Mean-value identities as an opportunity for Monte Carlo error reduction

L. A. Fernandez[1,2] and V. Martin-Mayor[1,2]
[1]*Departamento de Física Teórica I, Universidad Complutense, 28040 Madrid, Spain*
[2]*Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), 50009 Zaragoza, Spain*
(Received 29 January 2009; revised manuscript received 12 March 2009; published 11 May 2009)

In the Monte Carlo simulation of both lattice field theories and of models of statistical mechanics, identities verified by exact mean values, such as Schwinger-Dyson equations, Guerra relations, Callen identities, etc., provide well-known and sensitive tests of thermalization bias as well as checks of pseudo-random-number generators. We point out that they can be further exploited as *control variates* to reduce statistical errors. The strategy is general, very simple, and almost costless in CPU time. The method is demonstrated in the two-dimensional Ising model at criticality, where the CPU gain factor lies between 2 and 4.

DOI: 10.1103/PhysRevE.79.051109 PACS number(s): 05.50.+q, 02.70.−c, 75.40.Mg

I. INTRODUCTION

Monte Carlo simulation [1,2] is one of the handful of general methods in the theoretical physicist toolbox that can be applied to nonperturbative problems. In spite of this, it is a very inefficient method; the computational effort needed to get yet another decimal significant figure grows by a factor of 100.

Yet, there are alternatives to brute force when more accuracy is needed. A classical strategy consists of looking for statistical estimators of the sought quantities, which have the same expectation value as the commonly used naive estimators, but a reduced variance. The *multihit* method [3] (and later developments [4]) for the Polyakov loop in lattice QCD is a conspicuous example of such an improvement. Now, the numerical error is proportional to the square root of variance for the considered estimator. It follows that reducing the variance by a factor of 2 reduces as well in the same factor the numerical effort needed to achieve the desired statistical accuracy. Even a modest factor of variance reduction can be a significant improvement; the CPU time needed in application to lattice gauge theory or to condensed-matter physics (think for instance of spin-glass simulations [5]) often lies in the range $10-10^4$ processor years.

Here we propose a general road to variance reduction based on known identities between exact mean values. In spite of its usefulness, this strategy, known as *control variates* in the mathematical literature [6,7], is still not commonly used in the framework of Monte Carlo simulations in Physics (at the practical level, it only requires standard Monte Carlo data-analysis tools). In fact, it is fairly common to find in field theory or in statistical mechanics that a particular linear combination of nontrivial expectation values vanishes exactly (we provide specific examples below). There are different ways of finding such identities; Schwinger-Dyson equations exploit invariances of the integration measure [8]; Callen identities are derived by integrating in the functional integral some variable while holding all the others fixed [9] (multihit operators [3,4] belong to this category); Guerra relations are somehow specific to disordered systems [10]; in models where a cluster method works [11] cluster estimators with the same expectation value than their spin counterparts can be found (see, e.g., [12] and references therein). It is fair to say that for any problem amenable to a path-integral formulation, each of the above strategies will provide at least one identity—the vanishing of a precise linear combination of expectation values of nontrivial observables.

Researchers performing Monte Carlo simulations are acutely aware of the advantages provided by mean-value identities. If the numerically obtained expectation values do not verify them within errors, this will most probably be due to a thermalization bias [13,14] or to a failure of the used pseudo-random-number generator [15] (or to a programming bug!). We remark here that mean-value identities provide statistical estimators with reduced variance as well. The method is exemplified in the standard benchmark of the two-dimensional Ising model at its critical point.

We note finally that in previous work [16,17] covariance error reduction was presented for the finite-size scaling analysis of phase transitions. Indeed, covariance analysis improves the computation of the critical temperature and the leading scaling-correction exponent from the data on finite lattices [16]. It provides as well the optimal combination of different estimates of the sought critical exponent (each individual estimate being previously extrapolated to infinite volume) [17]. As we discuss in Sec. II D, covariance error reduction (specially as presented in Ref. [17]) is a particular case of the present approach.

The layout of the rest of this paper is as follows. In Sec. II we recall the error-reduction strategy in a general setting (without reference to any specific model). The reader merely interested in a practical recipe may proceed directly to Sec. II C. In Sec. III, we briefly describe the model and the observables, as well as the used mean-value identities. We present our numerical results in Sec. IV while our conclusions are in Sec. V. In the Appendix we present some technical results which are specific for the Swendsen-Wang cluster algorithm as applied to the Ising model.

II. COVARIANCE ERROR REDUCTION

We first discuss the problem as if the exact covariance matrix were accessible (Sec. II A). The effects of time correlations are described in Sec. II B. Real-life complications arise from the fact that the covariance matrix needs to be

estimated from a finite sample of Monte Carlo data, which fortunately does not induce any significant bias, Sec. II C. Finally, we discuss in Sec. II D how the general approach relates with the problem of finding the optimal linear combination of several estimates for the very same expectation value. We discuss some of the very counterintuitive features of this problem as well.

A. Minimal error

Let A, B_1, B_2, …, B_R be stochastic variables. We assume that a set of mean-value identities appropriate for the problem at hand tells us that $\langle B_i\rangle=0$ for $i=1,2,\ldots,R$. We assume as well that $\langle A^2\rangle$ and all the $\langle B_i^2\rangle$ are finite. We wish to profit from the covariance between A and the B_i to obtain the best determination (in the sense of minimal variance) of $\langle A\rangle$.

Before going on, it is useful to note that the operation of computing the covariance between real-valued stochastic variables X and Y,

$$\sigma_{XY} \equiv \langle (X-\langle X\rangle)(Y-\langle Y\rangle)\rangle, \tag{1}$$

has the structure of a scalar product. Indeed the four following properties are easy to establish: (i) it is symmetric, $\sigma_{XY}=\sigma_{YX}$, (ii) it is linear on each of its arguments, $\sigma_{X(\lambda_1 Y_1+\lambda_2 Y_2)}=\lambda_1\sigma_{XY_1}+\lambda_2\sigma_{XY_2}$, (iii) $\sigma_{XX}\geq 0$, and (iv) if $\langle X\rangle=0$ and $\sigma_{XX}=0$ it follows that $X=0$ with probability one. For later use, we introduce the correlation coefficient between X and Y,

$$r_{XY} \equiv \frac{\sigma_{XY}}{\sqrt{\sigma_{XX}\sigma_{YY}}}. \tag{2}$$

Using the B_i, it is straightforward to define stochastic variables with expectation value $\langle A\rangle$,

$$\tilde{A}(\lambda_1,\lambda_2,\ldots,\lambda_R) = A + \sum_{i=1}^{R} \lambda_i B_i. \tag{3}$$

Our task is to find the coefficients $\{\lambda_i^*\}_{i=1}^R$ that minimize the $\tilde{A}$ variance,

$$\sigma_{\tilde{A}\tilde{A}} = \sigma_{AA} + 2\sum_{i=1}^{R} \lambda_i \sigma_{AB_i} + \sum_{i=1}^{R} \lambda_i^2 \sigma_{B_iB_i}, \tag{4}$$

that has a minimum at

$$\lambda_i^* = -\sum_{i,i'=1}^{R} (\Sigma^{-1})_{i,i'}\sigma_{AB_{i'}}, \quad \Sigma_{ii'} = \sigma_{B_iB_{i'}} \tag{5}$$

In the following, we will denote the optimal random variable as

$$A^* = \tilde{A}(\lambda_1^*,\ldots,\lambda_2^*), \tag{6}$$

whose variance is

$$\sigma_{A^*A^*} = \sigma_{AA} - \sum_{i,i'=1}^{R} \sigma_{AB_i}(\Sigma^{-1})_{ii'}\sigma_{AB_{i'}}. \tag{7}$$

Note that rescaling any of the B_i, $B_i\to\alpha_i B_i$, would leave A^* unchanged. For $R=1$, Eq. (7) reads

$$\sigma_{A^*A^*} = \sigma_{AA}(1-r_{AB}^2). \tag{8}$$

In particular, whether A and B are correlated or anticorrelated is immaterial.

In a nutshell, we face a standard problem of best approximation in an Euclidean space; we are decomposing the fluctuating part of A, $A-\langle A\rangle$, on its components parallel and orthogonal with respect to the linear space generated by $\{B_i\}_{i=1}^R$. The best approximation, A^*, is found when the parallel component is made to vanish. The minimal variance is the norm squared of the orthogonal component. If we compute in a Monte Carlo simulation A^* rather than A, we are rewarded with a CPU gain factor of $\sigma_{AA}/\sigma_{A^*A^*}$.

B. Covariance and time correlations

The stochastic variables $X,Y,Z,\ldots$, considered in Sec. II C are actually Monte Carlo time averages. Indeed, the Monte Carlo dynamics can be regarded as a Markovian random walk in configuration space [2]. Let Θ be one of such spin (or gauge-field) configurations, and let $\Theta_{t=0}$, $\Theta_{t=1}$, …, be the time sequence of configurations visited by the random walker. We consider functions of the fields configuration $\mathcal{X},\mathcal{Y},\mathcal{Z},\ldots$ (observables hereafter) and use the shorthand $\mathcal{X}^{(t)}=\mathcal{X}[\Theta(t)]$, $t=0,1,\ldots,T-1$. Hence, our stochastic variable X will be (and similarly for $Y,Z,\ldots$)

$$X = \frac{1}{T}\sum_{t=0}^{T-1} \mathcal{X}^{(t)}. \tag{9}$$

The *Markovian* random walk in configuration space is fully determined by a transition matrix, $\mathcal{P}_{\Theta_{t+1}\Theta_t}$, namely, the conditional probability of reaching Θ_{t+1} from Θ_t in a single step. The transition matrix verifies the *balance* condition with respect to the equilibrium distribution function $\pi[\Theta]$,

$$\pi[\Theta_{t+1}] = \sum_{\Theta_t} \mathcal{P}_{\Theta_{t+1}\Theta_t}\pi[\Theta_t]. \tag{10}$$

In this work, we shall always consider that at $t=0$, equilibrium has been already reached. Thus, the expectation value for X is the Boltzmann average for $\mathcal{X}$.

It is convenient to consider the equilibrium (symmetrized) time-correlation function for two real observables, $\mathcal{X}$ and $\mathcal{Y}$ (*autocorrelation* if $\mathcal{X}=\mathcal{Y}$),

$$C_{\mathcal{X}\mathcal{Y}}(t) = \frac{1}{2}\langle \mathcal{X}^{(0)}\mathcal{Y}^{(t)} + \mathcal{X}^{(t)}\mathcal{Y}^{(0)}\rangle - \langle\mathcal{X}\rangle\langle\mathcal{Y}\rangle. \tag{11}$$

Note that $C_{\mathcal{X}\mathcal{Y}}(t)=C_{\mathcal{Y}\mathcal{X}}(t)=C_{\mathcal{X}\mathcal{Y}}(-t)$ and that it is bilinear in $\mathcal{X}$ and $\mathcal{Y}$. $C_{\mathcal{X}\mathcal{Y}}(0)$ is named *static covariance*, since it can be computed from equal-time expectation values. $C_{\mathcal{X}\mathcal{Y}}(t)$ allows one to compute σ_{XY}, since one straightforwardly obtains from Eq. (9) that

$$\sigma_{XY} = \frac{1}{T^2}\sum_{t,t'=0}^{T-1} C_{\mathcal{X}\mathcal{Y}}(t'-t). \tag{12}$$

We define the integrated correlation time (autocorrelation time, $\tau_{\mathrm{int},\mathcal{X}}$, if $\mathcal{X}=\mathcal{Y}$) as

$$\tau_{\mathrm{int},\mathcal{XY}} = \frac{\sum_{t=-\infty}^{t=\infty} C_{\mathcal{XY}}(t)}{2\sqrt{C_{\mathcal{XX}}(0)C_{\mathcal{YY}}(0)}}. \tag{13}$$

Now, a standard argument [2] tells us that if $\Sigma_{t=1}^{\infty} t|C_{\mathcal{XY}}(t)| < \infty$, the covariance of X and Y is

$$\sigma_{XY} = \frac{2\tau_{\mathrm{int},\mathcal{XY}}\sqrt{C_{\mathcal{XX}}(0)C_{\mathcal{YY}}(0)}}{T} + \mathcal{O}(T^{-2}). \tag{14}$$

For instance, the r_{AB} in Eq. (8) is just

$$r_{AB} = \frac{\tau_{\mathrm{int},\mathcal{AB}}}{\sqrt{\tau_{\mathrm{int},\mathcal{A}}\tau_{\mathrm{int},\mathcal{B}}}}. \tag{15}$$

Hence, the effectiveness of a particular control variate, B, does depend on the autocorrelation and correlation times of the chosen Monte Carlo algorithm [18].

We finally recall some well-known results [2]. $C_{\mathcal{XY}}(t)$ can be computed from the tth power of the transition matrix and the equilibrium distribution as

$$C_{\mathcal{XY}}(t) = \sum_{\Theta_t,\Theta_0} \frac{1}{2}(\mathcal{X}[\Theta_0]\mathcal{Y}[\Theta_t] + \mathcal{X}[\Theta_t]\mathcal{Y}[\Theta_0]) \times ([\mathcal{P}]^{|t|}_{\Theta_t\Theta_0} - \pi[\Theta_t])\pi[\Theta_0]. \tag{16}$$

At this point, an analogy with quantum mechanics is in order. Up to now, we have been working in the Schrödinger picture, where the probabilities evolve in time while the operators remain constant. Yet, it is best to work in an equivalent Heisenberg picture where only observables evolve in time. We define a time transformation, P, that transforms the observable $\mathcal{X}$ in the observable $P\mathcal{X}$. The value taken by $P\mathcal{X}$ for the configuration Θ is a conditional expectation value,

$$P\mathcal{X}[\Theta] = E(\mathcal{X}[\Theta_{t+1}]|\Theta_t = \Theta) = \sum_{\Theta'} \mathcal{X}[\Theta']\mathcal{P}_{\Theta'\Theta}. \tag{17}$$

Mind that if the Monte Carlo dynamics is composed of consecutive steps (in the Swendsen-Wang dynamics, for instance, one first update the bonds, then the spins: $\mathcal{P}^{\mathrm{SW}} = \mathcal{P}_{\mathrm{spin}}\mathcal{P}_{\mathrm{bond}}$), the evolution operators in the Heisenberg picture appear in reversed order (e.g., $P^{\mathrm{SW}} = P_{\mathrm{bond}}P_{\mathrm{spin}}$). We introduce a scalar product for *equal-time* real observables $(\mathcal{X},\mathcal{Y}) \equiv \langle \mathcal{X}(t)\mathcal{Y}(t)\rangle$. The correlation function is

$$C_{\mathcal{XY}}(t) = \frac{(\mathcal{X}, P^{|t|}\mathcal{Y}) + (P^{|t|}\mathcal{X}, \mathcal{Y})}{2} - \langle\mathcal{X}\rangle\langle\mathcal{Y}\rangle. \tag{18}$$

Thus the problem of computing correlation times is reduced to the spectral analysis of the operator P.

C. Practical recipes

In a Monte Carlo calculation, the stochastic variables A and B_i discussed in Sec. II A are directly related to some functions of the spin (or gauge-field) configuration, $\mathcal{A}$ and $\mathcal{B}_i$, where $i=1,2,\ldots,R$. One stores in disk T consecutive measurements of these functions $\{\mathcal{A}^{(t)},\mathcal{B}_1^{(t)},\ldots,\mathcal{B}_R^{(t)}\}_{t=1}^{T}$. We assume that autocorrelation times in Eq. (13) for these measurements are finite. Their Monte Carlo averages

$$\bar{A} = \frac{1}{T}\sum_t \mathcal{A}^{(t)}, \quad \bar{B}_i = \frac{1}{T}\sum_t \mathcal{B}_i^{(t)}, \quad i = 1,2,\ldots,R, \tag{19}$$

are just instances (i.e., disorder realizations) of the random variables A and B_i.

Let us form N *data blocks* $\{A_j, B_{i,j}\}_{j=1}^{N}$ by averaging sets of T/N consecutive measurements $\{\mathcal{A}^{(t)},\mathcal{B}_1^{(t)},\ldots,\mathcal{B}_R^{(t)}\}$. The basic assumption underlying the Monte Carlo error analysis [19] is that, provided that T/N is large enough as compared to Monte Carlo autocorrelation times, the $\{A_j, B_{i,j}\}_{j=1}^{N}$ are identically distributed and statistically independent for different j. Furthermore, one assumes that T/N is so large that the blocked data are not only independent but also Gaussian distributed,

$$A_j = \langle A\rangle + \eta_j^A\sqrt{N\sigma_{AA}},$$

$$B_{i,j} = \eta_j^{B_i}\sqrt{N\sigma_{B_iB_i}}, \quad i = 1,2,\ldots,R. \tag{20}$$

The η are Gaussian random numbers, with zero mean and covariance matrix,

$$\langle \eta_j^A \eta_{j'}^A\rangle = \delta_{jj'},$$

$$\langle \eta_j^A \eta_{j'}^{B_i}\rangle = \delta_{jj'} r_{AB_i},$$

$$\langle \eta_j^{B_i} \eta_{j'}^{B_{i'}}\rangle = \delta_{jj'} r_{B_iB_{i'}}, \tag{21}$$

where δ_{jj} is the Kronecker's delta. Note as well that one gets exactly the same numbers for $\bar{A}$ and $\bar{B}_i$ either by averaging over j the $\{A_j, B_{i,j}\}$ or using Eq. (19). For later use, we define also the jackknife blocks as (see, e.g., [19])

$$A_j^{\mathrm{JK}} = \frac{N\bar{A} - A_j}{N-1},$$

$$B_{i,j}^{\mathrm{JK}} = \frac{N\bar{B}_i - B_{i,j}}{N-1}, \quad i = 1,2,\ldots,R. \tag{22}$$

Our statistical estimators for the covariances are

$$\overline{\sigma_{AA}} = \sum_{j=1}^{N} \frac{(A_j - \bar{A})^2}{N(N-1)} = \sum_{j=1}^{N} \frac{(A_j^{\mathrm{JK}} - \bar{A})^2}{N/(N-1)},$$

$$\overline{\sigma_{AB_i}} = \sum_{j=1}^{N} \frac{(A_j - \bar{A})(B_{i,j} - \bar{B}_i)}{N(N-1)} = \sum_{j=1}^{N} \frac{(A_j^{\mathrm{JK}} - \bar{A})(B_{i,j}^{\mathrm{JK}} - \bar{B}_i)}{N/(N-1)},$$

$$\overline{\sigma_{B_iB_{i'}}} = \sum_{j=1}^{N} \frac{(B_{i,j} - \bar{B}_i)(B_{i',j} - \bar{B}_{i'})}{N(N-1)} = \sum_{j=1}^{N} \frac{(B_{i,j}^{\mathrm{JK}} - \bar{B}_i)(B_{i',j}^{\mathrm{JK}} - \bar{B}_{i'})}{N/(N-1)}. \tag{23}$$

At variance with the *numbers* σ_{AA}, σ_{AB_i}, or $\sigma_{B_iB_j}$, our estimators $\overline{\sigma_{AA}}$, $\overline{\sigma_{AB_i}}$, or $\overline{\sigma_{B_iB_{i'}}}$ are *random variables*. It is straightforward to show that their expectation values are the sought covariances, but they are subject to statistical errors whose (relative) size is of order $1/\sqrt{N}$. In fact, since one needs to

keep the data-block size T/N as large as possible to ensure the correctness of Eq. (20), the typical number of blocks is kept low, say $N \sim 100$. Incidentally, the second equality in each one of Eq. (23) is an algebraic one; we get the same numerical covariance estimates from the standard or the jackknife blocks.

At this point, we may trade the inaccessible minimization Eqs. (5) and (6) by the computable

$$A^* = A - \sum_{i,i'=1}^{R} (\bar{\Sigma}^{-1})_{ii'} \overline{\sigma_{AB_{i'}}} B_i, \quad \bar{\Sigma}_{ii'} = \overline{\sigma_{B_i B_{i'}}}. \tag{24}$$

The very same procedure is performed block by block, thus obtaining $\{A_j^*\}_{j=1}^N$. Errors are computed in a standard way from these blocks.

The reader might question the validity of Eq. (24) because the vanishing of $\langle B_i \rangle$ does not imply $\langle \Sigma_{i'} (\bar{\Sigma}^{-1})_{ii'} \overline{\sigma_{AB_{i'}}} B_i \rangle = 0$. This is specially worrying since, as we said above, the relative errors for σ_{AB_i} or $\sigma_{B_i B_{i'}}$ are $\sim 10\%$ in real-life calculations. The way out is in Eqs. (20) and (21). If in a particular simulation one finds the Gaussian fluctuations $\{\eta_j^A, \eta_j^{B_1}, \ldots, \eta_j^{B_1}\}_{j=1}^N$, the sign-reversed fluctuations $\{-\eta_j^A, -\eta_j^{B_1}, \ldots, -\eta_j^{B_1}\}_{j=1}^N$ are just as probable. One immediately notices that the covariance estimators, Eq. (23), are invariant under sign reversal of fluctuations. This means that $\bar{\sigma}_{AB_i}$, the matrix $\bar{\Sigma}$, and its inverse are also invariant, while the B_i transform to $-B_i$. Hence, if the probability distribution function of $\{\eta_j^A, \eta_j^{B_1}, \ldots, \eta_j^{B_1}\}_{j=1}^N$ is invariant under sign reversal, it follows that the expectation value for A^* in Eq. (24) is still $\langle A \rangle$ (according to Rubinstein [7], this fact was first noticed for the particular case of Gaussian distributed fluctuations in [20]). However, even in the absence of sign-reversal invariance, the bias induced is of order $1/T$ while the statistical error is of order $1/\sqrt{T}$.

As for functions of expectation values, let us explain the procedure by considering the second moment correlation length Eq. (30) that depends on the expectation values of two variables, $m(0)$ and $m(\vec{k}_{\min})$. One first transforms using Eq. (24) the estimates and the jackknife blocks of each of the needed quantities, e.g., $\overline{m^*(0)}$, $\overline{m^*(\vec{k}_{\min})}$, and $\{m_j^{\mathrm{JK},*}(0), m_j^{\mathrm{JK},*}(\vec{k}_{\min})\}_{j=1}^N$. Then we use Eq. (30) to obtain our best estimate of the correlation length from $\overline{m^*(0)}$ and $\overline{m^*(\vec{k}_{\min})}$. To estimate the errors, we first form N jackknife blocks by computing the correlation length from each of the N pairs $\{m_j^{\mathrm{JK},*}(0), m_j^{\mathrm{JK},*}(\vec{k}_{\min})\}$, then use the standard formulas [19].

D. Several observables with the same expectation value

Given a set of random variables $A_1, A_2, \ldots, A_{R+1}$ with a common expectation value, $\langle A_i \rangle = a$, one may wonder how to get the best possible estimate of a. This was precisely the case considered in [16,17]. We only discuss here the relationship with the (closer in spirit) approach of [17], where the A_i were estimates of the critical exponent ν for an Ising model at its critical point. The obvious way of addressing the problem is considering a linear combination

$$\tilde{A}(p_1, p_2, \ldots, p_{R+1}) = \sum_{i=1}^{R+1} p_i A_i, \quad \sum_{i=1}^{R+1} p_i = 1, \tag{25}$$

then minimizing $\sigma_{\tilde{A}\tilde{A}}$. This is a particular case of the optimization problem that we have already discussed at length in Secs. II A and II C. In fact, note that $p_{R+1} = 1 - p_1 - p_2 - \cdots - p_R$ and then, keeping an eye on Eq. (3), we write $A \equiv A_{R+1}$, with $\{\lambda_i = p_i, B_i = A_i - A_{R+1}\}_{i=1}^R$.

However, this optimization problem produced some counterintuitive results [17]. All five computed ν estimates for the two-dimensional Ising model lied *above* the exact value. In spite of this, the improved estimate was *below* the exact value. This apparent paradox can be easily explained in our language by considering the simpler case $R=1$ (so we have A_1 and A_2). Using the results reviewed in Sec. II A one easily finds that the minimal squared error is

$$\sigma_{A^*A^*} = \frac{\sigma_{A_1A_1}\sigma_{A_2A_2}(1 - r_{A_1A_2}^2)}{\sigma_{A_1A_1} + \sigma_{A_2A_2} - 2r_{A_1A_2}\sqrt{\sigma_{A_1A_1}\sigma_{A_2A_2}}}. \tag{26}$$

Hence, if $r_{A_1A_2}$ tends to one *and* if $\sigma_{A_1A_1} \neq \sigma_{A_2A_2}$ an errorless estimator exists. In fact, in the $r_{A_1A_2} \to 1$ limit we have $A_1 = a + \eta\sqrt{\sigma_{A_1A_1}}$ and $A_2 = a + \eta\sqrt{\sigma_{A_2A_2}}$ with η the *same* Gaussian random number for both variables (of course $\langle \eta \rangle = 0$ and $\langle \eta^2 \rangle = 1$). In other words, if for a particular simulation A_1 lies below (above) a, the same will be true for A_2. In spite of this, if we write $pA_1 + (1-p)A_2 = a + \eta[\sqrt{\sigma_{A_2A_2}} + p(\sqrt{\sigma_{A_2A_2}} - \sqrt{\sigma_{A_1A_1}})]$ and set $p = \sqrt{\sigma_{A_2A_2}}/(\sqrt{\sigma_{A_1A_1}} - \sqrt{\sigma_{A_2A_2}})$, an exact answer is found. Note, however, that the problem becomes ill conditioned when $\sigma_{A_1A_1}$ approaches $\sigma_{A_2A_2}$. In fact, if the two variances coincide we gain nothing by considering A_2 in addition to A_1, since in this case one would have $A_1 = A_2$ with probability 1.

III. MODEL, OBSERVABLES, AND MEAN-VALUE IDENTITIES

We shall put to work the strategy in Sec. II, in the standard benchmark, the Ising model in two dimensions, for which many exact results exist, including exact computations of some quantities in *finite systems* [21] that can be directly confronted with the Monte Carlo simulation. The spins $\mathcal{S}_{\vec{x}}$ are placed in the nodes of a square lattice of side L with periodic boundary conditions. The interaction is restricted to lattice nearest neighbors, with the partition function being ($\Sigma_{\{\mathcal{S}_{\vec{x}}\}}$: summation over the 2^{L^2} spin configurations)

$$Z = \sum_{\{\mathcal{S}_{\vec{x}}\}} \exp\left[\kappa \sum_{\|\vec{x}-\vec{y}\|=1} \mathcal{S}_{\vec{x}}\mathcal{S}_{\vec{y}}\right]. \tag{27}$$

The system undergoes a second-order phase transition at $\kappa_c = \log(1+\sqrt{2})/2$.

The main functions of the spins that we are considering are the energy and the Fourier transform of the spin field at zero and minimal momenta [$\vec{k} = (0,0)$ or $\vec{k}_{\min} = (2\pi/L, 0)$]

$$e = \frac{1}{L^2} \sum_{\|\vec{x}-\vec{y}\|=1} \mathcal{S}_{\vec{x}}\mathcal{S}_{\vec{y}}, \quad m(\vec{k}) = \frac{1}{L^2} \sum_{\vec{x}} \mathcal{S}_{\vec{x}} e^{i\vec{k}\cdot\vec{x}}. \tag{28}$$

From $m(\vec{k})$ we define the magnetic susceptibility

$$\chi = L^2 \langle [m(0)]^2 \rangle, \tag{29}$$

the second moment correlation length [22] (we gain statistics by averaging $[m(\vec{k}_{\min})]^2$ over $(2\pi/L,0)$ and $(0,2\pi/L)$)

$$\xi = \sqrt{\frac{\langle [m(0)]^2 \rangle - \langle [m(\vec{k}_{\min})]^2 \rangle}{4 \sin^2 \frac{\pi}{L} \langle [m(\vec{k}_{\min})]^2 \rangle}}, \tag{30}$$

and the renormalization-group invariant ratio

$$U_4 = \frac{\langle [m(0)]^4 \rangle}{\langle [m(0)]^2 \rangle^2}. \tag{31}$$

Our first mean-value identity comes from the Fortuin-Kasteleyn formulation (see, e.g., [2,19] for details). Given a decomposition of the lattice in $\mathcal{N}$ connected components (*clusters*), each containing n_c spins, it is easy to show that (see the Appendix for a quick review)

$$\chi = \frac{1}{L^2} \left\langle \sum_c n_c^2 \right\rangle. \tag{32}$$

Hence, our first control variate is

$$\mathcal{B}_{\mathrm{SW}} = [m(0)]^2 - \sum_c \frac{n_c^2}{L^4}. \tag{33}$$

A second control variate comes from a Callen identity [9]. Let the local field acting over site $\vec{x}$ be

$$h_{\vec{x}} = \sum_{\|\vec{x}-\vec{y}\|=1} \mathcal{S}_{\vec{y}}. \tag{34}$$

Then, if $\|\vec{x}-\vec{y}\| > 1$,

$$\langle \mathcal{S}_{\vec{x}} \mathcal{S}_{\vec{y}} \rangle = \langle \tanh(\kappa h_{\vec{x}}) \tanh(\kappa h_{\vec{y}}) \rangle. \tag{35}$$

Hence,

$$\mathcal{B}_{\mathrm{CI}} = \frac{1}{L^4} \sum_{\|\vec{x}-\vec{y}\|>1} [\tanh(\kappa h_{\vec{x}}) \tanh(\kappa h_{\vec{y}}) - \mathcal{S}_x \mathcal{S}_y], \tag{36}$$

which can be computed with $\mathcal{O}(L^2)$ operations as

$$\mathcal{B}_{\mathrm{CI}} = \frac{1}{L^4} \left[\left(\sum_{\vec{x}} \tanh(\kappa h_{\vec{x}}) \right)^2 - \left(\sum_{\vec{x}} \mathcal{S}_{\vec{x}} \right)^2 - \sum_{\vec{x}} [\tanh^2(\kappa h_{\vec{x}}) - 1] - \sum_{\|\vec{x}-\vec{y}\|=1} [\tanh(\kappa h_{\vec{x}}) \tanh(\kappa h_{\vec{y}}) - \mathcal{S}_{\vec{x}} \mathcal{S}_{\vec{y}}] \right]. \tag{37}$$

Finally, a Schwinger-Dyson equation [15] provides a third control variate,

$$\mathcal{B}_{\mathrm{SD}} = 1 - \frac{1}{L^2} \sum_{\vec{x}} e^{-2\kappa h_{\vec{x}}}. \tag{38}$$

IV. RESULTS

We have simulated the model on systems $L=16$, 128, and 512 using the Swendsen-Wang algorithm [11]. For each lattice size, we traced clusters 10^6 times taking measurements each time that the clusters were traced. We discarded the first 10% of measurements for thermalization (which, on the view of the autocorrelation times for this model and algorithm [19], is extremely conservative), hence formed $N=100$ data blocks of 9000 measurements each (we expect to be well in the Gaussian fluctuations regime). The jackknife error was used throughout for error computations. The used programs were minor modifications of the sample programs in [19].

As in Sec. II C, we name B_i $(i=\mathrm{SW},\mathrm{CI},\mathrm{SD})$ the block average of consecutive Monte Carlo measurements of $\mathcal{B}_i$. The results of the analysis using B_{SW} and/or B_{CI} as control variates are shown in Table I. We detect no bias when comparing with exact results or with previously published (and more precise) computations. When using the two control variates together, a CPU factor gain larger than 2 is achieved for χ, $\langle e \rangle$, and ξ for all values of L. This gain is largest for $L=16$ and deteriorates somewhat in going to $L=128$, but then stabilizes and does not significantly deteriorate further when going to $L=512$. For instance, for $L=512$ the CPU gain in the computation of the susceptibility is a factor 2.3 when comparing with the standard spin estimate [Eq. (29)] or a factor 2.0 when comparing with the cluster estimate [Eq. (32)].

Rather smaller gains are obtained by using B_{SW} and/or B_{CI} individually; for instance, in the χ computation using B_{SW} alone, the CPU gain factor is 2.8 for $L=16$, but it deteriorates to 1.56 for $L=128$ and 1.42 for $L=512$. The fact that we do significantly better by combining the two control variates (rather than using only one of them) suggests that the orthogonal component of B_{CI} with respect to B_{SW} is sizeable (and that this component still strongly correlates with the squared magnetization).

There are some interesting issues regarding the usefulness of B_{SW} as a control variate for χ. This is an instance of the problem considered in Sec. II D; we are after the optimal linear combination between Eqs. (29) and (32). In the Appendix we show that the optimal choice is very close to the cluster-based susceptibility, Eq. (32) [the optimum is exactly Eq. (32) if successive measurements are separated by a time interval of many autocorrelation times, so that they are essentially statistically independent]. This statement can be reworded as the use of the spin-based susceptibility via the control variate B_{SW} barely improves the cluster-based susceptibility [the usefulness of B_{SW} decreases with growing autocorrelation times, Eq. (A17)].

A related, yet different, issue is the temperature evolution of the efficiency of the cluster estimator for χ. At κ_c, Table I, errors for the spin- and cluster-based estimates are similar. This is in marked contrast with the situation in the paramagnetic scaling region ($\kappa < \kappa_c$, $1 \ll \xi \ll L$), see, e.g., [23]. In Eq. (A16), we give the (squared) ratio of statistical errors for the two estimators in terms of an autocorrelation time and of several expectation values of the *static* cluster-size distribution. At κ_c, a giant cluster dominates sums such as that in Eq. (32), see Table II in the Appendix. As a consequence, the squared error ratio at κ_c, Eq. (A16), is $\sim 1 + \frac{1.15}{\tau_{\mathrm{int},C}}$, never very large since $\tau_{\mathrm{int},C} \geq 1/2$, and decreasing with growing L due to critical slowing down. On the other hand, in the scaling region the largest cluster is not dramatically large, and a major

TABLE I. Comparison of numerical results for the quantities defined in Eqs. (28)–(31), namely, the internal energy, the magnetic susceptibility, the correlation length, and the dimensionless ratio U_4, as obtained in the two-dimensional Ising model at its critical point, for different lattice sizes. For the susceptibility we show also the cluster estimate, Eq. (32), that improves less than a 20% in terms of CPU time over the standard estimator Eq. (29). In contrast, the covariance improved estimates obtained from the mean-value identities [Eqs. (33) and (36)] do save more than a factor of 2 in computer cost. To check for the possibility of a bias induced by the covariance error reduction, we compare also with exact results (for the internal energy) or with independent and longer Monte Carlo simulations.

L	$\langle e \rangle$	χ	ξ	U_4
16				
Standard	1.45339(47)	139.719(155)	14.601(31)	1.16502(74)
Cluster		139.713(127)		
B_{SW} improved	1.45334(32)	139.700(93)	14.597(19)	1.16510(51)
B_{CI} improved	1.45316(24)	139.652(104)	14.590(25)	1.16524(63)
B_{SW} and B_{CI} improved	1.45319(18)	139.666(73)	14.594(18)	1.16517(50)
Others	1.453065… [21]	139.546(77) [23]	14.566(14) [23]	1.16586(34) [12]
128				
Standard	1.419052(100)	5316.6(76)	115.77(28)	1.16789(89)
Cluster		5317.7(70)		
B_{SW} improved	1.419101(94)	5321.7(60)	115.97(21)	1.16735(75)
B_{CI} improved	1.419047(79)	5316.4(68)	115.77(26)	1.16791(86)
B_{SW} and B_{CI} improved	1.419095(66)	5321.4(51)	115.96(19)	1.16736(71)
Others	1.419076… [21]	5318.1(28) [23]	115.81(13) [23]	1.16763(32) [12]
512				
Standard	1.415407(36)	60180(94)	463.62(115)	1.16809(89)
Cluster		60168(88)		
B_{SW} improved	1.415397(34)	60134(80)	462.99(92)	1.16852(76)
B_{CI} improved	1.415429(26)	60230(78)	464.14(101)	1.16768(78)
B_{SW} and B_{CI} improved	1.415421(24)	60183(62)	463.51(76)	1.16812(64)
Others	1.415429… [21]	60209(34) [23]	463.82(51) [23]	1.16782(30) [12]

(static) variance reduction is achieved by averaging over the sign of the different clusters at a fixed time. This gain is at the level of a single measurement. Yet, Eq. (A16), the benefits remain after that the Monte Carlo time averaging.

As for the benefits of including B_{SD} in the covariance reduction procedure, they are marginal at the critical point (the CPU gained when adding B_{SD} to $\{B_{SW}, B_{CI}\}$ is less than a 10%). Nevertheless, in the scaling region it can pay to consider B_{SD}. For instance, in a $L=512$ lattice at $\kappa=0.42$, where $\xi \sim 12$, we obtain a CPU gain factor of 1.23 for the cluster estimator of the susceptibility and 1.6 factor for the energy.

V. CONCLUSIONS

For any problem amenable to a path-integral formulation there are well-known strategies (Schwinger-Dyson [8], Callen [9], etc.) to obtain identities, which imply the vanishing of a precise linear combination of expectation values of nontrivial observables. More often than not, researchers performing Monte Carlo simulations compute the quantities appearing in the identities, since the extra CPU costs are negligible and the identities provide important consistency tests. In particular, they allow one to easily detect problems as frightening as programming bugs, failure of the used pseudo-random-number generator, or thermalization bias. What we have pointed out here is that using the general and simple control variates strategy [6,7], these identities provide as well a significant error reduction in the final outcome of Monte Carlo simulations. This comes at a negligible CPU cost. The method has been exemplified in the standard benchmark, the two-dimensional Ising model at criticality.

We note nevertheless that less trivial applications of this technique already exist. In particular, we have found that a Schwinger-Dyson equation, providing a now standard thermalization test in spin-glass simulations [14], can gain an error-reduction factor of one half on some final quantities (e.g., the correlation length) [24].

ACKNOWLEDGMENT

We acknowledge partial financial support from Ministerio de Ciencia e Innovación (Spain) through research Contract No. FIS2006-08533.

APPENDIX: ON CLUSTER ESTIMATORS

We will answer here two related questions. (1) Why do the control variate B_{SW} improve so little the cluster estimate

TABLE II. Numerical determinations for different lattice sizes, both at κ_c and at $\kappa=0.42$ (where $\xi\sim 12$ for large L), of the dimensionless ratios $g_{\mathcal{C}}$ and $g_{\mathcal{S}}$, Eq. (A12), and the cluster-estimator merit number R, Eq. (A11), recall also Eq. (A7). Note that at $\kappa=0.42$ the advantages of using a cluster estimator grows fast with L, while it remains fairly modest at κ_c. We show as well the product $\langle n_1 L^{-(D+2-\eta)/2}\rangle$ at κ_c, where n_1 is the largest cluster, $D=2$, and $\eta=1/4$ is the anomalous dimension. We see that n_1^2 scales as the full sum $\Sigma_c n_c^2$ (indeed $\langle n_1\rangle^2<\langle n_1^2\rangle<\langle\Sigma_c n_c^2\rangle=L^D\chi\propto L^{D+2-\eta}$). On the contrary, at $\kappa=0.42$, n_1 grows only mildly with L. We can also compare for both κ the average ratio of the sizes of the second largest to the largest cluster (n_2/n_1), and that of the third largest to the largest (n_3/n_1). While at κ_c there is a L-invariant hierarchical structure $n_1\sim 8n_2\sim 16n_3\ldots$, at $\kappa=0.42$ the largest cluster becomes a typical one with growing L.

$\kappa=\kappa_c$, $y=(D+2-\eta)/2$					
L	$g_{\mathcal{C}}$	R	$\langle n_1\rangle L^{-y}$	$\langle n_2/n_1\rangle$	$\langle n_3/n_1\rangle$
16	0.11590(43)	1.19371(35)	1.00701(60)	0.12953(41)	0.06484(21)
128	0.12440(62)	1.16125(51)	1.00687(84)	0.12528(49)	0.06180(27)
512	0.12572(62)	1.15683(41)	1.00683(93)	0.12485(52)	0.06166(29)
$\kappa=0.42$, $y=0$					
16	0.28942(73)	1.22644(53)	137.68(12)	0.27402(53)	0.15343(34)
128	0.16134(80)	3.4069(80)	1016.78(66)	0.72070(22)	0.58116(24)
512	0.010212(22)	13.967(15)	1963.27(60)	0.82843(14)	0.73909(15)

of the susceptibility [Eq. (32)]? (2) Why, at the critical point, do the susceptibility cluster estimator barely improve over the spin one [Eq. (29)]? Both questions are specific to the Swendsen-Wang dynamics for ferromagnetic systems [25].

Under a simplifying assumption, question (1) is addressed in Appendix A, Sec. 1, while question (2) is considered in Appendix A, Sec. 2. The assumption is that successive measurements are separated by a time interval of many autocorrelation times, so that they are essentially statistically independent. The assumption is removed in Appendix A, Sec. 3 (largely inspired in Ref. [26]). Yet, the static variance ratio computed in Appendix A, Sec. 2 still plays a prominent role in the general case.

1. B_{SW} for independent measurements

For independent measurements, time-correlation functions, Eq. (11), vanish for all times $t\neq 0$. Hence, we only need to compute a static covariance.

Let $\mathcal{M}\equiv L^2 m(0)$ be the extensive magnetization (recall Sec. III). At time t in the Swendsen-Wang dynamics, the lattice will be decomposed in $\mathcal{N}_t$ connected components, of size n_c^t with $c=1,2,\ldots,\mathcal{N}_t$ (the ordering is such that $n_1^t\geq n_2^t\geq n_3^t\cdots$). All the spins belonging to cluster c are given a common sign, $\mathcal{S}_c^t$. The value $\mathcal{S}_c^t=\pm 1$ is chosen with 50% probability, independently for each cluster c [25].

The spin estimator for $\mathcal{M}^2$ is

$$\mathcal{M}_t^2=\sum_{c,c'=1}^{\mathcal{N}_t} n_c^t n_{c'}^t \mathcal{S}_c^t \mathcal{S}_{c'}^t. \tag{A1}$$

On the other hand, if one averages Eq. (A1) over the $2^{\mathcal{N}_t}$ equivalent choices for the $\mathcal{S}_c^t=\pm 1$, only the diagonal terms $c=c'$ survive. Hence, the natural cluster estimator for $\langle\mathcal{M}^2\rangle$ is the Monte Carlo average of

$$\mathcal{C}_t=\sum_{c=1}^{\mathcal{N}_t}(n_c^t)^2. \tag{A2}$$

It is illuminating to write Eq. (A1) as

$$\mathcal{M}_t^2=\langle\mathcal{M}^2\rangle+\eta_{\mathcal{C}}^t+\eta_{\mathcal{S}}^t, \tag{A3}$$

$$\eta_{\mathcal{C}}^t=\mathcal{C}_t-\langle\mathcal{M}^2\rangle, \tag{A4}$$

$$\eta_{\mathcal{S}}^t=\sum_{c\neq c'} n_c^t n_{c'}^t \mathcal{S}_c^t \mathcal{S}_{c'}^t, \quad \mathcal{B}_{\mathrm{SW},t}=\frac{\eta_{\mathcal{S}}^t}{L^4}. \tag{A5}$$

Of course, $\langle\eta_{\mathcal{C}}^t\rangle=\langle\eta_{\mathcal{S}}^t\rangle=0$, but the statistical independence of $\mathcal{S}_c^t$ also implies $\langle\eta_{\mathcal{C}}^t\eta_{\mathcal{S}}^t\rangle=0$. Therefore,

$$\sigma_{\mathcal{C}\mathcal{C}}=\langle\eta_{\mathcal{C}}^2\rangle, \quad \sigma_{\mathcal{M}^2\mathcal{M}^2}=\langle\eta_{\mathcal{C}}^2\rangle+\langle\eta_{\mathcal{S}}^2\rangle. \tag{A6}$$

Let us try to improve $\mathcal{C}$ using $\mathcal{B}_{\mathrm{SW}}$ as control variate. We find $C_{\mathcal{C}\mathcal{B}_{\mathrm{SW}}}(0)=\langle\eta_{\mathcal{C}}^t\eta_{\mathcal{S}}^t\rangle/L^4=0$. It follows that the improved estimator $\mathcal{C}^*$ obtained using $\mathcal{B}_{\mathrm{SW}}$ as control variate is just $\mathcal{C}$. Using the language of Sec. II D, with no time correlations, the optimal linear combination between $L^2[m(0)]^2$ and $L^{-2}\Sigma_c n_c^2$ is just $L^{-2}\Sigma_c n_c^2$.

2. Static variance

Under the independent measurement assumption, the (squared) error ratio for the spin [Eq. (29)] and cluster [Eq. (32)] susceptibility estimators equals the static variance ratio

$$R^2=\frac{C_{\mathcal{M}^2\mathcal{M}^2}(0)}{C_{\mathcal{C}\mathcal{C}}(0)}. \tag{A7}$$

To relate R^2 with the cluster size distribution, we start from Eq. (A3) and a trivial relation between $C_{\mathcal{M}^2\mathcal{M}^2}(0)$ and the dimensionless ratio U_4, Eq. (31),

$$\frac{C_{\mathcal{M}^2\mathcal{M}^2}(0)}{\langle \mathcal{M}^2\rangle^2} = \frac{\langle \mathcal{M}^4\rangle - \langle \mathcal{M}^2\rangle^2}{\langle \mathcal{M}^2\rangle^2} = U_4 - 1. \tag{A8}$$

Hence, in the scaling region, where $U_4 \approx 3$, the spin estimator will be remarkably noisier than at κ_c, see Table I.

The covariance matrix for the $\eta_{\mathcal{S}}$, $\eta_{\mathcal{C}}$ can be expressed in terms of the n_c [27],

$$\langle \eta_{\mathcal{C}}^2\rangle = \left\langle \left(\sum_c n_c^2\right)^2 \right\rangle - \left\langle \sum_c n_c^2 \right\rangle^2, \tag{A9}$$

$$\langle \eta_{\mathcal{S}}^2\rangle = 2\left\langle \left(\sum_c n_c^2\right)^2 - \sum_c n_c^4 \right\rangle, \tag{A10}$$

so that

$$R = \sqrt{1 + \frac{\langle \eta_{\mathcal{S}}^2\rangle}{\langle \eta_{\mathcal{C}}^2\rangle}}. \tag{A11}$$

Introducing the dimensionless ratios

$$g_{\mathcal{C}} = \frac{\langle \eta_{\mathcal{C}}^2\rangle}{\langle \mathcal{M}^2\rangle}, \quad g_{\mathcal{S}} = \frac{\langle \eta_{\mathcal{S}}^2\rangle}{\langle \mathcal{M}^2\rangle}, \tag{A12}$$

we note that

$$U_4 - 1 = g_{\mathcal{C}} + g_{\mathcal{S}}, \quad R = \sqrt{1 + \frac{g_{\mathcal{S}}}{g_{\mathcal{C}}}}. \tag{A13}$$

Now, in the paramagnetic scaling region ($1 \ll \xi \ll L$) the thermodynamic limit of $g_{\mathcal{S}}$ is 2. Indeed, the two terms in the difference $\langle \eta_{\mathcal{S}}^2\rangle = 2\langle (\Sigma_c n_c^2)^2 - \Sigma_c n_c^4\rangle$ scale differently; when $\xi \ll L$ the first grows as the system volume *squared*, while the second scales linearly with the volume [28]. As a consequence, R *diverges* if one takes the large-L limit at fixed $\kappa < \kappa_c$. Since the susceptibility $\chi = \langle \mathcal{M}^2/L^2\rangle$ remains finite for large L, the error incurred when estimating the susceptibility from a *single* measurement, $\mathcal{C}_t$, vanishes in the large-L limit. Quite on the contrary, Eq. (A13), considered precisely at κ_c, strongly suggests that both $g_{\mathcal{C}}$ and $g_{\mathcal{S}}$ have a finite, nonvanishing, large-L limit (and hence a finite R).

We display in Table II our results for $g_{\mathcal{C}}$ and U_4 both at the critical point and at $\kappa = 0.42$, where $\xi \approx 12$. Indeed $R(\kappa_c) \sim 1.15$ remains bound. As we show in Table II, the average ratios n_2/n_1 and n_3/n_1 at κ_c are surprisingly small and size independent. In other words, the two sums in Eq. (A10) are dominated by n_1, causing a massive cancellation that diminish $g_{\mathcal{S}}$ as compared to $g_{\mathcal{C}}$.

3. Monte Carlo time correlations

We now drop the assumption of independent measurements. The (squared) ratio of the errors of the spin and susceptibility estimators is no longer R^2, Eq. (A7), but

$$\widetilde{R}^2 = \frac{\sum_{t=-\infty}^{t=\infty} C_{\mathcal{M}^2\mathcal{M}^2}(t)}{\sum_{t=-\infty}^{t=\infty} C_{\mathcal{C}\mathcal{C}}(t)}. \tag{A14}$$

Similarly, Eq. (8), the efficiency of B_{SW} as a control variate to improve the cluster susceptibility estimator is ruled by the correlation coefficient

$$r_{\mathcal{C}\mathcal{B}_{\mathrm{SW}}} = \frac{\sum_{t=-\infty}^{t=\infty} C_{\mathcal{C}\mathcal{B}_{\mathrm{SW}}}(t)}{\left[\sum_{t=-\infty}^{t=\infty} C_{\mathcal{C}\mathcal{C}}(t)\right]^{1/2}\left[\sum_{t=-\infty}^{t=\infty} C_{\mathcal{B}_{\mathrm{SW}}\mathcal{B}_{\mathrm{SW}}}(t)\right]^{1/2}}. \tag{A15}$$

Arguing as in Ref. [26] will lead us to our main result,

$$\widetilde{R}^2 = 1 + \frac{1}{2\tau_{\mathrm{int},\mathcal{C}}}[R^2 + 1], \tag{A16}$$

$$r_{\mathcal{C}\mathcal{B}_{\mathrm{SW}}} = \frac{1}{[2\tau_{\mathrm{int},\mathcal{C}}(R^2 - 1)]^{1/2}}. \tag{A17}$$

Since $R^2(\kappa_c) \sim 1.3$, the efficiency of the cluster estimator at κ_c is ruled by $\tau_{\mathrm{int},\mathcal{C}}$. Indeed, the (mild) critical slowing down can be traced in Table I. The usefulness of B_{SW} as a control variate, Eq. (A17), deteriorates as well with growing $\tau_{\mathrm{int},\mathcal{C}}$.

On the other hand, in the paramagnetic scaling region ($\kappa < \kappa_c$, $1 \ll \xi \ll L$), one easily has $\widetilde{R}^2 \sim 100$ or larger. Given Eq. (A16), and since $\tau_{\mathrm{int},\mathcal{C}} \geq 1/2$ [because $C_{\mathcal{C}\mathcal{C}}(t) > 0$, see below], this is due to the large R^2 that are to be expected, recall Appendix A, Sec. 2 (we expect $\tau_{\mathrm{int},\mathcal{C}}$ to be upper bounded in the large-L limit, for $\kappa < \kappa_c$). However, Eq. (A17), in the scaling region, B_{SW} behaves poorly as a control variate, since $\tau_{\mathrm{int},\mathcal{C}}$ is lower bounded while R^2 diverges in the large-L limit.

To derive Eqs. (A16) and (A17) we first note that (in space dimension D)

$$L^{2D} C_{\mathcal{C}\mathcal{B}_{\mathrm{SW}}}(t) = C_{\mathcal{C}\mathcal{M}^2}(t) - C_{\mathcal{C}\mathcal{C}}(t), \tag{A18}$$

$$L^{4D} C_{\mathcal{B}_{\mathrm{SW}}\mathcal{B}_{\mathrm{SW}}}(t) = C_{\mathcal{M}^2\mathcal{M}^2}(t) + C_{\mathcal{C}\mathcal{C}}(t) - 2C_{\mathcal{C}\mathcal{M}^2}(t). \tag{A19}$$

Equation (18) suggests that it will be fruitful to recall the main properties of the operator $P^{\mathrm{SW}} = P_{\mathrm{bond}} P_{\mathrm{spin}}$. The two operators P_{bond} and P_{spin} are of heat-bath type, and their action is quite simple [25]; for any observable O, $P_{\mathrm{spin}} O = E(O|\{b\})$ and $P_{\mathrm{bond}} O = E(O|\{\mathcal{S}\})$. In particular, we have

$$P_{\mathrm{spin}}\mathcal{M}^2 = \mathcal{C}, \quad P_{\mathrm{spin}}\mathcal{C} = \mathcal{C}, \quad P_{\mathrm{bond}}\mathcal{M}^2 = \mathcal{M}^2. \tag{A20}$$

All heat-bath operators, P^{HB}, share some nice features; they are self-adjoint, $(O_1, P^{\mathrm{HB}} O_2) = (P^{\mathrm{HB}} O_1, O_2)$ and idempotent $[P^{\mathrm{HB}}]^2 = P^{\mathrm{HB}}$. Furthermore, they preserve expectation values $\langle O\rangle = \langle P^{\mathrm{HB}} O\rangle$ [29].

Combining $P^{\mathrm{SW}} = P_{\mathrm{bond}} P_{\mathrm{spin}}$ with $[P_{\mathrm{spin}}]^2 = P_{\mathrm{spin}}$ (hence $[P_{\mathrm{bond}} P_{\mathrm{spin}}]^{t>0} = [P^{\mathrm{SW}}]^t P_{\mathrm{spin}}$) and with the self-adjointedness of P_{spin} and P_{bond}, we get for $t > 0$

$$\begin{aligned}(\mathcal{M}^2, [P^{\mathrm{SW}}]^t \mathcal{M}^2) &= (\mathcal{M}^2, [P^{\mathrm{SW}}]^t P_{\mathrm{spin}} \mathcal{M}^2),\\ &= (\mathcal{M}^2, [P^{\mathrm{SW}}]^t \mathcal{C}),\\ &= (P_{\mathrm{spin}} P_{\mathrm{bond}} \mathcal{M}^2, [P^{\mathrm{SW}}]^{t-1} \mathcal{C}),\\ &= (\mathcal{C}, [P^{\mathrm{SW}}]^{t-1} \mathcal{C}),\end{aligned} \tag{A21}$$

$$(\mathcal{M}^2,[P^{\mathrm{SW}}]^t\mathcal{C}) = (\mathcal{C},[P^{\mathrm{SW}}]^{t-1}\mathcal{C}), \tag{A22}$$

$$(\mathcal{C},[P^{\mathrm{SW}}]^t\mathcal{M}^2) = (\mathcal{C},[P^{\mathrm{SW}}]^t\mathcal{C}). \tag{A23}$$

Now, Eqs. (A21)–(A23) tell us that ($\delta_{t,0}$ stands for Kronecker's delta, we assume $t>0$)

$$C_{\mathcal{M}^2\mathcal{M}^2}(t) = \delta_{t,0}C_{\mathcal{M}^2\mathcal{M}^2}(0) + (1-\delta_{t,0})C_{\mathcal{C}\mathcal{C}}(t-1), \tag{A24}$$

$$C_{\mathcal{C}\mathcal{M}^2}(t) = \delta_{t,0}C_{\mathcal{C}\mathcal{C}}(0) + (1-\delta_{t,0})\frac{C_{\mathcal{C}\mathcal{C}}(t-1)+C_{\mathcal{C}\mathcal{C}}(t)}{2}. \tag{A25}$$

Deriving at this point, Eqs. (A16) and (A17), is straightforward.

We note, finally, that

$$(\mathcal{C},[P^{\mathrm{SW}}]^t\mathcal{C}) = (\mathcal{C},[P_{\mathrm{spin}}P_{\mathrm{bond}}P_{\mathrm{spin}}]^t\mathcal{C}), \tag{A26}$$

which implies that $C_{\mathcal{C}\mathcal{C}}(t)>0$, and hence $\tau_{\mathrm{int},\mathcal{C}} \geq 1/2$.

[1] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).

[2] A. D. Sokal, in *Functional Integration: Basics and Applications (1996 Cargèse School)*, edited by C. DeWitt-Morette, P. Cartier, and A. Folacci (Plenum, New York, 1997).

[3] G. Parisi, R. Petronzio, and F. Rapuano, Phys. Lett. B **128**, 418 (1983).

[4] M. Lüscher and P. Weisz, J. High Energy Phys. 09 (2001) 010.

[5] See, e.g., F. Belletti *et al.* (the Janus Collaboration), Phys. Rev. Lett. **101**, 157201 (2008).

[6] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Fletcher and Son, Methuen, London, 1964).

[7] R. Y. Rubinstein, *Simulation and the Monte Carlo Method* (Wiley and Sons, New York, 1981).

[8] J. S. Schwinger, Proc. Natl. Acad. Sci. U.S.A. **37**, 452 (1951); F. J. Dyson, Phys. Rev. **75**, 1736 (1949).

[9] H. B. Callen, Phys. Lett. **4**, 161 (1963).

[10] F. Guerra, Int. J. Mod. Phys. B **10**, 1675 (1996).

[11] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[12] J. Salas and A. D. Sokal, J. Stat. Phys. **98**, 551 (2000).

[13] E. Marinari, G. Parisi, and J. J. Ruiz-Lorenzo, Phys. Rev. B **58**, 14852 (1998).

[14] H. G. Katzgraber, M. Palassini, and A. P. Young, Phys. Rev. B **63**, 184422 (2001).

[15] See, e.g., H. G. Ballesteros and V. Martin-Mayor, Phys. Rev. E **58**, 6787 (1998), and references therein.

[16] H. G. Ballesteros, L. A. Fernandez, V. Martin-Mayor, and A. Munoz Sudupe, Phys. Lett. B **387**, 125 (1996); H. G. Ballesteros, L. A. Fernandez, V. Martin-Mayor, A. Munoz Sudupe, G. Parisi, and J. J. Ruiz-Lorenzo, Phys. Rev. B **58**, 2740 (1998).

[17] M. Weigel and W. Janke, Phys. Rev. Lett. **102**, 100601 (2009).

[18] For disordered systems, the time average in Eq. (9) is followed by a disorder average, which strongly diminishes the influence of the particular Monte Carlo dynamics.

[19] D. Amit and V. Martin-Mayor, *Field Theory: The Renormalization Group and Critical Phenomena*, 3rd ed. (World-Scientific, Singapore, 2005).

[20] S. S. Lavenberg and P. D. Welch, IBM Corporation, Yorktown Heights, New York Research Report No. RC8161, 1980 (unpublished).

[21] A. E. Ferdinand and M. E. Fischer, Phys. Rev. **185**, 832 (1969).

[22] F. Cooper, B. Freedman, and D. Preston, Nucl. Phys. B **210**, 210 (1982).

[23] L. A. Fernandez, V. Martin-Mayor, and D. Yllanes, Nucl. Phys. B **807**, 424 (2009).

[24] L. A. Fernandez, V. Martin-Mayor, S. Perez-Gaviro, A. Tarancon, and A. P. Young e-print arXiv:0905.0322.

[25] We shall need to recall the main features of the Swendsen-Wang dynamics, as formulated in Ref. [30]. The configuration space is expanded; to the original spins $\{\mathcal{S}_{\vec{x}}\}$ in the lattice sites, we add bond variables $b_{\vec{x}\vec{y}}=0,1$ in the lattice bonds joining nearest-neighbor sites $\vec{x}$ and $\vec{y}$ (the marginal probability for the spins is the Boltzmann weight of the original Ising model). The $b=1$ bonds are said to be occupied. Two lattice sites connected by a chain of occupied bonds are said to belong to the same cluster. The clusters partition the lattice into connected components. A crucial role is played by the conditional probabilities. That of the *spins, given the bonds* is: all the spins in a cluster take the same value, the spin values in different clusters are statistically independent and equal ± 1 with 50% probability. For that of the *bonds, given the spins*, different bonds are statistically independent, and a given bond can be one only if it connects spins of equal sign and, in that case, with probability $1-e^{-2\kappa}$. The two Schrödinger-type operators, $\mathcal{P}_{\mathrm{spin}}$ and $\mathcal{P}_{\mathrm{bond}}$ belong to the heat-bath category; $\mathcal{P}_{\mathrm{spin}}$ ($\mathcal{P}_{\mathrm{bond}}$) leave the bonds (spins) unchanged and choose the new spin (bond) configuration according to the conditional probability of the spins (bonds), given the bonds (spins). As all heat-bath dynamics, both $\mathcal{P}_{\mathrm{bond}}$ and $\mathcal{P}_{\mathrm{spin}}$ verify detailed balance. Their product $\mathcal{P}^{\mathrm{SW}}=\mathcal{P}_{\mathrm{spin}}\mathcal{P}_{\mathrm{bond}}$ verifies only the softer balance condition, Eq. (10).

[26] X. J. Li and A. D. Sokal, Phys. Rev. Lett. **63**, 827 (1989).

[27] Combining $\langle \eta'_{\mathcal{C}}\eta'_{\mathcal{S}}\rangle=0$ with Eqs. (A3) and (A8)–(A10), we recover the well-known cluster estimator for U_4: $U_4 = \langle 3(\Sigma_c n_c^2)^2 - 2\Sigma_c n_c^4\rangle / \langle \Sigma_c n_c^2\rangle^2$.

[28] For finite L, $g_{\mathcal{C}}>0$ while the large-L limit for U_4 is 3.

[29] For instance, $\langle \mathcal{M}^2\rangle = \langle P_{\mathrm{spin}}\mathcal{M}^2\rangle = \langle \mathcal{C}\rangle$. A telegraphic proof of $C_{\mathcal{C}\mathcal{B}_{\mathrm{SW}}}(0)=0$ is also obtained by combining $\langle \mathcal{C}\mathcal{M}^2\rangle = \langle P_{\mathrm{spin}}\mathcal{C}\mathcal{M}^2\rangle$ with $P_{\mathrm{spin}}\mathcal{C}\mathcal{M}^2 = \mathcal{C}P_{\mathrm{spin}}\mathcal{M}^2 = \mathcal{C}^2$ and with Eq. (A18).

[30] R. G. Edwards and A. D. Sokal, Phys. Rev. D **38**, 2009 (1988).