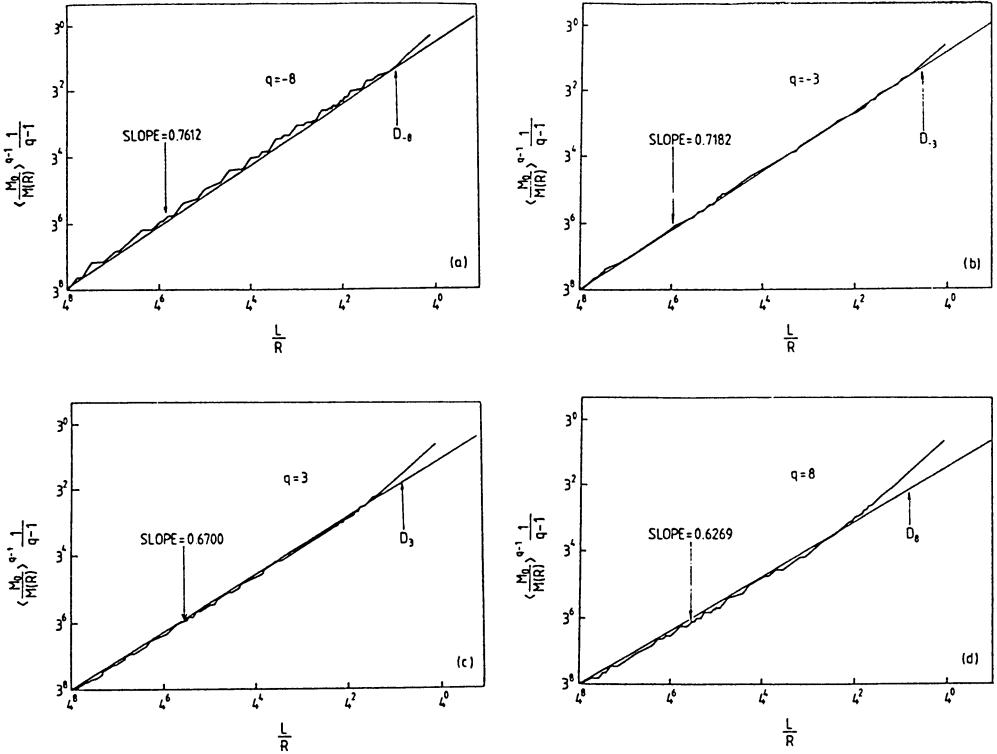


Figure 3. Estimation of the generalized dimensions D_q for $q = -8, -3, 3$ and 8 (a, b, c, and d, respectively) from the slopes of the plots $\ln\langle[M_0/M(R)]^{q-1}\rangle/(q-1)$ versus L/R . The averaging was made over 75 randomly positioned centres.

D_q^{bc} are shown as a function of $\ln \epsilon$. The q values were changed between -8 and 8. The curves we obtained have a region where they approach the value corresponding to the true D_q calculated from the exact solution. The length of this region strongly depends on both q and the size of the object. For larger q values the numerical results fit the exact one better. On the other hand, for $q = -8$ there is only a relatively narrow region (between 2^{60} and 2^{80}), where the box counting method gives reasonable results. Fig. 2b demonstrates that for smaller objects the convergence to the exact values is considerably worse.

The *sand box method* was applied using many randomly chosen centres [36]. In this case the number of particles was determined within a region of length R , where R was incremented from one particle size ($a = 1$) to W , the latter denoting the distance between a given centre and the most distant point belonging to the fractal. As expected, averaging over many centres gives well defined scaling. In this case we plotted $\langle(M_0/M(R))^{q-1}\rangle/(q-1)$ versus L/R (Fig. 3). Our numerical test suggests that the slope of these curves is in good agreement with the exact D_q values. The error is less than $\pm 3\%$ even for $\pm q = 8$ and $n = 8$. Note that for a structure of the same size box counting gives worse results (Fig. 2b).

4. $f(\alpha)$ for the mass distribution in diffusion-limited aggregates

We have seen that fractals generated in a recursive process can be multifractals in a geometrical sense. The big question, however, still remains: are there any phenomena in nature leading to mass multifractals? One expects that DLA clusters may have this property since they are produced during a process in which the growth probabilities are distributed according to a fractal measure.

It is far from trivial to answer the above question because of the condition (4). If one would like to be in the region where $l/a > 100$ and $L/l > 100$ the clusters have to have a diameter larger than 10^4 . Such linear sizes are just somewhat bigger than those of the the largest DLA clusters which have ever been generated. Obviously, results obtained on much smaller aggregates are not expected to give reasonable estimates for D_q .

In the present calculations coordinates of 5 off lattice DLA clusters of one million particles were used to obtain estimates of the mass multifractal spectrum. We followed a method which was recently argued [37] to be useful for the determination of the associated $f(\alpha)$. In this approach one first defines

$$\mu_i(q, l) = \frac{P_i^q(l)}{\sum_j P_j^q(l)} \quad (14)$$

with $P_i = M_i/M_0$, and then uses the expressions

$$f(q) = \lim_{l \rightarrow 0} \sum_i \mu_i(q, l) \ln \mu_i(q, l) \frac{1}{\ln l} \quad (15)$$

and

$$\alpha(q) = \lim_{l \rightarrow 0} \sum_i \mu_i(q, l) \ln P_i(q, l) \frac{1}{\ln l} \quad (16)$$

to evaluate the dependence of f on α as a function of the parameter q . The main advantages of the above procedure are that it provides $f(\alpha)$ directly, without the intermediate Legendre transformation of the values obtained for the D_q spectrum and it does not suffer from the poor sampling statistics which occur in other methods used for calculating $f(\alpha)$.

The results are presented in Fig. 4 , where the $f(\alpha)$ curves are shown for $q \geq 0$ and various box sizes l . The right part of the curves is not displayed because the data for $q < 0$ are not reliable due to the contribution from boxes having anomalously small number of particles in them. These boxes are present because in general one is not able to center them according to the underlying multiplicative process.

Fig. 4 suggests that as the box size is decreased, the $f(\alpha)$ curves tend to scale onto the same non-trivial function. The fact that the limiting $f(\alpha)$ seems to be a smooth curve instead of a single point (or a nearly vertical interval) indicates that the local singularity of the mass distribution in diffusion-limited aggregates can be described in terms of a fractal measure.

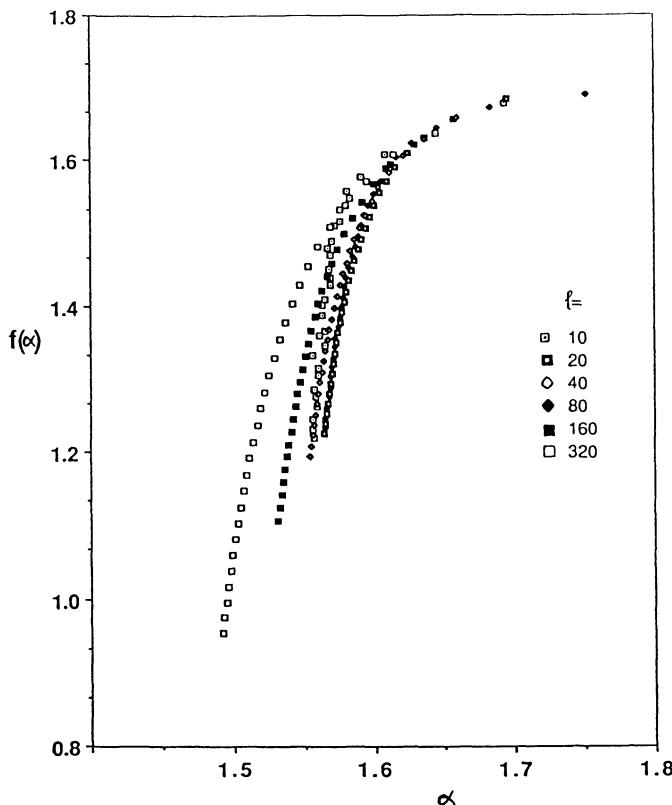


Figure 4. The spectrum of mass indexes α corresponding to the multifractal distribution of particle density in large off lattice diffusion-limited aggregates.

5. Conclusions

In conclusion, we have demonstrated that for geometrical multifractals the standard methods of determining fractal dimensions have to be applied with some precaution. In particular, the box counting algorithm gives reliable results only for cluster sizes *much larger* than those usually obtained in numerical simulations. During the application of the sand box algorithm it is essential to *average* over many randomly selected centres. Our results indicate that for geometrical multifractals the sand box method provides better estimates of the generalized dimensions, however, because of the necessary averaging this method requires considerably more computing time.

The mass multifractal spectrum of DLA clusters seems to be non-trivial, however, additional studies are needed to confirm our preliminary results which are in disagreement with the ones obtained from much smaller simulations. Further improvement of the results is expected from the application of the sand box approach to this problem.

Acknowledgements

Much of the work described here was carried out in collaboration with T. T  l and   . F  lop. The extremely large off-lattice DLA clusters were provided by P. Meakin. I have benefitted from the hospitality by F. Family extended to me during my stay at Emory University.

References

- [1] Mandelbrot, B. B., 1982 *The Fractal Geometry of Nature* (Freeman, San Francisco)
- [2] H. Takayasu, *Fractals in Physical Sciences* (Asakura Shoten Publishers, Tokio, 1985) (in Japanese)
- [3] J. Feder *Fractals* (Plenum Press, New York, 1988)
- [4] T. Vicsek *Fractal Growth Phenomena* (World Scientific, Singapore, 1989 jan.)
- [5] R. M. Brady and R. C. Ball, *Nature* **309**, 225 (1984)
- [6] M. Matsushita, M. Sano, Y. Hayakawa, H. Honjo and Y. Sawada, *Phys. Rev. Lett.* **53**, 286 (1984)
- [7] H. Honjo, S. Ohta and M. Matsushita, *J. Phys. Soc. Japan* **55**, 2487 (1986)
- [8] Gy. Radnocy, T. Vicsek, L. M. Sander and D. Grier, *Phys. Rev. A* **35**, 4012 (1987)
- [9] L. Niemeyer, L. Pietronero and H. J. Wiesmann, *Phys. Rev. Lett.* **52**, 1033 (1984)
- [10] Daccord, G., Nittmann, J., and Stanley, H. E., 1986 *Phys.Rev. Lett.* **56**, 336
- [11] Maloy, K. J., Feder J. and Jossang, J., 1985 *Phys. Rev. Lett.* **55**, 2681
- [12] H. Herrmann, *Phys. Rep.* **136**, 154 (1986)
- [13] P. Meakin, in *Phase Transitions and Critical Phenomena* edited by C. Domb and J. L. Lebowitz, Vol. 12, (Academic Press, New York, 1987) p. 365
- [14] Stanley, H. E. and Ostrowsky, N., editors, *Random Fluctuations and Pattern Growth* (Kluwer, Dordrecht, 1988)
- [15] R. Jullien and R. Botet *Aggregation and Fractal Aggregates* (World Scientific, Singapore, 1987)
- [16] T. C. Halsey, P. Meakin and I. Procaccia, *Phys. Rev. Lett.* **56**, 854 (1986)
- [17] C. Amitrano, A. Coniglio and F. di Liberto, *Phys. Rev. Lett.* **57**, 1016 (1986)
- [18] Y. Hayakawa, S. Sato and M. Matsushita, *Phys. Rev. A* **36**, 1963 (1987)
- [19] J. Nittmann, H. E. Stanley, E. Torboul and G. Daccord, *Phys. Rev. Lett.* **58**, 619 (1987)
- [20] S. Ohta and H. Honjo, *Phys. Rev. Lett.* **60**, 611 (1988)
- [21] B. B. Mandelbrot, *J. Fluid. Mech* **62**, 331 (1974)
- [22] U. Frisch and G. Parisi, in *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics* edited by M. Ghil, R. Benzi and G. Parisi (North-Holland, Amsterdam, 1985)
- [23] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia and B. Shraiman, *Phys. Rev. A* **33**, 1141 (1986)

- [24] T. Tél, *Z. Naturforsch*, **43a**, 1154, 1988
- [25] B. B. Mandelbrot in *Random Fluctuations and Pattern Growth* edited by H. E. Stanley and N. Ostrowsky (Kluwer, Dordrecht, 1988) p.333
- [26] T. A. Witten and L. M. Sander, *Phys. Rev. Lett* **47**, 1400 (1981)
- [27] F. Argoul, A. Arneodo, G. Grasseau and H. L. Swinney, *Phys. Rev. Lett.* **61**, 2558 (1988)
- [28] T. Nagatani, preprint
- [29] J. P. Hansen, J. L. McCauley, J. Muller and A. Skjeltorp, in *Random Fluctuations and Pattern Growth* edited by H. E. Stanley and N. Ostrowsky (Kluwer, Dordrecht, 1988) p.369
- [30] P. Meakin, private communication
- [31] T. Tél and T. Vicsek, *J. Phys. A* **20**, L835 (1987)
- [32] T. Vicsek, *J. Phys. A* **16**, L647 (1983)
- [33] I. Procaccia and R. Zeitak, *Phys. Rev. Lett.* **60**, 2511 (1988)
- [34] P. Garik, R. Richter, J. Hautman and P. Ramanlal, *Phys. Rev. A* **32**, 3156 (1985)
- [35] F. Family, D. Platt and T. Vicsek, *J. Phys. A* **20**, L1177 (1987)
- [36] T. Tél, A. Fülöp and T. Vicsek, preprint
- [37] A. Chhabra and R. V. Jensen, preprint

Noise Reduction in DLA

M.A. Novotny

Supercomputer Computations Research Institute, Florida State University,
Tallahassee, FL32306, USA

Abstract. The question of whether noise reduction changes the asymptotic behavior of DLA-type aggregates is tested. The models used are DLA with short range interactions and DLA with a nearest-neighbor dependent sticking probability. It is shown that in some cases noise reduction can change the asymptotic behavior of DLA-type aggregates.

1. Introduction.

In recent years there has been a great deal of experimental and theoretical interest in growth processes and associated pattern formation. Excellent overviews are provided in conference proceedings^{1,2,3} and recent books.^{4,5} Much of the computer simulation work on growth models has been done on Diffusion Limited Aggregation (DLA) models, the first of which was introduced in 1981 by Witten and Sander.⁶ The quantity of work on DLA is easily justified both in terms of using DLA as a ‘fruit fly’ to understand general pattern formation models, and by the applicability of DLA-type models to various physical phenomena. These physical phenomena include the irreversible kinetic aggregation of gold colloids⁷ and polystyrene spheres,⁸ electrodeposition,⁹ electric breakdown,¹⁰ viscous fingering in Hele-Shaw cells,^{11,12} the formation of snowflakes,¹³ and fluid-fluid displacement in porous media.^{14,15}

Most of the theoretical work which has been performed on DLA-type models has been done using computer simulations. Given a DLA-type rule (which should be motivated by physical considerations), how does one extract the relevant physical quantities, such as the fractal or multifractal nature of the model? In principle the answer to this is trivial: grow large enough aggregates so that the finite size effects are made arbitrarily small. Unfortunately, due to limitations in the amount of available computer time and memory this is not always possible. Nowhere is this more evident than in the study of the DLA model on a square lattice. In this case, small to medium sized aggregates are approximately circularly symmetric [see fig. 2(a)], while larger aggregates show a pronounced tendency to develop a Swiss cross shape.¹⁶ Even world-record size aggregates (which have a few times 10^6 particles) have a fractal dimension which is still changing with the size of the aggregate — so finite size effects are still present.

In order to overcome the computer limitations, a method which has come to be called noise reduction has been introduced for DLA-type algorithms.¹⁷⁻¹⁹ The idea is to introduce a modification to the DLA algorithm which would reduce the finite size effects without affecting the asymptotic behavior of the model. There has been some arguments^{20,21} about whether noise reduction changes the asymptotic behavior of the DLA-type model. This is the question that will be addressed in this paper.

2. Models and Methods.

Although DLA-type models can be implemented either on a given lattice or without any lattice, in this paper I will limit myself to discussions of DLA-type models on a square lattice. In the original formulation of the DLA algorithm,⁶ a seed particle is placed at the origin of the lattice. Then a random walker is released from infinity (or uniformly from a circle with a radius larger than the aggregate size and centered at the origin). When the random walker strikes the growing aggregate (which it will always do in two dimensions), it sticks to the aggregate at the point of contact. Then another random walker is released. Continuing this process gives aggregates like the one shown in fig. 2(a).

In order to decrease the average amount of computer time required before a walker sticks, a semi-lattice modification to DLA-type algorithms has been introduced.^{22,23} All simulations in this paper have been performed on a square lattice using the semi-lattice method. The basic idea is illustrated in fig. 1. The current aggregate is surrounded by a square lattice up to a certain distance. (I have used a distance of 15 lattice units.) Once a walker is released from the starting circle, it can move in one step to any uniformly chosen position on a circle which is centered on the current position of the walker and has a radius which overlaps the lattice surrounding the aggregate by one lattice unit. Once the walker moves onto the lattice, it moves with an equal probability to each of the 4 nearest-neighbor sites. It is possible for the walker to repeatedly go from on lattice moves to off lattice moves — as illustrated in fig. 1. In order to prevent walks which wander out to infinity, if the walker gets too far from the aggregate it is killed, and another one is released from the starting circle. (The radius of the kill circle used in these simulations was 20 times the maximum aggregate size.) Considerations for implementing DLA-type algorithms on vector and parallel computers have recently been described.²⁴

Noise reduction¹⁷⁻¹⁹ is implemented by placing a counter, which is initially set to zero, at each of the possible growth sites. Once a particle has decided to stick to the aggregate at that growth site, the counter is incremented by one. If the value of the counter is greater than or equal to some threshold value, m , the particle grows at the chosen site. Otherwise, the particle is discarded and another walker is released from the starting circle. It is important to remember that when a particle sticks, all the counters at other growth sites are unchanged.

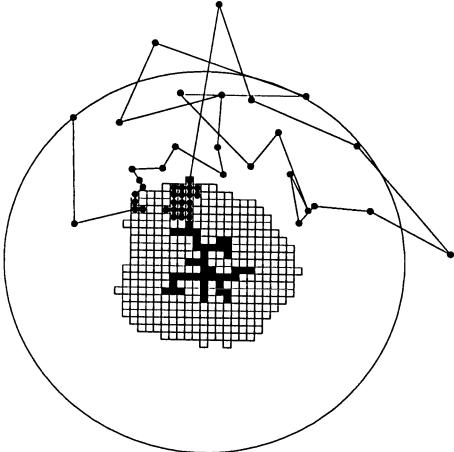


Fig. 1 The semi-lattice growth algorithm for a square lattice is shown. The black squares in the center are particles which have already grown to form the aggregate. The surrounding white squares comprise the lattice which is grown around the aggregate within 7 lattice units from any aggregate particles. In this region the incoming random walker makes on-lattice steps. In the surrounding region, the walker moves off-lattice with a step size determined by the distance between the walker and the lattice grown around the aggregate. The large circle shown is the launch circle for the random walker. The path of one incoming random walker is shown.

DLA aggregates for three values of m are shown in fig. 2, including $m=1$ which corresponds to normal DLA. To understand how noise reduction works, imagine that there were only two possible growth sites, with arrival probabilities for the random walker of a and b . (The walker must hit one site, so $a+b=1$.) If $m=2$, the probability of growth at a is $p_a = aa + 2a^2b$ and at b is $p_b = 1 - p_a$. This is because the sequence of arrivals which can give growth at site a is $aa+aab+aba$, with the latest arrival given by the left-most member of the sequence. For general m the growth probability at site a is given by

$$p_a = a^m \sum_{i=0}^{m-1} \binom{m+i-1}{i} b^i.$$

Assuming $p_a > p_b$, the way p_a increases with m is shown in fig. 3. Thus for growth probabilities, noise reduction causes ‘the rich to get richer and the poor to get poorer’.

One modification of DLA which has been introduced is to allow the incoming walker to have an interaction with the aggregate particles.²⁵ This interaction changes the probability that the walker will move to each of the 4 nearest plaquettes on the lattice. Each plaquette is assigned the number 1 initially. Then when the plaquette gets within a certain distance from the lattice, the number

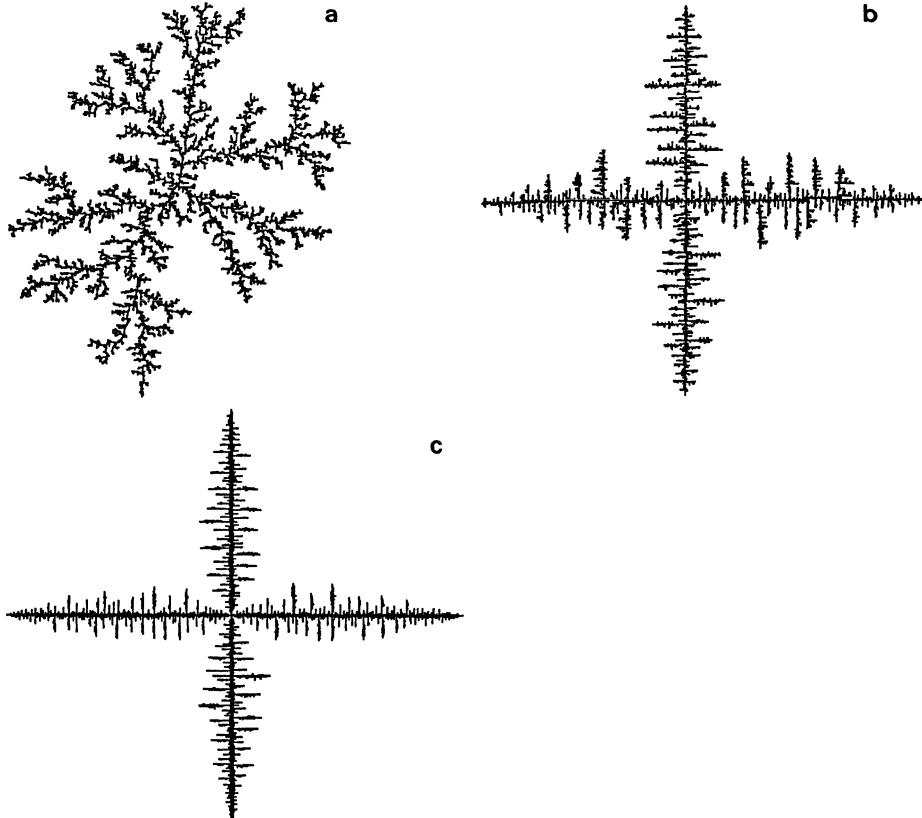


Fig. 2 DLA aggregates are shown on a square lattice. (a) shows a normal DLA aggregate, grown without any noise reduction ($m=1$). (b) shows a DLA aggregate grown with $m = 10$, and (c) shows a DLA aggregate grown with $m=100$. Each aggregate was grown on a 500×500 lattice.

in it is multiplied by the appropriate interaction strength, w . The probability that the walker moves to a given nearest neighbor plaquette is equal to the number in that plaquette divided by the sum of the numbers in the 4 nearest neighbor plaquettes. In this paper, a nearest neighbor 2-body repulsive force and 3 different 3-body forces have been included. See fig. 4 for representative aggregates. Is it possible to change the predominate direction of growth from along the diagonals as in fig. 4(a) to along the axes as in fig. 4(b)? This question will be addressed in Sec. 3. Recently, surface tension has been introduced in DLA by allowing the interactions to depend on the local curvature.²⁶

An alternative method of introducing surface tension into DLA is to change the probability that a particle sticks to the aggregate when it strikes the aggregate. If the sticking probability depends on the local curvature, this incorporates surface tension into DLA.^{27,28} An elementary estimation of surface tension can

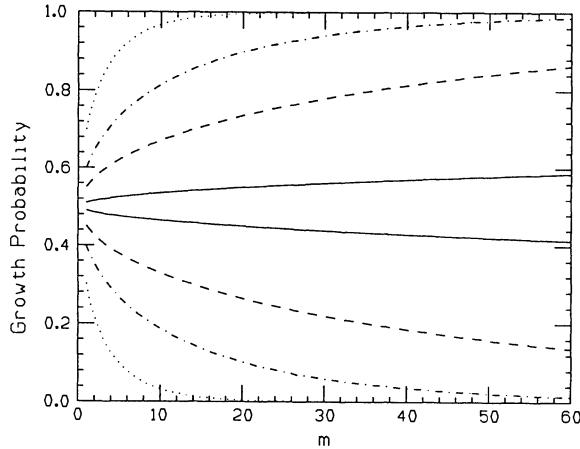


Fig. 3 The growth probability as a function of the noise reduction parameter m is shown for a system with two growth sites. The solid, dashed, dot-dashed, and dotted lines correspond respectively to initial values at $m=1$ of 0.51, 0.55, 0.60 and 0.70. The upper (lower) curves are the growth probabilities for the site with the largest (smallest) initial growth probability.

be done by including only nearest neighbors effects. A model for this, where the sticking probability depends on a parameter α as $p = \alpha^{3-B}$, with B the number of nearest neighbor plaquettes occupied by aggregate particles, was introduced by Witten and Sander²⁹ and further investigated by Banavar *et al.*³⁰ and Aukrust *et al.*³¹ Note that when $\alpha = 1$ this model is normal DLA, and on a square lattice the predominate direction of growth is along the axes.¹⁶ However, it has recently been shown³¹ that for α sufficiently small the predominant direction of growth is along the diagonals. Thus there is some critical value α_c at which the growth changes from along the axes to along the diagonals. Two aggregates grown with noise reduction for different values of α are shown in fig. 5. Fig. 5(a) has $\alpha < \alpha_c$ and fig. 5(b) has $\alpha > \alpha_c$. In the next section, the dependence of α_c on m will be presented.

3. Data and Analysis.

From fig. 5, for the α model of DLA there exists a critical value α_c at which a change from growth along the diagonals to growth along the axes occurs. In ref. 31 it was also found that the scaling behavior of the effective fractal dimension, D_f^{eff} , as a function of the size of the aggregate is different above and below α_c . Since both the scaling behavior and the preferred direction of growth change, α_c can be regarded as a critical point. It is possible, however, that there is more than one critical point lying near α_c . As α_c is approached, larger aggregates must be grown before the aggregate decides to grow along the

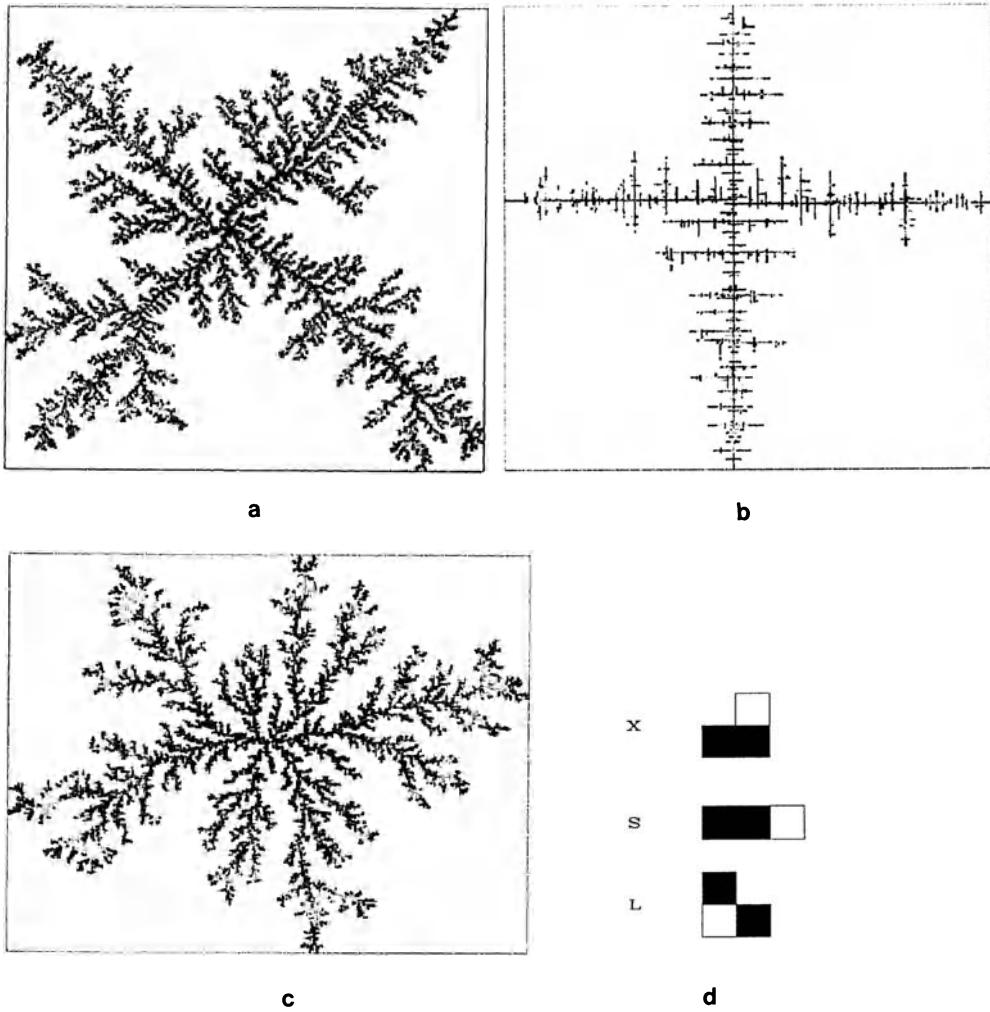


Fig. 4 Three different aggregates for interacting DLA on a 1000×1000 lattice are shown. Each was grown with a 2-body repulsive interaction of 0.01 and 3-body attractive interactions with strength 10^4 . The 3-body interactions used are shown in (d). The aggregates in (a), (b), and (c) correspond to the interactions labeled (L and x), (L and s), and (x and s). The number of particles in the aggregates are 39485, 9060, and 34387 respectively. In (d) the 3-body interactions are shown, with filled squares corresponding to sites on the aggregate and the empty square is where the 3-body interaction strength w is applied. These are only representative interactions, all interactions related to these by reflection and rotation symmetries are also included.

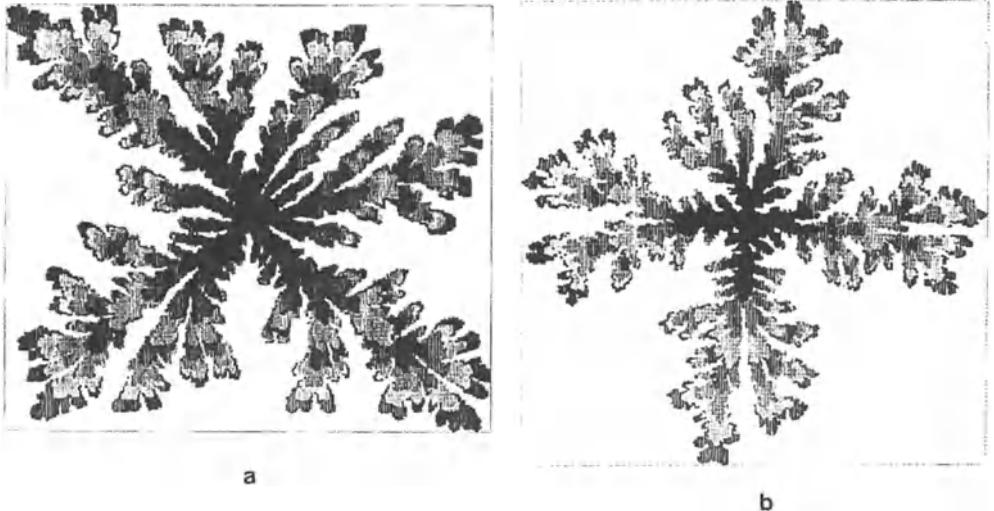


Fig. 5 Two aggregates for the α model which were grown with a noise reduction parameter $m = 10$ are shown. (a) shows a cluster of 1.1×10^5 particles grown with $\alpha = 0.3$. (b) shows a cluster of 5.5×10^4 particles grown with $\alpha = 0.4$. Both aggregates are shown on lattices which are about 1000×1000 . The different shades of grey correspond to different arrival times for the particles.

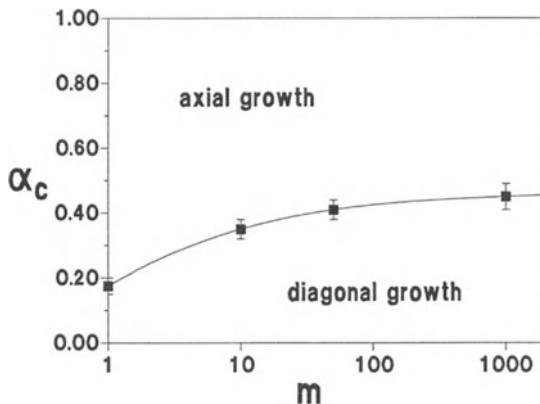


Fig. 6 The location of α_c for the α model as a function of the noise reduction parameter m is shown.

axes or along the diagonals. Hence, from simulations on finite size systems, only a bracket on α_c can be obtained. Fig. 6 shows³¹ the dependence of α_c on the amount of noise reduction, m . Clearly, α_c increases with m . Consequently, in this case there is a measurable dependence of the asymptotic behavior on the amount of noise reduction. For a fixed value of α it is thus possible that noise reduction will change the asymptotic behavior of the DLA aggregates.

For the interacting DLA model, as the amount of different 3-body interactions are changed, again the aggregates change from growth predominately along the axes [fig. 4(b)] to growth predominately along the diagonals [fig. 4(a)]. Hence there should be a critical value of the 3-body interaction strength. Rather than investigate the entire two dimensional phase diagram in the two 3-body interactions that are different in fig. 4(a) and 4(b), I choose to keep the total amount of the total 3-body interactions constant. (Similar results are expected for other cuts through the critical line.) Thus the 3-body interactions x and s given in fig. 4(d) were taken to satisfy $10^6 = w_s^2 + w_x^2$. The strength w_L was kept fixed at 10^3 . The critical value of the strength of the w_s interaction is somewhere between 100 and 450 for the case of no noise reduction. As the amount of noise reduction increases, the size of the error estimates for the critical value of w_s decreases. However, as seen in the fig. 7, for the lattice sizes used it is not clear whether or not the critical value of w_s changes with m .

It is also possible to use noise reduction to give spiral aggregates. (I thank R. K. P. Zia for this suggestion.) In fig. 8, representative aggregates for two values of noise reduction m are shown. In this simulation, the amount of noise

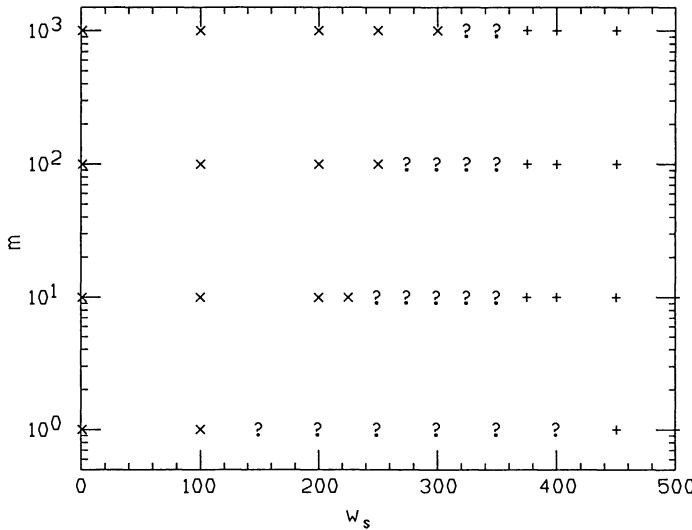


Fig. 7 The location of change from axial to diagonal growth as a function of the noise reduction parameter m is shown for interacting DLA. The 2-body interactions were equal to 0.1, and the 3-body interaction labeled L in fig. 4(d) is equal to 10^3 . The horizontal axis is the value of the 3-body interaction labeled s in fig. 4(d), and the strength of the 3-body interaction labeled x in fig. 4(d) varies to keep the sum of the squares of the 3-body interactions w_x and w_s equal to 10^6 . The symbol \times (+) stands for growth predominately along the diagonals (axes). The symbol ? stands for clusters for which aggregates grown on a 500×500 lattice did not have a clear predominate direction of growth.



Fig. 8 Two spiral DLA aggregates grown with noise reduction implemented as described in the text are shown. (a) shows an aggregate of 1.1×10^4 particles with $m=10$. (b) shows an aggregate of 1.1×10^4 particles with $m=100$.

reduction at a newly grown site was taken to be a vector \vec{m} with magnitude m at an angle $-\pi/2$ from the vector from the initial seed to the newly grown site. The amount of noise reduction at newly formed growth sites was then given by the components of \vec{m} . As seen in fig. 8, larger values of m give tighter spirals.

4. Discussion and Conclusion.

The two DLA-type algorithms studied in this paper both have a critical value of some parameter at which growth changes from predominately along the axes to predominately along the diagonals. For the case of α -DLA, inclusion of noise reduction changes α_c . For the interacting DLA studied, the critical value was not measurably changed with the amount of noise reduction.

This study demonstrates that the amount of noise reduction can change the asymptotic behavior, or loosely the ‘universality class’, of DLA aggregates. Consequently, noise reduction must be used *with extreme care* in the study of growth models. For the particular model of interest, one *must* extrapolate to the $m = 1$ limit. Although this study has been for two specific DLA models, the note of caution should be extended to studies of other growth models. This includes DLA on two^{12,17–19} and higher dimensional lattices,³² dielectric breakdown models,³³ growth using the Laplace equation,³⁴ Eden models,^{35,36} and other growth models.

Which way does noise reduction change the critical value of the model? Does increasing m increase the region of the axial growth or of the diagonal growth? As shown in fig. 3 the effect of noise reduction is to increase the growth probability of the growth sites that already have the largest probability of growing. For a DLA aggregate that first grows into a compact square, as

in the α model for small α , the largest growth probability is at the corners of the square. For a model where the typical finger thickness is about one lattice unit, such as for normal DLA [fig. 2(a)] or for some cases of interacting DLA [fig. 4(b)], the largest growth probability is at the tips of the longest fingers. A plausible hypothesis is that if the DLA model has a finger thickness greater than one lattice unit, increasing noise reduction will change the critical value to give a larger region of growth along the diagonals. Conversely, if the finger thickness is one lattice unit, increasing noise reduction will change the critical value to give a larger region of growth along the axes. It would be interesting to find an example where the last situation is realized — since this would mean that for normal DLA noise reduction may be a *relevant* parameter.

Acknowledgements. This research is supported in part by the Florida State University Supercomputer Computations Research Institute which is partially funded through Contract # DE-FC05-85ER25000 by the U.S. Department of Energy. Useful discussions with T. Aukrust, D. A. Browne, K. Kaski, D. P. Landau, P. Meakin, R. Tao, T. Vicsek and R. K. P. Zia are acknowledged. Special thanks to T. Aukrust for allowing the use of the figures for α -DLA.

References

- [1] *Kinetics of Aggregation and Gelation*, ed. F. Family and D. P. Landau, (North-Holland, Amsterdam, 1984).
- [2] *On Growth and Form, Fractal and Non-fractal Patterns in Physics*, ed. H. E. Stanley and N. Ostroskey, (Martinus Nijhoff, Boston, 1986).
- [3] *Time-Dependent Effects in Disordered Materials*, ed. R. Pynn and T. Riste, (Plenum Publishing Corporation, New York, 1987).
- [4] *Fractal Growth Phenomena*, T. Vicsek, (World Scientific, London, 1988).
- [5] *Fractals*, J. Feder, (Plenum Publishing Corporation, New York, 1988).
- [6] T. A. Witten and L. M. Sander, *Phys. Rev. Lett.* **47**, 1400 (1981).
- [7] D. A. Weitz, J. S. Huang, M. T. Lin and J. Sung, *Phys. Rev. Lett.* **54**, 1416 (1985).
- [8] A. T. Skjeltorp, in *Time-Dependent Effects in Disordered Materials*, ed. R. Pynn and T. Riste, (Plenum Publishing Corporation, New York, 1987).
- [9] R. M. Ball and R. C. Brady, *Nature* **309**, 255 (1984).
- [10] Y. Sawada, S. Ohta, M. Yamazaki and H. Honjo, *Phys. Rev. A* **26**, 3557 (1982).
- [11] J. Nittmann, G. Daccord and H. E. Stanley, *Nature* **314**, 141 (1985).

- [12] C. Tang, *Phys. Rev. A* **31**, 1977 (1985).
- [13] S. Miyazima and T. Tanaka, *J. Phys. Soc. Japan* **56**, 441 (1987).
- [14] L. Paterson, *Phys. Rev. Lett.* **52**, 1621 (1984).
- [15] K. J. Måløy, J. Feder and T. Jøssang, *Phys. Rev. Lett.* **55**, 2688 (1985).
- [16] P. Meakin, R. C. Ball, P. Ramanlal and L. M. Sander,
Phys. Rev. A **35**, 5233 (1987).
- [17] J. Szép, J. Cserti and J. Kertész, *J. Phys. A* **18**, L413 (1985).
- [18] J. Kertész and T. Vicsek, *J. Phys. A* **19**, L257 (1986).
- [19] P. Meakin, *Phys. Rev. A* **36**, 332 (1987).
- [20] J. Kertész, T. Vicsek and P. Meakin, *Phys. Rev. Lett.* **57**, 3303 (1986).
- [21] L. A. Turkevich and H. Scher, *Phys. Rev. Lett.* **57**, 3304 (1986).
- [22] P. Meakin, *J. Phys. A* **18**, L661 (1985).
- [23] R. Ball, R. M. Brady, G. Rossi and B. R. Thompson,
Phys. Rev. Lett. **55**, 1406 (1985).
- [24] M. A. Novotny and T. Aukrust, in *Proc. of the Second International Conference on Vector and Parallel Computing Issues in Applied Research and Development*, ed. P. Gaffney, (Ellis Horwood, London, in press).
- [25] M. A. Novotny, in *Time-Dependent Effects in Disordered Materials*,
ed. R. Pynn and T. Riste, (Plenum Publishing Corporation,
New York, 1987).
- [26] R. Tao, M. A. Novotny, and K. Kaski, *Phys. Rev. A* **38**, 1019 (1988).
- [27] T. Vicsek, *Phys. Rev. Lett.* **53**, 2281 (1984);
Phys. Rev. A **32**, 3084 (1985).
- [28] L. P. Kadanoff, *J. Stat. Phys.* **39**, 267 (1985).
- [29] T. A. Witten and L. M. Sander, *Phys. Rev. B* **27**, 5686 (1983).
- [30] J. R. Banavar, M. Kohmoto and J. Roberts, *Phys. Rev. A* **33**, 2065 (1986).
- [31] T. Aukrust, M. A. Novotny, D. A. Browne and K. Kaski,
Phys. Rev. A **39**, 2587 (1989).
- [32] P. Meakin, *Phys. Rev. A* **27**, 604 (1983).
- [33] J. Nittmann and H. E. Stanley, *Nature* **321**, 663 (1986).
- [34] J. Nittmann and H. E. Stanley, *J. Phys. A* **20**, L981 (1987).
- [35] J. Kertész and D. E. Wolf, *J. Phys. A* **21**, 747 (1988);
D. E. Wolf and J. Kertész, *J. Phys. A* **20**, L257 (1987);
D. E. Wolf and J. Kertész, *Europhys. Lett.* **4**, 651 (1987).
- [36] J. M. Kim and J. M. Kosterlitz, *Phys. Rev. A*, submitted.

Damage Spreading

H.J. Herrmann

SPhT, CEN Saclay, F-91191 Gif-sur-Yvette Cedex, France

Abstract. The concept of damage spreading in the context of Monte Carlo is introduced. Subtle differences between heat bath and Glauber dynamics give rise to totally different behavior in the damage spreading. In general, damage spreading defines new dynamical phase transitions which in some cases can be related exactly to usual thermal critical points. In spin-glasses, damage spreading can even generate multifractality.

1. Introduction

In the last years a new way of looking at Monte Carlo has led to the concept of damage spreading. The idea came from work that had been done on cellular automata and that ultimately originated from dynamical systems theory although there are some important differences.

Monte Carlo can be viewed as a dynamical process in phase space, i.e., starting from a given initial configuration one follows a trajectory in phase space under the application of the Monte Carlo procedure. This trajectory will of course depend on the specific type of Monte Carlo (heat bath, Glauber, Metropolis, Kawasaki, Kadanoff-Swift, etc.) but also on the sequence of random numbers. The trajectories will always go towards equilibrium which means that the configurations will be visited by the trajectory with a probability proportional to the Boltzmann factors. Once in equilibrium, the trajectories will stay there as assured by the detailed-balance condition.

The interesting question that we will ask now and which is the central issue of dynamical system theory is how much will a trajectory depend on the initial condition. More precisely: Suppose we make a small perturbation in the initial condition. Will the new trajectory be just slightly different or will it be totally different. The second case in which two initially close trajectories will become very different very fast is generically called chaos. The detailed definition of chaos may in some cases also include the speed with which the trajectories separate but in our context we do not want to make this distinction.

In order that the concept of closeness of trajectories be meaningful one must define what is a distance in phase space. Suppose we consider a system of Ising variables then a useful metric can be given by the "Hamming distance" or "damage"

$$\Delta(t) = \frac{1}{2N} \sum_i |\sigma_i(t) - \rho_i(t)| \quad (1)$$

where $\{\sigma_i(t)\}$ and $\{\rho_i(t)\}$ are the two (time-dependent) configurations in phase space and N is the number of sites, labeled by i . This definition is certainly arbitrary and the results that we present in the following depend on this definition. Physically it just measures the fraction of sites for which the two configurations are different. The definition of Eq. (1) can easily be generalized to q -state variables or continuous variables.

According to the discussion above, one would then call a dynamical behavior chaotic when in the thermodynamic limit $\Delta(t)$ goes to a finite value for large times if $\Delta(0) \rightarrow 0$. The opposite of chaotic is called frozen, i.e., when $\Delta(\infty) = 0$ if $\Delta(0) \rightarrow 0$. If one wants to get the limit $\Delta(0) \rightarrow 0$ properly in a numerical simulation one can use the following trick [1]: Consider three configurations $\{\sigma_A\}$, $\{\sigma_B\}$ and $\{\sigma_C\}$ with $\Delta_{AB}(0) = \Delta_{BC}(0) = 1/2\Delta_{AC}(0) = S$ where S is a fixed small number and Δ_{AB} denotes the distance between $\{\sigma_A\}$ and $\{\sigma_B\}$. Then

$$\Delta(t) = \Delta_{AB}(t) + \Delta_{BC}(t) - \Delta_{AC}(t) \quad (2)$$

is a very good extrapolation to $\Delta(0) \rightarrow 0$.

It is now clear why the procedure that we will follow has been called "damage spreading." One creates at $t = 0$ an initial damage $\Delta(0)$ in one configuration (which gives a second configuration) and watches if the damage spreads (chaotic situation) or if it heals (frozen situation). This kind of spreading has been investigated before [2] in cellular automata to describe for instance the influence of perturbations in genetic regulatory systems.

The crucial idea in order to study damage spreading in Monte Carlo is therefore to take two configurations and to apply on them the same sequence of random numbers in the Monte Carlo algorithm. In this way one is taking the same dynamics for both configurations. To get statistically meaningful results one must then average over many initial configurations of equal initial distance and over many different sequences of random numbers. It is also important to note that as in deterministic dynamical systems, once the two configurations become identical they will always stay identical.

2. Numerical Results for the Ising Model

In order to see if Monte Carlo can generate chaotic behavior in the sense outlined above, Glauber dynamics was applied in Ref. 1 on the two dimensional Ising model. Using Eq. (2) the distance of initially close equilibrium configurations was calculated after 10^4 updates per site, i.e., after a long time. The result is shown in Fig. 1 as a function of temperature. Let us remark that of course since the calculation was performed in a finite system of size L the final distance would have vanished after a time of the order of $\exp(L^2)$ because eventually two uncorrelated trajectories in a finite phase space will always meet but the times considered in Fig. 1 are much smaller than these "Poincaré" times. We see from Fig. 1 that above a certain temperature T_s the Glauber dynamics is chaotic while below T_s it is frozen. The order parameter for this transition between chaotic and frozen is just the quantity plotted in Fig. 1. The data seem to indicate that T_s is very close if not identical to the critical temperature T_c of the Ising model. It has not been possible up to now to determine the critical exponent β of this transition because the statistical fluctuations are very strong.

In three dimensions using Glauber dynamics qualitatively the same picture emerges as in two dimensions only the temperature T_s seems to be 4% below T_c as found in Ref. 3.

Particularly interesting is the fact that one finds a transition line even in a homogeneous field [4] as shown in Fig. 2. This indicates clearly that the dynamic transition between chaotic and frozen is not identical to the transition between the ferromagnetic and the paramagnetic phase. The transition line of Fig. 2b does not seem to agree either with the percolation transition of minority spins [4]. It is thus one of the open, challenging questions if this dynamical transition is related to any known property of the Ising model or if it is a totally novel phenomenon.

It was a lucky coincidence that independently the same questions were asked using a slightly different dynamics, namely heat bath. The authors [5] calculated the fraction $P(t)$ of the pairs of configurations that had not yet become identical after a time t and the average distance $D(t)$ between only those not yet identical configuration, so that $\Delta(t) = P(t).D(t)$. They started with not thermalized configurations and looked at different values of initial damage. Their result is shown in Fig. 3 after a time of 500 updates per site. One sees that the survival of damage depends on the initial damage (Fig. 3a) while when the configurations are different their distance does not depend on the initial damage (Fig. 3b). We also see that if the initial distance goes to zero the final distance is also going to vanish because of $P(t)$. This is in agreement with the result found for Glauber dynamics in the ferromagnetic phase. In the paramagnetic phase, however, the result for heat bath is strikingly different from Glauber dynamics because as seen from Fig. 3 heat bath does not show chaotic behavior but on the contrary is in a frozen phase in which the final distance vanishes even if the initial distance was big. The opposite behavior of the two dynamics is very surprising because normally it is thought that heat bath and Glauber dynamics are identical for the Ising model. We will investigate this question in the next section.

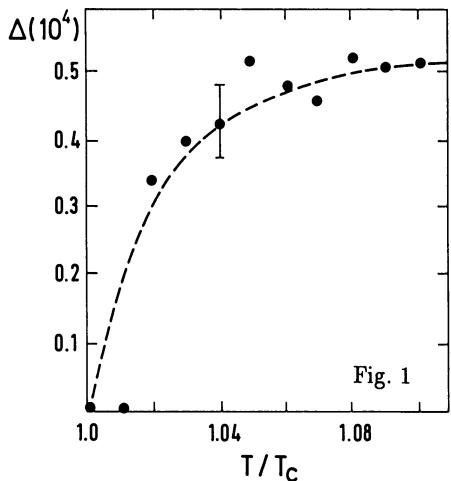


Fig. 1

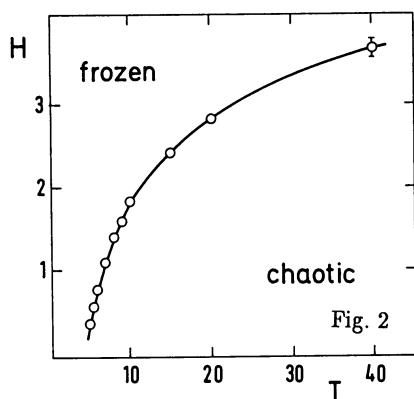


Fig. 2

Fig. 1 Hamming distance after 10^4 time steps for $\Delta(0) \rightarrow 0$ using Glauber dynamics as a function of T/T_c for the two dimensional Ising model (from Ref. 1).

Fig. 2 Phase diagram in the field-temperature plane between the chaotic and the frozen phase for the three-dimensional Ising model obtained by Le Caer [4] for Glauber dynamics in a 10^3 lattice.

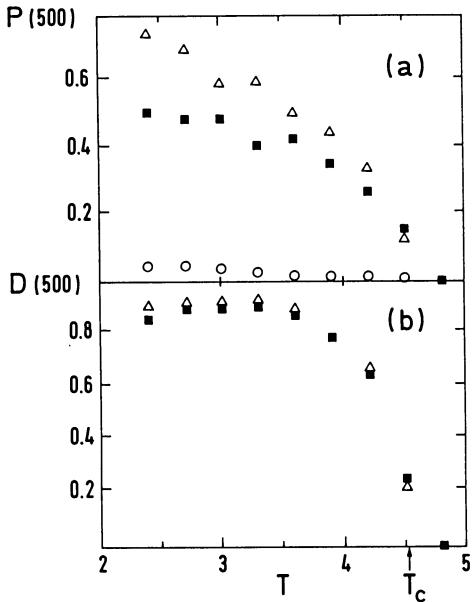


Fig. 3 Damage after $t = 500$ time steps in the 3d Ising model as a function of temperature for $L = 12$ from Ref. 5 for various initial damages: $\Delta(0) \rightarrow 0$ (\circ), $\Delta(0) = 1/2$ (\blacksquare) and $\Delta(0) = 1$ (\triangle); (a) fraction $P(t)$ of nonidentical configurations and (b) distance $D(t)$ between two configurations provided they are still different.

3. Heat Bath vs Glauber Dynamics in the Ising Model

Let us consider variables $\sigma_i = \pm 1$ and define as $h_i = \sum_j \sigma_j$ the local field acting on σ_i that comes from its nearest neighbors. Then one update in heat bath is given by setting the new value σ'_i of the spin to be $+1$ with probability p_i and -1 with probability $1-p_i$ where

$$p_i = \frac{e^{2\beta h_i}}{1+e^{2\beta h_i}}. \quad (3)$$

On the computer this is implemented by choosing a random number $z \in [0,1]$ and setting

$$\sigma'_i = \text{sign}(p_i - z) \quad (4)$$

In Glauber dynamics a spin is flipped with a probability

$$p(\text{flip}) = \frac{e^{-\beta \Delta E}}{1+e^{-\beta \Delta E}} \quad (5)$$

where ΔE is the difference between the energy of the would-be new configuration and of the old configuration. For the Ising model it is just $\Delta E = 2\sigma_i h_i$. On the computer one implements the Glauber dynamics via

$$\sigma'_i = -\sigma_i \text{ sign}(p(\text{flip}) - z) \quad (6)$$

Using Eqs. (3) and (5) one can express $p(\text{flip})$ in terms of p_i by

$$p(\text{flip}) = \begin{cases} 1-p_i & \text{if } \sigma_i = +1 \\ p_i & \text{if } \sigma_i = -1 \end{cases} \quad (7)$$

It can easily be seen from Eq. (7) that for the Glauber dynamics σ'_i is set $+1$ with probability p_i and -1 with probability $1-p_i$ just as in heat bath so that both dynamics have exactly the same probabilities.

Inserting Eq. (7) into Eq. (6) one finds how Glauber dynamics is implemented on the computer:

$$\begin{aligned} \sigma'_i &= \text{sign}(z - (1-p_i)) && \text{if } \sigma_i = +1 \\ \sigma'_i &= \text{sign}(p_i - z) && \text{if } \sigma_i = -1 \end{aligned} \quad (8)$$

This means that depending on the value of σ_i one uses the random number differently. So, if in one configuration the site σ_i was $+1$ and in the other configuration it was -1 , but the nearest neighbors in both configurations are the same, the damage at site i will probably survive in Glauber dynamics while it will certainly heal in heat bath. This gives rise to the different behavior in damage spreading between the two methods.

4. Relationship Between Damage and Thermodynamic Properties

Damage can be related to correlation functions as has been shown in Ref. 6. In the following we will briefly report on these relationships. We will consider Ising variables $\sigma_i = \pm 1$ which can also be expressed as usual by $\pi_i = 1/2 (1 - \sigma_i) = 0, 1$.

We want to discuss the damage between configurations $\{\sigma_A\}$ and $\{\sigma_B\}$ and in order to produce a small damage we will fix the spin at the origin of configuration $\{\sigma_B\}$ to be always

$$\sigma_0^B = -1 \quad (9)$$

This condition is different from the ones considered before because now the damage is fixed constituting thus a source of damage. In principle two types of damage can be imagined at site i: damage $++$ where $\sigma_i^A = +1$ and $\sigma_i^B = -1$ or damage $--$ where $\sigma_i^A = -1$ and $\sigma_i^B = +1$. The probabilities of finding a certain type of damage at site i can then be expressed as:

$$d_i^{++} = \langle\langle (1 - \pi_i^A) \pi_i^B \rangle\rangle \text{ and } d_i^{--} = \langle\langle \pi_i^A (1 - \pi_i^B) \rangle\rangle \quad (10)$$

where $\langle\langle \dots \rangle\rangle$ denotes a time average. Let us define the difference between the damages:

$$\Gamma_i = d_i^{++} - d_i^{--} = \langle\langle \pi_i^A \rangle\rangle - \langle\langle \pi_i^B \rangle\rangle \quad (11)$$

where we have used Eq. (10).

We want to express the damage through thermodynamic quantities defined on an unconstrained system $\{\pi_i\}$ with averages taken over many configurations, i.e., samples. So we translate condition (9) by a conditional probability and use ergodicity to go from time averages to thermal averages:

$$\langle\langle \pi_i^A \rangle\rangle = \langle \pi_i \rangle \text{ and } \langle\langle \pi_i^B \rangle\rangle = \frac{\langle \pi_i (1 - \pi_0) \rangle}{\langle 1 - \pi_0 \rangle} \quad (12)$$

where $\langle \dots \rangle$ denotes a thermal average. Inserting this into Eq. (11) one finds:

$$\Gamma_i = \frac{\langle \pi_i \pi_0 \rangle - \langle \pi_i \rangle \langle \pi_0 \rangle}{1 - \langle \pi_0 \rangle} = \frac{C_{oi}}{2(1-m)} \quad (13)$$

where

$$C_{oi} = \langle \sigma_i \sigma_0 \rangle - \langle \sigma_0 \rangle^2 \text{ and } m = \langle \sigma_0 \rangle \quad (14)$$

are just the correlation function and the magnetization. Relation (13) gives us an equality of a certain combination of the damages with thermodynamic functions. In its derivation we did not make any assumptions on the dynamics, the random numbers or the type of interaction and it is therefore of a very general validity.

If one would have chosen another form of fixed damage the result would have changed slightly. If one would for instance fix $\sigma_0^B = -1$ and $\sigma_0^A = +1$ then one would find

$$\Gamma_i = C_{oi}/(1-m^2).$$

If one considers variables with more degrees of freedom than Ising variables things can become more complicated but are still in principle feasible. As an example, let us look at the Ashkin-Teller model where on each site one has two binary variables $\sigma_i = \pm 1$ and $\tau_i = \pm 1$ which follow a Hamiltonian per site

$$\mathcal{E}_{ij} = -K(\sigma_i \sigma_j + \tau_i \tau_j) - 2L \sigma_i \sigma_j \tau_i \tau_j \quad (15)$$

In this case there are twelve possible damages per site which we label such that left means configuration A, right configuration B, top means σ and bottom means τ ; example $\uparrow\downarrow$ is $\sigma_i^A = +1, \tau_i^A = -1, \sigma_i^B = -1$, and $\tau_i^B = +1$. It can be shown [7] that if one fixes $\sigma_0^B = -1$ and $\tau_0^B = +1$ or $\sigma_0^B = +1$ and $\tau_0^B = -1$ then

$$\Gamma_i = \frac{\langle \sigma_i \tau_i \sigma_0 \tau_0 \rangle - \langle \sigma_i \tau_i \rangle \langle \sigma_0 \tau_0 \rangle}{2(1 - \langle \sigma_0 \tau_0 \rangle)} \quad (16)$$

with

$$\Gamma = \left[d_i^{++} + d_i^{+-} + d_i^{-+} + d_i^{--} \right] - \left[d_i^{++} + d_i^{+-} + d_i^{-+} + d_i^{--} \right] \quad (17)$$

and if one fixes $\sigma_0^B = -1$, then

$$\tilde{\Gamma}_i = \frac{\langle \sigma_i \sigma_0 \rangle - \langle \sigma_i \rangle \langle \sigma_0 \rangle}{2(1 - \langle \sigma_0 \rangle)} \quad (18)$$

with

$$\tilde{\Gamma}_i = \left[d_i^{+-} + d_i^{-+} + d_i^{++} + d_i^{--} \right] - \left[d_i^{+-} + d_i^{-+} + d_i^{++} + d_i^{--} \right] \quad (19)$$

This shows that both types of correlation functions that one has in the Ashkin–Teller model can be expressed as a combination of damages.

The damage for which we have presented numerical data in the preceding section was not the quantity Γ_i but it was the sum of all the damages:

$$\Delta = \sum_i \Delta_i \quad \text{with } \Delta_i = d_i^{+-} + d_i^{-+}. \quad (20)$$

In order to express these quantities in terms of thermodynamic quantities it is necessary to make some assumptions. Let us therefore restrict ourselves now to ferromagnetic interactions, heat bath dynamics and the use of the same random numbers for both configurations. We consider again only Ising variables and fix the damage as in Eq. (9), i.e., $\sigma_0^B = -1$. Since at $t = 0$ the only damage one has is the one of type $+-$ at the origin we have

$$p_i^A \geq p_i^B \quad \forall i \quad (21)$$

where p_i^A is the value defined in Eq. (3) for configuration A. Suppose one would try to create a damage of type $-+$ at site i. Then one would need, in order to produce $\sigma_i^A = -1$, a random number z which fulfills $z \geq p_i^A$ according to heat bath. Since one is using the same random number for configuration B this means using Eq. (21) that $z \geq p_i^B$ and therefore $\sigma_i^B = -1$. It is therefore impossible to create a damage of type $-+$ and therefore Eq. (21) will be valid also at the next time step. By induction one can conclude now that Eq. (21) will always be valid and that a damage of type $-+$ cannot be created under the conditions that we had imposed. Consequently, we have proved that $d^{-+} = 0$ and it follows for the damage

$$\Delta_i = \frac{C_{oi}}{2(1-m)} \text{ and } \Delta = \frac{\chi}{2(1-m)} \quad (22)$$

where $\chi = \sum_i C_{oi}$ is the susceptibility.

In Figs. 4 and 5 we see how the two exact relations of Eq. (22) are realized numerically for the 2d Ising model. In Fig. 4 we see the susceptibility obtained from Eq. (22) and obtained from the fluctuations of the magnetization as usual taken in a small system and the data agree very nicely. In Fig. 5 we see the correlation function obtained via Eq. (22) (circles) and in the usual way (triangles) using for both methods roughly the same computational effort. One sees that in the usual method, once the values are less than about 10^{-3} , the statistical noise takes over and the curve flattens. On the other hand, using Eq. (22) one gets to much smaller values without feeling substantial noise. The reason why the use of damage is superior numerically comes from the fact that this method looks just at the difference between two configurations subjected to the same thermal noise so that this noise is effectively canceled.

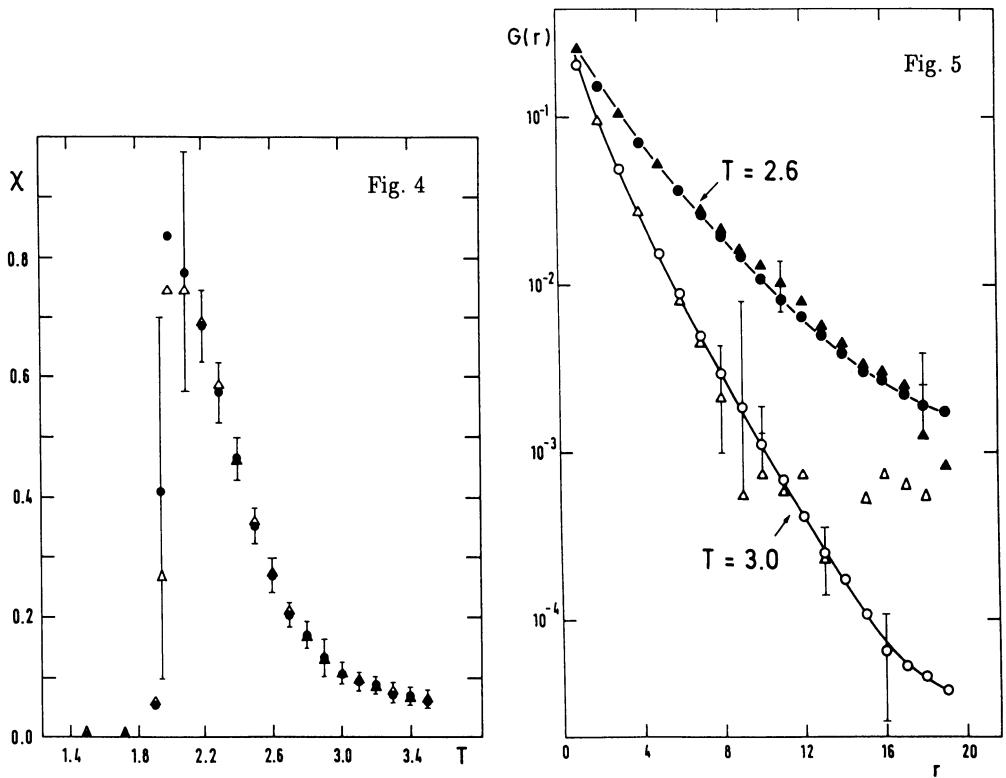


Fig. 4 Susceptibility $\chi(\bullet)$ and $2\Delta(1-m)(\Delta)$ as function of temperature from 30 systems of size 10×10 (taken from Ref. 6) for the 2d Ising model.

Fig. 5 Correlation function $G(r) = \sum_{|i|=r} C_{oi}$ (\blacktriangle for $T = 2.6$ and Δ for $T = 3.0$) and $2\bar{\Delta}(r)(1-m)(\bullet$ for $T = 2.6$ and \circ for $T = 3.0$) where $\bar{\Delta} = \sum_{|i|=r} \Delta_i$ as a function of r in a semi-log plot. The data came from 10 systems of size 40×40 and were taken from Ref. 6 for a 2d Ising model.

5. Damage Clusters

The damage that was fixed at the origin at time $t = 0$ acts as a source from which constantly damage spreads away. At a given instant one can look at all the sites that are damaged and one will find a cloud or cluster of sites around the origin. These clusters fluctuate in size and shape and are not necessarily connected. In Figs. 6a and 6b we see two of these clusters for the two dimensional Ising model at T_c which are just 38 time steps apart. Since heat bath was used these studies represent according to Eq. (22) the fluctuation of the magnetization due to the application of a local field at the origin. In Fig. 6c we see what happens if a damage

$\sigma_i^B = -1$ is fixed all along the boundary of the system. Then clusters of the type shown in

Figs. 6a and 6b overlap and one finds fluctuations that seem to be of all sizes. Using the equivalence of the Ising model to a lattice gas one can interpret these fluctuations as the fluctuations in local density which have actually been measured in a recent experiment of critical opalescence [8].

Using Eq. (22) it is actually possible to calculate the dependence between radius and number of sites for the damage clusters. One finds that at the critical point the clusters are fractal, that means that if L is the size of the smallest box into which the cluster fits and s is the number of sites in the cluster and has a relation

$$s \propto L^{d_f} \quad (23)$$

Using Eq. (22), i.e., $s = \frac{\int_{\text{vol}} \Delta_1 d^3 r \alpha \int_0^L G(r) r^{d-1} dr}{\sum_{|i|=r} C_{oi}}$ where $G(r) \propto \frac{1}{r^{d-1}}$ and using the fact that at T_c one has $G(r) \propto r^{2-d-\eta}$ one obtains $d_f = 2-\eta$. This can then be transformed into

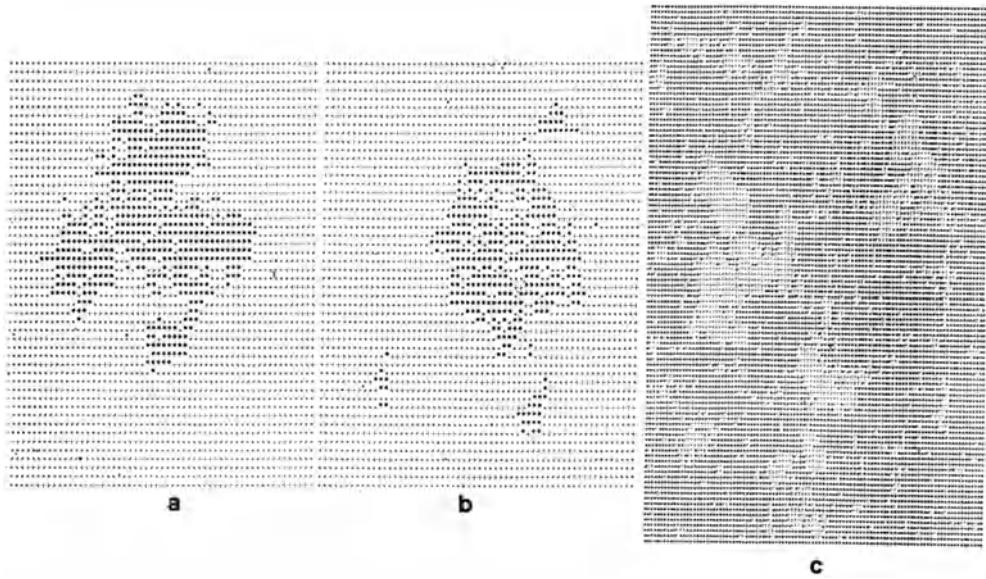


Fig. 6 Damage clusters of the 2d Ising model at T_c . In (a) and (b) the damage is fixed to be $\sigma_0^B = -1$ at the origin and the size of the system is $L = 60$. In (c) the same kind of damage is fixed all along the boundary and $L = 100$. Heat bath and the same random number sequence were applied on both configurations (taken from Ref. 6).

$d_f = d - 2\beta/\nu$ using the hyperscaling relation where β and ν are the critical exponents for the magnetization and for the correlation length. The above argument is only valid if one really measures directly the damage Δ_i , i.e., if one has the condition that the origin be permanently damaged. We note that in percolation theory a similar situation arises but there the mass of the clusters is defined differently so that the above condition has to be put in by hand which gives an additional contribution of β/ν in the exponent.

Since one does not expect several length scales in the problem one can also replace in Eq. (23) L by the radius of gyration R . In Fig. 7 we see results for the numerical determination of d_f for the 2d Ising model at T_c and an agreement with $d_f = 7/4$ is reasonably good.

It is also possible to make similar arguments for the case when the damage is fixed on the boundary as shown in Fig. 6c. There the density of damaged sites in the center of the box decreases with a powerlaw in the size of the box and consequently one finds again a fractal dimension as shown also in Fig. 7.

The fractal dimension of the clusters can also be measured by the touching method that has been widely used in cellular automata. One lets the damage spread until it touches the boundary of the system. The number of sites s that are damaged at this time scales like $s \propto L^{d_f}$ with the size L of the system, where d_f has the same value of the d_f in Eq. (23) if again one site is permanently damaged. In Fig. 8 we see the result for the three dimensional Ising model and

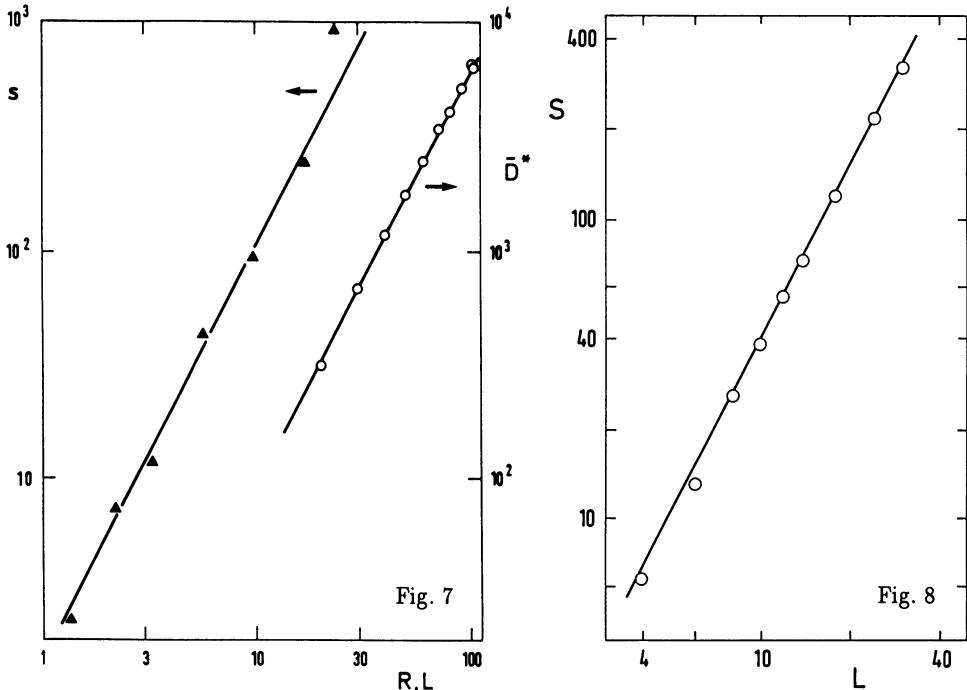


Fig. 7 Log-log plot of the number s of damaged sites against the radius R of the cluster (\blacktriangle) for the 2d Ising model at T_c with $\sigma_0^B = -1$ fixed. We also show the total number D^* of damaged sites if the damage is fixed on the boundary of a system of size L as a function of L (\circ) (taken from Ref. 6).

Fig. 8 Log-log plot of the number s of damaged sites that one has for the three dimensional Ising model at T_c when the damage first touches all the sides of the box of length L as a function of L . $\sigma_0^B = -1$ was fixed and heat bath dynamics was used (taken from Ref. 9).

finds numerically 1.95 in good agreement with $d_f = d - 2\beta/\nu$. Figure 8 was obtained by the condition that the damage touches all the sides of the box, if instead one demands that the damage touch only one side of the box the same result is found. Let us note that if one does not fix a damaged site and if one uses Glauber dynamics instead of heat bath the result changes and one seems to find compact clusters at T_c (Ref. 1).

6. Damage in Spin Glasses

Numerical work on spin glasses has been so challenging and frustrating in the past that any new method that gives some hope of improving the results should be tested. The application of the ideas of damage spreading has been done in Ref. 5 using heat bath and they have been very interesting. The results are shown in Fig. 9 which in their spirit are analogous to the data of the pure Ising model shown in Fig. 3. In three dimensions one believes that about $T_{SG} = 1.2$ there is a transition to a spin glass phase for the spin glass with randomly chosen coupling constants of +1 or -1. At $T_G = 4.5$, the critical temperature of the pure Ising model, one believes that the paramagnetic phase changes into a so-called Griffiths phase in which correlations decay a little slower than exponential; but this phase is very difficult to detect or to discern from the paramagnetic behavior.

From Fig. 9 one sees indeed two characteristic temperatures not too far from T_{SG} and T_G which could be interpreted as separating three phases. As opposed to the ferromagnetic phase of Fig. 3 both low temperature phases here are chaotic because the probability of two configurations to be still different after some time is not zero for times that are large but much shorter than the Poincaré time. In the low temperature phase the final distance does depend on the initial distance while in the intermediate phase the final distance is independent of the initial distance. One can indeed argue that this behavior reflects the phase space structure of the spin-glass phase and of the Griffiths phase. Unfortunately one finds qualitatively in two dimensions [10] the same picture as the one seen in Fig. 9 but it is believed that there is no spin-glass phase in two dimensions. This can be due to the fact that all these data are extremely far from an equilibrium-like state and must be treated with much care or it could be that the dynamical, chaotic phases of Fig. 9 cannot be related directly to the thermodynamic phases.

There exists, however, a way to single out T_G of the three-dimensional spin glass. During a certain interval of time a site will be damaged and healed again several times. One can look for each site i which is the number of times f_i that it is damaged again during a fixed, long time interval. Next one can look at the distribution $P(f)$, i.e., the number of sites that have a certain frequency f and analyze the moments of this distribution:

$$M_q = \sum_f f^q P(f) \quad q = 0, \pm 1, \dots \quad (24)$$

and normalize them:

$$m_q = (M_q / M_0)^{1/q} \quad (25)$$

These moments will scale with the size of the system L like

$$m_q \propto L^{x_q} \quad (26)$$

Usually the distribution is self-averaging and scales such that all x_q (with $q \neq 0$) are the same. But in specific cases, when the distribution has a particularly long tail it can exhibit what has been called multifractality (see contribution of T. Vicsek in this volume) which means that the x_q vary with q and one has therefore an infinity of different scaling exponents.

In Ref. 11 the moments of Eq. (25) have been calculated for various models. For the three-dimensional spin glass it has been found that the distribution is multifractal at T_G as

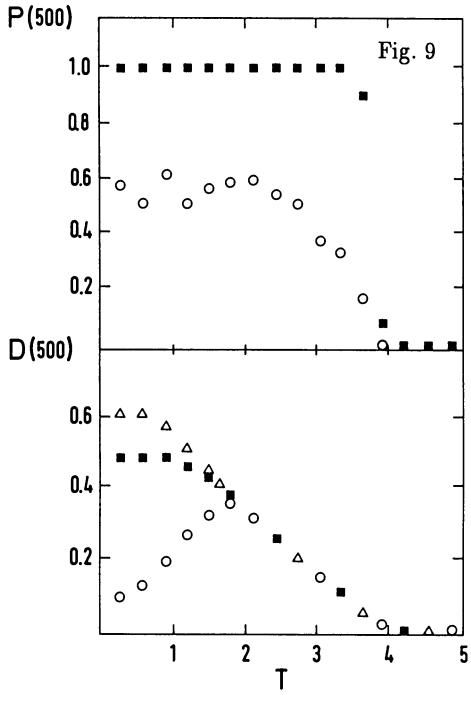


Fig. 9

Damage in the three dimensional \pm spin glass after 500 time steps as a function of temperature for $L = 12$ taken from Ref. 5 for various initial damages: $\Delta(0) \rightarrow 0$ (\circ), $\Delta(0) = 1/2$ (\blacksquare) and $\Delta(0) = 1$ (\triangle); (a) fraction $P(t)$ of nonidentical configurations and (b) distance $D(t)$ between two configurations provided they are still different.

Fig. 10

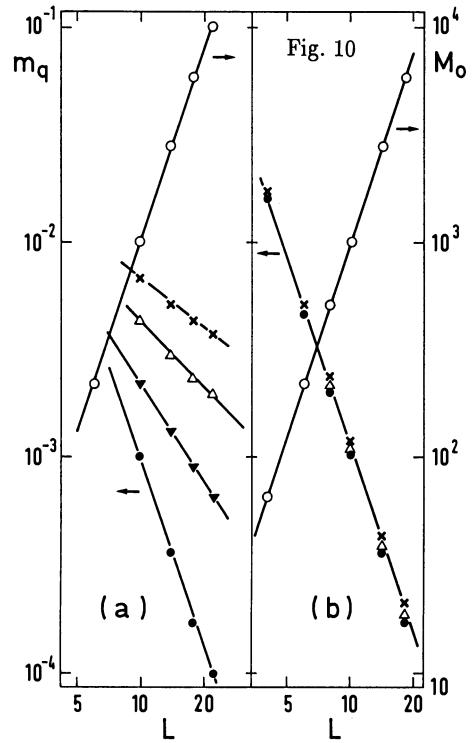
Log-log plot of the moments of the distribution $P(f)$ of the redamaging of sites as a function of the system size L for (a) the 3d \pm spin glass at T_G (b) the standard 3d Ising model at T_c . The moments are: $M_0(\circ)$, $m_1(\bullet)$, $m_2(\blacktriangle)$, $m_3(\triangle)$ and $m_4(x)$. Taken from Ref. 11.

can be seen from Fig. 10a. This is in marked contrast to what has been found at T_{SG} for the 3d spin glass, at T_G or T_{SG} for the 2d spin glass or at the critical temperature T_c of the 3d Ising model because in all these cases the lines for the different moments in the log-log plot are parallel to each other which means that x_q is the same for all q . As an example, we show in Fig. 10b the data for the 3d Ising model.

We can conclude that the temperature $T_G = 4.5$ of the 3d \pm spin glass is particular because there the distribution of the redamaging frequency is multifractal while in the usual case, particularly at T_G in the 2d \pm spin glass one has a simple one-exponent scaling behavior. This feature might be useful to single out models that show spin glass behavior.

7. Conclusion

We have seen that the concept of damage spreading that has been widely used for cellular automata can be applied to Monte Carlo. One finds dynamical phase transitions which only in



specific cases like heat bath can be related to well-known thermal critical points. Much has to be understood yet about these dynamical transitions in the general case.

Up to now damage spreading techniques seem to be useful for two purposes: More precise calculation of correlation functions and eventually dynamical information about spin glasses.

I thank Lucilla de Arcangelis, Antonio Coniglio and Ananias Mariz for many interesting discussions and D. P. Landau and the Center for Simulational Physics at the University of Georgia for the friendly hospitality.

References

1. H. E. Stanley, D. Stauffer, J. Kertész and H. J. Herrmann, Phys. Rev. Lett. **59**, 2326 (1987).
2. S. A. Kauffman, J. Theor. Biol. **22**, 437 (1969).
3. U. Costa, J. Phys. A**20**, L583 (1987).
4. G. Le Caér, in preparation (1989).
5. B. Derrida and G. Weisbuch, Europhys. Lett. **4**, 657 (1987).
6. A. Coniglio, L. de Arcangelis, H. J. Herrmann and N. Jan, Europhys. Lett. **8**, 315 (1989).
7. A. Mariz, L. de Arcangelis and H. J. Herrmann, in preparation (1989).
8. P. Guenoun, F. Perrot and D. Beysens, preprint (1989).
9. H. J. Herrmann et al., in preparation (1989).
10. A. U. Neumann and B. Derrida, J. Physique **49**, 1647 (1988).
11. L. de Arcangelis, A. Coniglio and H. J. Herrmann, in preparation (1989).

Molecular Dynamics Simulations of Proteins and Protein-Protein Complexes

J.E. Wampler, D.E. Stewart, and S.L. Gallion

School of Chemical Sciences and Advanced Computational Methods Center,
University of Georgia, Athens, GA 30602, USA

Advances in supercomputer technology have allowed problems involving the structure and energetics of biologically relevant macromolecules (proteins and nucleic acids) to be studied using computational methods such as molecular mechanics and molecular dynamics. In this paper, a major type of biological macromolecules, proteins, is briefly described and the empirical force fields which are used to study them at the atomic level are discussed. Results are summarized which show that supercomputer versions of molecular mechanics/dynamics programs can allow effective modeling of large proteins and protein complexes, useful prediction of protein structures based on existing structure data bases, and considerable insight into the nature of the interaction of proteins on the picosecond time scale.

1. Introduction

Biochemistry has evolved from what was basically a science of separation into a complex field dealing with the mechanism of the living process. The key component of these systems are the catalysts, which control, direct and moderate the biochemical reactions, the enzymes. These catalytically active proteins have characteristics which demand a role for dynamic flexibility. They often bind their chemical substrates in obligatory sequence or respond to binding of effector molecules with reversible changes in catalytic activity. Enzymes and enzyme-like proteins act as regulatable carriers for small molecules across cell membranes. Many of the responses to binding of substrates or effectors or to changes of environment result in gross structural changes that can be easily seen spectroscopically or chromatographically. However, in spite of the obvious role of dynamic flexibility in their function, most experimental biochemical tools and the most sophisticated instrumental methods of our day tend to give a static picture of enzymes. Certainly some techniques exhibit temporal resolution, but on a time scale so long that information on atomic motions is averaged out. Thus, for a very long time the culmination of an enzymological study has been the determination of a static X-ray crystal structure.

Now such structures are just the start. Computational chemistry brings the biochemist a tool which for the first time allows detailed examination of the dynamics of protein molecules. If these simulations are correct, they open many new avenues of study and have direct bearing on our understanding of structure/function relationships. Early indications from such studies raise important questions about the role of adaptive changes in structures when enzymes interact with their substrates, about dynamic processes that may actually be part of biocatalytic events and about roles of alternative structures. They prompt the molecular geneticist to alter proteins both to probe structure/function relationships and to engineer better or more stable catalytic molecules based on an "educated guess" from simulation rather than the more limited static view.

Dynamic simulations of intramolecular motions of proteins and protein complexes require a number of computational tools. The purpose of this paper is five fold.

- a) to briefly review and describe the complex, heterogeneous chemical composition of proteins.

- b) to describe and contrast the approaches that have been used to develop an empirical force field in order to manipulate protein structures computationally.
- c) to illustrate the success of these force fields in predicting static models of proteins.
- d) to discuss the role of supercomputer technology in making use of these force fields in a practical way for studies of protein folding and dynamics.
- e) to show the use of one of these empirical force fields (AMBER) in analyzing the movements of residues within the interface between interacting proteins.

2. Composition and Physical Properties of Proteins

Proteins are a class of biopolymers which are characterized most by the broad heterogeneity of their physical properties. Some proteins exhibit anionic character in physiological solutions; some exhibit cationic character. Some are soluble in the very hydrophobic lipid membranes of cells; some have a very hydrophilic character. Some exhibit strong optical properties, absorbing UV light or exhibiting dichroism; some exhibit little of such properties. Some exhibit rigid, regular structural organization; some are nearly amorphous. Yet despite this heterogeneity, the primary structure of all proteins is, very simply, a polymer chain consisting of sequential amide linkages between L-alpha-amino acids. The polymer chain has the redundancy shown in Figure 1A, carbonyl carbon linked to amide nitrogen linked to aliphatic carbon, repeated over and over from one end to the other. This main chain, the peptide backbone, has no branches and it is restricted in folding due to the planarity of the amide bond (Figure 1B) caused by the delocalization of the pi carbonyl electrons and the non-bonding pair on the nitrogen atom.

The heterogeneity of natural proteins described above is not due to the flexibility of the chain to fold in many different ways. In fact, examination of a large number of known X-ray structures reveals a fairly limited number of secondary folds of the main chain, the alpha helix and beta sheet structures so insightfully predicted from studies of solid models by Pauling [1] and several types of bends and turns [2]. Most water soluble enzymes consist of partially ordered arrangements of alpha helix and/or beta sheet sequences linked by various types of

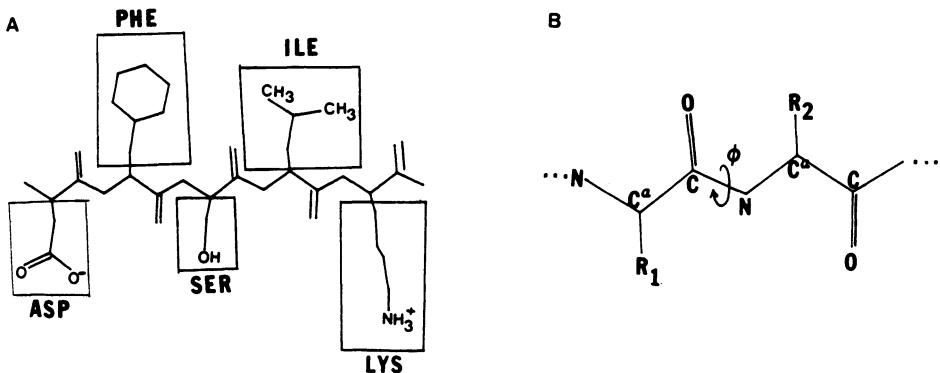


Figure 1. A - The polymeric character of the protein is revealed by the regular repeating backbone chain. The short side chains extend from each amino acid alpha carbon atom. It is the variability of these side chains that give proteins their wide range of chemical characteristics. B - Detail of the Amide Bond. The polymer backbone has a repeating amide (or peptide) bond which because of delocalization of the non-bonding electrons on the nitrogen and the pi electrons of the carbonyl has severely restricted rotation about the dihedral angle ϕ .

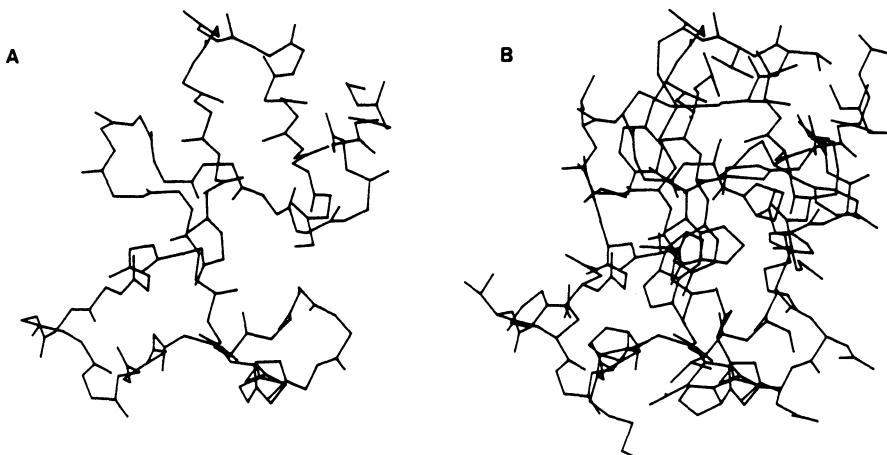
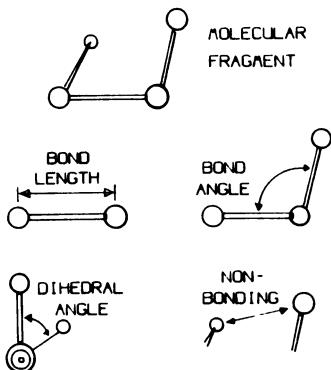


Figure 2. A - The backbone atoms of the small protein, rubredoxin, show the regular arrangement of folding. B - When the structure is shown with all side chain atoms, the globular nature of the overall structures is obvious.

turns and bends [2]. The real heterogeneity in structural and chemical properties for proteins lies in the composition and arrangement of the amino acid side chains. The building blocks for protein structure consist of some twenty different L-a-amino acids, the R-groups of Figure 1. It is the properties of these groups which are so variable and disparate. Some, such as the amino containing lysine and guanidinium containing arginine, have a formal positive charge under physiological conditions of pH. Glutamic acid and aspartic acid, on-the-other-hand, have a full negative charge under the same conditions due to a terminal carboxyl group of their side chains. Other amino acids such as cysteine with its terminal thiol, serine and threonine with their alcohol substituents and asparagine and glutamine with their amides have polarizable groups with reasonably large partial charges distributed on the sidechain atoms. Several amino acids have aromatic sidechain groups (tyrosine, phenylalanine, tryptophan and histidine). In addition, histidine has a very weak acid group that can readily exchange protons with donor and acceptor groups at physiological pH's. A number of side chains consist of uncharged aliphatic carbons (alanine, valine, leucine, isoleucine, proline and methionine) and tend to form a hydrophobic core (along with the aromatics) in the folded protein structures. Proline and cysteine have specific roles in determining the folding of the main chain. Proline contains a five membered aliphatic ring which includes the alpha carbon atom and the alpha amino group. This prevents the normal, free rotation of the alpha-carbon to alpha-amine bond and further limits folding. Two cysteines can form a crosslinking disulfide bond, a covalent tie across a protein molecule. Thus on one scale a protein is a fairly simply folded structure of main chain atoms (Figure 2A) and on another it is a complex collection of widely different chemical groups (Figure 2B).

3. The Molecular Mechanics Approach

The molecular mechanics approach to understanding the forces which act to fold proteins and to drive their dynamic behaviour depends on the Born-Oppenheimer approximation. Separation of nuclear motions from electronic motions and, in practical terms, treatment of the net electron density as a point charge located at the atomic center, make it possible to view the internal energy and the interatomic forces acting on each atom in a molecule as a simple linear summation of individual, discrete interactions. Referring to Figure 3, if we consider interatomic interactions of protein atoms, there are several obvious classifications



$$\Delta E = \Delta E_{BL} + \Delta E_{BA} + \Delta E_{DA} + \Delta E_{NB} \dots$$

Figure 3. The intramolecular forces considered in a molecular dynamics simulation.

of the forces involved. The covalent linkages have bond lengths which are heavily constrained as if the atoms were linked by a strong spring. Even minor shortening or lengthening of these bonds requires a very large energy input. Bond angles are slightly less constrained and dihedral angles about single bonds are typically free to change (except on the interior where steric constraints are high). Weak non-bonding forces between atoms have energies determined by van der Waals interactions. Coulombic interactions between non-bonding charged groups complete the basic picture. Thus, the force acting on any atom can be viewed as the vector sum of forces due to each of these terms for all appropriate interacting atoms. The total energy of the molecule is then the sum of all interaction energies for all appropriate combinations of atoms.

Several empirical force fields of this general type have been developed and successfully used to study the forces involved in folding proteins and in driving their dynamic behaviour. Levitt and Lifson [3] described the first empirical force field specifically designed to accomodate the multitude of atom types found in proteins. Allinger [4] developed a more rigorous function with a simpler set of atom types to model hydrocarbon compounds. Variations on these formulations have lead to several major, commercially available software packages which include similar empirical force field calculations with full sets of parameters for protein modeling [for reviews see 5-7]. Only two of these will be discussed in this paper, the AMBER force field developed for protein and nucleic acid modeling by Kollman and coworkers at the University of California at San Francisco [8] and the new extended version of the Allinger force field, MM2-87, which was recently used to minimize the protein crambin [9].

The AMBER potential function [8] consists of seven terms summarized in Table 1. The bond length and bond angle terms are simple harmonic potentials. The dihedral angle term includes a periodicity factor, n , to control the number of minima per 360° of full rotation. Non-bonding interactions are accounted for using a simple 6-12 van der Waals potential, a coulombic electrostatic potential and a separate 10-12 hydrogen bonding potential. The seventh term allows for application of any number of harmonic constraints. Several of these terms have optional forms. For example, a constant dielectric constant may be used in the electrostatic potential calculation or a distance dependent dielectric may be selected. In order to reduce computation times, a user-selected spherical cutoff distance may be applied in calculation of non-bonding interactions.

AMBER includes several optional selections to further speed calculation of the potential function. Hydrogen atoms may be treated explicitly or combined with the heavy atom to which they are attached (united atom representation). In AMBER, the aliphatic hydrogens are given a united atom treatment and polar hydrogens are explicitly represented. The list of atoms involved in non-bonded interactions is

Table 1. Potential Functions Used in AMBER 3.0 and MM2

Energy Term	AMBER	MM2
V _{bond}	$K_B(r - r_0)^2$	$K_B(r - r_0)^2(1 - 2(r - r_0))$
V _{angle}	$K_\Theta(\Theta - \Theta_0)^2$	$K_\Theta(\Theta - \Theta_0)^2(1 + K'_\Theta(\Theta - \Theta_0)^4)$
V _{dihedral}	$\left(\frac{V_n}{2}\right)(1 + \cos(n\Phi - \gamma))$	$\left(\frac{V_1}{2}\right)(1 + \cos(\Phi)) + \left(\frac{V_2}{2}\right)(1 - \cos(2\Phi)) + \left(\frac{V_3}{2}\right)(1 + \cos(3\Phi))$
V _{stretch-bend}		$K_{sb}(\Theta - \Theta_0)(r - r_0)(r' - r'_0)$
V _{van der Waals}	$\left(\frac{A}{r^{12}}\right) - \left(\frac{B}{r^6}\right)$	$K_{vdw} \left[k'e^{-12.5\left(\frac{r_u}{r}\right)} - k''\left(\frac{r_u}{r}\right)^6 \right]$ for $\left(\frac{r_u}{r}\right) < 3.311$ $K_{vdw} k'''\left(\frac{r_u}{r}\right)^2$ for $\left(\frac{r_u}{r}\right) \geq 3.311$
V _{electrostatic}	$\frac{K_e Q_i Q_j}{\epsilon r}$	$K_\mu [\mu \cdot \mu' (\cos(\chi) - 3\cos(\alpha)\cos(\alpha'))] / \epsilon r^3$
V _{hydrogen-bonding}	$\left(\frac{C}{r^{12}}\right) - \left(\frac{D}{r^{10}}\right)$	
V _{harmonic-constraints}	$K_\beta(\beta - \beta_0)^2$	

Definition of Symbols: r , distance; r_0 and Θ_0 , standard values for bond length and bond angle; K_B , K_Θ , bond and angle constants, $K'_\Theta = 7 \times 10^{-8}$; Φ , torsion angle; $\left(\frac{V_n}{2}\right)$, n -th fold barrier height; n , periodicity; γ , phase; K_{sb} , stretch-bend coupling constant; r' , distance between atoms j and k in the angle Θ_{ijk} ; A and B , Lennard-Jones constants; $K_{vdw} = (\epsilon_i \epsilon_j)^{1/2}$ where ϵ_i is the well-depth of two atoms of type i ; $k' = 2.9 \times 10^5$; $k'' = 2.25$; $r_v = r_i + r_j$ where r_i is the optimal distance of van der Waals interaction between two atoms of type i ; $k''' = 336.176$; $K_e = 330.32$; Q_i and Q_j are electrostatic point charges; ϵ , dielectric constant; $K_\mu = 14.39418$; μ and μ' are bond dipoles at angle χ ; α , angle between μ and the vector connecting the midpoints of μ and μ' ; α' , the angle between this vector and μ' ; C and D , constants for hydrogen-bond geometry correction term; K_β , the constant associated with harmonic constraint of general coordinate β to the value β_0 . Units of constants are such that energies are expressed in terms of kcal/mole.

not updated with every iteration of the program. This saves a large amount of time since generation of the non-bonded list is the most computationally expensive part of a molecular dynamics program. Additional harmonic constraints allow the user to specify full minimization (or dynamics) for only a portion of a molecule, holding the rest rigid but including these atoms as partners in all appropriate calculations. AMBER allows water to be included in the calculation with its contribution to the energetics evaluated using the non-bonding terms of a TIP3P potential [10].

The MM2 force field has fewer terms than AMBER, but each of the six terms it uses are more complex (Table 1). MM2 also includes a number of small correction terms (not shown) such as a term for the effect of delocalization at the anomeric carbon atom of sugars. MM2 is an all-atom force field; i.e. all hydrogen atoms must be included explicitly. It makes no provisions for arbitrary constraints but does include some restrictions on overall motion. MM2 has long been used successfully in the study of a variety of small organic molecules [11,12]. However, the force field has only recently been updated, MM2-87, to model amides [13] and proteins [9]. In

particular, the radii used in the non-bonded interactions for hydrogen atoms involved in hydrogen bonds were reduced and the energy of interaction was increased by 1 to 3 kcal/mole to correctly model hydrogen bonding in simple amides. In addition non-bonded electron pairs on nitrogens and oxygens were treated as pseudoatoms. These parameters were further refined in a minimization study using two low temperature X-ray structures of small cyclic peptides. Since there was a high degree of confidence in these experimental data, the parameters were adjusted so that minimized structures would not drift substantially away from the initial, experimental coordinates.

4. Minimization as a Test of Potential Functions

Molecular mechanics potential functions such as those used in AMBER and MM2 can be used to predict minimum energy configurations of molecules. Non-linear least squares minimization of the potential function as a function of the atomic positions of all atoms leads to this structure and an energy minimum. It should be pointed out, however, that the complexity of the multidimensional space involved ($3N$ dimensions where N is the number of atoms) dictates that minimization by available procedures leads only to a local minimum with a configuration near to that of the starting structure. Thus, minimization can lead to new, but similar structures when changes are made in previously known structures. It can not fold a protein *de novo*. In AMBER the user has a choice between two minimization approaches, steepest descents or conjugate gradient. MM2 uses a block-diagonal Newton-Raphson approach.

4.1 Success of MM2 in minimizations

Critical evaluations of the capabilities of the MM2 Force Field clearly demonstrate its overall success with small organic molecules. One such study [12] came to the following conclusion:

"We now support Allinger's conclusion that the molecular mechanics method, in principle, must be considered to be competitive with experimental determinations of the structures and enthalpies of molecules."

In this study of predicted and experimental properties for 55 acyclic, cyclic and polycyclic alkanes, Engler *et al.* [12] found agreement between predicted and experimental data as follows:

bond lengths	0.01 Angstroms
bond angles	1-2°
heats of formation	1 kcal

Recently, as discussed above, MM2 was extended to allow protein modeling. The optimized MM2(87) force field has been used to minimize crambin, a small, hydrophobic protein consisting of 46 residues and containing a total of 671 atoms, including 27 lone pairs. Crambin is among the highest resolution protein structures currently available [14]. The differences between the minimized and X-ray structures were compared with the results of other minimization studies using AMBER [9,15]. These results are listed in Table 2. Only those AMBER calculations which employed a dielectric constant of either 4.0 or 4xR and did not use a cutoff in the determination of the nonbonded forces were selected for comparison. The MM2 minimized structure showed the smallest root-mean-square (rms) positional deviation from the experimental structure giving a structure somewhat larger (1.1%). In comparison, all minimized structures from AMBER were compacted to some extent (between 1 and 7%). The rms deviation in the backbone torsion angles (PHI between the alpha carbon and the amide nitrogen, and PSI between the alpha carbon and the carboxyl carbon) were somewhat large in all the cases (between 11 and 19 degrees). Most of this may be attributed to the lack of crystal contacts in the model,

TABLE 2. Comparison of Results for Crambin.⁺

Force Field	Dielectric	RMS (Å)	$\frac{1}{2}R_g^{**}$	PHI	PSI
MM2	4.0	0.31	1.1	15.1	18.1
AMBER(J)*	4.0R	0.35	-1.1	11.5	12.1
AMBER(A)*	4.0	0.50	-7.4	14.1	13.1
AMBER(J)*	4.0	0.51	-2.1	14.6	15.8
AMBER(A)*	4.0R	0.82	-7.4	18.6	14.2

⁺ Taken from [9].

* J=Jorgensen radii or A=AMBER radii [9].

** Data from [14]

resulting in large discrepancies for torsion angles in the turn regions between secondary structure elements. The inclusion of neighboring residues from symmetry related molecules in these regions are expected to improve the results [9]. Because MM2 performs well for a wide variety of small organic molecules which may act as substrates for proteins, the results obtained are certainly encouraging and provide a basis for continued studies involving protein-substrate complexes.

4.2 Tests of AMBER in minimization

As shown in Table 2, above, AMBER 3.0 also maintains crambin near its high resolution crystal structure. In this case as well, the minimization changes can be attributed to differences due to crystal packing. The advantage with AMBER lies in a shorter calculation time, its proven ability to model many different proteins and its molecular dynamics capabilities. A second such test is illustrated by Table 3. In this case the results are compared using a united atom force field versus the all atom force field. While an all atom representation should be more accurate, inclusion of these extra atoms greatly increases the time required for the calculations. The difference between the two approaches is illustrated by the data of Table 3. These two proteins show about 67% amino acid sequence homology. The rms deviations shown were calculated between the mainchain atoms only. In this case, with lower resolution X-ray structures than the crambin example above and with a cutoff on non-bonded interactions (8.0 Å), minimization gave larger changes. The procedure typically involved 250 cycles of constrained minimization followed by several thousand cycles of unconstrained minimization until the convergence criterion was satisfied (rms gradient < 0.1 kcal/Å). The all atom force field required approximately 3 times more computational time.

TABLE 3. RMS Deviation Between X-Ray Crystal Structures and the AMBER 3.0 Minimized Structures of Rubredoxins (In Å).

STRUCTURE	UNITED ATOM	ALL-ATOM
D. vulgaris	0.8502	0.6995
C. pasteurianum	0.9350	0.5250

Another test of the accuracy of the AMBER force field comes from a recent study in our laboratory of its use in predicting structures of homologous proteins [16,17]. Pairs of related protein structures were obtained from the Brookhaven Protein Data Bank [18] and "mutated" each into the other. The mutant protein was then minimized and compared with its X-ray crystal structure either as originally obtained from Brookhaven or after AMBER minimization.

Protein Data Bank coordinate files contain a line for each atom of the protein, each line containing the atom name and number, the residue name and number, and the cartesian coordinates of the atom. The approach used was as follows: one structure file of each pair was modified so that the sequences became the same, i.e. the residue names were changed appropriately and any non-common atoms were deleted from the file. This results in a file containing the same residue sequence as the target protein but with atoms missing. A routine within the AMBER suite of programs [19] supplies missing atomic coordinates prior to minimization. Hence, these "mutated" structures were processed to replace the missing atoms and then minimized with varying levels of constraints. With proteins which are over 60% homologous in amino acid sequence the rms deviations for the mainchain atoms compared to the original crystal structure following this procedure was invariably less than 1 Å. This was true even when the original pair of X-ray structures differed by a rms deviation of as much as 2.7 Å [16,17].

To some extent the comparison of a minimized "mutant" structure to its x-ray crystal structure is an inappropriate comparison. The minimized structure approximates the dilute solution structure where the protein interacts only with solvent. As pointed out above, the crystal structure is determined with the protein under the influence of crystal packing forces [20]. These forces may distort the structure and minimization, even of high resolution crystal structures, leads to changes as described above. Thus, when the minimized "mutant" protein is compared to the minimized X-ray structure, such environmental factors should be removed. Table 4 shows such a comparison for the structures used for the data of Table 3. Obviously, the all atom force field gives a much better prediction when compared to the similarly minimized X-ray structure. The results suggest that although up to 3 times as much processor time may be required, the results are significant enough to justify the additional resources.

5. The Role of the Supercomputer in Protein Mechanics and Dynamics

The computational time required for evaluation of the potential energy function is the limiting factor in studies involving macromolecules. For systems containing more than about 500 atoms, nearly 90% of the total cpu time consumed involves calculation of the nonbonded (NB) energy. Approximation schemes such as the neighbor list technique [21,22] vastly decrease the amount of time spent in evaluating the potential energy function and its derivatives. Typically, the use of a list reduces the time for evaluation of the NB terms by a factor of 4 or 5 [23]. However, even with these approximations, the computational times required to yield significant results becomes prohibitive when the calculations are performed on conventional computers such as the VAX 11/780. Energy minimization of a 3000 atom system may require over 100 VAX cpu hours to achieve convergence. A 100 ps trajectory of a protein-solvent system consisting of 4500 atoms will require several thousand hours of VAX cpu time [24]. Productive research in the field of macromolecular modeling is therefore quite dependent on the ability to perform computations as efficiently as possible. For this reason, many groups involved with modeling of macromolecular systems have taken advantage of the increasing availability during the early and mid-1980's of vector supercomputers. This momentum has extended to the vector/parallel architectures such as exhibited by the Cray-2, ETA-10, Convex C-2, and Alliant machines.

TABLE 4. RMS Deviations of Predicted Vs. Minimized and Crystal Structures (In Å)

PREDICTED STRUCTURE	United-Atom		All-Atom	
	MIN.	XTAL	MIN.	XTAL
D. <u>vulgaris</u>	0.9209	1.1550	0.5390	0.9357
C. <u>pasteurianum</u>	0.8507	1.2378	0.2371	0.5425

The efficient use of any new architecture generally involves some degree of reprogramming effort. The amount of effort expended is largely determined by the performance goals set for any particular research project. There are very few sophisticated programs in this field that have been written explicitly for these new machine architectures. The major macromolecular mechanics and dynamics programs have evolved from previous programs written on scalar architectures. Thus, maximum exploitation of vector and parallel processing environments typically requires major changes in data structures and even algorithms. Such alterations are time consuming and are generally attempted only to solve specific problems where computational bottlenecks are well-localized. It is important to note that this strategy of limited vectorization often requires extra code to restructure data for the vectorized code segment as well as mapping the results back to a scalar data structure after vector computations are completed.

A number of papers have been published concerning the use of supercomputers for calculations of the molecular mechanics type [9, 23-29]. It is interesting to note in passing that much of the earlier work was done by European research groups, reflecting the earlier acceptance and wider distribution of supercomputers there.

Vogelsang, et. al., [25] reported on the vectorization of NB list and force calculations on the CYBER 205 for simulations of monoatomic liquids. The paper offered excellent insight to the basic concepts involved in vectorization of these types of calculations. Important aspects such as the use of bit vectors and increasing the effective vector length in condensed DO-loop structures were discussed. While this research addressed the specific implementation of a vectorized Lennard-Jones potential, the NB energy functions used in most major macromolecular mechanics programs are more complex, involving electrostatic terms, excluded atoms, switching functions, and explicit terms for hydrogen bonding. It was therefore not clear how well programs of this type would perform on vector processors.

van Gunsteren et. al., [23] reported on the optimization of a widely-distributed molecular mechanics program, GROMOS (GROningen MOlecular Simulation) for a CRAY-1 supercomputer. The paper was primarily concerned with techniques of neighbor list formation and compared atom-based, residue-based, and grid schemes for systems of various sizes. The results indicated that, for systems containing less than 6000 atoms, the grid techniques were somewhat slower than the other methods if the neighbor list is only updated every 10 steps in the simulation. A moderate amount of semantic vectorization led to an increase in speed of about 12 times that of a VAX 11-780 whereas a more thorough use of assembly language led to another factor of seven in time savings.

A detailed outline of the steps involved in the conversion and optimization of the AMBER 2.0 energy calculation for a CYBER 205 supercomputer were given by Gallion et al. [29]. Benchmarks were performed using the enzyme carboxyl protease from Rhizopus Chinensis with a bound inhibitor, pepstatin [30]. The complex contains a total of 2488 atoms, 2893 bonds, 4161 bond angles, and 6447 torsion angles.

It was noted that the first efforts to convert AMBER from the VAX to the CYBER 205 led to different rates of enhancement among the various subroutines. Of particular interest was the fact that the NB energy routine executed a factor of 37 faster than the VAX version whereas the NB list routine exhibited a performance increase of only 17. The difference was attributed to the ratio of the number of logical operations to the number of floating point operations. Code that contains more floating point operations than logical operations will exhibit the greater rate enhancement. Inner DO-loop condensations led to a performance increase of 7 for the NB list formation but only a 20% increase in speed for the NB energy evaluation. The poor performance of the latter was due to the fact that vector lengths reflected the number of NB pairs for each atom which, for a typical cutoff of 8 Å, was about 50 for the test system. In comparison, the average vector length in the NB list routine is N/2.

A block structure was implemented in the NB energy section to increase the average vector length so that, for each pass through the vectorized loop, calculations would involve a set of interaction spheres rather than a single, isolated sphere. The size of the block structure was dictated by paging and scheduling considerations and a maximum vector length of 3500 was selected as providing reasonable throughput. The block structure led to a NB energy evaluation

executing 133 times faster the same calculation on a VAX. With these changes, the computational bottleneck for a single function evaluation migrated to the routine responsible for determination of the dihedral and 1-4 energies. The final program reduced the time to perform a single function evaluation, complete with list formation, by a factor of 106 over a VAX 11-780 equipped with a floating point accelerator. Individual routines had performance increases ranging from 117 (NB list) to 211 (dihedral). It was noted that additional effort would probably result in another factor of two in increased performance [29].

It is important to note the performance of the optimized code as a function of the size of the system being modeled. Figure 4 shows the time required for a single function evaluation as a function of the cutoff radius (R_c) used to generate the list. The time to calculate the list was not included. With an 8 Å cutoff the vectorized code on the CYBER is about 80 times faster than the scalar code on the VAX 11-780. A 14 Å cutoff yielded a factor of near 120. The increase in the relative speed of the CYBER 205 when compared to the VAX for the larger calculation ($R_c = 14 \text{ \AA}$) is due to an increase in the proportion of time spent in the vector processor relative to the scalar processor on the CYBER and a corresponding reduction in the percentage of time spent in vector startup operations.

The results from this study [29] are typical of what is to be expected of a reasonably straightforward syntactical vectorization effort. For instance, the vectorization of MM2 for the CYBER 205 described briefly above led to an energy evaluation roughly 200 times as fast as the VAX version [9]. The time for 5 iterations of block diagonal Newton-Raphson minimization was decreased by a factor of 7 over the scalar CYBER 205 version upon vectorization of the code concerned with the determination of the first and second derivatives with respect to the nonbonded potential.

As mentioned previously, the degree of optimization is highly dependent on the amount of effort expended. As an example the NB list and energy routines in AMBER 3.0 have been vectorized for an ETA-10Q [31]. By separating code concerned with specific loop-independent functions (such as a constant or distance-dependent dielectric, atcm- or residue-based cutoff, presence of periodic boundaries, etc.)

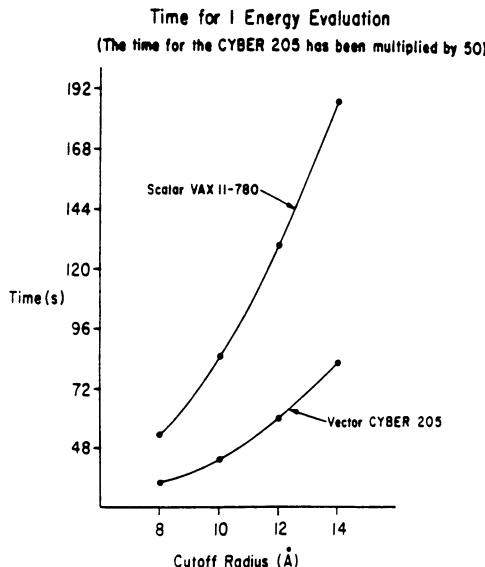


Figure 4. Total time for one energy evaluation as a function of the cutoff radius used to generate the nonbonded pair list. The number of nonbonded pairs included in the list and the cutoffs used for these calculations are: (122,379 pairs, $R_c = 8 \text{ \AA}$), (229,127 pairs, $R_c = 10 \text{ \AA}$), (375,742 pairs, $R_c = 12 \text{ \AA}$), (557,414 pairs, $R_c = 14 \text{ \AA}$).

and by storing the list of excluded atoms and the NB pair list as bit vectors, significant gains in speed were achieved. For a 1755 atom system (alpha-lytic protease), the time to generate the NB partner list (atom-based cutoff, no periodic boundaries) was decreased by a factor of 669 relative to a VAX 11/780. The time needed to evaluate the NB energy of the system was decreased by a factor of 237. No attempt was made to explicitly exploit the parallel nature of the ETA-10.

AMBER has also been optimized for the Alliant FX-series of supercomputers. These machines acquire most of the performance gains through large-grained parallelism. For instance, the set of NB partners corresponding to a set of atomic centers may be computed independently and concurrently on different processors for different sets of centers. The same principle applies to the evaluation of the NB energy of the system. The optimized program executes approximately 18 times faster than a VAX 8650 for a 554 atom system when using an 8 processor FX80-8 model. The results with other parallel systems are also encouraging [32,33]. As more processors become available on machines of this type it is quite feasible that the time needed to determine the conformational energy of a complex molecular system will decrease by another order of magnitude.

6. Molecular Dynamics Using AMBER, an example

For molecular dynamics simulations, the empirical force field of AMBER is used to calculate the acceleration applied to randomized, temperature equilibrated velocities. The new velocities are then applied to the current set of atomic coordinates for a short period, typically 0.0005 picoseconds (ps), in order to obtain a new coordinate set. The equations of motion are integrated using either a second order predictor two-step method [21] or a fifth order predictor-corrector multivalue method [34]. The user may optionally employ the van Gunsteren and Berendsen [35] SHAKE algorithm which fixes bond distances at their optimal values. With the SHAKE option the step can be increased to 0.001 ps. Temperature dependent velocities are assigned from the appropriate Boltzmann distribution function.

Research in this laboratory has involved investigations of the interactions between various electron transport proteins [17,36,37]. We have proposed the structure of a number of these complexes by attempting to correlate experimental results to structural features using computer graphics and computational methods. Interest in molecular dynamics simulations of these complexes was fueled by a recent MD study on the interaction between cytochromes c and b₅ [38]. In this study it was found that the nature of the intermolecular interactions allowed significant exploration of local conformational space and, in fact, that the simulation seemed to lead to geometries more suitable to electron transfer than the starting structure. This suggests that molecular dynamics simulations can be used not only for studies of atomic motions, but to optimize structures of model complexes.

A second observation from the Wendoloski et al. [38] study was that an aromatic amino acid side chain localized in the interface of the complex moved to a position where it might serve as an electron "bridge" during electron transfer. Site directed mutagenesis studies [39] have also implicated such a role for this residue.

In this case [38], simulations were performed both in vacuo and with explicit solvent atoms included. In both simulations the behavior of the molecules was quite similar, except that the movements seen in vacuo were more rapid. For example, the movement of the aromatic mentioned above occurred after only 5 ps of simulation in vacuo, but required 98 ps with explicit solvent. Thus, it is clear that solvent tends to dampen the atomic motions relative to an in vacuo simulation.

The intriguing results of Wendoloski et al. [38] prompted an MD study of the cytochrome c₃/rubredoxin complex from the sulfate reducing bacterium, *Desulfovibrio vulgaris* [17,37]. The crystal structures of both proteins were available from the Protein Data Bank [16] and a large body of experimental evidence could be used to guide creation of a potential protein-protein complex (Stewart et al., manuscript in preparation). The complex proposed is shown in Figure 5. Using the scalar Cyber 205 version of AMBER 3.0, this complex was first minimized (initially using the VAX) and then used to generate a 65 ps in vacuo molecular dynamics simulation. For this calculation a united atom representation was used with explicit hydrogens on polar groups, resulting in a total of 269 hydrogen atoms + 1392 heavy atoms (1661 total atoms). The timing data for the calculations is given in Table 5.

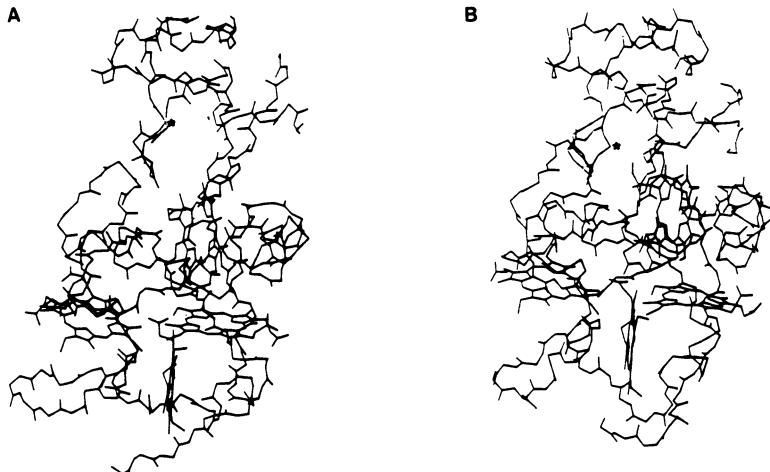


Figure 5. Structure of the proposed rubredoxin (top) - cytochrome c_3 (bottom) complex. Only the alpha-carbons of the protein are shown. The heme groups of cytochrome c_3 and the iron atom are also displayed. A - Structure after minimization. B - Structure after 65.0 ps of molecular dynamics simulation in vacuo.

TABLE 5. Timing Data for the Calculations.

RUN	COMPUTER	ITERATIONS	SEC/ITERATION
Constrained minimization	VAX 11/780	125	108.0
Unconstrained minimization	CYBER 205	10770	2.3*
Molecular Dynamics	CYBER 205	65000	4.8*

* Scalar version of AMBER 3.0 converted for operation on the Cyber 205 by Gallion and Flurhick [31].

Comparing the constrained minimization on the VAX 11/780 and the CYBER 205 it is evident that a supercomputer is needed for such studies, with a speedup of 50 observed here. In the molecular dynamics calculations, using a 0.001 ps time step 1000 steps are required for each 1 ps of simulation. This would be impractical on a computer with the speed of a VAX 11/780.

In this simulation the starting structure was minimized to an rms gradient of 0.025 kcal/ \AA . The energy minimized structure was used as the starting coordinates for the dynamics run. Initial velocities assigned to the energy minimized coordinates corresponded to a mean temperature of 50 K. The heating phase consisted of incrementing the temperature by 30 K every 200 femtoseconds (fs) to a final temperature of 300 K. Equilibration at 300 K was ensured by repeatedly reassigning random velocities from the appropriate Maxwellian distribution function every 200 fs for a period of 2.0 ps. The system was then allowed to stabilize for a period of 4.0 ps after which data collection was begun. A 9.5 \AA residue-based nonbonded cutoff was used, resulting in 360,000-370,000 nonbonded pairs. This was updated every 50 steps. Bondlengths were fixed at their optimal values using the SHAKE algorithm [35]. This calculation was done in vacuo with a distance dependent dielectric constant. This weights long range interactions less heavily than short range ones allowing for polarization and damping of interactions similar to those that should be observed if explicit solvent atoms were present [40]. However, not

including explicit solvent can result in some artifacts on the exterior of the protein such as improper hydrogen bond formation and artificially high atomic mobility [41].

Perhaps one of the best ways to examine the results from a molecular dynamics run is to display the trajectory on a computer graphics screen. Animation of the dynamics run using the commercial molecular modeling program QUANTA (Polygen, Inc.) revealed that the two proteins shift in relation to one another away from the minimized configuration. In addition major structural changes were seen to occur in both proteins in a concerted, non-random manner. Beta sheet regions on both proteins extend to one another with formation of new intermolecular hydrogen bonds. While this is an intriguing result causing distortions in both molecules, it may well be a result of this simulation being performed *in vacuo*. It has been shown that hydrogen bonds tend to be overweighted in such simulations [41]. Alternatively, this simulation could be revealing a natural adaptation of the two proteins upon binding.

The reorientation of the two molecules with respect to each other does increase favorable electrostatic and steric interactions [17]. The interaction energies of the complex gradually decrease during the initial steps of the simulation, until at 30 ps there is a marked decrease. Following this interaction, energies remain stable for the remainder of the simulation [17]. This suggests that the proteins translate over their surfaces until, at 30 ps, the optimal binding orientation is found. This transition is apparent in all the interaction energy terms (electrostatic, van der Waals, and hydrogen bonding). The number of ionic and hydrogen bonds between the two proteins increase during the initial steps of the simulation. Thus, dynamics simulation has led to an objectively determined model of low-energy interaction between cytochrome c₃ and rubredoxin with adaptive binding, an induced fit (Figure 5B).

The dihedral rotations of aromatic side chains in proteins have been well studied [42-51]. In these studies it has been found that the dihedral side chains of aromatic residues fluctuate from 30 to 90° depending on their location in the protein and that ring flips occur occasionally. Aromatic residues in electron transport proteins are especially interesting, since they have the potential to act as bridges for electron transfer via their electron-rich pi orbitals [52-53]. Evidence for such a role of aromatic residues perhaps may be found in the interaction between cytochrome c and cytochrome c peroxidase. Poulos & Kraut [53] proposed a complex between these two proteins with a large heme-to-heme distance [16.5 Å], but proposed that intervening aromatic residues could facilitate electron transfer. To examine this Liang and coworkers [39] performed site-directed mutagenesis studies on one of these intervening residues, phenylalanine-87 of cytochrome c. This residue was changed to tyrosine, serine, and glycine. Binding studies showed that the tyrosine and serine mutants formed apparently identical complexes with cytochrome c peroxidase; however, the rate of electron transfer in the serine mutant was 10⁴ slower than in the proteins with an aromatic residue at that position. While this does not prove that this residue participates in an electron "bridge", it does imply that it plays either a mechanistic or structural role [39]. A second indication of a possible role for aromatics in electron transfer between proteins came for the MD simulation of the cytochrome c/cytochrome b₅ complex discussed above. These considerations stimulated our interest in examining the behavior of the numerous aromatic residues in the cytochrome c₃/rubredoxin complex.

Probably the most intriguing set of aromatic residues are Tyr11 of rubredoxin (rd) and Tyr66 and His70 of cytochrome c. His 70 ligates heme 1 and is packed next to Tyr66. Tyr11 (rd) is located in the interface between the two proteins and is in the vicinity of Tyr66 (c₃). The difference between the two structures are striking since in the 65.0 ps structure Tyr11 (rd) has flipped up and is within 30-45 degrees to being parallel and is in contact with Tyr66 (c₃). Between 21 and 27 ps the dihedral angle of the sidechain of Tyr11 (rd) changes from an average value of 24 degrees to a new average value of -56 degrees. Since the starting dihedral angle was approximately 40 degrees this is an overall change of well over 90 degrees. The fluctuations prior the flip are around 50-60 degrees, but after the flip these have been damped to 20-30 degrees. An examination of the dihedral angle behavior of Tyr66 (c₃) reveals that at 27 ps its fluctuations are suddenly markedly

damped, decreasing from 40–60 degrees to 20–30 degrees. Considering this along with the structural data, it is evident that in this new orientation, Tyr11 (rd) is strongly interacting with Tyr66 (c₃). It is also evident that Tyr13 (rd) must at first be packed with Tyr11 (rd) until 27 ps, since at this time Tyr11 (rd) appears to become quite free and fluctuate substantially.

Since the residues are in the interface between the two proteins, this interaction could serve a mechanistic role in at least two ways. The stacking of aromatics between the two redox centers involving the histidine residue ligating heme 1, could provide an electron-rich path, or "bridge" for electron transfer. Alternatively, the role of this stack might simply lie in the reduced, coupled motions of the residues involved and the redox centers they link. The geometry is "solidified" by these interactions. The fact that the rings are not exactly parallel leaves open both possibilities, since there still is some pi-orbital overlap. In other proteins where adjacent aromatic residues play a structural role they are usually off-parallel by as much as 90 degrees [54].

These observations must be considered preliminary since the simulation was carried out *in vacuo*. However, we might expect that the interaction in the protein-protein interface would exclude water and thus be little effected by its presence. In addition, several studies have suggested that the motions and structural changes seen *in vacuo* using a distance dependent dielectric are quite similar to those in solution but that they occur much more rapidly without solvent damping [38,42,44, 48].

7. Summary and Conclusions

The ability to probe the dynamic behaviour of macromolecules at atomic resolution is central to the understanding of the mechanisms of ligand binding, catalysis, allostery and macromolecular association. The method of molecular dynamics allows such processes to be studied, but is restricted by the size of the systems of interest since computational demands increase rapidly as system size increases.

The rigorous approach of *ab initio* offers the most accurate and general solution to the problem of calculating conformational energies. However, the method cannot be feasibly used for systems of more than about 100 atoms due to the fact that computational time increases as M^4 or M^5 , where M is the total number of orbitals used. Semiempirical methods are also too time consuming to be of practical use for energy calculations involving biomolecules. Thus, the valence bond force fields of molecular mechanics currently offer the best solution to the problem of determining the conformational energy of a macromolecule in a reasonable amount of time. For this approach, the time for a single energy function evaluation increases as N^2 or N^3 , where N is the total number of atoms being considered. The problem is made even more tractable by employing the use of supercomputers to perform the calculations and by using approximations such as truncating the pairwise nonbonded interactions beyond a certain distance.

A number of force fields have been developed specifically for the purpose of studying proteins and nucleic acids. In this paper, we have discussed the form of the potential energy expression used in one of these, AMBER, and have reviewed some results obtained by using this force field in minimization and molecular dynamics simulations of electron transfer proteins. The results suggest that the processes of adaptive binding and induced fitting may be observed with simple *in vacuo* simulations of less than 100 ps. The stacking of aromatic residues in the protein-protein interface was noted and may point to a possible mechanistic role of these residues in electron transfer processes.

While results such as these are encouraging, there remains a great deal of work which is necessary to improve the performance, reliability, and utility of simulations using molecular mechanics force fields. The ability of programs such as MM2 to accurately handle highly strained molecules is of major importance to the rational design of pharmacological agents. [55] In addition, the treatment of processes involving electronic rearrangement which must be handled quantum mechanically, require the combination of molecular and quantum mechanics. [56–58] The ability to perform these calculations quickly is of great importance. Improvements in speed will make longer simulations on more complex systems using more accurate force fields or combined calculations more feasible.

Another recent development has been the incorporation of high performance graphics facilities in the design of supercomputers. The graphics supercomputer integrates computational power with high resolution, interactive computer graphics on the same machine rather than relying on the networked coupling between two or more machines. This merging permits graphical analysis of calculations while they are in progress [59]. This is of particular importance in molecular dynamics simulations where a visual monitoring of the trajectory may be alternated with interactive structural changes in order to achieve a more thorough understanding of the behaviour of the system under a variety of circumstances. These advances in hardware and software will also enable larger, more realistic systems to be studied using more accurate force fields and will undoubtedly contribute to the enhancement of our knowledge of sophisticated biological processes.

ACKNOWLEDGEMENTS.

The work described in this paper was supported by NIH (Grant GM41482), by NATO (5-2-05/RG #0404/88) and by the University of Georgia Research Foundation, the Advanced Computational Methods Center (ACMC) and the School of Chemical Sciences. Computer time and support from ETA Systems (1450 Energy Park Dr., St. Paul, Minnesota) and Alliant Computer Corp. (Littleton, Massachusetts) are gratefully acknowledged. SLG was supported by Alliant Computer Corp., ETA Systems and ACMC. DES was supported by ACMC, the School of Chemical Sciences and NIH. The authors are grateful to Professor N.L. Allinger (University of Georgia) for his pioneering efforts in molecular mechanics and helpful discussions and advice. We also gratefully acknowledge the help and support of Dr. Paul Weiner (Alliant) and Dr. Ken Flurhick (ETA Systems).

References

1. L. Pauling, R.B. Corey, H.R. Branson: Proc. Natl. Acad. Sci. USA **37**, 205-211 (1951)
2. M. Levitt, C. Chothia: Nature **261**, 552-558 (1976)
3. M. Levitt, S. Lifson: J. Mol. Biol. **46**, 269-279 (1969)
4. N.L. Allinger: J. Am. Chem. Soc. **99**, 8127-8134 (1977)
5. G. Levy: Chem. Design Auto. News **3**, 6-8 (1988)
6. R. Langridge, T.E. Ferrin, I.D. Kuntz, M.L. Connolly: Science **211**, 661-666
7. P. Gund, J.D. Andose, J.B. Rhodes, G.M. Smith: Science **208**, 1425-1431
8. S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., P. Weiner: J. Am. Chem. Soc. **106**, 765-784 (1984)
9. J. Lii, S. Gallion, C. Bender, H. Wikstrom, N.L. Allinger, K. Flurhick, M. Teeter: J. Comp. Chem., In Press (1988)
10. W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, M. Klein: J. Chem. Phys. **79**, 926-935 (1983)
11. U. Burkert, N.L. Allinger: Molecular Mechanics (American Chemical Society, Washington, D.C. 1982)
12. E.M. Engler, J.D. Andose, P. von R. Schleyer: J. Am. Chem. Soc. **95**, 8005-8025 (1973)
13. N.L. Allinger, R.A. Kok, M.R. Iman: J. Comp. Chem. **9**, 591-595 (1988)

14. M. Whitlow, M.M. Teeter: J. Am. Chem. Soc. **108**, 7163-7172 (1986)
15. D. Hall, M. Pavitt: J. Comp. Chem. **5**, 441-450 (1984)
16. D.E. Stewart, P.K. Weiner, J.E. Wampler: J. Mol. Graph. **5**, 137-144 (1987)
17. D.E. Stewart: The Structure, Interactions, and Dynamics of Electron Transport Proteins from Desulfovibrio: A Molecular Modeling and Computational Study; Ph.D. Dissertation, University of Georgia, Athens (1989)
18. F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.B. Rogers, D. Kennard, T. Shimanouchi, M. Tasumi: J. Mol. Biol. **112**, 535-542 (1977)
19. S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case: J. Comp. Chem. **7**, 230-252 (1986)
20. M. Karplus: Ann. N.Y. Acad. Sci. **482**, 107-123 (1986)
21. L. Verlet: Phys. Rev. **159**, 98-103 (1967)
22. M. Levitt: J. Mol. Biol. **168**, 595-620 (1983)
23. W.F. van Gunsteren, H.J.C. Berendsen, F. Colonna, D. Perahia, J.P. Hollenberg, D. Lelouch: J. Comp. Chem. **5**, 272-279 (1984)
24. H.J.C. Berendsen: in Proceedings of the Workshop on Molecular Dynamics and Protein Structure ed. by J. Hermans (University of North Carolina, Chapel Hill 1985) p. 18
25. R. Vogelsang, M. Schoen, M., C. Hoheisel: Comp. Phys. Comm. **30**, 235-241 (1983)
26. D.N.J. White: in Supercomputers in Chemistry (American Chemical Society, Washington, D.C. 1981)
27. C. Levinthal, R. Fine, G. Dimmler: in Proceedings of the Workshop on Molecular Dynamics and Protein Structure ed. by J. Hermans (University of North Carolina, Chapel Hill 1985) p. 126
28. B.R. Brooks: in Supercomputer Research in Chemistry and Chemical Engineering (American Chemical Society, Washington, D.C. 1987)
29. S.L. Gallion, R.M. Levy, P.K. Weiner, F. Hirata: Comput. Chem. **10**, 164-173 (1986)
30. R. Bott, E. Subramanian, D.R. Davies: Biochemistry **21**, 6956-6962 (1982)
31. S.L. Gallion, K.M. Flurchick: unpublished results (1988)
32. E. Clementi, G. Corongiu, J.H. Detrich, H. Khammohammabaidi, S. Chin, L. Domingo, A. Laaksonen, H.L. Nguyen: IBM Technical Report (POK-40) (1984)
33. J.B. Drake, A.K. Hudson, E. Johnson, D.W. Noid, G.A. Pfeffer, S. Thompson: Comput. Chem. **12**, 15-20 (1988)
34. J.W. Gear: Numerical Initial Value Problems in Ordinary Differential Equations (Prentice-Hall, New York 1971)
35. W.F. van Gunsteren, H.J.C. Berendsen: Mol. Phys. **34**, 1311-1327 (1977)
36. D.E. Stewart, J. LeGall, I. Moura, J.J.G. Moura, H.D. Peck, Jr., A.V. Xavier, P.K. Weiner, J.E. Wampler: Biochemistry **27**, 2444-2450 (1988)

37. D.E. Stewart, J. LeGall, I. Moura, J.J.G. Moura, H.D. Peck, Jr., A.V. Xavier, P.K. Weiner, J.E. Wampler: submitted (1988)
38. J.J. Wendoloski, J.B. Matthew, P.C. Weber, F.R. Salemme: *Science* **238**, 794-797 (1987)
39. N. Liang, G.J. Pielak, A.G. Mauk, M. Smith, B. Hoffman: *Proc. Natl. Acad. Sci. USA* **84**, 1249-1252 (1987)
40. J.M. Blaney, P.K. Weiner, A. Dearing, P.A. Kollman, E.C. Jorgensen, S.J. Oatley, J.M. Burridge, C.C.F. Blake: *J. Am. Chem. Soc.* **104**, 6424-6434 (1982)
41. M. Levitt, R. Sharon: *Proc. Natl. Acad. Sci. USA* **85**, 7557-7561 (1988)
42. B.R. Gelin, M. Karplus: *Biochemistry* **18**, 1256-1268 (1979)
43. J.A. McCammon, P.G. Wolynes, M. Karplus: *Biochemistry* **18**, 927-942 (1979)
44. S. Swaminathan, T. Ichiye, W.F. van Gunsteren, M. Karplus: *Biochemistry* **21**, 5230-5241 (1982)
45. M. Karplus, J.A. McCammon: *Ann. Rev. Biochem.* **53**, 263-300 (1983)
46. M. Levitt: *J. Mol. Biol.* **168**, 621-657 (1983)
47. C.B. Post, B.R. Brooks, M. Karplus, C.M. Dobson, P.J. Artymiuk, J.C. Cheetham, D.C. Phillips: *J. Mol. Biol.* **190**, 455-479 (1986)
48. L.X.Q. Chen, R.A. Engh, A.T. Brunger, D.T. Nguyen, M. Karplus, G.R. Fleming: *Biochemistry* **27**, 6908-6921 (1988)
49. R.F. Tilton, Jr., U.C. Singh, I.D. Kuntz, Jr., P.A. Kollman: *J. Mol. Biol.* **199**, 195-211 (1988)
50. M.E. Winfield: *J. Mol. Biol.* **12**, 600-611 (1965)
51. B.R. Gelin, M. Karplus: *Proc. Natl. Acad. Sci. USA* **72**, 2002-2006 (1975)
52. R.E. Dickerson, T. Takano, O.B. Kallai, L. Samson: Structure and Function of Oxidation Reduction Enzymes (Pergamon Press, Oxford 1972)
53. T.L. Poulos, J. Kraut: *J. Biol. Chem.* **255**, 10322-10330 (1980)
54. S.K. Burley, G.A. Petsko: *Science* **229**, 23-28 (1985)
55. R.S. Struthers, A.T. Hagler: in Conformationally Directed Drug Design ed. by J.A. Vida, M. Gordon (American Chemical Society, Washington, D.C. 1984) p. 239
56. N.L. Allinger, J.T. Sprague: *J. Am. Chem. Soc.* **95**, 3893-3907 (1973)
57. A. Warshel, R. Weiss: *J. Am. Chem. Soc.* **102**, 6218-6226 (1980)
58. U.C. Singh, P.A. Kollman: *J. Comp. Chem.* **7**, 718-730 (1986)
59. S.L. Gallion, R. Caltabiano, T.J. O'Donnell, U.C. Singh, P.K. Weiner: in preparation (1988)

Computer Simulation of Rotating Molecules in Condensed Phases

R.W. Gerling

Center for Simulational Physics, University of Georgia, Athens, GA 30602, USA and
Institut für Theoretische Physik I der Universität Erlangen-Nürnberg,
Glückstr.6, D-8520 Erlangen, Fed. Rep. of Germany

Abstract. The application of the Langevin-formalism for the study of rotating molecules is demonstrated with three examples:

- The rotational motion of the NH₄ tetrahedron in the molecular crystal NH₄Cl.
- A numerical solution of the nonlinear Langevin equation.
- A simple model for translational-rotational coupling.

1. Introduction

During the "séance du 9 Mars 1908" of the "academie des sciences" Paul Langevin presented a short paper "Sur la théorie du mouvement brownien" [1]. In this paper he published the now famous equation of motion

$$m \frac{d^2x}{dt^2} = -6\pi\mu a \frac{dx}{dt} + X \quad (1.1)$$

for a particle of mass m moving in a viscous medium under the influence of a stochastic force X . From Stokes law the resistance (friction) resulting from the viscosity of the medium is $-6\pi\mu a \frac{dx}{dt}$. a is the radius of the spherical particle and μ is the viscosity of the medium. Today we call this the Langevin equation and write, usually in a more modern notation (assuming n -dimensional vectors),

$$\ddot{\vec{x}} = -\Gamma \dot{\vec{x}} + \frac{G}{m} \vec{F}(t). \quad (1.2)$$

The acceleration $\ddot{\vec{x}}$ of a free particle is given by a friction term $-\Gamma \dot{\vec{x}}$ and a stochastic force $G\vec{F}(t)$. The strength of the stochastic force is G and $\vec{F}(t)$ is a fluctuating function of time which is characterized by

$$\begin{aligned} < \vec{F}(t) > &= 0 \\ < \vec{F}(t) \cdot \vec{F}(t') > &= \delta(t - t'). \end{aligned} \quad (1.3)$$

Normally the Langevin model is applied to a heavy particle in a bath of light particles. This means that the velocity of the heavy particle is small compared with the velocities of the bath particles. Thus the separation of time scales is justified.

In the case of rotational motion the situation is quite different. The particle of interest is of the same size (often it is exactly the same molecule) as the bath particles. Hence the conventional premise of a heavy particle in a bath of light ones is not correct. The interaction energy of two molecules m and n with orientations denoted collectively by Ω_m and Ω_n and a center of mass distance \vec{r}_{mn} is usually [2] split into its isotropic and anisotropic parts

$$V(r_{mn}, \Omega_m, \Omega_n) = V_i(r_{mn}) + V_a(r_{mn}, \Omega_m, \Omega_n). \quad (1.4)$$

By definition V_i is the spherical average of V , and therefore $V_i = \langle V \rangle$ and $\langle V_a \rangle = 0$. For globular molecules the anisotropic part V_a of the potential is much weaker than the isotropic part. This is clearly demonstrated by the existence of ODIC phases, where the central forces have established a crystalline order of the centers of mass of the molecules, but the weak anisotropic forces cannot achieve an orientational order of the molecules. In dense phases of globular molecules the rotational agitation of an individual molecule is permanent but weak at the same time. Therefore the angular momentum of a molecule can change only little on the time scale of the fluctuations of the torques created by its neighbors. Thus the separation of time scales leading to the Langevin picture is justified also for rotating molecules.

2. The Langevin Equation

To integrate the Langevin equation (1.2) we used a very simple and straightforward difference scheme

$$\dot{x}_i^{n+1} = \ddot{x}_i^n \Delta + \dot{x}_i^n \quad (2.1a)$$

$$x_i^{n+1} = \frac{1}{2} \ddot{x}_i^n \Delta^2 + \dot{x}_i^n \Delta + x_i^n. \quad (2.1b)$$

n labels the integration step and Δ is the length of the time interval in the simulation. We cannot use more sophisticated integration schemes, because the stochastic force prohibits that. This technique is correct up to order Δ^2 .

After every time step the conditions

$$|\vec{x}| = 1 \quad \vec{x} \cdot \dot{\vec{x}} = 0 \quad (2.2)$$

must be satisfied. This is necessary because we have the problem of the motion of a point on a sphere. To keep the point on the sphere the length of the radius-vector and the direction of the velocity are rescaled properly. Up to here it is a conventional integration technique of a system of differential equations. Now we have to introduce a stochastic noise source. This is modeled by a vector $\vec{F}_0(t)$, whose components are Gaussian distributed random numbers at any time t . $\vec{F}(t)$ is the projection of $\vec{F}_0(t)$ on a plane tangential to the sphere at the point $\vec{x}(t)$. In a Gauss-Markov process the probability that after time t a collision has occurred is given by [3]

$$p(t) = \frac{1}{T} \exp\left(-\frac{t}{T}\right), \quad (2.3)$$

where T is the mean time between two collisions. The probability that a collision has happened after a time Δ is

$$w(\Delta) = \int_0^\Delta p(t)dt = 1 - \exp\left(-\frac{\Delta}{T}\right) \approx \frac{\Delta}{T}, \quad (2.4)$$

where the last approximation is for a small time interval Δ . We have chosen a random torque of the following form:

$$F_i(t) = \sum_n f(t - n\Delta) \eta_i(n) \Theta(w - a_i(n)), \quad (2.5)$$

where the sum over n runs over all time steps. The $\eta_i(n)$ are Gaussian distributed random numbers; the $a_i(n)$ are equally distributed random numbers from the interval $[0, 1]$. $\Theta(x)$ is Heaviside's step function. The function $f(x)$ is unity for arguments between zero and Δ and zero elsewhere. The prefactor of the stochastic force G is given by the fluctuation-dissipation theorem

$$G = \sqrt{\frac{2I\Gamma T}{\beta\Delta^2}}. \quad (2.6)$$

β is the inverse temperature, measured in units of k_B . The factor $\frac{T}{\Delta^2}$ has its origin in the finite time steps in the computer simulation, leading to a torque which is not strictly δ -correlated.

We generated a pair of independent Gaussian distributed random numbers with the method of Box and Muller [4]:

$$\eta_1 = \sqrt{-2 \ln x_1} \cos(2\pi x_2) \quad (2.7a)$$

$$\eta_2 = \sqrt{-2 \ln x_1} \sin(2\pi x_2). \quad (2.7b)$$

This method requires two independent equally distributed random numbers x_1 and x_2 and provides two Gaussian distributed random numbers η_1 and η_2 with zero mean and unit variance. This method is exact and has also numerical advantages [5].

3. NH₄Cl

The Langevin model has been quite successfully used to study rotational motion of molecules in gases, liquids, and solids [6-13]. It is an excellent tool to study reorientational motion of molecules and molecular groups in molecular crystals. In this example the molecule NH₄ experiences a orientational potential which stems from its interaction with the Cl-neighbors in the molecular crystal NH₄Cl.

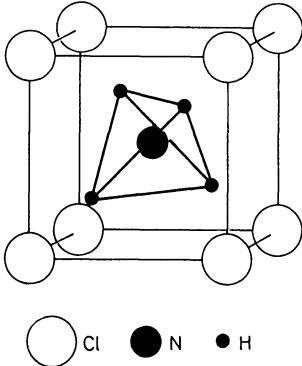


Figure 1: The lattice structure of NH_4Cl . The Cl^- ions are arranged in a cubic structure. The figure shows one of the two possible equilibrium orientations.

Hüller and Kane [14] used an electrostatic model to calculate the rotational potential of the ammonium ion in the cubic field of the surrounding halide ions and in the field of the surrounding ammonium ions. The structure of NH_4Cl can be seen in Fig. 1. The rotational potential of the ammonium ion is expanded in terms of symmetry adapted functions, in this case in terms of the cubic rotator functions $H_{\mu\nu}^{(\ell)}(\xi, \eta, \zeta)$ [15]. To avoid numerical problems integrating the differential equation quaternions $\vec{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4)$ [9,16] instead of Euler angles (ξ, η, ζ) were used. The rotational potential now reads

$$V_{rot}(\vec{\tau}) = B_{11}^3 H_{11}^{(3)}(\vec{\tau}) + B_{11}^4 H_{11}^{(4)}(\vec{\tau}) + B_{11}^6 H_{11}^{(6)}(\vec{\tau}). \quad (3.1)$$

The cubic rotator function $H_{11}^{(\ell)}(\vec{\tau})$ is a polynomial in the τ_i of order 2ℓ . The coefficients B_{11}^ℓ can be found in Ref. 14. B_{11}^3 is zero in the disordered phase, and non-zero in the ordered phase.

The simulation of the Langevin equation containing the potential (3.1) can be done by using the difference scheme (2.1) identifying x_i with τ_i ; the equation describes the motion of a point on the four-dimensional unit-sphere in the potential (3.1) driven by a stochastic force and hindered by a friction term. For the stochastic force the form (2.5) was used. The friction constant Γ was varied from 0.25 to 3.75. The temperature was varied between about 287 Kelvin and 1575 Kelvin. For each data point 10 runs of $2.5 \cdot 10^5$ time steps were averaged.

Töpler et al. [17] were able to measure the jump-rates ν_i of the ammonium tetrahedron around its four-fold and three-fold symmetry-axes separately. Their finding was that $\nu_4 \approx \frac{2}{3}\nu_3$. From the static potential (3.1) one can estimate that $\nu_4 \approx 60\nu_3$. The experimental result can be understood from the simulation.

The results from the simulation allow us to follow the path of the tetrahedron, so one can easily determine the jump rate around the different symmetry-axes by analysing the motion. Figure 2 shows typical results for the motion. The distance from a given minimum is plotted as a function of time. The time scale is in units of 10Δ . The number above the abscissa denotes the minimum. If the

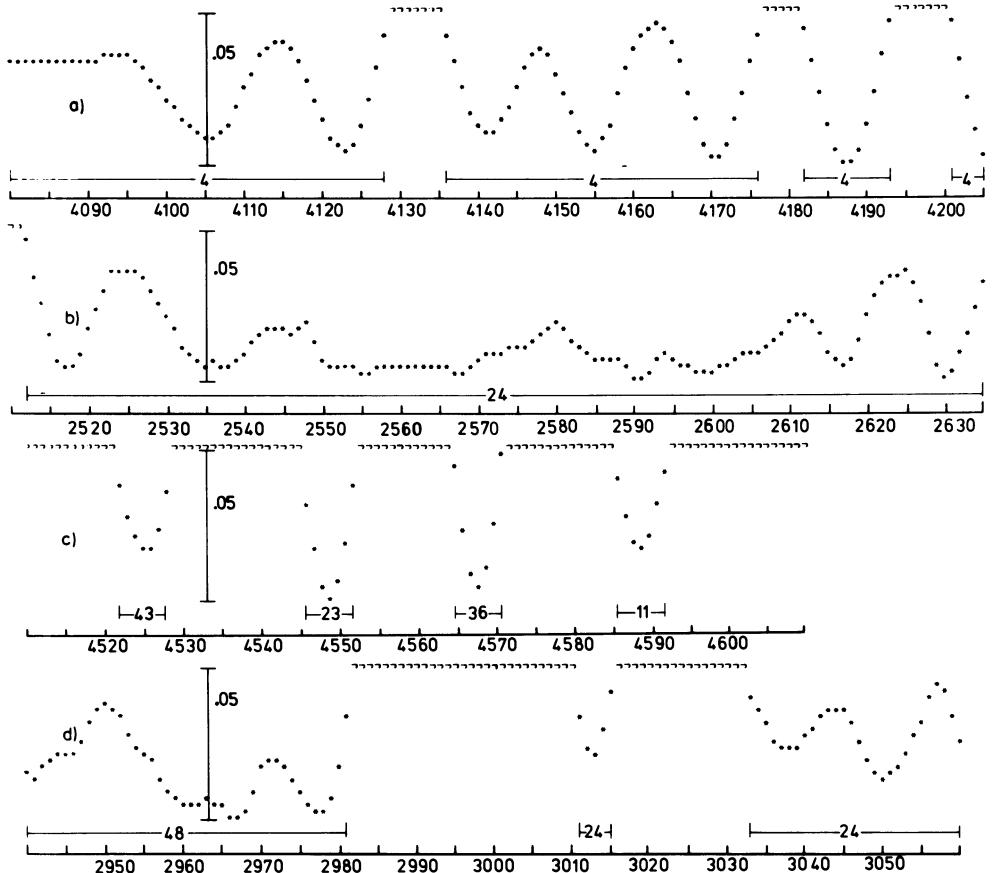


Figure 2: The figure shows the square of the distance of the tetrahedron from the minimum. The abscissa shows the time in units of 10Δ ; the number above the abscissa denotes the minimum.(a) The motion is oscillatory and $\Gamma = 0.25$. (b) The friction parameter $\Gamma = 2.50$ and the motion is stochastic. (c) Now the figure shows a multiple jump. All three jumps are 90° jumps around a (100) axis. The friction $\Gamma = 0.25$. (d) This figure shows a single jump. The jump is a 90° jump around a (100) axis. $\Gamma = 2.50$.

number changes the particle has jumped from one minimum to another. Figure 2a shows a result for small friction; the motion is nearly oscillatory. Figure 2b shows results for a ten times larger friction: the motion is stochastic.

In Figure 2c the key result can be seen: the tetrahedron, once jumping, does not jump just to the next minimum, but it makes multiple jumps through several minima. When the particle has picked up enough energy to cross a potential barrier, it does not lose its energy until it has passed through several other minima. A jump is finished when the total energy of a particle is less than the lowest barrier (barrier for 90° -jumps), because in this case the particle

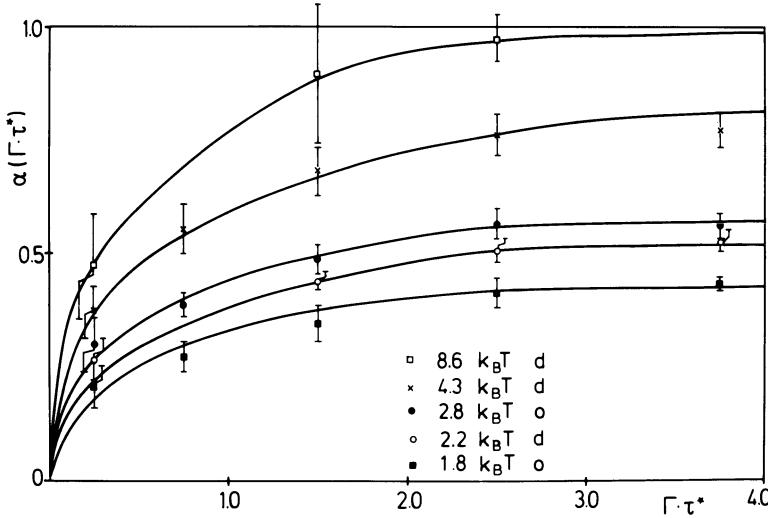


Figure 3: The figure shows the proportion α of single jumps as a function of the friction parameter Γ for different barrier heights as indicated.

cannot leave the minimum without any further agitation from the stochastic force. The percentage of multiple jumps can be changed almost continuously by varying the friction parameter. Figure 3 shows the fraction of single jumps as a function of the friction parameter.

If a given jump is a single jump, the probability that the jump is a 90° jump is almost independent of the friction. We found from the simulation

$$\begin{aligned} p_s^{90} &= 0.88 & p_s^{120} &= 0.13 \\ p_m^{90} &= 0.29 & p_m^{120} &= 0.40. \end{aligned} \quad (3.2)$$

Mixing single and multiple jumps we can produce the experimental results. The proportion of single jumps can be determined by

$$\begin{aligned} 0.88\alpha + 0.29(1 - \alpha) &= \nu_4 = \frac{2}{3}\nu_3 \\ 0.13\alpha + 0.40(1 - \alpha) &= \nu_3. \end{aligned} \quad (3.3)$$

The solution is $\alpha = 0.40$. Thus we have been able to explain the experiment by the dynamical effect of multiple jumps. The neutron scattering experiment cannot see the difference between multiple jumps and single jumps, because the jump duration in both cases is short compared to the residence time in a minimum.

Figure 4 shows a comparison of the quasi-elastic line-width with a first order approximation extracted from the computer simulation. The correlation function $\langle \vec{x}(t) \cdot \vec{x}(0) \rangle$ was fitted with an exponential for times larger than the inverse friction parameter. The decay-constant was used as the estimate for the

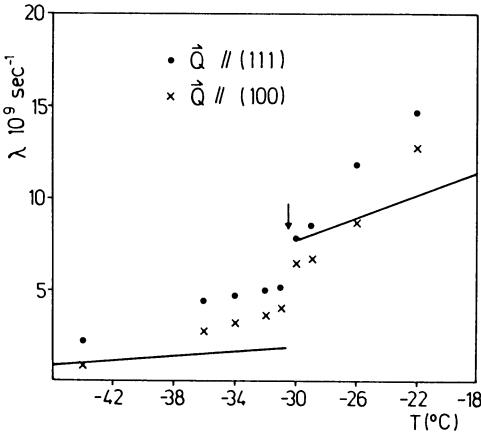


Figure 4: The width λ of the Lorentzians as a function of temperature. The points are from experimental data [17]. The lines are from the numerical simulation. The arrow indicates the phase transition.

width. The constants follow an Arrhenius law, which allowed us to extrapolate into the experimental temperature range. This simple estimate does not include a wave-vector dependence. The activation energies from the the Arrhenius law were

$$E_{ordered} = 214 \text{ meV}, \quad E_{disordered} = 181 \text{ meV}, \quad (3.4)$$

which agrees very well with the values of Hüller and Kane [14] and Töpler et al. [17].

4. Nonlinear Langevin Equation

For translational motion it is expected that the friction increases more strongly than linearly with the velocity. This is not the case for rotational motion. A molecule rotating around an n -fold symmetry-axis produces an alternating field of frequency $\omega = n\dot{\varphi}$. For large angular velocities $\dot{\varphi}$ the surroundings cannot respond to this high frequency field. From this we have to conclude that the friction term has to approach zero as $\dot{\varphi}$ goes to infinity.

Because the stochastic force and the friction have their origin in the same kind of fluctuations, both should have the same qualitative dependence on the velocity. This means that for large velocities the strength of the stochastic force should also approach zero in the rotational case.

We assume the nonlinear Langevin equation

$$\ddot{x} = -\Gamma(\dot{x})\dot{x} + \frac{G(\dot{x})}{m}F(t). \quad (4.1)$$

Because of mathematical difficulties we could not derive a fluctuation-dissipation theorem from equipartition as is done usually [18]. Instead we wrote down the

Fokker Planck equation for the distribution function $w(\dot{x}, t)$, which is equivalent to (4.1) [19]. We postulated (from physical reasons) that the Maxwell distribution should be the stationary solution of the Fokker Planck equation. This postulate allowed us to derive the nonlinear fluctuation dissipation theorem

$$\Gamma(\dot{x}) = \frac{1}{2m^2} \left(\frac{m}{k_B T} - \frac{1}{\dot{x}} \frac{\partial}{\partial \dot{x}} \right) G^2(\dot{x}). \quad (4.2)$$

This relation ensures thermal equilibrium for the nonlinear Langevin system in the same way the linear version of (4.2) does for the regular Langevin system. It is important to note that this is valid only in the Ito interpretation [18,20]. The calculation can be performed in the Stratonovich interpretation [18,21] as well, but for the numerical simulation the Ito interpretation is more convenient [19]. An additional restriction for $\Gamma(\dot{x})$ and $G(\dot{x})$ is that they have to be even functions of \dot{x} , because they should not depend on the direction of the velocity. One possible solution of (4.2) is

$$\begin{aligned} G(\dot{x}) &= \sqrt{2m\Gamma_0 k_B T} \exp\left(\frac{b\dot{x}^2}{4}\right) \\ \Gamma(\dot{x}) &= \frac{\Gamma_0 k_B T}{m} \left(\frac{m}{k_B T} - b \right) \exp\left(\frac{b\dot{x}^2}{2}\right) \\ b &< \frac{k_B T}{m}. \end{aligned} \quad (4.3)$$

The parameter b is the nonlinearity parameter. The case $b = 0$ reproduces the linear case. If $b < 0$ the friction term approaches zero for large \dot{x} : the rotational case. If $0 < b < \frac{k_B T}{m}$ the friction term increases faster than linear: the translational case. Figure 5 shows $-\Gamma(\dot{x})\dot{x}$ as a function of \dot{x} for several values of b .

Equation (4.1) cannot be solved for all possible values of the parameter b using standard integration routines. therefore we construct a special integration routine for this problem. The formal solution of Eq. (4.1) is

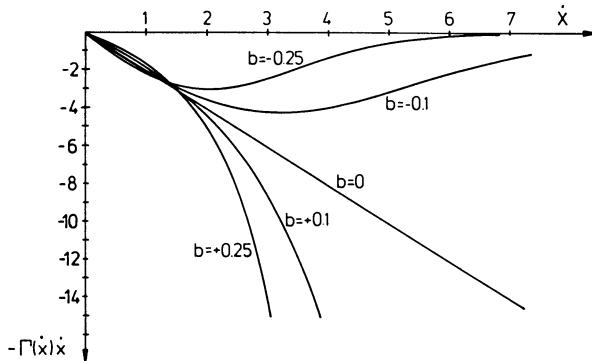


Figure 5: This figure shows the friction term $-\Gamma(\dot{x})\dot{x}$ as a function of \dot{x} . The parameters are: $k_B T = 1.0$; $\Gamma_0 = 2.0$; b as indicated.

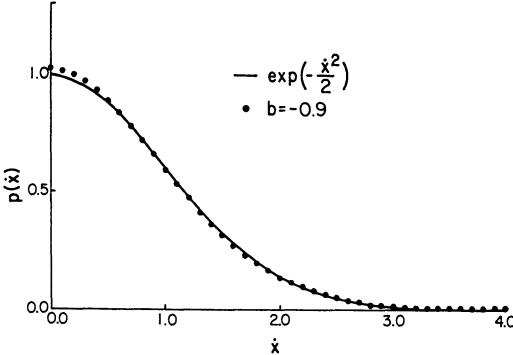


Figure 6: This figure shows the Maxwell distribution for positive velocities. The solid line is the function $\exp\left(-\frac{k_B T}{2}\dot{x}^2\right)$. The points are from the simulation. The curves are normalized, so that the area under the curve is the same. The parameters are: $k_B T = 1.0$; $\Gamma_0 = 2.0$; $3 \cdot 10^7$ time integration steps.

$$\begin{aligned} \dot{x}(t) &= \dot{x}(0) \exp\left(-\int_0^t \Gamma(\dot{x}(\tau)) d\tau\right) \\ &+ \exp\left(-\int_0^t \Gamma(\dot{x}(\tau)) d\tau\right) \int_0^t \exp\left(-\int_0^{\tau'} \Gamma(\dot{x}(\tau')) d\tau'\right) g(\dot{x}(\tau)) F(\tau) d\tau. \end{aligned} \quad (4.4)$$

For small time intervals Δ the integrals can be approximated by the integrands at the lower boundary times the length of the interval. This yields the integration scheme:

$$\dot{x}_{n+1} = \left(\exp(-\Gamma(\dot{x}_n)\Delta) \right) (\dot{x}_n + G(\dot{x}_n)F(n\Delta)\Delta), \quad (4.5)$$

where n labels the integration step. Expanding the exponential one gets back the standard integration scheme (2.1a). The advantage of (4.5) over (2.1a) is that it explicitly uses exponential decay due to friction.

We performed $3 \cdot 10^7$ integration steps with $\Delta = 10^{-3}$. Figure 6 shows a comparison of the simulation result with the Maxwell distribution. One sees that it is reproduced rather well. In Fig. 7 results for the correlation function $\langle \dot{x}(t)\dot{x}(0) \rangle$ are shown. In the rotational case one clearly sees several correlation times. Surprisingly we see only one correlation time in the translational case. This might be the reason for the success of the linear Langevin picture in describing translational motion. The error, within which the correlation functions can be reproduced from different runs, is of the order of the symbol size.

5. Translational Rotational Coupling

We have simulated a simple planar model consisting of four harmonic oscillators moving on the four sides of a square with a free dumbbell-rotator in the center of the square. The four oscillators are coupled to a heat bath by a stochastic

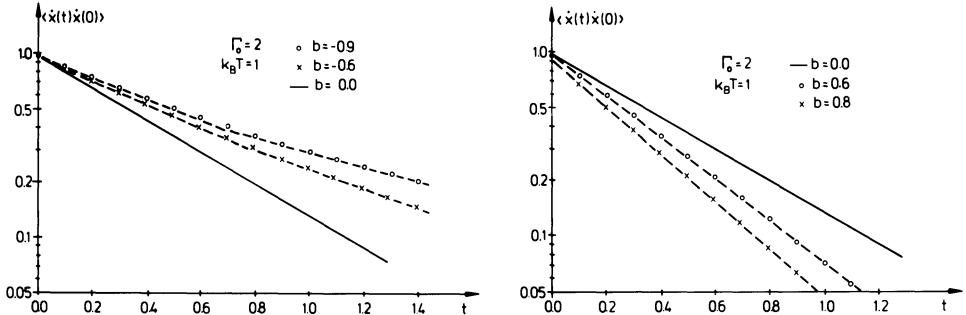


Figure 7: The results for the correlation function $\langle \dot{x}(t)\dot{x}(0) \rangle$. Left part: rotational case; b negative. Right part: translational case; b positive.

force and a linear friction. Fig. 8 outlines the geometrical relations. The four oscillators interact with the dumbbell in the center via a simple potential. One can think of several different interactions between the central particle and the outer particles. We chose an attractive $\frac{1}{r}$ potential. The two parameters of the model are the length of the rotator $2a$ and the length of a side of the square $2b$. The potential is given by

$$V(x_1, x_2, x_3, x_4, \varphi) = -k \left(f_1(x_1, \varphi) + f_1(x_3, \varphi) + f_2(x_1, \varphi) + f_2(x_3, \varphi) \right. \\ \left. f_3(x_2, \varphi) + f_3(x_4, \varphi) + f_4(x_4, \varphi) + f_4(x_4, \varphi) \right) \quad (5.1)$$

where the x_i ($i = 1, 2, 3, 4$) are the displacements of the oscillators measured from the middle of the four sides in a counter-clockwise direction and φ is the rotational angle of the dumbbell rotator. For simplicity we have introduced the four functions

$$f_1(x, \varphi) = (a^2 + b^2 + x^2 - 2a(b \cos \varphi + x \sin \varphi))^{-1/2} \quad (5.2a)$$

$$f_2(x, \varphi) = (a^2 + b^2 + x^2 + 2a(b \cos \varphi + x \sin \varphi))^{-1/2} \quad (5.2b)$$

$$f_3(x, \varphi) = (a^2 + b^2 + x^2 - 2a(b \sin \varphi - x \cos \varphi))^{-1/2} \quad (5.2c)$$

$$f_4(x, \varphi) = (a^2 + b^2 + x^2 + 2a(b \sin \varphi - x \cos \varphi))^{-1/2} \quad (5.2d)$$

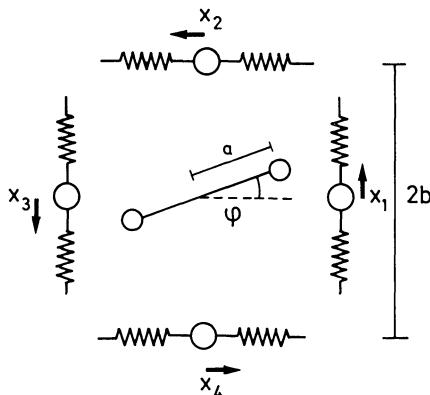


Figure 8: The figure shows the geometrical outline of the system. The masses indicated by the spheres were treated as point-masses.

The equations of motion for the system are

$$\ddot{x}_i = -\Omega^2 x_i - \frac{1}{m} \nabla_{x_i} V(x_i, \varphi) - \Gamma \dot{x}_i + G^{1/2} \frac{F_i(t)}{m}, i = 1 \dots 4 \quad (5.3a)$$

$$\ddot{\varphi} = -\frac{1}{I} \nabla_{\varphi} V(x_i, \varphi) \quad (5.3b)$$

where m and I are the mass of the identical oscillators and the moment of inertia of the rotator, respectively. The equations were integrated by using the difference scheme (2.1). The stochastic force pulses were modeled by the form given in (2.5).

The static potential for the rotator is shown in the bottom part of Fig. 9. In the upper part of Fig. 9 the time-average of the potential is shown:

$$\langle V(x_i(t), \varphi) \rangle_t = \frac{1}{M} \sum_{n=1}^M V(x_i^n, \varphi). \quad (5.4)$$

Due to the symmetry of our model the potential is always π -periodic with respect to φ . The rotator will in most cases be observed around the minima at $\varphi = n \cdot \frac{\pi}{4}, n \in N$. To observe its behavior best we will from now on fold φ in the interval $]-\frac{\pi}{4}, \frac{3\pi}{4}]$. Our investigations [22] show that the effects we want to study are most obviously observed if we use $b = 2a$.

For these values of the parameters the height of the potential wall for the static case becomes

$$V_b = V(x_i = 0, \varphi = \frac{\pi}{4}) - V(x_i = 0, \varphi = 0) \approx \frac{k}{a} \cdot 0.3115. \quad (5.5)$$

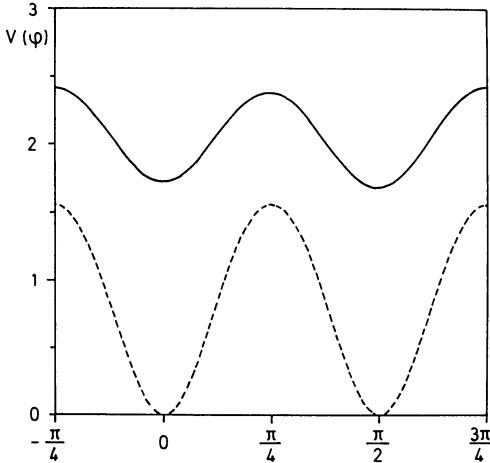


Figure 9: The figure shows the static potential in the lower part and the time averaged potential in the upper part.

We set the energy of the bath to $k_B T = 1$, $a = 1$ and $k = 5$ which results in $V_b = \frac{3}{2}k_B T$. The transition-rate for the rotator in the static case can be approximated by the probability of finding it at the maximum of the potential. That is equal to $\exp(\frac{-V_b}{k_B T}) = \exp(-1.5) = 0.22$ which is necessary for numerical simulations. In reality $\frac{V_b}{k_B T}$ is far larger, resulting in much lower jump-rates. To keep the calculation-times on the computer within reasonable limits this ratio must be smaller than in reality.

The only factor which could change the probability of a transition is a substantially lower mean energy barrier in the dynamical system. This can in fact be observed.

One possibility of studying the dynamics is to look at the distribution of the energy-barriers to be passed by the rotator. Our studies have shown that most of the time the energy-barriers are of different height. The rotator normally crosses the lower energy-barrier. Taking this into account we generated two graphs (Fig. 10) showing the height-distribution of the energy-barrier for the regular system and for the case where the higher potential-wall is neglected. The average barrier-height is in the regular system $\langle E_1 \rangle = 1.237k_B T$ and in the other case $\langle E_2 \rangle = 1.001k_B T$. Thus we see that the dynamics of the system causes a shift of the mean energy-barrier from $V_b = 1.5k_B T$ in the static case to $\langle E_2 \rangle = 1.0k_B T$ which can be regarded as the mean energy-barrier for the rotator to be crossed.

Another approach to this problem consists of simply observing the position of the minima as a function of time. It happens that (resulting from the displacement of the oscillators) one minimum vanishes and changes to e.g. a saddlepoint. In this case a transition of the rotator from one orientation to another orientation is primarily determined by its environment and not by its kinetic energy.

In Fig. 11 we see a generic situation in which the position of the rotator and the positions of the minima are shown. We see that some transitions are simply caused by the vanishing of a minimum where the particle was oscillating.

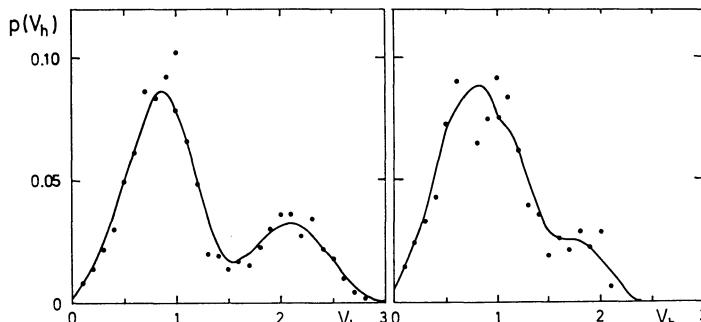


Figure 10: The figure shows the distribution of the barrier heights. In the left part both barriers are counted, in the right part only the lower of the two barriers is taken into account.

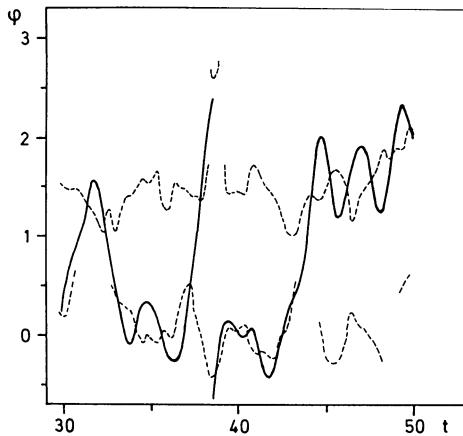


Figure 11: The figure shows the time evolution of the position of the minima (dashed lines) and the orientation of the rotator (solid line). Note the disappearance of the minima.

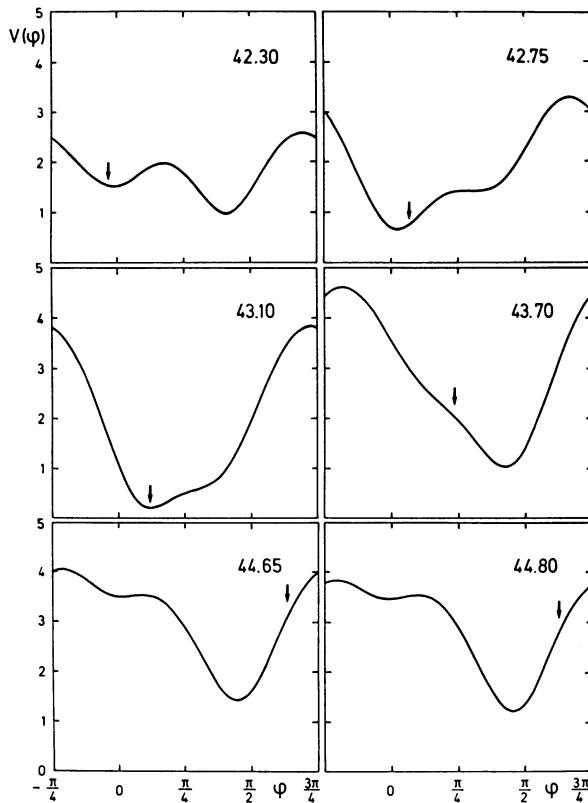


Figure 12: The figure shows the potential with the orientation of the rotator marked by an arrow for six different times. The times are chosen such that the disappearance of one minimum can be seen clearly.

In order to provide a better insight into the dynamics of one transition Fig. 12 shows a time-sequence of several pictures of the potential including the position of the rotator. This is another way to see that the transition-rate is also mainly determined by the transitions caused by the vanishing of the energy-barrier.

6. CONCLUSIONS

The simulation of the Langevin equation has the advantage that one very easily gets information about the detailed dynamics of the system. Such information is not available from other methods of solving the problem. All other methods calculate directly averaged quantities, e.g. correlation functions, where the numerical Langevin approach probes one particular realisation of the stochastic force. The direct access to the trajectory of the system gives additional insight, making the simulation of the Langevin equation a very powerful tool in understanding the detailed dynamics of a lot of systems, of which the rotational problems presented here are only one example.

7. ACKNOWLEDGEMENTS

I want to thank the Alexander von Humboldt Stiftung for the award of a Feodor-Lynen-fellowship. Helpful discussions and fruitful collaborations with A. Hüller, D. Klotz and A. Rau are acknowledged. The work is partly supported by the National Science Foundation (NSF) Grant No. DMR-86-03605 and by the "Advanced Computational Methods Center" at the University of Georgia.

8. REFERENCES

1. P. Langevin, Compt. Rend. **43**, 530 (1908).
2. I.F. Silvera, Rev. Mod. Phys. **52**, 393, (1980).
3. T. Schneider and E. Stoll, Phys. Rev. B **17**, 1302 (1978).
4. G.E.P. Box and M.M. Muller, Ann. Math. Stat. **29**, 610 (1958).
5. R.W. Gerling, Thesis, University of Erlangen, unpublished.
6. R.E.D. McClung, J. Chem. Phys. **73**, 2435 (1980); **75**, 5503 (1981).
7. G. Levi, J.P. Marsault, F. Marsault-Hérail, and R.E.D. McClung, J. Chem. Phys. **73**, 2443 (1980).
8. R.M. Levy, M. Karplus, and J.A. McCammon, Chem. Phys. Lett. **65**, 4 (1979).
9. R.W. Gerling and A. Hüller, Z. Phys. B **40**, 209 (1980).
10. R.W. Gerling, Z. Phys. B **45**, 39 (1981).
11. R.W. Gerling and B. de Raedt, J. Chem. Phys. **77**, 6263 (1982).
12. R.W. Gerling and A. Hüller, J. Chem. Phys. **78**, 446 (1983).
13. D. Klotz and A. Hüller, Z. Phys. B **63**, 69 (1986).
14. A. Hüller and J.W. Kane, J. Chem. Phys. **61**, 3594 (1974).
15. C. Brot and I. Darmon, Mol. Phys. **21**, 785 (1971).
16. F. Hund, Z. Physik **51**, 1 (1928); W.H. Furry, Phys. Rev. **107**, 7 (1957).
17. J.M. Töpler, D.R. Richter, and T. Springer, J. Chem. Phys. **69**, 3170 (1978).
18. N.G. van Kampen, **Stochastic Processes in Physics and Chemistry**, North Holland, Amsterdam (1981).
19. R.W. Gerling, Z. Phys. B **55**, 263 (1984).
20. K. Ito, Proc. Imp. Acad. Tokyo **20**, 519 (1944); Mem. Am. Mathem. Soc. **4**, 51 (1951).
21. R.L. Stratonovich, SIAM J. Control **4**, 362 (1966).
22. A. Rau and R.W. Gerling, preprint (1989).

Novel Contexts of Computer Simulation

H.L. Frisch

Department of Chemistry, State University of New York, Albany, NY 12222, USA

Abstract. In this review paper we illustrate the application of standard simulation methods in statistical physics in two novel contexts. The first deals with the extraction of a dynamic relaxation time from a biased sampling Monte Carlo computation on equilibrium data. The specific example reviews a previously published Rouse model polymer chain dynamics derived from the Kramers potential with and without excluded volume interaction. The second deals with the approximate representation of continuous functions of many variables of physical interest in information retrieval or condensed matter/statistical physics in terms of continuous functions of one variable. The arguments of the latter can be stored once and for all in a computer memory.

1. Introduction

Our first application of computer simulation asks whether a Monte Carlo computation can be employed to obtain dynamical relaxation times in certain selected situations. Typically the system under study should evolve in time according to some simplified, stochastic dynamics, e.g. Glauber or Kawasaki dynamics in spin systems or Brownian dynamics (described by a Fokker Planck equation) in polymer chain systems. One is interested in finding a transport coefficient (the simplest topic in non-equilibrium statistical mechanics) because it suffices to consider only the stationary state arising from a weakly perturbed equilibrium state. We assume we know the effective "potential" from which one can derive the perturbing force(s) acting on the equilibrium state which induce the appropriate stationary state which relaxes with the characteristic time we wish to study. This effective "potential" is used to bias the Monte Carlo calculation of a typical equilibrium property which is chosen to relax with this characteristic relaxation time. In general, we expect that such calculations will be valid only in a restricted interval of values of a physical parameter which limits both the description of the perturbation of the system by the chosen "effective" potential and

simultaneously lies in the validity range of the biased Monte Carlo calculation.

The specific example treated in Section 2 illustrating this procedure is a precis of a published calculation [1] of the viscosity relaxation time of a single Rouse model chain. We actually quote almost without modification some paragraphs of this publication; for a more extended treatment including some generalizations of this problem we refer our readers to reference [1]. Over 40 years ago Kramers [2] suggested a novel approach to the calculation of the intrinsic viscosity and other hydrodynamic properties of dilute polymer solutions at low shear rate. This was based on replacing the simple shear flow $\mathbf{v} = \dot{\gamma}ye_x$ (e_i Cartesian unit vector in i th direction, $\dot{\gamma}$ shear rate), by its irrotational part, which is derivable from a velocity potential $\phi = 1/2 \dot{\gamma}xy$. It turns out to lowest order in $\dot{\gamma}$ that this replacement does not affect the behavior of certain hydrodynamic properties, e.g., the intrinsic viscosity, if the long range hydrodynamic forces and Brownian motion are neglected. This method has been employed recently by Fesjian and Frisch [3] to find the rotational relaxation time of rods of very large aspect ratio (the Onsager [4] problem) in a simple shear flow. We explore the applicability of the Kramers potential to find dynamical relaxation times by carrying out a biased sampling Monte Carlo computation [6] on equilibrium data. We compare analytically the configurational distribution functions (CDF) [5] of a Rouse chain without excluded volume in dilute solution in the following flows: (1) simple shear; (2) Kramers potential streaming (i.e., elongation flow); and (3) pure rotation. In particular we show that the exact CDF of simple shear can be written as a product of three factors describing equilibrium, Kramers potential streaming, and rotation. As a consequence to the lowest order in $\dot{\gamma}$ (i.e., terms of order $\dot{\gamma}^2$), the radius of gyration of a Rouse chain subject to simple shear is the same as that of one subject only to Kramers potential streaming. We use these relations in the biased sampling Monte Carlo calculations of the radius of gyration of chains of $N = 8, 16, 32, 64, 100$, and 150 monomer beads on a simple cubic lattice. Except for the excluded volume effect these chains mimic Rouse chains and we obtain the relaxation time τ associated with the viscosity difference $\eta - \eta_s$ (η_s "solvent" viscosity), divided by the number density. This τ scales as expected with $N^{1+2\nu}$.

In section 3 we briefly summarize a recent suggestion [7] for representing approximately continuous functions of several (say s) variables on a compact domain, such as the s -dimensional unit hypercube, in terms of a superposition of continuous functions of one variable obtained by a few iterations of an

explicit, numerical construction. The constructive approximation of functions of two variables ($s = 2$) works astonishingly well and gives an excellent geometrical representation of the original function. We make a number of concluding remarks in the last section.

2. The Viscosity Relaxation Time of a Rouse Chain from a Biased Monte Carlo Calculation.

Consider a Rouse chain consisting of N beads with friction coefficient ζ connected by Hookean springs with spring constant H , the beads being located at \underline{r}_v , $v = 1, 2, \dots, N$. We define bond vectors $\underline{Q}_{k+1} = \underline{r}_{k+1} - \underline{r}_k$, $k = 1, 2, \dots, N-1$ and introduce normal coordinates \underline{Q}'_k by

$$\underline{Q}_j = \sum_k \Omega_{jk} \underline{Q}'_k. \quad (1)$$

It can be shown that in these coordinates the CDF of a Rouse chain in a velocity field $\underline{v} = \underline{v}_0 + \underline{\kappa} \cdot \underline{r}$ factorizes [6]

$$\psi(\underline{Q}'_1, \underline{Q}'_2, \dots, \underline{Q}'_{N-1}, t) = \prod_k \psi_k(\underline{Q}'_k, t) \quad (2)$$

and each ψ_k satisfies a diffusion (Smoluchowski) equation:

$$\frac{\partial \psi_j}{\partial t} = -\frac{\partial}{\partial \underline{Q}'_j} \left[\left[\underline{\kappa} \cdot \underline{Q}'_j \right] \psi_j - \frac{a_j}{\zeta} \left[k_B T \frac{\partial}{\partial \underline{Q}'_j} \psi_j + H \underline{Q}'_j \psi_j \right] \right], \quad (3)$$

where a_j is the j th eigenvalue of the Rouse Matrix [8]

$$a_j = 4 \sin^2 \left[\frac{j\pi}{2N} \right] = \frac{1}{c_j}. \quad (4)$$

The c_j 's are the eigenvalues of the Kramers matrix which is the inverse of the Rouse matrix. Both matrices are explicitly written down in Ref. [6].

The flows we will consider are special cases of the velocity field

$$\underline{v} = \frac{1}{2} \kappa_1 y \underline{e}_x + \frac{1}{2} \kappa_2 x \underline{e}_y \quad (5a)$$

with

$$\underline{v}_0 = 0 \text{ and } \underline{\kappa} = \begin{pmatrix} 0 & \frac{1}{2} \kappa_1 & 0 \\ \frac{1}{2} \kappa_2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (5b)$$

(a) simple shear

$$\kappa_1 = 2\dot{\gamma}, \kappa_2 = 0, \underline{v} = \dot{\gamma} y \underline{e}_x;$$

(b) Kramers potential flow

$$\kappa_1 = \kappa_2 = \dot{\gamma}, \quad \underline{v} = \frac{1}{2} \dot{\gamma} (y \underline{e}_x + x \underline{e}_y) = \nabla (\frac{1}{2} \dot{\gamma} xy);$$

(c) rotation

$$\kappa_1 = -\kappa_2 = \dot{\gamma}, \mathbf{v} = \frac{1}{2} \dot{\gamma} (\mathbf{y}\mathbf{e}_x - \mathbf{x}\mathbf{e}_y).$$

The stationary solution of Eq. (3) with flows as given in Eq. (5) can be written in the form [1]

$$\psi_j = \frac{1}{J_j} \exp \left\{ -\frac{H}{2k_B T} \times [A_j X_j'^2 + B_j X_j' Y_j' + C_j Y_j'^2 + Z_j'^2] \right\}, \quad (6)$$

where $\Omega_j' = X_j' \mathbf{e}_x + Y_j' \mathbf{e}_y + Z_j' \mathbf{e}_z$ and J_j is the normalization constant. The coefficients A_j , B_j , C_j are displayed in Table I for our three special flows ($\lambda_j = \zeta c_j / 2H$). Substitution of Eq. (6) into Eq. (2) enables us to write the non-equilibrium CDFs with equilibrium CDF,

$$\psi_0 = \frac{1}{J_0} \exp \left\{ -\frac{H}{2k_B T} \sum_n (X_n'^2 + Y_n'^2 + Z_n'^2) \right\} \quad (7)$$

factored out, viz., for shear:

$$\psi_s = \frac{J_0}{J_s} \psi_0 \exp \left\{ \frac{H}{2k_B T} \times \sum_n \left(\frac{\lambda_n^2 \dot{\gamma}^2 X_n'^2 + 2\lambda_n \dot{\gamma} X_n' Y_n' - \lambda_n^2 \dot{\gamma}^2 Y_n'^2}{1 + \lambda_n^2 \dot{\gamma}^2} \right) \right\}, \quad (8)$$

for Kramers potential flow:

$$\psi_p = \frac{J_0}{J_p} \psi_0 \exp \left\{ \frac{H}{2k_B T} \sum_n 2\lambda_n \dot{\gamma} X_n' Y_n' \right\}, \quad (9)$$

and for rotation:

$$\psi_r = \frac{J_0}{J_r} \psi_0 \exp \left\{ \frac{H}{2k_B T} \sum_n \left(\frac{\lambda_n^2 \dot{\gamma}^2 X_n'^2 - 2\lambda_n^3 \dot{\gamma}^3 X_n' Y_n' - \lambda_n^2 \dot{\gamma}^2 Y_n'^2}{1 + \lambda_n^2 \dot{\gamma}^2} \right) \right\}. \quad (10)$$

TABLE I. The coefficients of Eq. (6).

	Equilibrium	Shear	Kramers potential	Rotation
A_j	1	$\frac{1}{1 + \lambda_j^2 \dot{\gamma}^2}$	1	$\frac{1 + 2\lambda_j^2 \dot{\gamma}^2}{1 + \lambda_j^2 \dot{\gamma}^2}$
B_j	0	$-\frac{2\lambda_j \dot{\gamma}}{1 + \lambda_j^2 \dot{\gamma}^2}$	$-2\lambda_j \dot{\gamma}$	$\frac{2\lambda_j^3 \dot{\gamma}^3}{1 + \lambda_j^2 \dot{\gamma}^2}$
C_j	1	$\frac{1 + 2\lambda_j^2 \dot{\gamma}^2}{1 + \lambda_j^2 \dot{\gamma}^2}$	1	$\frac{1}{1 + \lambda_j^2 \dot{\gamma}^2}$

It should be noted that

$$\psi_s = \frac{J_0 J_p J_r}{J_s} \psi_0 \psi_p \psi_r \quad (11)$$

and that [1]

$$\psi_s = \frac{J_0}{J_s} \psi_0 \exp \left\{ \sum_v \frac{U_{Kv}(\mathbf{r}_v - \mathbf{r}_c)}{k_B T} \right\} \exp \left\{ \frac{H}{2k_B T} \times \sum_v \sum_\mu \left([D_1]_{\nu\mu} (x_\nu x_\mu - y_\nu y_\mu) - [D_2]_{\nu\mu} x_\nu y_\mu \right) \right\}, \quad (12)$$

where $U_{Kv}(\mathbf{r})$ is the Kramers potential

$$U_{Kv}(\mathbf{r}) = \frac{1}{2} \zeta \dot{\gamma} xy \quad (13)$$

and \mathbf{r}_c is the location of the center of mass of the chain

$$\mathbf{r}_c = \frac{1}{N} \sum_v \mathbf{r}_v. \quad (14)$$

The matrices D_1 and D_2 in the correction term are of order $\dot{\gamma}^2$ and $\dot{\gamma}^3$, respectively [1], Eq. (12) shows explicitly that to lowest order in $\dot{\gamma}$ the CDFs of the simple shear flow and the Kramers potential flow are identical. Specifically if one computes the radius of gyration then one finds ($N \rightarrow \infty$)

$$\begin{aligned} \langle R_{gyr}^2 \rangle &= \frac{1}{2N^2} \sum_\mu \sum_v \langle (\mathbf{r}_\mu - \mathbf{r}_v)^2 \rangle \\ &= \langle R_{gyr}^2 \rangle_0 \left[1 + \frac{16}{945} N^4 \lambda_H^2 \dot{\gamma}^2 \right], \end{aligned} \quad (15)$$

where $\langle R_{gyr}^2 \rangle_0$ is the equilibrium value of the radius of gyration and $\lambda_H = \zeta/4H$. This is the exact result for simple shear [9] and is correct for Kramers potential flow in lowest order of $\dot{\gamma}$ [1]. The viscosity difference relaxation time τ then becomes [6]

$$\tau = \sqrt{\frac{16}{945}} N^2 \lambda_H = \sqrt{\frac{16}{105}} \left(\frac{\eta - \eta_s}{nk_B T} \right) \quad (16)$$

with n the number of density of chains in the solution of viscosity η [1].

All previous calculations refer to the ideal Rouse model with a Gaussian chain configuration resulting in the absence of excluded volume interaction. We now consider the Rouse model with excluded volume interaction, where the mean square end-to-end distance $\langle R^2 \rangle$ and mean-square gyration radius $\langle R_{gyr}^2 \rangle$ are described by an exponent v ($v \approx 0.59$ in three dimensions [10]):

$$\langle R^2 \rangle \sim N^{2v}, \quad (17)$$

$$\langle R_{gyr}^2 \rangle \sim N^{2v} \quad (18)$$

and the distribution function $P(r_{ij})$ for the distance r_{ij} between monomers i,j along the chain no longer is a Gaussian [11]. In this case the eigenvalue spectrum is no longer given by Eq. (4), and in fact it is not explicitly known. Scaling arguments, however, predict that the relaxation time τ should scale with chain length N as [11]

$$\tau \propto N^{1+2\nu} (\approx N^{2.18} \text{ for } d=3) \quad (19)$$

and this result is confirmed by various renormalization group treatments and dynamic Monte Carlo simulations. For dynamic Monte Carlo simulations of the Rouse model with excluded volume, one typically models the polymer chain as a self-avoiding walk of N steps on a lattice, e.g., the simple-cubic lattice, and performs random motions of groups of neighboring bonds to new positions which satisfy the excluded volume constraint [5]. Measuring "time" in units of Monte Carlo steps per bond, one then obtains, for instance, the "time-correlation" of the end-to-end vector $\underline{R}(t)$ of the chain

$$\langle \underline{R}(0) \cdot \underline{R}(t) \rangle \approx \frac{1}{t_{\text{obs}} - t} \int_0^{t_{\text{obs}} - t} dt' \underline{R}(t') \cdot \underline{R}(t' + t). \quad (20)$$

In Eq. (20) we have already anticipated that in the context of such a dynamic simulation the ensemble average $\langle \dots \rangle$ really is replaced by a time average up to some observation time t_{obs} . Of course, Eq. (20) makes sense only for $t_{\text{obs}} \gg t$ otherwise the numerical results are "noisy" and unreliable.

Now the desired relaxation time τ_R is defined as [5,13]

$$\tau_R \equiv \int_0^{\infty} dt \langle \underline{R}(0) \cdot \underline{R}(t) \rangle / \langle R^2 \rangle. \quad (21)$$

Now it is clear that the application of Eq. (24) is somewhat difficult, since the integrand is known with limited accuracy for large times, and an extrapolation is needed in order that one can carry out the time integration, up to infinite time. Much care is needed in doing this and the analysis of errors involved in this procedure is not as straightforward as the error analysis of purely static averages [5,13].

The huge advantage of the approach based on the Kramers potential, however, is that we are able to extract a relaxation time τ from the computation of a purely static average. Defining the coordinate of the i th monomer of a chain in its center of gravity system as (x_i, y_i, z_i) , we compute the following components of the gyration radius square of the chain ($g = \zeta \gamma / 2k_B T$):

$$\langle X_{\text{gyr}}^2 \rangle_g \equiv \sum_{\text{conf}} \frac{1}{N} \sum_{i=1}^N x_i^2 e^{g \Sigma r_i^2} / \sum_{\text{conf}} e^{g \Sigma r_i^2}, \quad (22)$$

$$\langle Y_{\text{gyr}}^2 \rangle_g \equiv \sum_{\text{conf}} \frac{1}{N} \sum_{i=1}^N y_i^2 e^{g \sum_j x_j y_j} / \sum_{\text{conf}} e^{g \sum_j x_j y_j}, \quad (23)$$

$$\langle Z_{\text{gyr}}^2 \rangle_g \equiv \sum_{\text{conf}} \frac{1}{N} \sum_{i=1}^N z_i^2 e^{g \sum_j x_j y_j} / \sum_{\text{conf}} e^{g \sum_j x_j y_j}. \quad (24)$$

The sampling of the configurations over which the average in Eqs. [22]-[24] is performed can now be done with any method which is convenient. In practice, we have performed it by a dynamic method (including end bond rotation, kink jump motion, and 90° crankshaft rotation on the simple cubic lattice, for details see Ref. [1], since we wish to extend this approach to many-chain systems. For the study of these properties of isolated chains [Eq. (25)] much more efficient simulation techniques [5] can be employed if one wishes to perform a high precision study.

From Eq. [22], a time τ can be defined as follows:

$$\begin{aligned} \tau^2 &\equiv \lim_{g \rightarrow 0} [\langle X_{\text{gyr}}^2 \rangle_g / \langle X_{\text{gyr}}^2 \rangle_0 - 1] / g^2 \\ &= \lim_{g \rightarrow 0} [\langle Y_{\text{gyr}}^2 \rangle_g / \langle Y_{\text{gyr}}^2 \rangle_0 - 1] / g^2. \end{aligned} \quad (25)$$

In practice, we evaluate in a single simulation run the averages defined in Eq. [22] for a large array of values for g , chosen such that the order of magnitude of the weighting factor $\exp[g \sum_j x_j y_j]$ for the biased sampling is of order unity for typical chain configurations. Thus we consider only the case of a fairly weak bias, and thus this does not introduce the sort of accuracy problems that occur for strongly biased sampling methods [5].

Performing between $n = 150,000$ attempted moves ($N = 8$) and $n = 3 \times 10^6$ attempted moves ($N = 150$), we obtain the data shown in Fig. 1. Note that $\langle X_{\text{gyr}}^2 \rangle$ plotted there is actually $(\langle X_{\text{gyr}}^2 \rangle_g + \langle Y_{\text{gyr}}^2 \rangle_g) / 2$; the difference between $\langle X_{\text{gyr}}^2 \rangle_g$ and $\langle Y_{\text{gyr}}^2 \rangle_g$ is almost of the order of the size of the symbols, and is representative of the statistical error. The data are consistent with the limiting behavior for small g implied by Eq. (25), while for larger values of g^2 deviations from this behavior can be recognized. But from such data it is straightforward to identify the slope at $g^2 \rightarrow 0$ needed to obtain τ . Figures 2 and 3 show both $\langle Y_{\text{gyr}}^2 \rangle_0$ and τ on log-log plots vs. N to demonstrate the expected power law behaviors. While systematic deviations occur for very small N (using data for too small N one would yield enhanced "effective exponents" due to corrections to scaling [5,12], for $N > 32$ the numerical data are consistent with the expected behavior: $\langle X_{\text{gyr}}^2 \rangle_0$ should vary as $N^{-1.18}$ [Eq. (17)], τ should behave as $N^{-2.18}$ [Eq. (19)]. Within the accuracy of our exponent estimates (which is of the order of a few percent) we find exactly those values. Note that in

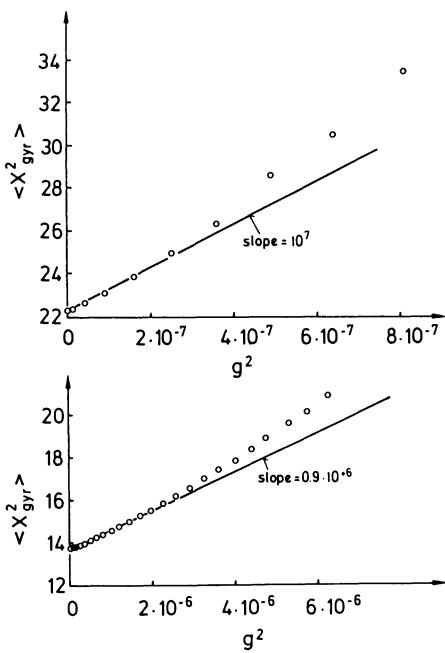


Fig. 1 Plot of $\langle X_{gyr}^2 \rangle$ vs g^2 for $N = 150$ (upper part) and $N = 100$ (lower part), for self-avoiding walks on the simple cubic lattice.

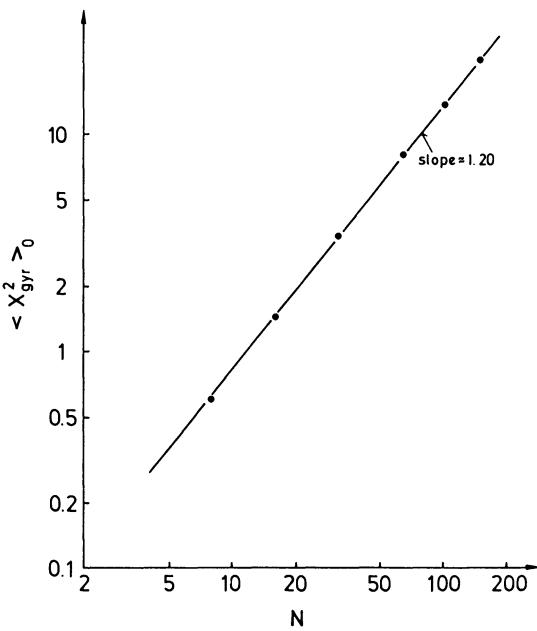
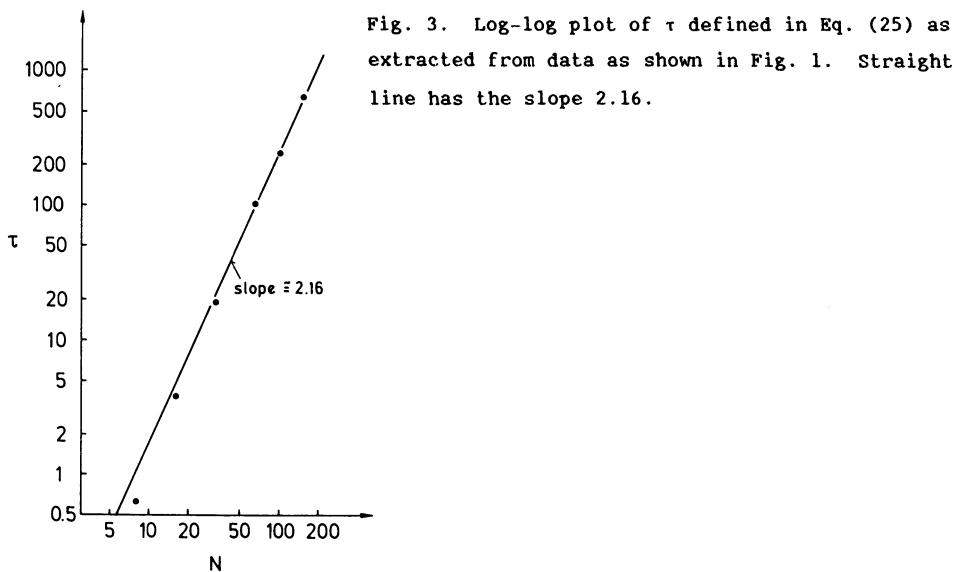


Fig. 2. Log-log plot of $\langle X_{gyr}^2 \rangle_0$ (gyration radius square component in the absence of the Kramers potential) plotted vs N . The straight line indicates a variation of $\langle X_{gyr}^2 \rangle \propto N^{2\nu}$ with $\nu \approx 0.6$.



this approach it is not difficult to estimate τ and $\langle X_{gyr}^2 \rangle_0$ with the same relative accuracy.

Eq. (19) can also be justified by the present approach in terms of the Kramers potential: since Eqs. (22) (26) relate the time τ to purely static averages. We can make this point completely obvious by a Taylor expansion of Eq. (22) in powers of g , which yields

$$\begin{aligned} \tau^2 &= \langle X_{gyr}^2 \rangle_0^{-1} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \langle x_i^2 x_j x_k y_j y_k \rangle_0 \\ &\quad - \sum_{j=1}^N \sum_{k=1}^N \langle x_j x_k y_j y_k \rangle_0. \end{aligned} \quad (26)$$

Since typical distances from the center of gravity x_i, y_j are of order N^ν , and there are of order N^2 terms in the sums, we estimate $\tau^2 \propto N^{-2\nu} N^{2+6\nu} = N^{2+4\nu}$ which is Eq. (19).

For further extensions and generalizations we refer our readers to Ref. 1.

3. Approximate Representation of Functions of Several Variables in Terms of Functions of One Variable

In many physical contexts one deals with functions of several variables ranging from very many in the statistical mechanics of many body problems or conformational analysis of polyatomic molecules to two in the coding of satellite image data. This multiplicity of variables obscures access to

practical solutions of such problems. We have recently shown how to effectively, numerically approximate continuous functions of several variables by finite superpositions of continuous functions of one variable. We outline briefly here the method and how it can be used to simplify besides the representation, the integration of functions of several variables or aid in the process of finding global maxima or minima of functions of several variables.

Our starting point is the Arnol'd [14] and Kolmogorov's [15] solution of Hilbert's 13th [16] problem. We employ a version of the proofs presented by Lorentz [17] of the basic theorem: given a continuous function of s variables $f = f(x_1, \dots, x_s)$ defined on the unit hypercube

$$f \in C[\underline{R}^s], f : [0,1]^s \rightarrow \underline{R}$$

there exists a continuous function of a single argument $g(z) : [0,s] \rightarrow \underline{R}$ such that

$$f(x_1, \dots, x_s) = \sum_{q=0}^{2s} g(z_q) \quad (27)$$

where the argument z_q is an embedding of the originally bounded multi-dimensional domain onto a one-dimensional bounded interval $z_q : [0,1]^s \rightarrow [0,1]$ defined by

$$z_q = \sum_{j=1}^s \lambda_j \phi_q(x_j). \quad (28)$$

In Eq. (28) λ_j are rationally independent numbers and the $\phi_q(x)$ are monotonically increasing continuous functions (defined on special self-scaled, multiply refined partitions of the unit interval) which are tilted versions of functions introduced by Natanson [18] which can be iteratively constructed on a computer. The whole convergent process consists of the iteration of Eq. (27) on the error committed at the previous step in the manner explicitly described in Ref. [17].

The principal observation made in Ref. [14] is that at least for functions of two variables ($s=2$), a finite iteration of this process consisting of three refinements of the partition (the stored $\phi_q(x)$) and a few iterations of the g 's in Eq. (27) (from one to five), results in a surprisingly good approximation of a number of smooth target functions chosen as solutions of differential equations. The criteria of goodness used are: (1) the overall geometrical appearance and (2) the relative norm of the error, $\Theta = (\text{norm of}$

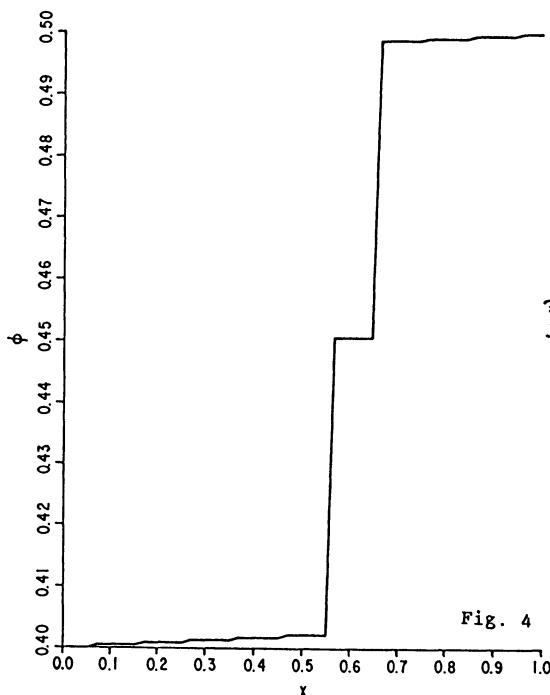


Fig. 4

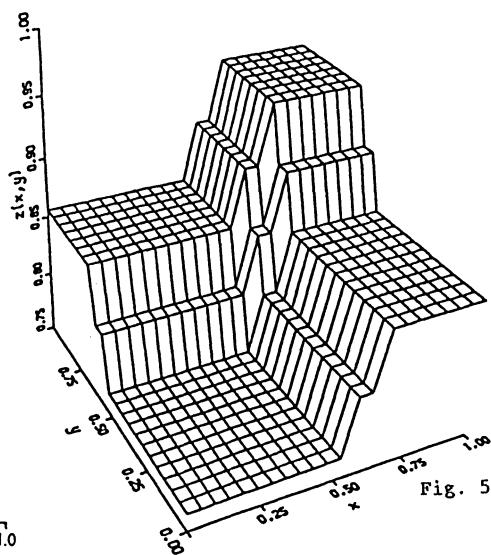


Fig. 5

Fig. 4. The tilted Natanson construction for $q=2$ for the second refinement of the partition of the unit interval, $\phi_2(x)$ vs x .

Fig. 5. Embedding of the unit square given by $z_2 = \phi_2(x) + \lambda\phi_2(y)$ vs x and y with $\phi_2(x)$ as given in Figure 4 and $\lambda = 0.912354278987$.

the approximation)/(norm of the original function) which is Kolmogorov's convergence criterion [15]. The refinements of the $\phi_q(x)$ are tedious and time consuming but once done can be stored at each rank of refinement (they take about 5 hours on our Vax 8650). Figure 4 shows $\phi_2(x)$ for the second refinement of the partition. Figure 5 shows the corresponding $z_2(x,y) = \phi_2(x) + x\lambda\phi_2(y)$; $x_1 = x$, $x_2 = y$, $\lambda_1 = 1$, $\lambda_2 = \lambda$ an irrational number approximated for our purposes by $\lambda = 0.912354378987$. We store the λ and ϕ 's up to the third refinement. Since we will ultimately carry out integrations we store also the Jacobians $J(x_1, x_2/z_q)$ which are given in sufficient approximation by a volume, $v^{(2)}$, of a partition of the domain defined by Eq. (28) (for $s = 2$) conveniently enlarged to fill the entire domain with them which introduces only an error of a few percent in the integrals. Thus denoting by z_q^k the image of the small k^{th} hypercube of the refinement of a partition given by Eq. (28), we can write

$$\int_{x_1^1}^{x_1^2} \int_{x_2^1}^{x_2^2} f(x_1, x_2) dx_1 dx_2 = \int_{x_1^1}^{x_1^2} \int_{x_2^1}^{x_2^2} \sum_{q=0}^4 g(z_q(x_1, x_2)) dx_1 dx_2 \quad (29)$$

$$= \sum_{q=0}^4 \int_{z_q^1}^{z_q^2} g(z_q) J(x_1, x_2 / z_q) dz_q \approx v^{(2)} \sum_{q=0}^4 \sum_k g(z_q^k)$$

where $0 \leq x_i^1 < x_i^2 \leq 1$ ($i = 1, 2$) and $0 \leq z_q^1 < z_q^2 \leq 2$.

We will only give one illustration of this procedure using ϕ 's obtained up to a third refinement. The target function is

$$f(x_1, y) = \sigma^{-1} \exp[-\{(x-x_0)^2 + (y-y_0)^2\}/\sigma^2] \sin(ax+by) \quad (30)$$

with $a = 7.0$, $b = 15.0$, $\sigma = 0.2$, $x_0 = y_0 = 0.5$ shown in Figure 6.

Figure 7 shows $g(z)$ (which is useful for encoding image data) obtained from the first iteration of Eq. (27). This function enables us very easily to locate the global minima of the function among multiple minima, which is relevant to the configurational analysis of polyatomic molecules. The shape of $g(z)$ will be similar to that depicted in Figure 7 independently of the number of original variables (the number of cumulations will be $2 s + 1$ for s variables). One finds the global minimum of $g(z)$ (up to the error of the order of the refinement of the partition) at each cumulation, z_q^{\min} say, and obtains the image via the inverse of Eq. (28). The location of the minimum given in Table 2 is obtained from the average of $z_0^{\min}, \dots, z_q^{\min}$

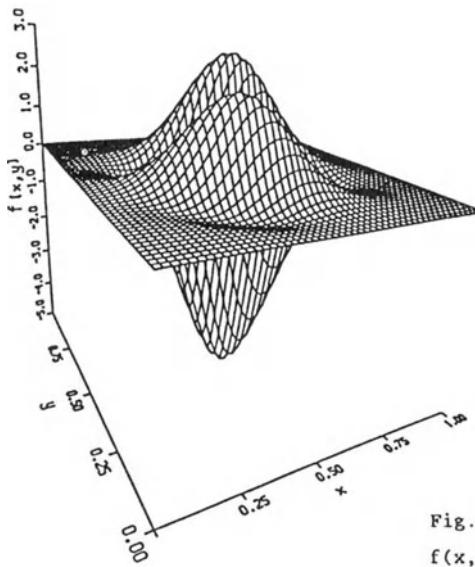


Fig. 6. The target function given by Eq. (30), $f(x, y)$ vs x and y .

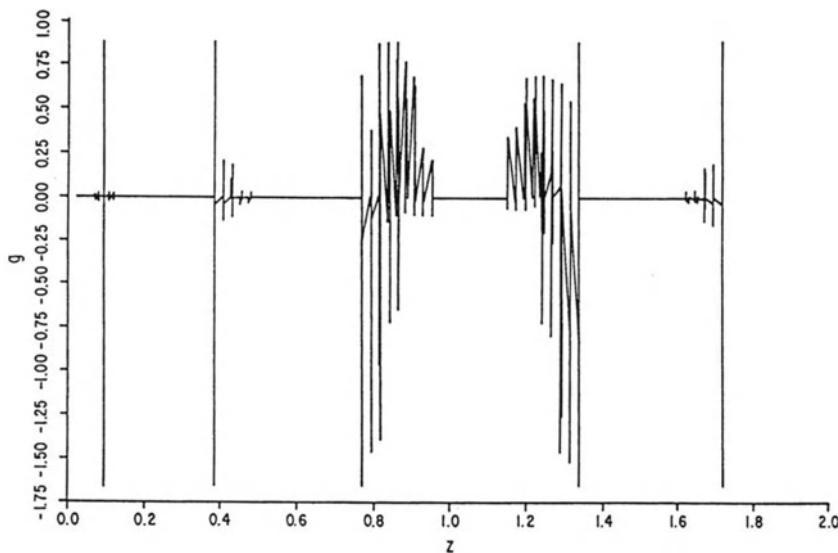


Fig. 7. The approximating function $g(z)$ vs z of the target function described by Eq. (30) obtained from the first iteration of Eq. (27).

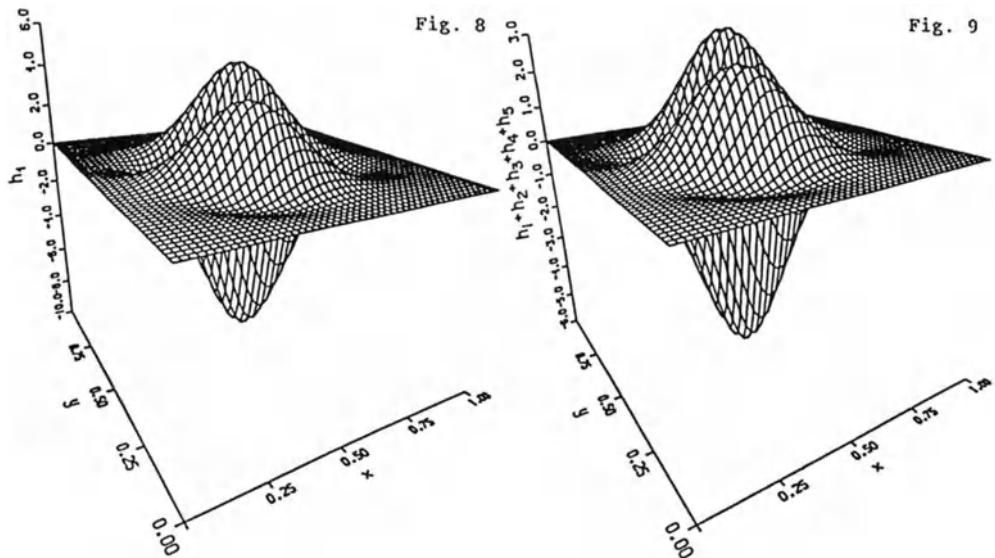


Fig. 8. The first iteration h_1 vs x, y of the target function shown in Figure 6.

Fig. 9. The fifth iteration $h_1 + h_2 + h_3 + h_4 + h_5$ vs x, y of the target function shown in Figure 6.

and this is compared to the exact value. Figure 6, 8 and 9 are the original target function given by Eq. (30), its geometrical reconstruction after one iteration h_1 and its geometrical reconstruction after five iterations,

Table 2. The relative norm, location of minimum and the integral of the functions shown in Figures 6, 8 and 9.

Function	θ	location of minimum	relative	integral	relative
			error	in minimum	in integral
			in minimum		
Original Eq. (3)	-	x_{\min} 0.501 y_{\min} 0.501	-	-0.036	-
First iteration, h_1	0.8633	-	-	-	-
Fifth iteration	0.1285	$x_{\min} = 0.495$ $y_{\min} = 0.491$	0.009 0.019	-0.033	0.08
$h_1 + h_2 + h_3 + h_4 + h_5$					

$h_1 + h_2 + h_3 + h_4 + h_5$, respectively. The time taken to produce these figures together with finding the minimum and the integral over the unit interval $0 \leq x, y \leq 1$, together with the relative norm θ , all shown in Table 2, was a couple of minutes of cpu time on the Vax 8650. Similar results were found with other target functions [2].

4. Concluding Remarks

The assessment of the usefulness of Monte Carlo simulations to obtain dynamical relaxation times awaits further well worked examples. Similarly the representation of functions of several variables as finite superpositions of continuous functions of one variable is currently under investigation for more than two variables. Clearly this calls for a more extended computer capacity but we do not foresee any difficulties in realizing the procedures outlined in Section 3. Since we are not interested in a proof of convergence of such a representation but only in the numerical accuracy of the asymptotic, iterated approximation one could attempt to replace the continuous functions of one variable by a suitably chosen, "optimal", set of discontinuous functions. All this awaits further investigations.

5. Acknowledgement

This work was supported by the National Science Foundation Grant DMR 8515519 and the US Army Research Office Grant DAAL 0388KO198.

References

1. H.L. Frisch, N. Pistoor, A. Sariban, K. Binder and S. Fesjian, J. Chem. Phys. 89, 5194 (1988).
2. H.A. Kramers, J. Chem. Phys. 14, 415 (1946).
3. S. Fesjian and H.L. Frisch, J. Chem. Phys. 80, 4410 (1984).
4. L. Onsager, Ann. N.Y. Acad. Sci. 51, 627 (1949).
5. For a general review of the Monte Carlo technique for the simulation of lattice models for macromolecules see A. Baumgärtner, in Applications of the Monte Carlo Method in Statistical Physics, edited by K. Binder (Springer, Berlin, 1984); K. Kremer and K. Binder, Comput. Phys. Rep. 7, 259 (1988).
6. R.B. Bird, C.F. Curtiss, R.C. Armstrong, and O. Hassager Dynamics of Polymeric Liquids, 2nd ed. (Wiley, New York, 1987), Vol. 2, Chap. 15.
7. H.L. Frisch, C. Borzi, G. Ord, J.K. Percus and G.O. Williams, "Approximate Representation of Functions of Several Variables", unpublished manuscript, submitted to Phys. Rev.
8. Reference 6, Chap. 11.
9. N. Pistoor, Diplomarbeit, Johannes-Gutenberg-Universität Mainz (unpublished); see also N. Pistoor and K. Binder, Colloid Polym. Sci. 266, 132 (1988).
10. J.C. le Guillou and J. Zinn-Justin, Phys. Rev. B 21, 3976 (1980).
11. N. Madras and A. Sokal, J. Stat. Phys. 50, 109 (1988).
12. P.G. de Gennes, Scaling Concepts in Polymer Physics (Cornell University, Ithaca, 1979).
13. A. Berretti and A.D. Sokal, J. Stat. Phys. 40, 483 (1985).
14. V.I. Arnol'd, Dokl. Akad. Nauk SSR, 108, 179 (1956).
15. A. Kolmogorov, Dokl. Akad. Nauk SSR, 114, 679 (1957).
16. D. Hilbert, Bull. Amer. Math. Soc. 8, 461 (1902).
17. G.G. Lorentz, "Representations of Functions", Chelsea Publishing Co., New York, 1986.
18. I.P. Natanson, "Theory of Functions of a Real Variable", Vol. 1, Ungar, New York, 1955.
19. K.D. Gibson and H.A. Scheraga, in "Structure and Expression", Vol. 1, edited by M. Sarma and R. Sarma, Adenine Press, Guilderland, N.Y., 1988.

Part II

Quantum Systems

Monte Carlo Studies of Many-Electron Systems

R.L. Sugar

Department of Physics, University of California, Santa Barbara, CA 93106, USA

Abstract. The Monte Carlo simulation of strongly correlated electron systems is described. Emphasis is placed on the study of two-dimensional systems at low temperature, and results are presented for the two-dimensional Hubbard model both on and off half-filling.

1. Introduction

Many problems of interest in condensed matter physics involve systems of strongly interacting electrons which move on regular lattices. At present the most prominent of these problems is high temperature superconductivity, but one should also include the heavy fermion problem, the formation of magnetic order and metal-insulator transitions. Because the interactions of the electrons in these problems are strong, perturbative techniques are generally not applicable, and numerical simulations have the potential to play an important role.

The simulation of many fermion systems is notoriously difficult because of the antisymmetry of the fermion wave function. This problem is of great importance in high energy, nuclear, and atomic physics, as well as in condensed matter physics, and in recent years a number of promising algorithms have been developed.^{1,2,3,4,5} In this paper I will describe one of these algorithms, which is particularly suited for the simulation of two-dimensional fermion systems at low temperatures.^{2,6} This is an important class of problems in condensed matter physics at the present time because the high-temperature superconductors are quasi-two-dimensional, and in the models currently being studied superconductivity, if it occurs at all, does so at temperature scales which are considerably smaller than those set by the input parameters.

I shall present the algorithm and show numerical results for the two-dimensional Hubbard model, although the same techniques can be used to study a variety of other two-dimensional models.

Although the Hubbard model is very simple in structure, it has been used to study phenomena as diverse as magnetism, superconductivity, and metal insulator transitions. Many physicists working on the high temperature superconductors have suggested that the two-dimensional Hubbard model contains the basic physics needed to understand them. The Hubbard model has been studied extensively for over twenty-five years with techniques including mean field theory, diagram summation, variational calculations and most recently numerical simulations. Despite very substantial efforts, questions as fundamental as the nature of the phase diagram remain open.

The Hamiltonian for the Hubbard model is

$$H = -t \sum_{\langle ij \rangle} \sum_{\sigma} (c_{i\sigma}^{\dagger} c_{j\sigma} + c_{j\sigma}^{\dagger} c_{i\sigma}) + U \sum_i (n_{i+} - \frac{1}{2})(n_{i-} - \frac{1}{2}) - \mu \sum_i (n_{i+} + n_{i-}). \quad (1)$$

Here $c_{i\sigma}^{\dagger}$ and $c_{i\sigma}$ are creation and annihilation operators for an electron on the i^{th} lattice site with z-component of spin σ , and $n_{i\sigma} = c_{i\sigma}^{\dagger} c_{i\sigma}$. The first term in Eq. (1) represents the kinetic energy of the electrons. The symbol $\langle ij \rangle$ indicates that the summation is over all nearest neighbor lattice sites. The hopping parameter, t , sets the scale for the kinetic energy and for the bandwidth. The second term in Eq. (1) is an onsite coulomb interaction. The factors of $\frac{1}{2}n_{i\sigma}$ are included to make the Hamiltonian invariant under a particle-hole transformation in the absence of a chemical potential ($\mu = 0$). This insures that for $\mu = 0$ the band is half-filled, in other words there is on average one electron per lattice site. The last term in Eq. (1) is the chemical potential. μ can be adjusted to obtain any band filling. This model can be generalized to include non-nearest neighbor hopping terms, off-site coulomb interactions, more than one band of electrons, and phonon interactions. However, even in its simplest form the Hubbard model is rich in physics and presents a major calculation problem.

In the path integral formulation of field theory, fermions are represented by anti-commuting c-numbers, elements of a Grassmann algebra. Thus, fermion path integrals cannot be evaluated

directly by numerical techniques. One must integrate out the fermion degrees of freedom before beginning the simulation. One can only do this if the Hamiltonian is quadratic in the creation and annihilation operators; however, the coulomb interaction is quartic in these operators. For the Hubbard model, and many other condensed matter models, it is possible to reduce the Hamiltonian to quadratic form by introducing an auxiliary field, that is, by making a Hubbard-Stratonovich transformation. Integration of the fermion degrees of freedom then leaves a highly nonlocal interaction among the Hubbard-Stratonovich variables involving the famous fermion determinant. This is a problem that can be dealt with by importance sampling techniques. In Section II I describe the Hubbard-Stratonovich transformation for the Hubbard model, and discuss the structure of the fermion determinant. In Section III I describe an algorithm for the simulation of the Hubbard model with special attention to the difficulties encountered at low temperatures.^{2,6} Finally, in Section IV I present some recent numerical results, and discuss prospects for future work.

2. The Structure of the Fermion Determinant

We are interested in performing calculations at finite temperature, so the expectation value of a physics observable, O , is given by

$$\langle O \rangle = \frac{\text{Tr} (e^{-\beta H} O)}{\text{Tr} e^{-\beta H}}. \quad (2)$$

Here H is the Hamiltonian, which for the sake of definiteness I will take to be that of Eq. (1), and β is the inverse temperature.

Just as in the derivation of path integrals we divide the interval $0 \leq \tau \leq \beta$ into L steps of width $\Delta\tau$. The variable τ will be referred to as the imaginary time because of the analogy between the operator $\exp(-\beta H)$ and the quantum mechanical time evolution operator $\exp(-itH)$. Each interval $\Delta\tau$ will be referred to as an imaginary time slice. Making use of the Trotter approximation one can write

$$\text{Tr} e^{-\beta H} = \text{Tr} \left(e^{-\Delta\tau H} \right)^L \simeq \text{Tr} \left(e^{-\Delta\tau V} e^{-\Delta\tau K} \right)^L, \quad (3)$$

where the kinetic energy operator, K , and the potential energy operator, V , are given by

$$K = -t \sum_{\langle ij \rangle} \sum_{\sigma} (c_{i\sigma}^\dagger c_{j\sigma} + c_{j\sigma}^\dagger c_{i\sigma}) \quad (4)$$

$$V = +U \sum_i (n_{i+} - \frac{1}{2})(n_{i-} - \frac{1}{2}) - \mu \sum_i (n_{i+} + n_{i-}). \quad (5)$$

The Trotter approximation gives rise to errors in physical quantities of order $\Delta\tau^2$. These are the only systematic errors in this approach. One must either perform calculations at several values of $\Delta\tau$ and then extrapolate to the limit $\Delta\tau = 0$, or work at sufficiently small values of $\Delta\tau$ so that the systematic errors are smaller than the statistical ones that arise in the simulation from finite sample sizes. In practice the $\Delta\tau$ errors do not cause a significant difficulty.

The purpose of employing the Trotter approximation is to isolate the potential energy, which is a sum of commuting operators. In order to reduce the Hamiltonian to a quadratic form in the electron creation and annihilation operators we make use of Hirsch's discrete Hubbard-Stratonovich transformation⁷

$$e^{-\Delta\tau U(n_{i+} - \frac{1}{2})(n_{i-} - \frac{1}{2})} = e^{-\Delta\tau U/4} \frac{1}{2} \sum_{s_{i,\ell}=\pm 1} e^{-\Delta\tau s_{i,\ell} \lambda(n_{i+} - n_{i-})} \quad (6)$$

at each lattice point i and imaginary time slice ℓ . Here $\cosh(\Delta\tau\lambda) = \exp(\Delta\tau U/2)$. Then undoing the Trotter breakup we have on each time slice

$$H(\ell) = \sum_{\langle ij \rangle} \sum_{\sigma} c_{i\sigma}^\dagger h_{i,j}^\sigma(\ell) c_{j\sigma} = \sum_{\sigma} H_{\sigma}(\ell) \quad (7)$$

with

$$h^{\sigma}(\ell) = k + v^{\sigma}(\ell) \quad (8)$$

and

$$k_{i,j} = \begin{cases} -t & \text{if } i \text{ and } j \text{ are nearest neighbors} \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

$$v^{\sigma}(\ell)_{i,i} = \delta_{i,i} [\sigma \lambda s_{i,\ell} - \mu] \quad (10)$$

The partition function can now be written as

$$Z = \text{Tr} e^{-\beta H} = \sum_{s_{i,\ell}=\pm 1} z_+(s) z_-(s). \quad (11)$$

with

$$z_\sigma = \text{Tr} \left\{ T \prod_\ell e^{-\Delta\tau H_\sigma(\ell)} \right\}. \quad (12)$$

The symbol T signifies an imaginary time ordered product over the factors $\exp[-\Delta\tau H_\sigma(\ell)]$. z_σ is clearly the partition function for electrons with z-component of spin σ in the external field of the Hubbard-Stratonovich spins $s_{i,\ell}$.

If the single particle Hamiltonian $h^\sigma(\ell)$ were independent of ℓ , then the evaluation of $z_\sigma(\ell)$ would be trivial. Denoting the k^{th} eigenvalue of h^σ by $\epsilon_{k,\sigma}$, we could then write

$$z_\sigma(s) = \prod_k (1 + e^{-\beta\epsilon_{k,\sigma}}) = \det [I + e^{-\beta h^\sigma}]. \quad (13)$$

Unfortunately the $h^\sigma(\ell)$ do depend on ℓ since the spin variables on each time slice are independent. The generalization of Eq. (13) is obvious although the algebraic proof is somewhat lengthy²

$$\begin{aligned} z_\sigma(s) &= \det M^\sigma = \det \left[1 + T \prod_\ell e^{-\Delta\tau h^\sigma(\ell)} \right] \\ &= \det [1 + B_L^\sigma B_{L-1}^\sigma \dots B_1^\sigma]. \end{aligned} \quad (14)$$

The B_ℓ^σ are the single particle imaginary time evolution operators for a single time slice.

$$B_\ell^\sigma = e^{-\Delta\tau h^\sigma(\ell)}. \quad (15)$$

It should be evident that the structure of Eqs. (11), (14) and (15) is independent of the specific model we happen to be studying. As long as the model can be reduced to a quadratic form in the electron creation and annihilation operators by the introduction of one or more auxiliary fields, we always obtain these equations, albeit with different expressions for $h^\sigma(\ell)$. These equations contain all of the information concerning Fermi statistics. For future reference it should be noted that $\det M^\sigma$ is invariant under a cyclic permutation of the matrices B_ℓ^σ :

$$\det M^\sigma = \det [I + B_\ell^\sigma \dots B_1^\sigma B_L^\sigma \dots B_{\ell+1}^\sigma] = \det [I + A^\sigma(\ell)]. \quad (16)$$

I have concentrated on the partition function, but the same set of manipulations must be applied to the trace in the numerator of Eq. (2). It can be written in the form²

$$\text{Tr} \left[e^{-\beta H} O \right] = \sum_{s_i, \ell = \pm 1} \det M^+(s) \det M^-(s) \tilde{O}(s), \quad (17)$$

where $\tilde{O}(s)$ can generally be expressed in terms of matrix elements of the finite temperature Green's functions for a single electron in the presence of the spin field, s . The equal time Green's functions are merely $[I + A(\ell)]^{-1}$, and the unequal-time Green's functions are obtained by multiplying these by appropriate products of the B_ℓ^σ .²

3. An Exact Updating Algorithm

We have seen in the last section that the expectation value of a physical observable, O , can be written in the form

$$\langle O \rangle = \frac{\sum_{s_i, \ell = \pm 1} \det M^+(s) \det M^-(s) \tilde{O}(s)}{\sum_{s'_i, \ell = \pm 1} \det M^+(s') \det M^-(s')} \quad (18)$$

So, if we generate a set of spin configurations with a probability distribution

$$P(s) = \frac{\det M^+(s) \det M^-(s)}{\sum_{s'_i, \ell = \pm 1} \det M^+(s') \det M^-(s')}, \quad (19)$$

then a measurement of $\tilde{O}(s)$ in this distribution will yield the desired expectation value. This presupposes that the product of the fermion determinants is positive definite, which is the case for the Hubbard model with a half-filled band.⁷ I will therefore start by considering the half-filled band and discuss the modifications which must be made for other fillings latter. (In this paper I will only consider a repulsive coulomb interaction, $U > 0$. For $U < 0$ the product of fermion determinants is positive definite for all fillings.)

We make use of standard importance sampling techniques to generate the desired spin configurations. We sweep through the lattice bringing each spin in turn into equilibrium with its sur-

roundings by the heatbath algorithm. Suppose that under the spin flip $s_{i,\ell} \rightarrow -s_{i,\ell}$. The fermions matrices undergo the change $M^\sigma \rightarrow M'^\sigma$. Then in order to implement the heatbath algorithm we must calculate the determinant ratio

$$R^\sigma = \frac{\det M'^\sigma}{\det M^\sigma}. \quad (20)$$

If the proposed spin flip is accepted with the probability

$$p(s_{i,\ell} \rightarrow -s_{i,\ell}) = \frac{R^+ R^-}{1 + R^+ R^-}, \quad (21)$$

then detailed balance will be satisfied, and we are assured of obtaining the distribution of Eq. (19). One must, of course, perform a number of lattice sweeps initially to bring the system into equilibrium, and between measurements to insure that independent configurations are generated . The equilibration and correlation times depend on the couplings and the temperature.

In implementing this algorithm one cannot afford to recalculate the fermion determinant from scratch each time a spin is updated. This would take of order N^3 numerical operations, where N is the number of spatial lattice sites. Fortunately, this is unnecessary. We update all spin variables on the same imaginary time slice consecutively. For the ℓ^{th} time slice we make use of Eq. (16), and the Trotter formula to pull the factor of $\exp[-\Delta\tau v^\sigma(\ell)]$ to the left of the matrix $A^\sigma(\ell)$. We then see that under the spin flip $s_{i,\ell} \rightarrow -s_{i,\ell}$

$$A^\sigma \rightarrow A^\sigma(\ell)' = [I + \Delta^\sigma(i, \ell)] A^\sigma(\ell), \quad (22)$$

where the matrix $\Delta^\sigma(i, \ell)$ has only one non-zero element

$$\Delta^\sigma(i, \ell)_{i,i} = (e^{-2\sigma\Delta\tau\lambda s_{i,\ell}} - 1) = \delta^\sigma(i, \ell). \quad (23)$$

Because the matrix $\Delta^\sigma(i, \ell) A^\sigma(\ell)$ has only one non-zero row, the determinant ratio R^σ can be written in the form

$$\begin{aligned} R^\sigma &= \det \left[\frac{I + A^\sigma(\ell)'}{I + A^\sigma(\ell)} \right] \\ &= \det [I + G^\sigma(\ell) \Delta^\sigma(i, \ell) A^\sigma(\ell)] \\ &= 1 + [1 - G^\sigma(\ell)_{i,i}] \delta^\sigma(i, \ell), \end{aligned} \quad (24)$$

where

$$G^\sigma(\ell)_{i,j} = \langle T \left[c_{i,\sigma}(\ell\Delta\tau) c^\dagger_{j,\sigma}(\ell\Delta\tau) \right] \rangle = [I + A^\sigma(\ell)]_{i,j}^{-1} \quad (25)$$

is the equal time, finite temperature Green's function in the presence of the spin field. Thus, if the equal time Green's functions are known, the evaluation of the determinant ratios is trivial. In addition, if the $G^\sigma(\ell)$ are known at one point in the simulation, they can be re-evaluated after each spin flip by again taking advantage of the sparceness of $\Delta^\sigma(i, \ell)$. Under the change $s_{i,\ell} \rightarrow -s_{i,\ell}$, $G^\sigma(\ell) \rightarrow G^\sigma(\ell)'$ with

$$\begin{aligned} G^\sigma(\ell)' &= G^\sigma(\ell) - G^\sigma(\ell)\Delta^\sigma(i, \ell)A^\sigma(\ell)G^\sigma(\ell)' \\ &= G^\sigma(\ell) - \frac{G^\sigma(\ell)\Delta^\sigma(i, \ell)[I - G^\sigma(\ell)]}{[1 + \{1 - G^\sigma(\ell)\}_{i,i}\delta^\sigma(i, \ell)]}. \end{aligned} \quad (26)$$

Thus, the updating of all N^2 elements of $G^\sigma(\ell)$ requires “only” of order N^2 numerical operations.

Once all spins on the ℓ^{th} time slice have been updated, one moves to the $\ell + 1$ time slice using the relation

$$G^\sigma(\ell + 1) = B_{\ell+1}^\sigma G^\sigma(\ell) B_{\ell+1}^{\sigma-1}. \quad (27)$$

One can therefore update all NL spins on the lattice with of order N^3L numerical operations. The rapid growth in numerical operations with spatial lattice size makes this algorithm impractical for three-dimensional systems; however, it is quite feasible to apply it to the two-dimensional models of current interest in condensed matter physics. A number of fermion algorithms such as the pseudo-fermion method,¹ the Langevin equation,³ and hybrid-molecular dynamics^{4,5} have operation counts that grow only linearly with the lattice volume. However, to date it has not been possible to implement these algorithms efficiently for condensed matter models at low temperatures due to the ill-conditioned nature of the fermion determinant.

At the beginning of the simulation it is necessary to calculate the electron Green's functions, $G^\sigma(\ell)$, from scratch using a standard matrix inversion algorithm such as Gaussian elimination.

This requires of order N^3 operations, so one does not want to do it often. The repeated updating of the $G^\sigma(\ell)$ on a single time slice via Eq. (26) does not lead to numerical instabilities, but the movement to a new time slice via Eq. (27) does. Therefore, it is occasionally necessary to recalculate the $G^\sigma(\ell)$ from scratch to prevent buildup of errors. This must be done with increasing frequency as the temperature is lowered. For sufficiently low temperatures the fermion matrices become so ill-conditioned that it is impossible to calculate their inverses by standard techniques. Until quite recently these difficulties prevented simulation of the two-dimensional Hubbard model for values of $t\beta$ much larger than 4.

The origin of the problem can already be seen for the free theory, $U = 0$. For this case the eigenvalues of βh^σ range between $-4t\beta$ and $+4t\beta$. We are interested in performing simulations for $t\beta \gtrsim 10$. It is clear that in this temperature regime one cannot even compute the matrices $A^\sigma(\ell)$ by straightforward multiplication of the B^σ without losing numerical control of the contributions of the eigenvalues which are of order unity, and it is these eigenvalues that play the crucial role in determining the behavior of the $G^\sigma(\ell)$. In order to overcome this difficulty we perform a factorization of the $A^\sigma(\ell)$ which separates the large and small eigenvalues.^{6,8,9,10} We first multiply together a set of m B^σ matrices, where m is not so large that we lose numerical control of the contributions of the small eigenvalues to the product:

$$a_1^\sigma(\ell) = B_{\ell+m}^\sigma B_{\ell+m-1}^\sigma \cdots B_{\ell+1}^\sigma. \quad (28)$$

Typically $mt\Delta\tau \approx 1.5$. We then use the modified Gram-Schmidt orthogonalization procedure to write

$$a_1^\sigma(\ell) = U_1^\sigma D_1^\sigma R_1^\sigma, \quad (29)$$

where U_1^σ is an orthogonal matrix, D_1^σ is a diagonal matrix, and R_1^σ a right triangular matrix. The large variations in the magnitude of matrix elements is isolated in D_1^σ , which is the only one of the three matrices that is ill-conditioned. We now multiply by another set of m B^σ matrices, and perform another Gram-Schmidt orthogonalization

$$\begin{aligned} a_2^\sigma(\ell) &= B_{\ell+2m}^\sigma \cdots B_{\ell+1}^\sigma \\ &= B_{\ell+2m}^\sigma \cdots B_{\ell+m+1}^\sigma U_1^\sigma D_1^\sigma R_1^\sigma \\ &= U_2^\sigma D_2^\sigma R_2^\sigma. \end{aligned} \quad (30)$$

Continuing in this manner we finally obtain

$$A^\sigma(\ell) = U^\sigma(\ell) D^\sigma(\ell) R^\sigma(\ell), \quad (31)$$

where again the diagonal matrix, D^σ is the only ill-conditioned matrix, and contains all the large variation in scale. To evaluate the equal time Green's functions we write

$$\begin{aligned} G^{\sigma -1}(\ell) &= I + A^\sigma(\ell) = I + U^\sigma(\ell) D^\sigma(\ell) R^\sigma(\ell) \\ &= U^\sigma(\ell) \left[D^\sigma(\ell) + U^{\sigma -1}(\ell) R^{\sigma -1}(\ell) \right] R^\sigma(\ell). \end{aligned} \quad (32)$$

$U^\sigma(\ell)$ and $R^\sigma(\ell)$ are, of course, trivial to invert. The matrix in square brackets is strongly diagonal dominated, and can be stably inverted by, for example, Gaussian elimination. Using this simple trick we have been able to eliminate numerical instabilities in the algorithm. For half-filling we are limited in the values of β for which we can carry out simulations only by the availability of computer time. The time required to perform a simulation grows roughly linearly with β .

4. Numerical Results

I will conclude by presenting some results which we have recently obtained with the algorithm described above. I will start with the case of the half-filled band for which the product of fermion determinants is positive definite.

An important question is whether the ground state has long range antiferromagnetic order. Of course a phase transition to an antiferromagnetic state can only occur at zero temperature in two-dimensions. At finite temperature we can obtain a signal for such a transition when the correlation length becomes comparable to the size of the lattice. Operators of particular interest are the spin-spin correlation function

$$\begin{aligned} C(\ell_x, \ell_y) &= \frac{1}{N} \sum_i \langle (c_{i+\ell}^\dagger \sigma_z c_{i+\ell}) (c_i^\dagger \sigma_z c_i) \rangle \\ &= \frac{1}{N} \sum_i \langle (n_{i+\ell\uparrow} - n_{i+\ell\downarrow}) (n_{i\uparrow} - n_{i\downarrow}) \rangle \end{aligned} \quad (33)$$

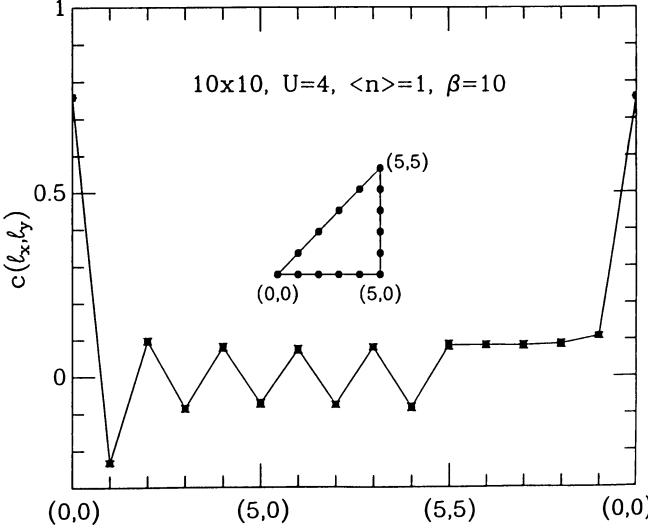


Fig. 1: The spin-spin correlation function $C(\ell_x, \ell_y)$ on a 10×10 lattice at $\beta = 10$ for a half-filled band. The horizontal axis traces out the triangular path shown in the center of the figure. Strong antiferromagnetic correlations are evident.

and its fourier transform

$$S(\vec{q}) = \sum_{\ell} e^{i\vec{q} \cdot \vec{\ell}} C(\ell_x, \ell_y). \quad (34)$$

In Figure 1 I show $C(\ell_x, \ell_y)$ as a function of ℓ for a 10×10 lattice at $\beta = 10$. All results are for $t = 1$ and $U = 4$. As indicated in the insertion, data are shown starting at the point $\ell_x = 0, \ell_y = 0$, first moving in the x direction, then the y direction, and finally returning to the origin along a diagonal, $\ell_x = \ell_y$. The oscillations in $C(\ell_x, \ell_y)$ while moving in the x and y directions and the lack of them while moving along the diagonal are clear indications of the formation of antiferromagnetic order. For comparison Figure 2 shows $C(\ell_x, \ell_y)$ on a 8×8 lattice with $\beta = 10$ for quarter-filling, where there is no antiferromagnetic order.

Spin-wave theory predicts that in the limit that the spatial lattice volume, $N \rightarrow \infty$

$$\frac{\langle S(\pi, \pi) \rangle}{N} = \frac{m^2}{3} + O\left(\frac{1}{\sqrt{N}}\right), \quad (35)$$

where m is the antiferromagnetic order parameter. Similarly

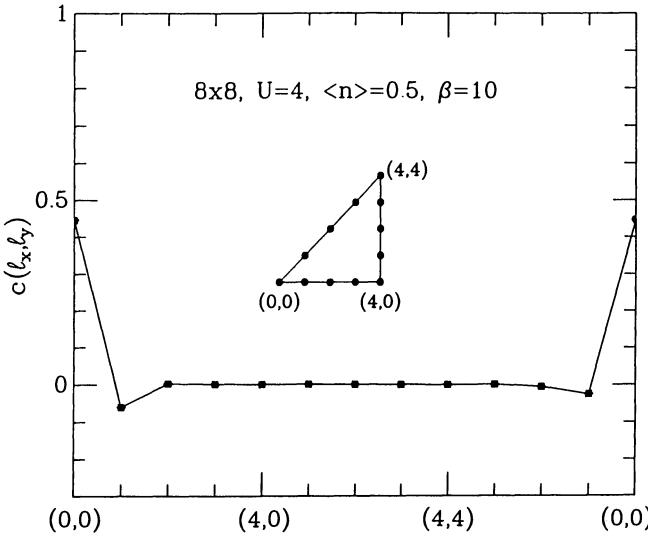


Fig. 2: The spin-spin correlation function $C(\ell_x, \ell_y)$ on an 8×8 lattice at $\beta = 10$ for a quarter-filled band.

$$\langle c\left(\frac{\sqrt{N}}{2}, \frac{\sqrt{N}}{2}\right) \rangle = \frac{m^2}{3} + 0\left(\frac{1}{\sqrt{N}}\right). \quad (36)$$

These quantities are plotted versus $1/N_x = 1/N^{1/2}$ in Figure 3. Our results are in good agreement with earlier work by Hirsch and Tang.¹¹ One can define an effective hopping parameter through the relation

$$\frac{t_{\text{eff}}}{t} = \frac{\langle c_{i\sigma}^\dagger c_{j\sigma} + c_{j\sigma}^\dagger c_{i\sigma} \rangle_U}{\langle c_{i\sigma}^\dagger c_{j\sigma} + c_{j\sigma}^\dagger c_{i\sigma} \rangle_0}. \quad (37)$$

The subscripts U and 0 indicate averages with finite U and with $U = 0$ respectively. As antiferromagnetic order develops one expects t_{eff}/t to decrease. In Figure 4 t_{eff}/t is plotted versus U on a 4×4 lattice at $\beta = 16$. One sees that this expectation is confirmed. The solid curve is from a strong coupling approximation and the dashed curve from a weak coupling one. Note how well the numerical results interpolate between these two approximations.

Away from half-filling the product of the fermion determinants is no longer positive definite, so we cannot use $P(s)$ as defined in Eq. (19) as the probability function for the simulation. Instead we use

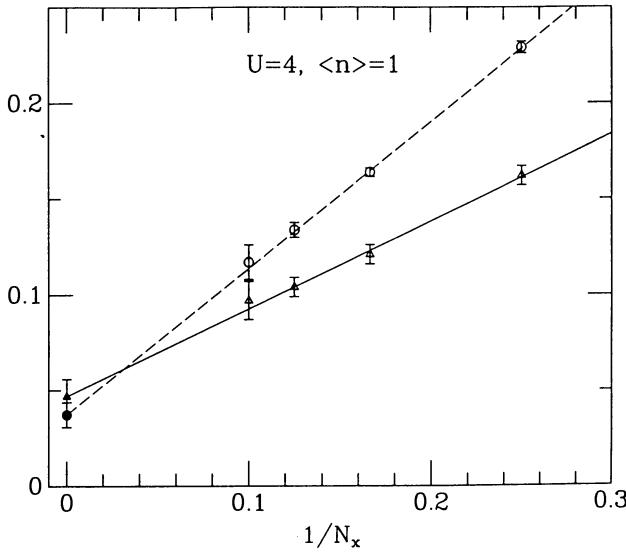


Fig. 3: $S(\pi, \pi)/N$ (circles) and $C(N_x/2, N_x/2)$ (triangles) vs. $1/N_x$. The data was taken from runs at $\beta \geq 16$. The lines are least-squares fits to the data. The extrapolation of these fits to the infinite system are shown as the solid symbols.

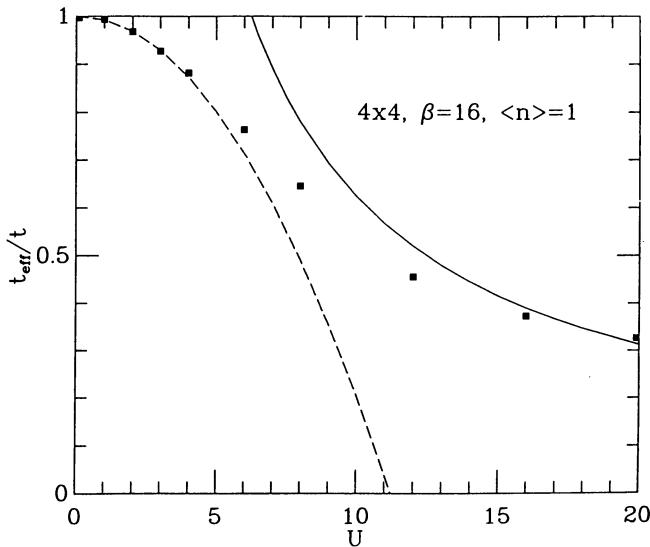


Fig. 4: The effective hopping parameter as a function of U on a 4×4 lattice for $\beta = 16$. The solid line is a strong coupling result. The dashed line is obtained from perturbation theory.

$$\tilde{P}(s) = \frac{|\det M^+(s)\det M^-(s)|}{\sum_{s'_{i,\epsilon}=\pm 1} |\det M^+(s')\det M^-(s')|} \quad (38)$$

for this purpose, and incorporate the sign of the product of the fermion determinants, $S = \det M^+ \det M^- / |\det M^+ \det M^-|$, into the measurements. The expectation value of a physics observable is now given by

$$\langle O \rangle = \frac{\langle OS \rangle_{\tilde{P}}}{\langle S \rangle_{\tilde{P}}}, \quad (39)$$

where the subscript \tilde{P} indicates that the average is to be taken in the probability distribution \tilde{P} .

The approach will be useful if the expectation value of S is not too small. When $\langle S \rangle$ does become small, one is trying to compute the ratio of two small quantities, and the statistical fluctuations become large. In Figure 5 I plot the expectation value of the sign of the determinants as a function of band filling $\langle n \rangle$, on a 4×4 lattice at $\beta = 6$ and 8, and in Figure 6 on 6×6 and 8×8 lattices at $\beta = 6$. Note that $\langle S \rangle$ must be unity at $\langle n \rangle = 0$ and 1. It is difficult to carry out simulations for $\langle S \rangle \leq 0.1$, so there is a

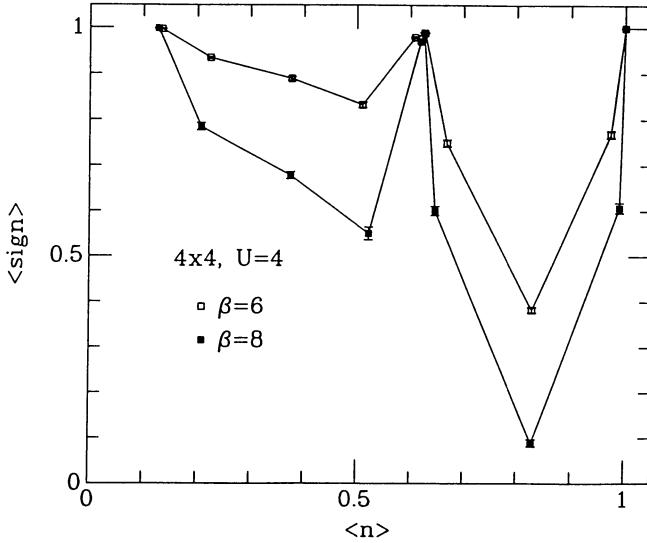


Fig. 5: The average sign of the fermion determinants, $\langle S \rangle$ as a function of band filling on a 4×4 lattice at $\beta = 6$ and 8.

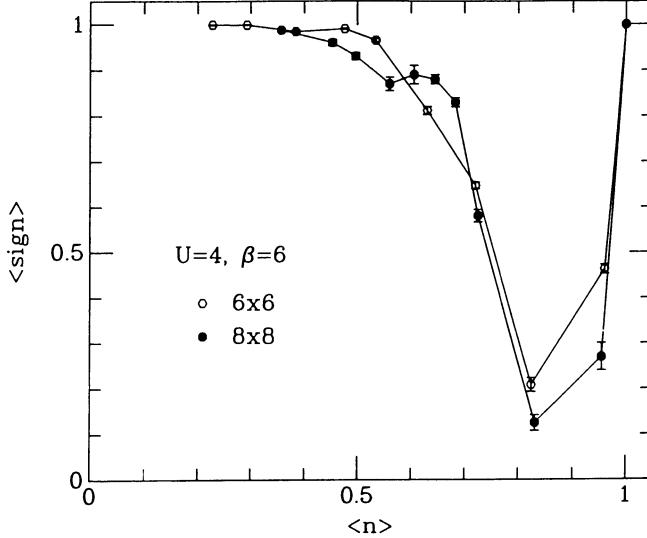


Fig. 6: The average sign of the fermion determinants, $\langle S \rangle$, as a function of band filling on 6×6 and 8×8 lattices at $\beta = 6$.

wide range of band filling for which simulations can be done, but a very interesting range for which they cannot be performed with the present algorithm.

As an example of the type of results one can obtain I plot in Figure 7 the fourier transform of the number density $n(\vec{k})$,

$$n(\vec{k}) = \sum_{\sigma} \langle c_{k\sigma}^{\dagger} c_{k\sigma} \rangle, \quad (40)$$

with $k_y = k_x$ as a function of k_x for a quarter-filled band. The data is from lattices as large as 16×16 . Data from different sized lattices has been combined in order to obtain many k_x values. This is possible because finite sized effects are quite small. The results are consistant with the absence of a gap. For contrast I plot the same quantity for a half-filled band on a 16×16 lattice in Figure 8. The filled circles are Monte Carlo points and the open circles are for free fermions with a gap of $\Delta = 1.3812$. The fit is excellent, giving further evidence for an antiferromagnetic state.

An important question concerning the two-dimensional Hubbard model off half-filling is whether there is a superconducting phase. If such a phase exists, it is expected to occur near half-filling, just in the region that small values of $\langle S \rangle$ make simulations

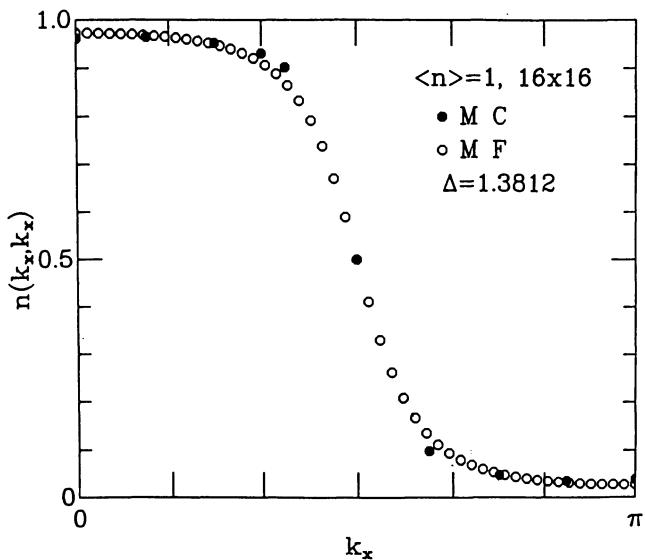


Fig. 7: The momentum distribution $n(\vec{k})$ for a quarter-filled band.

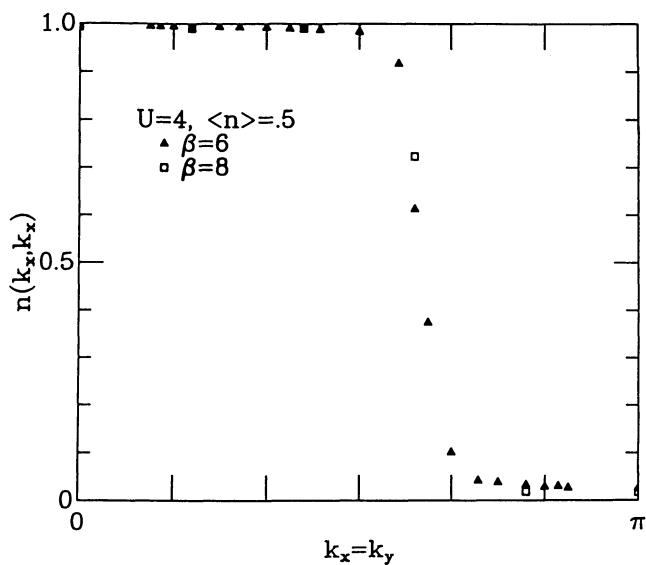


Fig. 8: The momentum distribution $n(\vec{k})$ for a half-filled band.

difficult. Since one is restricted to high temperatures, one can at best hope to see an enhancement in one of the pairing channels. The quantities to be studied are the pair-field susceptibilities

$$P_{\delta',\delta} = \frac{1}{N} \int_0^\beta d\tau \left\langle \Delta_{\delta'}(\tau) \Delta_\delta^\dagger(0) \right\rangle, \quad (41)$$

where

$$\Delta_\delta^\dagger = \sum_\ell c_{\ell+\delta,\uparrow}^\dagger c_{\ell,\downarrow}^\dagger. \quad (42)$$

The index δ runs over the nine relative lattice spacings shown in Figure 9. After integrating out the fermion degrees of freedom the pair-field susceptibilities can be written as averages over products of electron Green's functions

$$P_{\delta',\delta} = \frac{1}{N} \int_0^\beta d\tau \left\langle G_{\uparrow\ell'+\delta',\ell+\delta}(\tau) G_{\downarrow\ell',\ell}(0) \right\rangle_s, \quad (43)$$

where the subscript s indicates an average over spin configurations with due regard for the signs of the determinants. In order to determine whether there is an attractive pairing interaction in a given channel, we compare the $P_{\delta',\delta}$ with the uncorrelated pair-field susceptibilities

$$\bar{P}_{\delta',\delta} = \frac{1}{N} \int_0^\beta d\tau \left\langle G_{\uparrow\ell'+\delta',\ell+\delta}(\tau) \right\rangle_s \left\langle G_{\downarrow\ell',\ell}(0) \right\rangle_s. \quad (44)$$

Since the averages over spin configurations are taken separately for

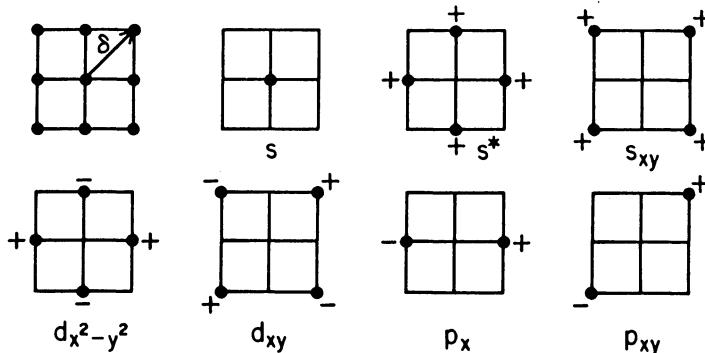


Fig. 9: The internal pair coordinate δ used in defining Δ_δ^\dagger in Eq. (42). Various pair-field modes discussed in the text are shown.

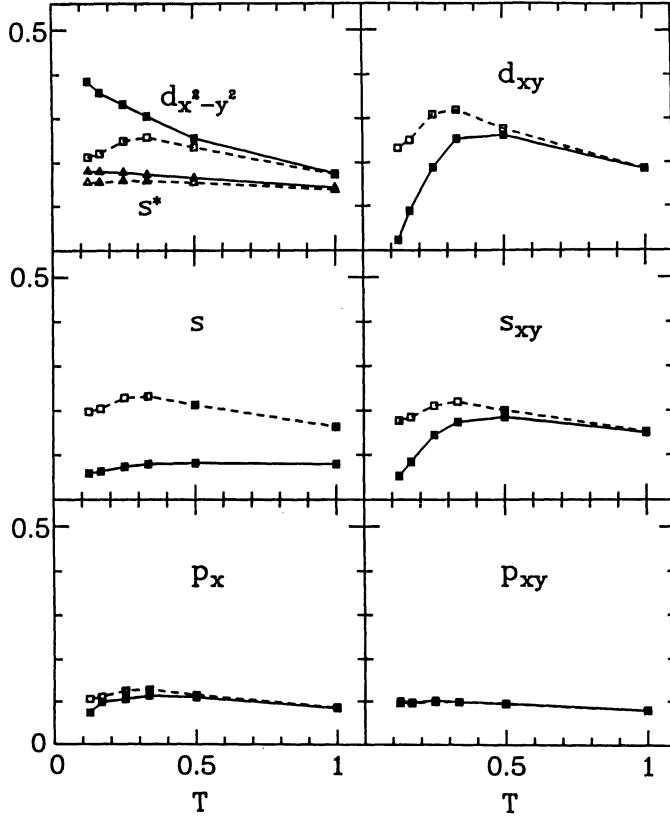


Fig. 10: The pair-field susceptibility, $P_{\delta',\delta}$ and the uncorrelated pair-field susceptibility, $\bar{P}_{\delta',\delta}$ as a function of temperature at half-filling.

the Green's functions in $\bar{P}_{\delta',\delta}$, the electron lines are fully dressed but non-interacting. In Figure 10 I plot P and \bar{P} as a function of T for a variety of pair field modes. This calculation is at half-filling on a 4×4 lattice. Only the $d_{x^2-y^2}$ and s^* modes show an enhancement due to interactions. In Figure 11 the same quantities are shown at a filling of $\langle n \rangle = 0.875$. The magnitude of the d-wave susceptibility increases as the system is doped away from half-filling, although the enhancement relative to \bar{P} is greatest at $\langle n \rangle = 1$. These calculations have been repeated on 6×6 and 8×8 lattices with similar qualitative results. Our results are consistent with a phase diagram for the two-dimensional Hubbard model in which strong antiferromagnetic correlations are suppressed by doping, and *d*-wave superconducting correlations become important at

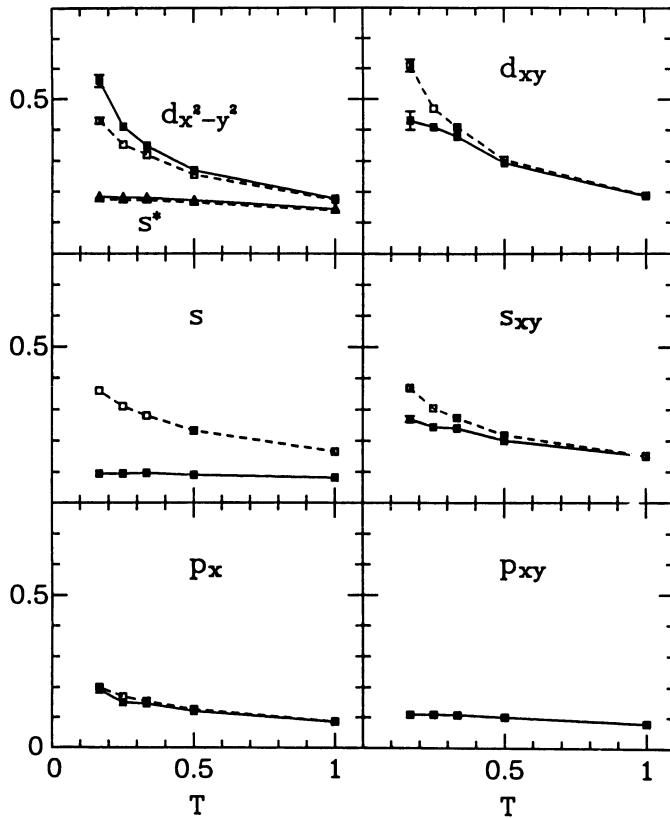


Fig. 11: The pair-field susceptibility, $P_{\delta',\delta}$ and the uncorrelated pair-field susceptibility, $\bar{P}_{\delta',\delta}$ as a function of temperature at quarter-filling.

temperatures significantly lower than those which characterize the half-filled antiferromagnetic correlations. Such a phase diagram would be consistent with recent RPA calculations¹², as well as with weak coupling renormalization group calculations¹³ and recent results from self-consistent perturbation theory.¹⁴ Simulations at significantly lower temperatures will be necessary to confirm this picture, and work is presently in progress.

Acknowledgments

Work described in this paper was carried out in collaboration with R. Blankenbecler, E. Dagotto, J. Gubernatis, E. Loh, A. Moreo,

D. Scalapino, R. Scalettar, D. Toussaint and S. White. This work was supported in part by National Science Foundation Grant PHY-86-14185, Department of Energy Grant DE-FG03-88ER45197, and by the San Diego Supercomputer Center, where all numerical simulations were carried out.

References

1. F. Fucito, E. Marinari, G. Parisi, and C. Rebbi, *Nucl. Phys. B* **180**, 369, (1981).
2. R. Blankenbecler, D.J. Scalapino, and R.L. Sugar, *Phys. Rev. D* **24**, 2278, (1981).
3. G.G. Batrouni, G.R. Katz, A.S. Kronfeld, G.P. Lepage, B. Svetitsky, and K.G. Wilson, *Phys. Rev. D* **32**, 2736, (1985).
4. S. Duane, *Nucl. Phys. B* **257**, 652, (1985); S. Duane and J. Kotug, *Phys. Rev. Lett.* **55**, 2774, (1985); and *Nucl. Phys. B* **275**, 398, (1986); S. Gottlieb, W. Liu, D. Toussaint, R.L. Renken and R.L. Sugar, *Phys. Rev. D* **35**, 2531, (1987).
5. R.T. Scalettar, D.J. Scalapino, R.L. Sugar and D. Toussaint, *Phys. Rev. B* **36**, 8632, (1987).
6. S.R. White, D.J. Scalapino, R.L. Sugar, E.Y. Loh, J.E. Gubernatis, and R.T. Scalettar, UCSB preprint UCSBTH-88-18, to be published in *Phys. Rev. B*.
7. J.E. Hirsch, *Phys. Rev. B* **31**, 4403, (1985).
8. G. Sugiyama and S.E. Koonin, *Annals of Phys.* **168**, 1, (1986).
9. S. Sorella, S. Baroni, R. Car and M. Parrinello, to be published in *Proceedings of the Adriatico Research Conference “Towards the theoretical understanding of the High- T_c Superconductors”*, Modern Physics B, S. Lundquist, E. Tosatti, M. Tosi and L. Yu eds. (World Scientific Press, Singapore, 1988).
10. E.Y. Loh, J.E. Gubernatis, R.T. Scalettar, S.R. White, and R.L. Sugar, to be published in *Workshop on Interacting Electrons in Reduced Dimension*, D. Baeriswyl and D.K. Campbell eds. (Plenum, New York, 1989).

11. J.E. Hirsch and S. Tang *Phys. Rev. Lett.* **62**, 591, (1989).
12. N.E. Bickers, D.J. Scalapino and R.T. Scalettar, *Int. J. Mod. Phys.* **B1**, 687, (1987).
13. I.E. Dzyaloshinskii, *Zh.E.T.F.* **93**, 1487, (1987); I.E. Dzyaloshinski and V.M. Yakovenko, *Zh.E.T.F.* **94**, 344, (1988); H.J.-Schmidt, to be published.
14. N.E. Bickers, D.J. Scalapino and S.R. White, to be published.

Universal Reduction of T_c in Strong Coupling Superconductors due to Anderson/Kondo Impurities

M. Jarrel

The Ohio State University, Columbus, OH 43210, USA

The first exact calculation of $(\partial T_c / \partial c)_{c=0}$, the initial depression of the superconducting T_c , due to a small concentration, c , of magnetic impurities is presented. $N(0)(\partial T_c / \partial c)_{c=0}$ can be expressed as the product of *two* functions $-f(\lambda_o)$, and $g(T_K/T_{co})$. I find that f decreases rapidly and monotonically with λ_o , indicating that magnetic impurities reduce T_c most effectively in weak coupling superconductors. I also find that, in contrast to previous results, the maximum $|(\partial T_c / \partial c)_{c=0}|$ occurs when $T_{co} = T_K$, thus resolving a longstanding experimental puzzle.

It is well known that magnetic impurities have a profound effect upon the properties of superconductors. In the Kondo limit they form a magnetic moment which spin-flip scatters the conduction electrons through an antiferromagnetic coupling. This scattering breaks apart the singlet Cooper pairs which characterize the superconducting state. Thus a small concentration of impurities can severely reduce the superconducting transition temperature [1] and even destroy superconductivity. As the temperature is lowered toward the Kondo temperature, T_K , the scattering rate increases logarithmically. However, as the temperature is lowered further, such that $T \ll T_K$, the conduction electrons screen the impurity moment. Thus one might expect the impurities to inhibit superconductivity most effectively in some intermediate regime.

In addition, the impurities exert a depairing potential on the superconductor which must compete with the phonon-mediated pairing potential. Thus the

magnitude of the inhibition of T_c must depend upon the relative strength of these potentials. Since the pairing potential is proportional to λ_o one might expect the impurities to inhibit most effectively in the weak coupling limit (λ_o small).

Here I present the first exact calculation of $(\partial T_c / \partial c)_{c=0}$, the initial depression of the superconducting transition temperature, T_c , by a small concentration, c , of paramagnetic impurities. The calculations are made with a novel combination of quantum Monte Carlo simulation and Eliashberg-Migdal perturbation theory [2,3]. The Monte Carlo allows one to account for the Kondo effect in an exact, non-perturbative way, hence the exact nature of this calculation. By using an Eliashberg analysis of the Monte Carlo data, I was able to see the effect of magnetic impurities on superconducting hosts with $0.5 > \lambda_o > 3.0$.

It was found that $N(0)(\partial T_c / \partial c)_{c=0} = -f(\lambda_o)g(T_K/T_{co})$, where T_K is the Kondo temperature, and λ_o is the electron-phonon coupling constant. Here, g is a universal function of T_K/T_{co} (see Figs. 1 and 4) and $f = -(\partial T_c / \partial c)_{c=0 \text{ max}}$, the maximum initial depression (see Figs. 3 and 4). Two new features were found in this study. First, the maximum initial depression occurs when $T_K = T_{co}$ in rough agreement with experiment[4], but in extreme quantitative disagreement with previous approximate calculations[5,6,7,8] which placed the maximum between 5 and 12. Second, the magnitude of the depression is strongly dependent upon λ_o and is reduced from its weak coupling limit (λ_o small) for finite λ_o .

In superconductors doped with magnetic impurities $(\partial T_c / \partial c)_{c=0}$ can be measured as a function of T_K/T_{co} if T_K can be changed in a controlled manner. Typically this is done either by the application of pressure[9,4], or by changing the type of impurity[10] or the composition of the superconducting host[11,12]. Interpretation of the data requires a quantitative calculation of $(\partial T_c / \partial c)_{c=0}$. An unequivocal measure of T_K/T_{co} , and hence T_K , is then obtained by fitting this data to a theoretical curve which gives $(\partial T_c / \partial c)_{c=0}$ as a function of T_K/T_{co} .

However, as shown in table 1, due to the breakdown of the previous theories when $T_K/T_{co} \approx 1$, the values of T_K obtained in this way are much larger than those estimated from the resistivity data[4,13,10]. In addition, as shown in Figs. 3 and 4, the initial depression of T_c depends strongly upon λ_o . Thus the magnitude of the initial depression is also an indirect measure of λ_o .

In the model used here the conduction electrons interact with Einstein phonons with a coupling strength λ_o , and frequency ω_o , resulting in a transition temperature, T_{co} , of the pure system. In terms of Eliashberg parameters $\alpha^2(\omega)F(\omega) =$

Table 1. Kondo temperatures, T_K , of Ce impurities in La-Th alloys. The values of T_K in the second and third columns were obtained by fitting the result of MZ and this result (MJ) to $N(0)(\partial T_c/\partial c)_{c=0}$ data reported by Luengo *et al*[11]. The values in the fourth column were estimated from the resistivity, $\rho(T)$, data of Peña and Meunier [12]. T_K was taken as the temperature half way up the logarithmic rise of $\rho(T)$ of samples driven normal by 2 → 3% Ce impurities. The values in parenthesis were estimated by Meunier *et al*[13] from their own and other data.

% Th	T_K (MZ) (°K)	T_K (MJ) (°K)	T_K (exp.) (°K)
0	1	—	— (0.1)
10	5	—	—
15	8	1.1	2.5
20	14	2.1	3.5
25	18	2.6	3.5 (< 3)
34	55	4.5	4.3
41	96	8.0	5.0 (≈ 10)

$\lambda_o \frac{\omega}{2} \delta(\omega - \omega_o)$, where F is the phonon density of states and α^2 describes the electron-phonon interaction. For all values of λ_o it has been shown that $T_{co} \approx 0.25\omega_o (e^{2/\lambda_o} - 1)^{-1/2}$ [15]. This temperature is lowered by a small concentration of Anderson impurities characterized by a hybridization width Δ , an on-site repulsion U , and a Kondo temperature T_K . For the symmetric Anderson model a perturbation expansion gives $T_K = 0.364 (2\Delta U/\pi)^{1/2} e^{-\pi U/8\Delta}$ [14,16].

Calculation of $(\partial T_c / \partial c)_{c=0}$ required a combination of Monte Carlo simulation and diagrammatic perturbation theory. In the limit $c \rightarrow 0$ the impurities are uncorrelated. So that each impurity makes an independent contribution to $(\partial T_c / \partial c)_{c=0}$, and the net contribution is simply cN times the contribution of a single impurity (where N is the number of lattice sites). Thus it is only necessary to simulate a single impurity embedded in the metallic host.

Using standard Eliashberg-Migdal diagrammatic techniques one finds that

$$N(0) (\partial T_c / \partial c)_{c=0} = -T^2 \Delta \sum_n \frac{C_n^2}{|\omega_n|^2} (Im|G_d(i\omega_n)| - \Delta|G_d(i\omega_n)|^2) + T^4 \Delta^2 \sum_{n,m} \frac{C_n C_m}{|\omega_n||\omega_m|} (\Gamma_d(i\omega_n, i\omega_m) - \beta^2 |G_d(i\omega_n)|^2 \delta_{n,m}) \quad (1)$$

where ω_n and ω_m are fermion Matsubara frequencies at temperature $T = 1/\beta$, $N(0)$ is the conduction band density of states at the fermi energy, and G_d and Γ_d [17] are the one and two particle greens functions, respectively, for the isolated impurity. The function C_n is obtained from the Eliashberg equation of the pure host lattice. C_n approaches a constant of order unity for $|\omega_n| \ll \omega_o$ and falls off as ω_n^{-2} for $|\omega_n| \gg \omega_o$, thus providing a cutoff for the n and m summations.

The one- and two-particle single impurity d -greens functions, G_d and Γ_d , are obtained from a Monte Carlo simulation of the Anderson model developed by Hirsch and Fye[18]. This simulation is quite similar in form to that discussed by R.L. Sugar elsewhere in this monograph. The essential difference is that here there are strong Coulomb interactions only on the d-site. Thus the simulation

must be formulated so that the non-interacting fermion degrees of freedom may be integrated out.

The basic idea is to recast a d -dimensional quantum problem into a $d + 1$ -dimensional classical problem with a discrete path integral formulation. The Anderson model Hamiltonian is $H = H_1 + H_2$.

$$H_1 = \frac{V}{\sqrt{N}} \sum_{\mathbf{k}\sigma} \left(c_{\mathbf{k}\sigma}^\dagger d_\sigma + d_\sigma^\dagger c_{\mathbf{k}\sigma} \right) + \sum_{\mathbf{k}\sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \left(\epsilon_d + \frac{U}{2} \right) (n_{d\uparrow} + n_{d\downarrow}) \quad (2)$$

$$H_2 = U n_{d\uparrow} n_{d\downarrow} - \frac{U}{2} (n_{d\uparrow} + n_{d\downarrow}) \quad (3)$$

In a discrete path integral formulation

$$\mathcal{Z} = Tr_f e^{-\beta H} = Tr_f \left(e^{-\Delta\tau H} \right) \approx Tr_f \left(e^{-\Delta\tau H_1} e^{-\Delta\tau H_2} \right)^L \quad (4)$$

Before we integrate out the fermions, we need to remove the quartic terms. This is done with a discrete version of the Hubbard-Stratonovich transformation.

$$e^{-\Delta\tau H_2} = \frac{1}{2} Tr_\sigma e^{\alpha\sigma(n_{d\uparrow} - n_{d\downarrow})} \quad (5)$$

where $\cosh(\alpha) = e^{(\Delta\tau U/2)}$, and $\sigma = \pm 1$. making this substitution, we have recast the problem into one of free fermions interacting with a time dependent field σ_l . The price we must pay is having to integrate over all possible configurations of this field.

$$\mathcal{Z} \approx Tr_{f\sigma} (e^{-\Delta\tau H'})^L \quad (6)$$

where

$$\begin{aligned} H' &= \frac{V}{\sqrt{N}} \sum_{\mathbf{k}\sigma} \left(c_{\mathbf{k}\sigma}^\dagger d_\sigma + d_\sigma^\dagger c_{\mathbf{k}\sigma} \right) + \sum_{\mathbf{k}\sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} \\ &+ \left(\epsilon_d + \frac{U}{2} \right) (n_{d\uparrow} + n_{d\downarrow}) + \frac{\alpha}{\Delta\tau} \sigma_l (n_{d\uparrow} - n_{d\downarrow}) \end{aligned} \quad (7)$$

Since H' is quadratic, we may integrate out the fermion degrees of freedom.

$$\mathcal{Z} = Tr_\sigma \prod_\mu \det(\mathbf{O}_\mu) \quad \mathbf{O}(\{\sigma_l\})_\mu = \mathbf{G}_\mu^{-1} \quad \mu = \uparrow, \downarrow \quad (8)$$

\mathbf{O}_μ is a matrix function of the set of Ising variables $\{\sigma_l\}$.

$$\mathbf{O}_\mu = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \dots & +\mathbf{B}_L^\mu \\ -\mathbf{B}_1^\mu & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & -\mathbf{B}_2^\mu & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{B}_{L-1}^\mu & \mathbf{I} \end{pmatrix} \quad (9)$$

Here each entry is itself a $(N + 1)$ by $(N + 1)$ matrix with \mathbf{I} being the identity, $\mathbf{0}$ being a matrix of zeroes, and $\mathbf{B}_l^\mu = e^{-\Delta\tau\mathbf{K}}e^{\mathbf{V}_l^\mu}$. \mathbf{K} is defined by the relation

$$H_{1,s} = \mathbf{f}_s^\dagger \mathbf{K} \mathbf{f}_s \quad \mathbf{f}_s = \begin{pmatrix} d_s \\ c_{k_{1,s}} \\ c_{k_{2,s}} \\ \vdots \\ c_{k_{N,s}} \end{pmatrix} \quad (10)$$

and \mathbf{V}_l^μ by

$$\mathbf{V}_l^\mu = (\sigma_l \alpha \mu) \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \quad (11)$$

Notice that the 1,1 element of each submatrix corresponds with the d -site.

The simulation proceeds by proposing changes in the Ising variables

$$\sigma_l \rightarrow \sigma'_l = -\sigma_l. \quad (12)$$

This change is accepted with probability

$$R = \frac{\det(\mathbf{O}'_\uparrow) \det(\mathbf{O}'_\downarrow)}{\det(\mathbf{O}_\uparrow) \det(\mathbf{O}_\downarrow)} \quad (13)$$

For such a local change in the Ising field, the ratio of determinants is given by

$$\frac{\det(\mathbf{O}'_\mu)}{\det(\mathbf{O}_\mu)} = 1 - (G_d^\mu(l, l) - 1)(e^{\alpha\mu(\sigma'_l - \sigma_l)} - 1) \quad (14)$$

Note that only the d-Greens functions are needed to determine the ratio of determinants.

If the change is accepted, then $\mathbf{G}_\mu = \mathbf{O}_\mu^{-1}$ must be updated accordingly. From Eqs. (9–12) it follows that, in general, the Greens function matrix is obeys the Dyson-like equation

$$\mathbf{G}'_\mu = \mathbf{G}_\mu + (\mathbf{G}_\mu - 1) (e^{\mathbf{V}'_\mu - \mathbf{V}_\mu} - \mathbf{I}) \mathbf{G}'_\mu. \quad (15)$$

We only want to measure the d-Greens functions, and from Eq. (14) we see that only the d-Greens functions are needed to determine the ratio of determinants. As only the d-elements of $(e^{\mathbf{V}'_\mu - \mathbf{V}_\mu} - \mathbf{I})$ are nonzero, we may write an equation for the d-Greens functions

$$\mathbf{G}'_{d,\mu} = \mathbf{G}_{d,\mu} + (\mathbf{G}_{d,\mu} - 1) (e^{\mathbf{V}'_\mu - \mathbf{V}_\mu} - \mathbf{I})_d \mathbf{G}'_{d,\mu}. \quad (16)$$

This is an L by L matrix equation where only the nonzero d-elements of $(e^{\mathbf{V}'_\mu - \mathbf{V}_\mu} - \mathbf{I})$ are used.

From Eqs. (12–14 and 16) it is possible to construct an algorithm involving L by L matrices of only the d-greens function. The non-interacting lattice degrees of freedom have been integrated out. With this algorithm I have taken $\Delta\tau = 0.25$ and 0.1875 and studied β values as large as 25 , in units where $N(0) = .25$ and $4 < U < 8$. Larger values of β were avoided since the computer time required by the algorithm scales like L^3 . The systematic errors associated with the finite value of $\Delta\tau$ were estimated to be typically of order 2% but could be twice that for Γ_d at the largest values of U reported here.

The values of U , Δ , and T_{co} were picked so as to remain in the universal Kondo regime as defined by Krishnamurthy, *et al*[16], while keeping U small enough, and Δ and T_{co} large enough to make the simulation feasible. Thus, although I was able to explore a wide range of T_K/T_{co} , I was limited to relatively high T_{co} , and consequently large values of λ_o and/or ω_o .

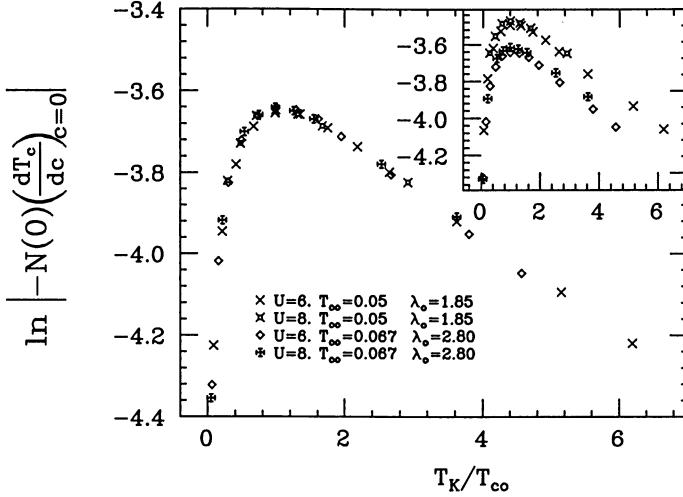


Figure 1: Rescaled $\ln |N(0) (\partial T_c / \partial c)_{c=0}|$ vs T_K / T_{co} when $\omega_o = 0$. The raw data (insert) was rescaled to fit a single function of T_K / T_{co} . The maximum inhibition of the transition temperature always occurs when $T_K = T_c$.

Fig. 1 shows the underlying universal nature of $N(0) (\partial T_c / \partial c)_{c=0}$ as a function of T_K / T_{co} for a wide range of U , T_{co} , and λ_o , when the host Einstein frequency is held fixed at $\omega_o = 0.25$. The data has been rescaled to overlap the $U = 6$, $T_{co} = 0.067$, and $\lambda_o = 2.80$ data by adding or subtracting a constant to obtain the best fit (the unscaled data is shown in the insert). It is clear that the shape of the curve is universal in that it depends only upon the ratio T_K / T_{co} . The only divergence from this universality was found for extremely large ω_o [19]. The maximum depression always occurs when $T_K / T_{co} = 1$ in contrast to previous theories which placed the maximum between 5 and 12.

This will alleviate problems encountered in the past, when experimental data was fit to theory, and unreasonably large values of T_K were inferred which did not agree with those estimated from resistivity data[13,10]. This is illustrated in table 1. Here T_K 's of Ce impurities in a superconducting $\text{La}_{1-x}\text{Th}_x$ alloy are tabulated. In this alloy T_K increases monotonically with Th concentration, x , and $N(0) (\partial T_c / \partial c)_{c=0}$ exhibits a peak at $x \approx 0.45$ [11]. The second column

contains the data of Luengo *et al*[11]. They fit their data to the result of Müller-Hartmann and Zittartz, (MZ)[5] to obtain T_K as a function of x . However, as pointed out by Meunier *et al*[13], the values of T_K they obtained are about an order of magnitude larger than those estimated from the resistivity (shown in the fourth column of table 1), even when $T_K/T_{co} \ll 1$. At first this seems surprising since the MZ result is correct when $T_K/T_{co} \ll 1$. However, the reason for this discrepancy is clear: the data must be fit to the MZ curve by making the points of maximum initial depression coincide. As I have shown, the MZ theory predicts a point of maximum depression which is an order of magnitude too high. Thus all of the inferred values of T_K are off by a corresponding factor. However, when this data is fit to the curve in Fig. 1, the values of T_K obtained, and displayed in the third column, are in much closer agreement with the resistivity data in the fourth column.

In Fig. 2. $N(0)(\partial T_c/\partial c)_{c=0}$ is plotted versus T_K/T_{co} for different values of all the parameters, but for only three values of the host electron-phonon coupling strength λ_o . The data falls on one of three identically shaped curves determined

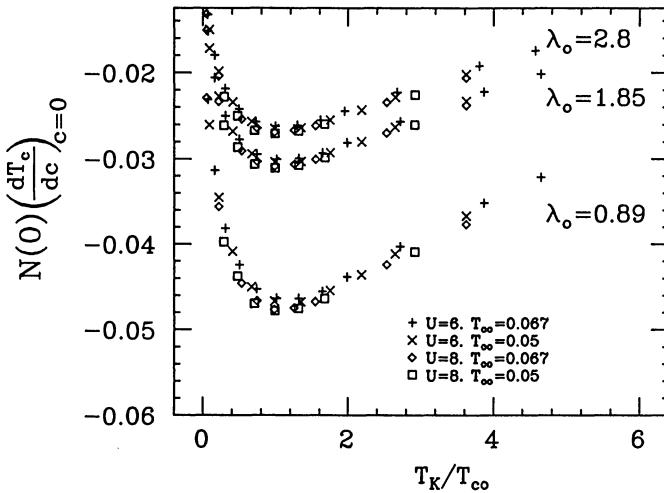


Figure 2: $N(0)(\partial T_c/\partial c)_{c=0}$ vs T_K/T_{co} for three values of the electron-phonon coupling strength λ_o . Although U , and T_{co} varied, λ_o determines which of the three curves the data falls on.

by λ_o , with $\lambda_o = 2.8$ for the upper curve, $\lambda_o = 1.85$ for the middle curve, and $\lambda_o = 0.89$ for the lower curve. The spread in the data on each curve is systematic, due to the finite values of $\Delta\tau$, which becomes worse as β and U increase. Thus $N(0) (\partial T_c / \partial c)_{c=0}$ may be expressed as the product of a universal function $g(T_K/T_{co})$, which contains the shape of the curve, multiplied by a scale factor $-f(\lambda_o)$ which is a function of only λ_o . If we define $g(1) = 1$, then $-f(\lambda_o)$ is the maximum initial depression.

For example, in Fig. 3 I have plotted the maximum initial depression versus λ_o , when $U = 8$ and ω_o is adjusted to keep $T_c = T_K = 0.05$ (ω_o ranged from $\omega_o = .2$ for $\lambda_o = 2.54$ to $\omega_o = 1.0$ for $\lambda_o = 0.56$). The maximum depression is strongly dependent on λ_o , especially when λ_o becomes small. In the weak coupling limit ($\lambda_o \rightarrow 0$) MZ find the maximum initial depression $-N(0)(\partial T_c / \partial c)_{c=0 \max} = f(0) \approx 0.125$. If the curve in Fig. 3 extrapolates to this value, then $N(0)(\partial T_c / \partial c)_{c=0}$ will be strongly dependent upon the electron-phonon coupling strength over the range $0 < \lambda_o < 3.0$. Most normal superconductors have electron-phonon couplings within this range and hence the initial depression will depend strongly upon the electron phonon coupling, and $-N(0)(\partial T_c / \partial c)_{c=0 \max} < 0.125$. For example, Luengo *et al*[11] find a maximum

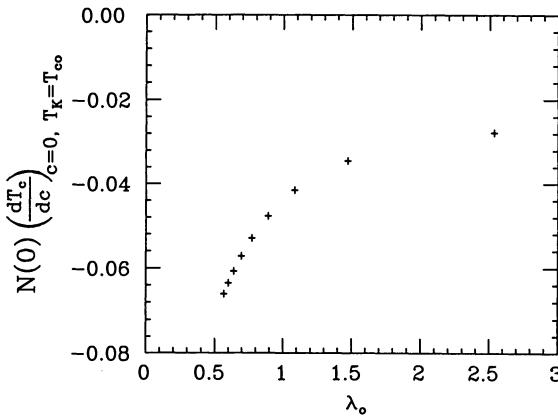


Figure 3: Maximum initial depression versus λ_o when $U = 8$ and $T_c = T_K = 0.05$. The maximum depression decreases monotonically with λ_o , and changes most rapidly when $0.5 < \lambda_o < 1.5$.

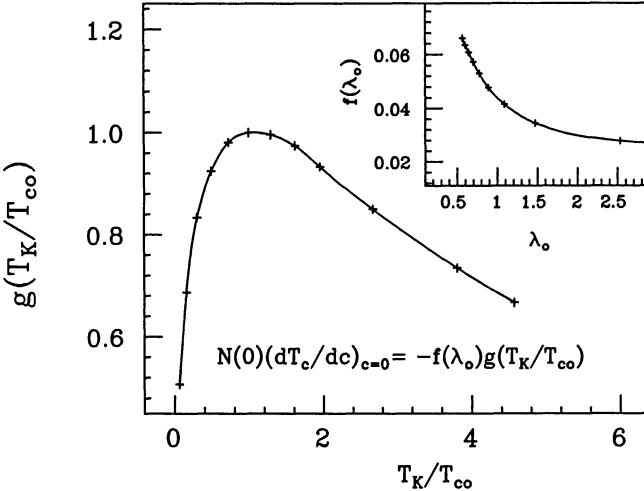


Figure 4: Plots of $g(T_K/T_{co})$ and $f(\lambda_o)$ obtained from a fit to data.

initial depression $N(0)(\partial T_c/\partial c)_{c=0, \max} \approx 0.091$ due to Ce impurities in a La-Th superconductor.

In Fig. 4 $f(\lambda_o)$ and $g(T_K/T_{co})$ are plotted. They were obtained from a fit to all the $(\partial T_c/\partial c)_{c=0}$ data, $N(0)(\partial T_c/\partial c)_{c=0} = -f(\lambda_o)g(T_K/T_{co})$. The function $f(\lambda_o)$ is the maximum initial depression found in Fig. 3, and $g(T_K/T_{co})$ is a universal function of T_K/T_{co} . The function $f(\lambda_o)$ could not be obtained for $\lambda_o < \frac{1}{2}$ (the weak coupling limit) as this would require either unphysically large values of ω_o or values of T_{co} too small for the Monte Carlo simulation. Note that the impurities inhibit superconductivity most effectively in weak coupling superconductors. This is due to the competition between the pair breaking potential and the pairing potential. If we hold T_K fixed, and increase the pairing potential (which is proportional to λ_o), then the effect of the magnetic impurities diminishes.

In conclusion, I have made the first exact calculation of $(\partial T_c/\partial c)_{c=0}$, and shown that it is a universal function of T_K/T_{co} multiplied by a scaling function of λ_o . $(\partial T_c/\partial c)_{c=0}$ has been calculated from the pair breaking regime ($T_K/T_{co} \ll 1$) where the impurities are magnetic, to the pair weakening regime ($T_K/T_{co} \gg 1$)

where the impurities are non-magnetic due to the screening of the conduction electrons. Two new features were found. First, the maximum depression always occurs when $T_K/T_{co} = 1$ in quantitative disagreement with previous theories which break down as T_K/T_{co} approaches one. It was this large quantitative discrepancy which resulted in an extreme overestimation of Kondo temperatures when data was fit to these theories. I have shown that when data is fit to the curves shown here, more reasonable values of T_K are obtained. Second, the maximum initial depression $f(\lambda_o)$ changes rapidly with λ_o in the region $0.5 < \lambda_o < 3.0$, and $f(\lambda_o)$ becomes larger as λ_o becomes smaller, indicating that the impurities inhibit superconductivity most effectively in weak coupling superconductors.

I am pleased to acknowledge useful conversations with J. W. Allen, D.L. Cox, C. Jayaprakash, H.R. Krishnamurthy, M.B. Maple, D.J. Scalapino, and J. W. Wilkins. I am especially indebted to Bernd Schüttler who developed the algorithm used to analyze the Monte Carlo data[3,2].

This work was supported by the DOE-Basic Energy Science, Division of Materials Research, and the Ohio Supercomputer Center.

References

- [1] M.B. Maple, *Magnetism: A Treatise on Modern Theory and Materials*, ed. H. Suhl, (Academic Press, New York, 1973), Vol. V, p. 289-325.
- [2] H. B. Schüttler, M. Jarrell, and D. J. Scalapino, Phy. Rev. Lett. **58**, 1147, (1987).
- [3] H. B. Schüttler, M. Jarrell, and D. J. Scalapino, J. Low Temp. Phys. (1987).
- [4] W. Gey, and E. Umlauf, Z. Physik **242**, 241 (1970).
- [5] E. Müller-Hartmann, and J. Zittartz, Phys. Rev. Lett. **26**, 428, (1970).

- [6] A.A. Abrikosov and L.P. Gorkov, J. Exp. i Theor. Phys. **39** (1960), 1781
 [Soviet Phys.-JETP **12** (1961), 1243].
- [7] T. Matsuura, S. Ichinose, and Y. Nagaoka, Prog. of Theor. Phy. **57**, 713
 (1977).
- [8] N.E. Bickers and G.E.Zwicknagl, Phys. Rev. B **36**, 6746 (1987). This cal-
 culation was for the six-fold degenerate Anderson model, which may quan-
 titatively effect the results.
- [9] M.B. Maple, and K.S. Kim, Phys. Rev. Lett. **23**, 118, (1969).
- [10] S. Takayanagi *et.al*, J. Low Temp. Phys. **16**, 519, (1974).
- [11] C.A. Luengo, J.G. Huber, M.B. Maple, and M. Roth, Phys. Rev. Lett. **32**,
 54 (1974).
- [12] O. Peña and F Meunier, Sol. State Comm. **14**, 1087 (1974)
- [13] F. Meunier, S. Ortega, O. Peña, M. Roth, and B. Coqblin Sol. State Comm.
14, 1091 (1974).
- [14] F.D. Haldane, J. Phys. C **11**, 5015 (1978).
- [15] V.Z. Kresin, Phys. Lett. A /bf 122, 434 (1987).
- [16] H.R. Krishna-murthy, J.W. Wilkins, and K.G. Wilson, Phys. Rev. B **21**,
 1003, (1979).
- [17] Γ_d is defined by

$$\Gamma_d(i\nu, i\nu') = \int_0^\beta d\tau_1 \cdots d\tau_4 e^{[i\nu'(\tau_1 - \tau_2) - i\nu(\tau_3 - \tau_4)]} \\ \langle T d_\uparrow(\tau_1) d_\downarrow(\tau_2) d_\downarrow^\dagger(\tau_4) d_\uparrow^\dagger(\tau_3) \rangle$$
 where d_σ destroys a spin- σ fermion on the impurity site.
- [18] J.E. Hirsch, and R.M. Fye, Phys. Rev. Lett. **56**, 2521, (1986).
- [19] M. Jarrell, Phys. Rev. Lett. **61**, 2612, (1988).

Part III

Contributed Papers

New Interfacial Phenomena in a Driven Diffusive System

J.L. Vallés¹, K.-t. Leung², and R.K.P. Zia²

¹School of Physics and Astronomy, University of Minnesota,
Minneapolis, MN 55455, USA

²Physics Department, Virginia Polytechnic Institute, Blacksburg, VA 24061, USA

Abstract. Non-equilibrium stationary states constitute a subject of increasing interest where computer simulation has proven to be an invaluable tool. In this paper we describe the behavior of a driven lattice gas when we impose shifted periodic boundary conditions, which are used in equilibrium to generate tilted interfaces. In the case of small shifts, our study shows a non-analyticity in the internal energy, in agreement with previous results which indicate that the driving field suppresses the interface roughness. When we increase the value of the shift beyond a critical angle, we first obtain a splitting into a multistrip configuration and then a series of merging transitions.

Driven diffusive systems present stationary states which remain far from equilibrium under the action of an external agent. Even the simple case of a lattice gas with an electric field (E) that generates a current along one of the principal directions contains very interesting new features [1], and has been the subject of much analytical [2] and numerical [3-5] work during the past few years. Most of these recent studies were mainly focused on the nature of the non-equilibrium phase transitions and the associated critical behavior. In two dimensions, for instance, extensive Monte Carlo simulations [3] strongly indicate the existence of a new non-classical universality class, and provide the values of the new critical exponents, thus qualifying as a most valuable tool in these non-equilibrium problems, when the use of the traditional Gibbs' formalism is invalidated. Numerical studies have also found in these systems a distinct power-law decay of the correlations above the critical temperature [5].

Below the transition temperature, the driven lattice gas configurations present a coexistence of anisotropic domains which belong to the ion-rich and ion-poor phases. The study of the interfaces separating these domains is very important, as in the case of equilibrium phase transitions, but very few analytical results exist [6].

The interface of a three dimensional lattice gas in equilibrium is known to undergo a roughening phase transition [7], which is the mechanism associated with the disappearance of facets in a crystal. When the temperature is raised through a temperature T_R the smooth interface becomes rough, this being characterized by a divergence of the statistical width in the thermodynamic limit. The unbound proliferation of the excitations of step formation that this involves is only possible because the excess step free energy vanishes. In contrast, for the two dimensional lattice gas $T_R = 0$ and interfaces are rough at all temperatures.

If the action of an electric field drives the system to a stationary state far from equilibrium, the behavior of the interface can be substantially different. This motivated our recent Monte Carlo simulations [8] which analysed the lattice gas in two dimensions for a wide range of values of E , and indicate a strong suppression of the interface roughness. While this resembles the effects of a gravitational field on the interface in equilibrium, in our case the interface is parallel to the field and we also find subtle differences when comparing the structure factors.

When one studies a system on a finite lattice in equilibrium, an interface with a single step can be interpreted also as the limiting case of a small tilt θ . The excess step free energy is associated then with the linear term in the expansion of the anisotropic surface tension τ in powers of the tilt angle, i.e., when normalized by the area $f_s = \frac{\partial \tau}{\partial \theta}|_0$. One can obtain these tilted interfaces by imposing shifted periodic boundary conditions (SPBC) in one of the principal directions of the lattice, and they are often used in Monte Carlo studies to explore f_s in equilibrium systems [9]. Although we cannot compute the free energy in our non-equilibrium model, the study of the internal energy $u(\theta)$ of the tilted interfaces generated through SPBC will

provide [10] unexpected new results clearly associated with the presence of the field. Moreover, the existence of SPBC will prove a new external agent that competes with the field, and one observes interesting new phenomena as the configurations tend to become commensurate with the boundary conditions.

The system we have studied is based on the half-filled two-dimensional Ising model with attractive interactions. The lattice sites ($i, j = 1, 2, \dots L$) can be empty ($n_{i,j} = 0$) or occupied by an ion ($n_{i,j} = 1$), and the action of a uniform electric field E pointing in the -x direction induces a particle-conserving hopping dynamics governed by a transition probability per unit time $p = \min(1, \exp[-(\delta H + \epsilon E)/k_B T])$, where δH is the change in the lattice gas configurational energy,

$$H(\{n_{i,j}\}) = -4J \sum_{n.n.} n_{i,j} n_{i',j'} , \quad J > 0, \quad (1)$$

and $\epsilon = 1, 0, -1$ when the jump is against, transverse to or favored by E respectively. In order to study tilted interfaces we impose SPBC in the x direction while keeping PBC in the y direction. This is expressed as:

$$n_{i,j+L} = n_{i,j} \quad n_{i+L,j} = n_{i,j-h} \quad (2)$$

for a shift of h lattice spacings. For systems in equilibrium the free energy can be obtained through a thermodynamic integration in T from the internal energy u , which is related to the average number of ion-hole bonds per lattice site,

$$u(\theta, T) = \frac{\langle H \rangle}{2JN} + 2. \quad (3)$$

Since we cannot use the free energy in our driven system, we shall try to characterize the interface by the excess internal energy, i.e. the energy difference between systems with and without an interface,

$$\Delta u(\theta, T) = \frac{1}{2} \{u(\theta, T) - u_B(\theta, T)\}. \quad (4)$$

For systems in equilibrium the two interfaces are equivalent and the bulk energy u_B does not depend on θ . Since u_B is a constant and the θ dependence is not affected by the integration

$$\frac{\partial u(0, T)}{\partial \theta} \neq 0 \Leftrightarrow \frac{\partial \tau(0, T)}{\partial \theta} \neq 0, \quad (5)$$

and the slope of the energy-vs-angle curve will determine whether the interface is rough or smooth. Because we expected that also for a driven system the internal energy $u(\theta)$ would contain significant quantitative information about the interface we carried out a simulation for small h in a system with $L = 100$ and $E = 15$ (∞ in practice) at $T=0.8$ (in units of $T_c = 2.2692 J/k_B$). The runs extended over 4×10^5 MCS. When compared with the results for PBC, this non-equilibrium case with SPBC reveals several surprising new features. As one can see in Fig. 1, the two interfaces are not equivalent: the 'leading edge' appears rough while the 'trailing edge' appears smooth. Moreover, the bulk energy density u_B and the particle density depend on the shift h , and there is a gradient in the particle density, induced by the inequivalence of the two interfaces. A finite-size analysis can help us separate the bulk and interface effects. With this aim, we performed simulations with $L = 20, 36, 48$ and 100 at the same $T=0.8$ and obtained $u_B(\theta)$ and $u_I(\theta)$ from the fit

$$u\left(\frac{1}{L}, \theta\right) = u_B(\theta) + u_I(\theta) \frac{\sec \theta}{L}, \quad (6)$$

where the second term is the interfacial contribution. To first order in $1/L$ we find that u_B is non-analytic in θ , namely

$$u_B(\theta) \approx 0.08 + 0.75 |\theta| + \dots \quad u_I(\theta) \approx 2.68, \quad (7)$$

which is a completely unexpected result. Also, for equilibrium systems bulk properties do not depend on the existence of a single step on the interface. In our system, in contrast, even at

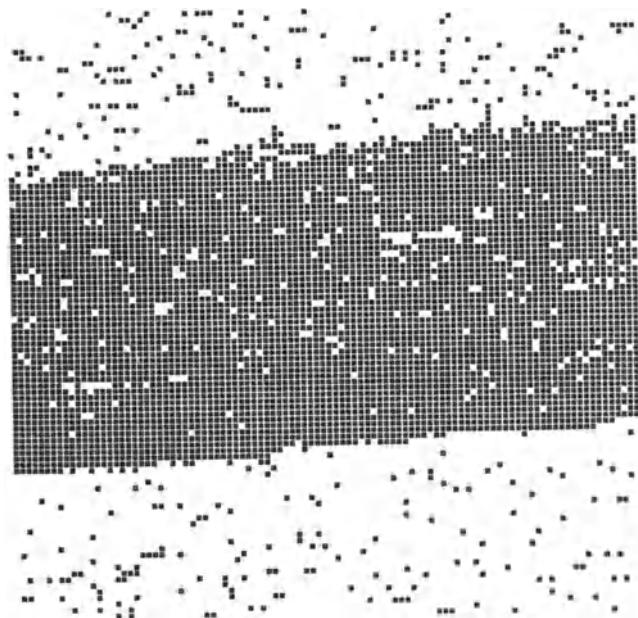


Fig. 1. Configuration for $L = 100$ and $T = 0.8$ corresponding to SPBC with a shift $h = 8$. One can see that the leading edge (the upper interface) appears rough while the trailing edge (the lower one) appears smooth. The electric field points from right to left.

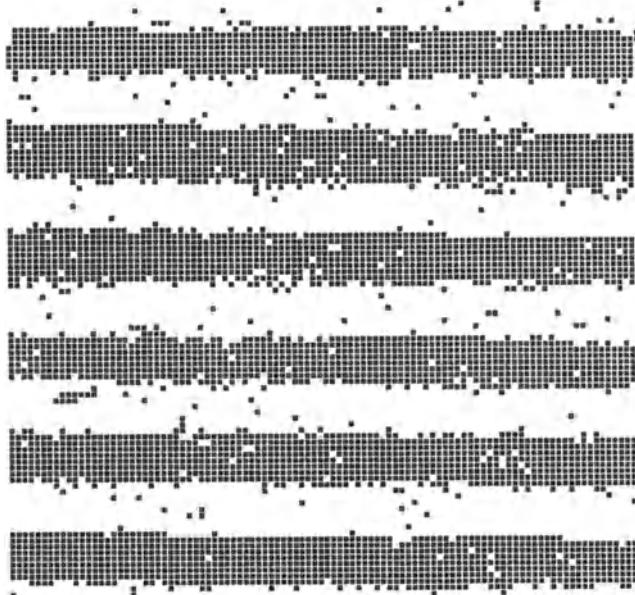


Fig 2. Multistrip configuration for $L = 100$ and $T = 0.8$ corresponding to SPBC with a shift $h = 15$, which is above the critical angle.

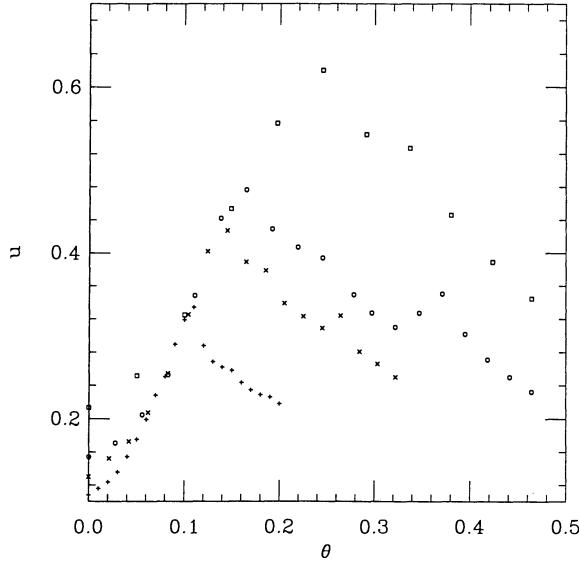


Fig. 3. The energy per site is plotted versus $\theta = \tan^{-1}(h/L)$ for systems of $L = (\square)20$, $(\circ)36$, $(\times)48$, and $(+)100$. The temperature is $T = 0.8$.

$T = 0$ the leading edge is unstable due to the joint effect of the field and the SPBC, and acts as a source of particles while the trailing edge acts as a sink, so that a steady state is reached.

When we studied larger values of the shift h , for sizes $L = 20, 36, 48$ and 100 at $T = 0.8$, the competition between E and the SPBC proved to have even stronger effects. In equilibrium the stable configuration always consists of a single strip with interfaces tilted at an angle $\tan^{-1}(h/L)$. Under the action of E , however, this only occurs for small shifts. Above a critical angle $\theta_c(L)$ one finds multistrip configurations instead, like the one shown in Fig. 2. As h is increased further, the splitting transition is followed by a series of merging transitions, from n -strip configurations to $(n-1)$ -strip ones, as the system adopts more stable configurations with a smaller number of strips. The lowest energy for those n -strip states corresponds to the shift closest to L/n , when the strips match the SPBC. If we change h away from one of those exact divisors of L the energy increases, as one can see in Fig. 3, where we have plotted the energy u versus the angle θ for several L . Simulations performed at higher temperatures indicate that the critical angle is less well defined at higher T , as is also the case for the interface.

The study of this driven lattice gas shows that the combined effect of the field E and the SPBC originates a new and surprising behavior both for small and large values of the shift. This is a good example of the relevance of the boundary conditions in non-equilibrium stationary states. We have also found that although we do not have a free energy, the internal energy provides a good quantitative description of the new non-equilibrium interfacial phenomena.

Acknowledgements: J.L.V. was supported by the C.I.R.I.T. de la Generalitat de Catalunya, Spain, and would also like to thank the Computational Many-Body Theory Group at Courant Institute for its hospitality. K.-t.L. was supported in part by the Advanced Computational Methods Center of the University of Georgia. R.K.P.Z. was supported in part by the N.S.F. through the Division of Materials Research. This research was conducted using the Cornell National Supercomputer Facility, a resource of the Center for Theory and Simulation In Science and Engineering at Cornell University, which is funded in part by the National Science Foundation, New York State, and the IBM Corporation.

REFERENCES

1. S. Katz, J.L. Lebowitz and H. Spohn, *Phys. Rev. B* **28**, 1655 (1983); *J. Stat. Phys.* **34**, 497 (1984).
2. H. van Beijeren and L. S. Schulman, *Phys. Rev. Lett.* **53**, 806 (1984); J. Krug, J. L. Lebowitz, H. Spohn and M. Q. Zhang, *J. Stat. Phys.* **44**, 535 (1986); K.-t. Leung and J. L. Cardy, *J. Stat. Phys.* **44**, 567 (1986); H. K. Janssen and B. Schmittmann, *Z. Phys. B* **64**, 503 (1986).
3. J. L. Vallés and J. Marro, *J. Stat. Phys.* **43**, 441 (1986); *J. Stat. Phys.* **49**, 89 (1987).
4. J. Marro and J. L. Vallés, *J. Stat. Phys.* **49**, 121 (1987); J. Marro, J. L. Lebowitz, H. Spohn and M. H. Kalos, *J. Stat. Phys.* **38**, 725 (1985).
5. M. Q. Zhang, J. S. Wang, J. L. Lebowitz and J. L. Vallés, *J. Stat. Phys.* **52**, 1461 (1988).
6. K.-t. Leung, *J. Stat. Phys.* **50**, 405 (1988); A. Hernandez-Machado and D. Jasnow, *Phys. Rev. A* **37**, 626 (1988).
7. For a review, see H. van Beijeren and I. Nolden in *Topics in Current Physics*, **43**, ed. W. Schommers and P. von Blankenhagen (Springer, Berlin 1987), pp. 259-300.
8. K.-t. Leung, K. K. Mon, J. L. Vallés and R. K. P. Zia, *Phys. Rev. Lett.* **61**, 1744 (1988); *Phys. Rev. B* **39**, - (1989).
9. K. K. Mon, S. Wansleben, D. P. Landau and K. Binder, *Phys. Rev. Lett.* **60**, 708 (1988); *Phys. Rev. B* **39**, 7089 (1989).
10. J. L. Vallés, K.-t. Leung and R. K. P. Zia, *J. Stat. Phys.* **56**, 43 (1989).

Spin Dynamics of Cuprate Superconductors: Exact Results from Numerical Continued Fraction Expansions

C.-X. Chen and H.-B. Schüttler

Center for Simulational Physics, University of Georgia, Athens, GA 30602, USA

We present the first results for the $T = 0$ dynamical structure factor of finite-sized 2D square-lattice spin-1/2 antiferromagnetic Heisenberg systems with up to 16 spins, obtained by numerical continued fraction expansion techniques. Our results are compared to a recently proposed Schwinger boson mean-field theory. We find that the mean-field theory represents an excellent approximation to the exact spin excitation spectra and to the static structure factor for all system sizes studied. We also show the results for the spin dynamics in the presence of a dopant induced hole carrier, describe by either single or multi-orbital Hubbard models.

The coexistence of superconductivity and strong antiferromagnetic correlations in the high- T_c cuprate materials¹ has renewed our theoretical interest^{2–15} in the two-dimensional (2D) spin-1/2 Heisenberg and strong coupling Hubbard models. Here we report the first exact numerical results for the dynamical structure factor of the 2D spin-1/2 Heisenberg antiferromagnet on finite-sized lattices at zero temperature. For clusters with up to 16 lattice sites, we have applied a Mori-type¹⁶ continued fraction expansion (CFE) technique^{17–19} to calculate real-frequency spectra from Lanczos-generated groundstate wavefunctions²⁰.

We compare our results to the $SU(2)$ Schwinger boson mean-field theory which was recently proposed by Arovas and Auerbach (AA)¹⁴ and is here applied to the finite-sized systems solved numerically. This approximation preserves the spin rotational invariance of the model Hamiltonian in 2D at $T = 0$, and predicts the long-range antiferromagnetic order, consistent with the recent numerical results³. Furthermore, it does not exhibit spontaneous symmetry breaking at $T > 0$, consistent with the Mermin-Wagner theorem²¹. On the other hand, the approximation is uncontrolled, and detailed comparison with an exact solution of the model is therefore a crucial step to establish its validity.

The Hamiltonian is in the following assumed to be of the form

$$H = \sum_{\langle l,m \rangle} J \vec{S}_l \cdot \vec{S}_m$$

with \vec{S}_l denoting the spin-1/2 vector operator at site l . The $\langle l, m \rangle$ -summation runs once over all nearest neighbor bonds on a square-shaped square lattice cluster with periodic boundary conditions and $N_L = L \times L$ sites. The lattice sizes N_L have been selected such that the imposition of periodic boundary conditions does not frustrate the antiferromagnetic correlations, namely, $N_L = 2 \times 2$, $\sqrt{8} \times \sqrt{8}$, $\sqrt{10} \times \sqrt{10}$ and 4×4 . The units of energy and length are chosen such that the exchange constant $J = 1$ and lattice constant $a = 1$, respectively. The dynamical spin-spin structure factor $S(q, \omega)$ at wavevector q and frequency ω is given by

$$S(q, \omega) = N_L^{-1} \sum_{l,m} \int dt e^{i\omega t - iq(l-m)} \langle S_l^{(z)}(t) S_m^{(z)}(0) \rangle,$$

where $S_l^{(z)}(t)$ is the spin z-component at site l after time-evolution in real-time t and $\langle \dots \rangle$ denotes the thermal average at temperature T . For finite systems, there is only a finite set of allowed q-vectors in the first Brillouin zone and, as a function of ω , each $S(q, \omega)$ can be written as a superposition of δ -function terms of the form $S_\nu(q)\delta(\omega - \omega_\nu(q))$ with discrete excitation energies $\omega_\nu(q)$ and intensities $S_\nu(q)$. These frequency δ -functions are in the following Figs.1-3 displayed as Lorentzian peaks with a finite width (FWHM) $\epsilon = 0.03$.

To generate exact results for $S(q, \omega)$ at $T = 0$, one needs to solve first for the groundstate of H . Numerically, this can be done by using a modified Lanczos algorithm^{6,20}. The method is described as follows: We start with a trial wavefunction ψ_0 which should have nonzero projection over the true ground state ϕ_0 . Applying H over ψ_0 we can generate another vector $\psi_1 = (H - \langle \psi_0 | H | \psi_0 \rangle) \psi_0$, which is orthogonal to ψ_0 . After normalization, ψ_0 and

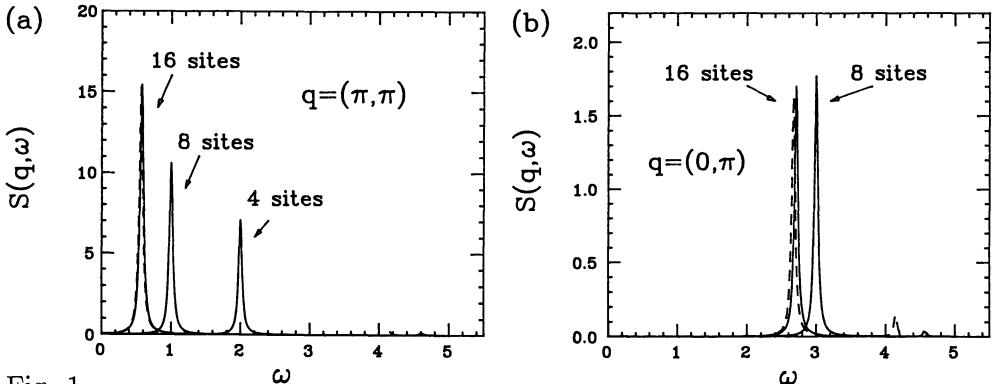


Fig. 1.

$S(q, \omega)$ for lattice sizes $N_L = 4, 8, 16$ at q -vectors $q = (\pi, \pi)$ (a) and $(q = \pi, 0)$ (b). Exact (full) and mean-field (dashed) curves have been plotted for each N_L and q . In most cases, mean-field and exact curves cannot be discerned on the scale of this plot.

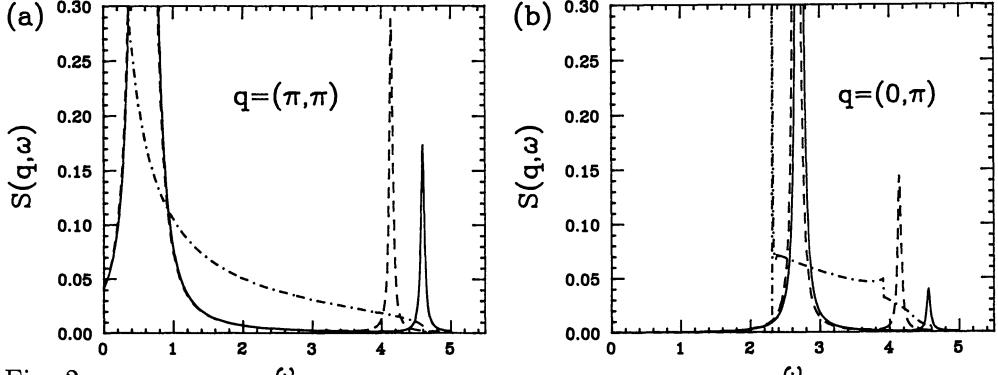


Fig. 2.

$S(q, \omega)$ vs ω for 16 sites lattice at different q -vectors. Exact (full) and mean-field (dashed) curves for 16 sites and infinite system mean-field curves (dot-dash) are shown.

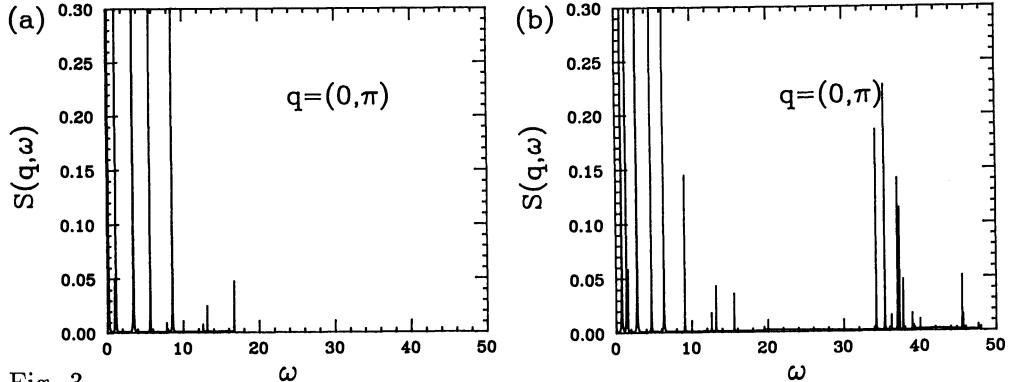


Fig. 3.

(a) $S(q, \omega)$ of the strong coupling single orbital Hubbard model vs ω , for 8 sites lattice. (b) the same as (a), but for three orbital Hubbard model.

ψ_1 form a basis of a 2×2 representation of H , which can be easily diagonalized. Among its two eigenvectors, we choose the one with the lowest eigenvalue as our new trial state, and repeat the process. We have obtained both the groundstate energy and normalized wavefunction amplitudes (relative to a set of site-diagonal spin basis states) with absolute numerical errors no larger than 10^{-5} and in agreement with available published results ⁶. As described in refs.(17-19), the CFE expansion of $S(q, \omega)$ is

$$S(q, \omega) = \frac{1}{\pi} \text{Im} \frac{b_0}{z - a_0 - \frac{b_1}{z - a_1 - \frac{b_2}{z - \dots}}}$$

where $z = \omega - E_0 + i\epsilon$, and E_0 is the ground state energy. The CFE coefficients a_n and b_n are determined by

$$a_n = \langle f_n | H | f_n \rangle / \langle f_n | f_n \rangle, \\ b_n = \langle f_{n+1} | f_{n+1} \rangle / \langle f_n | f_n \rangle, \quad b_0 = \langle f_0 | f_0 \rangle,$$

The intermediate states $|f_n\rangle$ are related to the ground state $|0\rangle$ by the recursion relation:

$$|f_{-1}\rangle = 0, \quad |f_0\rangle = S^{(z)}|0\rangle, \\ |f_{n+1}\rangle = H|f_n\rangle - a_n|f_n\rangle - b_n|f_{n-1}\rangle,$$

Thus the problem of finding $S(q, \omega)$ is reduced to performing further matrix multiplication with the Hamiltonian and can be implemented as efficiently as a Lanczos ground state calculation.

From the groundstate wavefunction, CFE coefficients of (a_n, b_n) were generated up to $n = 200$. For the lattice sizes studied, a CFE representation truncated after $n = 5 - 10$ was found to give converged results for $S(q, \omega)$ to within 1 – 2%.²⁵ Following ref. (14b), we compare our results to a renormalized mean-field structure factor, $S(q, \omega) = (2/3)S^{MF}(q, \omega)$ where $S^{MF}(q, \omega)$ denotes the "bare" mean-field result,²⁶ given by eq.(6) of ref. (14b). The normalization factor 2/3 was introduced *ad hoc* in ref.(14b) so that $S(q, \omega)$ obeys the susceptibility sum rule $\sum_q \int \frac{d\omega}{2\pi} S(q, \omega) = 1/4$. The comparision between MF and exact results is shown in Figs. 1 and 2. Clearly, there is excellent agreement between both sets of data. In fact, most of the main excitation peak shown in Fig. 1 the deviations between MF and exact results are so small that they cannot be discerned on the scale of the plot. Remarkably, the MF results reproduce also qualitatively correct, the appearance of the higher-energy side peaks on larger (4×4) lattice and they give a reasonably accurate quantitative representation of their position and intensity at all wavevectors, as shown in Fig.2. Thus, for small clusters even lower-intensity details are fully accounted for by the mean-field theory.

We have compared our exact diagonalization data and the MC data of Reger and Young³ for the integrated spectral intensity, i. e. the static structure factor $S_s(q) = \int \frac{d\omega}{2\pi} S(q, \omega)$, evaluated at $q = (\pi, \pi)$ with the meanfield results. Again, we find remarkably close agreement between all three sets of data. Consistent with the MC data³, the exact results and analytical arguments⁹, the mean-field results for $S_s(\pi, \pi)/N_L$ fall on a straight line when plotted vs. the inverse linear system size $1/L = N_L^{-1/2}$, and give a mean-field value for the staggered magnetization $m^{(MF)} = 0.306$, in excellent agreement with the MC value³ $m^{(MC)} = 0.30 \pm 0.02$.

We have also studied the effects of dopant induced holes on the antiferromagnetic background. Fig.3 shows results obtained for a single hole carrier in a $\sqrt{8} \times \sqrt{8}$ system for a strong-coupling single orbital (t-J) model^{12,27} (Fig.

3a) and for a strong coupling three orbital Hubbard ²⁸ (Fig.3b), in which oxygen p-orbitals are explicitly taken into account. Consistent with the theory by Zhang and Rice,²⁹ there is a close correspondence in the low-energy properties of the two models. In addition, the three-orbital model shows higher energy spin excitations which can be consistently interpreted as singlet-triplet breakup transitions of the Zhang-Rice singlet quasiparticle.

We would like to thank Drs. M. H. Lee and E.Y. Loh for useful discussions concerning the CFE method. This work was supported by Faculty Fellowships from the Advanced Computational Methods Center and from the Office of the Vice President for Research at the University of Georgia.

REFERENCES

- (1) For a review see: R. J. Birgeneau and G. Shirane, in "Physical Properties of High Temperature Superconductors", D.M. Ginsberg Ed., World Scientific Publishing, Feb. 1989 (in press).
- (2) E. Manousakis and R. Salvador, Phys. Rev. Lett. **60**, 840(1988); Phys. Rev. B **38** (1988) (in press).
- (3) (a) J. D. Reger and A. P. Young, Phys Rev B **37**, 5978 (1988). (b) J. D. Reger, J. A. Riera, and A. P. Young, preprint.
- (4) S. Tang and H. Lin, Phys. Rev. B **38**, 6863 (1988).
- (5) T. Barnes and E. S. Swanson, Phys. Rev. B **37**, 9405 (1988).
- (6) E. Dagotto and A. Moreo, Phys. Rev. B **38**, 5087 (1988).
- (7) J. E. Hirsch and S. Tang, UCSD preprint (1988) (to be published); S. Tang and J. E. Hirsch, UCSD preprint (1988) (to be published).
- (8) S. Chakravarty, B. I. Halperin and D. R. Nelson, Phys Rev. Lett. **60**, 1057 (1988).
- (9) D. A. Huse Phys. Rev. B **37**, 2380 (1988); D.A. Huse and V. Elser, Phys. Rev. Lett. **60**, 2531 (1988).
- (10) S. Miyashita, J. Phys. Soc. Jap. **57**, 1934 (1988).
- (11) J. Oitmaa and D. D. Betts, Can. J. Phys. **56**, 897 (1978).
- (12) P. W. Anderson, Science **235**, 1196 (1987); G. Baskaran, Z. Zou, and P.W. Anderson, Solid State Commun. **63**, 973 (1987).
- (13) I. Affleck and J. B. Marston, Phys. Rev. B **37**, 3774 (1988).
- (14) (a) D. Arovas and A. Auerbach, Phys. Rev. B **38**, 316 (1988); (b) A. Auerbach and D. P. Arovas, Phys. Rev. Lett. **61** 617 (1988).
- (15) M. Takahashi, Phys. Rev. Lett. **58**, 168 (1987); Phys. Rev. B **36**, 3791 (1986); University of Tokyo, ISSP, preprint Ser. A No. 2091 (1989) (to be published).
- (16) H. Mori, Prog. Theor. Phys. **34**, 399 (1965); *ibid.* **33**, 423 (1965).

- (17) M. H. Lee, Phys. Rev. Lett. **49**, 1072 (1982); Phys. Rev. B **26**, 1072 (1982); J. Math. Phys. **24**, 2512 (1983).
- (18) E. R. Gagliano and C. A. Balseiro, Phys. Rev. Lett. **59**, 2999 (1987); Phys. Rev. B **38**, 11766 (1988).
- (19) E. Y. Loh and D. Campbell, Synth. Metals bf 27, 499 (1988)
- (20) E. R. Gagliano, E. Dagotto, A. Moreo, and F. C. Alcaraz, Phys. Rev. B **34**, 1677 (1986).
- (21) N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17** 1133 (1966).
- (22) T. Oguchi, Phys. Rev. **117**, 117 (1960).
- (23) N. E. Bickers, Rev. Mod. Phys. **59**, 845 (1987).
- (24) Statical correlation functions and susceptibilities can be straightforwardly expressed in terms of frequency integrals over the corresponding dynamical correlation functions.
- (25) An n -fold continued fraction with numerators $b_0 \dots b_n > 0$ (c.f. refs. 17-19) has $(n+1)$ distinct poles. Numerically, we find that for large n these poles contain either negligibly small spectral weight or they are very densely clustered around a few frequency values, the actual excitation energies $\omega_\nu(q)$. This spurious splitting of the excitation energies (typically 1% or less) is a result of finite numerical accuracy and determines the limits of resolution of the CFE method. When broadened with a finite width $\epsilon = 0.03$, our results are numerically stable (i.e. show very little n -dependence) between $n = 5$ and $n = 200$.
- (26) We have carefully re-derived the results of AA, ref. (14b), and believe that the left-hand side of their eq.(6) for S^{MF} is in error by omission of an overall factor $(2/N_L)$. In all results shown in the present paper this missing factor has been included.
- (27) B. I. Shraiman and E. Siggia, Phys. Rev. Lett. **60**, 740(1988); S. Trugman, Phys. Rev. **37**, 1597(1988).
- (28) H. B. Schüttler and A. Fedro, J. Appl. Phys. **63**, 4209(1988); P. Prelovsek, Phys. Lett. **A126**, 287(1988); J. Zaanen and A. M. Oles, Phys. Rev. **B37**, 9423(1988).
- (29) F. C. Zhang and T. M. Rice, Phys. Rev. **B37**, 3759(1988).

Statistical Mechanics in Optimization Problems

Pik-Yin Lai

Center for Simulational Physics, University of Georgia, Athens, GA 30602, USA

The idea of treating certain difficult combinatorial optimization problems using the concepts and techniques in statistical mechanics is discussed. The NP-complete problem of graph coloring is used as a simple illustration.

1. NP-complete Problems [1,2]

For many years, researchers have been trying to find 'efficient' algorithms for solving various combinatorial optimization problems. 'Efficient algorithm' means an algorithm requiring a number of steps that grows as a polynomial function in the size of the input. These optimization problems are referred to as polynomial-time problems or just P-class problems. Efficient algorithms are known for many combinatorial optimization problems, but not for each and every one of them. Many problems arising in areas such as computer science, operations research, electrical engineering, number theory and other branches of discrete mathematics have defied solution in spite of the innumerable efforts and attempts over many decades. This class of hard problems includes about three hundred problems which have been attempted individually, and yet no polynomial algorithm has been found to solve any one of them. Furthermore, if a polynomial-time solution of any one member of this class is found, this will imply the existence of polynomial algorithms for the whole class. It should be noticed that there is no direct proof that members in this class of problems are hard to solve, i.e. no polynomial algorithm exists, but the circumstantial evidence suggests that such an efficient solution is highly unlikely to exist. This class of hard problems are classified as NP-complete problems. Thus from the practical point of view, the time needed to find a numerical solution to the NP-complete problems grows faster than any polynomial function of the size of the problem and for large systems, direct approach becomes very quickly intractable.

2. Statistical Mechanics and NP-complete Problems

In a physical system, the free energy is always minimized at equilibrium, thus nature somehow knows its way to find the optimized solution. This is the principle used in formulating optimization problems using statistical mechanics. In recent years, there has been much interest in investigating various combinatorial optimization problems using the concepts and methods of statistical mechanics. Special attention has been concentrated on the so-called NP-complete optimization problems which are very hard

to deal with. Optimization problems such as the travelling salesman problem [3,4], graph partitioning [5,6,7], graph coloring [6,8] and the Bipartite Matching Problem [9] (a P-class problem) were analysed using the ideas and techniques of disordered systems such as spin glasses. It is highly unlikely that polynomial algorithms will exist for these NP-complete problems, thus if the size of the problem is large, things becomes intractable. Fortunately, statistical mechanics often deals with systems of large sizes (thermodynamic limit) and thus may provide a general understanding of the nature of the NP-complete problems. In the following subsection, we will briefly discuss the studies of some famous NP-complete problems with the use of statistical mechanics. Some well-known physical statistical mechanical systems like the spin glass and random field problems were shown to be related to NP-complete problems. On the other hand, some famous NP-complete problems like the travelling salesman problem and graph partitioning problem can be formulated and investigated by statistical mechanics.

There are basically two strategies for the heuristics to search for the optimal feasible solution to a problem, the 'divide and conquer' scheme and iterative improvement. For the first strategy, one divides the given problem into a number of subproblems of tractable sizes and solves them separately. The solutions of the subproblems must then be patched back together. For this method to give good solutions, the subproblems must be naturally disjoint. This agrees with the favorable idea of parallel processing nowadays and with the present powerful parallel computers, this is quite a good method in practice. However if the problem cannot be naturally dispatched, then one needs to arrange the schedules of the subproblems which may in turn be a NP-complete problem by itself! In iterative improvement, one begins with some initial configuration, a systematic rearrangement operation is applied to every part of the system in turn until the rearranged configuration has an improved cost function. Then this rearranged configuration becomes the new initial configuration and the same process is repeated iteratively until no further improvement can be obtained. Thus iterative improvement is essentially a search in the phase space for rearrangement steps which lead downhill. An obvious drawback is that this search usually gets stuck in a local minimum, but what one wants is the global one. To avoid the above mentioned drawback, one observes that trapping in local minima is analogous to supercooling a liquid and the liquid solidifies in a metastable state. So a 'temperature' T is introduced in the optimization problem which does not correspond to anything in the actual optimization problem, but is just a parameter which enables the configuration to have a chance to jump out of local minima in a controlled fashioned. The idea of the Metropolis algorithm was borrowed from statistical mechanics to generalise the iterative improvement with the cost function corresponding to the energy. This is the idea of simulated annealing [10]. The best near-optimal solution is obtained with the temperature lowered slowly and eventually annealed to $T = 0$. In practice, one has to devise the annealing schedule to make sure everything works fine and a knowledge of the energy landscape is an important piece of information (e.g. when the system will 'freeze' and whether there is a phase transition etc.) which may be provided by using the standard methods in statistical mechanics.

Many of the NP-complete combinatorial optimization problems have much in common with the frustrated system with quenched randomness like the spin glass model. The conflicting constraints in the optimization problem cause frustration. Also the infinite range spin glass model at low temperature is shown to have a complex phase

space structure with many minima each separated by an energy barrier that diverges in the thermodynamic limit ($N \rightarrow \infty$). If one starts with a false minimum, in order to overcome this energy barrier, one has to flip a number of spins, say $g(N)$, where g is some increasing function, which takes a time scale at least $\sim e^{O(g(N))}$ in some thermal algorithm like the Metropolis procedure. This is to be compared with the difficulty of no known polynomial algorithm in finding the global optimal solution for an NP-complete problem. In fact this is part of the motivation to study NP-complete problems using the language of statistical mechanics of quenched disordered systems. Also knowledge of the phase space structure of these frustrated random systems, which can be achieved by analytic techniques used in random frustrated system, will be very helpful for improving the heuristics in practice.

The analytic approach we use, as shown in the next section, is to treat the optimization problem as a random system with the instances of the problem obeying a certain probability distribution. Then one can extract the generic or average information and obtain the most probable results by performing some kind of average on the random variables. Thus one obtains results that do not depend on the particular realization of the problem but are applicable to 'almost all' realizations as the size of the problem becomes very large. Of course, one can always make up a 'worse case' scenario which violates the results from the probabilistic analysis, however, if we draw an infinite sequence of problem instances, then with probability one, the predictions would be observed to fail only finitely often. This approach may suffer from the drawback of the fact that one seldom knows the probability distribution in the realistic situations. Nevertheless, this approach yields valuable insights and in many cases helps to improve the heuristics of the algorithm in searching for the optimal solution.

3. The Chromatic Number Problem

The problem is specified as follows: Given a graph $G(V, E)$ where V is the set of vertices ($|V| = N$) and E is the set of edges, what is the minimum number of colors (the chromatic number χ) needed to color the vertices such that no two vertices connected by an edge will have the same color? The planar version of this problem is related to the famous 'four-color theorem' which states that for any planar map the chromatic number is at most four[11], i.e. $\chi_{\text{planar}} \leq 4$. However, the general problem of determining the chromatic number was shown to be NP-complete. In fact, the problem remains NP-complete even if one asks whether $\chi_G \leq 3$ and is still NP-complete even if one further restricts the above question to planar graphs. Monte Carlo simulations on randomly generated planar graphs [8] indicated an ultrametric structure for the optimal configurations (i.e. those that minimize the number of unsatisfied constraints) when the number of colors used is less than four. If four colors were used, this ultrametricity is lost, since there is no more frustration now from the 'four-color theorem'. This ultrametricity is similar to that of the SK model and it is believed that it is a general feature for NP-complete problems.

In the present case, one is interested in a given graph, which is usually quite arbitrary, and wants to determine the number of colors necessary to color it properly. A well-known application of this problem is given below: At the end of the semester, students should sit examinations within a short period of time. It is desirable that

each student sits daily at most one examination and all students having the same examination take it at the same time, and the number of days needed to finish all examinations should be minimized. Let V be the set of examinations and Y be the set of students. For each examination v , let $S(v) \subset Y$ be the set of students who must take v . Construct the graph $G(V, E)$ where $(v, v') \in E$ if $S(v) \cap S(v') \neq \emptyset$ (i.e. v and v' cannot be held at the same time). Then a coloring of G corresponds to a possible assignment for the set of examinations. The number of colors is the number of days needed and the problem reduces to finding $\chi(G)$.

In this work, we investigate the random version of the chromatic problem, i.e. the graph G is a random graph $G(N, p)$ where N is the number of vertices and p is the probability that any two vertices are connected by an edge (model A in graph theory). So every vertex has a finite probability p of linking with every other vertex. There have been some quite intensive investigations on this subject by applied mathematicians and graph theorists. Two important results due to Erdos and Spencer [12] and Bollobas and Erdos [13] are on the bounds of χ for almost every $G(N, p)$ (almost every means as $N \rightarrow \infty$ the probability of occurrence $\rightarrow 1$). For $p = 2c/N$ and c large,

$$(1 + o(1)) \frac{c}{\ln c} < \chi < (1 + o(1)) \frac{2c}{\ln c} \quad (1)$$

where $o(1)$ is with respect to c . In this case, the average coordination number of the graph is $2c$. For p fixed independent of N ,

$$\left(\frac{1}{2} - \epsilon\right) \frac{N \ln(\frac{1}{1-p})}{\ln N} < \chi < (1 + \epsilon) \frac{N \ln(\frac{1}{1-p})}{\ln N}. \quad (2)$$

The upper bound is obtained by the 'greedy algorithm'[14] which can color the graph satisfactorily, but the number of colors used may not be minimal.

It is believed [14] that most likely the greedy algorithm uses twice as many colors as necessary and hence χ should be closer to the lower bound. Furthermore, it was conjectured [15] that χ goes to the lower bound as $N \rightarrow \infty$ in (1) and (2), however, there has been no proof of this. To this date, there is no known polynomial algorithm which can approximate χ to within any constant factor [14,15]. Also there has not been much simulation work done for large N , after all, this is an NP-complete problem. One result is [16]

$$\text{For } G(1000, 0.5) ; \quad \chi \simeq 85 \pm 12. \quad (3)$$

In what follows, we show how methods of statistical mechanics will yield results that agree with (1), (2) and (3).

4. Potts Model and Statistical Mechanical Treatment

It has been shown that the graph coloring problem is related to the Potts model [17,18] and we shall make use of this idea by considering the q-state anti-ferromagnetic Potts Hamiltonian

$$H = \sum_{i < j}^N J_{ij} \delta_{\sigma_i \sigma_j} \quad (4)$$

where σ_i , which can take q values (colors), is the spin at the i th vertex. $J_{ij} > 0$ is the coupling between the vertices i and j (in graphical terms J_{ij} is an edge). N is the number of vertices. We consider a random graph $G(N, p)$, J_{ij} has the distribution

$$p(J_{ij}) = p\delta(J_{ij} - J) + (1 - p)\delta(J_{ij}) \quad ; \quad J > 0. \quad (5)$$

$J_{ij} = 0$ means that there is no edge joining vertices i and j . This type of anti-ferromagnetic model will favor the spins to lie in different states (directions) since spins that are along the same direction give $\delta_{\sigma_i \sigma_j} = 1$ and give a positive contribution to the energy. It can be easily shown that the zero temperature partition function

$$\lim_{\beta \rightarrow \infty} Z_G \equiv P_G(q, N) = \lim_{\beta \rightarrow \infty} \text{Tr} \exp(-\beta \sum_{i < j}^N J_{ij} \delta_{\sigma_i \sigma_j}) \quad (6)$$

gives the number of ways to color G with q colors such that no two vertices connected by a bond have the same color. Since $P_G(q, N)$ is a non-decreasing function of q , χ is given by the minimal positive integer value of q such that $P_G(q, N)$ attains its least positive integer value. Notice that this model (4) and (5) contains an important element, namely frustration, which gives rise to the interesting features.

We then compute the average P_G with probability distribution given by (5), or in spin glass terms, we calculate the 'annealed average'. Using (5), one has

$$\langle P_G(q, N) \rangle_J = \lim_{\beta \rightarrow \infty} \langle Z_G \rangle_J = \text{Tr} \exp\left(-\ln\left(\frac{1}{1-p}\right)\right) \sum_{i < j}^N \delta_{\sigma_i \sigma_j} \quad (7)$$

i.e. the average number of ways to color $G(N, p)$ using q colors can be written as the partition function of an infinite range anti-ferromagnetic Potts model with coupling constant $\ln(1/(1-p))$. Using the results of the infinite range anti-ferromagnetic Potts model, for $\beta H = K \sum_{i < j}^N \delta_{\sigma_i \sigma_j}$ with $K > 0$ as the coupling constant, the partition function is

$$\text{Tr} \exp\left(-\frac{K}{N} \sum_{i < j}^N \delta_{\sigma_i \sigma_j}\right) = q^N e^{-\frac{KN}{2q}} \quad (8)$$

We distinguish now the two cases $p = 2c/N$ and p independent of N . In the first case $K = 2c$ and

$$\langle P_G(q, N) \rangle_J = \left(q e^{-\frac{c}{q}}\right)^N \quad (9)$$

Thus the value q^* which makes P_G equal to 1 is given by

$$q^* \ln q^* = c \quad \text{or} \quad q^* = \frac{c}{\ln c} \left(1 + \frac{\ln \ln c}{\ln c} + O\left(\left(\frac{\ln \ln c}{\ln c}\right)^2\right)\right) \quad (10)$$

Note that in this case the "free energy" $\ln P_G$ is extensive ($\propto N$). χ is given by $\lceil q^* \rceil$ where $\lceil \cdot \rceil$ denotes the next higher integer.

Similarly in the case that p is independent of N , $K = N \ln(\frac{1}{1-p})$ and we find from equation (8) that

$$q^*(1-p)^{\frac{N}{\ln N}} = 1 \quad \text{which implies}$$

$$q^* = \frac{N \ln(\frac{1}{1-p})}{2 \ln N} (1 + \gamma) \quad (11)$$

where

$$\gamma = \frac{\ln \ln N}{\ln N} - \frac{\ln(\ln(\frac{1}{2(1-p)}))}{\ln N} + O\left(\left(\frac{\ln \ln N}{\ln N}\right)^2\right).$$

We see that the values in (10) and (11) are consistent with the bounds given by (1) and (2) and actually coincide with the lower bound. If one is to trust the annealed approximation, it yields also the $\text{o}(1)$ term in (1) as spelled out in eq.(10). Similarly, the leading corrections to the lower bound in (2) are given by equation (11). Our results coincide with the lower bounds as has been conjectured [12]. For $G(1000, 0.5)$ eq.(2) yields $50 < \chi < 100$ whereas our numerical result gives $\chi = 80$ which agrees very well with the numerical simulation value [16] of $\chi = 85 \pm 12$. For a discussion of performing a quench average, please refer to ref. 6.

This work was carried out in collaboration with Yadin Goldschmidt.

REFERENCES

1. Garey M.R. and Johnson D.S., in *Computers and Intractability*, (San Francisco: Freeman 1979)
2. Karp R.M., in *Complexity of Computer Computations* edited by R.E. Miller and J.W. Thatcher p.85 (Plenum Press 1972)
3. Vannimenus J. and Mezard M., *J. Physique Lett.*, **45**, L1145 (1984)
4. Mezard M. and Parisi G., *J. Physique*, **47**, 1285 (1986)
5. Fu Y. and Anderson P. W., *J. Phys. A*, **19**, 1605 (1986)
6. Lai P-Y and Goldschmidt Y. Y., *J. Stat. Phys.*, **48**, 513 (1987)
7. Goldschmidt Y. Y. and Lai P-Y, *J. Phys. A*, **21**, L1043 (1988)
8. Bouchaud J.P. and Le Doussal P., *Europhys. Lett.*, **1**, 91 (1986)
9. Orland H., *J. Physique Lett.*, **46**, L763 (1985)
10. Kirkpatrick S., Gelatt C. D. and Vecchi M. P. Jr., *Science*, **220**, 671 (1983)
11. Appel K. and Haken W., *Sci. Am.*, (Oct. 1977), *Bull. Am. Math. Soc.*, **82**, 711 (1976)
12. Erdos P. and Spencer J., in *Probabilistic methods in combinatorics* (New York and London: Academic Press 1974)
13. Bollobas B. and Erdos P., *Math. Proc. Camb. Phil. Soc.*, **80**, 419 (1976)
14. McDiarmid C., *Ann. Oper. Res.*, **1**, 183 (1984)
15. McDiarmid C., *SIAM J. Comput.*, **8**, 1 (1979)
16. Johri A. and Matula D.W. ,in *Technical Report 82-CSE-6*, Southern Methodist University, Dallas TX (1982)
17. Wu F.Y., *Rev. Mod. Phys.*, **54**, 235 (1982)
18. Baxter R., in *Exactly solved models in statistical mechanics*, p.33 (New York: Academic Press 1982)

Structural Anisotropy of Aggregation Clusters

H.M. Lindsay¹, R. Klein², D.A. Weitz³, M.Y. Lin⁴, and P. Meakin⁵

¹Department of Physics, Emory University, Atlanta, GA 30322, USA

²Fakultät für Physik, Universität Konstanz, Konstanz, Fed. Rep. of Germany

³Exxon Research and Engineering Co., Rt 22E, Annandale, NJ 08801, USA

⁴National Institute of Science and Technology, Gaithersburg, MD, USA

⁵E.I. DuPont Co., Wilmington, DE 19898, USA

Abstract. The fractal description of the structure of aggregation clusters is incomplete, as it does not encompass the anisotropy of the clusters. Using simulated DLCA, RLCA and DLA clusters, we calculate a set of multipole expansion terms, to quantitatively evaluate the cluster anisotropy. Through examination of the scaling of these multipole terms, we show that the anisotropy of cluster-cluster aggregates is scale-invariant, while for DLA clusters of less than 10000 particles, the anisotropy decreases with increasing cluster size. We show that no new scaling exponents are required to determine the wavevector dependence of the multipole terms for cluster-cluster aggregates; only the fractal dimension is needed.

Colloidal aggregation forms complex, highly disordered structures, which can be quantitatively characterized by a fractal dimension d_f .[1] The fractal description can be successfully applied to measurements of the scaling of cluster mass with radius, where $M \propto R^{d_f}$; to measurements of the internal structure of clusters, through the particle pair-correlation function, which scales as $g(r) \propto r^{d_f - 3}$; and to light and x-ray scattering measurements of the static structure factor of the aggregates, which scales as $S(k) \propto k^{-d_f}$. The fractal description is a powerful tool for understanding the growth processes in colloidal aggregation, but it does not fully describe the complex random structure of an aggregate. The structure factor, pair correlation function, and mass scaling, which are well described by a fractal measure, are all orientationally averaged quantities. Since a random aggregate does not have a radial symmetry by the very nature of its formation, a description beyond the fractal dimension is necessary to fully understand its structure.

A quantitative description of the asymmetry of an aggregation cluster can be obtained by expanding the structure into a series of spherical harmonics.[2] This expansion is conducted in momentum space rather than real space, allowing the calculation of the structure factor $S(k)$ rather than the pair-correlation function $g(r)$ from the expansion; this allows us to directly compare light scattering

experiments to our calculations. Additionally, we are still able to probe the length scale dependence of the structure of the aggregates through the dependence of the multipole expansion terms upon the wavevector k .

This paper examines the scaling of the multipole expansion terms with cluster mass and length scale. We use clusters generated through computer simulations of several off-lattice aggregation models.[3-5] To allow comparison with experiments we study aggregates generated by diffusion-limited cluster-cluster aggregation (DLCA) and reaction-limited cluster-cluster aggregation (RLCA). We then compare the results for cluster-cluster aggregates with those for clusters generated by single-particle diffusion-limited aggregation (DLA).

The multipole expansion terms are calculated for a cluster of N particles, with each particle having a displacement b_i and an orientation Ω_i from the center of mass of the cluster. The orientationally averaged scattering from the cluster may then be written as a series of multipole expansion terms

$$S(k) = \sum_l S_l(k) \quad (1)$$

where

$$S_l(k) = \sum_{m=-l}^l \left| \sum_{i=0}^N j_l(k b_i) Y_{lm}(\Omega_i) \right|^2 \quad (2)$$

with $j_l(k b_i)$ the spherical Bessel function of order l , and $Y_{lm}(\Omega_i)$ the spherical harmonics of order l,m . The $l>0$ multipole terms give a direct indication of the importance of the asymmetry of a cluster to its scattering, at a given length scale determined by the scattering wavevector k .

The dependence of the multipoles on cluster mass provides new insight to the scale-invariance of the clusters.[6] The clusters, due to their finite size, each have a characteristic length scale; we choose the radius of gyration R_g as that scale. If the clusters are self-similar below this length scale, then all lengths may be scaled by R_g , and the scaled properties of the cluster should follow universal scale-invariant functions. If the asymmetry of the clusters is scale-invariant, then the only scale-dependence found in Eq. (2) is in the spherical Bessel functions $j_l(k b_i)$, and in the number of particles in the cluster N . Scale-invariant asymmetry implies that if the magnitude of the multipoles terms is scaled by $N^2=S(0)$, and the scattering wavevector is scaled by the radius of gyration of the

cluster, then the multipole terms from clusters of all sizes would scale onto a single curve.

Figure 1 shows the results of the above scaling for the $l=2$ multipole terms for DLCA clusters of 100 to 10000 particles. The scaled amplitudes fall upon a single scaling curve. Similar results are obtained for all other multipole terms calculated. This scaling demonstrates that the structural asymmetry of these DLCA clusters is scale-invariant.

A similar scaling may be attempted for RLCA and DLA clusters. For the reaction-limited clusters, scaling behavior is found, demonstrating that the asymmetry of both types of cluster-cluster aggregates is scale-invariant. For single-particle DLA clusters, when the multipole expansion terms are scaled, they do not fall onto a single curve; instead, the scaled terms decrease as the size of the cluster increases, for clusters of 625 to 10000 particles. This would indicate that DLA clusters become less anisotropic as they grow larger. Quantitatively this behavior can be seen in Figure 2, in which the maximum value of each scaled multipole term is plotted as a function of cluster mass for $l=2$ to 7 for DLA clusters. The peak value clearly decreases as the size of the cluster is increased. The origin of this behavior for the DLA clusters is not clear. We emphasize that we have investigated clusters only up to $N=10000$, so our results may reflect a finite-size effect; they may also reflect the fact that the DLA clusters have a unique center.

The wavevector dependence of the multipole terms for the cluster-cluster aggregates has also been investigated. Each term with

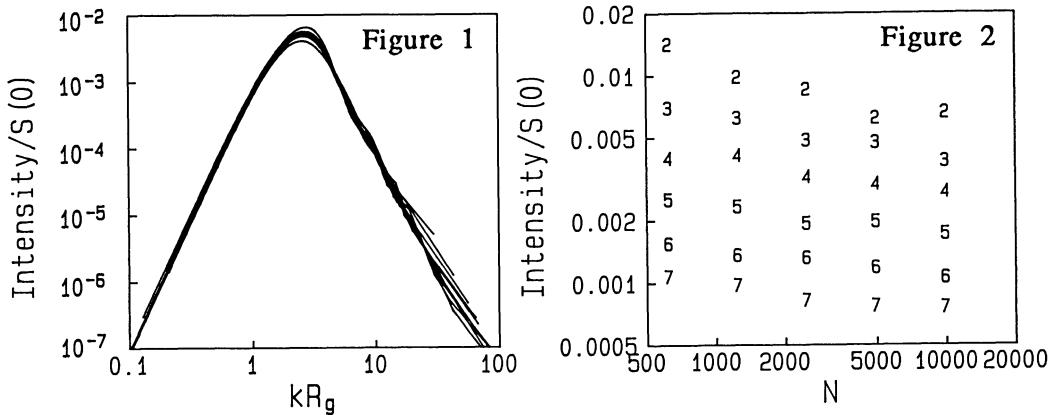


Figure 1. The $l=2$ multipole terms for DLCA clusters of 100 to 10000 particles. Each curve averages over twenty clusters of the same mass, with the amplitude of each term scaled by $S(0)$.

Figure 2. The peak values of the $l=2$ to 7 multipole terms, scaled by $S(0)$, as a function of size for DLA clusters.

$l > 2$ increases as k^{2l} at small wavevector, reaches a peak when $kR_g \approx l$, and then decreases. The behavior at small wavevectors can be predicted from the form of the spherical Bessel functions in Eq. (2). For $kR_g > 1$, the fractal structure of the cluster is resolved; in this regime, the static structure factor, which is the sum of all the multipole terms, scales as k^{-df} . This allows a simple scaling relation to be found for the form of the multipole terms at large wavevectors. Due to the large-amplitude form of $j_l(x)$, all the multipole terms which are at or past their peak values will have roughly the same amplitude. The initial k^{2l} behavior of the multipole terms ensures that only those terms which are at or past their peak contribute substantially to $S(k)$; therefore the number of contributing terms scales as k . Since $S(k) \propto k^{-df}$, and each contributing multipole term is equal, the multipole terms must scale as $k^{-(df+1)}$.

The calculated values of the multipole expansion terms show a wavevector dependence as predicted above. For DLCA clusters a logarithmic plot of peak height versus l gives a slope of -2.76, which is quite close to the expected value of $-(df+1)$, which would be 2.78. For RLCA clusters the results are similar, with a measured slope of -3.1, from clusters with $df=2.1$.

In conclusion, we have shown that cluster-cluster aggregates have scale-invariant asymmetry, while for $N < 10000$, single-particle DLA clusters become less asymmetric as they become larger. We have determined the wavevector dependence of the multipole terms for the cluster-cluster aggregates, and demonstrate that this dependence is determined by the ordinary fractal dimension df ; no new scaling exponent is needed.

References

1. *Kinetics of Aggregation and Gelation*, ed. F. Family and D. P. Landau (North-Holland, Amsterdam, 1984).
2. H. M. Lindsay, R. Klein, D. A. Weitz, M. Y. Lin and P. Meakin, Phys. Rev A38, 2614(1988).
3. P. Meakin, Phys. Rev. Lett. 51, 1119 (1987).
4. P. Meakin and F. Family, Phys. Rev. A36, 5498 (1987).
5. P. Meakin, Phys. Rev. A27, 604 (1983).
6. H. M. Lindsay, R. Klein, D. A. Weitz, M. Y. Lin and P. Meakin, Phys. Rev. A, to appear (1989).

Monte Carlo Studies of Spinodal Decomposition in the 2d Kinetic Ising Model

J.G. Amar

Department of Physics, Emory University, Atlanta, GA 30322, USA

Abstract. Results obtained from extensive Monte Carlo simulations of domain growth in the two dimensional *spin-exchange* kinetic Ising model with equal numbers of up and down spins are presented. Using different measures of domain size - including the pair-correlation function, the energy, and circularly-averaged structure factor - the domain size is determined as a function of time (at $T = 0.5 T_c$) for times up to 10^6 Monte Carlo steps. The growth law $R(t) = A + B t^{1/3}$ is found to provide an excellent fit to the data. Preliminary work, using the Connection Machine 2 (CM2) parallel-processor is also described.

1. Introduction

While the case of domain growth with non-conserved order parameter is well understood to follow the Allen-Cahn growth law [1] $R(t) \sim t^n$ (where $R(t)$ is the average size of an ordered domain) with $n = 1/2$ at long times, domain growth with a conserved order parameter (spinodal decomposition) is not as well understood. Thus, for example, the classic work of Lifshitz and Slyozov [2] predicts a long-time growth law of the form $R(t) \sim t^{1/3}$ in the limit of a dilute amount of one phase or small volume fraction. However, recent Monte Carlo studies [3-6] for the case of equal concentrations of the two phases in the two- and three-dimensional spin-exchange kinetic Ising model have found effective exponents in the range 0.17-0.25 when fitting the domain size to a power law form. In addition, recent work by Mazenko et al [3] on the two-dimensional spin-exchange Ising model at 50% concentration led to the prediction of a logarithmic growth law at long times in this case.

Recently, D.A. Huse [7] presented a generalisation of the Lifshitz-Slyozov theory using phenomenological arguments to apply to the case of equal fractions of the two phases. Huse's theory assumes that the rate of growth of the average domain size R can be written at late times, as

$$\frac{dR}{dt} = C_2 / R^2 (t) + C_3 / R^3 (t) + O(R^{-4}) \quad (1)$$

where C_2 corresponds to the contribution to the growth from bulk-diffusion and C_3 corresponds to the contribution due to surface-diffusion. Solving Eq.1 for large R and t , Huse obtained the growth law

$$R(t) = A + B t^{1/3} \quad (2)$$

where $A = C_3 / 2C_2$ and $B = (C_2)^{1/3}$. If one defines a time-dependent effective growth exponent $n_{\text{eff}}(t)$ as the logarithmic derivative of the domain size $R(t)$ e.g.

$$n_{\text{eff}}(t) = d \ln(R(t)) / d \ln t \quad (3)$$

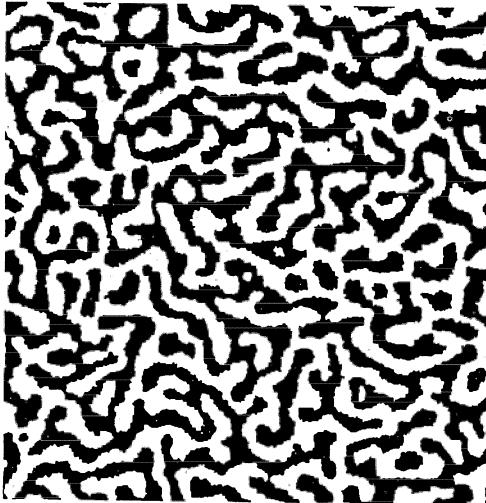


Fig.1 Typical configuration showing interconnected structure formed at critical concentration at a time 100,000 MCS after quench to $T = 0.5 T_c$.

then Eq.2 implies:

$$n_{\text{eff}}(t) = \frac{1}{3} - C/R(t) \quad (4)$$

where $C = C_3 / 6C_2$. Thus, although in the limit of infinite time the effective exponent would approach $1/3$, for any finite time one would observe an effective exponent which varied linearly with $1/R$. Huse argued that for a dilute system (for which growth is in the form of isolated droplets) surface diffusion would be negligible while for the case of large volume fraction (for which the domains are interconnected as in Fig.1) surface diffusion would be significant at early times, and thus a correction to the growth law of the form of Eq.4 would be required.

Huse also presented Monte Carlo data for the two-dimensional spin-exchange kinetic Ising model on a 192×192 lattice consisting of an average of 21 runs up to 4000 Monte Carlo steps (MCS) and 7 runs from 4000 MCS to 40,000 MCS for a quench from infinite temperature to a temperature of $T = 0.9 T_c$, and one run (of 4000 MCS) at $T = 0.5 T_c$ in order to support his conjecture. While Huse's Monte Carlo data indicated an increase in $n_{\text{eff}}(t)$ with time, in agreement with his predictions, his extrapolated long-time exponent $n(t = \infty) = 0.29 \pm .04$ was significantly *less* than the predicted value of $1/3$. In addition, there were large fluctuations in the later-time data.

Thus, it was decided that a more extensive Monte Carlo study with a larger lattice (512^2), a larger number of runs (to get better statistics) and a variety of different measures of the domain size would be useful.

2. Monte Carlo Simulations

Our Monte Carlo simulations [8] were conducted, using the standard Metropolis algorithm, on a 512×512 square lattice with isotropic nearest-neighbor couplings J and periodic

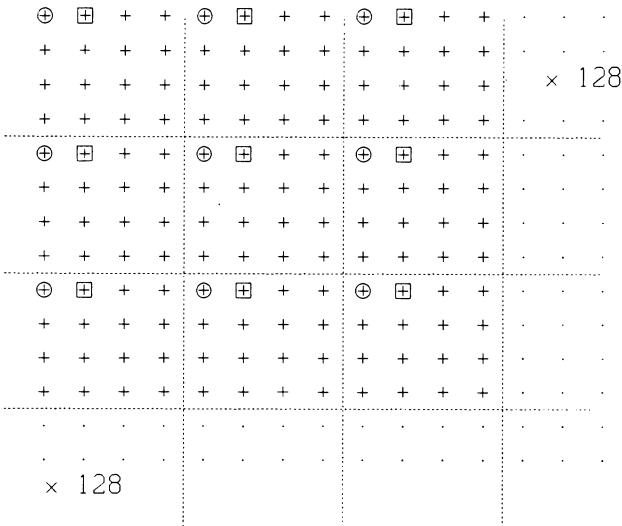


Fig.2 Schematic showing decomposition of 512×512 lattice into 16 (128×128) sublattices. The circled sites correspond to one (128×128) sublattice while the sites labeled by a square correspond to another.

boundary conditions on the Cyber 205 at the National Institute for Standards and Technology. Using the binary representation of an Ising spin as 1 bit of a 64-bit word, we were able to store the entire lattice as one 4096-word binary array. However, instead of attempting one spin exchange at a time, we used an efficient vectorisable algorithm [9] involving a form of multispin coding which is an extension of the "checkerboard" algorithm used in multispin simulations of the Ising model with spin-flip dynamics. In our scheme, the 512×512 lattice is broken up into 16 independent 128×128 sublattices (see Fig.2), with each sublattice corresponding to a vector on the Cyber 205. One "vectorised exchange" corresponds to allowing all the spins from one sublattice to attempt to exchange with all the spins from another nearest-neighbor sublattice, using the Metropolis algorithm. The choice of which pair of sublattices or vectors to try to "exchange" is made randomly and, since there are 32 different pairs of nearest-neighbor sublattices, each 32 sets of "attempted vectorised exchanges" corresponds to one Monte Carlo step. Our scheme was made efficient by the use of a modification of the "demon bit" technique used by Bhanot et al.[10] for equilibrium simulation of the three-dimensional Ising model. Our use of this technique is further described in Refs. 8 and 12. The use of this technique allowed our program (except for the generation of demon bits) to be efficiently vectorised using hardware logical vector operations. Using these techniques we achieved a simulation rate of approximately 15×10^6 attempted exchanges per second or about 28.5 MCS per second.

At the beginning of each run our lattice of spins was prepared in a random (infinite-temperature) state, after which spins were randomly flipped until a state consisting of an equal number of up and down spins was achieved. A run then consisted of quenching the system to a temperature below T_c and conducting the simulation. Our simulations consisted of 100 runs of 100,000 MCS at $T = 0.5 T_c$ and 10 runs of 10^6 MCS at $T = 0.5 T_c$. In addition, we have conducted one additional run of 10^6 MCS at $T = 0.3 T_c$. Fig. 1 shows a typical configuration of the lattice during the course of a run.

Data taken included the spin pair-correlation function $g(r;t) = \langle s(0,t) s(r,t) \rangle$, the two-dimensional structure factor $S(k,t) = \langle |(1/N) \sum s(r_i,t) \exp(ik \cdot r_i)|^2 \rangle$, the energy per spin

$E(t)$ and the fluctuations of these quantities. The domain size $R(t)$ was monitored using four different measures of domain size. From the energy we obtained the measure, (equivalent to the inverse perimeter density) $R_E(t) = 2 J / (E(t) - E_0)$ where E_0 is the equilibrium energy per spin at temperature T . From the zero-crossing of the pair-correlation function along the axes and diagonals we obtained the measure $R_G(t)$. From the first and second moments $k_1(t)$ and $k_2(t)$ of the circularly-averaged structure factor we obtained the measures $R_1(t) = 2\pi/k_1(t)$ and $R_2(t) = 2\pi/\sqrt{k_2(t)}$.

3. Results

Figure 3 shows a log-log plot of $R_G(t)$ and $R_E(t)$ from the 10^6 -MCS data. We see that even at 10^6 MCS the slope of the curves is increasing, indicating an increasing effective exponent $n_{\text{eff}}(t)$ with time. Similar behavior was found for the other two measures of domain size $R_1(t)$ and $R_2(t)$. Figure 4 shows the behavior of $n_{\text{eff}}(R_E)$ as a function of $1/R_E$. The data fit very well the form of Eq.4 with an asymptotic value of $n_{\text{eff}}(\infty) = 0.326$. Again similar plots were obtained for the other three measures of domain size. Table 1 summarizes our results for the extrapolated long-time exponents using fits of the form of Eq.4 to our data. We see that our data all strongly support an extrapolated long-time exponent close to $1/3$ - except possibly for R_2 - which is strongly affected by noise in the tail of the structure factor, and for which the extrapolated value is somewhat *below* $1/3$. Averaging over the other three measures of domain size for the 100-run data we obtain the estimated long-term exponent of $n(\infty) = .330 \pm .005$.

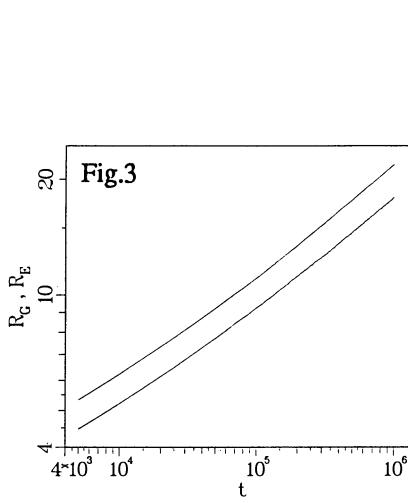


Fig.3 Log-log plot of $R_G(t)$ and $R_E(t)$, at intervals of 5000 MCS up to 10^6 MCS, averaged over 10 runs. Upper curve is $R_G(t)$.

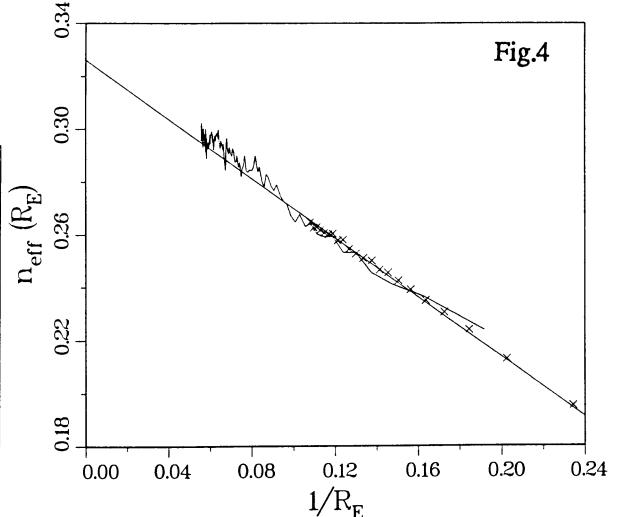


Fig.4 Plot of effective exponent $n_{\text{eff}}(R_E)$ vs $1/R_E$. Crosses correspond to averages over 100 runs (100,000 MCS) while solid curve corresponds to 10 runs of 10^6 MCS. Straight line shown is least-squares fit to the 100-run data, with a y-intercept of $0.326 \pm .003$.

Table 1 Extrapolated long-time exponents at $T = 0.5 T_c$ using least-squares fits of the form $n_{\text{eff}}(1/R) = D - C/R$ to data

	100 runs	10 runs
$n_{\text{eff}}(R_E)$	0.326 ± 0.003	0.337 ± 0.001
$n_{\text{eff}}(R_G)$	0.328 ± 0.0008	0.336 ± 0.001
$n_{\text{eff}}(R_1)$	0.336 ± 0.0007	0.344 ± 0.0009
$n_{\text{eff}}(R_2)$	0.310 ± 0.0008	0.315 ± 0.0008

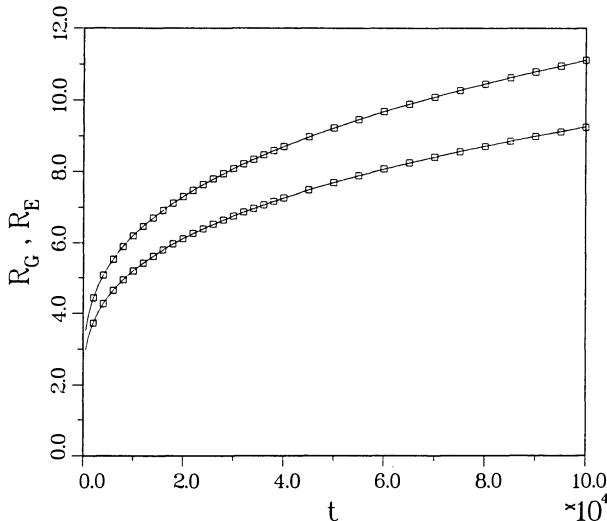


Fig.5 Plot of $R_G(t)$ and $R_E(t)$ for 100-run data with accompanying fits of form $R(t) = A + B t^{1/3}$ (solid lines). Upper curve is $R_G(t)$.

Figure 5 shows fits of the form $R(t) = A + B t^{1/3}$ for $R_G(t)$ and $R_E(t)$ for the 10^5 -MCS data. The fits are surprisingly good, even for early times, and the average relative error of the fits is of the order of the average estimated statistical error for the simulations. We have also conducted one run at $T = 0.3 T_c$ for which fluctuations were surprisingly small. Data for this run, when analysed similarly, also gives a very good fit of the form of Eq.4, with a long-time asymptotic exponent of $.329 \pm .001$.

We have also looked at the scaling of the circularly-averaged structure factor at $T = 0.5 T_c$. We find that scaling does hold over a range of from 4000 MCS to 10^6 MCS. However, the structure factor itself is anisotropic, which is not surprising given the anisotropy of the surface tension below the critical temperature.

Finally we have looked at the variation of the coefficients B using our data for $T = 0.5 T_c$ and $T = 0.3 T_c$. We find rough agreement [8] with a phenomenological expression given in Ref.7, while for higher temperatures the situation is not so clear.

4. Conclusion

Thus, we have shown that, even at 'critical' concentration, the long-time growth exponent for domain growth is $1/3$. Even more strikingly, our data can all be quite accurately fit except for the earliest times (within the accuracy of our simulation) by the simple form $R(t) = A + B t^{1/3}$.

One question which is not yet fully understood is the variation of the long-term growth coefficient B and the correction term C_3 , with temperature and with volume fraction. With this in mind we have conducted preliminary simulations [11] using our code for the two-dimensional spin-exchange kinetic Ising model for volume fractions different from $1/2$. The results of our simulations indicate that for small volume fractions, the contribution due to the "surface diffusion term" C_3 remains rather large. In these simulations, the asymptotic behavior of the form of Eq.4 develops at *later* times, thus requiring possibly *longer* runs than in the case of critical concentration. In addition, further corrections appear to be needed. For low volume fractions, fluctuations are quite large and much better statistics are required.

Recently, two of us (J.A. and F.S.) have developed a program [12] to do Monte Carlo simulations on the Connection Machine 2 at Naval Research Laboratory which is about 9 times faster than our fully optimised code on the Cyber 205. The algorithm for this code is essentially the same as described above, with one 4×4 array of spins allotted to each processor, with 16,384 processors in the machine and is fully described in Ref.12. With this code, we plan to further study the variation of domain growth rate and of correction terms such as C_3 as a function of temperature and volume fraction.

Acknowledgments: This work was done in collaboration with Francis E. Sullivan and Raymond D. Mountain at the National Institute for Standards and Technology.

References

1. S.M. Allen and J.W. Cahn, *Acta Metall.* **27**, 1085 (1979).
2. I.M. Lifshitz and V. V. Slyozov, *J. Phys. Chem. Solids* **19**, 35 (1961).
3. G.F. Mazenko, O.T. Valls, and F. C. Zhang, *Phys. Rev. B* **31**, 4453 (1985); **32**, 5807 (1985).
4. M. Rao, M.H. Kalos, J.L. Lebowitz and J. Marro, *Phys. Rev. B* **13**, 4328 (1976).
5. J.L. Lebowitz, J. Marro, and M.H. Kalos, *Acta Metall.* **30**, 297 (1982).
6. G.S. Grest and D.J. Srolovitz, *Phys. Rev. B* **30**, 5150 (1984).
7. D.A. Huse, *Phys.Rev. B* **34**, 7845 (1986).
8. J.G. Amar, F.E. Sullivan, and R.D. Mountain, *Phys.Rev. B* **37**, 196 (1988).
9. F. Sullivan and R.D. Mountain, in Proceedings of the Sixth International Phoenix Conference on Computers and Communications, Scottsdale, AZ (1987). IEEE Computer Society Press, IEEE Computer Society, 1730 Massachusetts Ave. NW, Washington, D.C. 20036
10. G. Bhanot, D. Duke, and R. Salvador, *J. Stat. Phys.* **44**, 985 (1986).
11. J.G. Amar and R.D. Mountain, unpublished.
12. J.G. Amar and F.E. Sullivan, to be published in *Computer Physics Communications*.

Histogram Methods for Monte Carlo Data Analysis

R.H. Swendsen¹ and A.M. Ferrenberg²

¹Department of Physics, Carnegie-Mellon University, Pittsburgh, PA 15213, USA

²Center for Simulational Physics, Department of Physics and Astronomy,
University of Georgia, Athens, GA 30602, USA

Abstract. In this talk, I will discuss new methods for optimizing the analysis of data from Monte Carlo computer simulations. The heights and locations of peaks in the specific heat and the susceptibility can be found with high accuracy for relatively short simulations. For multiple simulations to cover a wide range of values, explicit error estimates allow objective planning of the lengths of runs and the parameter values to be simulated. Since information can cover a large temperature range, it is possible to include both zero and infinite temperature to generate the free energy and entropy over the full range. The methods are applicable to simulations in a variety of fields related to statistical mechanics.

1. Introduction

Recently, we showed that histograms can be used to greatly increase the amount of information obtained from a single computer simulation in the neighborhood of a critical point.[1] Oddly enough, the equations necessary for this improvement have been available for many years[2], but it does not seem to have been realized that they could be used as a generally applicable method for determining the the location and height of maxima and minima with higher accuracy.

Even more recently, we have developed a multiple-histogram method[3] to provide information over a wide range of parameters, beyond the region of validity of the single histogram method.

In this talk, I will give a sketch of both methods, along with a discussion of their advantages and limitations.

2. Single Histogram Method

Consider a Hamiltonian $H(\sigma)$, defined on the state σ . In the following, we shall deal with spin models, so that σ represents a set of spins $\sigma = \{\sigma_i\}$, although this is not a restriction of the method. We shall discuss the behavior of the system as a function of $\beta = 1/k_B T$, but the generalization to an arbitrary number of parameters is straightforward. The partition function is given by

$$Z(\beta) = \sum_{\sigma} \exp[-\beta H(\sigma)] = \sum_E W(E) \exp[-\beta E] \quad (1)$$

where $W(E)$ is the density of states.

We normally perform Monte Carlo (MC) simulations at a set of R values $\{\beta_r\}$ and store the data as histograms, $\{N_r(E)\}$, with total numbers of values $\{n_r\}$. For each of

these simulations, we can then calculate the corresponding distributions for neighboring temperatures as

$$P_r(\beta, E) = N_r(E) \exp[-(\beta - \beta_r)E] z_r^{-1}(\beta) \quad (2)$$

where

$$\bar{z}_r(\beta) = \frac{1}{n_r} \sum_E N_r(E) \exp[-(\beta - \beta_r)E] \quad (3)$$

and $\bar{z}_r(\beta)$ is related to the true partition function by

$$\bar{z}_r(\beta) = n_r Z(\beta) / Z(\beta_r) \quad (4)$$

where we have used a bar over an expression to indicate the expectation value with respect to all MC simulations of length n_r .

The free energy is given by

$$-\beta F(\beta) + \beta_r F(\beta_r) = \ln \bar{z}_r(\beta) - \ln n_r \quad (5)$$

The density of states is related to the histogram by

$$W(E) = \bar{N}_r(E) n_r^{-1} \exp[-f_r + \beta_r E] \quad (6)$$

where $f_r = \beta_r F(\beta_r)$ is a parameter evaluated at β_r .

The expectation value of any function of the energy $A(E)$, can then be obtained as a function of β .

$$\langle A(E) \rangle = \sum_E A(E) P_r(\beta, E) \quad (7)$$

We have shown in Ref.1 that these equations allow very accurate determination of the height and location of peaks in the specific heat. For a 16x16 Ising model, we were able to determine the height of the peak to within 0.2% and the location of the maximum to within 0.04% of the exact values.

We also noted that the region of validity of the histogram method coincides with the scaling region, so that information about the critical region does not deteriorate with increasing system size. The maximum shift of the energy scales as the width of the energy distribution, which is related to the specific heat, C . Since the specific heat also relates the change in energy to the corresponding change in temperature, the maximum temperature change can be easily seen to scale as $C^{-1/2} L^{-d/2}$, where L is the size, and d the dimension of the system. At the critical point, C diverges as $L^{\alpha/\nu}$, so that the range of temperatures scales as $L^{-1/\nu}$, which coincides with the scaling region. Similar consideration of other variables shows that this scaling is a general result.

3. Multiple Histogram Method

For more general problems, it is often desired to investigate the behavior of the system over a wide range of temperatures.

The following is an optimized method for combining the data from an arbitrary number of simulations to obtain information over a wide range of parameter values. Specifically, we combine the results of several simulations such as to minimize the error in the final estimate for the density of states. An essential feature of our approach is that the optimization is performed independently for each value of the energy E.

For each MC simulation, errors in the histograms are given by

$$\delta^2 N_r(E) = g_r \overline{N_r(E)} \quad (8)$$

If successive MC configurations are independent, then $g_r=1$, otherwise $g_r=1+2\tau_r$, where τ_r is the correlation time.

We can now combine the histograms to form a general expression which leads to an improved estimate for $W(E)$. This gives us

$$W(E) = \sum_{r=1}^R p_r(E) \overline{N_r(E)} n_r^{-1} \exp[-f_r + \beta_r E] \quad (9)$$

with

$$\sum_{r=1}^R p_r(E) = 1 \quad (10)$$

If we minimize the error in the resultant estimate for $W(E)$, and define

$$P(E, \beta) = W(E) \exp[-\beta E] \quad (11)$$

we obtain the essential multiple-histogram equations.

$$P(E, \beta) = \frac{\sum_{r=1}^R g_r^{-1} N_r(E) \exp[-\beta E]}{\sum_{r=1}^R g_r^{-1} n_r \exp[-\beta_r E + f_r]} \quad (12)$$

where

$$\exp[-f_r] = \sum_E P(E, \beta_r) \quad (13)$$

Any operator can then be evaluated as a function of β using

$$\langle A(E) \rangle = \sum_E A(E) P(E, \beta) / z(\beta) \quad (14)$$

where

$$z(\beta) = \sum_E P(E, \beta) \quad (15)$$

The values of f_r are found self-consistently by iterating Eqs. (12) and (13). Convergence can be greatly accelerated using the derivatives of the new values of f_n as functions of the old values. Note that from the form of Eqs. (12) and (13), an arbitrary constant can be added to all f_n 's without affecting the solution. This constant is determined by evaluating the free energy at zero or infinite temperature, where it is known. Alternatively, it can be set to zero at some reference point, which is useful when applying the method to the calculation of interfacial free energies.

If the method is restricted to two MC simulations, the calculated difference in the free energies between the simulated points is identical to that obtained by Bennett's method.[4,5]

The statistical error in $P(E,\beta)$ is given by

$$\delta P(E,\beta) = \left[\sum_r g_r^{-1} N_r(E) \right]^{-1/2} P(E,\beta) \quad (16)$$

This expression provides a clear guide for planning a series of simulations. The locations and heights of peaks in the relative error, plotted as a function of E , give direct quantitative indications of the optimum locations and lengths of additional MC simulations.

As a practical matter, it is useful to handle most of the calculations in terms of the logarithms of the various quantities in these equations. Also, if there is insufficient storage to save an entire two-dimensional histogram, the magnetization and its square can be stored as a functions of S to determine the magnetic susceptibility. It is sometimes preferable to store MC data in terms of lists rather than histograms, with obvious modifications of the equations.

4. Examples

We have tested this multiple-histogram method against the exact solution of the $d=2$ Ising model with $L=16$. The Hamiltonian is given by

$$H = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j \quad (17)$$

where the spins take on the values +1 and -1, the sum is over nearest-neighbor pairs, and we take $J=1.0$ for convenience.

For temperatures close to the critical temperature, a single simulation is sufficient for high accuracy.[1] However, for temperatures more than about 20% away from the critical temperature, additional simulations are necessary. To improve the accuracy away from β_c , we added simulations at $\beta=0.3$ and 0.64 to the first simulation at $\beta_c = 0.4406868$. Then, using plots of the relative error from Eq.(15) as a guide, we studied the results of adding simulations at $\beta = 0.0, 0.1, 0.2, 0.375$, and 0.525 for a total of eight. We used the Swendsen-Wang algorithm,[6] taking two million MC sweeps at each temperature, except for T_c , where we used 9.5 million sweeps. The value of g_r at T_c was estimated at 6, with smaller values at the other temperatures.

The difference between the calculated and exact values of the specific heat over the full temperature range is less than 0.3%, and would differ from the exact values[7] by less than the width of the line if we were to include a plot with this paper.

The results for the entropy show dramatic improvement as data from additional MC simulations are included in the calculation. The entropy is given by the expression

$$S = \ln z(\beta) + T^{-1} z^1(\beta) \sum_E E P(E, \beta) \quad (18)$$

so that entropy differences can be calculated directly. As an example, consider the d=2 Ising model on an L by L lattice, for which the total difference in the entropy per site between zero and infinite temperature is $(1-L^{-2}) \ln 2$. For L=16, this has the value 0.69044. When the data is restricted to a simulation at the critical temperature, the entropy difference was 0.591, which is off by 15%. Adding data from the simulations at $\beta=0.3$ and 0.64 gave a value of 0.6746, which already reduced the error to 2.3%. Finally, using data from all eight simulations, we obtained 0.69030 with a remaining error of only 0.02%. This is an order of magnitude better than would be necessary to see the L^{-2} term due to the two-fold degeneracy at $T=0$.

In summary the multiple-histogram method provides an optimized combination of data from different sources, produces results in the form of continuous functions for all values of interest, and can be applied to an arbitrary number of simulations. Errors can be calculated and provide a clear and simple guide to optimizing the length and location of additional simulations to provide maximum accuracy.

5. Acknowledgements

We would like to thank H. Park for helpful discussions, and for calling our attention to the relationship between our results and those of Bennett. This work was supported by the National Science Foundation Grant No. DMR-8613218.

6. References

- [1] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. 61, 2635 (1988).
- [2] Z.W. Salsburg, J.D. Jackson, W.Fickett, and W.W. Wood, J. Chem. Phys. 30, 65 (1959); D.A. Chesnut and Z.W. Salsburg, J. Chem. Phys. 38, 2861 (1963); I.R. McDonald and K. Singer, Disc. Far. Soc. 43, 40 (1967), J. Chem. Phys. 47, 4766 (1967) 50, 2308 (1969); J.P. Valleau and D.N. Card, J. Chem. Phys. 57, 5457 (1972); G. Torrie and J.P. Valleau, Chem. Phys. Lett. 28, 578 (1974); M. Falcioni, E. Marinari, M.L. Paciello, G. Parisi, and B. Taglienti, Phys. Lett. 108B, 331 (1982); E. Marinari, Nucl. Phys. B235[FS11], 123 (1984); G. Bhanot, S. Black, P. Carter and R. Salvador, Phys. Lett. B183, 331 (1987); G. Bhanot, K.M. Bitar, S. Black, P. Carter and R. Salvador, Phys. Lett. B187, 381 (1987); G. Bhanot, K.M. Bitar, and R. Salvador, Phys. Lett. B188, 246 (1987).
- [3] A. M. Ferrenberg and R. H. Swendsen, unpublished.
- [4] C.H. Bennett, J. Comput. Phys. 22, 245 (1976).
- [5] H. Park, private communication.
- [6] R.H. Swendsen and J.S Wang, Phys. Rev. Lett. 58, 86 (1987).
- [7] A.E. Ferdinand and M.E. Fisher, Phys. Rev. B 185, 832 (1969).

Theoretical Models for Adsorption and Diffusion of In on Cu(100)

M.S.S. Challa¹, Yi Li¹, M.R. Press¹, S.N. Khanna¹, P. Jena¹, and M. Yousouff²

¹Department of Physics, Virginia Commonwealth University,
Richmond, VA 23284, USA

²Fakultät für Physik, Universität Konstanz, Konstanz, Fed. Rep. of Germany and
Department of Physics, Indian Institute of Technology, Kanpur, India

Using ab-initio molecular cluster calculations we show that In adsorbed on Cu(100) surface occupies a position above the bridge sites of the Cu plane and that the electric field gradient felt by the In atoms is appreciable only at Cu vacancies. To explain data from PAC experiments we propose that the radioactive In probes diffuse to the Cu plane during the decontamination process. Monte Carlo simulations of the In cluster kinetics show this model to be feasible and also that the free In monomer hopping rate is about 30 times greater than previous estimates.

1. Introduction and experimental results

The mobility of atoms on surfaces is of fundamental interest and plays an important role in surface growth and reaction phenomena. Recently the PAC (perturbed photon-photon angular correlation) technique has been developed [1] to accurately determine the adsorption of In impurities on various surfaces. In these experiments, a small quantity of radioactive ^{111}In probes are introduced on to the host surface. When the probes decay to Cd by emitting two photons, the angular correlation between the photons is perturbed by the electric field gradient (efg) produced by the surrounding atoms and thus this correlation is a sensitive test of the local environment of the In probes. Two quantities are principally measured in PAC experiments: V_{zz} , the largest component of the efg tensor and the asymmetry parameter, η , which describes the orientation of the efg tensor. For In on a clean Cu(100) surface, it has been established [2] that $V_{zz} = +1.0 \times 10^{18} \text{ V/cm}^2$ and that $\eta = 0$. This value of η shows unambiguously that In atoms occupy an axially symmetric (4-fold) site on the Cu surface.

Another set of PAC experiments have been conducted recently by Klas et al [3] to examine the temperature dependence of the clustering of In impurities on the Cu(100) surface. These authors have found 3 types of signals (V_{zz} 's) denoted by f_0 , f_1 and f_2 and their experimental results are reproduced in Fig. 1. Klas et al [3] have interpreted the above signals as arising from probes with 0 In neighbors, 1 In neighbor and more than 1 In neighbor respectively. Thus, Fig. 1 shows that at low temperatures most of the In atoms exist as part of large clusters which break up as the temperature increases.

Previous theoretical work has yielded contradictory pictures. Using ab-initio molecular cluster calculations, Lindgren [4] found that the experimental results for V_{zz} are explained well only if the In atoms occupied the Cu vacancies; V_{zz} rapidly diminished to zero when the In atoms moved out of the Cu plane. On the other

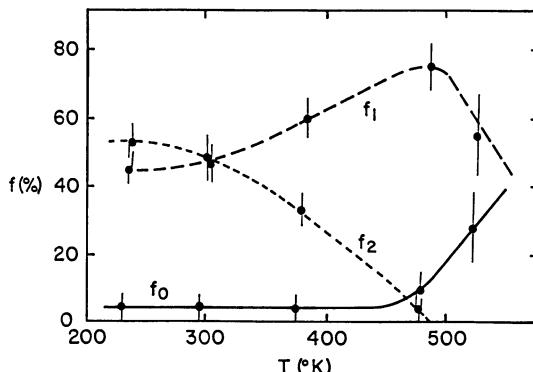


Fig.1 Experimental results on the variation of the f_0 , f_1 and f_2 signals with temperature (from [3]).

hand, Blender et al. [5] (hereafter referred to as BDFY) fitted the kinetics results reasonably well by simulating the diffusion of the In atoms in a plane above the Cu plane. To resolve this contradiction, we show here results from two types of calculations. First, we used ab-initio calculations to determine not only the efg's but also the binding energies of In at the allowed sites since it is the energetics which determines the adsorption sites. Then we used Monte Carlo simulations of the kinetics to show the feasibility of our model.

2. Ab-initio cluster calculations

We modelled the Cu(100) surface by configurations of up to 12 Cu atoms confined to the bulk geometry. As shown in Fig. 2, there are 3 allowed 4-fold sites for the chemisorption of the In atom; we refer to these as substitutional, on-top and bridge sites. For each configuration, the In atom is at a variable distance d above the Cu plane. For various values of d the molecular orbitals of the In-Cu cluster and the energies were evaluated in a self-consistent fashion using linear combination of atomic orbitals and a local density approximation to density functional theory. The charge density computed from the molecular orbitals was then used to calculate V_{zz} at the In position. (Full details are given in reference 6). For each configuration, the equilibrium site for In was identified by the value of d with the highest In binding energy and the results are summarized in Fig. 2. We conclude from the energetics that the bridge site is clearly preferred by In with $d \approx 3.25$ Å. We shall henceforth refer to the plane of these adsorption sites as the top plane.

Comparing our efg results (see Fig. 2) with those of Lindgren [4], we find qualitative agreement in that V_{zz} is large and positive only at the substitutional site. The numerical differences with Lindgren's results, however, are not unexpected since estimates of the 2nd-derivatives of the potential are involved and the results are influenced by a number of parameters (including even the numerical integration scheme). After detailed examination of several other types of clusters [6], we concluded that the PAC signals arise from In atoms at Cu vacancies.

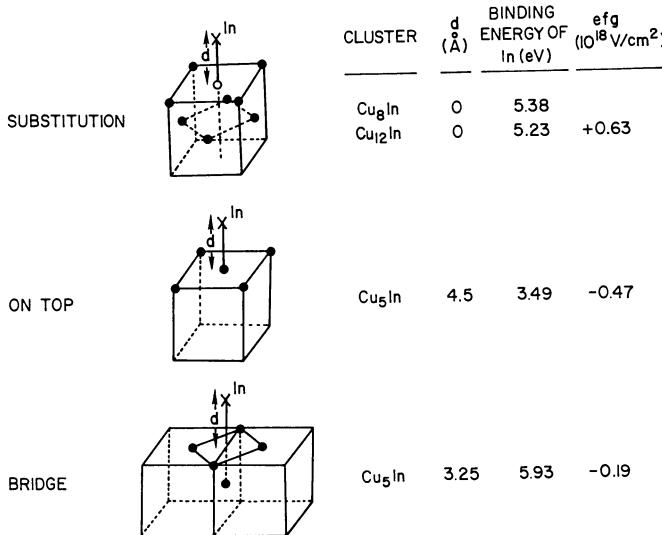


Fig. 2

The possible 4-fold adsorption sites for In on Cu(100) and the geometries used in the cluster calculations. d is the height of In above the Cu plane. Filled circles represent Cu atoms while the open circle in the substitutional site geometry indicates a Cu vacancy. For each geometry, the ab-initio results corresponding to the maximum binding energy are shown. The substitutional-site results for the Cu₁₂In cluster show that size effects are small in energy calculations.

3. Our model for the cluster kinetics

To explain the cluster data of Klas et al. [3] (Fig.1), it is necessary to examine the experimental set up more closely. In these experiments, less than 1% of a monolayer of ¹¹¹In probes were deposited on to the Cu surface which was then heated to 600°K to desorb the chlorine contaminants. The system was then cooled to 77°K and then 7% of a monolayer of normal In atoms were introduced. The signals were then monitored after annealing at various higher temperatures.

We conjecture that the probes, although in the top plane initially, diffuse to the energetically less-favorable Cu vacancies during chlorine desorption; the probes are then rendered immobile by the low concentration of vacancies. The normal In atoms (added at low temperatures), however, remain in the top plane during most of the experiment and are free to diffuse in this plane. Thus the signals arise from the normal In atoms "seen" by the probes. This is markedly different from the BDFY model where the probes and the normal In atoms behave identically; the problem with their picture is that diffusion in the Cu plane requires a very large number (>50%) of Cu vacancies.

To reproduce the data of Fig. 1, we performed Monte Carlo simulations as follows. The Cu(100) surface was represented in the computer by a square array (200x200) of lattice points and 1% of

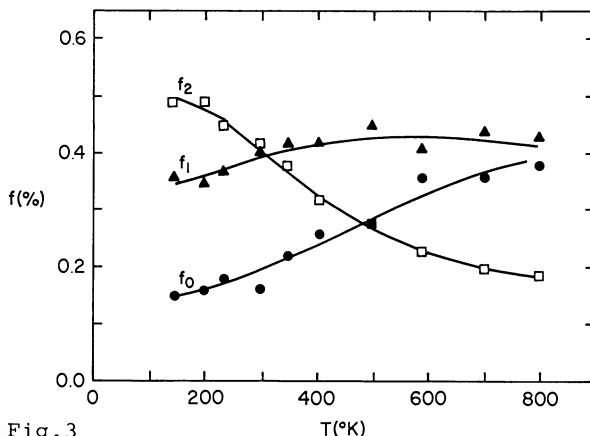


Fig. 3

T(°K)

Results for the temperature-variation of $\{f_i\}$ obtained from Monte Carlo simulations of the kinetics. The temperature scale obtained by fitting to the experimental data of Fig. 1 (see text).

these were occupied randomly by the probes. The probes were then fixed at these positions for the duration of the simulation. An identical array was used to represent the top plane and 7% of these sites were occupied randomly by the normal In atoms. In the top plane, two diffusion rates govern the hopping of a particle: (1) if the particle is isolated it hops randomly to a vacant nearest-neighbor site with probability unity and (2) if any of the nearest neighbor sites are occupied it hops with probability β ($10^{-4} < \beta < 1$). Attempting to move every atom in the top plane once constitutes a Monte Carlo step/site (MCS). The signals are determined by the occupancy of those sites in the top plane which are nearest-neighbors of the probes. For a given value of β , typical runs consisted of 10^5 MCS with the $\{f_i\}$ monitored after every 500 MCS; the averaged values of the $\{f_i\}$ during a run were then studied as a function of β .

To convert β to temperature, T , we adopted the phenomenology used by BDFY, i.e., we assumed that $\beta = \beta_0 \exp(-E_b/kT)$ and obtained $\beta_0 = 0.45$ and $E_b = 0.077$ eV by fits to the experimental data. (In comparison, BDFY obtained $\beta_0 = 5.7$ and $E_b = 0.17$ eV). Our results are shown in Fig. 3 and, while we found the quality of the fits to be comparable to those of BDFY, we estimate a much larger free-monomer hopping rate, α . By comparing results from simulation and experiment for the rate of decay of f_0 , we found that $\alpha = 0.0083$ s⁻¹ at 180°K and that the migration energy, E_a , was 0.47 eV. In comparison, BDFY obtained $\alpha = 0.0003$ s⁻¹ and $E_a = 0.52$ eV.

4. Summary and future research

In conclusion, our results show that the In are preferentially adsorbed at the bridge sites while the efg is significant only at the Cu vacancies. We explain the experimental results by conjecturing that the In probes diffuse to the Cu vacancies during the decontamination process which is carried out at high

temperatures. While the agreement between our simulation results for the cluster kinetics and PAC data is comparable to that obtained by BDFY [5], an important prediction of our simulations is that the hopping rate for the In monomers should be much larger than the previous estimate.

An experimental test of our model can be realized by repeating the experiments of Klas et al [3] but without adding the normal In atoms. In such a set-up, we predict that there should be little variation of the $\{f_i\}$ in the range of temperatures of Fig. 1. On the theoretical side, we are formulating model potentials for In on Cu in order to determine the monomer hopping rate by numerical integration of the dynamics. Another possibility is to refine the model kinetics to include hopping rates which depend on the number of neighbors.

5. Acknowledgements

This work was supported by a grant from the Army Research Office (DAAL03-89-K-0015).

References

1. W. Körner, W. Keppner, B. Lehndorff-Junges and G. Schatz, Phys. Rev. Lett., 49, 1735 (1982).
2. T. Klas, J. Voigt, W. Keppner, R. Wesche and G. Schatz, Phys. Rev. Lett., 57, 1068 (1986).
3. T. Klas, J. Voigt, R. Platzer, R. Wesche and G. Schatz, Hyper. Inter., 34, 577 (1987).
4. B. Lindgren, Hyper. Int., 34, 217 (1987).
5. R. Blender, W. Dietrich, H. L. Frisch and M. Yussouff, Z. Phys. B, 71, 103 (1988).
6. Yi Li, M. R. Press, S. N. Khanna, P. Jena and M. Yussouff, submitted to Phys. Rev. B.

Index of Contributors

- | | | |
|----------------------|-------------------|-------------------------|
| Amar, J.G. 173 | Khanna, S.N. 184 | Schüttler, H.-B. 1, 157 |
| Burkitt, A.N. 16 | Klein, R. 169 | Stauffer, D. 26 |
| Challa, S.S. 184 | Lai, Pik-Yin 163 | Stewart, D.E. 68 |
| Chen, C.-X. 157 | Landau, D.P. 1 | Sugar, R.L. 116 |
| Ferrenberg, A.M. 179 | Leung, K.-t. 152 | Swendsen, R.H. 179 |
| Frisch, H.L. 99 | Li, Yi 184 | Vallés, J.L. 152 |
| Gallion, S.L. 68 | Lin, M.Y. 169 | Vicsek, T. 34 |
| Gerling, R.W. 85 | Lindsay, H.M. 169 | Wampler, J.E. 68 |
| Heermann, D.W. 16 | Meakin, P. 169 | Weitz, D.A. 169 |
| Herrmann, H.J. 56 | Mon, K.K. 1 | Yousouff, M. 184 |
| Jarrell, M. 137 | Novotny, M.A. 45 | Zia, R.K.P. 152 |
| Jena, P. 184 | Press, M.R. 184 | |
| | Rapaport, D.C. 4 | |