Open problems in Monte Carlo renormalization group: Application to critical phenomena (invited)

Cite as: Journal of Applied Physics **61**, 3605 (1987); https://doi.org/10.1063/1.338688
Published Online: 04 June 1998

Rajan Gupta

Open problems in Monte Carlo renormalization group: Application to critical phenomena (invited)

Rajan Gupta[a)]
Theoretical Division, MS-B285, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

The Monte Carlo renormalization group (MCRG) methods and the underlying theory is reviewed. The Gupta–Cordery improved MCRG method is described and compared with the standard one. The emphasis is on the progress made in understanding the truncation errors in the linearized transformation matrix and on open problems. The problem of deteriorating exponents with long-range transformations is highlighted. Lastly, some of the existing methods for calculating the renormalized Hamiltonian are reviewed and evaluated.

This paper was not proofread by the author; however, it has been proofread by one of the Publication Chairmen.

I. INTRODUCTION

The development of Monte Carlo renormalization group method (MCRG) was essentially complete in 1979 with the work of Wilson,[1] Swendsen[2] and Shenker and Tobochnik.[3] Prior to this Ma[4] and Kadanoff[5] had provided key ingredients. There already exists extensive literature on MCRG and I direct the reader to it[1,3,6–8] for details and for a wider exposure. Similarly, the reviews[9,10] are a good starting point for background on spin systems. The topics I shall cover are (1) introduction to MCRG and its methodology; (2) improved Monte Carlo renormalization group (IMCRG); (3) comparison of the standard MCRG method and IMCRG with emphasis on the truncation errors; (4) renormalized Hamiltonians and methods to calculate them; and (5) open problems.

II. INTRODUCTION TO MCRG

Renormalization group[10–13] (RG) is a general framework for studying systems near the critical surface (defined by a divergent correlation length) where singularities in thermodynamic functions arise from coherence at all length scales. The MCRG method was developed to handle this problem of infinitely many coupled degrees of freedom so that sensible results can be obtained from finite lattices. There are two central ideas behind MCRG: The first is to average over the infinitely many degrees of freedom in discreet steps. The block degrees of freedom on the coarse lattice are the ones relevant to the description of the physical quantities of interest. The interaction between these averaged (block) fields is described by an infinite set of couplings that get renormalized at each blocking step. The second point is that there are no singularities in the coupling constant space even though the correlation length and thermodynamic quantities diverge on the critical surface.

The MCRG methods discussed here have a fundamental assumption that there exists a fixed point of the transformation and this is short ranged. Thus, even though an infinite number of couplings are generated under renormalization, we shall assume that only a few short-range ones are sufficient to simulate the system at a given scale and preserve the long distance physics.

A. Standard Monte Carlo

Consider a magnetic system consisting of spins $\{s\}$ on the sites of a d-dimensional lattice L described by a Hamiltonian H. From the outset, H will include all possible couplings $\{K_\alpha\}$. The behavior of all thermodynamic quantities can be determined from a detailed knowledge of the partition function

$$Z = \sum e^{-H} = \sum e^{K_\alpha S_\alpha}, \tag{1}$$

where S_α are the interactions. In Monte Carlo, configurations of spins on the original lattice are generated by the Metropolis,[14] heat bath,[15] molecular dynamics (alias Microcanonical[16] or the Langevin[17,18] algorithm with a Boltzmann distribution $e^{-H} \equiv e^{K_\alpha S_\alpha}$). All thermodynamic quantities are given as simple averages of correlation functions over these "importance sampled" configurations. The accuracy of the calculations depend on the size of the statistical sample and on the lattice size L used. Both these quantities depend on the largest correlation length ξ in the system. Near the critical temperature T_c associated with second-order phase transitions, the correlation length and consequently thermodynamic quantities like the specific heat, etc., diverge as functions of $(T - T_c)$ with universal critical exponents. These have been calculated for many systems analytically or by Monte Carlo using finite size scaling or by the MCRG method. Because ξ diverges at T_c, long runs are needed to counter the critical slowing down. Also, to control finite size effects the lattice size has to be maintained at a few times ξ. The problem of critical slowing down is addressed by analyzing update algorithms (Metropolis versus heat bath versus Microcanonical versus Langevin with acceleration techniques like multigrid,[19] Fourier acceleration,[18,20] etc.). The optimum method is, of course, model dependent and has to take care of metastability (local versus global minima) and global excitations like vortices, instantons, etc., that are not efficiently handled by local changes. This last feature has not received adequate attention. To control the second problem in standard Monte Carlo, effects of a finite lattice especially

[a)] J. Robert Oppenheimer Fellow.

 0021-8979/87/083605-07$02.40

as $\xi \to \infty$, finite size scaling[10] has been used with success. In this review I shall concentrate on MCRG. First I shall describe how universality and scaling are explained by the renormalization group.

The renormalization group transformation (RGT) is an operator R defined on the space of coupling constants $\{K_\alpha\}$. In practice the RGT is a prescription to average spins over a region of size b, the scale factor of the RGT, to produce the block spin which interacts with an effective theory $H^1 = R(H)$. The two theories H and H^1 describe the same long distance physics but the correlation length in lattice units $\xi \to \xi/b$. If this RGT has a fixed point H^* such that $H^* = R(H^*)$, then clearly the theory is scale invariant at that point and ξ is either 0 or ∞. An example of a fixed point with $\xi = 0$ is $T = \infty$ and these are trivial. The interesting case is $\xi = \infty$ about which the theory is governed by a single scale ξ. I will discuss this assumption of hyperscaling, i.e., a single scale controlling all physics, later. If this fixed point is unstable in 1 direction only [this direction is called the renormalized trajectory (RT)], then noncritical H close to H^* will flow away from H^* along trajectories that asymptotically converge to the RT. Thus the long distance physics of all the trajectories that converge is identical and is controlled by the RT. Similarly, points ϵ away from H^* on the $\infty - 1$ dimension hypersurface at which $\xi = \infty$ (the critical surface) will converge to H^*. The fact that the fixed point with its associated RT control the behavior of all H in the neighborhood of H^* is universality.

Next, consider a noncritical H that approaches H^* along the RT. Thermodynamic quantities depend on a single variable, i.e., distance along the RT. This is scaling. Corrections to scaling occur when H does not lie on the RT. These are governed by the irrelevant eigenvalues of the RGT which give the rate of flow along the critical surface towards H^* and for H not on the RT, the rate of convergence towards it. The relevant eigenvalue gives the rate of flow away from the fixed point along the unstable direction RT and is related to the critical exponent ν. This terse exposé ends with a word of caution; all these statements have validity close to H^*.

B. Standard MCRG method

In the MCRG method, configurations are generated with the Boltzmann factor $e^{K_\alpha S_\alpha}$ as in standard Monte Carlo. The RGT, $P(s^1,s)$, is a prescription for averaging variables over a cell of dimension b. The blocked variables $\{s^1\}$ are defined on the sites of a sublattice L^1 with lattice spacing b times that of L. They interact with a priori undetermined couplings $\{K^1_\alpha\}$, and the configurations are distributed according to the Boltzmann factor e^{-H^1}, i.e.,

$$e^{-H^1(s^1)} = \sum P(s^1,s) e^{-H(s)} . \tag{2}$$

All expectation values, with respect to the Hamiltonian H^1, can be calculated as simple averages on the blocked configurations. The blocking is done n times to produce a sequence of configurations distributed according to the Hamiltonians H^n. They all describe the same long distance physics but on increasingly coarse lattices. The fixed point H^*, the RT, and the sequence of theories H^n generated from a given starting H depend on the RGT.

The RGT should satisfy the Kadanoff constraint

$$\sum^{1} P(s^1,s) = 1 , \tag{3}$$

independent of the state $\{s\}$. This guarantees that the two theories H and H^1 have the same partition function. The RGT should also incorporate the model's symmetry properties; a notable example is the choice of the block cell in the antiferromagnetic Ising model. Usually, there exists considerable freedom in the choice of the RGT. In fact, many different RGT can be used to analyze a given model. In such cases a comparison of the universal properties should be made and the RGT dependent quantities isolated. I defer discussion on how to evaluate the efficiency of a RGT to Sec. II E.

C. Methods to calculate the critical exponent

There are two methods to calculate the critical exponents from expectation values calculated as simple averages over configurations. In both there is an implicit assumption that the sequence H^n stays close to H^*. The more popular method is due to Swendsen[2,7] in which the critical exponents are calculated from the eigenvalues of the linearized transformation matrix $T^n_{\alpha\beta}$ which is defined as

$$T^n_{\alpha\beta} = \frac{\partial K^n_\alpha}{\partial K^{n-1}_\beta} = \frac{\partial K^n_\alpha}{\partial \langle S^n_\sigma \rangle} \frac{\partial \langle S^n_\sigma \rangle}{\partial K^{n-1}_\beta} . \tag{4}$$

Each of the two terms on the right is a connected two-point correlation matrix

$$U^n_{\alpha\beta} \equiv \frac{\partial \langle S^n_\sigma \rangle}{\partial K^{n-1}_\beta} = \langle S^n_\sigma S^{n-1}_\beta \rangle - \langle S^n_\sigma \rangle \langle S^{n-1}_\beta \rangle \tag{5}$$

and

$$D^n_{\alpha\beta} \equiv \frac{\partial \langle S^n_\sigma \rangle}{\partial K^n_\beta} = \langle S^n_\sigma S^n_\beta \rangle - \langle S^n_\sigma \rangle \langle S^n_\beta \rangle . \tag{6}$$

Here $\langle S^n_\sigma \rangle$ are the expectation values on the nth renormalized lattice and K^n_σ are the corresponding couplings. The relevant exponent ν is found from the leading eigenvalue λ_t of $T^n_{\alpha\beta}$ as

$$\nu = \ln b / \ln \lambda_t , \tag{7}$$

where b is the scale factor of the RGT. The magnetic exponent is given by replacing λ_t by λ_h in Eq. (7) where λ_h is the largest eigenvalue of T constructed from odd interactions. I have restricted the discussion to the special case of one relevant eigenvalue. In general, systems can have multicritical points with more than one relevant interaction. Next, the eigenvalues which are smaller than one (called irrelevant) yield exponents that control corrections to scaling. An eigenvalue of exactly one is called marginal. Lastly, there is an additional class of eigenvalues, the redundant eigenvalues, that are not physical. Their value depends on the RGT, so one way to isolate them is to repeat the calculation with a different RGT. I shall return to these in Sec. II E.

The accuracy of the calculation of exponents improves when they are evaluated close to the fixed point. This can be

achieved by starting from a critical point and blocking the lattice a sufficient number of times, i.e., H^n for large n. In this case the convergence is limited by the starting lattice size and how close the starting H^c is to H^*. This method can be improved if the renormalized couplings $\{K^n\}$ are determined starting from a known critical Hamiltonian. We assume that the couplings fall off exponentially with the range, so that H^* can be approximated by a small number of short-range couplings. For calculations in models for which the critical coupling is not known exactly, and when using a truncated H^n the system will flow away from H^* under blocking. This flow away from H^* can be avoided by first putting H back on the critical surface by Wilson's two-lattice method described in Sec. II D. In Sec. III and V, I describe a few methods to calculate the renormalized couplings.

A second possible improvement is to tune the RGT so that the convergence to H^* from a starting H^c takes fewer blocking steps. This is discussed in Sec. II E.

The practical limitation to calculate the exponents is that the two matrices U and D can only be determined in a truncated subspace. Further, in order to set up T, the matrix D has to be inverted. Thus the determination of exponents has two types of truncation errors: The elements of the truncated T differ from the true T due to the inversion of a truncated D and the second come from diagonalizing a truncated T. These errors will be analyzed in detail in Sec. IV.

The second method to calculate the leading relevant exponent is due to Wilson.[6] Consider once again the two-point connected correlation function (the derivative of an expectation value) $\langle S^i_\alpha S^j_\beta \rangle_c$ with $j > i$. Expand S^i_α in term of the eigenoperators O^i_α of the RGT. Close to H^* the level dependence in O^i_α (equivalently in the expansion coefficients $c^i_{\alpha\beta}$) can be neglected. Then to the leading order

$$\langle S^i_\alpha S^j_\beta \rangle_c \sim \lambda_t^{j-i} c_{\alpha,t} \langle O_t S^j_\beta \rangle , \tag{8}$$

where λ_t is the leading relevant eigenvalue and corrections are suppressed by $(\lambda/\lambda_t)^{j-i}$. Thus for each α and β, the ratio $\langle S^i_\alpha S^j_\beta \rangle / \langle S^{i+1}_\alpha S^j_\beta \rangle$ gives an estimate for the leading eigenvalue λ_t. This method works even when the starting coupling is not exactly critical. The accuracy of the method improves if $j-i$ is large (since nonleading terms are suppressed geometrically) and if used close to the fixed point.

I have compared the results for the two methods in the $d=2$ Ising model[28] using a 64^2 lattice and blocking three times starting from a 44-term Hamiltonian H^2. For $i=1$ and $j=2,3,4$, $\lambda_t = 2.00(3)$, $2.01(2)$, and $2.01(1)$, while $\lambda_h = 3.658(5)$, $3.660(5)$, and $3.663(5)$. Swendsen's method gave $1.998(2)$, $1.993(3)$, $1.990(3)$, and $3.666(1)$, $3.662(2)$, $3.660(2)$, respectively, and thus seems slightly better. However, the trends leave room for Wilson's method becoming better for large j. So, further tests in other models need to be made.

The calculation of ν from the leading eigenvalue does not assume hyperscaling. The relation between ν and the specific-heat index α, i.e., $\alpha = 2 - \nu d$ does. A known cause of hyperscaling violations are dangerous irrelevant operators.[10] In the presence of these, universal scaling functions have a power-law singularity $1/(ut^\phi)^\mu$ in the limit $u \to 0$ where u is the irrelevant scaling field and ϕ is the scaling exponent. The renormalization group approach is preserved but the hyperscaling law is modified to $\alpha = 2 - \nu d + \mu|\phi|$. However, to predict α we need μ, the power with which the scaling function diverges. It is not known how to calculate this with MCRG. A side remark: in applying finite size scaling analysis to this case (with an enhanced definition of the scaling functions for the specific heat data), we need to specify u to study the divergence in the limit $u \to 0$. But scaling fields are a function of the RGT. So a MCRG calculation is necessary to identify it. Thus at present it is an open problem.

On the critical surface the two-point correlation functions [like in Eqs. (5) and (6)] diverge in the thermodynamic limit. However, their ratio is the rate of change of couplings and these are well behaved provided one considers only short-ranged correlation functions as will be shown later. The reason that MCRG is assumed to have better control over finite size effects is that if H^* is short ranged then a truncated $T^n_{\alpha\beta}$ is sufficient to determine the leading eigenvalue. Also, the finite size contributions to the elements $T^n_{\alpha\beta}$ fall off like the couplings, i.e., exponentially. Thus reliable estimates may be obtained from small lattices.

D. Wilson's two-lattice method to find a critical point

The critical temperature is not known analytically for most models. Also, couplings calculated after blocking may not be critical due to truncation and statistical errors. The following method can be used to put H on to the critical surface.

Consider MCRG simulations L and S with the same starting couplings K^0_α but on lattice sizes $L = b^n$ and $S = b^{n-1}$. If K^0_α is critical and after a few blockings the two theories are close to H^*, then all correlation functions attain their fixed point values. For noncritical starting H, expand about H^* in the linear approximation

$$\langle L^m_\alpha \rangle - \langle S^{m-1}_\alpha \rangle = \frac{\partial}{\partial K^0_\beta} \{ \langle L^m_\alpha \rangle - \langle S^{m-1}_\alpha \rangle \} \Delta K^0_\beta$$
$$= \{ \langle L^m_\alpha L^0_\beta \rangle_c - \langle S^{m-1}_\alpha S^0_\beta \rangle_c \} \Delta K^0_\beta \tag{9}$$

to determine ΔK^0_α. To reduce finite size effects the compared expectation values are calculated on the same size lattices. The critical coupling is given by

$$K^c_\alpha = K^0_\alpha - \Delta K^0_\alpha \tag{10}$$

and this estimate should be improved iteratively.

E. Optimization of the RGT

The freedom to choose the RGT leads to the question: What are the criteria by which to decide what is the best RGT. I will first address the question—what is the effect of changing the RGT on the fixed point and on the RT. The answer is: Changing the RGT moves the fixed point on the critical surface but only along redundant directions. A simple argument is as follows[21]: Consider two different RGT, R_1 and R_2, and their associated fixed points H^*_1 and H^*_2. There are no nonanalytic corrections to scaling at either fixed points and the associated RT. If these two points are distinct, then under $R_1 H^*_2$ flows to H^*_1. Consequently there are no scaling violations along the flow. This is by definition a re-

dundant direction. This also implies that the associated RT differ by redundant operators.

The presence of redundant operators does not effect the physics, however, it can obscure the results.[22] The redundant eigenvalues are not physical, depend on the RGT, and can be relevant or irrelevant. If a relevant redundant operator is present then the flows will not converge to the H^* or to the RT. Thus the first criterion in picking a RGT is that the redundant eigenvalues be less than one and small. Then the irrelevant redundant operators can be used to make the Hamiltonian short ranged (local).

We desire the convergence to H^* be fast. This gives the second criterion: the coefficients of the leading irrelevant operators in $(H^* - H_c)$ should be made small.

Swendsen[23] has conjectured that the fixed point can be moved anywhere on the critical surface by tuning the RGT. In particular, if a given H^c is made H^*, then that RGT is optimal. There is some support for this in spin systems, where by making the RGT long ranged, one can successively kill terms in the renormalized Hamiltonian. In the $d=2$ Ising model we[24] used a 21-term RGT to bring H^1 close to the nearest-neighbor H^c. We looked at all 44 even interactions that fit in a 3×3 square of spins and found they were small compared to the simple majority rule. Swendsen[23] had found a similar behavior for the $d=2$ Ising model with a tuned 10-term RGT and measuring the 10 dominant couplings. He also found that the thermal eigenvalue is improved significantly compared to the majority rule with the update Hamiltonian taken to be nearest-neighbor H_c and in separate calculation a 10-term truncated renormalized Hamiltonian H^1. The second comparison is not complete because his determination of renormalized couplings have large truncation errors. More careful tests with the $d=2$ Ising model show that the improvement in the thermal exponent is not systematic.[24] In all cases (different range of the RGT), the value of ν increases and in most cases it overshoots the known exact result. Thus the improvement seen by Swendsen in $d=3$ is misleading because the simple majority rule RGT gives too low a value and adding more terms increases it; but this process is not convergent. That this is a problem is clear from the results for the magnetic exponent. In $d=2$, the magnetic eigenvalue at first blocking with the majority rule is[24] 3.683(2), which agrees with the earlier result of Swendsen.[7] Gausterer and Lang[25] find 3.692(3) with a three-parameter RGT of slightly larger range. Umrigar and I[24] find 3.713(2) with a 21-parameter RGT. Since the exact result is 3.668, we conclude that the eigenvalue becomes worse on going beyond the RGT based on the block cell. This is surprising especially because the fixed point is at zero odd couplings and these remain zero under tuning the RGT.

The deterioration of exponents raises a fundamental question about the validity of MCRG. Are there long-range couplings being induced that will surface in the calculations if the $\{K_\alpha\}$ are measured after many blockings? Thus there are two additional things to check in this approach: first, whether the coefficients of the RGT terms fall off like the couplings with the range, i.e., exponentially, and second, whether the long-range untuned couplings continue to fall off at least as fast as before after many blockings. At present, we[24] only have results in the $d=2$ Ising model, keeping all terms that fit in a 3×3 square after one blocking step. This is not enough.

To summarize, the criterion for an optimum RGT is to make the H^* and the RT as short ranged as possible (kill the leading irrelevant operators) and to have small redundant eigenvalues. In critical phenomena, the improvement can be quantified by measuring the convergence of the exponents as a function of the blocking level. The present status of understanding of how best to optimize MCRG is ambiguous. For the moment let me conclude this section by starting: In light of the above contradiction, we need to investigate further because one of the basic assumptions of MCRG may be wrong.

III. IMPROVED MONTE CARLO RENORMALIZATION GROUP[26]

I shall describe the Gupta–Cordery MCRG method (IMCRG) in some detail. In this method, too, the renormalized Hamiltonian and the linearized transformation matrix T are determined in some truncated space of interactions. However, in this subspace they have no additional truncation errors, i.e., the determined quantities have their infinite component values. Next, there are no long time correlations even on the critical surface and the block n-point correlation functions like $\langle S^1_\alpha S^1_\beta\rangle - \langle S^1_\alpha\rangle\langle S^1_\beta\rangle$ are calculable numbers. Because of these properties, the method allows a careful error analysis in the determination of the exponents from a truncated T.

In the IMCRG method the configurations $\{s\}$ are generated with the weight

$$P(s^1,s)e^{-H(s)+H^g(s^1)}, \tag{11}$$

where H^g is a guess for H^1. Note that both the site and block spins are used in the update of the site spins. In analogue to Eq. (2), the distribution of the block spins is given by

$$e^{-H^1(s^1)+H^g(s^1)} = \sum P(s^1,\ s)e^{-H(s)+H^g(s^1)}. \tag{12}$$

If $H^g = H^1$, then the block spins are completely uncorrelated and the calculation of the n-point functions on the block lattice is trivial:

$$\langle S^1_\alpha\rangle = 0, \quad \langle S^1_\alpha S^1_\beta\rangle = n_\alpha\delta_{\alpha\beta}, \tag{13}$$

where for the Ising model (and most other models) the integer n_α is simply a product of the number of sites times the multiplicity of the interaction type S_α. When $H^g \neq H^1$, then to first order

$$\langle S^1_\alpha\rangle = \langle S^1_\alpha S^1_\beta\rangle_{H^n=H^1}(K^1 - K^g)_\beta. \tag{14}$$

Using Eqs. (13) and (14), the renormalized couplings $\{K^1_\alpha\}$ are determined with no truncation errors

$$K^1_\alpha = K^g_\alpha + (\langle S^1_\alpha\rangle/n_\alpha). \tag{15}$$

This procedure can be iterated—use H^{n-1} as the spin H in Eq. (11) to find H^n. If the irrelevant eigenvalues are small, then after two or three repetitions of the RGT, the sequence H^n converges to the fixed point Hamiltonian H^* which is assumed to be short ranged. For the $d=2$ Ising model, the

method has been shown to be extremely stable.[27] The linearity approximation, Eq. (14), is under control. An iteration process using a few thousand sweeps suffices to determine successively improved H^g up to an accuracy of $O(10^{-4})$. Beyond that the errors fall as $\sqrt{N}$ and the number of interactions that have to be included grows rapidly.

The one remaining approximation is in the use of a truncated H^{n-1} for the spin Hamiltonian in the update to find H^n. This is solved formally in a straightforward manner: In Eq. (11) use H^g as the guess for H^n. The update now involves the original spins and all block spins up to the nth level in the Boltzmann weight

$$P(s^n,s^{n-1})\ldots P(s^1,s)e^{-H(s)+H^g(s^n)}. \tag{16}$$

The four Eqs. (12)–(15) are unchanged except that the *level* superscript is replaced by n, i.e., the nth level block-block correlation matrix is diagonal and given by Eq. (13). With this modification, the H^n is calculated directly. The limitation on n is the size of the starting lattice. The other practical limitation is the complexity of the computer program. I have made the following comparison in the $d=2$ Ising model[28]: H^2 was calculated using (2.2) and by iterating, i.e., $H_c \to H^1 \to h^2$ in which case all interactions of strength $>5\times10^{-4}$ are retained in H^1. The statistical accuracy in both cases is $O(10^{-5})$. I find that the iterated answer is good to only 10^{-4}. Thus the truncation errors do conspire and get magnified. The lesson learned from the simple case of $d=2$ Ising model is that in order to get couplings correct to one part in 10^{-5} at $n=2$, it is necessary to include all couplings of strength $\geqslant 10^{-5}$ in H^1.

The calculation of the T matrix proceeds exactly as in the standard MCRG, i.e., Eqs. (4)–(6). However, in the limit $H^g = H^1$, the block-block correlation matrix D is diagonal and given by Eq. (13). Thus it has no truncation errors, can be inverted with impunity, and the final elements of T are free of all truncation errors. This is the key feature of IMCRG. The only error comes from finding the eigenvalues from a truncated T matrix. These errors can be estimated and the results improved perturbatively as explained in Sec. IV.

In addition to the advantages mentioned above, simulating with IMCRG, the system does not have critical slowing down. The correlation length ξ can always be made of $O(1)$, so finite size effects are dominated by the range of interactions, which by assumption of a short range H^* fall off exponentially. Thus, critical phenomenon can be studied on small lattices and with no hidden sweep to sweep correlations that invalidate the statistical accuracy of the results. Using H^0 as the known nearest-neighbor critical point $K_{nn}^c = 0.440\,686\,8$, we[24] find that the IMCRG results[27] for H^1 are independent (within the statistical accuracy $\approx 10^{-5}$) of finite size effects for lattice sizes 16, 32, 64, and 128. Again, only those couplings that fit into a 3×3 square were included.

IMCRG is in practice very similar to MCRG though a little more complicated because it requires a simultaneous calculation of a many term $H(s)$ and H^g at update. However, conceptually it is very different and powerful. I believe that IMCRG provides a complete framework to analyze the critical behavior of spin systems.

IV. TRUNCATION ERRORS IN THE LTM

Consider the matrix equation for T in block form

$$\begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} = \begin{pmatrix} U_{11} & U_{21} \\ U_{21} & U_{22} \end{pmatrix}, \tag{17}$$

where D_{11} and U_{11} are the two derivative matrices calculated in some truncated space of operators that are considered dominant. The elements of the submatrix T_{11} will have no truncation errors provided we can calculate

$$T_{11} = D_{11}^{-1}(U_{11} - D_{12}T_{21}). \tag{18}$$

In the IMCRG method the matrix D is diagonal and known, so D_{12} is 0. Thus elements of T_{11} determined from U_{11} have no truncation errors. The errors in the eigenvalues and eigenvectors arise solely from diagonalizing T_{11} rather than the full matrix T. Calculations in the $d=2$ Ising model have shown that these errors are large (of order 10%), and the convergence is not systematic, i.e., the result fluctuates about 2. This may be because all operators of a given range are not included. An open problem therefore is a robust criterion for classifying operators into sets such that including successive sets decreases the truncation error geometrically by a large factor.

The errors arising from using a submatrix T_{11} can be reduced significantly by diagonalizing

$$\begin{aligned} &T_{11} + T_{11}^{-1}T_{12}T_{21} \\ &\quad = D_{11}^{-1}U_{11} + (-D_{11}^{-1}D_{12} + T_{11}^{-1}T_{12})T_{21}, \end{aligned} \tag{19}$$

as shown by Shankar, Gupta, and Murthy.[29] The correction term $T_{11}^{-1}T_{12}T_{21}$ is the second-order perturbation result valid for all eigenvalues that are large compared to those of T_{22}. The matrix $T_{12}T_{21} \approx (T^2)_{11} - (T_{11})^2$ can be calculated approximately in IMCRG. There are errors (which I have ignored) due to the RG flow, because of which T^2 is evaluated at a different point than T. These errors depend on how close to H^* the calculation is done. For the $d=2$ Ising model we[28] find that the perturbative correction significantly decreases truncation errors in the relevant eigenvalues. Second, when multilevel IMCRG is used, Eq. (16), the exponents have much smaller fluctuations at earlier levels and are close to those from MCRG. So, the present status is that MCRG works just as well but with far less effort. The other thing we have learned from this study is that the difference between the calculated eigenvalue at $n=1$ ($1.97 \pm .01$) and the exact result, 2, does not seem to be due to truncation errors or statistics. It is most likely due to irrelevant operators causing corrections to scaling.

In standard MCRG, the calculations with $T = D_{11}^{-1}U_{11}$ have shown good convergence once few operators, $O(5$–$10)$, are included. The reason for this is an approximate cancellation between the two types of truncation errors. If in Eq. (17) we ignore terms with T_{22} and approximate $T_{11} = D_{11}^{-1}U_{11}$ then

$$-D_{11}^{-1}D_{12} + T_{11}^{-1}T_{12} \sim -D_{11}^{-1}D_{12} + U_{11}^{-1}U_{12}.$$

Usually the derivative matrices are roughly proportional, i.e., $U \sim \lambda_t D$ and the corrections fall off as the ratio of nonleading eigenvalues to the leading one λ_t. This statement can be checked by expanding operators in term of eigenopera-

tors. Thus Swendsen,[7] by calculating just $D_{11}^{-1}U_{11}$ and ignoring all truncation problems, was in effect canceling a large part of the truncation error [2nd term in Eq. (19)] against the error arising from diagonalizing a truncated matrix [perturbative correct, 3rd term in Eq. (19)]. This explains his success. Shankar[30] has found a correction term to further decrease the truncation effects in MCRG. Given the assumptions, the flow under a RG and the success of the procedure as it exists, an improvement will be hard to evaluate. However, the check needs to be made.

Thus, at present the best way to get accurate results is to use IMCRG to calculate the renormalized couplings and Swendsen's MCRG method to calculate the eigenvalues.

V. DETERMINATION OF THE RENORMALIZED HAMILTONIAN

The advantage of using a Hamiltonian close to H^* in MC simulations is to reduce the effect of operators that lead to scaling violations. There are, to the best of my knowledge, 11 methods in existence to calculate the renormalized couplings. These have been reviewed previously.[8] I shall here briefly describe only three methods most relevant to spin systems.

The generic problem of systematic errors in the estimate of the couplings due to the finite number of couplings kept in the analysis will be referred to as "truncation errors." This is a serious drawback because the errors can be very large and there is no way of estimating them without a second long simulation. Unlike IMCRG, all the following methods have uncontrolled truncation errors.

A. Swendsen's method[31] using the Callen representation

The block expectations values of interactions are calculated in two ways. First as simple averages over block configurations, and second using the Callen representation[32] with a guess for the block couplings. From these two estimates, the block couplings at n levels are determined simultaneously. The estimate is improved iteratively. The method is fast and easy to implement but it does have undetermined truncation errors.

B. Callaway–Petronzio–Wilson[33,34] method of fixed block spins

This method is useful for discrete spin systems like the Ising model and models in the same universality class. A MCRG calculation is modified by fixing all the block spins except one such that only a controllable few block interactions are nonzero. The system is simulated with the RGT used as an additional weight in the Metropolis algorithm. The ratio of probability of this unfixed spin being up to it being down is equal to a determined function of a certain number (depending on how many block interactions are nonzero) of block couplings. By using different configurations of fixed block spins a system of linear equations is set up from which the block couplings are determined. The drawback of this method, even for the Ising model, is that it is hard to set up the block spins so that only a few (≈ 10) block interactions are nonzero. Wilson showed that this can be done if one uses the lattice gas representation, i.e., 0 or 1 for spin values. The couplings in the ± 1 representation are then given by an expansion in the lattice gas couplings. The second improvement due to Wilson is that instead of a MC determination of the ratio of probabilities, the exact result can be obtained in the transfer matrix formalism. In the $d = 2$ Ising model, the convergence of the ± 1 couplings in terms of the lattice gas couplings is slow.[34] About a 1000 lattice gas couplings were necessary for an accuracy of $\approx 10^{-4}$. However, the calculation is nonstatistical and very fast.

C. Microcanonical (Creutz's Demon) method[35]

This method is very efficient if from a previous MCRG calculation expectation values of m block interactions at each of the n block levels are determined. To determine the corresponding couplings at the nth level, a microcanonical simulation is done (on a same size lattice as on which the block expectation values were calculated) with the corresponding m energies fixed and with one demon per interaction. The desired m couplings are then determined from the distribution of demon energies. The accuracy has a fundamental limitation for discrete spin systems because the demon energy and the total energy is discrete. This method shares an advantage with Swendsen's method; a single original calculation is necessary to determine the block expectation values on many levels. Thus if the simulated H is critical, then at each blocking level H^n is also on the critical surface.

The three methods have similar truncation errors. The truncated H^n is not critical. Also, note that the renormalized couplings and H^* are not universal but depend on the specific RGT. Therefore, this improvement program is within the context of a particular RGT.

Umrigar and I (Ref. 24) have performed the following test in the $d = 2$ Ising model: We used IMCRG to determine H^1 in the subspace of all two-spin and four-spin interactions that exist in a 3×3 square. This was then used to perform a standard MCRG calculation for the eigenvalues. The result was remarkable; the thermal eigenvalue is 2.001 ± 0.001 and the magnetic 3.669 ± 0.001 at the first level. The exact answers are 2 and 3.668. The same is true when H^2 is used as the results given after Wilson's method in Sec. II C show. We are extending the calculation to the $d = 3$ Ising model. If these stability tests work then we shall feel confident that a good way to calculate the exponents is to first calculate the renormalized couplings using IMCRG and then calculate the exponents by MCRG.

VI. OPEN PROBLEMS

I shall just list the problems that have already been discussed before and elaborate on the rest.

(1) Why do the thermal and magnetic exponents become worse under an optimization of the RGT to improve convergence to H^*? One possible reason may be the generation of long-range interactions. This is a violation of a fundamental assumption of renormalization group and MCRG is

in trouble unless a RGT exists for which this does not happen.

(2) A result obtained from the study of the $d = 2$ Ising model is that the T matrix has elements that grow along rows and fall along columns.[29] The leading left eigenvector is normal to the critical surface. Its elements give an estimate of the growth in the elements along the rows of the T matrix. For two spin interactions these grow like $x^{3/4}$. Therefore, a priori, the matrix T is badly behaved. The reason one gets sensible results is because the elements along the columns are observed to fall off faster (presumably exponentially). In problems examined so far we can arrange T to look like

$$\begin{pmatrix} A & B \\ \epsilon & D \end{pmatrix}, \tag{20}$$

with A the minimal truncated $n \times n$ block matrix that should be calculated. The case $\epsilon = 0$ is simple; there are no truncation errors in either method and diagonalizing A gives the n largest eigenvalues. Otherwise, the truncation error depends on the dot product of terms in ϵ and B. The requirement of absolute convergence in the dot product only guarantees that this product is finite but it may be arbitrarily large, i.e., $O(1)$. Therefore, for each model a careful study of the signs and magnitude of the elements in ϵ as a function of the RGT is necessary. This should also give a handle on the generation of long-range interactions with bad RGT. So we need to develop a theory for how elements along the columns fall off.

(3) So far I have only talked about the leading thermal eigenvalue. In the $d = 2$ Ising model, the irrelevant eigenvalues are known to be $1/2^n$. These are not well reproduced. For example, we consistently find a value close to 0.4 rather than 0.5 for the leading irrelevant eigenvalue. The second unknown in this case is the statistical accuracy. While for the relevant eigenvalue determined by Swendsen's method there seems to be an amazing cancellation of sweep to sweep correlations between the matrices U and D, this is not true for the rest.

(4) A classification scheme for operators according to the range of the interactions. *Range* should include a definition of a measure of distance and the number of spins in the interaction. The criterion of success to use is that on including all operators of a given range there should be a geometric decrease in truncation errors.

(5) A way to include the calculation of $\mu|\phi|$ by MCRG in models with hyperscaling violations caused by dangerous irrelevant operators.

ACKNOWLEDGMENTS

It is my pleasure to thank D. P. Landau for inviting me to participate in the conference. I also thank M. Fisher, A. Patel, R. Shankar, R. Swendsen, C. Umrigar, and K. G. Wilson who have openly shared their thoughts with me and thereby influenced the writeup.

[1] K. G. Wilson, in *Recent Developments in Gauge Theories*, Cargese, 1979, edited by G. t'Hooft *et al.* (Plenum, New York, 1980).
[2] R. H. Swendsen, Phys. Rev. Lett. **42**, 859 (1979).
[3] S. H. Shenker and J. Tobochnik, Phys. Rev. B **22**, 4462 (1980).
[4] S. K. Ma, Phys. Rev. Lett. **37**, 461 (1976).
[5] L. P. Kadanoff, Rev. Mod. Phys. **49**, 267 (1977).
[6] K. G. Wilson, in *Progress in Gauge Field Theories*, edited by G. 't Hooft *et al.* (Plenum, New York, 1984).
[7] R. H. Swendsen, in *Real Space Renormalization, Topics in Current Physics*, edited by Th. W. Burkhardt and J. M. J. van Leeuwen (Springer, Berlin, 1982), Vol. 30, p. 57.
[8] R. Gupta, in *Proceedings of the Wuppertal Conference on Lattice Gauge Theories: A Challenge in Large Scale Computing* (Plenum, New York, 1986).
[9] K. Binder, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer, Berlin, 1979), Vol. 7 in *Applications of Monte Carlo Methods in Statistical Physics* (Springer, Heidelberg, 1983).
[10] M. Fisher, in *Advanced Course in Critical Phenomenon*, Vol. 186 in *Lecture Notes in Physics* (Springer, Berlin, 1983).
[11] K. G. Wilson and J. Kogut, Phys. Rep. **12C**, 76 (1974).
[12] P. Pfeuty and G. Toulouse, *Introduction to the Renormalization Group and Critical Phenomenon* (Wiley, New York, 1978).
[13] D. Amit, *Field Theory, the Renormalization Group and Critical Phenomenon* (World Scientific, Singapore, 1984).
[14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[15] M. Creutz, Phys. Rev. D **21**, 2308 (1980).
[16] D. Callaway and A. Rehman, Phys. Rev. Lett. **49**, 613 (1982); M. Creutz, Phys. Rev. Lett. **50**, 1411 (1983); J. Polonyi and H. W. Wyld, Phys. Rev. Lett. **51**, 2257 (1983).
[17] G. Parisi and Wu Yongshi, Sci. Sin. **24**, 483 (1981).
[18] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. D **32**, 2736 (1985).
[19] A. Brandt, in *Multigrid Methods, Lecture Notes in Math* (Springer, Berlin, 1982), p. 960, and references therein.
[20] The idea was first discussed by G. Parisi, in *Progress in Gauge Field Theories*, edited by G. 't Hooft *et al.* (Plenum, New York, 1984).
[21] A more careful statment is made by M. E. Fisher and M. Randeria, Phys. Rev. Lett. **56**, 2332 (1986).
[22] R. Shankar and R. Gupta, Phys. Rev. B **32**, 6084 (1985).
[23] R. H. Swendsen, Phys. Rev. Lett. **52**, 2321 (1984).
[24] R. Gupta and C. Umrigar (unpublished).
[25] H. Gausterer and C. B. Lang, Graz preprint UNIGRAZ-UTP 4/86.
[26] R. Gupta and R. Cordery, Phys. Lett. A **105**, 415 (1984).
[27] R. Gupta in *Proceedings of the Tallahassee Conference on Advances in Lattice Gauge Theory* (World Scientific, Singapore, 1985).
[28] R. Gupta (unpublished).
[29] R. Shankar, R. Gupta, and G. Murthy, Phys. Rev. Lett. **55**, 1812 (1985).
[30] R. Shankar, Phys. Rev. B **33**, 6515 (1986).
[31] R. H. Swendsen, Phys. Rev. Lett. **52**, 1165 (1984).
[32] H. B. Callen, Phys. Lett. **4B**, 161 (1961).
[33] D. Callaway and R. Petronzio, Phys. Lett. **139B**, 189 (1984).
[34] K. G. Wilson and C. Umrigar (unpublished).
[35] M. Creutz, A. Gocksch, M. Ogilvie, and M. Okawa, Phys. Rev. Lett. **53**, 875 (1984).