Symposium de superinformatique, Montréal, QC, 4-6 juin 1990

Monte Carlo method for strongly interacting electrons

A.-M.S. Tremblay

C. Boily, L. Chen, H. Nélisse, A. Reid

Département de physique et

Centre de recherche en physique du solide

Université de Sherbrooke

Sherbrooke, Québec J1K 2R1

Abstract

The discovery of high temperature superconductors has given strong impetus to the study of strongly interacting electrons moving on two-dimensional lattices. The simplest model for this kind of problem, the Hubbard model, is introduced. The computational diffculty of this deceptively simple looking problem is then pointed out. The most efficient numerical approach for it, the Monte Carlo method, is described and its main shortcomings are discussed: a) The fermionic nature of electrons leads to non-positive weights for the Monte Carlo; b) The low temperature limit leads to ill-conditioned matrices. The last of these problems has been solved, but the first one is still outstanding. Nevertheless, the approach works well for many cases of interest. A short summary of our results and references to the literature are given.

1. Introduction: The Hubbard model as a computational problem.

The discovery of high temperature superconductors has motivated renewed interest in the problem of strongly interacting electrons moving on two-dimensional lattices. This is a problem which is well suited for this Supercomputing Symposium. In this talk, we intend to present this problem as a computing challenge, without going into the physical motivation for studying it. We will also quote only a few results at the very end. After introducing the model and its inherent difficulties, we will proceed in section 2 to present the Monte Carlo approach to this problem which has been developed mainly in California during the last decade. We will describe the general approach, give reference to the original literature and point to the main computational diffculties. We have developed a code in Sherbrooke which seems competitive with the best available ones (less than ten) in the world and which is currently being used to study various physical aspects of the problem. We have used the principles described in the literature to develop the code but the implementation (a few thousand lines) is entirely ours.

a) Computing environment: CPU time is the limiting factor in obtaining results. Memory and input-output requirements are modest. Typically, after one hour of computation on the CRAY, we have to transfer a few kilo-bytes of data at most. The telecommunication links are DATAPAC and BITNET. The programs have been developed on IBM-PC's with 0.2 Mflops DEFINICON numeric cards. The results are usually analyzed and plotted on that equipment. A SUN Sparc Station 1 has recently been added to our equipment. It allows quicker program development mainly through its increased speed and debugging tools. The larger memory and execution speed also allows us to run realistic test cases locally. The Cray is however essential to obtain publishable results in

a reasonable time.

b) The problem: Assume a two-dimensional lattice of N sites. Each site can be either empty, occupied with an electron of spin up, occupied with an electron of spin down, or occupied by two electrons of opposite spins. This means that each site can be in one of 4 states and hence that for N sites, the many-electron state space has a size 4^N. For N = 16, meaning a 4x4 lattice, this already means 2^{32} states (one Megabyte is about 2^{20} bytes). The Hubbard Hamiltonian defines the matrix which must be diagonalized in this state space. Grouping states with a fixed number of electrons block diagonalizes this Hamiltonian. Only fully antisymmetric state vectors are physically allowed (Pauli principle). Nevertheless, for finite densities of electrons this state space is still enormous, and there are non-zero off-diagonal matrix elements between any pair of vectors which differ only by spin preserving displacements of one electron by one lattice spacing. State vectors where two electrons of the same spin occupy the same state are excluded by the Pauli principle. Physically, the Hubbard Hamiltonian contains kinetic energy and short range repulsion. This means that double occupancy of electrons with opposite spins is given a diagonal contribution U times the number of doubly occupied sites, while off-diagonal matrix elements are -t when there is a spin preserving electron hop leading from the column state vector to the row state vector. The sign of these off-diagonal matrix elements is also influenced by antisymmetry. The Hamiltonian matrix is thus sparse but has no simple structure and the state space is exponentially large. In practice then, exact diagonalizations must heavily rely on symmetries to simplify the problem, and they cannot be done for systems much larger than 4x4. Furthermore, the complexity of the problem clearly increases exponentially with size. By contrast, with the Monte Carlo approach one does not need to store the whole

state space and, furthermore, the computational task increases only as a polynomial of the lattice size (and inverse temperature). There exists other approaches[1] to Quantum Monte Carlo problems such as Monte Carlo Green's functions or world line algorithms but none of these will be described here.

2. BBSH Monte Carlo algorithm for the Hubbard model.

This section is a summary of the algorithm in its present state. The main ideas originated with Blankenbecler, Scalapino and Sugar.[2] This approach was extensively applied to the Hubbard model and further developed by Hirsch.[3] White et al.[4] recently found the solution to the long standing problem of low temperature calculations. Loh and Gubernatis[5] have recently written a thorough review of the algorithmic details. Important results obtained with this algorithm have been recently reviewed by Scalapino.[6] This section is in a sense only a brief survey of these references. Part a) below is a very general approach which is used in almost all computer simulations of quantum problems. After parts b) and c), the reader will realize how the problem has been reduced to a simple trace over classical bi-valued variables. Part d) shows how physical observables are in principle computed. Part e) will make clear what is the remaining algorithmic challenge for this approach: the so-called sign problem. In f), the full Monte Carlo updating cycle is described, and finally, g) contains ideas which are of more general applicability, in particular ideas on stabilizing products and inversions of matrices by using orthogonalization.

a) Trotter formula (path integral formulation).

The quantum statistical problem one faces is to compute physical observables $\hat{O}$ using the formula:

$$\langle O \rangle = \mathrm{Tr}[\exp(-\beta(\hat{H}-\mu\hat{N}))\hat{O}] \,/\, \mathrm{Tr}[\exp(-\beta(\hat{H}-\mu\hat{N}))] \qquad (1)$$

where $\hat{H}$ is the Hubbard Hamiltonian described above, $\hat{N}$ gives the number of electrons in the state on which it acts while the trace is over the state space described in the introduction. $\hat{H} - \mu\hat{N}$ contains a diagonal part $\hat{V}$ which does not commute with the off-diagonal part $\hat{K}$. $\hat{K}$ or $\hat{V}$ by themselves though are easily diagonalized. There is thus an advantage to separate these two terms using the Trotter decomposition (first suggested by M.Suzuki in the context of quantum Monte Carlo problems):

$$e^{-\beta(\hat{H}-\mu\hat{N})} = \prod_{\ell=1}^{N_\tau} e^{-\Delta\tau(\hat{K}_\ell+\hat{V}_\ell)} \approx \prod_{\ell=1}^{N_\tau} e^{-\Delta\tau\hat{K}_\ell}\, e^{-\Delta\tau\hat{V}_\ell} + O(\Delta\tau^2) \qquad (2)$$

where $N_\tau\Delta\tau \equiv \beta$ and N_τ is the number of so-called "imaginary-time slices". Formally the error is of order $\Delta\tau$, but by careful consideration of the complete problem, Fye[7] and Suzuki[8] have shown that the error is indeed of order $\Delta\tau^2$. The operators $\hat{K}_\ell$ are the same for every value of ℓ, but we have now effectively added N_τ lattice points for every original spatial dimension of our problem. We now effectively have a lattice in imaginary-time and real space.

b) Hubbard-Stratonovich decoupling of the interactions.

The interacting part of the Hamiltonian, which is proportional to U and contained in V, can be now written in a form which commutes with K by introducing one bi-valued classical variable s_ν for every point of the imaginary-time and space lattice. This is the so-called discrete Hubbard Stratonovich transformation introduced by Hirsch.[8] In the case of a

repulsive potential, it takes the form:

$$e^{-\Delta\tau U(\hat{n}_{\nu+}-\frac{1}{2})(\hat{n}_{\nu-}-\frac{1}{2})} = e^{-\Delta\tau U/4} \frac{1}{2} \sum_{s_\nu = \pm 1} e^{\lambda s_\nu(\hat{n}_{\nu+} - \hat{n}_{\nu-})} \tag{3}$$

with + refering to up spins and - to down spins and where λ is defined by $\cosh \lambda = e^{\Delta\tau U/2}$.

c) Trace over fermion degrees of freedom.

The quantum mechanical part of the problem is now left in a trace which can be performed exactly because the electrons are not interacting with each other anymore. They are interacting only with the Hubbard-Stratonovich field s_ν. In other words, if we can do the following trace for + or - electrons (take $\sigma = \pm$) for any operator $\hat{0}$,

$$\mathrm{Tr}\left\{ \prod_{l=1}^{N_\tau} e^{-\Delta\tau\hat{K}^\sigma} e^{\Delta\tau\mu\hat{N}^\sigma} e^{\lambda\sigma\hat{N}^\sigma s_\nu} \hat{0} \right\} \tag{4}$$

then we are left only with a trace over what we will call from now on the field s_ν, and this can be handled with standard importance-sampling techniques. Let us first do the case where $\hat{0}$ is the identity. The general problem is usually treated with Grassmann variables,[2] but we sketch instead the more accessible proof given by Hirsch.[8] Using spinless fermions to simplify the notation, one first notes that

$$e^{-\Delta\tau\hat{K}} = e^{-\Delta\tau \sum c_i^+ K_{ij} c_j} = e^{-\Delta\tau \sum c_\alpha^+ D_\alpha c_\alpha} \tag{5}$$

where the sum is over repeated indices while the creation-annihilation

operators c_α^+ and c_α are simply the c_i^+ and c_i in the basis where K_{ij} is a diagonal matrix with eigenvalues D_α. In that basis,

$$\mathrm{Tr}\left\{e^{-\Delta\tau \sum c_\alpha^+ D_\alpha c_\alpha}\right\} = \mathrm{Tr}\left\{\prod_\alpha e^{-\Delta\tau\, c_\alpha^+ D_\alpha c_\alpha}\right\} = \prod_\alpha \mathrm{Tr}\left\{e^{-\Delta\tau\, c_\alpha^+ D_\alpha c_\alpha}\right\}$$

$$= \prod_\alpha \left(1 + e^{-D_\alpha} \right) = \det\left(I + e^{-K} \right) \qquad (6)$$

where the last result, just as the trace, is independent of the basis in which it is computed. This result can be generalized to many products of exponentials so that the final result is

$$\mathrm{Tr}\left\{\prod_{\ell=1}^{N_\tau} e^{-\Delta\tau \hat{K}^\sigma}\, e^{\Delta\tau\mu\hat{N}^\sigma}\, e^{\lambda\sigma\hat{N}^\sigma s_\nu}\right\} = \det\left[I + B^\sigma_{N_\tau} B^\sigma_{N_\tau - 1} \cdots B^\sigma_1 \right] \qquad (7)$$

where, the B^σ_ℓ are defined by,

$$B^\sigma_\ell = e^{-\Delta\tau K}\, e^{V^\sigma_\ell}$$

$$\begin{cases} K_{ij} = -t \text{ when } i \text{ and } j \text{ are near n ighbour and } 0 \text{ otherwise.} \\ (V_{ij})^\sigma_\ell = \delta_{ij} \left[\sigma\lambda s_{i\ell} + \Delta\tau\mu \right] \end{cases} . \qquad (8)$$

d) Computation of obsevables.

The problem now involves non-interacting electrons moving in a space and time dependent field. All observables (i.e. $\hat{O}$ different from unity in (4)) can be computed by using Wick's theorem for every space-time configuration of the fields s_ν and then taking the trace over these s_ν. This means that any observable can be computed once we know that the Green's functions take the form:

$$\langle c_{i\sigma}(\ell_1) c^{+}_{j\sigma}(\ell_2)\rangle = \left[B^{\sigma}_{\ell_1} B^{\sigma}_{\ell_1-1} \cdots B^{\sigma}_{\ell_2+1} \frac{1}{I + B^{\sigma}_{\ell_2} \cdots B^{\sigma}_{1} B^{\sigma}_{N_\tau} \cdots B^{\sigma}_{\ell_2+1}} \right]_{ij} \quad (9)$$

e) Weight for importance sampling.

Denoting the determinant in Eq.(7) by $\det(M^{\sigma})$, the trace which must be performed to obtain observables is

$$\langle \hat{O} \rangle = \frac{\sum_{s_\nu} \det[M^{+}]\det[M^{-}] O(s_\nu)}{\sum_{s_\nu} \det[M^{+}]\det[M^{-}]} \quad (10)$$

where $O(s_\nu)$ is the expression for the operator after the quantum mechanical trace has been performed. As in usual Monte Carlo approaches, the trace over the variables s_ν is evaluated by sampling. The product of determinants could be used as a weight for importance sampling, were it not for the fact that it can sometimes be negative (reflecting the antisymmetric nature of the original state vectors). Nevertheless, one can go quite far in using the absolute value of the product of the determinants as a weight for the importance sampling while keeping track of the sign of that product in the observables (including the denominator!). The variables s_ν are first chosen at random, then they are flipped one at a time. A change of one of the s_ν is accepted with a probability given by

$$P = R^{+}R^{-}/(1+R^{+}R^{-}) \quad ; \quad R^{\pm} = \left| \det[M^{\pm}]^{new} / \det[M^{\pm}]^{old} \right| . \quad (11)$$

While exact in principle, this approach runs into problems when the average sign of the product of determinants becomes small, simply because the sampling of state space becomes inefficient (there are large fluctuations from one

realization of the s_ν field to the other.)

f) A full updating cycle.

Let us first define,

$$G^\sigma = \begin{bmatrix} I & 0 & \dots & B^\sigma_{N_\tau} \\ -B^\sigma_1 & I & \dots & 0 \\ 0 & \dots & \ddots & \\ 0 & \dots & I & 0 \\ 0 & \dots & -B^\sigma_{N_\tau - 1} & I \end{bmatrix}^{-1} \qquad (12)$$

This is the Green's function in imaginary time and space. Note that $\det[M^\sigma] = \det[(G^\sigma)^{-1}]$ and that the quantity defined in (9) and from which one can compute observables is simply a matrix element of the G^σ defined in Eq.(12). Hence Eq.(12) contains all the information we need. Let us see how it is changed when one Hubbard-Stratonovich field s_ν is updated. For this, following Fye and Hirsch,[10] it is easiest to define the intermediate quantity $G'^\sigma = [\ (G^\sigma)^{-1} e^{-V'^\sigma}\ ]^{-1}$. When one field variable s_ν is updated the change in G'^σ comes from the change of a single element of the diagonal matrix V^σ. The new matrix G'^σ is obtained from Dyson's equation,

$$G'^\sigma = G'^\sigma - G'^\sigma [\ e^{-V'} - e^{-V}\ ] G'^\sigma \ . \qquad (13)$$

We are now ready to go through a full updating cycle. The spins are updated by scanning the spatial lattice at a fixed imaginary-time and then moving at the next time.

A) One flips a field variable s at space point i, in time slice ℓ. The flip is accepted with a probability which can be computed from Eq.(11) and from

$$\frac{\det M'^\sigma}{\det M^\sigma} = \frac{\det G'^\sigma}{\det G^\sigma} = 1 + (\ 1 - G^\sigma_{ii}(\ell,\ell)\ )\ [\ e^{(V'_{ii})^\sigma_\ell - (V_{ii})^\sigma_\ell} - 1\ ] \qquad (14)$$

The latter equation follows from Dyson's equation (13) and from some standard algebra with determinants.

B) G itself must be updated if the flip is accepted, so as to be able to use (14) again. While all space-time matrix elements of G^σ are modified by a flip, within a time slice one needs only the matrix elements of G^σ which are diagonal in that time index. Given that the matrix elements which are diagonal in time index decouple from all other times in Dyson's Eq.(13), it thus suffices to update the Green's function at all space points in a given time slice. The corresponding equation is easy to deduce from the general Dyson equation (13). Since each space point of the Green's function needs to be updated, this gives order N^2 operations to perform.

C) The steps A) and B) above are repeated for every space point within a time slice, i.e. N times.

D) When all spins in a time slice have been updated, observables are computed. The results are accumulated to be averaged. Partial averages in groups of 100 measurements are also computed to obtain an estimate of the Monte Carlo error from the fluctuations from one group of 100 measurements to the other (which are usually statistically independent). (To help remove statistical dependencies, every other complete space-time lattice update is done without measurement.)

E) To move to the next time slice, one needs the Green's function which is diagonal in that new time index. That new Green's function is obtained either from

$$G^\sigma(\ell+1,\ell+1) = (B_\ell^\sigma)\, G(\ell,\ell)\, (B_\ell^\sigma)^{-1} \qquad (15)$$

or, for reasons of numerical stability, from a computation of G^σ from scratch using its definition (9) with $\ell_1 = \ell_2$ (there are no B_ℓ^σ matrices in the

numerator in that case). As discussed in White et al.,[4] it is necessary to do the latter after one has moved roughly a number of time slices N_m equal to one unit of β, divided by $\Delta\tau$. Furthermore, at low temperature, that recomputation from scratch must be done with great care, as discussed below.

We are now in a position to see that the computation time scales as the number of time slices N_τ times N^2 times N, the latter factors coming from the Green's function update at each space point within a time slice. In a typical calculation, the initial s_ν field configuration is random and one third to one-half of the time is spent sweeping the lattice without measurement, just to reach equilibrium. It seems that shorter equilibration times could be used without damaging the calculation.

g) Improvements: low temperature stabilization and sparse matrices.

The single particle energies of the Hubbard model range from -4t to 4t even for the non-interacting (U = 0) case. When β = 10/t say, i.e. at low temperature, this means that the exponentials where β times energy appear are spread over a very wide range, in turn leading to numerical instabilities. The problem is in the computation of quantities such as $B^\sigma_{N_\tau} B^\sigma_{N_\tau-1} \cdots B^\sigma_1$, which appear in the Green's functions. Following a previously known procedure which was adapted to the present problem by White et al.,[4] N_m matrices B, starting with B_1, are multiplied together, then the columns of the resulting product, say C, are orthogonalized following a modified Gram-Schmidt procedure. This allows C to be written in the form, UDV where U is the matrix of orthonormalized column vectors, D is their norm and V is an upper triangular matrix with unit elements on the diagonal. The different numerical scales are in D. To multiply the following set of N_m matrices B, one first multiplies them from the left with the previous U, then by D. At this point, the

numerical scales are still separate (Each column has a scale determined by D). One can thus orthogonalize again, obtaining U'D'V'. The last V' can multiply the previous V from the left in a stable manner, and we are left with a UDV decomposition which allows us to continue that process untill all B's have been multiplied. Adding the 1 which appears in the Green's function definition (9), one makes a further UDV decomposition before inverting. Indeed, U and V can be inverted stably, and all the different numerical scales are in D, which can be inverted trivially. Let us note that this procedure allows computation of equal-time observables. For unequal-time observables these ideas must be generalized, as has been described by White et al.[4] The latter generalization is not however as stable as the former. The stabilization allows one to go to arbitrarily low temperature, but on the other hand the Gram-Schmidt procedure is time consuming, (order N^3) and it must be done a number of times which scales as N_τ, leading to an algorithm which now scales as $N_\tau{}^2 N^3$. However, the stabilization can be speeded up by storing UDV decompositions, as described by White et al.[4]

A final trick which speeds up the computation, is to work with sparse B matrices. Indeed, despite the fact that K is sparse, $e^{-\Delta\tau K}$ is not a sparse matrix, and unless something is done, multiplication by these full matrices, which scales as N^3, is time consuming. The trick is to write K as the sum of four matrices: For each lattice direction, one writes K as the sum of two matrices, one for even bonds, and one for odd bonds. Each of the resulting matrices is in separated blocks of 2x2 matrices, meaning that the exponential of any of these four matrices stays sparse. An error is introduced when $e^{-\Delta\tau K}$ is written as a product of four exponentials but following the same ideas as those described at the beginning, this error is of order $\Delta\tau^2$, which is anyway the order to which we are working. The time saving in fact allows one to work

with smaller $\Delta\tau$, and this rapidly makes up for the inaccuracy.

On this subject, it should be clear to the reader that the $\Delta\tau^2$ error is a systematic one which is easy to control by making plots of any observable quantity[7] as a function of $\Delta\tau^2$. These plots allow extrapolation and are straight lines for surprisingly large values of $\Delta\tau$.

3. Some results and research program.

We have studied the jump in the momentum distribution of interacting electrons in one dimension.[11] In so doing, we have been able to resolve a controversy concerning the existence or not of a Fermi liquid in one dimension. We are presently completing a study of the tendency to antiferromagnetism in models for highly anisotropic organic conductors which should be able to explain the superconductivity in these compounds.[12] We[13] are also studying the normal state resistivity of the two dimensional model. If the Physics of high-temperature superconductivity is in the Hubbard model, as has been repeatedly suggested , we should be able to reproduce the observed linear temperature dependence of the normal state resistivity of high temperature superconductors. To do this study, we have to use recently developed techniques for the extraction of real-time results from imaginary-time Monte Carlo data.[14]

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada, and the Ministère de l'éducation du Québec. A computing time grant from Cray Research Canada, and technical assistance from the OCLSC staff in Toronto is also gratefully acknowledged.

REFERENCES

(1) *Monte Carlo Methods,* by M.H. Kalos, P.A. Whitlock, (Wiley, N.Y., 1986); *Quantum Monte Carlo Methods,* ed. M. Suzuki, (Springer Verlag, Berlin, 1987); *Quantum Simulations of Condensed Matter Phenomena,* ed. J.D. Doll, and J.E. Gubernatis (World Scientific, Singapore, 1990).

(2) R. Blankenbecler,, D.J. Scalapino, and R.L. Sugar, Phys. Rev. D24, 2278 (1981).

(3) J.E. Hirsch, Phys. Rev. B31, 4403 (1985).

(4) S.R. White, D.J. Scalapino, R.L. Sugar, E.Y. Loh, J.E. Gubernatis, R.T. Scalettar, Phys. Rev. B40, 506 (1989).

(5) E.Y. Loh, and J.E. Gubernatis, "Stable numerical simulations of models of interacting electrons in Condensed-Matter Physics" in *Electron Phase Transitions*, ed. W. Hauke, and Y.V. Kopaev, (Elsevier, N.Y. 1990).

(6) D.J. Scalapino, "Results form Numerical Simulations of the 2D Hubbard Model", UCSB preprint UCSBTH-89-67.

(7) R.M. Fye, Phys. Rev. b33, 6271 (1986); R.M. Fye and R.T. Scalettar, Phys. Rev. B36, 3833 (1987)

(8) M. Suzuki, Phys. Lett. 113A, 299 (1985).

(9) J.E. Hirsch, Phys. Rev. B28, 4059 (1983).

(10) R.M. Fye, and J.E. Hirsch, Phys. Rev. B38, 433 (1988).

(11) C. Bourbonnais, H. Nélisse, A. Reid, and A.-M.S. Tremblay, Phys. Rev. B40, 2297 (1989).

(12) C. Bourbonnais, C. Boily, L. Chen, H. Nélisse, A. Reid, and A.-M.S. Tremblay, in preparation.

(13) P. Bénard, and A.-M.S. Tremblay, in preparation.

(14) S.R. White, "Dynamical pair susceptibilities for the two-dimensional Hubbard Model", University of California at Irvine preprint; R.N. Silver, D.S. Sivia, and J.E. Gubernatis, Phys. Rev. B41, 2380 (1990).