Transfer-matrix Monte Carlo estimates of critical points in the simple-cubic Ising, planar, and Heisenberg models

M. P. Nightingale

Department of Physics, University of Rhode Island, Kingston, Rhode Island 02881

H. W. J. Blöte

Faculty of Applied Physics, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

The principle and the efficiency of the Monte Carlo transfer-matrix algorithm are discussed. Enhancements of this algorithm are illustrated by applications to several phase transitions in lattice spin models. We demonstrate how the statistical noise can be reduced considerably by a similarity transformation of the transfer matrix using a variational estimate of its leading eigenvector, in analogy with a common practice in various quantum Monte Carlo techniques. Here we take the two-dimensional coupled XY-Ising model as an example. Furthermore, we calculate interface free energies of finite three-dimensional O(n) models, for the three cases $n=1$, 2, and 3. Application of finite-size scaling to the numerical results yields estimates of the critical points of these three models. The statistical precision of the estimates is satisfactory for the modest amount of computer time spent. [S0163-1829(96)01126-5]

I. INTRODUCTION

Many important problems in computational physics and chemistry can be reduced to the computation of dominant eigenvalues of matrices of high or infinite order. Among the numerous examples of such matrices are quantum-mechanical Hamiltonians and transfer matrices. The latter were introduced in statistical mechanics by Kramers and Wannier in 1941 to study the two-dimensional Ising model,[1] and ever since, important work on lattice models in classical statistical mechanic has been done with transfer matrices, producing both exact and numerical results.

The analogy of the time-evolution operator in quantum mechanics and the transfer matrix in statistical mechanics allows the two fields to share numerous techniques. Specifically, a transfer matrix $\mathbf{T}$ of a statistical-mechanical lattice system in d dimensions often can be interpreted as the evolution operator in discrete, imaginary time t of a quantum-mechanical analog, as is well known. That is, $\mathbf{T} \approx \exp(-t\mathcal{H})$, where $\mathcal{H}$ is the Hamiltonian of a system in $d-1$ dimensions, the quantum-mechanical analog of the statistical-mechanical system. From this point of view, the computation of the partition function and of the ground-state energy are essentially the same problems: finding the largest eigenvalue of $\mathbf{T}$ and of $\exp(-t\mathcal{H})$, respectively.

The transfer-matrix Monte Carlo method used in this paper employs an algorithm as simple as the diffusion Monte Carlo algorithm, which was developed to compute the dominant eigenvalue of the evolution operator $\exp(-t\mathcal{H})$. In contrast to the diffusion Monte Carlo algorithm, the transfer-matrix Monte Carlo method provides exact eigenvalues, subject only to statistical noise and as qualified below in Sec. II. More specifically, unlike the transfer-matrix Monte Carlo method, the diffusion Monte Carlo algorithm suffers from a systematic error, the time-step error, because of the necessity to employ an approximate, short-time evolution operator.

Similar errors are also found in the path-integral Monte Carlo method and, in general, in all approaches based on the Trotter formula.[2] An alternative, related approach, viz., Green function Monte Carlo, used to compute the dominant eigenvalue of $(\mathcal{H}-E)^{-1}$, where E is close to the ground-state energy, does not suffer from a time-step error, and, from that point of view, the Green function Monte Carlo algorithm is more elegant than the diffusion Monte Carlo algorithm. However, the Green function Monte Carlo algorithm is considerably more complicated, and enhancement of that algorithm by the variance reduction techniques discussed below has its limitations.

From an orthodox complexity theory point of view, *exact* numerical transfer-matrix computations for lattices in more than one dimension are intractable, since the order of transfer matrices grows exponentially with the number of lattice sites in a transfer slice. Standard Monte Carlo methods in statistical mechanics, on the other hand, statistically sample the Boltzmann distribution, typically employing some variant of the Metropolis algorithm. One can argue that Monte Carlo methods are of polynomial complexity in system size, at least for certain important physical observables. This raises the question of the ultimate utility of the transfer matrix for computational purposes.

In many cases, one is interested in the behavior of systems in the thermodynamic limit. For critical systems in particular, one has to rely on finite-size scaling and extrapolation methods to extract the relevant information from the computations. The transfer-matrix method has advantages in both respects. In the first place, one can compute the spectrum of the transfer-matrix method virtually to machine precision, which permits extrapolation without serious loss of numerical accuracy. Secondly, a large body of numerical evidence suggests that the transfer-matrix spectrum has weaker corrections to scaling than quantities commonly computed by the standard Monte Carlo method. Clearly, also the transfer-matrix Monte Carlo method takes advantage of

the weakness of the corrections to scaling. Unfortunately, statistical noise is introduced, but this can be reduced substantially by the use of optimized trial eigenvectors, by virtue of which the Monte Carlo process is in effect only used to compute *corrections* to an already sophisticated approximation.

If one could neglect the correlations introduced by the reweighting step of the transfer-matrix Monte Carlo algorithm [see the split and join steps (2a) and (2b) in the algorithm given in Sec. II] and if one could ignore the resulting loss of efficiency of the transfer-matrix Monte Carlo algorithm, this method would be a solution to the exponential growth problem mentioned above.[3] In addition, the transfer-matrix Monte Carlo method would be completely free of critical slowing down, since the correlation *time* of the algorithm is equal to the correlation *length* of the slices used in the definition of the transfer matrix. Again, the use of optimized trial eigenvectors can serve to reduce the detrimental effect of the multiplicative reweighting.

Another feature of the Monte Carlo transfer matrix, which can contribute to a reduction of the correlation time of the stochastic process, is that moves are effectively made at surface sites. This makes it much easier to overcome the barriers some systems present to standard Monte Carlo algorithms. An example of such a system is the XY-Ising model discussed in Ref. 4.

The layout of this paper is as follows. In Sec. II we review the basic Monte Carlo algorithm to determine transfer-matrix eigenvalues by means of a statistical implementation of the power method. Apart from relatively minor details, the algorithm given in Sec. II is the same as the one discussed in Refs. 5–7. Section III describes the similarity transformation of the transfer matrix, which leads to a pronounced decrease of the statistical errors of the Monte Carlo process. Section III in particular describes in detail the construction of a variational approximation of the eigenstate associated with the largest eigenvalue. This approximate eigenstate yields the similarity transformation used to reduce the statistical errors of the algorithm. Details of the speedup of the algorithm are presented at the end of Sec. III, a coupled XY-Ising model in two dimensions. Finally, Sec. IV contains applications of the transfer-matrix Monte Carlo method to three-dimensional $O(n)$ models for $n = 1$, 2, and 3. Preliminary discussions of the the work discussed in Secs. III and IV were published elsewhere.[6,4]

II. MONTE CARLO IMPLEMENTATION OF THE POWER METHOD

Consider an operator $\mathbf{T}$ of which we want to compute the dominant eigenvalue. Let $\mathbf{T}$ be represented by matrix elements $\langle R|\mathbf{T}|S\rangle = T_{RS}$, where $|R\rangle$ and $|S\rangle$ are basis states of the physical system under consideration. These states will be treated here as discrete. For Monte Carlo calculations, the distinction between continuous and discrete states is a minor technicality; in the discussion below, generalization to the continuous case follows immediately by replacing the appropriate sums by integrals and replacing Kronecker by Dirac δ functions.

Perhaps the simplest way to calculate the dominant eigenvalue of a matrix or integral kernel is the power method.

That is, choose an arbitrary initial state $|u^{(0)}\rangle$ and compute iteratively:

$$|u^{(t+1)}\rangle = \frac{1}{c_{t+1}}\mathbf{T}|u^{(t)}\rangle, \tag{1}$$

where c_{t+1} is a constant chosen so that $|u^{(t+1)}\rangle$ is normalized or in some other convenient standard form. For $t\rightarrow\infty$, the constants c_t approximate the dominant eigenvalue λ_0 of $\mathbf{T}$ and the vectors $|u^{(t)}\rangle$ converge to the corresponding eigenvector.

To implement Eq. (1) by a Monte Carlo method, $|u^{(t)}\rangle$ is represented by a sequence of N_t walkers. Each of these walkers is a pair (R_α, w_α), $\alpha = 1, \ldots, N_t$. The variable R_α of a walker represents a possible configuration of the system described by $\mathbf{T}$, and w_α represents its statistical weight. The latter quantity is subject to the condition $w_l < w_\alpha < w_u$, where w_l and w_u are bounds introduced so as to keep all weights w_α of the same order of magnitude, which improves the efficiency of the algorithm. This sequence of walkers represents a (sparse) vector with components

$$\underline{u}_R^{(t)} = \sum_{\alpha=1}^{N_t} w_\alpha \delta_{R,R_\alpha}, \tag{2}$$

where δ is the usual Kronecker δ function. The underbar is used to indicate that the $\underline{u}_R^{(t)}$ represent a stochastic vector $|\underline{u}^{(t)}\rangle$. A stochastic process will be defined presently with transition probabilities such that $c_{t+1}|\underline{u}^{(t+1)}\rangle$ has a conditional expectation value equal to $\mathbf{T}|\underline{u}^{(t)}\rangle$ for any given sequence of walkers representing $|\underline{u}^{(t)}\rangle$. In practice, one has to average over the stationary state of a stochastic process in which the constants c_t are determined using *a posteriori* statistics, so that c_{t+1} and $|\underline{u}^{(t+1)}\rangle$ are correlated. As a consequence, there is no guarantee that the stationary-state expectation value of $|\underline{u}^{(t)}\rangle$ is *precisely* an eigenstate of $\mathbf{T}$, at least not for finite N_t. The same mathematical problem occurs if one takes the time average of Eq. (1) in the presence of noise correlated to the c_t. The resulting bias[8,9] has also been discussed in the context of the diffusion Monte Carlo method.[7]

To define the stochastic process, Eq. (1) is rewritten as

$$u_R^{(t+1)} = \frac{1}{c_{t+1}}\sum_S P_{RS}D_S u_S^{(t)}, \tag{3}$$

where

$$D_S = \sum_R T_{RS} \qquad \text{and} \qquad P_{RS} = T_{RS}/D_S. \tag{4}$$

Equation (3) describes a process represented by a Monte Carlo run which, in addition to a few initial equilibration sweeps, consists of a time series of a little over M_0 sweeps over all walkers at times labeled by $t = \ldots, 0, 1, \ldots, M_0$. The sweep at time t consists of two steps designed to perform stochastically the matrix multiplications in Eq. (3). Following Nightingale and Blöte,[9] the process is defined by the following steps, which transform the generation of walkers at time t into the the generation at time $t + 1$. Variables pertaining to times t and $t + 1$ will be denoted, respectively, by unprimed and primed symbols.

(1) Update the old walker (S_α, w_α) to yield a temporary walker (S'_α, w'_α) according to the transition probability $P_{S'_\alpha S_\alpha}$, where $w'_\alpha = D_{S_\alpha} w_\alpha / c'$, for $\alpha = 1, \ldots, N_t$. The next step can change the number of walkers. To maintain their number close to a target number, say N_0, choose $c' = \hat{\lambda}_0 (N_t/N_0)^{1/s}$, where $\hat{\lambda}_0$ is a running estimate of the eigenvalue λ_0 to be calculated, where $s \geq 1$ (see below).

(2) From the temporary walkers construct the new generation of walkers as follows.

(a) Split each walker (S', w') for which $w' > b_u$ into two walkers $(S', \frac{1}{2}w')$. The choice $b_u = 2$ is a reasonable one.

(b) Join pairs (R'_α, w'_α) and (R'_β, w'_β) with $w'_\alpha < b_l$ and $w'_\beta < b_l$ to produce a single walker $(R'_\gamma, w'_\alpha + w'_\beta)$, where $R'_\gamma = R'_\alpha$ or $R'_\gamma = S'_\beta$ with relative probabilities w'_α and w'_β. We chose $b_l = 1/2$.

(c) Any temporary walker left single in step (2b), or for which $b_l < w'_\alpha < b_u$, becomes a permanent member of the new generation of walkers.

The algorithm described above was constructed so that for any given realization of $|\underline{u}^{(t)}\rangle$, the expectation value of $c_{t+1} |\underline{u}^{(t+1)}\rangle$, in accordance with Eq. (1), satisfies

$$E(c_{t+1}|\underline{u}^{(t+1)}\rangle) = \mathbf{T}|\underline{u}^{(t)}\rangle, \qquad (5)$$

where $E(\cdot)$ denotes the conditional average over the transitions defined by the above stochastic process. More generally by p-fold iteration one finds[5]

$$E\left(\left[\prod_{b=1}^{p} c_{t+b}\right]|\underline{u}^{(t+p)}\rangle\right) = \mathbf{T}^p|\underline{u}^{(t)}\rangle. \qquad (6)$$

The stationary state average of $|u^{(t)}\rangle$ is close to the dominant eigenvector of $\mathbf{T}$, but, as mentioned above, it has a systematic bias when the number N_t of walkers is finite. For increasing p, components of nondominant eigenvectors can be projected out and thus the bias is reduced, in principle. Unfortunately, the variance of the corresponding estimators increases as their bias decreases. The reader is referred to Refs. 8, 5, 6, and 3 for a more detailed discussion of this problem. Suffice it to mention here, first, that s is the expected number of time steps it takes to restore the number of walkers to its target value N_0 and, second, that strong population control ($s = 1$) tends to introduce a stronger bias than weaker control ($s > 1$).[10]

With Eq. (6) one constructs an estimator[5] of the dominant eigenvector $|u^{(\infty)}\rangle$ of the matrix $\mathbf{T}$:

$$|\hat{u}^{(p)}\rangle = \frac{1}{M_0}\sum_{t=1}^{M_0}\left(\prod_{b=0}^{p-1} c_{t-b}\right)|\underline{u}^{(t)}\rangle. \qquad (7)$$

More practically, suppose that $\langle\psi_T|$ is an approximate leading eigenbra of $\mathbf{T}$, and that $\mathbf{O}$ is an arbitrary operator. The mixed expectation value of $\mathbf{O}$ can be approximated as

$$\frac{\langle\psi_T|\mathbf{O}|u^{(\infty)}\rangle}{\langle\psi_T|u^{(\infty)}\rangle} \approx \frac{\langle\psi_T|\mathbf{O}|\hat{u}^{(p)}\rangle}{\langle\psi_T|\hat{u}^{(p)}\rangle}. \qquad (8)$$

An important special case is obtained by choosing in this expression $\mathbf{O} = \mathbf{T}$ and $\langle\psi_T|R\rangle = 1$ for all R. The latter corre-
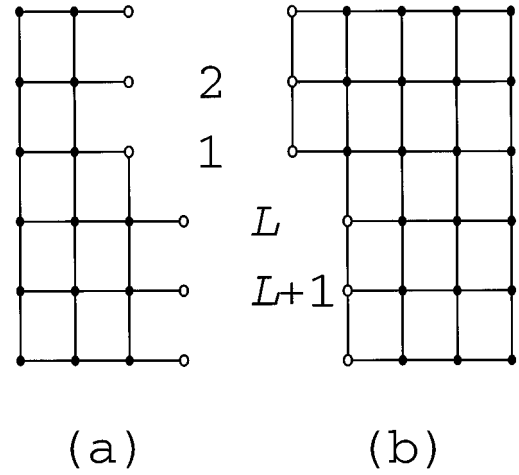

FIG. 1. Illustration of left and right eigenvectors of the transfer matrix.

sponds to the infinite-temperature approximation for the trial state, and in that case, Eq. (8) reduces to an estimator for the dominant eigenvalue of $\mathbf{T}$:

$$\lambda_0 \approx \frac{\sum_{t=1}^{M_0}(\Pi_{b=0}^{p}c_{t-b})W^{(t)}}{\sum_{t=1}^{M_0}(\Pi_{b=0}^{p-1}c_{t-b})W^{(t-1)}}, \qquad (9)$$

where

$$W_t = \langle\psi_T|\underline{u}^{(t)}\rangle = \sum_{\alpha=1}^{n_t} w_\alpha^{(t)}. \qquad (10)$$

For the above special choice of the trial bra $\langle\psi_T|$, Eq. (8) becomes the expression for the surface expectation value of $\mathbf{O}$ in the geometry shown on the right in Fig. 1. Although we have used the transfer-matrix algorithm only for the computation of the dominant eigenvalue of the transfer matrix for the applications discussed in this paper, it should be mentioned for completeness that one can also compute bulk expectation values, at least asymptotically, as follows.

One can represent the Kramers-Wannier transfer matrix by the graph shown in Fig. 2(a). This matrix transfers from an old slice to a new one, with slices represented, respectively, by small solid and large open circles. The process adds only one new site: the open circle labeled 1. One site, the small solid circle labeled L, is about to disappear into the bulk. Coincidences of both types of circles represent Kronecker δ functions in the transfer matrix [see Eq. (13)]. The solid lines stand for interactions added in one transfer operation. One can define a transfer matrix with extended slices consisting of m of the original, minimal slices. The dominant eigenvector of this extended transfer matrix is simply the original eigenvector multiplied by the Boltzmann weight associated with the portion of the lattice containing variables that have not yet been summed over. Equation (8), used with any operator in which occur only variables of slice m, becomes a bulk expectation value for $m \to \infty$. The implementation of this concept is called *forward walking* in the context of the quantum Monte Carlo method,[11,12] and this only requires extending the walkers so that their states correspond to the extended slices introduced above. This increases the
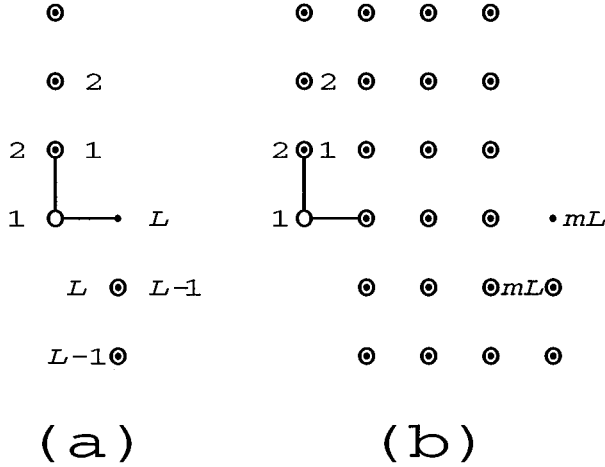
FIG. 2. Illustration of the calculation of correlation functions involving spins in the bulk below the surface layer. Site labels before the addition of the new spin (open circle) appear to the right, and the new labels to the left of a lattice point.

memory requirements and the cost of splitting a walker, but otherwise the efficiency of the algorithm is not affected.

III. VARIANCE REDUCTION (IMPORTANCE SAMPLING) AND TRIAL VECTORS

In principle, if $\langle \psi_T |$ equals an exact eigenbra of the operator **O** in Eq. (8), the right-hand side of the expression is a zero-variance estimator. In general, no exact eigenvectors are known, but even an approximation may yield a substantial reduction of statistical noise. A more efficient well-known[13] way to exploit an approximate left eigenbra $\langle \psi_T |$ to reduce variance works by application of the method described above to a similarity transform of the original operator **T**. This transformation is defined by

$$\widetilde{\mathbf{T}} = \mathbf{I} \mathbf{T} \mathbf{I}^{-1}, \qquad (11)$$

where **I** is diagonal in the configuration presentation, and is defined as

$$\mathbf{I} = \sum_R |R\rangle \langle \psi_T | R \rangle \langle R|. \qquad (12)$$

Ideally, $\langle \psi_T |$ would equal the exact dominant eigenbra of **T**. In that case, the stochastic process defined as above, but with **T** replaced by $\widetilde{\mathbf{T}}$, would become optimally efficient and in fact would lack critical slowing down. For such an ideal processes $\widetilde{\mathbf{D}}$, defined as in Eq. (4) as a function of $\widetilde{\mathbf{T}}$, would be a constant times the unit matrix. The walker weights would no longer fluctuate so that birth and death processes would no longer occur. The walkers would evolve into a statistically independent ensemble. The estimator given in Eq. (8), appropriately transformed, would have zero variance. The transformed bra $\langle \widetilde{\psi}_T | = \langle \psi_T | \mathbf{I}^{-1}$ would have all elements equal to unity in the configuration representation. In other words, $\widetilde{\mathbf{T}}$ would be represented by a stochastic ma-

trix, which would eliminate reweighting of walkers and the concomitant split and join steps in the algorithm.

In the absence of *exact* eigenbras, approximations may be obtained by variational methods. The variational expression for the leading eigenbra $\langle \psi_T |$ can conveniently be cast in the form of an effective surface Hamiltonian with pair interactions between nearest neighbors, next-nearest neighbors, and so on. These interactions are treated as variational parameters and can be determined from analysis of the walker population.[14]

Since generalization to higher dimensions and models with different microscopic variables is straightforward, it will suffice to consider the Kramers-Wannier transfer matrix for the two-dimensional Ising model to explain the construction of trial vectors used in the applications discussed in Sec. IV.

For a simple-quadratic lattice of M sites, wrapped on a cylinder with a circumference of L spins and helical boundary conditions, the transfer matrix for the Ising model is

$$T_{S,R} = e^{K(s_1 r_1 + s_1 r_L)} \prod_{i=1}^{L-1} \delta_{s_i, r_{i+1}}, \qquad (13)$$

with $S = (s_1, s_2, \ldots, s_L)$ and $R = (r_1, r_2, \ldots, r_L)$, where the $s_i = \pm 1$ and $r_i \pm 1$. The conditional partition function of the lattice of M sites, subject to the restriction that the spins on the left-hand edge be in state R, as illustrated in Fig. 1, is denoted $Z_M(R)$. One has

$$Z_{M+1}(S) = \sum_R T_{S,R} Z_M(R). \qquad (14)$$

Obviously, for $M \to \infty$ the restricted sums $Z_M(R)$ are proportional to the components $u_R^{(\infty)}$ of the dominant right eigenvector of the transfer matrix. The eigenvector is represented by the graph on the right in Fig. 1. Solid circles indicate spins that have been summed over, while the fixed surface spins are represented by the open circles; each bond represents a factor $\exp(Ks_i s_j)$. The left eigenvector, which is the one that has to be approximated by an optimized trial vector, is represented by the graph on the left. In passing, we mention the following relation between left and right eigenvectors, which follows by inspection of the graphs:

$$\langle u^{(\infty)} | S \rangle = \prod_{i=1}^{L-1} e^{Ks_i s_{i+1}} \langle U(S) | u^{(\infty)} \rangle, \qquad (15)$$

where U is the reflection operator, $U(S) = (s_L, s_{L-1}, \ldots, s_1)$.

A similarity transformation of the transfer matrix **T** can be introduced by dividing up the interaction energies between the columns differently. That is, h is introduced by writing

$$T_{S,R} \equiv e^{h(S,R)}. \qquad (16)$$

A transformation $h \to \widetilde{h}$ is defined by

$$\widetilde{h}(S,R) = g(S) + h(S,R) - g(R), \qquad (17)$$

$$\widetilde{T}_{S,R} = \hat{\psi}_T(S) T_{S,R} / \hat{\psi}_T(R), \qquad (18)$$

$$\hat{\psi}_T(S) = e^{g(S)}. \qquad (19)$$

For purposes of variance reduction, versatile trial vectors that capture some of the essential physics without seriously slowing down computations can be chosen of the form

$$\hat{\psi}_T(S) = \exp\left(\sum_{i,j}^{*} K_{ij} s_i s_j \right), \qquad (20)$$

a form reminiscent of the Jastrow functions used for quantum many-body systems. The asterisk in the sum over pairs indicates that the K_{ij} are truncated for distances greater than a couple of lattice spacings.

The couplings K_{ij} in Eq. (20) are variational parameters. They can be determined efficiently with the Monte Carlo scheme introduced by Umrigar, Wilson, and Wilkins,[14] i.e., by minimization of the variance of $\widetilde{D}(S)$, where the variance is approximated by a weighted sum over the states of the walkers of one generation, during the initial stage of the Monte Carlo run. This procedure is efficient and stable as long as the K_{ij} are truncated with care, in which case it is perfectly feasible to use as many as 50–100 different parameters.

The magnitude of the K_{ij} is expected to increase with the strength of the correlations between surface spins. Since all correlations between surface spins for the left eigenvector have to be propagated through the lattice on the left, as illustrated in Fig. 1, one expects that for high temperatures, i.e., small K,

$$K_{ij} \propto K^{d_{ij}}, \qquad (21)$$

where d_{ij} is the length of the shortest path along edges connected by bonds between sites i and j. By inspection of the graph in Fig. 1, we therefore expect the following partial ordering in decreasing strength of interaction and increasing d_{ij}:

$$d_{1L} = 2,$$
$$d_{12} = d_{23} = \cdots = d_{L-1,L} = d_{1,L-1} = 3,$$
$$d_{13} = d_{24} = \cdots = d_{L-2,L} = d_{1,L-2} = 4,$$
$$d_{14} = d_{25} = \cdots = d_{L-3,L} = d_{1,L-3} = d_{2,L} = 5,$$
$$\vdots . \qquad (22)$$

It is important to note that if $K_{ij} = K_{i+1,j+1}$ the corresponding factors cancel in the transformed transfer matrix $\widetilde{\mathbf{T}}$ for $2 \leq i \leq L-2$, since $s_i = t_{i+1}$ for nonvanishing transfer-matrix elements. For reasons of efficiency it is therefore advantageous to have this equality satisfied as often as possible. Unfortunately, helical boundary conditions introduce a step which destroys translation symmetry on the surface and renders the partial ordering in Eq. (22) insufficient. For example, sites 1 and 2 are more strongly correlated than sites 2 and 3, and correlations keep decreasing through pair $(L-1,L)$. Consequently, $K_{12} > K_{23} > \cdots > K_{L-1,L}$.

In practice, the *differences* between the K_{ij} with $d_{ij} = 3$ are frequently greater than the higher-order K_{ij}. Then, it is necessary to treat K_{12} and K_{23} as different parameters of the trial vector. An efficient compromise is to treat K_{ij} in which site 1 or L participate as different. The same applies to all K_{ij} for which the shortest path between i and j straddles the step on the surface. To summarize, we distinguish different

TABLE I. Estimated eigenvalue and standard deviations for the XY-Ising model. These data apply to the point $(A = 1.005, C = -0.2285)$ [cf. Eq. (23)] on the line where Ising and XY transitions coincide. Results are shown for various values of d_m, the path length of the cutoff in Eq. (24). The results are for a strip of width $L = 20$ and were obtained with a target number of walkers $N_0 = 10\,000$ and $M_0 = 1250L$ generations of which an initial 10% were discarded. The last column shows the computer time in arbitrary units needed per time step of one walker.

λ_0	σ	d_m	μs
34.17406	0.0071	0	15
34.20875	0.0052	2	15
34.21658	0.0015	3	17
34.21418	0.00083	4	19
34.21384	0.00052	5	21
34.21366	0.00049	6	23
34.21379	0.00041	7	26

types of pairs of sites (i,j) both on the basis of the distance d_{ij} and to some extent on the location of the pair, enforcing as much translation invariance as possible.

Clearly, the above depends only on lattice geometry and not on the Ising nature of the variables. In general, the only requirement is to generate lists of lattice sites separated by various distances d_{ij}. Once the incidence matrix has been defined for the pertinent lattice, it is possible to proceed in a fashion independent of these geometrical details.

To illustrate the efficiency and flexibility of this technique for constructing trial vectors, we use the XY-Ising model. It consists of coupled Ising and planar rotator degrees of freedom on a simple quadratic lattice. On each lattice site there are two variables $s_i = \pm 1$ and $\mathbf{n}_i$, a two-component unit vector. The Hamiltonian divided by $-k_B T$ is given by

$$H = \sum_{(i,j)} (A\mathbf{n}_i \cdot \mathbf{n}_j + B\mathbf{n}_i \cdot \mathbf{n}_j s_i s_j + C s_i s_j). \qquad (23)$$

We consider the special case $A = B$ and only from the point of view of the performance of transfer-matrix Monte Carlo algorithm. For a discussion of the physics of this model the reader is referred to Ref. 4. The trial vectors discussed above for the Ising model have an immediate generalization

$$\psi_T = \exp\left(\sum_{i,j}^{*} (A_{i,j}\mathbf{n}_i \cdot \mathbf{n}_j + B_{i,j}\mathbf{n}_i \cdot \mathbf{n}_j s_i s_j + C_{i,j} s_i s_j) \right). \qquad (24)$$

The truncation scheme introduced above for the Ising model is purely geometrical, and therefore carries over without changes to the XY-Ising model. It should, however, be noted that there are models and choices of transfer matrices to which the above scheme is not applicable. Reference 15 contains a discussion and an example of such a case.

Table I shows the estimates of the dominant eigenvalue of XY-Ising model for a trial vector truncated at different values of d_{ij}. As can be seen by comparing the first and last lines of the table, the variance in the estimate of the eigenvalue is reduced by a factor 300 for a fixed number of Monte
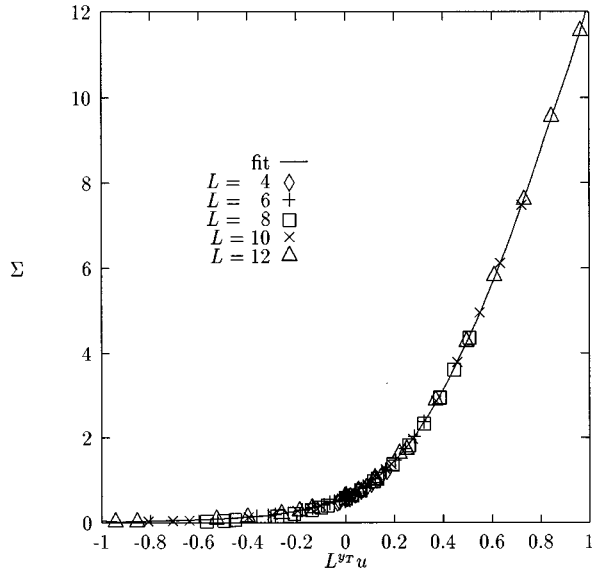
FIG. 3. Finite-size scaling plot for the interface free energy of the three-dimensional Ising model.

Carlo steps. Taking into account that the computer time per step doubles, this constitutes a speedup by a factor of 150.

IV. APPLICATIONS

As an illustration of the transfer-matrix technique we apply the method to three-dimensional $O(n)$ models for $n=1$, 2, and 3, i.e., the Ising, planar, and Heisenberg models. In particular the significance of the results of the planar and Heisenberg models goes beyond mere illustrations. These results are sufficiently accurate to be of some relevance for the location of the critical points.

The $O(n)$ spins are located on the simple cubic lattice. The transfer matrix for an $L \times L \times \infty$ system, with helical boundary conditions and layers of $N=L^2$ sites each, is a straightforward generalization of Eq. (13) and reads

$$T_{\mathbf{S},\mathbf{R}} = \prod_{i=1}^{N-1} \delta_{\mathbf{s}_i, \mathbf{r}_{i+1}} \exp[K\mathbf{s}_1 \cdot (\mathbf{r}_1 + \mathbf{r}_L + \mathbf{r}_N)], \quad (25)$$

where the $\mathbf{s}_i$ and $\mathbf{r}_i$ are n-component unit vectors, $\mathbf{S}=(\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_N)$ and $\mathbf{R}=(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)$.

As discussed above, the transfer-matrix Monte Carlo algorithm is designed to compute the dominant eigenvalue λ_0 of the transfer matrix. The reduced free energy per site is $f=-\ln\lambda_0$. From the free energy one can calculate the surface tension as the difference in free energy of two systems, one with ferromagnetic and the other with antiferromagnetic interactions, if the dimensions are chosen so as to force an interface in the antiferromagnetic system. For $L \times L \times \infty$ systems with helical boundary conditions, to which the present calculations are restricted, this means that L has to be even.

Renormalization group theory predicts that the values of Δ, the reduced interface free energy per lattice site, as a function of coupling K and system sizes L collapse onto a single curve, at least close to the critical point K_c and for sufficiently large systems. In terms of the nonlinear thermal scaling field
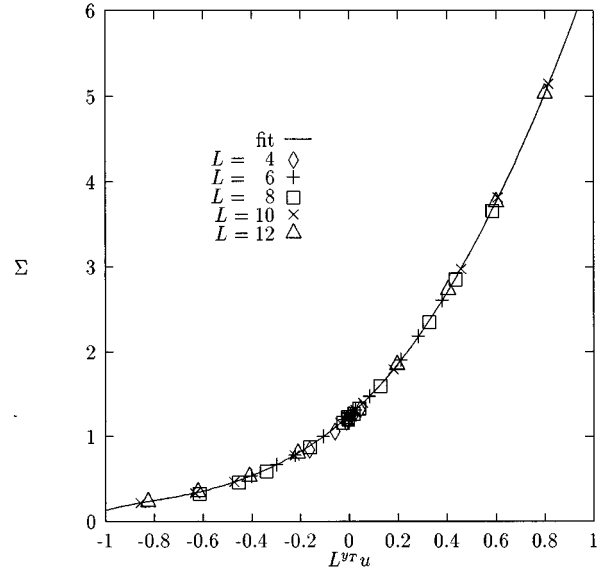

FIG. 4. Finite-size scaling plot for the interface free energy of the three-dimensional planar model.

$$u(K)=K-K_c+a(K-K_c)^2+\cdots, \quad (26)$$

this curve $\Sigma(x)$ is determined by

$$\Delta(u,L)=L^{1-d}\Sigma(L^{y_T}u), \quad (27)$$

for a d-dimensional system with a thermal scaling exponent $y_{\mathbf{T}}$. The function Σ can be expanded in a series

$$\Sigma(x)=\sum_{l=0}^{\infty} \Sigma_l x^l, \quad (28)$$

and for $O(n)$ models behaves for large x as

$$\Sigma(x)=A_\Sigma x^{[d-1-p(n)]/y_{\mathbf{T}}}, \quad (29)$$

where $p(1)=0$ and $p(2)=p(3)=1$.

Equations (27) to (29) are useful for the interpretation of the $O(n)$ transfer-matrix Monte Carlo results for the interface free energy. These results were obtained using finite sizes up to $L=12$, and populations typically consisting of 2500 or 5000 walkers. Typical run lengths are 5000 steps, where each step means the addition of a surface layer of $L \times L$ spins. Variance-reducing trial vectors [see Eq. (20)] were constructed for path lengths up to 5. As before, the variance of the Monte Carlo process was observed to decrease considerably with increasing path length. For each system size, interface free energies were obtained for approximately ten different couplings in a range of about 10% around the critical points of the Ising and planar models, and about 1% for the case of the Heisenberg model.

On the basis of these results for the Ising ($n=1$) case, the function Σ is shown in Fig. 3. This data collapse is achieved by means of a least-squares fit with parameters K_c, y_T, and a and 13 Taylor coefficients Σ_l, a generalization of a technique used in the past.[5]

To check if the system sizes were in the asymptotic finite-size scaling regime, fits were done both with and without the $6 \times 6 \times \infty$ data. The results of these fits are displayed in Table
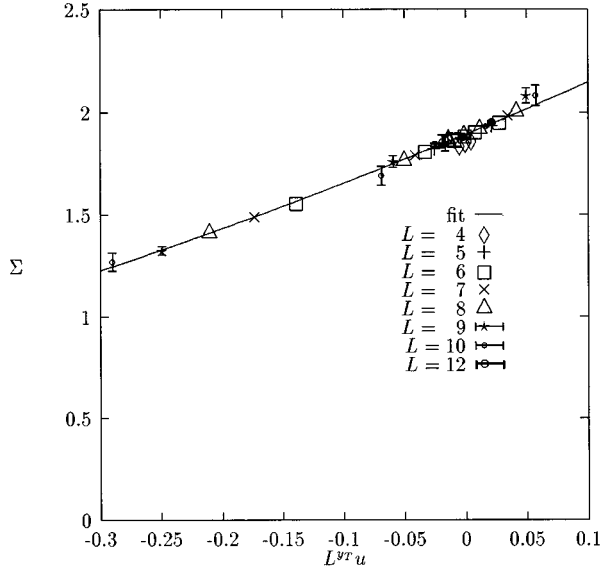
FIG. 5. Finite-size scaling plot for the interface free energy of the three-dimensional Heisenberg model.

TABLE II. Parameters, as defined in the text, and their standard errors for the scaling function of the interfacial free energy of the three-dimensional Ising model.

	$n \geq 6$	$n \geq 8$
K_c	0.22162 ± 0.00002	0.22165 ± 0.00003
y_T	1.583 ± 0.004	1.594 ± 0.009
Σ_0	$0.6171 \pm 0.000\,7$	0.6194 ± 0.0025
Σ_1	2.6111 ± 0.0176	2.5650 ± 0.0505
Σ_2	6.0475 ± 0.1001	5.8047 ± 0.2565
Σ_3	9.2362 ± 0.3052	8.4073 ± 0.6724
Σ_4	6.0087 ± 0.6350	4.8601 ± 1.0249
Σ_5	-13.6165 ± 0.9830	-10.8331 ± 1.7414
Σ_6	-33.2578 ± 4.0574	-24.7108 ± 5.6481
Σ_7	4.4790 ± 3.3849	1.4272 ± 4.2080
Σ_8	70.6918 ± 11.4018	46.6641 ± 14.0107
Σ_9	28.0314 ± 7.9088	21.4636 ± 8.5327
Σ_{10}	-69.1548 ± 14.3387	-40.6066 ± 15.7549
Σ_{11}	-43.4754 ± 10.0932	-27.7038 ± 10.3867
Σ_{12}	25.2137 ± 6.4840	13.2245 ± 6.4356
Σ_{13}	18.8658 ± 4.8588	10.5010 ± 4.6564
a	-2.65 ± 0.16	-2.65 ± 0.37

II in the Appendix. To summarize, the results are $K_c = 0.221\,62 \pm 0.000\,02$ and $y_T = 1.584 \pm 0.004$ using data with $L = 6-12$, and $K_c = 0.221\,67 \pm 0.000\,04$ and $y_T = 1.584 \pm 0.014$ if the $L = 6$ are omitted. These results agree well with accurate determinations using other methods (see, e.g., Refs. 16–18 and references therein) which appear to cluster about $K_c = 0.221\,655$ (with a margin of about 10^{-6}) and $y_T = 1.586$ (with a precision of a few times 10^{-3}).

It is remarkable that the corrections to scaling appear to be very small, as appears from the data shown in Fig. 3. In standard Monte Carlo analyses[19] of $L \times L \times L$ systems these corrections are quite prominent, and form an obstacle to the accurate determination of critical parameters.

The scaling plot shown in Fig. 3 can be used to determine the amplitude A_Σ graphically: On a double-logarithmic plot the asymptotic slope of the curve follows from the known value of the thermal exponent y_T; cf. Eq. (29). The problem of calculating this amplitude has attracted considerable attention lately and the reader is referred to a paper by Shaw and Fisher[20] for details and further references to the literature. For the largest values of the scaled temperature variable x, we find $A'_\Sigma = A_\Sigma K_c^{2/y_T} = 1.8$, while the trend with x is an increasing one. This value is somewhat larger than Mon's[21] estimate $A_\Sigma = 1.58 \pm 0.05$, but still in the range $1.4 \leq A_\Sigma \leq 2.0$ obtained by Shaw and Fisher. As a final comment we note that Mon's method requires systems of linear dimensions in excess of 48 to reach the asymptotic infinite-size regime, with an increasing trend of the estimates of A_Σ with increasing $x = L^{y_T} u$.

A similar analysis was performed for the planar model ($n = 2$). In comparison with the Ising case, the scaling function Σ behaves more smoothly as a function of x, so that a satisfactory fit could be obtained with fewer Taylor coefficients. The fitted parameters, which are K_c, y_T, and a and eight Taylor coefficients Σ_n, are shown in Table III of the Appendix. Our results for the critical point are

$K_c = 0.454\,10 \pm 0.000\,03$ for system sizes $L = 6-12$, and $K_c = 0.454\,13 \pm 0.000\,05$ for $L = 8-12$. These values are close to results from series expansions[22] $K_c = 0.453\,86$ and more recently[23] $K_c = 0.454\,14 \pm 0.000\,007$; and from cluster Monte Carlo calculations[24] $K_c = 0.454\,08 \pm 0.000\,08$. Also our results for the temperature exponent, namely, $y_T = 1.491 \pm 0.003$ for $L \geq 6$ and $y_T = 1.487 \pm 0.006$ for $L \geq 8$, are in a good agreement with existing results; we quote the coupling-constant-expansion value[27] $y_T = 1.495 \pm 0.005$ and the Monte Carlo result[24] $Y_T = 1.492\,6 \pm 0.0037$.

Fitted with these parameters the data collapse very well onto the function Σ, as shown in Fig. 4. Again, this scaling plot can be used to determine the amplitude A_Σ graphically: In this case the asymptotic power-law exponent is $1/y_T$. A fit of the data at the highest available values of $x = L^{y_T} u$ leads to $A_\Sigma = 5.9$, while the trend is still increasing with x.

The calculations for the Heisenberg case $n = 3$ were clustered in a narrow interval around the critical temperature, and were not aimed at an accurate determination y_T. Thus, the transfer-matrix Monte Carlo data could be analyzed by means of a least-squares fit with less parameters: K_c, and y_T and three Taylor coefficients Σ_n. The parameters are shown in Table IV in the Appendix. The result for y_T is well within the statistical accuracy, equal to the known coupling-constant-expansion value[25] $y_T = 1.418$. Including the latter value as a known variable in the fits leaves our results for the critical point practically unchanged. These are $K_c = 0.692\,91 \pm 0.000\,04$ for system sizes $L = 6-12$, and $K_c = 0.692\,94 \pm 0.000\,08$ for $L = 8-12$. These values are close to results from series expansions,[22] $K_c = 0.692\,94$, and more recently,[23] $K_c = 0.6929 \pm 0.001$; and from Monte Carlo calculations,[27] $K_c = 0.692\,88 \pm 0.000\,04$; see also Ref. 26.

The data collapse for the $n = 3$ case onto the function Σ as determined by the least-squares fit is shown in Fig. 5.

TABLE III. Parameters, as defined in the text, and their standard errors for the scaling function of the interfacial free energy of the three-dimensional planar model.

	$n \geqslant 6$	$n \geqslant 8$
K_c	0.45410 ± 0.00003	0.45413 ± 0.00005
y_T	1.491 ± 0.003	1.487 ± 0.006
Σ_0	1.2448 ± 0.0010	1.2469 ± 0.0033
Σ_1	2.5592 ± 0.0144	2.5929 ± 0.0345
Σ_2	2.4285 ± 0.0439	2.4738 ± 0.0796
Σ_3	0.9881 ± 0.0623	0.9544 ± 0.1031
Σ_4	-0.5096 ± 0.0664	-0.4292 ± 0.1050
Σ_5	-0.7770 ± 0.1579	-0.5171 ± 0.2741
Σ_6	0.1737 ± 0.0754	0.0793 ± 0.1204
Σ_7	0.4346 ± 0.1503	0.1847 ± 0.2567
a	-0.7805 ± 0.1159	-0.9245 ± 0.2259

Finally we remark that, although in each of the cases $n = 1$, 2, and 3 the finite-size effect appears to be small for $L \geqslant 6$, it is large for $L = 4$. For this reason the $L = 4$ data were not included in the fits.

ACKNOWLEDGMENTS

TABLE IV. Parameters, as defined in the text, and their standard errors for the scaling function of the interfacial free energy of the three-dimensional Heisenberg model. The Monte Carlo data were taken relatively close to K_c, so that the temperature exponent y_T is not accurately determined. The accuracy of K_c is unaffected.

	$n \geqslant 6$	$n \geqslant 8$
K_c	0.69291 ± 0.00004	0.69294 ± 0.00008
y_T	1.44 ± 0.07	1.55 ± 0.18
Σ_0	1.8919 ± 0.0015	1.8933 ± 0.0043
Σ_1	2.4563 ± 0.3123	1.9036 ± 0.7665
Σ_2	0.7991 ± 0.3005	0.5097 ± 0.4651

APPENDIX: SCALING PLOT PARAMETER ESTIMATES

Tables II – IV contain estimates of the parameters used in the finite-size scaling plots for the interface free energy of $O(n)$ models, as discussed in Sec. IV.

[1] H.A. Kramers and G.H. Wannier, Phys. Rev. **60**, 252 (1941).

[2] H.F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959).

[3] For a discussion of these correlations see K.J. Runge, Phys. Rev. B **45**, 12 292 (1992).

[4] M.P. Nightingale, E. Granato, and J.M. Kosterlitz, Phys. Rev. B **52**, 7402 (1995).

[5] M.P. Nightingale and H.W.J. Blöte, Phys. Rev. Lett. **60**, 1562 (1988).

[6] M.P. Nightingale, in *Finite-Size Scaling and Simulation of Statistical Mechanical Systems,* edited by V. Privman (World Scientific, Singapore, 1990), pp. 287–351.

[7] C.J. Umrigar, M.P. Nightingale, and K.J. Runge, J. Chem. Phys. **99**, 2865 (1993).

[8] J.H. Hetherington, Phys. Rev. A **30**, 2713 (1984).

[9] M.P. Nightingale and H.W.J. Blöte, Phys. Rev. B **33**, 659 (1986).

[10] D.M. Ceperley and C.J. Umrigar (private communication).

[11] M.H. Kalos, J. Comput. Phys. **1**, 257 (1966); the original idea of ''forward walking'' predates this paper [M.H. Kalos (private communication)]. For further references see Ref. 11 of Ref. 12.

[12] K.J. Runge, Phys. Rev. B **45**, 7229 (1992).

[13] D.M. Ceperley and M.H. Kalos, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer, Berlin, 1979).

[14] C.J. Umrigar, K.G. Wilson, and J.W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988); in *Computer Simulation Studies in Condensed Matter Physics*, edited by D.P. Landau, K.K. Mon, and H.-B. Schüttler, Springer Proceedings in Physics Vol. 33 (Springer-Verlag, Berlin, 1988), p. 185.

[15] E. Granato and M.P. Nightingale, Phys. Rev. B **48**, 7438 (1993).

[16] H.W.J. Blöte, E. Luijten, and J.R. Heringa, J. Phys. A **28**, 6289 (1995).

[17] H.W.J. Blöte, J.R. Heringa, A. Hoogland, E.W. Meyer, and T.S. Smit, Phys. Rev. Lett. **76**, 2613 (1996).

[18] R. Gupta and P. Tamayo (unpublished); Int. J. Mod. Phys (to be published).

[19] A.M. Ferrenberg and D.P. Landau, Phys. Rev. B **44**, 5081 (1991).

[20] L. Shaw and M.E. Fisher, Phys. Rev. A **39**, 2189 (1989).

[21] K.K. Mon, Phys. Rev. Lett. **60**, 2749 (1988).

[22] K. Ohno, Y. Okabe, and A. Morita, Prog. Theor. Phys. **71**, 714 (1984).

[23] J. Adler, C. Holm, and W. Janke, Physica A **20**, 581 (1993).

[24] W. Janke, Phys. Lett. A **148**, 306 (1990).

[25] J.C. Le Guillou and J. Zinn-Justin, Phys. Rev. B **21**, 3976 (1980).

[26] C. Holm and W. Janke, Phys. Lett. A **173**, 8 (1993); Phys. Rev. B **48**, 936 (1993).

[27] K. Chen, A.M. Ferrenberg, and D.P. Landau, J. Appl. Phys. **73**, 5488 (1993).