

A Variational Approach to Monte Carlo Renormalization Group

Yantao Wu¹ and Roberto Car^{1,2}

¹The Department of Physics, Princeton University

²The Department of Chemistry, Princeton University

(Dated: September 29, 2017)

We present a Monte Carlo method for computing the renormalized coupling constants and the critical exponents within renormalization theory. The scheme, which derives from a variational principle, overcomes critical slowing down, by means of a bias potential that renders the coarse grained variables uncorrelated. The 2D Ising model is used to illustrate the method.

Since Wilson's seminal contribution [1] to renormalization group (RG) theory [2], there has been strong interest in methods to compute the renormalized coupling constants and the critical exponents in a non-perturbative fashion. This goal has been achieved with the Monte Carlo (MC) RG approach of Swendsen. In 1979, he introduced a method to compute the critical exponents, which did not require explicit knowledge of the renormalized Hamiltonian [7]. A few years later, he solved the problem of calculating the renormalized coupling constants, using an equality due to Callen [8] to write the correlation functions in a form explicitly depending on the couplings. By imposing that the standard MC expression of a correlation function and its corresponding Callen form be equal, he derived equations whose iterative solution led to the coupling constants [9]. Finding the renormalized Hamiltonian is an example of inverse statistical mechanical problem [10]. MCRG has been used successfully in many applications but difficulties related to sampling efficiency may be severe. Typically, the evaluation of the correlation functions near a critical point suffers from critical slowing down and is affected by large sampling errors in large systems. This difficulty can be alleviated with ingenious cluster algorithms [11], which, however, are limited to specific models.

Here we present an MCRG framework based on a variational principle for a biasing potential acting on the coarse grained degrees of freedom of a RG transformation. In our approach, the coupling constants and the critical exponents derive from the same unifying principle. Swendsen's formulae emerge as a special case, but our scheme also leads to formulations exempt from critical slowing down. In addition, it permits to estimate variationally the effect of truncating the Hamiltonian.

Although the approach is rather general, as we will briefly discuss in the conclusion, here we limit ourselves, for concreteness, to lattice models with discrete spin degrees of freedom, $\{\sigma\}$. A generic Hamiltonian has the form

$$H(\sigma) = \sum_{\alpha} K_{\alpha} S_{\alpha}(\sigma), \quad (1)$$

where the K_{α} are coupling constants and the S_{α} are operators acting on the spins σ , such as sums or products

of spins or combinations thereof.

RG considers a flux in the space of Hamiltonians (1) under scale transformations that reduce the linear size of the original lattice by a factor b . The rescaled degrees of freedom take the same discrete values of the original spins, to which they are related by a coarse graining transformation, $\sigma' = \tau(\sigma)$, satisfying a semigroup property [12]. For example, τ can be the block spin transformation of Kadanoff [13].

The distribution of the σ' is obtained from the distribution of the σ by tracing out the original degrees of freedom while keeping the σ' fixed:

$$p(\sigma') = \frac{\sum_{\sigma} \delta_{\tau(\sigma), \sigma'} e^{-H(\sigma)}}{Z} = \frac{e^{-H'(\sigma')}}{Z'}. \quad (2)$$

Here δ is the discrete Kronecker-delta function, Z and Z' are partition functions that ensure the normalization of the corresponding distributions. While the partition function Z' is invariant under RG transformations, the renormalized Hamiltonian H' is not, except at fixed points of the RG flow:

$$Z = \sum_{\sigma} e^{-H(\sigma)} = \sum_{\sigma'} e^{-H'(\sigma')} = Z' \quad (3)$$

and

$$H'(\sigma') = -\log \sum_{\sigma} \delta_{\tau(\sigma), \sigma'} e^{-H(\sigma)} \quad (4)$$

Repeated *ad infinitum*, the RG transformations generate a flux in the space of Hamiltonians, in which all possible coupling terms appear, unless forbidden by symmetry. For example, in an Ising model with no magnetic field, only even spin products appear. The space of the coupling terms is, in general, infinite. However, perturbative and non-perturbative calculations suggest that only a finite number of couplings should be sufficient for a given degree of accuracy.

In the proximity of a critical point, the distribution (2) of the block spins σ' displays a divergent correlation length, originating critical slowing down of local MC updates. This can be avoided by modifying the distribution of the σ' by adding to the Hamiltonian $H'(\sigma')$ a biasing potential $V(\sigma')$ to force the biased distribution of the

block spins, $p_V(\boldsymbol{\sigma}')$, to be equal to a chosen *target distribution*, $p_t(\boldsymbol{\sigma}')$. For instance, p_t can be the constant probability distribution. Then the $\boldsymbol{\sigma}'$ have the same probability at each lattice site and act as uncorrelated spins, even in the vicinity of a critical point.

It turns out that $V(\boldsymbol{\sigma}')$ obeys a powerful variational principle that facilitates the sampling of the Landau free energy [14]. In the present context, we define the functional $\Omega[V]$ of the biasing potential $V(\boldsymbol{\sigma}')$ by:

$$\Omega[V] = \log \frac{\sum_{\boldsymbol{\sigma}'} e^{-[H'(\boldsymbol{\sigma}') + V(\boldsymbol{\sigma}')]} }{\sum_{\boldsymbol{\sigma}'} e^{-H'(\boldsymbol{\sigma}')}} + \sum_{\boldsymbol{\sigma}'} p_t(\boldsymbol{\sigma}') V(\boldsymbol{\sigma}'), \quad (5)$$

where $p_t(\boldsymbol{\sigma}')$ is a normalized known target probability distribution. As demonstrated in [14], the following properties hold:

1. $\Omega[V]$ is a convex functional with a lower bound.
2. The minimizer, $V_{\min}(\boldsymbol{\sigma}')$, of Ω is unique up to a constant and is such that:

$$H'(\boldsymbol{\sigma}') = -V_{\min}(\boldsymbol{\sigma}') - \log p_t(\boldsymbol{\sigma}') + \text{constant} \quad (6)$$

3. The probability distribution of the $\boldsymbol{\sigma}'$ under the action of V_{\min} is:

$$p_{V_{\min}}(\boldsymbol{\sigma}') = \frac{e^{-(H'(\boldsymbol{\sigma}') + V_{\min}(\boldsymbol{\sigma}'))}}{\sum_{\boldsymbol{\sigma}'} e^{-(H'(\boldsymbol{\sigma}') + V_{\min}(\boldsymbol{\sigma}'))}} = p_t(\boldsymbol{\sigma}') \quad (7)$$

The above three properties lead to the following MCRG scheme.

First, we approximate $V(\boldsymbol{\sigma}')$ with $V_{\mathbf{J}}(\boldsymbol{\sigma}')$, a linear combination of a finite number of terms $S_{\alpha}(\boldsymbol{\sigma}')$ with unknown coefficients J_{α} , forming a vector $\mathbf{J} = \{J_1, \dots, J_{\alpha}, \dots, J_n\}$.

$$V_{\mathbf{J}}(\boldsymbol{\sigma}') = \sum_{\alpha} J_{\alpha} S_{\alpha}(\boldsymbol{\sigma}') \quad (8)$$

Then the functional $\Omega[V]$ becomes a convex function of \mathbf{J} , due to the linearity of the expansion, and the minimizing vector, \mathbf{J}_{\min} , and the corresponding $V_{\min}(\boldsymbol{\sigma}')$ can be found with a local minimization algorithm using the gradient and the Hessian of Ω :

$$\frac{\partial \Omega(\mathbf{J})}{\partial J_{\alpha}} = -\langle S_{\alpha}(\boldsymbol{\sigma}') \rangle_{V_{\mathbf{J}}} + \langle S_{\alpha}(\boldsymbol{\sigma}') \rangle_{p_t} \quad (9)$$

$$\frac{\partial^2 \Omega(\mathbf{J})}{\partial J_{\alpha} \partial J_{\beta}} = \langle S_{\alpha}(\boldsymbol{\sigma}') S_{\beta}(\boldsymbol{\sigma}') \rangle_{V_{\mathbf{J}}} - \langle S_{\alpha}(\boldsymbol{\sigma}') \rangle_{V_{\mathbf{J}}} \langle S_{\beta}(\boldsymbol{\sigma}') \rangle_{V_{\mathbf{J}}} \quad (10)$$

Here $\langle \cdot \rangle_{V_{\mathbf{J}}}$ is the biased ensemble average under $V_{\mathbf{J}}$ and $\langle \cdot \rangle_{p_t}$ is the ensemble average under the target probability distribution p_t . The first average is associated to the Boltzmann factor $\exp\{-[H'(\boldsymbol{\sigma}') + V(\boldsymbol{\sigma}')]\} = \sum_{\boldsymbol{\sigma}} \delta_{\tau(\boldsymbol{\sigma}), \boldsymbol{\sigma}'} \exp(-H(\boldsymbol{\sigma})) \exp(-V(\tau(\boldsymbol{\sigma})))$ and

can be computed with MC sampling. The second average can be computed analytically if p_t is simple enough. $\langle \cdot \rangle_{V_{\mathbf{J}}}$ always has inherent random noise, or even inaccuracy, and some sophistication is required in the optimization problem. Following [14], we adopt the stochastic optimization procedure of [15], and improve the statistics by running independent MC simulations, called *multiple walkers*, in parallel. For further details, consult [14] and the Supplementary Material (SM) [16].

The renormalized Hamiltonian $H'(\boldsymbol{\sigma}')$ is given by Eq. 6 in terms of $V_{\min}(\boldsymbol{\sigma}')$. Taking a constant p_t , we have modulo a constant:

$$H'(\boldsymbol{\sigma}') = -V_{\min}(\boldsymbol{\sigma}') = \sum_{\alpha} (-J_{\min, \alpha}) S_{\alpha}(\boldsymbol{\sigma}') \quad (11)$$

In this finite approximation the renormalized Hamiltonian has exactly the same terms of $V_{\min}(\boldsymbol{\sigma}')$ with renormalized coupling constants

$$K'_{\alpha} = -J_{\min, \alpha}. \quad (12)$$

The relative importance of an operator S_{α} in the renormalized Hamiltonian can be estimated variationally in terms of the relative magnitude of the coefficient $J_{\min, \alpha}$. When $J_{\min, \alpha}$ is much smaller than the other components of \mathbf{J}_{\min} , the corresponding $S_{\alpha}(\boldsymbol{\sigma}')$ is comparably unimportant and can be ignored. The accuracy of this approximation could be quantified by measuring the deviation of $p_{V_{\min}}(\boldsymbol{\sigma}')$ from $p_t(\boldsymbol{\sigma}')$.

To illustrate the method, we present a study of the Ising model on a $2D$ square lattice in the absence of a magnetic field. We adopt 3×3 block spins with the majority rule. 26 coupling terms were chosen initially, including 13 two-spin and 13 four-spin products. One preliminary iteration of variational RG (VRG) was performed on a 45×45 lattice starting from the nearest-neighbor Hamiltonian. The coupling terms with renormalized coupling constants smaller than 0.001 in absolute value were deemed unimportant and dropped from further calculations. 13 coupling terms, including 7 two-spin and 6 four-spin products, survived this criterion and were kept in all subsequent calculations [16]. Each calculation consisted of 5 VRG iterations starting with nearest-neighbor coupling, K_{nn} , only. All the subsequent iterations used the same lattice of the initial iteration. Standard Metropolis MC sampling [17] was adopted, and the calculations were done at least twice to ensure that statistical noise did not alter the results significantly.

In Fig. 1, results are shown for a 300×300 lattice with two initial K_{nn} , equal to 0.4355 and to 0.4365, respectively. When $K_{nn} = 0.4365$, the renormalized coupling constants increase over the five iterations shown, and would increase more dramatically with further iterations. Similarly, they decrease when $K_{nn} = 0.4355$. Thus, the critical coupling K_c should belong to the window $0.4355 - 0.4365$. The same critical window is found

for the 45×45 , 90×90 , 150×150 , and 210×210 lattices [16]. Because each iteration is affected by truncation and finite size errors, less iterations for the same rescaling factor would reduce the error. For example, 4 VRG iterations with a 2×2 block have the rescaling factor of a 16×16 block. The latter is computationally more costly than a calculation with 2×2 blocks, but can still be performed with modest computational resources. Indeed, with a 16×16 block, RG iterations on a 128×128 lattice gave a critical window $0.4394 - 0.4398$ [16], very close to the exact value, $K_c \sim 0.4407$, due to Onsager [18].

The statistical uncertainty of the calculated renormalized coupling constants is smaller with the variational method than with the standard method [19]. For example, using VRG and starting with $K_{nn} = 0.4365$ on a 300×300 lattice, we found a renormalized nearest-neighbor coupling equal to 0.38031 ± 0.00002 after one RG iteration with 3.968×10^5 MC sweeps. Under exactly the same conditions (lattice size, initial K_{nn} , coupling terms and number of MC sweeps) we found instead a renormalized nearest-neighbor coupling equal to 0.3740 ± 0.0003 with the standard method. In the VRG calculation we estimated the statistical uncertainty with the block averaging method [20], while we used the standard deviation from 14 independent calculations in the case of the standard method. A small difference in the values of the coupling constants calculated with VRG and the standard method is to be expected, because the two approaches are different embodiments of the truncated Hamiltonian approximation.

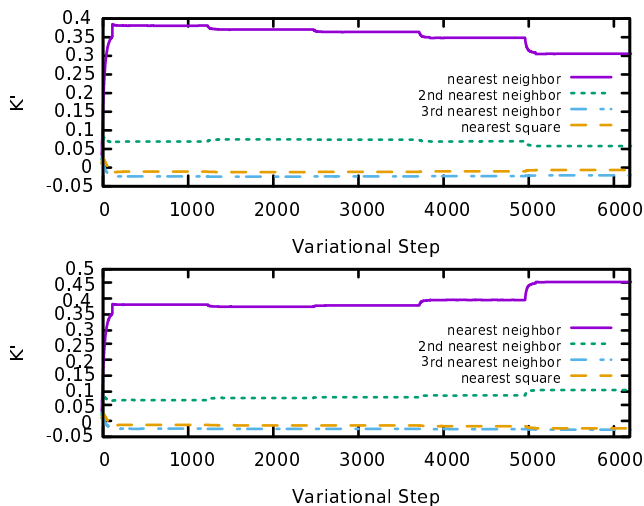


FIG. 1. (color online). Variation of the renormalized coupling constants over five VRG iterations on a 300×300 lattice. Each iteration has 1240 variational steps, each consisting of 20 MC sweeps. 16 multiple walkers are used for the ensemble averages in Eqs. 9 and 10. For clarity, we only show the four largest renormalized couplings after the first iteration. Full plots are reported in the SM [16]. Top: Simulation starting with $K_{nn} = 0.4365$. Bottom: Simulation starting with $K_{nn} = 0.4355$.

According to theory [21], the critical exponents are obtained from the leading eigenvalues of $\frac{\partial K'_\alpha}{\partial K_\beta}$, the Jacobian matrix of the RG transformation, at a critical fixed point. In order to find $\frac{\partial K'_\alpha}{\partial K_\beta}$ near a fixed point, we need to know how the renormalized coupling constants K'_α from a RG iteration on the Hamiltonian $H = \sum_\beta K_\beta S_\beta$, change when K_β is perturbed to $K_\beta + \delta K_\beta$, for fixed target probability p_t and operators S_α . The minimum condition, Eq. 9, implies $\frac{d\Omega}{dJ_\alpha} = 0$, i.e. for all γ :

$$\frac{\sum_{\sigma} S_\gamma(\sigma') e^{-\sum_{\beta} (K_{\beta} S_{\beta}(\sigma) - K'_{\beta} S_{\beta}(\sigma'))}}{\sum_{\sigma} e^{-\sum_{\beta} (K_{\beta} S_{\beta}(\sigma) - K'_{\beta} S_{\beta}(\sigma'))}} = \langle S_\gamma(\sigma') \rangle_{p_t}, \quad (13)$$

and

$$\frac{\sum_{\sigma} S_\gamma(\sigma') e^{-\sum_{\beta} ((K_{\beta} + \delta K_{\beta}) S_{\beta}(\sigma) - (K'_{\beta} + \delta K'_{\beta}) S_{\beta}(\sigma'))}}{\sum_{\sigma} e^{-\sum_{\beta} ((K_{\beta} + \delta K_{\beta}) S_{\beta}(\sigma) - (K'_{\beta} + \delta K'_{\beta}) S_{\beta}(\sigma'))}} = \langle S_\gamma(\sigma') \rangle_{p_t}. \quad (14)$$

Expanding Eq. 14 to linear order in $\delta K'_\alpha$ and δK_β , we obtain ([16])

$$A_{\beta\gamma} = \sum_{\alpha} \frac{\partial K'_\alpha}{\partial K_\beta} \cdot B_{\alpha\gamma}, \quad (15)$$

where

$$A_{\beta\gamma} = \langle S_\beta(\sigma) S_\gamma(\sigma') \rangle_V - \langle S_\beta(\sigma) \rangle_V \langle S_\gamma(\sigma') \rangle_V, \quad (16)$$

and

$$B_{\alpha\gamma} = \langle S_\alpha(\sigma') S_\gamma(\sigma') \rangle_V - \langle S_\alpha(\sigma') \rangle_V \langle S_\gamma(\sigma') \rangle_V. \quad (17)$$

Here $\langle \cdot \rangle_V$ denotes average under the biased Hamiltonian, $\tilde{H} = \sum_{\beta} [K_{\beta} S_{\beta}(\sigma) - K'_{\beta} S_{\beta}(\sigma')]$.

If we required the target average of $S_\gamma(\sigma')$ to coincide with the unbiased average under $H = \sum_{\beta} K_{\beta} S_{\beta}$, K' would necessarily vanish and Eqs. 16-17 would coincide with Swendsen's formulae [7]. If we used a uniform target probability, the σ' at different sites would be uncorrelated, and critical slowing down would be absent.

In practice, in order to compute the critical exponents, we first need to locate K_c . From the above calculations on the 45×45 , 90×90 , and 300×300 lattices with a 3×3 block spin, we expect that $K_c = 0.436$ should approximate the critical nearest-neighbor coupling in our model. Indeed an RG iteration starting from this value gives couplings that remain essentially constant, as illustrated in Figs. S11-S13 of the SM [16].

Then, we use Eqs. 15-17 to compute the Jacobian of the RG transformation by setting $K_c = 0.436$. The renormalized coupling constants after the first RG iteration represent K_α , and those after the second RG iteration represent K'_α . The results for biased and unbiased ensembles are shown in Table I, which reports the leading

even (*e*) and odd (*o*) eigenvalues of $\frac{\partial K'_\alpha}{\partial K_\beta}$ when including 13 coupling terms for the three $L \times L$ lattices with $L = 45, 90$, and 300 . As seen from the table, biased and unbiased calculations give slightly different eigenvalues, as one should expect, given that the respective calculations are different embodiments of the truncated Hamiltonian approximation. For $L = 300$ the results are well converged in the biased ensemble. By contrast, we were not able to obtain converged results for this lattice in the unbiased ensemble on the time scale of our simulation. The absence of critical slowing down in the biased simulation is demonstrated in Fig. 2, which displays the time decay of a correlation function in the biased and unbiased ensembles. See also Figs. S14-S15 of the SM [16].

	L	λ_1^e	λ_1^o
unbiased	45	2.970(1)	7.7171(2)
	90	2.980(3)	7.7351(1)
biased	45	3.045(5)	7.858(4)
	90	3.040(7)	7.870(2)
	300	3.03(1)	7.885(5)
Exact		3	7.8452

TABLE I. Leading even (*e*) and odd (*o*) eigenvalues of $\frac{\partial K'_\alpha}{\partial K_\beta}$ at the approximate fixed point found with VRG, in both the unbiased and biased ensembles. The number in parentheses is the statistical uncertainty on the last digit, obtained from the standard error of 16 independent runs. 13 (5) coupling terms are used for even (odd) interactions. The calculations used 10^6 MC sweeps for the 45×45 and 90×90 lattices, and 5×10^5 sweeps for the 300×300 lattice.

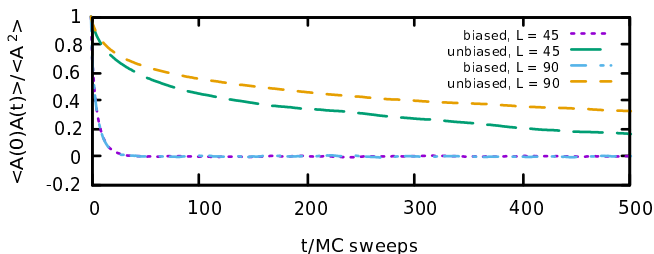


FIG. 2. (color online). Time correlation of the estimator $A = S_0(\boldsymbol{\sigma})S_0(\boldsymbol{\sigma}')$ on 45×45 and 90×90 lattices (Eq. 16). S_0 is the nearest neighbor term in the simulations of Table I.

The fixed point used for Table I is approximate, and we did not make any effort to fine tune the approximation. Refinements could be done iteratively using Eqs. 15-17, as we will discuss in a future paper. There is an important benefit in knowing accurately the location of the fixed point, because then a single RG iteration, instead of multiple implicit iterations would suffice to compute the Jacobian. Moreover, one could use small block spins, having a smaller statistical uncertainty than larger block spins.

In summary, we have unified the calculation of critical exponents and renormalized couplings within the same

framework. A key feature of our approach is that we adopt a biased ensemble, $\langle \cdot \rangle_V$, for the averages. This not only simplifies the algorithm, but also enhances the sampling. In fact, the original motivation for the variational principle [14] was to overcome the long correlation time in first-order phase transitions. The bias potential constructed by optimizing the functional acquires a history-dependence that discourages the sampling of previously visited configurations [14], thereby breaking the long correlation time of the unbiased simulation. In the RG context, enhanced sampling eliminates critical slowing down. We expect that it should be also helpful in systems with deep local free energy minima, as the variational method was originally designed to deal precisely with such systems.

The finite size of the numerical samples is a source of error. If the RG iterations are carried out on a single $L \times L$ lattice, the coarse grained lattice will have size $\frac{L}{b} \times \frac{L}{b}$. Then, as noted in [19], the calculated renormalized couplings will have different size errors on the $L \times L$ and $\frac{L}{b} \times \frac{L}{b}$ lattices. A better way, as suggested in [22], would be to perform calculations on two lattices, $L \times L$ and $\frac{L}{b} \times \frac{L}{b}$, so that the coarse grained lattice rescaled by b^n , at the n th iteration starting from $L \times L$, would coincide with the lattice rescaled by b^{n-1} , at the $(n-1)$ th iteration starting from $\frac{L}{b} \times \frac{L}{b}$. In this way, two successive RG iterations have the same lattice size, with a significant cancellation of finite size errors. We plan to discuss in a future paper how this idea could be implemented within VRG.

Finally some considerations are in order. It was noted in [23] that RG ideas can be used to obtain the full thermodynamic functions, not just their singular components. In fact one has [23]:

$$\exp [H'(K'; \boldsymbol{\sigma}') + Ng(K)] = \sum_{\boldsymbol{\sigma}} \delta_{\tau(\boldsymbol{\sigma}), \boldsymbol{\sigma}'} \exp [H(K; \boldsymbol{\sigma})] \quad (18)$$

Here $g(K)$ is the regular component of the free energy, which can also be accessed within our approach, as shown in the SM [16] where we demonstrate that $g(K)$ is precisely the thermodynamic free energy per site in the biased ensemble. It is thus interesting, and somewhat surprising, that the information on the critical behavior is fully contained in the statistical ensemble $\langle \cdot \rangle_V$, even though $g(K)$ is a regular function and $\langle \cdot \rangle_V$ does not show singular behavior.

Our approach can be extended beyond lattice models with discrete (spin) degrees of freedom. Field theoretical models, like the Ginzburg-Landau free energy functional $\Phi[\psi]$ ([24][25]) could be studied by representing the field $\psi(x)$ on a regular lattice, either in direct or reciprocal space. The values of the field on the lattice are a set of continuous and unbounded degrees of freedom. RG transformations on these variables can be studied with VRG, by conveniently adopting a Gaus-

sian rather than a constant probability p_t . We note that two levels of coarse-graining are required to study critical phenomena starting from the atomistic scale. First, one constructs a Ginzburg-Landau functional by coarse-graining the atomistic model and then one applies RG to the Ginzburg-Landau model. Interestingly, both levels of coarse graining are greatly facilitated by the same variational principle [14] adopted here to formulate VRG. The first step of this program, i.e. extracting a Ginzburg-Landau model from atomistic simulations has recently been presented in the literature [26].

In this paper we have focused on static phenomena, but MCRG approaches exist also for dynamic critical phenomena (see e.g. [27]). Formulating VRG for dynamics is beyond the scope of the present study and is left to future investigation.

All the codes used in this project were written in C++, and would be available upon request. The authors would like to thank C. Castellani and L. Pietronero for discussions. Partial support for this work was provided by the Department of Energy under Grant no. DE-FG02-05ER46201. Part of this work was done while R.C. was on sabbatical leave at the Sapienza University in Rome (Italy). The authors are grateful for the hospitality of the Beijing Computational Science Research (CSRC) during the final stage of manuscript preparation.

[1] K. G. Wilson, Phys. Rev. B **4**, 3174 (1971).

[2] RG theory originated in the context of quantum field theory ([3]–[6]). For a brief history of RG theory, including the early papers, see [12].

[3] E. C. G. Stueckelberg and A. Peterman, Helv. Phys. Acta **26**, 499 (1953).

- [4] M. Gell-Mann and F. E. Low, Phys. Rev. **95**, 1300 (1954).
 [5] N. N. Bogoliubov and D. V. Shirkov, Dokl. Akad. Nauk SSSR **103**, 203 (1955).
 [6] N. N. Bogoliubov and D. V. Shirkov, Nuovo Cimento **3**, 845 (1956).
 [7] R. H. Swendsen, Phys. Rev. Lett. **42**, 859 (1979).
 [8] H. Callen, Physics Letters **4**, 161 (1963), ISSN 0031-9163.
 [9] R. H. Swendsen, Phys. Rev. Lett. **52**, 1165 (1984).
 [10] J. Albert and R. H. Swendsen, Physics Procedia **57**, 99 (2014), ISSN 1875-3892.
 [11] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
 [12] K. G. Wilson and J. B. Kogut, Phys. Rept. **12**, 75 (1974).
 [13] L. P. Kadanoff, Physics **2**, 263 (1966).
 [14] O. Valsson and M. Parrinello, Phys. Rev. Lett. **113**, 090601 (2014).
 [15] F. Bach and E. Moulines, in *Advances in Neural Information Processing Systems 26*, edited by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Curran Associates, Inc., 2013), pp. 773–781.
 [16] Supplementary Material.
 [17] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, The Journal of Chemical Physics **21**, 1087 (1953).
 [18] L. Onsager, Phys. Rev. **65**, 117 (1944).
 [19] R. H. Swendsen, Phys. Rev. B **30**, 3866 (1984).
 [20] H. Flyvbjerg and H. G. Petersen, The Journal of Chemical Physics **91**, 461 (1989).
 [21] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).
 [22] D. P. Landau and R. H. Swendsen, Phys. Rev. Lett. **46**, 1437 (1981).
 [23] M. Nauenberg and B. Nienhuis, Phys. Rev. Lett. **33**, 1598 (1974).
 [24] L. Landau, Zh. Eksp. Teor. Fiz. **7**, 19 (1937).
 [25] V. Ginzburg and L. Landau, Zhurnal Eksp. Yheoret. Physics **29**, 1064 (1950).
 [26] M. Invernizzi, O. Valsson, and M. Parrinello, Proceedings of the National Academy of Sciences **114**, 3370 (2017).
 [27] J. Tobochnik, S. Sarker, and R. Cordery, Phys. Rev. Lett. **46**, 1417 (1981).