CHAPTER 4

Stable Numerical Simulations of Models of Interacting Electrons in Condensed-Matter Physics

E.Y. LOH Jr and J.E. GUBERNATIS

Theoretical Division and Center for Nonlinear Studies
Los Alamos National Laboratory
Los Alamos, NM 87545, USA

Electronic Phase Transitions
Edited by
W. Hanke and Yu.V. Kopaev

Contents

1. Introduction

With the explosive development of large-scale computers over the past decade, there has been a parallel increase of interest in simulating models of interacting electrons. The desire is not only to "benchmark" analytic approaches to problems – which, except for a few special cases, are perturbation methods based on extreme parameter values and whose accuracy and radii of convergence are unknown – but also to develop a new predictive approach, valid on its own, that is capable of treating intermediate regions of parameter space where existing analytic approaches fail. Indeed, many interesting and novel materials, such as the heavy-fermion and high-temperature superconducting materials, seem to fall in these regimes.

Simulation techniques are still being developed. In many cases, where results are reported, they are as much a demonstration of an algorithm as they are an exploration of new physics. Difficulties in performing the simulations include excessive computation time, probabilities that become negative, fermion wavefunctions that turn bosonic, and numerical instabilities at low temperatures. Progress, however, is being made, and part of our objective is to describe some recent progress in stabilizing simulations at low and zero temperatures (Loh Jr et al. 1989).

In order to eliminate all but the essential electronic degrees of freedom, we work in a limited basis of single-electron orbitals, giving the Hamiltonian

$$H = -\sum_{ij}\sum_{\sigma} T_{ij}(c^{\dagger}_{i\sigma}c_{j\sigma} + c^{\dagger}_{j\sigma}c_{i\sigma}) + \sum_{ij} V_{ij}n_i n_j, \tag{1.1}$$

where the T_{ij} are the hopping integrals and $c_{i\sigma}$ is the annihilation operator for an electron of spin $\sigma = \uparrow$ or $\downarrow$ in orbital i. This Hamiltonian models the kinetic energy of the particles through the inter-orbital hoppings as $-T_{ij}c^{\dagger}_{i\sigma}c_{j\sigma}$. The Coulomb interaction is modelled as a functional $\sum_{ij} V_{ij}n_i n_j$ of the charge density $n_i = c^{\dagger}_{i\uparrow}c_{i\uparrow} + c^{\dagger}_{i\downarrow}c_{i\downarrow}$, where the V_{ij} include the effects of screening. This Hamiltonian is a generic form that includes such condensed-matter physics standards as the Hubbard and extended-Hubbard Hamiltonians and the Anderson Hamiltonian. There is great interest in studying approximate models at zero and finite temperatures to see if they contain features that

explain antiferromagnetism, superconductivity, or other low-temperature phase transitions observed in nature.

In a path-integral representation of the problem, we review how one introduces auxiliary, Hubbard–Stratonovich fields (Hubbard 1959) to eliminate the self-interactions of the electrons. These auxiliary fields mediate the screened interaction V_{ij}, somewhat like photons mediating the standard electromagnetic interaction. Once the direct interactions among the electrons have been eliminated, the resulting "free-fermion" problem is readily solved, albeit only formally, in terms of determinants of single-electron wavefunctions. Finally, we perform "importance sampling" – with Monte Carlo, Langevin, or molecular-dynamics techniques – over the auxiliary bosonic fields.

One difficulty in thermodynamic studies of many-electron systems is that the lowest-energy states are assigned exponentially large weights in the low-temperature limit. Due to the Pauli exclusion of electrons, however, these states are not macroscopically occupied. Instead, low-energy states are filled up to some "Fermi energy", which controls the physics of the system. Numerically, information about states around this "Fermi energy", which is exponentially suppressed relative to the bottom of the "band", must be extracted as small differences of large numbers – a hopelessly noisy procedure for finite-precision computers. Fortunately, recent algorithmic developments allow the explicit separation of exponentially divergent numerical scales associated with different energies, stabilizing numerical simulation at a small computational cost (Loh Jr et al. 1989).

In this chapter, we will discuss finite-temperature and zero-temperature formalisms for simulating systems of fermions that are coupled only with bosonic degrees of freedom – either physical electron–phonon interactions or couplings to a fictitious field that eliminates direct electronic correlations. In section 2, we establish the theoretical framework for performing fermionic simulations by presenting the transformations which replace electronic correlations with intermediate bosonic fields, by characterizing path integrals of coupled electron–boson wavefunctions, and by discussing the importance sampling of bosonic fields. In this section, we also depart from common practice in that we devote considerable attention to developing a specific insight into what goes on in these simulations, where intuition is often a poor guide. This discussion is presented to motivate the generally tedious mathematics that underlies the algorithms and to help the reader in further algorithmic development. In section 3, we formulate algorithms in mathematical terms, giving the details necessary to carry out a calculation.

Unfortunately, the diverse numerical scales present in the simulations make calculations unstable at low temperatures on any computer of finite precision. In section 4, we discuss techniques for decoupling these scales and stabilizing simulations. In section 5, we present results for sampling models with strong

electronic correlations. Finally, in section 6, we discuss and give suggestions for future areas of development.

2. *Theoretical framework*

In this section, we present the theoretical framework for simulating systems of coupled electrons and bosons. The Trotter approximation (Trotter 1959, Suzuki 1976) is used to discretize quantum path integrals. Intermediate bosonic fields replace direct electronic interactions according to the Hubbard–Stratonovich transformation (Hubbard 1959). We try to give the reader a feeling for the characteristics of the bosonic and fermionic degrees of freedom. Within the approach we describe, only the fermionic degrees of freedom are summed "out" in a formally exact manner. Thus, we discuss importance-sampling strategies for the bosonic fields.

2.1. *Path integrals in imaginary time*

Feynman formulated quantum mechanics as an integral over all paths of a physical system through phase space (Feynman and Hibbs 1965). For real-time dynamics, any particular path contributes to the integral with a phase that depends exponentially on the integrated action along the path. Hence, two very similar paths can interfere destructively and provide no net contribution to the path integral. On the other hand, near a path of stationary phase, paths have similar phases and can add up constructively. In the classical limit $\hbar \to 0$, only the stationary-phase path is important, and so in this limit Feynman's path-integral formulation of quantum mechanics reduces to the principle of least action.

Statistical mechanics, however, entails the study of path integrals in imaginary time (Feynman and Hibbs 1965), in which contributions vary exponentially in magnitude, but not in phase. Therefore, path integrals are dominated by paths of large magnitude, which can be identified by importance-sampling techniques (Creutz and Freedman 1981). For quantum simulations, then, one generally studies the operator $\exp(-\beta H)$ by writing each matrix element as a path integral

$$\langle \psi_L | e^{-\beta H} | \psi_R \rangle = \sum_{|\psi_1\rangle, |\psi_2\rangle, \ldots |\psi_{N_t-1}\rangle} \langle \psi_L | e^{-\Delta\tau H} | \psi_1 \rangle \langle \psi_1 | e^{-\Delta\tau H} | \psi_2 \rangle \cdots \times \langle \psi_{N_t-1} | e^{-\Delta\tau H} | \psi_R \rangle \tag{2.1}$$

in imaginary time $\beta = it/\hbar$, where the path integral has been discretized in small "time" steps $\Delta\tau = \beta/N_t$. In contrast, the real-time dynamics of quantum systems is characterized by many interfering paths which resist similar importance-sampling treatments.

For finite-temperature statistical mechanics, the partition function is

$$Z = \mathrm{Tr}\, \mathrm{e}^{-\beta H} = \sum_{|\psi\rangle} \langle \psi | \mathrm{e}^{-\beta H} | \psi \rangle, \tag{2.2}$$

where $\beta = 1/k_B T$ is the reciprocal temperature; i.e. the integral is over all paths that are periodic in imaginary time, with $|\psi\rangle = |\psi_L\rangle = |\psi_R\rangle$.

In the zero-temperature limit, $\beta = 1/k_B T$ diverges and the boundary condition in imaginary time becomes unimportant. Thus, we may fix the endpoints of the paths and study

$$Z = \langle \psi_L | \mathrm{e}^{-\beta H} | \psi_R \rangle, \tag{2.3}$$

which projects the ground state out of $|\psi_L\rangle$ and $|\psi_R\rangle$ for very large β.

We will develop finite- and zero-temperature formalisms for simulations based on eqs. (2.2) and (2.3), respectively.

2.2. *The Trotter approximation*

Aside from the physical intuition and elegant formalism that accompany path integrals (Feynman and Hibbs 1965), eq. (2.1) has the important computational advantage that it is possible to approximate the matrix elements of $\exp(-\Delta\tau H)$ for $\Delta\tau \to 0$ even though the matrix elements of $\exp(-\beta H)$ are, in practice, impossible to evaluate. Typically, one breaks the Hamiltonian H into any number of pieces $H = H_1 + H_2 + \cdots + H_n$, which possibly do not commute, and writes the Trotter approximation (Trotter 1959, Suzuki 1976)

$$\mathrm{e}^{-\Delta\tau H} = \mathrm{e}^{-\Delta\tau H_1} \mathrm{e}^{-\Delta\tau H_2} \cdots \mathrm{e}^{-\Delta\tau H_n} + \mathcal{O}(\Delta\tau)^2, \tag{2.4}$$

for each factor in eq. (2.1). In the limit $\Delta\tau \to 0$, the approximation for the path integral becomes exact. Since numerical calculations are performed for nonzero $\Delta\tau$, results must be extrapolated to $\Delta\tau = 0$.

2.3. *The Hubbard–Stratonovich transformation*

The Trotter approximation may be used to separate out direct electronic interactions. These interactions may then be replaced by couplings to bosonic fields using the Hubbard–Stratonovich transformation (Hubbard 1959). As an example, we consider the Hamiltonian

$$H = H_0 + U(n_\uparrow - \tfrac{1}{2})(n_\downarrow - \tfrac{1}{2}), \tag{2.5}$$

where n_σ is the number of electrons of spin $\sigma = \uparrow$ or $\downarrow$ in some impurity orbital. Hubbard (1959) originally used time-ordering conventions to handle the path integral. With eq. (2.4), the action in the path integral can be simplified. Through the Trotter approximation, $\exp(-\Delta\tau H) \approx \exp(-\Delta\tau H_0)\exp(-\Delta\tau U(n_\uparrow - \frac{1}{2})(n_\downarrow - \frac{1}{2}))$, the impurity-orbital repulsion is separated from the factor $\exp(-\Delta\tau H_0)$ at each time interval in eq. (2.1). This

can be written as the Gaussian integral

$$\begin{aligned}
&\exp(-\Delta\tau U(n_\uparrow - \tfrac{1}{2})(n_\downarrow - \tfrac{1}{2})) \\
&\quad = \exp(-\Delta\tau U((n_\uparrow - \tfrac{1}{2})^2 + (n_\downarrow - \tfrac{1}{2})^2)/2)\exp(\Delta\tau U(n_\uparrow - n_\downarrow)^2/2) \\
&\quad = \exp(-\Delta\tau U/4)\exp(\Delta\tau U(n_\uparrow - n_\downarrow)^2/2) \\
&\quad = \exp(-\Delta\tau U/4)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx \exp(-x^2/2 + \sqrt{\Delta\tau U}\, x(n_\uparrow - n_\downarrow)). \quad (2.6)
\end{aligned}$$

By casting the exponent in the form of a perfect square and introducing the auxiliary field x, one can rewrite $U(n_\uparrow - \frac{1}{2})(n_\downarrow - \frac{1}{2}) = U(c_\uparrow^\dagger c_\uparrow - \frac{1}{2})(c_\downarrow^\dagger c_\downarrow - \frac{1}{2})$, which is quartic in fermion creation and annihilation operators, in a form which is only quadratic in these operators. The repulsive Coulomb interaction has been replaced by an intermediate bosonic field x, which couples to the net spin $n_\uparrow - n_\downarrow$ in the impurity orbital. Physically, the effect of a positive value of U is to create a spin moment in this orbital.

We have engineered the transformation so that the auxiliary field is always real. Had we not done so, insurmountable computational difficulties would have arisen since, as we will see, we would like to have real exponentials. Had U been negative, we would have written

$$\begin{aligned}
\exp(-\Delta\tau U(n_\uparrow - \tfrac{1}{2})(n_\downarrow - \tfrac{1}{2})) = \exp(-\Delta\tau U/4)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} dx \exp((-x^2/2) \\
+ \sqrt{\Delta\tau|U|}\, x(n_\uparrow + n_\downarrow - 1)), \quad (2.7)
\end{aligned}$$

coupling the field to a charge degree of freedom. Physically, the effect of a negative value of U is to create charge fluctuations in the impurity orbital.

Since it must be effected at each value of the imaginary time, the Hubbard–Stratonovich transformation produces a functional integral over some field $x(\tau)$. By restricting fluctuations in this field, one may derive the Hartree–Fock and random-phase approximations RPA (Evenson et al. 1970, Hamann 1970, Hassing and Esterling 1973). By treating the fields numerically, one may sum over all fluctuations of the fields and so systematically improve these approximations to any desired accuracy.

All interactions in condensed-matter physics can, in principle, be derived by eliminating the bosonic fields that appear in the electromagnetic field equations. We are not interested in solving all of condensed-matter physics ab initio, however, but only a very approximate model that is nonrelativistic, tight-binding, screened, etc. The auxiliary, Hubbard–Stratonovich variables are the bosonic fields that, when eliminated, produce the reduced interaction. These fields have several characteristics. First, like the electromagnetic photon, the field introduced in eq. (2.6) is massless and so provides an instantaneous interaction. Second, the field is entirely localized at the site and so results only in an

on-site repulsion. Extended Coulomb interactions can be modelled either by bosonic fields that are localized on bonds between the orbitals or by coupling fields on different sites. (See Appendix I.) Third, the field could be discrete. This observation, due to Hirsch (1985), is motivated by the fact that the fermion occupation is also discrete: the occupation numbers $n_\uparrow$ and $n_\downarrow$ of up and down electrons can only be 0 or 1. Again, working only with real auxiliary fields, we use one of two different transformations,

$$\begin{aligned} e^{-\Delta\tau U(n_\uparrow - 1/2)(n_\downarrow - 1/2)} &= \tfrac{1}{2} e^{(-\Delta\tau|U|/4)} \sum_x e^{\alpha x(n_\uparrow - n_\downarrow)}, \qquad U > 0, \\ &= \tfrac{1}{2} e^{(-\Delta\tau|U|/4)} \sum_x e^{\alpha x(n_\uparrow + n_\downarrow - 1)}, \quad U < 0, \end{aligned} \tag{2.8}$$

depending on the sign of U. In both cases, $\cosh\alpha = \exp(\Delta\tau|U|/2)$ and the bosonic field is discrete: $x = \pm 1$. We remark again that for an on-site repulsion, the auxiliary field couples to the spin degree of freedom, while for an on-site attraction it couples to the charge. Simulations using discrete transformations produce results with error bars about half of those for the continuous transformation (Hirsch 1985, Buendia 1986) presumably due to the decreased phase space. In Appendix II, we will see that the discrete and continuous transformations are equivalent in the limit $\Delta\tau \to 0$.

Once the auxiliary fields have been introduced, the sums (2.2) and (2.3) become nested sums

$$Z = \sum_{\{x\}} \sum_{\text{fermionic}} \exp(-S) = \sum_{\{x\}} p[x] \tag{2.9}$$

over bosonic $\{x\}$ and fermionic degrees of freedom, where S is some fermion–boson action that we will write out explicitly for some cases later. For a fixed configuration of bosonic field variables x, we will sum over the fermions exactly to calculate the weight $p[x]$ of that configuration. Then, we will sum over the $x(\tau)$ stochastically using importance sampling.

2.4. Boson world lines

To develop a physical intuition for the configurations generated in the simulations, we now discuss world line pictures, the imaginary-time evolutions of bosons and fermions. For a bosonic field, the world line is the value of the field as a function $x(\tau)$ of time. These fields may be auxiliary, massless Hubbard–Stratonovich variables or they could be massive, physical phonons. For fermions, the world lines are the positions of the fermions in real space, again, as functions of time.

While the world line of a quantum particle may be continuous in imaginary time, it is never differentiable. There are always components of very high frequency, which cannot be represented by a finite number of degrees of freedom on a computer. On the other hand, these high-frequency components contribute

negligibly to most physical observables, such as the potential energy or the average displacement of the particle; hence, cutting off the very high frequency components causes no practical limitations in calculations (Loh Jr 1988). Specifically, the error caused by discretizing the path integral in finite imaginary-time units $\Delta\tau$ is due principally to the Trotter approximation (2.4) and not to the elimination of the high frequencies.

Let us now examine the effects of the mass of a quantum particle on the world lines of the particle. Consider massive Einstein phonons in the Holstein Hamiltonian

$$H = H_{\mathrm{el}} + \sum_i \left(\frac{p_i^2}{2m} + \frac{k x_i^2}{2} \right) + \lambda \sum_i x_i n_i , \tag{2.10}$$

where H_{el} is the purely electronic part of the Hamiltonian and x_i and p_i are the positions and momenta of the site phonons which couple to the electronic charge density n_i. The energy expended in lattice distortion is modelled with the harmonic term $k\sum x_i^2/2$. In this model, an electron distorts the ionic lattice through its coupling λ to the phonons. In certain parameter regimes, it is energetically favorable for a second electron to remain close to this first electron rather than to distort another portion of the crystal structure. Thus, the phonons induce an effective attraction between electrons, much as the massless Hubbard–Stratonovich fields mediate an attraction in eq. (2.7). Due to their finite mass, the physical phonons do not react instantaneously to electronic motions and so provide a retarded attraction.

For the massive phonons, the mass m governs the time variation of the phonon world lines through the velocity term

$$\dot{x}_i(\tau) = \frac{x_i(\tau) - x_i(\tau - \Delta\tau)}{\Delta\tau}$$

in the discretized action

$$S = \sum_{\tau=\Delta\tau}^{\beta} \Delta\tau \left(H_{\mathrm{el}} + \sum_i (\tfrac{1}{2} m \dot{x}_i(\tau)^2 + \tfrac{1}{2} k x_i(\tau)^2) + \lambda \sum_i x_i(\tau) n_i(\tau) \right). \tag{2.11}$$

In the limit $\Delta\tau \to 0$, then, the world line $x(\tau)$ must be continuous.

In contrast, eq. (2.7) has only distortion ($-x^2/2$) and fermion–boson ($\sqrt{\Delta\tau\, U}\, x(n-1)$) terms. Since there is no kinetic-energy term, field variables on adjacent time slices, τ and $\tau - \Delta\tau$, are not directly coupled: in contrast to physical, massive bosons, the world lines of the Hubbard–Stratonovich fields are not even continuous, let alone differentiable! Indeed, if we had employed Hirsch's discretized transformation (2.8), then $x(\tau) = \pm 1$ and the lines could not be continuous.

World lines, generated by purely bosonic actions, are shown in fig. 1. The world lines are piecewise constant for the duration $\Delta\tau = 0.04$ of a time slice, but

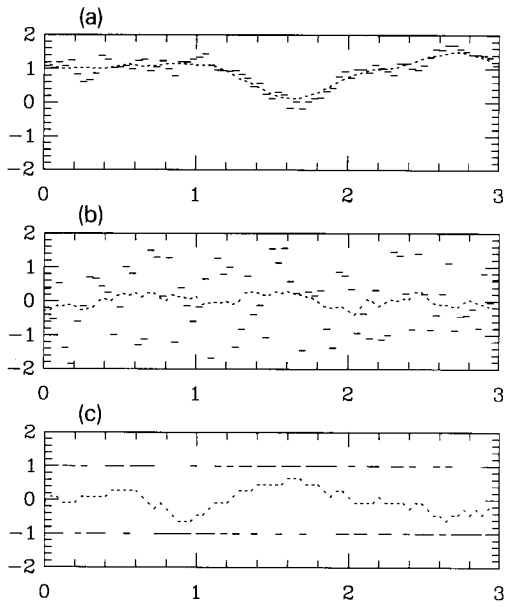

Fig. 1. Sample bosonic world lines: (a) massive field, (b) massless, continuous field, and (c) massless, discrete $x = \pm 1$ field. The massless fields have large fluctuations, which average out to produce a low-frequency average order $\sqrt{\Delta\tau}$. Since the coupling to massless fields is larger, however, the net effect of bosonic fields is always of the same order.

then jump to new constant values. Thus, each line is described by a sequence of values $X_l = x(l\Delta\tau)$ for $l = 1, 2, \ldots, N_t$. The world lines are generated by the weights

$$p[x] \sim \exp\left(-\Delta\tau \sum_l \left(\frac{m}{2} \left(\frac{X_l - X_{l-1}}{\Delta\tau} \right)^2 + \frac{k}{2} X_l^2 \right) \right),$$

with $m = k = 1$ for the massive boson in fig. 1a;

$$p[x] \sim \exp\left(-\sum_l \frac{X_l^2}{2} \right)$$

for the massless, continuous field in fig. 1b; and

$$p[x] \sim \exp\left(-\sum_l (\delta(X_l - 1) + \delta(X_l + 1)) \right)$$

for the massless, discrete field in fig. 1c.

Figure 1a shows the world line for a massive boson. Given that it is required to be piecewise constant, the line is reasonably "continuous": $x(\tau)$ is highly

correlated between successive time slices. Nevertheless, it is not "differentiable" (in the limit $\Delta\tau \to 0$). The field variable is typically of order 1. Hence, electron–boson and purely electronic contributions enter the action to the same order in $\Delta\tau$.

This balance between electron–boson and purely electronic contributions is somewhat more subtle for massless bosons. Consider a massless field $x(\tau)$ that mediates the single-orbital interactions as in eq. (2.5). Possible world lines for $x(\tau)$ are sketched in figs. 1b, c for continuous fields [eqs. (2.6) or (2.7)] an discrete fields $x = \pm 1$ [eq. (2.8)], respectively. Now, successive values $x(\tau - \Delta\tau)$ and $x(\tau)$ of the field variable are completely uncorrelated. Over some finite time τ, the electron feels $N_x = \tau/\Delta\tau$ random fields with a mean field of order $1/\sqrt{N_x} \sim \sqrt{\Delta\tau}$. Thus, for massless bosons the effective field on the electron is reduced by a factor $\sqrt{\Delta\tau}$ due to random fluctuations. The dotted lines in figs. 1b, c show the reduced, effective field after the high-frequency oscillations have been eliminated by averaging over a short imaginary time. On the other hand, the effective coupling of the electrons to this reduced field is much stronger. While the electron–boson contribution to the action goes as $\sum \Delta\tau x$ for the massive boson, it goes as $\sum \sqrt{\Delta\tau |U|}\, x$ in eqs. (2.6) and (2.7) and as $\sum \alpha x \sim \sum \sqrt{\Delta\tau |U|}\, x$, as $\Delta\tau \to 0$, in eq. (2.8). Hence, the effective coupling increases by a factor $\sqrt{\Delta\tau}$ to compensate for the reduction in the effective, time-averaged field. In short, the electron–boson contribution enters the action to the same order as the purely electronic contribution. For massive bosons, however, $x(\tau)$ guides the electrons smoothly, while for massless bosons the field achieves the same effect by jerking the electrons wildly.

(As we shall see, round-off errors in the numerical calculations pose severe difficulties for simulations. One effect of the wild fluctuations in massless fields is to accentuate these numerical instabilities since the fields require larger-scale computations to effect smaller-scale ($\sqrt{\Delta\tau}$) averages.)

The electronic degrees of freedom, then, are determined to equal order in $\Delta\tau$ both by purely electronic contributions and by the coupling to bosonic fields. We have shown this for massless and massive fields that were generated by purely bosonic actions. In Appendix II, we will see that the result remains the same even when the bosonic fields are generated by the full electron–boson action. While it appears that we have relied on a different electron–boson coupling for the massless- and massive-field cases, the redefinition of the coupling is, in fact, quite natural. If we had simply followed the conventions of the massive fields and let the mass go to zero, the electron–boson coupling and the time-averaged field would have been of order 1. Fluctuations in the field, however, would have grown to order $1/\sqrt{\Delta\tau}$. We choose, instead, to keep the field values of unit order. Thus, for massless bosons, we must rescale the field and the coupling and the time-averaged field drops to order $\sqrt{\Delta\tau}$.

2.5. The single-electron propagator

For any particular configuration of field variables, the single-electron, imaginary-time propagator $\mathbf{B}(\tau_2, \tau_1)$ for $\tau_2 \geqslant \tau_1$ is given by the matrix elements

$$B_{ij}(\tau_2, \tau_1) = \langle 0 | c_i \left(\mathscr{T} \exp\left(- \int_{\tau_1}^{\tau_2} S(\tau)\, \mathrm{d}\tau \right) \right) c_j^\dagger | 0 \rangle, \tag{2.12}$$

where $\mathscr{T}$ is the time-ordering operator, S the imaginary-time electron–boson action containing the dependence on the $x(\tau)$, and $|0\rangle$ is the vacuum state. This propagator obeys the identities

$$\mathbf{B}(\tau_3, \tau_2)\mathbf{B}(\tau_2, \tau_1) = \mathbf{B}(\tau_3, \tau_1), \quad \mathbf{B}(\tau, \tau) = \mathbf{1}. \tag{2.13}$$

Since the up and down electrons can couple differently to the bosonic fields, we will compute the single-particle propagator (2.12) separately for each electronic spin $\uparrow$ and $\downarrow$. In a homogeneous bosonic field, $\mathbf{B}^\sigma(\tau_2, \tau_1)$ causes the wavefunction of an electron with spin σ, injected at a specific site at imaginary time τ_1, to diffuse evenly throughout the spatial lattice and grow or decrease in magnitude. From eqs. (2.12) and (2.13), the solution of the equations of motion is

$$\begin{aligned} c_i(\tau_2) &= \sum_j B_{ij}(\tau_2, \tau_1) c_j(\tau_1), \\ c_i^\dagger(\tau_2) &= \sum_j c_j^\dagger(\tau_1) B_{ji}^{-1}(\tau_2, \tau_1). \end{aligned} \tag{2.14}$$

Notice that $c_i(\tau)$ and $c_i^\dagger(\tau)$ are not Hermitian conjugates since we are working in imaginary time.

2.6. The many-electron propagator

Up to this point, we have neglected the statistics of the identical fermions. While the spin σ of a fermion distinguishes it from a fermion of opposite spin $-\sigma$, all fermions of the same spin are indistinguishable. Indistinguishable fermions obey Fermi–Dirac statistics, i.e. the wavefunction changes sign under the exchange of any two identical fermions. An important consequence of this is the Pauli exclusion principle for identical fermions in the same state. When the Fermi–Dirac nature of the particles becomes important – at low temperatures and high densities – it is the delicate, nearly perfect cancellation of opposite-sign contributions to the partition function that gives rise to the distinctly fermionic phenomena that we hope to measure. Thus, while we can use importance sampling for the bosonic fields, summing stochastically over fermion configurations would be an ineffective means of capturing fermionic character. Our strategy will be to use importance sampling to sum over configurations of the bosons, but to sum, even if only formally, over all fermionic paths exactly. This,

we will see, does not eliminate the cancellation or "sign" problem completely, but it does make many calculations possible.

Of course, these is nothing "fermionic" about the single-particle propagator (2.12), since statistics is meaningless for a single particle. So we now consider the imaginary-time propagator for N_σ identical fermions, each of which has spin σ and couples identically to the external, time-varying field, but all of which are independent of each other except for their statistics. The advantage of having independent fermions is that many-fermion propagators are easily expressed in terms of single-particle propagators as (see Appendix III)

$$\langle 0|c_{i_1}c_{i_2}\cdots c_{i_{N_\sigma}}\left(\mathcal{T}\exp\left(-\int_{\tau_1}^{\tau_2}S(\tau)\,d\tau\right)\right)c^\dagger_{j_{N_\sigma}}\cdots c^\dagger_{j_2}c^\dagger_{j_1}|0\rangle$$
$$=\det\begin{pmatrix} B_{i_1j_1}(\tau_2,\tau_1) & B_{i_1j_2}(\tau_2,\tau_1) & \cdots & B_{i_1j_{N_\sigma}}(\tau_2,\tau_1)\\ B_{i_2j_1}(\tau_2,\tau_1) & B_{i_2j_2}(\tau_2,\tau_1) & \cdots & B_{i_2j_{N_\sigma}}(\tau_2,\tau_1)\\ \vdots & \vdots & \ddots & \vdots\\ B_{i_{N_\sigma}j_1}(\tau_2,\tau_1) & B_{i_{N_\sigma}j_2}(\tau_2,\tau_1) & \cdots & B_{i_{N_\sigma}j_{N_\sigma}}(\tau_2,\tau_1)\end{pmatrix}. \tag{2.15}$$

The determinant automatically incorporates the odd parity of particle exchange into the calculation. We remark that while the number $N!/N_\sigma!(N-N_\sigma)!$ of many-body states grows exponentially with the number N of possible single-particle states, the factorization of the problem into single-particle propagators reduces the calculation to a manipulation of $N\times N$ matrices and makes the calculation scale as N^3 at worst*. Thus, the effort of introducing auxiliary fields to eliminate direct electron–electron interactions has already brought us some advantages.

To summarize, since the evolving particles are "independent", the many-particle propagator should simply be a product of single-particle propagators. Having injected N_σ identical particles at time τ_1 into orbitals $j_1, j_2, \ldots, j_{N_\sigma}$, we do not, however, know which of these particles arrive at the various orbitals $i_1, i_2, \ldots, i_{N_\sigma}$ at time τ_2. The determinant (2.15) is a sum over all such arrivals with the appropriate minus signs inserted to account for an even or odd number of permutations of the identical fermions.

2.7. *The sign problem*

We now imagine that we inject two particles at τ_1 at sites j_1 and j_2, as in fig. 2. In the figure, the space axis is vertical and one assumes periodic boundary conditions. The imaginary-time axis runs from right to left. We assume the sign of the electron–boson coupling to be such that shaded regions of space are unfavorable for electron propagation. In the figure, the bosonic field variables happen to

* "Conjugate-gradient" methods, which we will not discuss here, scale nominally as N since they iterate with sparse representations of $N\times N$ matrices. Unfortunately, such methods require a divergent number of iterations at low temperatures, making them impractical.

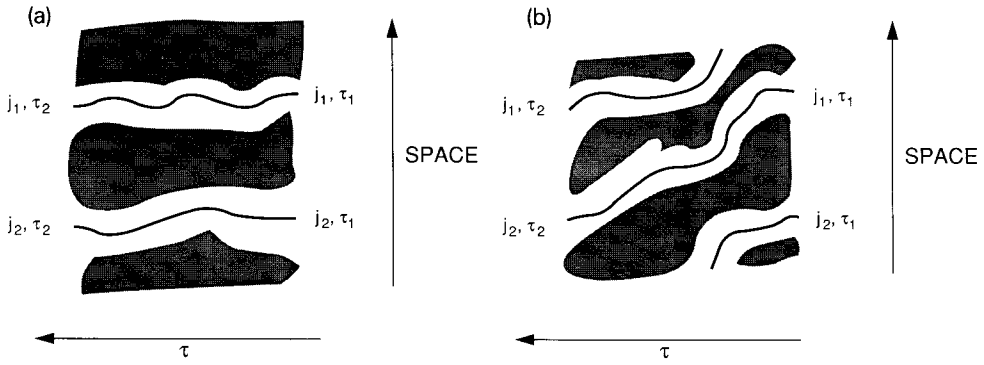

Fig. 2. Bosonic field configurations of different "sign". The electron–boson coupling is chosen such that shaded regions of space–time are unfavorable for electron propagation. In (a), well-defined channels lead electrons at j_1 and j_2 back to themselves. In (b), the channels force an exchange of the electrons. The fields in (b), therefore, would appear in a simulation with opposite signs relative to (a).

provide well-defined channels for the electrons. In fig. 2a, these channels guide the electrons from j_1 and j_2 back to themselves. Hence, the configuration of field variables represented in fig. 2a contributes to the partition function with a positive weight. In contrast, the auxiliary variables represented in fig. 2b force an exchange of the electrons and their contribution appears in the partition function with a negative weight. These pictures correspond to high temperatures (short imaginary times and few chances for exchanges), low particle densities, and massive bosons (slowly varying fields in time). In practice, particularly for massless bosonic fields, which mediate instantaneous interactions, the fields are very noisy and hence do not provide well-defined channels. Thus, in the limit of low temperature, high fermion densities, and massless fields, there are many exchanges and it is difficult to characterize the sign of the determinant. The picture becomes one of fermions tunnelling through an amorphous medium. Nevertheless, it is clear that the space–time structure of the x-fields can lead to determinants (2.15) of opposite signs.

This is unfortunate because we rely on importance sampling to make measurements in the hopelessly large phase space of field variables x, and importance sampling is incapable of keeping track of delicate cancellations of contributions with opposite signs. To a large extent, the fermionic nature of the particles has already been accounted for by the exact evaluation of determinants. Thus, "determinantal" simulations have had a number of successes while "world-line" methods*, which sum stochastically over assignments between

* "World-line" methods (Hirsch et al. 1981, 1982) work for fermions in one dimension, which are isomorphic via the Jordan–Wigner transformation to quantum spin chains. Above one dimension, these methods are useful for fermions only at high temperatures and low densities, when the short imaginary time $\beta = 1/k_B T$ and large interparticle spacing make fermion exchange improbable – in short, when the fermionic nature of the particles is unimportant.

identical particles, work only for systems with no essential fermionic characteristics. Nevertheless, as we will see, even "determinantal" methods will fail due to sign difficulties.

2.8. *"Nodal" surfaces*

The net weight with which a particular set of bosonic fields enters the partition function (2.9) takes the form of a determinant. This determinant, as we have seen, may be either a positive or a negative functional of the field variables $\{x\}$, particularly at low temperatures, for which a sufficiently long imaginary time is available for particles to exchange many times. Clearly, in the phase space of the $\{x\}$, the determinant has nodal surfaces that separate these positive and negative regions and on which a configuration of field variables would have vanishing weights. We may think of this vanishing weight as being due to an infinite "potential energy" for the configuration.

Further, for a simulation to produce unbiased averages, it must be ergodic, i.e. it must be able to sample all of configuration space and not just one subspace bounded by potential barriers. Such a sampling is difficult to achieve with a gradual evolution of the field variables in the presence of "nodal" surfaces. On the other hand, a large change in even a single field variable may cause a nodal crossing if the discretization parameter $\Delta\tau$ is sufficiently large.

2.9. *Importance sampling: single classical particle*

Importance sampling is the technique of generating states of a physical system stochastically with the same probability distribution as in thermal equilibrium. While many samples are generated to reduce the statistical fluctuations in this stochastic process, the number of samples needed to make reasonably good measurements on the physical system is many orders of magnitude smaller than the size of phase space, which is prohibitively large. Typically, importance sampling takes the form of a Markov chain, a trajectory through phase space in which each new sample is closely related to the previous one.

We begin by discussing importance sampling in the context of a very simple example: a single classical particle in a potential well at finite temperature. In thermodynamic equilibrium, the probability of the particle being at position x is $\exp(-\beta V(x))$, where $V(x)$ is the external potential and $\beta = 1/k_B T$ is, once again, the reciprocal temperature. Of course, for this classical model, the action is diagonal in any basis and we no longer have an imaginary-time dimension. While the partition-function integral is best evaluated with numerical quadratures, such methods are impractical in evaluating the high-dimensional integrals that occur in many-body problems. Hence, we will discuss strategies for impotance sampling that are appropriate for many-body problems.

The Monte Carlo method is perhaps the most popular approach for sampling phase space.* In this method one proposes a change in the system configuration and then accepts or rejects the change in accordance with "detailed balance". For example, in the Metropolis algorithm (Metropolis et al. 1953), given a particle at position x, one proposes a new position $x' = x + \delta x$, with δx chosen at random from some symmetric distribution, and accepts the point with probability

$$\begin{aligned} P(x \to x') &= 1 && \text{if } e^{-\beta V(x')} > e^{-\beta V(x)}, \\ &= e^{-\beta(V(x')-V(x))} && \text{if } e^{-\beta V(x')} < e^{-\beta V(x)}. \end{aligned} \tag{2.16}$$

Another popular choice for the transition probability is (Binder 1987)

$$P(x \to x') = \frac{e^{-\beta V(x')}}{e^{-\beta V(x)} + e^{-\beta V(x')}}, \tag{2.17}$$

which would arise if we had frozen the system so that the particle could only move between position x and the randomly chosen position x' and then put the system in contact with a heat bath. These two algorithms maintain detailed balance

$$P(x \to x')e^{-\beta V(x)} = P(x' \to x)e^{-\beta V(x')}, \tag{2.18}$$

ensuring that the equilibrium distribution $\exp(-\beta V(x))$ of x remains fixed. In simulation time, the classical particle would appear to jump around in phase space under these rules.

An alternative approach would be to let the particle evolve by simulating its dynamics. We have not specified the dynamics of the particle, since originally we were interested only in its equilibrium properties. Nevertheless, we may introduce a fictitious dynamics that does not affect the particle's equilibrium properties. For example, the addition of a kinetic energy $p^2/2m$ leaves the partition function

$$Z = \int dx \int dp\, e^{-\beta(p^2/2m + V(x))} \sim \int dx\, e^{-\beta V(x)} \tag{2.19}$$

unchanged to within some arbitrary multiplicative factor. If we choose some momentum p from the Gaussian distribution $\exp(-\beta p^2/2m)$, we may then let the particle evolve according to the equations of motion:

$$\dot{x} = \frac{p}{m}, \qquad \dot{p} = -\frac{\partial V(x)}{\partial x}. \tag{2.20}$$

In molecular dynamics, these equations are commonly integrated using the leap-frog algorithm, which is fast, stable, and easy to implement (Verlet 1968). In

* For more details on classical Monte Carlo methods, including both the Metropolis and heat bath algorithms, see, e.g. Binder (1987).

molecular dynamics, the classical particle would trace long, smooth trajectories through phase space. A many-body system would appear to undergo a complicated, "breathing" motion. To sample the thermodynamic ensemble, we must thermalize p periodically, i.e. we must pick p afresh. If we choose p from the Gaussian distribution at each time-evolution step, we are using the Langevin method, which leads to a Brownian motion in the degrees of freedom.

One drawback of using a continuous evolution of the $\{x\}$, such as in a molecular-dynamics or Langevin method, is that it is difficult to tunnel through the potential barriers of the nodal surfaces of the determinantal functional. In contrast, Monte Carlo updating may "step" across such a potential barrier with the update of even only one field variable if $\Delta\tau$ is not too small.

Finally, a combination of Monte Carlo, molecular-dynamics, or Langevin methods may be used to perform the importance sampling. Indeed, a popular "hybrid" combination is to let the system evolve according to some fictitious dynamics using molecular dynamics (updating the momenta in the system occasionally and so including occasional Langevin steps), but then to correct for step-size errors in the numerical integration by accepting or rejecting the evolved system according to the Metropolis rule (Scalettar et al. 1987)

2.10. The Green's function

The matrix elements of the Green's function $\mathbf{G}^{\sigma}$ are

$$G^{\sigma}_{ij}(\tau', \tau) = \langle c_{i\sigma}(\tau') c^{\dagger}_{j\sigma}(\tau) \rangle_x. \tag{2.21}$$

Here, the angular brackets denote averages over all electron paths for a *particular* configuration of the bosonic fields x. This average, of course, is taken with respect to whichever imaginary-time boundary condition [eq. (2.3) or eq. (2.2)] is appropriate and must be calculated separately for $\sigma = \uparrow$ and $\sigma = \downarrow$. While the Green's function is customarily defined with a time-ordering operator inside the brackets to serve field-theoretic purposes, this convention is not needed and will not be used in our numerical development.

The Green's function is the central object in numerical simulations, playing at least two essential roles. The first is in the importance sampling of the bosonic fields. For example, one may propose and then (using Metropolis or heat bath algorithms) accept or reject a change in a single field variable $x_i(\tau)$ at site i and imaginary time τ. In the spirit of perturbation theory, we could estimate the changes in the electron–boson contribution to the action, under the assumption that the electronic degrees of freedom remain fixed under this change in a single field variable. Then, since $x_i(\tau)$ couples only to $n_{i\uparrow}(\tau)$ and $n_{i\downarrow}(\tau)$, the only information we need about the electrons are the $G^{\sigma}_{ii}(\tau, \tau) = 1 - \langle n_{i\sigma}(\tau) \rangle_x$ for both spin-up and for spin-down electrons. Indeed, Blankenbecler, Scalapino, and Sugar (Scalapino and Sugar 1981, Blankenbecler et al. 1981a, b) have shown that the $G^{\sigma}_{ii}(\tau, \tau)$ are all that one needs to evaluate changes in the electron–boson

contribution to the action exactly. Similarly, the $G^{\sigma}_{ii}(\tau, \tau)$ provide all the information needed to perform Langevin or molecular-dynamics sampling, since the forces on the bosonic field variables depend only on local electronic densities.

The second important role that the Green's function plays in simulations is in measurements of the electronic degrees of freedom. If A is a measurement operator that contains electron creation and annihilation operators, then

$$\langle A \rangle = \frac{1}{N_m} \sum_{\{x\}}' \langle A \rangle_x, \tag{2.22}$$

where the primed sum is over N_m bosonic field variables sampled in proportion to their probabilities in thermal equilibrium, and $\langle A \rangle_x$ is the expectation value of A for a fixed set of fields x. Since $\langle A \rangle_x$ is an average for fermions which interact only with fixed bosons and not with each other, we may apply Wick's theorem (Fetter and Walecka 1971) to express it as a sum over products of contractions of fermion operators. Each such contraction, of course, is a specific element of the $\mathbf{G}^{\sigma}$.

One final note on these Green's functions is that their matrix elements need not fall in the numerical ranges we would normally expect. For example, $G^{\sigma}_{ii}(\tau, \tau) = 1 - \langle n_{i\sigma}(\tau) \rangle_x$ and so should fall in the range $0 \leqslant G^{\sigma}_{ii}(\tau, \tau) \leqslant 1$. For fermions evolving in imaginary time through time-varying fields, however, this is no longer the case. Indeed, near the nodal surfaces discussed earlier, the elements of $\mathbf{G}^{\sigma}$ diverge.

3. *Mathematical formulation*

In this section, we develop the simulation algorithms in detail, focussing on a particular example to make the discussion more transparent. We work with the two-dimensional Hubbard model, eliminating the interactions between the fermions using the discrete Hubbard–Stratonovich transformation (Hirsch 1985). Importance sampling is then performed using single-site-update Monte Carlo. After each change is accepted, the Green's function is updated by the procedure suggested by Blankenbecler, Scalapino, and Sugar (BSS) (Scalapino and Sugar 1981, Blankenbecler et al. 1981a, b). The two-dimensional Hubbard Hamiltonian is a convenient example. Other Hamiltonians may be treated by steps similar to those outlined here.

3.1. *Matrix representation of single-particle propagators*

The Hamiltonian we consider is

$$\begin{aligned} H &= H_t + H_U + H_\mu \\ &= -\sum_{ij,\sigma} T_{ij}(c^{\dagger}_{i\sigma} c_{j\sigma} + c^{\dagger}_{j\sigma} c_{i\sigma}) + U \sum_i (n_{i\uparrow} - \tfrac{1}{2})(n_{i\downarrow} - \tfrac{1}{2}) - \mu \sum_i n_i, \end{aligned} \tag{3.1}$$

where the T_{ij} are the hopping integrals, $U > 0$ is the strength of the on-site Hubbard repulsion, and μ is the chemical potential, chosen to be zero at half filling. Sums over i are taken over sites on a periodic, square lattice, and we will consider nonzero hopping T_{ij} only between nearest-neighbor sites i and j. We will use N for the number of sites on the lattice and work in the real-space basis of sites i.

Using the Trotter approximation, we break each of the N_t factors $\exp(-\Delta\tau H)$ in eq. (2.1) into two additional factors

$$e^{-\Delta\tau H} \approx e^{-\Delta\tau(H_U + H_\mu)} e^{-\Delta\tau H_t}, \tag{3.2}$$

which are diagonal and off-diagonal, respectively, in a site-occupation representation. Since the terms which make up H_U and H_μ are diagonal, they commute among themselves and their exponentials can be factored into

$$e^{-\Delta\tau(H_U + H_\mu)} = \prod_i e^{-\Delta\tau U(n_{i\uparrow} - 1/2)(n_{i\downarrow} - 1/2)} \prod_i e^{\Delta\tau\mu n_i}$$

without approximation. We reduce the exponential of the quartic terms $-\Delta\tau U(n_{i\uparrow} - \frac{1}{2})(n_{i\downarrow} - \frac{1}{2})$ by invoking Hirsch's discrete Hubbard–Stratonovich transformation (2.8).

The up and down fermions couple to the external fields differently. Hence, we compute two single-particle propagators $\mathbf{B}^\uparrow$ and $\mathbf{B}^\downarrow$. For a given configuration of field variables,

$$\mathbf{B}^\sigma(\tau_2, \tau_1) = \mathbf{B}^\sigma(\tau_2, \tau_2 - \Delta\tau)\mathbf{B}^\sigma(\tau_2 - \Delta\tau, \tau_2 - 2\Delta\tau) \cdots \mathbf{B}^\sigma(\tau_1 + \Delta\tau, \tau_1),$$

where

$$\mathbf{B}^\sigma(\tau, \tau - \Delta\tau) \approx \mathbf{A}^\sigma(\tau)\exp(\Delta\tau\mu)\exp(\Delta\tau\mathbf{T}) \tag{3.3}$$

gives eq. (3.2) in our single-particle, real-space basis. The particles diffuse in real space for a small imaginary time $\Delta\tau$ according to $\exp(\Delta\tau\mathbf{T})$, their propagators are amplified or attenuated through the scalar factor $\exp(\Delta\tau\mu)$, and finally are scattered by the external potentials described by

$$\mathbf{A}^\sigma(\tau) = \begin{pmatrix} e^{\sigma\alpha x_1(\tau)} & & & 0 \\ & e^{\sigma\alpha x_2(\tau)} & & \\ & & \ddots & \\ 0 & & & e^{\sigma\alpha x_N(\tau)} \end{pmatrix},$$

where the exponents $\sigma\alpha x = \pm\alpha x$ flip signs depending on whether $\sigma = \uparrow$ or $\downarrow$, and where $\cosh(\alpha) = \exp(\Delta\tau U/2)$ as in eq. (2.8). Since we couple the external fields only to local fermionic degrees of freedom, $\mathbf{A}^\sigma$ is diagonal. The σ-dependence of the $\mathbf{A}^\sigma$ reflects the fact that the two types of fermions, $\sigma = \uparrow$ or $\downarrow$, couple to the auxiliary fields differently. Indeed, it is only because of the $\mathbf{A}^\sigma$ that the propagators $\mathbf{B}^\uparrow$ and $\mathbf{B}^\downarrow$ are different.

3.2. Checkerboard breakup

Generally, we will choose to break the kinetic-energy matrix $\exp(\Delta\tau \mathbf{T})$ in eq. (3.3) further. While the computational cost of diagonalizing and then exponentiating $\mathbf{T}$ is relatively small, the resulting matrix is dense, making the number of operations needed to perform a matrix multiplication scale as N^3. Alternatively, for uniform hopping integrals, one may utilize fast-fourier transforms (FFT) and apply the kinetic-energy factors $\exp(\Delta\tau \mathbf{T})$ in momentum space, in which $\mathbf{T}$ is diagonal. This approach, unfortunately, does not allow for nonuniform hoppings, and efficient use of the FFT restricts the linear size of the system one studies to powers of 2.

A sparse and extremely convenient approximate form for $\exp(\Delta\tau \mathbf{T})$ results from a further application of the Trotter approximation to the kinetic energy. In the single-particle, real-space basis, the exponential of the hopping part of the Hamiltonian may be written as

$$\exp(\Delta\tau \mathbf{T}) = \exp\left(\Delta\tau \sum_{\langle ij\rangle} \mathbf{T}^{(ij)}\right) \approx \prod_{\langle ij\rangle} \exp(\Delta\tau \mathbf{T}^{(ij)}). \tag{3.4}$$

The sparse matrices $\mathbf{T}^{(ij)}$, with only $T^{(ij)}_{ij} = T^{(ij)}_{ji} = T_{ij}$ nonzero, are easily exponentiated, giving

$$\exp(\Delta\tau \mathbf{T}^{(ij)}) = \exp \Delta\tau \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & T_{ij} & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & T_{ij} & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cosh(\Delta\tau T_{ij}) & \cdots & \sinh(\Delta\tau T_{ij}) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \sinh(\Delta\tau T_{ij}) & \cdots & \cosh(\Delta\tau T_{ij}) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

Each $\exp(\Delta\tau \mathbf{T}^{(ij)})$ is also sparse, with only the *ii*, *ij*, *ji*, and *jj* elements differing from those of the unit matrix. If we replace the multiplication of the dense matrix by the series (3.4) of sparse-matrix multiplications, the number of operations for

multiplication onto an $N \times N$ matrix is reduced from N^3 to $N \times N_b$, where the number of bonds N_b grows linearly with the number of sites for local hoppings. For the square lattice, for example, $N_b = 2N$. For historical reasons, eq. (3.4) is referred to as the checkerboard breakup.*

Not only is the checkerboard breakup of the kinetic energy reasonably fast, it is extremely versatile and convenient. No diagonalization or FFT is required and changes in the hopping integrals – as when one incorporates a fixed lattice distortion – may be effected immediately. Further, evaluation of the inverse $\exp(-\Delta\tau\,\mathbf{T})$ requires no extra work: one has only to reverse the sign of the off-diagonal elements. In subsequent discussions, we may write $\exp(\Delta\tau\,\mathbf{T})$ even when we use the checkerboard, approximate form.

3.3. Bosonic-configuration weights

Within the zero-temperature framework suggested by eq. (2.3), let us write $|\psi_L\rangle$ and $|\psi_R\rangle$ as products of single-particle states. Then, using eq. (2.15), the field weights in eq. (2.9) become

$$p[x] = \prod_{\sigma=\uparrow,\downarrow} \det(\mathbf{P}_L^\sigma \mathbf{B}^\sigma(\beta, 0)\mathbf{P}_R^\sigma), \tag{3.5}$$

where the introduction of the auxiliary fields x has decoupled the system into separate $\sigma = \uparrow, \downarrow$ problems. The field-variable dependence of the right-hand side of eq. (3.5) enters through the single-particle propagator $\mathbf{B}$.

In eq. (3.5), we have written $|\psi_L\rangle$ and $|\psi_R\rangle$ as products of up- and down-electron single-particle states in order to use eq. (2.15). The rows (columns) of the rectangular $N_\sigma \times N$ ($N \times N_\sigma$) matrices $\mathbf{P}_L^\sigma$ ($\mathbf{P}_R^\sigma$) give the single-particle states of $|\psi_L\rangle$ ($|\psi_R\rangle$) of spin $\sigma = \uparrow, \downarrow$ in our N-state basis. For example, let $|\psi_R\rangle = |\psi_R^\uparrow\rangle|\psi_R^\downarrow\rangle$ and let

$$\begin{aligned}|\psi_R^\sigma\rangle = {} & (P_{11}c_{1\sigma}^\dagger + P_{21}c_{2\sigma}^\dagger + \cdots + P_{N1}c_{N\sigma}^\dagger) \\ & (P_{12}c_{1\sigma}^\dagger + P_{22}c_{2\sigma}^\dagger + \cdots + P_{N2}c_{N\sigma}^\dagger) \\ & \cdots \\ & (P_{1N_\sigma}c_{1\sigma}^\dagger + P_{2N_\sigma}c_{2\sigma}^\dagger + \cdots + P_{NN_\sigma}c_{N\sigma}^\dagger)|0\rangle,\end{aligned}$$

where $|0\rangle$ is the vacuum state. Then

$$\mathbf{P}_R^\sigma = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1N_\sigma} \\ P_{21} & P_{22} & \cdots & P_{2N_\sigma} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N1} & P_{N2} & \cdots & P_{NN_\sigma} \end{pmatrix}.$$

* The Trotter formula first became popular in numerical simulations for one-dimensional chains. There, the alternation of even and odd bonds laid out a checkerboard in space–time. See, e.g. Barma and Shastry (1978).

Typically, we will choose $|\psi_L\rangle = |\psi_R\rangle$, so that the $\mathbf{P}_L^\sigma$ are the Hermitian conjugates of the $\mathbf{P}_R^\sigma$. The numbers $N_\uparrow$ and $N_\downarrow$ of up- and down-spin electrons need not be equal. These numbers are fixed over the course of any one simulation however, so that the zero-temperature algorithm is carried out in the canonical ensemble.

For the finite-temperature partition function (2.2), we are no longer interested in matrix elements of $\exp(-\beta H)$ between specified initial and final many-body wavefunctions, but we would like to sum over all $|\psi_L\rangle = |\psi_R\rangle$ of all occupations $N_\uparrow$ and $N_\downarrow$. Now,

$$p[x] = \prod_{\sigma=\uparrow,\downarrow} \det(\mathbf{1} + \mathbf{B}^\sigma(\beta, 0)). \tag{3.6}$$

We derive this result in Appendix IV. Here, we only note that if we simply had $\det(\mathbf{B})$, we would have had eq. (2.15) with a packed lattice: a fermion on each site. On the other hand, $\det(\mathbf{1})$ is the many-particle propagator for zero particles. In eq. (3.6), $\det(\mathbf{1} + \mathbf{B})$ samples both of these terms and so generates the grand-canonical partition function for a particular set of field variables, summing over all states with all occupation numbers of electrons.

3.4. Matrix representation of the Green's function

To derive $\mathbf{G}$ in terms of the single-particle propagator, let us begin by examining matrix elements of the equal-time Green's function

$$G_{ij}^\sigma(\tau, \tau) = \langle c_{i\sigma}(\tau) c_{j\sigma}^\dagger(\tau)\rangle = \Delta_{ij} - \langle c_{j\sigma}^\dagger(\tau) c_{i\sigma}(\tau)\rangle,$$

where Δ_{ij} is the Kronecker delta function. The expectation $\langle c_{j\sigma}^\dagger(\tau) c_{i\sigma}(\tau)\rangle$ may be evaluated by coupling the action to $c_{j\sigma}^\dagger c_{i\sigma}$ for an instant at imaginary time τ:

$$\mathcal{T}\exp\left(-\int_0^\beta S(t)\,dt\right) = \mathcal{T}\exp\left(-\int_\tau^\beta S(t)\,dt\right)\exp\left(-\int_0^\tau S(t)\,dt\right)$$

$$\to \mathcal{T}\exp\left(-\int_\tau^\beta S(t)\,dt\right)\exp(h c_{j\sigma}^\dagger c_{i\sigma})\exp\left(-\int_0^\tau S(t)\,dt\right).$$

Under this transformation of the action, the weight $p[x]$ becomes a function $p_h[x]$ of the coupling constant h. For example, in the zero-temperature formalism,

$$p_h[x] = \det(\mathbf{P}_L^\sigma \mathbf{B}^\sigma(\beta, \tau) e^{h\mathbf{O}} \mathbf{B}^\sigma(\tau, 0)\mathbf{P}_R^\sigma)\det(\mathbf{P}_L^{-\sigma}\mathbf{B}^{-\sigma}(\beta, 0)\mathbf{P}_R^{-\sigma}),$$

where the only nonzero matrix element of $\mathbf{O}$ is $O_{ji} = 1$. With the definitions

$$\mathbf{L}^\sigma(\tau) = \mathbf{P}_L^\sigma \mathbf{B}^\sigma(\beta, \tau) \tag{3.7}$$

and

$$\mathbf{R}^\sigma(\tau) = \mathbf{B}^\sigma(\tau, 0)\mathbf{P}_R^\sigma, \tag{3.8}$$

the expectation becomes

$$\langle c^\dagger_{j\sigma}(\tau)c_{i\sigma}(\tau)\rangle = \frac{\partial}{\partial h}\ln p_h[x]|_{h=0} = \mathrm{Tr}\,\frac{\partial}{\partial h}\ln(\mathbf{L}^\sigma e^{h\mathbf{O}}\mathbf{R}^\sigma)|_{h=0}$$
$$= \mathrm{Tr}(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\mathbf{L}^\sigma\mathbf{O}\mathbf{R}^\sigma = (\mathbf{R}^\sigma(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\mathbf{L}^\sigma)_{ij}.$$

The equal-time Green's function in the zero-temperature formalism becomes

$$\mathbf{G}^\sigma(\tau,\tau) = \mathbf{1} - \mathbf{R}^\sigma(\tau)(\mathbf{L}^\sigma(\tau)\mathbf{R}^\sigma(\tau))^{-1}\mathbf{L}^\sigma(\tau). \tag{3.9}$$

In the finite-temperature formalism, we can play the same tricks, coupling to $c^\dagger_j c_i$, writing the resulting field weight as a function of h, and taking a logarithmic derivative. Now,

$$\mathbf{G}^\sigma(\tau,\tau) = (\mathbf{1} + \mathbf{B}^\sigma(\tau,0)\mathbf{B}^\sigma(\beta,\tau))^{-1}. \tag{3.10}$$

For unequal arguments, we transform the Green's function using the integrated equations of motion (2.14) for the fermion operators to get

$$\mathbf{G}(\tau',\tau) = \mathbf{B}(\tau',\tau)\mathbf{G}(\tau,\tau) = \mathbf{G}(\tau',\tau')\mathbf{B}(\tau',\tau). \tag{3.11}$$

3.5. *Metropolis algorithm*

As we have seen, the weight of a configuration of the bosonic field variables is a product [eq. (3.5) or (3.6)] of two determinants. Such determinants are too expensive to evaluate for each set $\{x\}$ of field variables that we encounter in the course of a simulation. Fortunately, importance sampling via either the Metropolis or heat bath algorithm requires only the ratio $\mathscr{R} = p[x']/p[x]$ of weights. As we will now see, these ratios are straightforward to evaluate when $\{x'\}$ differs from $\{x\}$ in the value of only one field variable.

The change of only one field variable, at site i and at imaginary time τ, from $x_i(\tau)$ to $x'_i(\tau)$ affects the time evolution of the fermions through the matrix $\mathbf{A}$:

$$\mathbf{A}^\sigma(\tau) \to \begin{pmatrix} e^{\sigma\alpha x_1(\tau)} & & & & 0 \\ & \ddots & & & \\ & & e^{\sigma\alpha x'_i(\tau)} & & \\ & & & \ddots & \\ 0 & & & & e^{\sigma\alpha x_N(\tau)} \end{pmatrix},$$
$$= (\mathbf{1} + \mathbf{\Delta}^\sigma(i,\tau))\mathbf{A}^\sigma(\tau),$$

where the only nonzero matrix element of $\mathbf{\Delta}^\sigma(i,\tau)$ is $\Delta^\sigma_{ii}(i,\tau) = \exp(\sigma\alpha(x'_i(\tau) - x_i(\tau))) - 1$. The accompanying change in the single-particle propagator $\mathbf{B}^\sigma$ is

$$\mathbf{B}^\sigma(\beta,0) = \mathbf{B}^\sigma(\beta,\tau)\mathbf{B}^\sigma(\tau,0)$$
$$\to \mathbf{B}^\sigma(\beta,\tau)(\mathbf{1} + \mathbf{\Delta}^\sigma(i,\tau))\mathbf{B}^\sigma(\tau,0).$$

For the zero-temperature algorithm, we now have $\mathscr{R} = \mathscr{R}^{\uparrow}\mathscr{R}^{\downarrow}$, with

$$\mathscr{R}^{\sigma} = \frac{\det(\mathbf{L}^{\sigma}(\tau)(\mathbf{1} + \mathbf{\Delta}^{\sigma}(i, \tau))\mathbf{R}^{\sigma}(\tau))}{\det(\mathbf{L}^{\sigma}(\tau)\mathbf{R}^{\sigma}(\tau))},$$

using the definitions (3.7) and (3.8). Dropping a number of indices to simplify the equations, the probability-ratio factors become

$$\begin{aligned}\mathscr{R}^{\sigma} &= \frac{\det(\mathbf{L}(\mathbf{1} + \mathbf{\Delta})\mathbf{R})}{\det(\mathbf{LR})} \\ &= \frac{\det(\mathbf{LR} + \mathbf{L\Delta R})}{\det(\mathbf{LR})} \\ &= \det(\mathbf{1} + (\mathbf{LR})^{-1}\mathbf{L\Delta R}). \end{aligned} \tag{3.12}$$

The last expression is the determinant of $N_{\sigma} \times N_{\sigma}$ matrices. What is remarkable is that this determinant is equal to that of a larger, $N \times N$, matrix:

$$\mathscr{R}^{\sigma} = \det(\mathbf{1} + \mathbf{\Delta R}(\mathbf{LR})^{-1}\mathbf{L}) = \det(\mathbf{1} + \mathbf{\Delta}(\mathbf{1} - \mathbf{G}^{\sigma})). \tag{3.13}$$

(See Appendix V.) We finally write the ratio of probabilities as

$$\mathscr{R} = \mathscr{R}^{\uparrow}\mathscr{R}^{\downarrow} = \prod_{\sigma = \uparrow, \downarrow} \det(\mathbf{1} + \mathbf{\Delta}^{\sigma}(i, \tau)(\mathbf{1} - \mathbf{G}^{\sigma}(\tau, \tau))). \tag{3.14}$$

For the finite-temperature algorithm, we proceed in a similar fashion:

$$\begin{aligned}\mathscr{R}^{\sigma} &= \frac{\det(\mathbf{1} + \mathbf{B}^{\sigma}(\beta, \tau)(\mathbf{1} + \mathbf{\Delta}^{\sigma}(i, \tau))\mathbf{B}^{\sigma}(\tau, 0))}{\det(\mathbf{1} + \mathbf{B}^{\sigma}(\beta, \tau)\mathbf{B}^{\sigma}(\tau, 0))} \\ &= \frac{\det(\mathbf{1} + \mathbf{B}^{\sigma}(\beta, 0) + \mathbf{B}^{\sigma}(\beta, \tau)\mathbf{\Delta}^{\sigma}(i, \tau)\mathbf{B}^{\sigma}(\tau, 0))}{\det(\mathbf{1} + \mathbf{B}^{\sigma}(\beta, 0))} \\ &= \det(\mathbf{1} + (\mathbf{1} + \mathbf{B}^{\sigma}(\beta, 0))^{-1}\mathbf{B}^{\sigma}(\beta, \tau)\mathbf{\Delta}^{\sigma}(i, \tau)\mathbf{B}^{\sigma}(\tau, 0)) \\ &= \det(\mathbf{1} + \mathbf{\Delta}^{\sigma}(i, \tau)\mathbf{B}^{\sigma}(\tau, 0)(\mathbf{1} + \mathbf{B}^{\sigma}(\beta, 0))^{-1}\mathbf{B}^{\sigma}(\beta, \tau)) \\ &= \det(\mathbf{1} + \mathbf{\Delta}^{\sigma}(i, \tau)(\mathbf{1} - \mathbf{G}^{\sigma}(\tau, \tau))). \end{aligned}$$

This time, the cyclic rearrangement of factors in the determinant is straightforward since the factors are square matrices with well-defined inverses. The ratio of probabilities has exactly the same form [eq. (3.14)] in terms of the Green's function as in the zero-temperature case.

At this point, one might seem hard pressed to justify all this formalism. In fact, however, eq. (3.14) provides us with a very fast means of evaluating the ratio $p[x']/p[x]$. Since field variables couple only to local fermionic degrees of freedom, a change in only field variable leads to correction matrices $\mathbf{\Delta}^{\sigma}$ which have only one nonzero matrix element. The determinants in eq. (3.14) become

very easy to evaluate and the Metropolis ratio reduces to

$$\mathscr{R} = \mathscr{R}^{\uparrow}\mathscr{R}^{\downarrow} = \prod_{\sigma=\uparrow,\downarrow} (1 + \Delta^{\sigma}_{ii}(i,\tau)(1 - G^{\sigma}_{ii}(\tau,\tau))). \tag{3.15}$$

This final expression is not a product of matrix determinants, but of simple scalars which are readily computed provided that the diagonal matrix elements of the equal-time Green's functions $\mathbf{G}^{\sigma}(\tau,\tau)$ are known. If each bosonic field were coupled to fermionic degrees of freedom on several near-neighbor sites or if several field variables were changed, then $\mathbf{\Delta}^{\sigma}$ would be less sparse and eq. (3.15) would be, correspondingly, more involved.

3.6. *Green's function updating*

At first glance, it appears that we have accomplished nothing. While the updating probability $\mathscr{R} = p[x']/p[x]$ can be simply expressed in terms of elements of the Green's functions (3.9) or (3.10), the evaluation of the $\mathbf{G}^{\sigma}$, requiring inverses of matrices, is at least as difficult as calculating the determinants (3.5) or (3.6) outright.

In simulations, however, we compute the inverses fairly infrequently. Whenever we change either an external field $x_i(\tau)$ or the time τ, we use the BSS algorithm (Scalapino and Sugar 1981, Blankenbecler et al. 1981a, b) to update the Green's functions in the finite-temperature approach. We will describe similar procedures for maintaining the relevant inverse for the zero-temperature algorithm. It must be conceded that all elements of the inverse must be updated – even if the field is changed on only one site in space–imaginary-time. Thus, the computation is still relatively expensive – of order N_{σ}^2 operations per update for the zero-temperature algorithm and of order N^2 operations at finite temperatures – and is the dominant portion of the computation. On the other hand, we will generally avoid the more expensive procedure of calculating the inverses from scratch, which would be of order N_{σ}^3 and N^3 operations at zero and finite temperature, respectively. The question is then, "If a change is made in the field variables, how may we update the inverses in eqs. (3.9) and (3.10) efficiently?" In answering the question we will again assume the changes in the single-particle propagators to be described in terms of the correction matrices $\mathbf{\Delta}^{\sigma}$ as $\mathbf{B}^{\sigma}(\tau,0) \to (\mathbf{1} + \mathbf{\Delta}^{\sigma}(i,\tau))\mathbf{B}^{\sigma}(\tau,0)$.

For the zero-temperature case,

$$\begin{aligned}(\mathbf{L}^{\sigma}\mathbf{R}^{\sigma})^{-1} &\to (\mathbf{L}^{\sigma}(\mathbf{1} + \mathbf{\Delta}^{\sigma})\mathbf{R}^{\sigma})^{-1} \\ &= (\mathbf{L}^{\sigma}\mathbf{R}^{\sigma} + \mathbf{L}^{\sigma}\mathbf{\Delta}^{\sigma}\mathbf{R}^{\sigma})^{-1} \\ &= (\mathbf{1} + (\mathbf{L}^{\sigma}\mathbf{R}^{\sigma})^{-1}\mathbf{L}^{\sigma}\mathbf{\Delta}^{\sigma}\mathbf{R}^{\sigma})^{-1}(\mathbf{L}^{\sigma}\mathbf{R}^{\sigma})^{-1}.\end{aligned}$$

Of course, we already have $(\mathbf{L}^{\sigma}\mathbf{R}^{\sigma})^{-1}$, it is inverse before updating. We compute the other inverse, $(\mathbf{1} + (\mathbf{L}^{\sigma}\mathbf{R}^{\sigma})^{-1}\mathbf{L}^{\sigma}\mathbf{\Delta}^{\sigma}\mathbf{R}^{\sigma})^{-1}$, by making use of the fact that the $\mathbf{\Delta}^{\sigma}$ are sparse. General solutions are given by the Sherman–Woodbury and

Morrison formulas (Press et al. 1988, Rice 1983). For our simple case, where the only nonzero matrix element of $\mathbf{\Delta}^\sigma$ is Δ_{ii}^σ, we assume

$$(\mathbf{1} + (\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\mathbf{L}^\sigma\mathbf{\Delta}^\sigma\mathbf{R}^\sigma)^{-1} = (\mathbf{1} + x(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\mathbf{L}^\sigma\mathbf{\Delta}^\sigma\mathbf{R}^\sigma),$$

where x is some scalar which we must determine. Since

$$\mathbf{\Delta}^\sigma\mathbf{R}^\sigma(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\mathbf{L}^\sigma\mathbf{\Delta}^\sigma = \mathbf{\Delta}^\sigma(\mathbf{1} - \mathbf{G}^\sigma)\mathbf{\Delta}^\sigma = \Delta_{ii}^\sigma(1 - G_{ii}^\sigma)\mathbf{\Delta}^\sigma$$

is simply a scalar multiple of $\mathbf{\Delta}^\sigma$, our Ansatz gives

$$\begin{aligned}\mathbf{1} &= (\mathbf{1} + (\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\mathbf{L}^\sigma\mathbf{\Delta}^\sigma\mathbf{R}^\sigma)(\mathbf{1} + x(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\mathbf{L}^\sigma\mathbf{\Delta}^\sigma\mathbf{R}^\sigma)\\ &= \mathbf{1} + (1 + x + x\Delta_{ii}^\sigma(1 - G_{ii}^\sigma))(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\mathbf{L}^\sigma\mathbf{\Delta}^\sigma\mathbf{R}^\sigma,\end{aligned}$$

or

$$x = -\frac{1}{1 + \Delta_{ii}^\sigma(1 - G_{ii}^\sigma)}.$$

This, of course, is simply $-1/\mathscr{R}^\sigma$. The updating equation is now

$$(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1} \to (\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\frac{1}{\mathscr{R}^\sigma}(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}\mathbf{L}^\sigma\mathbf{\Delta}^\sigma\mathbf{R}^\sigma(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1} \tag{3.16}$$

for single-variable updating. The updating is fast because the $\mathbf{\Delta}^\sigma$ are sparse. Since the Green's functions are $N \times N$ matrices while the $(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}$ are only $N_\sigma \times N_\sigma$, we will maintain only the $(\mathbf{L}^\sigma\mathbf{R}^\sigma)^{-1}$ in zero-temperature calculations, computing elements of the $\mathbf{G}^\sigma$ as they are needed. Nevertheless, one could still write updating equations for the real-space Green's functions. From eqs. (3.9) and (3.16),

$$\mathbf{G}^\sigma \to \mathbf{G}^\sigma - \frac{1}{\mathscr{R}^\sigma}\mathbf{G}^\sigma\mathbf{\Delta}^\sigma(\mathbf{1} - \mathbf{G}^\sigma). \tag{3.17}$$

The corresponding derivation for the finite-temperature algorithm is more straightforward:

$$\begin{aligned}\mathbf{G}^\sigma(\tau, \tau) &= (\mathbf{1} + \mathbf{B}^\sigma)^{-1}\\ &\to (\mathbf{1} + (\mathbf{1} + \mathbf{\Delta}^\sigma(\tau, \tau))\mathbf{B}^\sigma)^{-1}\\ &= (\mathbf{1} + \mathbf{B}^\sigma)^{-1}(\mathbf{1} + \mathbf{\Delta}^\sigma(\tau, \tau)\mathbf{B}^\sigma(\mathbf{1} + \mathbf{B}^\sigma)^{-1})^{-1}\\ &= \mathbf{G}^\sigma(\mathbf{1} + \mathbf{\Delta}(\mathbf{1} - \mathbf{G}))^{-1},\end{aligned}$$

writing $\mathbf{B}^\sigma(\tau, 0)\mathbf{B}^\sigma(\beta, \tau)$ simply as $\mathbf{B}^\sigma$. Using an Ansatz $(\mathbf{1} + x\mathbf{\Delta}^\sigma(\mathbf{1} - \mathbf{G}^\sigma))$ again for the inverse $(\mathbf{1} + \mathbf{\Delta}^\sigma(\mathbf{1} - \mathbf{G}^\sigma))^{-1}$, we find $x = -1/\mathscr{R}^\sigma$ and

$$\mathbf{G}^\sigma \to \mathbf{G}^\sigma - \frac{1}{\mathscr{R}^\sigma}\mathbf{G}^\sigma\mathbf{\Delta}^\sigma(\mathbf{1} - \mathbf{G}^\sigma)$$

as before.

3.7. "Wrapping" Green's functions

Up to this point, we have described algorithms for determining the Metropolis acceptance/rejection ratio $\mathscr{R}$ and for updating the Green's functions $\mathbf{G}^{\sigma}$ in the event of one accepting a proposed field-variable change. These algorithms are formally equivalent for the zero- and finite-temperature approaches. Our discussion, however, has always involved the equal-time Green's functions $\mathbf{G}^{\sigma}(\tau, \tau)$ at a particular imaginary time τ. What happens when we wish to consider Monte Carlo hits at other imaginary times?

For the zero-temperature formalism, we construct the Green's functions out of the products $\mathbf{L}(\tau)$ and $\mathbf{R}(\tau)$ from eqs. (3.7) and (3.8). The inverse

$$(\mathbf{L}(\tau)\mathbf{R}(\tau))^{-1} = (\mathbf{P}_{\mathrm{L}}\mathbf{B}(\beta, \tau)\mathbf{B}(\tau, 0)\mathbf{P}_{\mathrm{R}})^{-1}$$
$$= (\mathbf{P}_{\mathrm{L}}\mathbf{B}(\beta, 0)\mathbf{P}_{\mathrm{R}})^{-1},$$

that we must compute is independent of the imaginary time τ at which we perform our Metropolis updates. The way we use this inverse to evaluate elements of the Green's functions $\mathbf{G}(\tau, \tau) = \mathbf{1} - \mathbf{R}(\tau)(\mathbf{L}(\tau)\mathbf{R}(\tau))^{-1}\mathbf{L}(\tau)$, however, does depend on τ. Thus, when we move from one time to another, we must modify $\mathbf{L}(\tau)$ and $\mathbf{R}(\tau)$. For example, for $\tau' > \tau$,

$$\mathbf{L}(\tau') = \mathbf{L}(\tau)\mathbf{B}(\tau', \tau)^{-1}$$

and

$$\mathbf{R}(\tau') = \mathbf{B}(\tau', \tau)\mathbf{R}(\tau).$$

Modification of the left and right matrices $\mathbf{L}(\tau)$ and $\mathbf{R}(\tau)$ as we move from one value of τ to another, then, involves moving factors of $\mathbf{B}$ from the left side of $\mathbf{R}$ to the right side of $\mathbf{L}$ or vice versa. This can be achieved by one of two means. In the first option, suitable for small changes in τ, we tack on factors of $\mathbf{B}$ on the side by straightforward multiplication and delete these factors from the other side by multiplication of the corresponding inverses. From eq. (3.3) and the checkerboard breakup, we see that such multiplications are made up of factors $\mathbf{A}$ and $\exp(\Delta\tau\,\mathbf{T})$ that are sparse and trivial to invert.

The second option, suitable for large changes in τ, eliminates a number of these multiplications. In such a case, as one builds up $\mathbf{L}$ and $\mathbf{R}$ from $\mathbf{P}_{\mathrm{L}}$ and $\mathbf{P}_{\mathrm{R}}$, one stores partial products along the way. To string additional factors of $\mathbf{B}$ onto one side requires the multiplications we just discussed. Deletion of such factors from the other side, however, is performed more trivially by recalling a previously stored partial product. For example, in fig. 3a, the partial products $\mathbf{P}_{\mathrm{L}}$, $\mathbf{L}(\beta - \tau_0), \ldots, \mathbf{L}(\tau + \tau_0)$, and $\mathbf{L}(\tau)$, generated in the computation of $\mathbf{L}(\tau)$ are stored, as are the by-products $\mathbf{P}_{\mathrm{R}}$, $\mathbf{R}(\tau_0)$, $\mathbf{R}(2\tau_0), \ldots, \mathbf{R}(\tau)$ of the $\mathbf{R}(\tau)$ computation. To move from τ to a fairly distant imaginary time $\tau + \tau_0$, we need the factors $\mathbf{L}(\tau + \tau_0)$ and $\mathbf{R}(\tau + \tau_0)$. The right factor $\mathbf{R}(\tau + \tau_0) = \mathbf{B}(\tau + \tau_0, \tau)\mathbf{R}(\tau)$ is easily computed by successively multiplying by $\mathbf{B}(\tau + \tau_0, \tau)$, a series of sparse

$$\begin{array}{ccccccccccc}
\mathbf{L}(\beta)=\mathbf{P}_L & \mathbf{L}(\beta-\tau_0) & \ldots & \mathbf{L}(\tau+\tau_0) & \mathbf{L}(\tau) & \mathbf{R}(\tau) & \ldots & \mathbf{R}(2\tau_0) & \mathbf{R}(\tau_0) & \mathbf{R}(0)=\mathbf{P}_R \\
\Downarrow & \Downarrow & \ldots & \Downarrow & \Downarrow & \Downarrow & \ldots & \Downarrow & \Downarrow & \Downarrow \\
\mathbf{L}(\beta)=\mathbf{P}_L & \mathbf{L}(\beta-\tau_0) & \ldots & \mathbf{L}(\tau+\tau_0) & \mathbf{R}(\tau+\tau_0) & \mathbf{R}(\tau) & \ldots & \mathbf{R}(2\tau_0) & \mathbf{R}(\tau_0) & \mathbf{R}(0)=\mathbf{P}_R
\end{array}$$

Fig. 3. Computer storage scheme. As one builds up $\mathbf{L}(\tau)$ and $\mathbf{R}(\tau)$, needed for updating at time τ, one stores the intermediate partial products $\mathbf{P}_L, \ldots, \mathbf{L}(\tau)$ and $\mathbf{P}_R, \ldots, \mathbf{R}(\tau)$, as in (a). When one moves to imaginary time $\tau + \tau_0$, $\mathbf{L}(\tau + \tau_0)$ is already available as a previously stored partial product. Further, $\mathbf{R}(\tau + \tau_0)$ is easily computed from $\mathbf{R}(\tau)$ and may be stored in lieu of $\mathbf{L}(\tau)$, which is no longer needed, as in (b).

matrices, with the partial product $\mathbf{R}(\tau)$, which of course is already available. The left factor $\mathbf{L}(\tau + \tau_0)$ has already been stored in a previous calculation. At $\tau + \tau_0$, matrices are stored as in fig. 3b.

Storing partial products in this manner has the added advantage that roundoff errors will not accumulate from the repeated adding-on of factors and the subsequent deletion by inverse multiplication. As we will discuss, difficulties from roundoff errors can be devastating. In practice, we sweep over all times τ, through a combination of coarse and fine movements in imaginary time, using both of the previous approaches.

For the finite-temperature approach, we advance to a later imaginary time $\tau' > \tau$ by writing the Green's function as

$$\begin{aligned}
\mathbf{G}(\tau', \tau') &= (\mathbf{1} + \mathbf{B}(\tau', 0)\mathbf{B}(\beta, \tau'))^{-1} \\
&= (\mathbf{1} + \mathbf{B}(\tau', \tau)\mathbf{B}(\tau, 0)\mathbf{B}(\beta, \tau)\mathbf{B}(\tau', \tau)^{-1})^{-1} \\
&= \mathbf{B}(\tau', \tau)\mathbf{G}(\tau, \tau)\mathbf{B}(\tau', \tau)^{-1}. \qquad (3.18)
\end{aligned}$$

Of course, $\mathbf{B}(\tau', \tau) = \mathbf{B}(\tau', \tau' - \Delta\tau) \cdots \mathbf{B}(\tau + \Delta\tau, \tau)$. In eq. (3.18), we are simply taking factors of $\mathbf{B}(\tau + \Delta\tau, \tau)$ off the right side of $\mathbf{G}$ (by inverse multiplication) and "wrapping" them onto the left side. It is this picture that we describe as "wrapping" matrices around the Green's functions. Naturally, to go from larger imaginary times to smaller ones, we simply wrap matrices in the opposite direction. As before, multiplication by $\mathbf{B}$ matrices and inverse multiplication by $\mathbf{B}^{-1}$ matrices is straightforward and relatively fast, since the components of the $\mathbf{B}$'s are either diagonal or they are easy to approximate with the checkerboard breakup. Of course, we may also choose to compute and store partial products of $\mathbf{B}(\beta, \tau)$ and $\mathbf{B}(\tau, 0)$ in a way similar to that in the zero-temperature case.

3.8. Measurement estimators

Physical measurements of electronic properties appear in analytical calculations and in numerical simulations as expectation values of products of fermion creation and annihilation operators. In this section, we discuss several practical points as to how to measure such expectation values in simulations.

In simulations, as in analytical calculations, Wick's Theorem is used to convert expectation values of products of operators into sums of products of expectation values of all possible pair-wise contractions of creation and annihilation operators. The expectation values of these contractions are elements of the Green's functions, which must be calculated and maintained for updating field variables. (In the zero-temperature formalism, we maintain the inverse $(\mathbf{L}^\sigma \mathbf{R}^\sigma)^{-1}$, from which elements of the Green's functions are readily calculated.) Thus, Wick's Theorem expresses results of measurements in terms of the central objects of a simulation.

As an example, consider the z-component of the antiferromagnetic structure factor

$$S(\pi, \pi) = \frac{1}{N} \sum_{ij} (-)^{i-j} \langle (n_{i\uparrow} - n_{i\downarrow})(n_{j\uparrow} - n_{j\downarrow}) \rangle, \tag{3.19}$$

where N is the number of sites on the lattice, $n_{i\sigma}$ is the number of electrons with spin σ on site i, and $(-)^{i-j}$ is $+1$ if i and j are on the same sublattice, but -1 if they are not. We estimate S in eq. (3.19) by averaging measurements over N_m different samples of the bosonic field variables, where N_m is sufficiently large so that we can gather good statistics. Thus,

$$S(\pi, \pi) = \frac{1}{N_m} \sum_{\{x\}} \frac{1}{N} \sum_{ij} (-)^{i-j} (\langle n_{i\uparrow} n_{j\uparrow} \rangle_x + \langle n_{i\downarrow} n_{j\downarrow} \rangle_x - \langle n_{i\uparrow} n_{j\downarrow} \rangle_x - \langle n_{i\downarrow} n_{j\uparrow} \rangle_x), \tag{3.20}$$

where the expectation values $\langle\ \rangle_x$ are now for specific configurations of field variables $\{x\}$. For any such configuration, the electrons are completely decoupled from each other, which allows us to apply Wick's Theorem. Since the Hubbard–Stratonovich transformation block diagonalizes the problem in electron spin, contractions $\langle c_\sigma^\dagger c_{-\sigma} \rangle_x$ between operators of different spins are necessarily zero. Hence, the cross products

$$\langle n_{i\sigma} n_{j-\sigma} \rangle_x = \langle n_{i\sigma} \rangle_x \langle n_{j-\sigma} \rangle_x = (1 - G_{ii}^\sigma)(1 - G_{jj}^{-\sigma})$$

in eq. (3.20) decouple nicely. In contrast, products in eq. (3.20) of like-spin factors produce more nonzero contractions:

$$\begin{aligned} \langle c_{i\sigma}^\dagger c_{i\sigma} c_{j\sigma}^\dagger c_{j\sigma} \rangle_x &= \langle c_{i\sigma}^\dagger c_{i\sigma} \rangle_x \langle c_{j\sigma}^\dagger c_{j\sigma} \rangle_x + \langle c_{i\sigma}^\dagger c_{j\sigma} \rangle_x \langle c_{i\sigma} c_{j\sigma}^\dagger \rangle_x \\ &= (1 - G_{ii}^\sigma)(1 - G_{jj}^\sigma) + (\Delta_{ij} - G_{ji}^\sigma) G_{ij}^\sigma, \end{aligned} \tag{3.21}$$

where Δ_{ij} is the Kronecker delta function, which we use in reordering operators in the form required by our definition of the Green's function. Notice that in the case $i = j$, eq. (3.21) reduces to $(1 - G_{ii}^\sigma)$, which we would expect for $\langle n_{i\sigma} n_{i\sigma} \rangle_x = \langle n_{i\sigma} \rangle_x$.

Using Wick's Theorem, any electronic observable is easily reduced to averages of combinations of Green's function elements. Accordingly, the measurement process is straightforward in principle. There are several observations and points, however, that are useful to discuss.

In stochastic simulations, long computer runs are needed to collect many measurements and so reduce the statistical error. When the time that is needed to generate independent statistical samples is relatively large, it is necessary to gather as much data as possible from each sample. Thus, symmetries of the Hamiltonian are often exploited to construct many estimators of the same quantity. For example, to estimate the bond charge (kinetic energy)

$$\langle c_i^\dagger c_{i+\hat{\alpha}} \rangle,$$

we may average measurements of

$$\langle c_{i\sigma}^\dagger(\tau) c_{i+\hat{\alpha}\sigma}(\tau) \rangle$$

over many values of lattice position i, electronic spin σ, spatial neighbors $\hat{\alpha}$, and, in the case of the finite-temperature formalism which is periodic in τ, imaginary time τ. Due to fluctuations in the bosonic field variables $\{x\}$, individual configurations will not obey the symmetries of the Hamiltonian. Averages of measurements over many configurations will.

On the other hand, there may well be features in a simulation that break symmetry artificially. For example, in the Hubbard model, the staggered magnetization

$$M_{(\pi,\pi)}^\alpha = \sum_i (-)^i (c_{i\uparrow}^\dagger \quad c_{i\downarrow}^\dagger) \sigma^\alpha \begin{pmatrix} c_{i\uparrow} \\ c_{i\downarrow} \end{pmatrix},$$

is rotationally invariant, σ^α being the Pauli spin matrices. Due to the symmetry breaking of the Hubbard–Stratonovich transformation, however, the longitudinal estimator

$$\frac{1}{N} \langle (M_{(\pi,\pi)}^z)^2 \rangle$$

produces much noisier estimates of the antiferromagnetic structure factor in finite-temperature simulations (Hirsch 1987) than the rotationally equivalent transverse estimator

$$\frac{1}{2N} \langle (M_{(\pi,\pi)}^x)^2 + (M_{(\pi,\pi)}^y)^2 \rangle.$$

In ground-state calculations, symmetries may be broken still further by the trial wavefunction $|\psi\rangle$ (Sorella et al. 1988). Thus, the symmetrized measurement

$$\frac{1}{3N} \langle (M_{(\pi,\pi)}^x)^2 + (M_{(\pi,\pi)}^y)^2 + (M_{(\pi,\pi)}^z)^2 \rangle$$

may provide very good estimates of the structure factor (3.19), while a surprisingly long imaginary time β may be required to project a symmetric ground state out of $|\psi\rangle$. In summary, while the true quantum-mechanical wavefunction may be the symmetric superposition of many symmetry-breaking contributions, any one sample in a numerical simulation will not be symmetric. Hence, symmetrized estimators often provide great improvements in performance.

While it may be advantageous to collect formally redundant measurements, there is an added computational cost associated with computing the extra averages. In addition, these extra measurements are correlated to some degree so that they may give little additional information and hence may not be worth the extra cost. In practice, then, one averages over only a limited set of estimators. Further, while the elements of the Green's functions are always available during the course of a simulation, one typically chooses to employ several updating sweeps through the space–time lattice between measurements, since successive configurations may well be correlated.

Simulations generate extremely long streams of measurements. By "binning" these streams – dividing the streams into bins of equal lengths and reporting bin averages instead of individual measurements – one can both reduce the volume of output and estimate the statistical uncertainties in the measurements. The length of the bins must be long compared to the correlation time of the simulation. Thus, while consecutive samples of the simulations are generally highly correlated, averages from successive bins may be treated as being statistically independent. If A_i are the bin averages for a certain observable A, then

$$\bar{A} = \frac{1}{N_{\text{bin}}} \sum_i A_i \tag{3.22}$$

is, of course, the estimate of the observable, while

$$\sigma(A) \approx \frac{1}{\sqrt{N_{\text{bin}} - 1}} \sigma(A_i) = \frac{1}{\sqrt{N_{\text{bin}} - 1}} \sqrt{\frac{\sum_i A_i^2}{N_{\text{bin}}} - \bar{A}^2} \tag{3.23}$$

is the statistical uncertainty in the estimate (3.22). Here, N_{bin} is the number of bins.

A typical Monte Carlo run may entail 10 to 20 bins of 250 to 500 measurements each, with two to five sweeps of the space–time lattice between measurements. The configuration of the system is generally initialized randomly, so that 500–1000 sweeps should be performed to equilibrate the system before measurements are made. Experimentation and individual standards are the final arbiters in deciding how to set these parameters. Since different physical observables generally have different statistical properties, run parameters will also depend on the quantities to be measured.

The errors associated with measurements reported will be due to both the nonzero discretization parameter $\Delta\tau$ and to statistical fluctuations. Our use of

path integrals requires us to study the limit $\Delta\tau \to 0$. In practice, measurements obey small-$\Delta\tau$ scaling laws (Suzuki 1985, Fye 1986, Fye and Scalettar 19xx) – in most cases, either as $\Delta\tau$ or as $\Delta\tau^2$ – for surprisingly large values of $\Delta\tau$. This is fortunate since the reduced number of degrees of freedom for large-$\Delta\tau$, coarsely discretized path integrals allows faster sweeps through the space–time lattice and more movement through phase space for each sweep. The size of statistical errors depends greatly on the quantity being measured. For several thousand sweeps, we would expect path-integral simulations to estimate the energy – generally the best-behaved observable – to fractions of one percent. In contrast, long-range observables, such as the antiferromagnetic structure factor, may well have up to 10% fluctuations for the same number of sweeps.

3.9. Minus signs

Importance sampling requires that the weights $p[x]$ all be positive definite so that they may be interpreted as probabilities. In case they are not, the weights governing the importance sampling should be the absolute values $|p[x]|$. The expectation value (2.22) becomes

$$\frac{\sum_{\{x\}} \langle A \rangle_x p[x]}{\sum_{\{x\}} p[x]} = \frac{\sum_{\{x\}} \langle A \rangle_x s[x]|p[x]|}{\sum_{\{x\}} s[x]|p[x]|} = \frac{\sum'_{\{x\}} \langle A \rangle_x s[x]}{\sum'_{\{x\}} s[x]}, \tag{3.24}$$

where $s[x]$ is the sign of $p[x]$ and the primed sums are over configurations $\{x\}$ generated according to the weights $|p[x]|$.

Of course, the "minus-sign" problem in simulations of fermions is that the numerator and denominator in eq. (3.24) can be much smaller than the statistical fluctuations in the measurement process. In such a case, the simulation is defeated. Further, "nodal surfaces" can possibly exist in $\{x\}$-phase space, even if there is no sign problem. In eqs. (3.5) and (3.6), we see that each weight $p[x] = \prod_\sigma p^\sigma[x]$ is a product of $\sigma = \uparrow$ and $\downarrow$ contributions. As long as the individual p^σ can change sign, phase space will be partitioned by "potential barriers" or "nodal surfaces". If these surfaces coincide throughout phase space, however, the product $p[x]$ will have the same sign in every partition. This is in fact the case for the half-filled Hubbard model and the symmetric Anderson lattice.

4. Stabilization

As we have seen, the Green's function, and not the partition function, is the central object of the simulations. It is needed to perform importance sampling and to make measurements. As one sweeps from one imaginary time τ to another, **G** can be produced on the new time slice by "wrapping" matrices, as described in the previous section. Unfortunately, round-off errors accumulate in

this procedure, making it necessary occasionally to recompute the Green's function from scratch. The more serious difficulty is that at low temperatures, as β becomes large, the Green's function cannot be computed at all. The reason is that as many **B** matrices are multiplied together, the product becomes more and more ill-conditioned, with exponentially divergent numerical scales. In calculating the Slater determinants in eqs. (3.5) and (3.6), one computes small differences of large matrix elements. These differences are very inaccurate, dominated by the noise in the least significant bits of the matrix elements when the calculations are performed on finite-precision computers.

More physically, the single-particle propagator **B** amplifies "low-energy" states for any particular configuration of the bosonic fields while attenuating the "high-energy" states. In addition, the states near some intermediate "Fermi energy", buried exponentially by the states at the bottom of the "band", describe the important physical phenomena in fermionic systems. Although this correspondence to a single-particle band picture is loose – our strongly correlated systems are actually sums over *many* single-particle, time-varying problems – the fact remains that the fermionic behavior of a model gets "drowned" by other large numerical scales. Conventional simulations ultimately fail simply because these important small-scale features cannot be extracted from $\mathbf{B}(\beta, \tau)$ and $\mathbf{B}(\tau, 0)$ using computers with finite precision.

The sparse matrices (3.3), which make up **B**, can be represented with high precision on a finite-precision computer. Likewise the matrix elements of the Green's function can also be represented, since these elements are generally well-behaved numbers. It is only the intermediate products in the calculation of **G**, the single-particle propagator for large imaginary times, that are difficult to represent on finite-precision machines without loss of important information. It is our task, then, to organize the calculation of **G** such that no essential information is lost.

Several recent efforts (White et al. 1988, Hirsch 1988) at low-temperature stabilization use matrices of higher dimensions, at a substantial cost of computer time and memory. In this section, we describe even more recent developments (Loh Jr et al. 1989) in efficiently stabilizing the computation of the matrix products and the matrix inversions needed to get the Green's function. There are two keys to this approach. First, in forming matrix products, small scales are maintained explicitly, rather than implicitly as small differences of large numbers. Second, the varied numerical scales are combined only at the last step in the calculation of the Green's function, using some scales to cut off smaller, inconsequential scales. Resorting, once again, to a band picture, the numerical scale associated with a single-particle state of energy E relative to the chemical potential μ is $\exp(-\beta(E-\mu))$, which diverges or vanishes exponentially with β. On the other hand, the occupation of that state, a Green's function element, is $1/(\exp(\beta(E-\mu))+1)$, which ranges between 0 and 1. More generally, if we are able to keep the numerical scales $\exp(\beta(E-\mu))$ separated, we will be able to cut

off the ill-behaved scales in the final step in calculating **G** by adding terms of unit scale.

We will begin this section by discussing the stable multiplication of matrices. Then we will discuss various ways of cutting off the large and small scales to compute the Green's function for a variety of simulation applications.

4.1. Stable matrix multiplication

The *condition number* of a matrix is roughly the ratio of the largest singular value of the matrix to the smallest one and represents an upper bound to be amplification of errors in matrix multiplications. Our aim is to decompose ill-conditioned matrices by representing them in the form **UDV**, where the diagonal matrix **D** contains the diverging singular values explicitly and has the large condition number, but where **U** and **V** are "sufficiently well-conditioned", a property which will be made more precise later. If we choose both **U** and **V** to be orthogonal, the resulting decomposition is the singular-value decomposition (SVD), which is known to be very stable in performance. Unfortunately, the inner loops of SVD subroutines have vanishingly small lengths and so perform slowly on vector computers. In practice, the modified Gram–Schmidt (MGS) factorization **UDV**, where **U** is orthogonal and **D** is diagonal, but **V** is unit triangular, is preferable. The computation time for an MGS decomposition is, in some cases, up to 20 times less than that for SVD. Fortunately, unit-triangular matrices are sufficiently well-conditioned, so that large numbers of them can be multiplied in simulations without destroying stability.

To understand what we gain in decomposing an ill-conditioned matrix into such a **UDV** factorized form, consider the schematic multiplication

$$\mathbf{UDV} = \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix} \begin{pmatrix} X & & & \\ & X & & \\ & & X & \\ & & & x \end{pmatrix} \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix}$$

$$= \begin{pmatrix} X & X & X & X \\ X & X & X & X \\ X & X & X & X \\ X & X & X & X \end{pmatrix}.$$

Here, **U** and **V** are sketched as having only matrix elements of unit scale, while **D** has elements of many scales, which we have pivoted into descending order. Notice that once the multiplication has been performed, all of the elements of the product matrix are of the largest scale. The product is, essentially, an outer product of the first column of **U** and the first row of **V**. The smallest numerical

scales exist only implicitly, as small differences of large matrix elements, and they can never be recovered on a computer of finite precision.

In contrast, a matrix which displays its small scales explicitly can be factorized stably. Consider, for example, the factorization of a column-stratified matrix **M**:

$$\mathbf{U}^{-1}\mathbf{M}\mathbf{V}^{-1} = \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix} \begin{pmatrix} X & X & X & x \\ X & X & X & x \\ X & X & X & x \\ X & X & X & x \end{pmatrix} \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix}$$

$$= \begin{pmatrix} X & & & \\ & X & & \\ & & X & \\ & & & x \end{pmatrix} = \mathbf{D}.$$

Multiplication on the left of **M** by a transformation matrix only combines elements within a given column – elements of the same scale – and so causes no loss of information. Multiplication on the right by $\mathbf{V}^{-1}$ combines columns of different scales. This multiplication, however, does not overwrite any small-scale information so long as large-scale columns are scaled down appropriately before they are added into columns of smaller scale. Such factorization is in fact possible for both the SVD and MGS decompositions.

To compute the product of many matrices stably, we decouple the various scales present throughout the calculation. To illustrate this, we first imagine we have decomposed some partial product $\mathbf{B}(\tau, 0) = \mathbf{UDV}$. To extend the single-particle propagator to imaginary time $\tau + \tau_0$, we write

$$\mathbf{B}(\tau + \tau_0, 0) = \mathbf{B}(\tau + \tau_0, \tau)\mathbf{UDV} = (\mathbf{B}(\tau + \tau_0, \tau)\mathbf{UD})\mathbf{V}$$

$$= \left(\mathbf{B}(\tau + \tau_0, \tau)\mathbf{U} \begin{pmatrix} X & & & \\ & X & & \\ & & X & \\ & & & x \end{pmatrix} \right) \mathbf{V}$$

$$= \begin{pmatrix} X & X & X & x \\ X & X & X & x \\ X & X & X & x \\ X & X & X & x \end{pmatrix} \mathbf{V}$$

$$= (\mathbf{U}'\mathbf{D}'\mathbf{V}')\mathbf{V} = \mathbf{U}'\mathbf{D}'(\mathbf{V}'\mathbf{V}), \tag{4.1}$$

giving the decomposition of the new partial product $\mathbf{B}(\tau + \tau_0, 0)$. Here, τ_0 is the length of imaginary time for which the single-particle propagator can be

extended without swamping the machine precision. We have decomposed the stratified matrix $\mathbf{B}(\tau + \tau_0, \tau)\mathbf{UD}$ into $\mathbf{U'D'V'}$. The $\mathbf{V}$ matrices must be sufficiently well-conditioned in order that we can multiply many of them together stably; we can do this for both orthogonal (for SVD) and unit-triangular (for MGS) $\mathbf{V}$ matrices. The product $\mathbf{B}(\tau + \tau_0, \tau)\mathbf{U}$ is formed as a series of sparse-matrix multiplications (3.3) on $\mathbf{U}$.

Gram–Schmidt orthogonalization has long been used to stabilize matrix products (Wilkinson 1965). In the calculation of characteristic Lyapunov exponents from time series, transformations are linearized and represented as matrices at each time step. The stable computation of a long product of many such matrices is often carried out by decomposing partial products (Benettin and Galgani 1979, Shimada and Nagashima 1979). MacKinnon and Kramer (1983) have used orthogonalization in the scaling theory of electrons in disordered solids. Sugiyama and Koonin (1986) first proposed limited applications of these techniques to Monte Carlo simulations of fermions. Only recently (Loh Jr et al. 1989), have decomposition techniques been applied widely, and in a greater variety of contexts, to fermionic simulations. We will now describe how to use decomposed forms of the single-particle propagator to simulations of fermions.

4.2. Zero-temperature studies

Sugiyama and Koonin (1986) first used matrix-decomposition techniques in fermion Monte Carlo for equal-time measurements in a zero-temperature context. Recently, Sorella et al. (1989) used this stabilized approach with a Langevin updating of the auxiliary fields.

In the zero-temperature approach, the Green's function (3.9) is

$$\mathbf{G}^\sigma(\tau, \tau) = \mathbf{1} - \mathbf{R}^\sigma(\tau)(\mathbf{L}^\sigma(\tau)\mathbf{R}^\sigma(\tau))^{-1}\mathbf{L}^\sigma(\tau).$$

The factorized forms for the $\mathbf{L}$ and $\mathbf{R}$ matrices are

$$\mathbf{L}^\sigma = \mathbf{V}_\mathrm{L}^\sigma \mathbf{D}_\mathrm{L}^\sigma \mathbf{U}_\mathrm{L}^\sigma, \qquad \mathbf{R}^\sigma = \mathbf{U}_\mathrm{R}^\sigma \mathbf{D}_\mathrm{R}^\sigma \mathbf{V}_\mathrm{R}^\sigma,$$

where the order of the $\mathbf{U}$, $\mathbf{D}$, and $\mathbf{V}$ matrices has been reversed for $\mathbf{L}^\sigma = \mathbf{P}_\mathrm{L}^\sigma \mathbf{B}^\sigma(\beta, \tau)$ since the product is built up on the right side, instead of on the left side as in eq. (4.1). In the zero-temperature, canonical (fixed-number) approach, $\mathbf{L}^\sigma$ ($\mathbf{R}^\sigma$) are $N_\sigma \times N$ ($N \times N_\sigma$) rectangular matrices, where the $\mathbf{U}_\mathrm{L}^\sigma$ ($\mathbf{U}_\mathrm{R}^\sigma$) are also $N_\sigma \times N$ ($N \times N_\sigma$) rectangular matrices, while the $\mathbf{D}_\mathrm{L}^\sigma$ ($\mathbf{D}_\mathrm{R}^\sigma$) and $\mathbf{V}_\mathrm{L}^\sigma$ ($\mathbf{V}_\mathrm{R}^\sigma$) are "small" $N_\sigma \times N_\sigma$ square matrices. Using the factorized forms in the expression for the Green's function, we find that

$$\begin{aligned}\mathbf{G}^\sigma &= \mathbf{1} - (\mathbf{U}_\mathrm{R}^\sigma \mathbf{D}_\mathrm{R}^\sigma \mathbf{V}_\mathrm{R}^\sigma)((\mathbf{V}_\mathrm{L}^\sigma \mathbf{D}_\mathrm{L}^\sigma \mathbf{U}_\mathrm{L}^\sigma)(\mathbf{U}_\mathrm{R}^\sigma \mathbf{D}_\mathrm{R}^\sigma \mathbf{V}_\mathrm{R}^\sigma))^{-1}(\mathbf{V}_\mathrm{L}^\sigma \mathbf{D}_\mathrm{L}^\sigma \mathbf{U}_\mathrm{L}^\sigma) \\ &= \mathbf{1} - \mathbf{U}_\mathrm{R}^\sigma(\mathbf{U}_\mathrm{L}^\sigma \mathbf{U}_\mathrm{R}^\sigma)^{-1}\mathbf{U}_\mathrm{L}^\sigma,\end{aligned}$$

i.e. the Green's function depends only on the rectangular $\mathbf{U}_\mathrm{L}^\sigma$ and $\mathbf{U}_\mathrm{R}^\sigma$ matrices but not at all on the square $\mathbf{V}_\mathrm{L}^\sigma$, $\mathbf{D}_\mathrm{L}^\sigma$, $\mathbf{V}_\mathrm{R}^\sigma$, or $\mathbf{D}_\mathrm{R}^\sigma$ transformation matrices.

In zero-temperature simulations, then, we can stabilize the calculation by orthonormalizing the rows of $\mathbf{L}^\sigma$ and the columns of $\mathbf{R}^\sigma$ without ever keeping or manipulating the linear transformations which effect the orthonormalization.

4.3. *Finite-temperature* $\boldsymbol{G}(\tau, \tau)$*: I*

For finite-temperature simulations, however, we are not interested in the effect of the single-electron propagator $\mathbf{B}$ on the preselected set of single-particle states represented in $\mathbf{P}_L$ and $\mathbf{P}_R$. Rather, we trace over all possible states. Thus, the transformation matrices must be kept. Since only $\mathbf{G}(\tau, \tau)$ is needed to update the Hubbard–Stratonovich field variables, we will begin by examining the equal-time Green's function. Fortunately, there is a simple prescription for calculating the $\mathbf{G}^\sigma(\tau, \tau)$ that can be added as a module to stabilize standard BSS codes.

The equal-time Green's function (3.10) is

$$\mathbf{G}^\sigma(\tau, \tau) = (\mathbf{1} + \mathbf{B}(\tau, 0)\mathbf{B}(\beta, \tau))^{-1}. \tag{4.2}$$

We build up the decomposition for the product of sparse factors of which $\mathbf{B}(\tau, 0)\mathbf{B}(\beta, \tau)$ is composed in the manner suggested by eq. (4.1). Now the Green's function becomes

$$\begin{aligned}\mathbf{G} &= (\mathbf{1} + \mathbf{UDV})^{-1} = \mathbf{V}^{-1}(\mathbf{U}^{-1}\mathbf{V}^{-1} + \mathbf{D})^{-1}\mathbf{U}^{-1} \\ &= \mathbf{V}^{-1}(\mathbf{U}'\mathbf{D}'\mathbf{V}')^{-1}\mathbf{U}^{-1} = (\mathbf{V}'\mathbf{V})^{-1}(\mathbf{D}')^{-1}(\mathbf{U}\mathbf{U}')^{-1}.\end{aligned} \tag{4.3}$$

Once we have formed the sum $\mathbf{U}^{-1}\mathbf{V}^{-1} + \mathbf{D}$, we decompose it into $\mathbf{U}'\mathbf{D}'\mathbf{V}'$ and the remaining factors are easily inverted and combined in any order.

What we have done is to isolate the divergent scales in $\mathbf{D}$ until they are combined with the unit elements of $\mathbf{U}^{-1}\mathbf{V}^{-1}$, which cut off these divergent scales much like the unit term cuts off $\exp(\beta(E - \mu))$ in the standard band picture.

In all, we define numerical stability in an operational fashion. While the implication all along has been that we do not lose information when the numerical scales of the problem are maintained explicitly, in fact, much information is lost. For example, if we multiply the inverse $(\mathbf{U}^{-1}\mathbf{V}^{-1} + \mathbf{D})^{-1}$ in eq. (4.3) by the original matrix $(\mathbf{U}^{-1}\mathbf{V}^{-1} + \mathbf{D})$, we would never recover the identity matrix – even if the original matrix and its computed inverse were both kept in decomposed form. Similarly, if we perform many operations, such as "wrapping", on the "stably" computed Green's function, small errors that are present in our representation of the Green's function on any finite-precision machine – even if $\mathbf{G}$ is kept in decomposed form – would accumulate until the errors became larger than the calculated matrix elements. Our operational definition of stability, then, is simply that we should be able to calculate matrix elements of the Green's function, the only object needed for the simulation importance sampling and measurement process, to several significant figures. The overwriting of small scales by unit numbers and of unit scales by big numbers in the

addition $\mathbf{U}^{-1}\mathbf{V}^{-1} + \mathbf{D}$ does not introduce any significant errors into the computation of the Green's function. Rather, this addition simply cuts off certain numerical scales. In a simple band picture, it is not important whether a state deep in the band is amplified by 10^{100} or by 10^{1000} – all that is important is that states deep in the band are filled. Similarly, it is not important whether high-energy states are attenuated by 10^{-100} or by 10^{-1000} – it only matters that such states are cut off. In our calculation of the Green's function, we generate the single-particle propagator in decomposed form, identifying the "transformation" matrices $\mathbf{U}$ and $\mathbf{V}$, which describe the "single-particle" states for a given set of Hubbard–Stratonovich fields, and the scales $\mathbf{D}$ associated with the different states. At this point, addition of $\mathbf{U}^{-1}\mathbf{V}^{-1}$ cuts off the divergent scales to identify which states have been amplified and which have been attenuated. It is for this reason that the overwriting of scales, which we have so tediously been avoiding, can be performed in this last step on $\mathbf{U}^{-1}\mathbf{V}^{-1} + \mathbf{D}$ without introducing any errors into the computation of Green's function. Indeed, in practice the exponentially divergent elements of $\mathbf{D}$ cannot be stored on a computer of finite dynamic range and so we cut them off at, say, 10^{+100} and 10^{-100}. It is enough for us to know simply which scales are "big" and which are "small".

Again, even while we are able to compute the matrix elements of the Green's function to reasonably high accuracy, we may "wrap" $\mathbf{G}$ only for a short imaginary time – of the scale τ_0, which was introduced in the discussion of eq. (4.1). Thus, $\mathbf{G}$ must be recomputed from scratch from time to time. The number of times $\mathbf{G}$ must be recomputed in one sweep of all time slices goes as β/τ_0. The number of decompositions one must perform in each such recomputation also goes as β/τ_0. Thus, the overhead associated with stabilizing the computation scales as $(\beta/\tau_0)^2$. While this scaling may dominate the computation at very low temperatures (large β) or for unstable problems on machines of small world lengths (small τ_0), we have found that for sets of problems – such as sufficiently low temperatures to study the ground state of a single-band Hubbard model on up to 16×16 lattices on a 64-bit machine – the stabilization overhead represents only a fraction of the running time.

4.4. Finite-temperature $\boldsymbol{G}(\tau, \tau)$: II

An alternative to the great deal of recomputation involved in re-evaluating eq. (4.3) for many different τ is to use computer memory to store partial products. Now, we imagine we have the decompositions of *two* partial products. Building up a partial product from the right, we define

$$\mathbf{B}(\tau, 0) = \mathbf{U}_\mathrm{R}\mathbf{D}_\mathrm{R}\mathbf{V}_\mathrm{R}, \tag{4.4}$$

and building up the remaining factors from the left, we define

$$\mathbf{B}(\beta, \tau) = \mathbf{V}_\mathrm{L}\mathbf{D}_\mathrm{L}\mathbf{U}_\mathrm{L}, \tag{4.5}$$

where the order of the factorization is reversed. In contrast with the zero-temperature case, all these matrices are square. Now, from eq. (4.2), we obtain

$$\begin{aligned}\mathbf{G}(\tau,\tau) &= (\mathbf{1} + \mathbf{U}_R\mathbf{D}_R\mathbf{V}_R\mathbf{V}_L\mathbf{D}_L\mathbf{U}_L)^{-1} \\ &= \mathbf{U}_L^{-1}(\mathbf{U}_R^{-1}\mathbf{U}_L^{-1} + \mathbf{D}_R\mathbf{V}_R\mathbf{V}_L\mathbf{D}_L)^{-1}\mathbf{U}_R^{-1}. \end{aligned} \tag{4.6}$$

Again, the inverse of the ill-conditioned sum can be stabilized by decomposing the sum and then inverting its individual pieces. Schematically, the piece to be inverted is

$$\begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix} + \begin{pmatrix} X & & & \\ & X & & \\ & & X & \\ & & & x \end{pmatrix} \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix} \begin{pmatrix} X & & & \\ & X & & \\ & & X & \\ & & & x \end{pmatrix}$$

$$= \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix} + \begin{pmatrix} XX & XX & Xx & Xx \\ XX & XX & Xx & Xx \\ xX & xX & xx & xx \\ xX & xX & xx & xx \end{pmatrix}.$$

Because we have kept the diagonal matrices on the outsides of the terms, elements of different scales are added together only in the last step, to "cut off scales", as before.

In building up the partial product $\mathbf{B}(\tau, 0)$, one may store the decompositions of the partial products $\mathbf{B}(\tau_0, 0)$, $\mathbf{B}(2\tau_0, 0)$, $\mathbf{B}(3\tau_0, 0)$, and so on. Thereafter, to move from some imaginary time τ to $\tau + \tau_0$, one must perform the appropriate series of sparse-matrix multiplications and another decomposition, as in eq. (4.1). To move from τ backward to $\tau - \tau_0$, on the other hand, one need only recall a previously stored partial product. This procedure is both faster and accumulates fewer roundoff errors than when applying and then stripping off factors with multiplication of matrices and then their inverses. Similarly, one may store the partial products $\mathbf{B}(\beta, \beta - \tau_0)$, $\mathbf{B}(\beta, \beta - 2\tau_0,)$, $\mathbf{B}(\beta, \beta - 3\tau_0)$, and so on, in building up $\mathbf{B}(\beta, \tau)$. In all, storage of these partial products and utilization of eq. (4.6) brings the stabilization overhead down to scaling as (β/τ_0). While it becomes necessary to store the decompositions of many partial products, the memory costs are affordable.

4.5. Unequal-time Green's function

Finally, while equal-time measurements can always be expressed in terms of averages of the equal-time Green's function $\mathbf{G}(\tau, \tau)$, application of Wick's

Theorem to time-dependent quantities such as susceptibilities

$$\chi(i\omega_n) = \int_0^\beta \langle \mathbf{A}^\dagger(\tau)\mathbf{A}(0)\rangle e^{i\omega_n\tau}\,d\tau$$

will generate contractions between fermion operators at different imaginary times. Thus, we will require matrix elements of the unequal-time Green's function

$$\mathbf{G}(\tau', \tau) = \mathbf{B}(\tau', \tau)(\mathbf{1} + \mathbf{B}(\tau, 0)\mathbf{B}(\beta, \tau))^{-1}.$$

For small changes $\tau' \to \tau''$ in the arguments of $\mathbf{G}$, multiplication by the single-particle propagator $\mathbf{B}(\tau'', \tau')$ produces the appropriate Green's function:

$$\mathbf{G}(\tau'', \tau) = \mathbf{B}(\tau'', \tau')\mathbf{G}(\tau', \tau).$$

As with "wrapping", however, this procedure cannot propagate Green's functions for more than some imaginary time – again, roughly τ_0. In this section, we describe two methods of stably forming

$$\mathbf{G}(\tau, 0) = \mathbf{B}(\tau, 0)(\mathbf{1} + \mathbf{B}(\beta, 0))^{-1} = (\mathbf{B}(\tau, 0)^{-1} + \mathbf{B}(\beta, \tau))^{-1}. \tag{4.7}$$

First, we can pattern our approach as closely as possible to eq. (4.3). Using the decompositions (4.4) and (4.5), we get

$$\mathbf{G}(\tau, 0) = (\mathbf{V}_R^{-1}\mathbf{D}_R^{-1}\mathbf{U}_R^{-1} + \mathbf{V}_L\mathbf{D}_L\mathbf{U}_L)^{-1}.$$

We then isolate the most ill-conditioned diagonal matrix. For example, if $\tau > \beta - \tau$, we write

$$\mathbf{G}(\tau, 0) = \mathbf{U}_R(\mathbf{D}_R^{-1} + \mathbf{V}_R\mathbf{V}_L\mathbf{D}_L\mathbf{U}_L\mathbf{U}_R)^{-1}\mathbf{V}_R = \mathbf{U}_R(\mathbf{U}'\mathbf{D}'\mathbf{V}')^{-1}\mathbf{V}_R, \tag{4.8}$$

proceeding somewhat as before. If, on the other hand, $\tau < \beta - \tau$, one would isolate $\mathbf{D}_L$. In preliminary tests, this decomposition has proven to behave surprisingly well, even in the case of $\tau \approx \beta - \tau$, where it is not clear which diagonal matrix to isolate.

Alternatively, we can arrange the factors in eq. (4.7) so that the two terms in the final, "cutoff" addition are stratified matrices with explicitly maintained small scales. We write

$$\mathbf{G}(\tau, 0) = \mathbf{U}_L^{-1}(\mathbf{P}\mathbf{D}_R^{-1}\mathbf{U}_R^{-1}\mathbf{U}_L^{-1} + \mathbf{P}\mathbf{V}_R\mathbf{V}_L\mathbf{D}_L)^{-1}\mathbf{P}\mathbf{V}_R. \tag{4.9}$$

Since we have been pivoting the largest singular values to the upper left-hand corner of the diagonal matrices, (4.9) includes a pivoting matrix

$$\mathbf{P} = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}$$

to order the elements of $\mathbf{D}_R^{-1}$ in the same way. Schematically, the sum to be inverted in eq. (4.9) is

$$\begin{pmatrix} X & X & X & X \\ X & X & X & X \\ X & X & X & X \\ x & x & x & x \end{pmatrix} + \begin{pmatrix} X & X & X & x \\ X & X & X & x \\ X & X & X & x \\ X & X & X & x \end{pmatrix}.$$

Clearly, the small scales in the problem are kept explicitly until this addition, when we would expect that the cutting off of scales would not introduce any significant errors. Indeed, eq. (4.9) generally outperforms eq. (4.8) in tests of Green's function computation on finite-precision machines, calculating matrix elements of $\mathbf{G}$ with errors one order of magnitude below those from eq. (4.8). Unfortunately, the algorithm also seems to fail occasionally. This is not currently understood.

5. *Results for Hubbard models*

The stabilization algorithms of the previous section have been tested in a variety of circumstances. As we noted in section 2, the massless Hubbard–Stratonovich fields are characterized predominantly by their fluctuations. Thus, a simple test of stability is to calculate the matrix elements of the Green's function in both single and double precision for random configurations of the auxiliary variables. The agreement between the elements from these two different calculations is typically at least to ten significant figures when the tests are performed on 64-bit computers. Further, Monte Carlo results on very small lattices (4 sites) can be compared to results from exact diagonalizations. Results for energies and ground-state correlation functions have been checked down to temperatures of $T = 0.01$ ($\beta = 100$), almost two orders of magnitude beyond the temperatures at which unstable calculations fail or the temperatures that are needed to project out the ground-state properties of this small system.

At the time this chapter was finished (March, 1989), work utilizing the recently developed stabilization techniques were just reaching the published literature. Here, we report some results for Hubbard models in two dimensions (White et al. 1989a, Scalettar et al. 1989, Scalettar et al. 1991). Simulations have been run on lattices up to 18×18 sites at reciprocal temperatures as high as $\beta = 32$. The Green's functions were never propagated very far ($\tau_0 \sim 0.5$–1) in imaginary time before the computation was restabilized. These results come from simulations using the discrete Hubbard–Stratonovich transformation and Monte Carlo updating with both the finite- and zero-temperature formulations. The longest runs needed the order of 10 hours on a CRAY X-MP computer.

5.1. Repulsive Hubbard model

The single-band, repulsive Hubbard model is the model most studied by quantum Monte Carlo methods. The bulk of these studies has occurred in the past two years in connection with various proposals regarding mechanisms for high-temperature superconductivity. These recent studies have confirmed the basic results of the original study by Hirsch (1987): at half filling, the model is antiferromagnetic and this antiferromagnetism rapidly diminishes as one dopes with electrons or holes to move away from half filling. The newer studies have also, quite naturally, focussed on calculations of various superconducting pairing correlation functions and susceptibilities over a range of dopings and on-site Coulomb interaction strengths. All calculations to date have shown that these measures of superconductivity are suppressed relative to their values when the Coulomb interaction is zero. Recently, other procedures for calculating attractive pairings have been suggested, but even with their use the superconducting state is yet to be seen (White et al. 1989b).

A traditional difficulty with simulations of fermions has been achieving sufficiently low temperatures to study the ground-state properties of a system. With the stabilized methods discussed in section 4, not only is it now numerically possible to go to extremely low temperatures, but the overhead associated with the stabilization is negligible. By computing with both the zero-temperature and finite-temperature algorithms, White et al. (1989a) have shown that by $\beta = 20$ the finite-temperature simulation measures ground-state properties. Because of the efficiency of these methods, computation time can be devoted to studies on increased lattice sizes rather than on stabilizing small-lattice calculations. For example, in fig. 4 the antiferromagnetic structure factor (3.19) is plotted against the reciprocal temperature β for various sizes of square lattices at

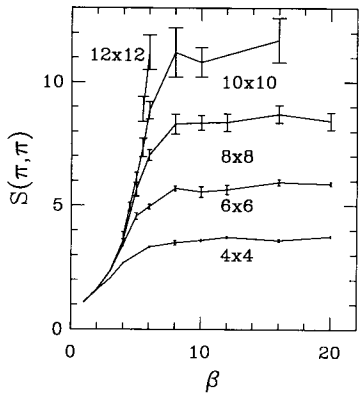

Fig. 4. The antiferromagnetic structure factor (3.19) as a function of reciprocal temperature β at $U = 4t$ for the repulsive Hubbard model in two dimensions (White et al. 1989a).

a coupling $U = 4t$, i.e. half of the bandwidth (White et al. 1989a). It is clear from this figure that there is no essential difficulty in going well beow the temperature that is needed to isolate the ground state. At high temperatures (low β) there is little size dependence. At low temperatures, however, the structure factor is seen to diverge with lattice size. Fits of the data to finite-size corrections predicted from spin wave theory suggest long-range antiferromagnetism in the ground state (White et al. 1989a).

The original work of Hirsch (1987) was also the first to discuss and illustrate the existence of a severe sign problem in simulations of this model. White et al. (1989a) also measured the average sign, the denominator in eq. (3.24), as a function of doping at low temperatures. At half filling, a particle–hole symmetry ensures that $\langle s \rangle = 1$. At extreme fillings – either very few particles or very few holes – there are so few opportunities for exchange that the fermionic nature of the model becomes unimportant and the average sign is close to 1. What is surprising is that the determinants incorporate so much fermionic nature into the calculations of configuration weights (3.5) and (3.6) that the "low-density" value of $\langle s \rangle \approx 1$ holds even up to quarter filling. What is disconcerting is that the drop off of the sign as one moves away from half filling is extremely fast. It is unfortunate that so near half filling, where the possibility of superconductivity in a repulsive model exists, the simulations are the most difficult to perform.

5.2. *Attractive Hubbard model*

While no convincing explanation has yet been given for the pairing mechanism binding charge carriers in the novel high-temperature superconductors, the short coherence lengths in these materials have prompted many researchers to explore local, real-space pairing models. Neglecting the question as to what gives rise to the effective interaction, one can, nonetheless, assume a local, on-site attraction between fermions and study the resulting $U < 0$ Hubbard model.

In addition to its possible relevance to high T_c, the negative-U model in two dimensions is interesting in its own right, displaying curious phenomenology near half filling. Due to the high degree of symmetry (O(3)) at half filling, the model is expected to have no finite transition temperature. (At half filling, the model is isomorphic to the half-filled, repulsive Hubbard model, which belongs to the same universality class as the Heisenberg antiferromagnet and has long-range order only in the ground state.) Doping off of half filling in the negative-U model, however, reduces the symmetry of the ground state and gives rise to a finite, Kosterlitz–Thouless transition temperature (Scalettar et al. 1989).

In fig. 5, the number of singlet pairs in the pair condensate at zero temperature, normalized to the number of sites in the system, is plotted against the reciprocal of the linear size of the system for $U = -2$, -4, and -8 at quarter filling on lattices of up to 18×18 sites (Scalettar et al. 1989). Again, finite-size corrections from spin wave theory predict a straight-line extrapolation of the

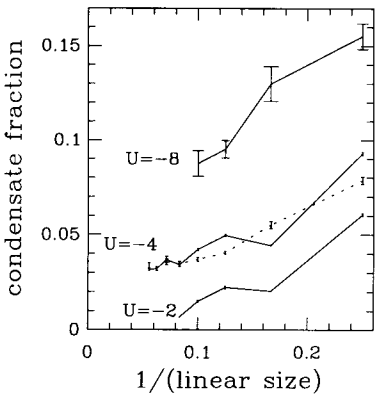

Fig. 5. Spin wave extrapolation of the ground-state fraction of singlet pairs in the pair condensate for the quarter-filled, two-dimensional, attractive Hubbard model (Scalettar et al. 1989). At weak coupling $U = -2$, finite-size modulations make extrapolations difficult. At a stronger coupling $U = -8$, error bars increase dramatically. The numerical simulation performs best at an intermediate coupling $U = -4$, for which analytical approaches are most limited. Data are shown for periodic boundary conditions (solid lines) and open boundaries (dotted line).

data to the infinite-system limit. Several features are apparent from the plot. First, at weak couplings $U = -2$ and -4, the data have a fairly irregular dependence on lattice size. The reason for this is easy to deduce from the $U = 0$ limit. In this limit, the fermions no longer interact and the Hamiltonian is easily diagonalized in momentum space. The single-particle density of states in this limit is very sparse for finite-size lattices. In order to achieve a particular filling for a given lattice size, the Fermi energy must sometimes lie on a single-particle energy and sometimes between levels. Thus, the effective density of states at the Fermi surface varies widely. This irregular dependence on lattice size, of course, diminishes with N, as the interlevel spacing gets smaller, and with $|U|$, as the single-particle levels get mixed.

Due to these irregularities, it is impossible to extrapolate the $U = -2$ data to the thermodynamic limit. By the time $U = -4$, however, it is clear that a macroscopic fraction of the fermions are in the pair condensate. Increasing the magnitude of the electron–electron coupling to $U = -8$, the pairs become more defined and the macroscopic fraction grows. The irregularities with lattice size have, presumably, been damped out by the electronic correlations. Now, however, this is difficult to tell since the error bars have increased dramatically. From the discrete Hubbard–Stratonovich transformation (2.8), we see that the coupling α of the electrons to the fluctuating field grows with $|U|$. Thus, for weak couplings $U = -2$ and $U = -4$, the statistical errors associated with sampling only a finite number of field configurations are small. As we see from fig. 5, however, these fluctuations increase dramatically as one goes to the strong-coupling limit.

In short, simulations are difficult to perform in weak coupling, due to the severe lattice-size dependence, and also in strong coupling, due to the increasing statistical fluctuations. The parameter regime in which numerical simulations perform best is one that is difficult to study with analytical techniques.

In fig. 6, the charge-density-wave structure factor

$$\mathrm{CDW}(q) = \frac{1}{N} \sum_{r_1 r_2} \mathrm{e}^{\mathrm{i}q\cdot(r_1 - r_2)} n_{r_1} n_{r_2} \tag{5.1}$$

is plotted as a function of wavenumber q for two-dimensional lattices of up to 18×18 sites for $U = -4$ (Scalettar et al. 1989). Other than for the smallest (4×4) lattice, results generally lie on top of one another for the different numbers of sites. While there is structure in the plot, the peaks do not diverge with lattice size. Thus, we characterize the system as a liquid, rather than as a solid. The peak in $\mathrm{CDW}(q)$ at half filling is, of course, at $q = (\pi, \pi)$ due to the perfect nesting of the Fermi surface. At quarter filling, the peak occurs on the lines $q = (q_x, \pi)$ and $q = (\pi, q_y)$, as predicted in the RPA.

Consider a charge excitation

$$|\psi_q\rangle = \rho_q|\psi_0\rangle / \langle\psi_0|\rho_{-q}\rho_q|\psi_0\rangle^{1/2}$$

of wavenumber q, where

$$\rho_q = \frac{1}{N} \sum_r \mathrm{e}^{\mathrm{i}q\cdot r} n_r$$

and $|\psi_0\rangle$ is the ground state. Then the energy of the excitation is

$$\omega(q) = \langle\psi_q|H|\psi_q\rangle - \langle\psi_0|H|\psi_0\rangle = \frac{2}{3}\frac{E_0}{N}\left(1 - \frac{1}{4}\varepsilon_q\right) \Big/ \mathrm{CDW}(q),$$

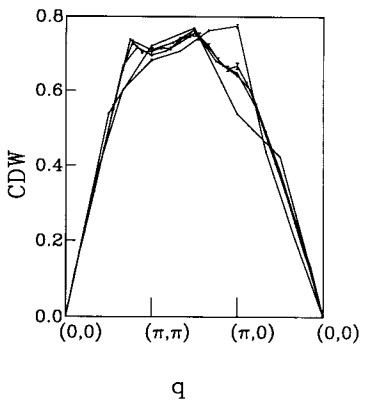

Fig. 6. The charge-density-wave structure factor (5.1) as a function of wave number q for quarter filled, two-dimensional, Hubbard lattices of up to 18×18 sites for $U = -4$ (Scalettar et al. 1989).

where (E_0/N) is the ground-state energy per site and $\varepsilon(q) = -2t(\cos q_x + \cos q_y)$. Thus, peaks in CDW($q$) can become dips in $\omega(q)$. The weak structure in fig. 6, however, is not strong enough to lead to a roton minimum in the excitation spectrum $\omega(q)$.

5.3. Copper-oxide clusters

Of course, more directly relevant to the high-temperature superconductors are multiband Hubbard models that have the structure of the CuO_2 planes. In the planes, each Cu site is surrounded by four nearest-neighbor O sites, while each O site has two nearest-neighbor Cu sites. While the debate over which orbitals of the Cu and O atoms are relevant – or even if the physics of the superconductivity is contained in the planes – continues, a great deal of attention remains focussed on the "Emery" model (Emery 1987) which considers only the $d_{x^2-y^2}$ Cu and p–σ bonding/antibonding O orbitals. The largest orbital overlap is between neighboring d- and p-orbitals and the largest Coulomb repulsion is the on-site Hubbard repulsion U_d. Thus, Scalettar et al. (1991) have studied

$$H = -t \sum_{\langle ij\rangle\sigma} (c^\dagger_{i\sigma} c_{j\sigma} + c^\dagger_{j\sigma} c_{i\sigma}) + \varepsilon \sum_j n_j + U_d \sum_i n_{i\uparrow} n_{i\downarrow}, \tag{5.2}$$

where the hopping sum is over nearest-neighbor Cu–O pairs, the site-energy sum is over O sites j, and the correlation sum is over Cu sites i. The creation and annihilation operators are actually for holes, with the reference "undoped" system having one hole per cell. We will choose $t = 1$, $\varepsilon = 1$, and $U_d = 4$, all energies nominally in eV.

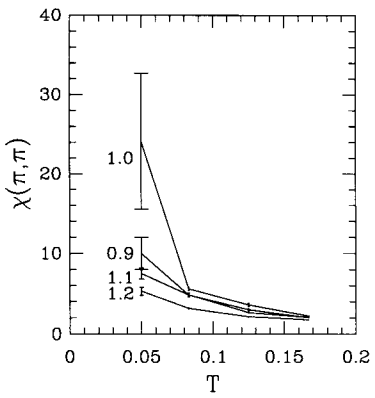

Fig. 7. The antiferromagnetic susceptibility $\chi(\pi, \pi)$ on the copper sublattice as a function of temperature for different lattice fillings on a two-dimensional CuO_2 cluster (Scalettar et al. 1991). Slight doping suppresses the antiferromagnetism.

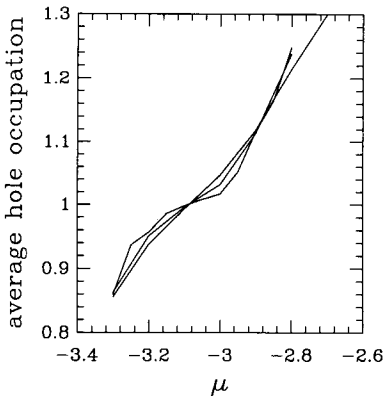

Fig. 8. The average number of holes per cell on the CuO_2 cluster as a function of the chemical potential μ for reciprocal temperatures $\beta = 12, 16, 20$ (Scalettar et al. 1991). As the temperature is lowered, a broad, flat plateau develops about the reference filling of 1 hole/cell.

In fig. 7, the antiferromagnetic susceptibility $\chi(\pi, \pi)$ on the copper sublattice is plotted as a function of temperature for different lattice fillings (Scalettar et al. 1991). The lattice is made up of 16 CuO_2 cells arranged in a 4×4 cluster. Thus, there are 48 sites in the system. This should be large enough to observe a bound-pair of holes since the coherence lengths in these systems are only several cell lengths. The effect of the site-energy difference is to place the holes predominantly on the Cu sites. Antiferromagnetic superexchange between Cu sites leads to the divergence in χ at low temperatures. Upon doping, however, itinerant holes wash out the staggered order. Figure 7 shows the low-temperature behavior of the antiferromagnetism as the doping is varied from 0.9, 1.0, 1.1, to 1.2 holes per unit cell.

In eq. (5.2), the filling of the lattice is controlled with a chemical potential term. In fig. 8, the average number of holes per cell is plotted as a function of the chemical potential μ for reciprocal temperatures $\beta = 12, 16, 20$ (Scalettar et al. 1991). As the temperature is lowered, a broad, flat plateau develops about the reference filling of 1 hole/cell. This is a clear signature of the stability of that filling.

6. Conclusion

In this chapter, we have briefly reviewed "determinantal" methods of simulating many-electron systems. In such methods, direct electronic interactions are replaced by couplings to auxiliary bosonic fields. Sums over the fermionic degrees of freedom are performed exactly for fixed configurations of the bosonic

fields. In contrast, sums over these fields are performed stochastically, using the determinantal weights to guide the importance sampling.

We have concentrated on providing a more heuristic exposition of the simulation formalism. Further, we have described recently developed methods for stabilizing the calculations at low temperatures.

Alternative stable algorithms do exist for specialized problems. The first approach is the world-line method for one-dimensional systems of interacting electrons (and quantum spin systems). This method has been reviewed by De Raedt and Lagendijk (1985). The second approach is the "impurity" algorithm of Hirsch and Fye (1986). In this approach, electrons interact on a small number N_i of impurity orbitals. The non-interacting part of the Hamiltonian is integrated out and the impurities may be regarded as being embedded in an infinite medium. Procedures exist for performing the updating (Hirsch and Fye 1986) and for restoring correlations between the impurities and the non-interacting degrees of freedom (Gubernatis 1987). Unfortunately, the updating process needs the elements of the Green's function between all imaginary times instead of just the equal-time elements. Hence, the computing time scales as $(N_i\beta)^3$. For the one and two-impurity Anderson Hamiltonians, the method is stable and efficient. For a lattice of impurity sites, however, this approach is very expensive. Other algorithms (White et al. 1988, Hirsch 1988) have reduced the prefactor for many-impurity problems, but methods that stabilize simulations by expanding Green's function matrices remain very costly.

In contrast, the overhead associated with explicitly maintaining essential small scales is modest. It is possible to stabilize a variety of simulation algorithms in this manner. Thus, one must choose a particular implementation. It appears that the Monte Carlo updating of discrete Hubbard–Stratonovich fields has the most desirable properties. Since the phase space of discrete fields is smaller than for continuous fields, the importance sampling is much more effective. The correlations (Buendia 1986, Lin and Hirsch 1986) and integration errors of Langevin and molecular-dynamics approaches make those methods less efficient and more difficult to implement well. Finally, Monte Carlo updating appears to be the most successful algorithm is crossing the "nodal barriers" which divide the bosonic-field phase space.

These efficient stabilization techniques have allowed relatively inexpensive explorations of the low-temperature phases of a variety of models of strongly correlated electrons. In most cases, it is no longer necessary to extrapolate to low temperatures. Rather, it is now possible simply to run at temperatures that are low enough to extract ground-state properties. Further, the overhead associated with the stabilization is only a small portion of the computational cost. Thus, low-temperature simulations (White et al. 1989a) of the two-dimensional Hubbard model can easily be realized at costs one order of magnitude below those previously possible (Tang and Hirsch 1988).

Nevertheless, important algorithmic difficulties remain in the numerical studies of fermions. Undoubtedly, the most important of these is the "minus-sign" problem. To date, this obstacle can really only be overcome by avoiding it. (Sorella et al. (1989) speculate that the sign problem can be ignored in the limit of zero temperature.) Hence, studies of the single-band Hubbard model are performed either for the $U < 0$ model or for the half-filled $U > 0$ model. Studies of the Anderson lattice are best limited to the symmetric case. These limitations can be extremely unfortunate. For example, the simplest explanation for high-temperature superconductivity would be the single-band Hubbard model. However, the validity of such a theory cannot be checked with numerical methods because it is just close to half filling, the only doping at which the repulsive Hubbard model could superconduct, that the sign problem is worst.

Large-scale simulations always benefit by faster algorithms. The most serious factor in limiting the speed of fermionic simulations is the dependence on lattice size. Determinantal methods scale as the cube of the number of lattice sites. The linear dependence of conjugate-gradient methods on N is highly attractive. Unfortunately, despite many attempts, such as "preconditioning", at controlling of the number of conjugate-gradient iterations needed at low temperatures, these techniques tend to lose their desirable scaling properties.

Acknowledgements

We would like to thank D.J. Scalapino, R.T. Scalettar, R.L. Sugar, and S.R. White for may useful discussions and productive collaborations and D.K. Campbell and R. Gupta for reading the manuscript. This work was supported by the U.S. Department of Energy.

Appendix I. Extended electronic interactions

In section 2, we considered only on-site interactions between electrons. Here, we briefly consider the case of interactions

$$\sum_{ij} V_{ij} n_i n_j \tag{I.1}$$

extended to other sites as well.

The standard approach is to rewrite each term $-\Delta\tau V_{ij} n_i n_j$ as a perfect square, which may be eliminated with an auxiliary field, and an on-site repulsion. Then, the on-site repulsions may be collected and eliminated with eq. (2.8). Unfortunately, the introduction of a new field for each term further expands the phase space that must be explored, making efficient simulations more difficult.

As the range of the interaction is increased, this approach simply becomes intractable.

The alternative is to couple auxiliary fields not only to the local electronic degrees of freedom but also to bosonic fields on other orbitals. For example, generalizing the Hubbard–Stratonovich transformation, we can write (Negele and Orlando 1988)

$$\exp(\tfrac{1}{2}\boldsymbol{n}^{\mathrm{T}}\mathbf{M}\boldsymbol{n}) \sim \int \mathrm{d}\boldsymbol{x} \exp(-\tfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}\mathbf{M}^{-1}\boldsymbol{x} + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{n})$$

$$\sim \int \mathrm{d}\boldsymbol{y} \exp(-\tfrac{1}{2}\boldsymbol{y}^{\mathrm{T}}\mathbf{M}\boldsymbol{y} + \boldsymbol{y}^{\mathrm{T}}\mathbf{M}\boldsymbol{n}), \tag{I.2}$$

where $\boldsymbol{n}^{\mathrm{T}} = (n_1, n_2, \ldots)$ is a vector of electron occupation numbers in orbitals 1, 2, ..., N. All of the eigenvalues of $\mathbf{M}$ must be positive in order that the integrals converge. In the first integral, the electronic charge $\boldsymbol{n}$ couples only to the local bosonic field $\boldsymbol{x}$, while the field x on each site couples to the field on each of the other sites through $\mathbf{M}^{-1}$. In the second integral, if $\mathbf{M}$ has a limited-range interaction, then the boson–electron coupling $\boldsymbol{y}^{\mathrm{T}}\mathbf{M}\boldsymbol{n}$ and boson–boson coupling $\boldsymbol{y}^{\mathrm{T}}\mathbf{M}\boldsymbol{y}$ are also of limited-range.

For a general repulsive interaction, however, the left-hand side of (I.2) is $\exp(-\Delta\tau\boldsymbol{n}^{\mathrm{T}}\mathbf{V}\boldsymbol{n})$ with $V_{ij} > 0$. To make the eigenvalues of $\frac{1}{2}\mathbf{M} = -\Delta\tau\mathbf{V}$ all positive in such cases, we must add an offset to the action. The contribution to the path integral from this generalized repulsion becomes

$$\exp\left(-\Delta\tau \sum_{ij} V_{ij} n_i n_j\right) = \exp\left(\Delta\tau V_0 \sum_i n_i^2 - \Delta\tau \sum_{ij} V_{ij} n_i n_j\right)$$

$$\times \exp\left(-\Delta\tau V_0 \sum_i n_i^2\right). \tag{I.3}$$

With the use of (I.2), the first factor in (I.3) may now be expressed in terms of a bosonic field whose components couple both to the local electronic charge density and to each other. The second factor may not be so treated as its "matrix" has only negative eigenvalues. This factor has the form of an on-site repulsion; as we have seen, it may be interpreted as resulting from a second intermediate field, whose components couple to the local electronic spin density but not to each other.

Again, we have engineered these transformations so that the auxiliary fields and the couplings are always real.

Appendix II. Bosonic world lines

In this appendix, we will continue the characterization of imaginary-time evolutions of bosonic field variables from section 2.

We first examine the relationship between the discrete [eq. (2.8)] and continuous [eq. (2.6)] transformations in the limit $\Delta\tau \to 0$, considering, as an example, the case $U > 0$. We average the external potential felt by the electrons over an imaginary time that is infinitesimally small and yet much larger than $\Delta\tau$. The coupling of the electrons to this locally time-averaged field is simply $\alpha x(n_\uparrow - n_\downarrow) \to \sqrt{\Delta\tau|U|}x(n_\uparrow - n_\downarrow)$, as $\Delta\tau \to 0$, where x is the averaged field. Phase space for the discrete field variables is uniform. Each such variable has zero mean and unit variance. A local time average x of many such fields, therefore, must have a Gaussian distribution $\exp(-x^2/2)/\sqrt{2\pi}$, from the central-limit theorem. Incorporating this weighting factor, we see that the elimination of high-frequency oscillations in the discrete field of (2.8) produces the continuous transformation (2.6). In the limit $\Delta\tau \to 0$, then, the discrete and continuous transformations are equivalent.

In section 2, we argued that the massless Hubbard–Stratonovich fields are essentially decoupled in time and that the time averages of such fields are necessarily of order $\sqrt{\Delta\tau}$, but that the electron–boson terms still enter the action to the same order in $\Delta\tau$ as the purely electronic terms, since the coupling to this weak average field is $1/\sqrt{\Delta\tau}$. We presented our case using fields that were generated by purely bosonic actions. Here, we justify our characterization of the massless fields rigorously.

We will examine discrete Hubbard–Stratonovich fields in the limit $\Delta\tau \to 0$. Further, we will work within the finite-temperature formalism. Our results remain valid within the zero-temperature formalism as well. The correlation of two distinct field variables is

$$\langle x(\tau)x(\tau')\rangle = \frac{\sum_{\{x\}} x(\tau)x(\tau')\,\mathrm{Tr}\,\mathscr{T}\exp(-\int_0^\beta S(\tau)\,\mathrm{d}\tau)}{\sum_{\{x\}}\mathrm{Tr}\,\mathscr{T}\exp(-\int_0^\beta S(\tau)\,\mathrm{d}\tau)}, \tag{II.1}$$

where the trace is over fermionic degrees of freedom and $S(\tau)$ is the action of the electrons evolving through the time-varying Hubbard–Stratonovich fields $\{x\}$. The field variable $x(\tau)$ enters the numerator of (II.1) as

$$\begin{aligned}\sum_{x(\tau)=\pm 1} x(\tau)\,\mathrm{e}^{-\Delta\tau S(\tau)} &= \sum_{x(\tau)=\pm 1} x(\tau)\,\mathrm{e}^{\alpha x(\tau)s(\tau)} \\ &= \tanh(\alpha s(\tau)) \sum_{x(\tau)=\pm 1} \mathrm{e}^{\alpha x(\tau)s(\tau)} \\ &= \tanh(\alpha s(\tau)) \sum_{x(\tau)=\pm 1} \mathrm{e}^{-\Delta\tau S(\tau)},\end{aligned}$$

where $s = n_\uparrow - n_\downarrow$ is the local electronic spin. Hence,

$$\langle x(\tau)x(\tau')\rangle = \langle \tanh(\alpha s(\tau))\tanh(\alpha s(\tau'))\rangle \to \Delta\tau|U|\langle s(\tau)s(\tau')\rangle = \mathcal{O}(\Delta\tau)$$

as $\Delta\tau \to 0$. Further, if $\tau = \tau'$, then $\langle x^2 \rangle = \langle 1 \rangle = 1$. The correlation of two field variables, then, drops to order $\Delta\tau$ for even an infinitesimal separation in imaginary time. As a corollary, while field-variable correlations are related to physical measurements, the fluctuations in these quantities make them impractical estimators in simulations.

We can now calculate the characteristic size of the mean field $N_x^{-1} \sum_\tau x(\tau)$, averaged over, again, some infinitesimal imaginary time $N_x \Delta\tau$ that is large compared to $\Delta\tau$. The mean-square field is

$$\left\langle \left(\frac{1}{N_x} \sum x(\tau) \right)^2 \right\rangle = \frac{1}{N_x^2} \sum_\tau \langle x(\tau)^2 \rangle + \frac{1}{N_x^2} \sum_{\tau \neq \tau'} \langle x(\tau) x(\tau') \rangle$$

$$= \frac{1}{N_x} + \frac{1}{N_x^2} \sum_{\tau \neq \tau'} \Delta\tau |U| \langle s(\tau) s(\tau') \rangle = \mathcal{O}(\Delta\tau).$$

Hence, the time-averaged field is of order $\sqrt{\Delta\tau}$, as in section 2.

Appendix III. Many-electron propagator

In this appendix, we examine the many-electron propagator (2.15). The many-electron propagator naturally takes the form of a determinant since the antisymmetry of electron exchange is automatically incorporated by the odd parity of the row or column exchange in determinantal evaluation.

The inner product of two N_σ-electron states is

$$\langle 0 | c_{i_1} c_{i_2} \cdots c_{i_{N_\sigma}} c^\dagger_{j_{N_\sigma}} \cdots c^\dagger_{j_2} c^\dagger_{j_1} | 0 \rangle = \det \begin{pmatrix} B_{i_1 j_1} & B_{i_1 j_2} & \cdots & B_{i_1 j_{N_\sigma}} \\ B_{i_2 j_1} & B_{i_2 j_2} & \cdots & B_{i_2 j_{N_\sigma}} \\ \vdots & \vdots & \ddots & \vdots \\ B_{i_{N_\sigma} j_1} & B_{i_{N_\sigma} j_2} & \cdots & B_{i_{N_\sigma} j_{N_\sigma}} \end{pmatrix}, \tag{III.1}$$

where the $B_{ij} = \langle 0 | c_i c^\dagger_j | 0 \rangle$ are the single-electron inner products. The proof proceeds easily via induction. Clearly, the result holds true for $N_\sigma = 1$. Assume eq. (III.1) holds for $N_\sigma - 1$. The anticommutation of electron operators gives

$$c_i c^\dagger_j = B_{ij} - c^\dagger_j c_i,$$

even if c_i and $c^\dagger_j$ are expressed in different bases. In eq. (III.1), we may exchange fermion operators to bring $c_{i_{N_\sigma}}$ all the way over to the right side until it multiplies the vacuum state $|0\rangle$ directly, producing a null result. The nonzero

remainders of the electron exchanges give

$$\begin{aligned}
&\langle 0|c_{i_1}c_{i_2}\cdots c_{i_{N_\sigma}}c^\dagger_{j_{N_\sigma}}\cdots c^\dagger_{j_2}c^\dagger_{j_1}|0\rangle \\
&\quad = B_{i_{N_\sigma}j_{N_\sigma}}\langle 0|c_{i_1}c_{i_2}\cdots c_{i_{N_\sigma-1}}c^\dagger_{j_{N_\sigma-1}}\cdots c^\dagger_{j_2}c^\dagger_{j_1}|0\rangle \\
&\qquad - B_{i_{N_\sigma}j_{N_\sigma-1}}\langle 0|c_{i_1}c_{i_2}\cdots c_{i_{N_\sigma-1}}c^\dagger_{j_{N_\sigma}}c^\dagger_{j_{N_\sigma-2}}\cdots c^\dagger_{j_1}|0\rangle \\
&\qquad - B_{i_n j_{N_\sigma-2}}\langle 0|c_{i_1}c_{i_2}\cdots c_{i_{N_\sigma-1}}c^\dagger_{j_{N_\sigma}}c^\dagger_{j_{N_\sigma-1}}c^\dagger_{j_{N_\sigma-3}}\cdots c^\dagger_{j_1}|0\rangle \\
&\qquad + \cdots . \qquad \text{(III.2)}
\end{aligned}$$

Expressing the $(N_\sigma - 1)$-electron inner products as determinants, we see that eq. (III.2) is simply the cofactor expansion of the $N_\sigma \times N_\sigma$ determinant in eq. (III.1).

Now consider the imaginary-time propagation of a N_σ-electron state. We may expand the imaginary-time propagator as

$$\mathcal{T}\exp\left(-\int_0^\beta S(\tau)\,d\tau\right) = \exp(-\Delta\tau S(\beta))\cdots\exp(-\Delta\tau S(2\Delta\tau)) \times \exp(-\Delta\tau S(\Delta\tau))$$

in terms of piecewise-constant contributions $S(l\Delta\tau)$ in the limit $\Delta\tau \to 0$. With the quartic terms in the electron operators reduced by the Hubbard–Stratonovich transformations, the contributions $S(l\Delta\tau)$ are at most quadratic in c and $c^\dagger$. Further, we will require that each piece $S(l\Delta\tau)$ is Hermitian and number-conserving. (While each $\exp(-\Delta\tau S(l\Delta\tau))$ must be Hermitian, the product of many such factors need not.) Then, the imaginary-time propagator becomes

$$\mathcal{T}\exp\left(-\int_0^\beta S(\tau)\,d\tau\right) = \text{constant} \times \exp(-\boldsymbol{c}^\dagger \mathbf{M}_{N_t}\boldsymbol{c})\cdots\exp(-\boldsymbol{c}^\dagger\mathbf{M}_2\boldsymbol{c}) \times \exp(-\boldsymbol{c}^\dagger\mathbf{M}_1\boldsymbol{c}),$$

where we have written the creation and annihilation operators

$$\boldsymbol{c}^\dagger = (c_1^\dagger \quad c_2^\dagger \quad \cdots \quad c_N^\dagger)$$

and

$$\boldsymbol{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix}$$

as vectors and the contributions to the action explicitly as number-conserving operators in matrix form.

Working in the basis in which $\mathbf{M}$ is diagonal, we can easily prove that

$$e^{-\boldsymbol{c}^\dagger\mathbf{M}\boldsymbol{c}}\boldsymbol{c}^\dagger = \boldsymbol{c}^\dagger e^{-\mathbf{M}}e^{-\boldsymbol{c}^\dagger\mathbf{M}\boldsymbol{c}}. \qquad \text{(III.3)}$$

Using (III.3) to propagate the N_σ-electron wavefunction

$$(e^{-c^\dagger \mathbf{M}_{N_t} c} \cdots e^{-c^\dagger \mathbf{M}_1 c}) c^\dagger \boldsymbol{v}_{j_{N_\sigma}} \cdots c^\dagger \boldsymbol{v}_1 |0\rangle$$
$$= (c^\dagger e^{-\mathbf{M}_{N_t}} \cdots e^{-\mathbf{M}_1} \boldsymbol{v}_{j_{N_\sigma}}) \cdots (c^\dagger e^{-\mathbf{M}_{N_t}} \cdots e^{-\mathbf{M}_1} \boldsymbol{v}_1)|0\rangle,$$

we may invoke eq. (III.1) to prove eq. (2.15).

Appendix IV. Finite-temperature field weights

Here, we show that eq. (3.6) is the weight of a configuration of field variables $\{x\}$ within the finite-temperature formalism. For a fixed configuration of bosonic field variables, the electrons evolve through a time-dependent field described by $\{x\}$. The electrons are decoupled from each other, however, so that the weight can be expressed in terms of the single-electron propagator $\mathbf{B}$, which depends on $\{x\}$. We would like to show that the trace

$$p[x] = \mathrm{Tr}\, \mathscr{T} \exp\left(- \int_0^\beta S(\tau)\, d\tau \right) \tag{IV.1}$$

over electronic degrees of freedom is $\det(\mathbf{1} + \mathbf{B}(\beta, 0))$, where $\mathbf{B}$ is given by eq. (2.12) and the action S depends on the configuration of the bosonic fields $\{x\}$. Here, we will consider electrons of only one particular spin σ since the generalization to electrons of many spins is trivial.

Here, we use the grand-canonical trace over all numbers N_σ of electrons. Since Pauli exclusion allows at most one electron per orbital, the number N_σ of electrons must fall in the range $0 \leqslant N_\sigma \leqslant N$, where N is the total number of orbitals. For each choice of N_σ, we must sum over the states

$$c^\dagger_{i_{N_\sigma}} \cdots c^\dagger_{i_2} c^\dagger_{i_1} |0\rangle,$$

where the sum over the $i_1, i_2, \ldots, i_{N_\sigma}$ must be over all distinct subsets of the orbitals $1, 2, \ldots, N$.

For a particular number N_σ of electrons and a particular subset $i_1, i_2, \ldots, i_{N_\sigma}$ of orbitals, the contribution

$$\langle 0 | c_{i_1} c_{i_2} \cdots c_{i_{N_\sigma}} \left(\mathscr{T} \exp\left(- \int_0^\beta S(\tau)\, d\tau \right) \right) c^\dagger_{i_{N_\sigma}} \cdots c^\dagger_{i_2} c^\dagger_{i_1} |0\rangle$$

$$= \det \begin{pmatrix} B_{i_1 i_1}(\beta, 0) & B_{i_1 i_2}(\beta, 0) & \cdots & B_{i_1 i_{N_\sigma}}(\beta, 0) \\ B_{i_2 i_1}(\beta, 0) & B_{i_2 i_2}(\beta, 0) & \cdots & B_{i_2 i_{N_\sigma}}(\beta, 0) \\ \vdots & \vdots & \ddots & \vdots \\ B_{i_{N_\sigma} i_1}(\beta, 0) & B_{i_{N_\sigma} i_2}(\beta, 0) & \cdots & B_{i_{N_\sigma} i_{N_\sigma}}(\beta, 0) \end{pmatrix}$$

to the trace is given by eq. (2.15).

Further, the determinant

$$\det(\mathbf{1} + \mathbf{B}(\beta, 0)) = \det \begin{pmatrix} 1 + B_{11} & B_{12} & \cdots & B_{1N} \\ B_{21} & 1 + B_{22} & \cdots & B_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ B_{N1} & B_{N2} & \cdots & 1 + B_{NN} \end{pmatrix} \tag{IV.2}$$

can be expanded as

$$\begin{aligned} &\det(1) \\ &+ \det(B_{11}) + \det(B_{22}) + \cdots + \det(B_{NN}) \\ &+ \det\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} + \det\begin{pmatrix} B_{11} & B_{13} \\ B_{31} & B_{33} \end{pmatrix} + \cdots \\ &\qquad + \det\begin{pmatrix} B_{N-1\,N-1} & B_{N-1\,N} \\ B_{N\,N-1} & B_{N\,N} \end{pmatrix} \\ &+ \cdots \\ &+ \det \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1N} \\ B_{21} & B_{22} & \cdots & B_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ B_{N1} & B_{N2} & \cdots & B_{NN} \end{pmatrix} \end{aligned}$$

in terms of the number of factors of B. We can easily see that this sum is precisely the grand-canonical trace that we are trying to evaluate. The first term is the contribution from the vacuum state. The next line is made up of contributions from all single-electron states. The following line is made up of contributions from all two-electron states. The last term, finally, is the contribution from the N-electron state, which packs each orbital with an electron.

In the special case that the fields are time independent, the trace (IV.1) can be evaluated in terms of stationary states. Once again, writing the electron creation and annihilation operators as vectors, the time-independent Hamiltonian

$$H = -\boldsymbol{c}^\dagger \mathbf{M} \boldsymbol{c} = -\boldsymbol{c}^\dagger (\mathbf{O}^\dagger \mathbf{D} \mathbf{O}) \boldsymbol{c} = -\boldsymbol{d}^\dagger \mathbf{D} \boldsymbol{d} = -\sum_i D_{ii} d_i^\dagger d_i$$

may be diagonalized with the unitary transformation $\mathbf{O}$ in terms of the stationary states $\boldsymbol{d} = \mathbf{O}\boldsymbol{c}$ of the system. The grand-canonical trace

$$\begin{aligned} \mathrm{Tr}\, e^{-\beta H} &= \mathrm{Tr} \exp\left(\beta \sum_i D_{ii} d_i^\dagger d_i \right) = \prod_i \sum_{d_i^\dagger d_i = 0,1} e^{\beta D_{ii} d_i^\dagger d_i} \\ &= \prod_i (1 + e^{\beta D_{ii}}) \end{aligned} \tag{IV.3}$$

is easily found using the diagonalized form. The final expression is just the partition function for free electrons. Rewriting it in basis-invariant form, we find

$$\mathrm{Tr}\, e^{-\beta H} = \det(\mathbf{1} + e^{\beta \mathbf{D}}) = \det(\mathbf{1} + e^{\beta \mathbf{M}}),$$

the time-independent form of (IV.2).

Appendix V. Determinants of different-size matrices

To produce eq. (3.13) from eq. (3.12), we used an identity

$$\det(\mathbf{1}_N + \mathbf{AB}) = \det(\mathbf{1}_M + \mathbf{BA}), \tag{V.1}$$

where $\mathbf{A}$ is an $N \times M$ rectangular matrix while $\mathbf{B}$ is an $M \times N$ rectangular matrix. To highlight the fact that (V.1) relates determinants of matrices of different sizes, we have written the $N \times N$ unit matrix as $\mathbf{1}_N$ and the $M \times M$ unit matrix as $\mathbf{1}_M$.

To prove (V.1), let us simply prove that $f_N(\lambda) = f_M(\lambda)$, where

$$f_N(\lambda) = \det(\mathbf{1}_N + \lambda \mathbf{AB})$$

and

$$f_M(\lambda) = \det(\mathbf{1}_M + \lambda \mathbf{BA}).$$

It is clear that f_N and f_M are polynomials in λ of order at most N and M, respectively, and that $f_N(0) = f_M(0) = 1$. Expanding the logarithm of $f_N(\lambda)$ as a power series about $\lambda = 0$ and using the relation $\det(\mathbf{M}) = \exp(\mathrm{Tr} \ln \mathbf{M})$, we find

$$\ln(f_N(\lambda)) = \mathrm{Tr} \ln(\mathbf{1}_N + \lambda \mathbf{AB}) = -\mathrm{Tr} \sum_{n=1}^{\infty} \frac{(-\lambda \mathbf{AB})^n}{n}.$$

Due to the cyclic nature of the trace, the order of $\mathbf{A}$ and $\mathbf{B}$ can be reversed. This gives the power-series expansion of $\ln(f_M(\lambda))$, proving the relationship between f_N and f_M.

References

Barma, M., and B.S. Shastry, 1978, Phys. Rev. B **18**, 3351.

Benettin, G., and L. Galgani, 1979, Lyapunov characteristic exponents and stochasticity, in: International Workshop on Intrinsic Stochasticity in Plasmas, June 17–23, Institut d'Etudes Scientifique de Cargese, Corse, France, eds G. Lavol and D. Grésillon (Editions de Physique, Orsay).

Binder, K., 1987, Monte Carlo Methods in Statistical Physics (Springer, Berlin).

Blankenbecler, R., D.J. Scalapino and R.L. Sugar, 1981a, Phys. Rev. D **24**, 2278.

Blankenbecler, R., D.J. Scalapino and R.L. Sugar, 1981b, Phys. Rev. B **24**, 4295.

Buendia, G.M., 1986, Phys. Rev. B **33**, 3519.

Creutz, M., and B. Freedman, 1981, Ann. Phys. **132**, 427.
DeRaedt, H., and A. Lagendijk, 1985, Phys. Rep. **127**, 233.
Emery, V.J., 1987, Phys. Rev. Lett. **58**, 2794.
Evenson, W.E., J.R. Schrieffer and S.Q. Wang, 1970, J. Appl. Phys. **41**, 1199.
Fetter, A.L., and J.D. Walecka, 1971, Quantum Theory of Many-Particle Systems (McGraw-Hill, New York).
Feynman, R.P., and A.R. Hibbs, 1965, Quantum Mechanics and Path Integrals (McGraw-Hill, New York).
Fye, R.M., 1986, Phys. Rev. B **33**, 6271.
Fye, R.M., and R.T. Scalettar, 1987, Phys. Rev. B **37**, 3833.
Gubernatis, J.E., 1987, in: Quantum Monte Carlo Methods, ed. M. Suzuki (Springer, Berlin) p. 216.
Hamann, D.R., 1970, Phys. Rev. B **2**, 1373.
Hassing, R.F., and D.M. Esterling, 1973, Phys. Rev. B **7**, 432.
Hirsch, J., R.L. Sugar, D.J. Scalapino and R. Blankenbecler, 1982, Phys. Rev. B **26**, 5033.
Hirsch, J.E., 1985, Phys. Rev. B **31**, 4403.
Hirsch, J.E., 1987, Phys. Rev. B **35**, 1851.
Hirsch, J.E., 1988, Phys. Rev. B **38**, 12023.
Hirsch, J.E., and R.M. Fye, 1986, Phys. Rev. Lett. **56**, 2521.
Hirsch, J.E., D.J. Scalapino, R.L. Sugar and R. Blankenbecler, 1981, Phys. Rev. Lett. **47**, 1628.
Hubbard, J., 1959, Phys. Rev. Lett. **3**, 77.
Lin, H.Q., and J.E. Hirsch, 1986, Phys. Rev. B **34**, 1964.
Loh Jr, E.Y., 1988, Multigrid Monte Carlo methods, in: Computer Simulation Studies in Condensed Matter Physics; Recent Developments, eds D.P. Landau and H.B. Schüttler, Springer Proc. Phys. Vol. 3 (Springer, Berlin).
Loh Jr, E.Y., J.E. Gubernatis, R.T. Scalettar, R.L. Sugar and S.R. White, 1989, Stable matrix-multiplication algorithms for the low-temperature numerical simulations of fermions, in: Interacting Electrons in Reduced Dimensions, eds D. Baeriswyl and D.K. Campbell (Plenum Press, New York).
MacKinnon, A., and B. Kramer, 1983, Z. Phys. B **53**, 1.
Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, 1953, J. Chem. Phys. **21**, 1087.
Negele, J.W., and H. Orlando, 1988, Quantum Many-Particle Systems (Addison-Wesley, Redwood City, CA).
Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1988, Numerical Recipes: The Art of Scientific Computing (Cambridge University Press, Cambridge).
Rice, J.R., 1983, Numerical Methods, Software and Analysis (McGraw-Hill, New York).
Scalapino, D.J., and R.L. Sugar, 1981, Phys. Rev. Lett. **46**, 519.
Scaletar, R.T., 1989, Physica C **162–164**, 313.
Scalettar, R.T., D.J. Scalapino, R.L. Sugar and D. Toussaint, 1987, Phys. Rev. B **36**, 8632.
Scalettar, R.T., E.Y. Loh Jr, J.E. Gubernatis, A. Moreo, S.R. White, D.J. Scalapino, R.L. Sugar and E. Dagotto, 1989, Phys. Rev. Lett. **62**, 1407.
Scalettar, R.T., S.R. White, D.J. Scalapino and R.L. Sugar, 1991, Phys. Rev. B, to be published.
Shimada, I., and T. Nagashima, 1979, Prog. Theor. Phys. **61**, 1605.
Sorella, S., E. Tosatti, S. Baroni, R. Car and M. Parrinello, 1988, Int. J. Mod. Phys. B **1**, 993.
Sorella, S., S. Baroni, R. Car and M. Parrinello, 1989, Europhys. Lett. **8**, 663.
Sugiyama, G., and S.E. Koonin, 1986, Ann. Phys. **168**(1), 1–26.
Suzuki, M., 1976, Commun. Math. Phys. **51**, 183.
Suzuki, M., 1985, Phys. Lett. A **113**, 199.
Tang, S., and E. Hirsch, 1988, Phys. Rev. B **37**, 9546.
Trotter, H.F., 1959, Proc. Am. Math. Soc. **10**, 545.
Verlet, L., 1968, Phys. Rev. **165**, 21.

White, S.R., R.L. Sugar and R.T. Scalettar, 1988, Phys. Rev. B **38**, 11665.
White, S.R., D.J. Scalapino, R.L. Sugar, E.Y. Loh Jr, J.E. Gubernatis and R.T. Scalettar, 1989a, Phys. Rev. B **40**, 506.
White, S.R., D.J. Scalapino, R.L. Sugar, N.E. Bickers and R.T. Scalettar, 1989b, Phys. Rev. B **39**, 839.
Wilkinson, J.H., 1965, The Algorithmic Eigenvalue Problem (Clarendon Press, Oxford) p. 607.