Comparative study of selected parallel tempering methods

A. Malakis and T. Papakonstantinou
Section of Solid State Physics, Department of Physics, University of Athens, Panepistimiopolis, GR 15784 Zografos, Athens, Greece
(Received 4 March 2013; revised manuscript received 21 May 2013; published 30 July 2013)

We review several parallel tempering schemes and examine their main ingredients for accuracy and efficiency. The present study covers two selection methods of temperatures and several choices for the exchange of replicas, including a recent novel all-pair exchange method. We compare the resulting schemes and measure specific heat errors and efficiency using the two-dimensional (2D) Ising model. Our tests suggest that an earlier proposal for using numbers of local moves related to the canonical correlation times is one of the key ingredients for increasing efficiency, and protocols using cluster algorithms are found to be very effective. Some of the protocols are also tested for efficiency and ground state production in 3D spin-glass models where we find that a simple nearest-neighbor approach using a local n-fold-way algorithm is the most effective. Finally, we present evidence that the asymptotic limits of the ground state energy for the isotropic case and for an anisotropic case of the 3D spin-glass model are very close and may even coincide.

DOI: 10.1103/PhysRevE.88.013312 PACS number(s): 02.70.Tt, 05.50.+q, 75.10.Nr, 64.60.Cn

I. INTRODUCTION

Monte Carlo (MC) sampling has dramatically increased our understanding of the behavior of statistical mechanics systems [1,2]. Importance sampling, that is, the Metropolis *et al*. [3] method and its variants [1,2] were, for many years, the main tools in condensed matter physics, particularly for the study of critical phenomena in Ising models. However, in many complex systems, effective potentials have a complicated rugged landscape with many minima and maxima which become more pronounced with increasing system size. Thus, any reasonable MC sampling has to overcome energy barriers and cross from one basin to another in the state space in order to obtain a representative set of configurations.

In order to overcome such problems, occurring mainly in the study of first-order phase transitions, spin glasses, and biomolecules, generalized ensembles have been developed with the two main categories known as the entropic sampling method and the parallel tempering sampling method [1]. Parallel tempering (PT), or replica exchange ensembles [4–9], are very effective alternatives for the study of spin glasses [10–30], protein folding [31], and biomolecules [32–34].

PT involves a mixture of MC moves at the individual temperatures (local moves) with exchange attempts between different replicas (swap moves). Local moves depend on the implemented local algorithm, that is on the algorithm used at the individual temperatures. Thus, a local move may be a spin-flip attempt (Metropolis algorithm), or a spin flip to a new state (n-fold-way algorithm) or a cluster flip to a new state (Wolff algorithm). An obvious question is how often we should perform the swap moves, or how many local moves should intervene between swap moves. The number of swap moves required to transfer any replica from the highest (lowest) temperature to the lowest (highest), and vice versa, defines the round-trip time and is one of the possible global measures characterizing the efficiency of a PT protocol. It is a measure indicating difficulties in the flow (bottleneck effects) as a replica moves from the high to low temperature and vice versa. Schemes minimizing the round-trip time are expected to improve the general sampling efficiency of the PT protocols, facilitating equilibration and transitions between multiple minima in spin glasses.

There are many different ways to construct such a PT protocol and the available freedom in choosing specific details makes their comparison a challenging job [35–39]. Our motivation in the present comparative study is to shed light on aspects that appear to be still unresolved. We attempt this by combining several features and testing the behavior of the resulting PT protocols. We consider two methods for the selection of temperature sequences and test their performance by varying the numbers of local moves used at each temperature. One of these methods is the simple constant acceptance exchange (CAE) method [8,9,34,40]. Using this selection method, Bittner *et al*. [40] have shown that if the numbers of local moves are related to the canonical correlation times, the resulting PT protocol appears to optimize the round-trip time of replicas. The calculation of correlation times needs additional costly preliminary runs, but the idea is of theoretical interest, and in some cases an inferred moderate approximation may be better than using an arbitrary not-too-small and not-too-big empirical mixture of local and swap moves.

The second method, for the selection of temperatures, was introduced by Sabo *et al*. [41]. This method requires a constant increase in entropy between successive temperatures. This constant entropy increase (CEI) approach is also supposed to optimize the performance of PT ensembles. The two methods of selection produce temperature sequences that are more concentrated in the temperature range where the specific heat has a maximum, with the CEI method producing a more dense set close to the maximum point. In systems with sharp specific heat peaks, the difference between the corresponding protocols may be considerable and one would like to know their relative efficiency and how this may be influenced by other features of the protocols. We will confirm that the proposal of Bittner *et al*. [40] for using numbers of local moves related to the canonical correlation times is the key ingredient for both selection methods. Thus, using appropriate cluster algorithms for the local moves one obtains very effective protocols optimizing the round-trip time.

There is also great freedom in the choice of the exchange of replicas, although most protocols use adjacent exchange attempts. Even in the case of adjacent exchange moves, the performance of PT is influenced by further details, such as

the numbers of local moves used between exchange attempts and the ordering or mixing of the exchange attempts. Nonadjacent exchange attempts can be incorporated in the PT schemes in a way that fully maintains detailed balance. However, remote exchanges have in general very small acceptance rates, and such schemes may well waste time in unsuccessful attempts, without any statistically significant increase in the sampling efficiency of the protocols. Such methods of nonadjacent exchange moves have also been tried in constructing PT protocols [42] and will also be included in our tests in this paper. An alternative procedure for overcoming the problem of low acceptance rates for remote exchanges has been recently adopted, by enforcing exchange of nonadjacent replicas using kinetic MC methods [43,44]. The novel all-pair exchange (APE) method of Brenner *et al*. [44] is such a case. This method maintains detailed balance, at least in determining the probabilities of generating exchange moves (between adjacent or nonadjacent replicas). We will consider this method as a representative of kinetic MC methods and contrast its accuracy and efficiency with the nearest neighbor (NN) PT protocols. In particular, the combination of the APE method with a Wolff (W) cluster algorithm [1,45] or an n-fold-way algorithm will be examined. This last algorithm is called also in the literature the BKL algorithm [1,46–48], after Bortz, Kalos, and Lebowitz [46] who invented it. Other names are the continuous time MC or the kinetic MC algorithm (see the discussion in Sec. II B). The recent infinite swapping method [49], will not be considered in this paper. This technique utilizes a symmetrization strategy, using all possible $(M!)$ temperature (replica) permutations. The application and the performance of this method to the problem of finding true ground states of the three-dimensional (3D) spin-glass model is in our future interests. Our numerical tests are carried out for the square Ising model with linear size $L = 50$ and $N = L \times L$ lattice sites, and involve accurate measurements for the specific heat errors and the efficiency of the PT protocols. Furthermore, detailed tests are carried out for efficiency and ground state production in 3D spin-glass models, where we find efficient low-temperature choices using, as a local algorithm, the n-fold-way algorithm.

The rest of the paper is laid out as follows: In Sec. II A we give a brief description of two basic methods for selecting the temperature sequences for PT sampling. The tested NN and APE exchange schemes are detailed in Sec. II B and the rest of the PT protocol details are defined in the elementary PT step in Sec. II C. In Sec. III we define measures for the specific heat errors and the efficiency and we present several tests of the PT protocols on the 2D Ising model. Finally, in Sec. IV we test the performance of some of the PT protocols for ground state production on 3D spin-glass models. In Sec. IV A we discuss the problem of ground states for the 3D Edwards-Anderson bimodal (EAB) model [10,12], while in Sec. IV B we consider a variant of this, with spatially uniaxial anisotropic exchange interactions and study the finite-size behavior of its ground state energy. Our conclusions are summarized in Sec. V.

II. PARALLEL TEMPERING SCHEMES

A. Selecting temperature sequences

In constructing an accurate and efficient PT protocol, the optimum selection of temperatures is still an open problem. There is a rather large number of ideas that have been proposed in the last decade to resolve this question [40–44,50–53]. According to the approach followed by Katzgraber *et al*. [53], optimal temperatures correspond to a maximum rate of round trips between low and high temperatures in temperature space and can be obtained using a recursive readjustment of temperatures. This feedback-optimized update scheme, is a sophisticated and appealing method, but because of its complexity, other simpler methods have been more often implemented in comparative studies and applications of PT.

Among these simpler methods, the CAE method, when used with appropriate number of sweeps between replica exchanges, has been illustrated to produce a similar approach that optimize the round-trip time [40]. To obtain the temperatures corresponding to a CAE rate r we follow here Ref. [40]. Starting from a chosen lowest temperature, adjacent temperatures are determined by calculating the acceptance exchange rate from

$$R(1 \leftrightarrow 2) = \sum_{E_1,E_2} P_{T_1}(E_1) P_{T_2}(E_2) p(E_1,T_1 \leftrightarrow E_2,T_2), \quad (1)$$

where $P_{T_i}(E_i)$ is the energy probability density function for replica i at temperature T_i and

$$p(E_1,T_1 \leftrightarrow E_2,T_2) = \min[1, \exp(\Delta\beta \Delta E)] \quad (2)$$

is the PT probability to accept a proposed exchange of two replicas, with $\Delta\beta = 1/T_2 - 1/T_1$ and $\Delta E = E_2 - E_1$. Demanding that $R(1 \leftrightarrow 2) = r$ for all adjacent replicas, we obtain the temperatures of the required CAE sequence (from the above equations), provided that the energy probability density functions (PDFs) are known, or can be reasonably well approximated by some preliminary MC runs.

Although the exact density of states (DOS) and the energy PDF for the square Ising model with linear size $L = 50$ is known [54], we have used this information only for the exact determination of specific heat errors and not for defining the CAE temperature sequence. Instead, we apply a simple histogram method [1,45] to find the energy PDFs at any temperature, using a preliminary (Metropolis or PT) run in an appropriate set of temperatures in the range of interest. Applying then a recursive scheme we calculate the CAE sequence, and by repeating a Metropolis run at these temperatures we estimate their canonical correlation times [1]. This practice can be applied to a general system for which the DOS is not known, possibly using in the first preliminary run (especially in a spin-glass system) a PT protocol in an *ad hoc* reasonable set of temperatures.

The second method of selection of the PT temperature sequence, requires a constant increase in entropy between successive temperatures [41]. To describe this method we follow Sabo *et al*. [41] and denote the M temperatures of the CEI sequence by $(T_m;\ m = 1,\ldots,M)$ and the total increase in entropy from T_1 to T_M by ΔS. Then the adjacent temperatures are determined successively, starting from the given T_1, from

$$\int_{T_m}^{T_{m+1}} dT \frac{C_u(T)}{T} = \frac{\Delta S}{(M-1)}, \quad (3)$$

where the specific heat at any temperature can be calculated from the above mentioned preliminary (Metropolis or PT) run

TABLE I. Temperature sequences for PT methods, canonical correlation times in units of lattice sweeps, and acceptance rates. The two cases shown are the constant acceptance rate (columns 1, 2, and 3) and the constant entropy increase (columns 3, 4, and 5) methods, as applied in a temperature range close to the critical point of the $L = 50$ square Ising model.

CAE			CEI		
T	τ	r	T	τ	r
1.9200	3.0	0.499	1.9200	3.0	0.361
1.9669	3.0	0.500	1.9825	3.0	0.403
2.0121	3.0	0.501	2.0375	3.6	0.428
2.0557	3.9	0.499	2.0875	4.0	0.452
2.0975	4.0	0.498	2.1325	4.8	0.480
2.1377	5.0	0.501	2.1725	6.2	0.509
2.1757	6.4	0.500	2.2075	10.0	0.551
2.2115	9.5	0.501	2.2375	13.7	0.594
2.2446	19.5	0.501	2.2625	25.6	0.607
2.2751	32.6	0.501	2.2850	39.0	0.617
2.3050	33.1	0.500	2.3075	31.3	0.601
2.3374	19.1	0.500	2.3325	20.3	0.572
2.3746	9.8	0.500	2.3625	11.2	0.531
2.4167	6.3	0.500	2.4000	7.3	0.519
2.4631	5.0	0.500	2.4425	5.6	0.489
2.5134	4.1	0.500	2.4925	4.5	0.447
2.5680	4.0	0.500	2.5525	4.0	0.434
2.6268	3.8	0.500	2.6200	3.9	0.410
2.6903	3.0		2.6975	3.0	

and a simple histogram method [1,45]. Repeating a Metropolis run at these temperatures we also estimate the corresponding canonical correlation times.

In Table I we display, for the CAE and CEI selection methods, the corresponding temperature sequences, their canonical correlation times, and the acceptance rates between adjacent replicas. As usual, to fix the temperature scale we set (the exchange interaction of the Ising model) $J/k_B = 1$. The canonical correlation times were estimated from the discrete form of the energy autocorrelation function, following the method described in Ref. [1], and are measured in units of lattice sweeps (N Metropolis attempts). The temperature range used is approximately centered around the pseudocritical temperatures of the specific heat and magnetic susceptibility of the $L = 50$ square Ising model, and includes the exact critical point. Both methods produce temperature sequences that are more concentrated in the temperature range where the specific heat has a maximum, with the CEI method producing a more dense set close to the maximum point, as can be seen from this table.

We point out here that, for the CAE selection, we start from a given lower temperature $T_1 = 1.9200$ and proceed to find higher temperatures, until a desired limit, with a given constant acceptance rate. For the construction of Table I, which is used in our tests presented in Sec. III, we have used $r = 0.5$ and the resulting higher temperature is $T_M = 2.6903$ with $M = 19$. Our choice here for the CAE rate ($r = 0.5$) follows Ref. [40], and is somewhat arbitrary. One could also use the value $r = 0.3874$ recommended in Ref. [55]. However, our test were repeated (on a lattice with linear size $L = 20$) using other values of the CAE rate, producing similar behavior. We discuss this point again in Sec. III, where we observe that the CAE and CEI selections of temperatures show comparable performance. Since we wish to compare the CAE and CEI schemes, we apply the CEI procedure starting from the same $T_1 = 1.9200$, set its final temperature $T_M = 2.6975$, and use the same number of replicas $M = 19$. The small difference between the two higher temperatures is due to the histogram data kept for the specific heat in the preliminary run, since both sequences were obtained in one unified algorithm.

Thus, the two schemes are defined approximately in the same temperature range with the same number of replicas, a practice that facilitates their comparison. The weaknesses or merits of the two selection methods may also be related to the choice of lower and higher temperatures and the value of the constant acceptance rate r, and thus will depend on the total number of replicas M. In general, all the details of the protocols may influence the round-trip time or the efficiency of the PT schemes. An interesting recent example is that of the PT cluster algorithm based on the CAE selection method presented by Bittner and Janke [56]. In their implementation, it was possible to study the critical range of the 2D and 3D Ising models with a rather small number of replicas. In the case of the 3D Ising model, it was argued that a PT scheme using only $M = 4$ replicas, for lattice sizes $L = 4$–80, was adequate to cover the critical range. Finally, these authors used in their study a more generous approach with $M = 7$–21 replicas, for lattice sizes $L = 4$–80.

B. Nearest neighbor and all-pair exchange schemes

Mixing local MC attempts at individual temperatures (local attempts) with exchange attempts between different replicas (swap attempts) is the essential procedure in PT. It is this feature that enables an ergodic walk in temperature space, transferring information between the highest and lowest temperatures. We provide in this section short descriptions of alternatives for the ordering of swap attempts, first for the NN exchange protocols and then for PT protocols using all-pair exchange methods. The rest of the details of an elementary PT step are described in the next section.

In a NN exchange protocol only adjacent exchange attempts are proposed. There are $M - 1$ different NN proposals which may be uniquely denoted by the lowest temperature index $i = 1, \ldots, M - 1$. Following Brenner *et al.* [44], we denote the replica configuration before a swap attempt by $A = \{x_1, x_2, \ldots, x_i, x_{i+1}, \ldots, x_M\}$, where x_i is the replica at the temperature T_i. Then, a NN exchange is denoted by $A \to B$, where $B = \{x_1, x_2, \ldots, x_{i+1}, x_i, \ldots, x_M\}$. The acceptance- rejection rule of an exchange attempt corresponds to an acceptance rate $P_{\rm acc}(A \to B)$, which is usually given by the Metropolis form of the swap attempt $p(x_i \leftrightarrow x_{i+1})$, as specified in Eq. (2). Since the generation of the various NN exchange attempts (proposals) proceeds with equal probabilities, $P_{\rm gen}(A \to B) = P_{\rm gen}(B \to A) = 1/(M - 1)$, the swap attempts satisfy the detailed balance condition $P(A)P_{\rm gen}(A \to B)P_{\rm acc}(A \to B) = P(B)P_{\rm gen}(B \to A)P_{\rm acc}(B \to A)$. The product probability distribution $P(A) = \rho(x_1)\rho(x_2)\cdots\rho(x_M)$ is stationary with respect to the swap attempts. We now specify four choices [$(\mathrm{NN})_a$, $(\mathrm{NN})_b$, $(\mathrm{NN})_c$, and $(\mathrm{NN})_d$] for the ordering of the

NN proposals that will be tested in Sec. III. In $(\mathrm{NN})_a$, a random permutation (say, $j_1,\ldots,j_{M-1}$) is generated, from the set $i=1,\ldots,M-1$, and this permutation is used in an exchange swap cycle of $M-1$ proposals. Thus, swap moves are organized in cycles of $M-1$ proposals, and as explained in the next section, such a swap cycle may be used in defining the unit of time of an elementary PT step. In $(\mathrm{NN})_b$, the lowest temperature index (i) is randomly chosen from the set $i=1,\ldots,M-1$ and thus multiple exchanges of the same pair are allowed in the swap cycle. In $(\mathrm{NN})_c$, the sequence of proposals is fully deterministic, consisting of two swap subcycles, starting from the odd proposals $i=1,3,\ldots$, in increasing order, and continuing with the even proposals $i=2,4,\ldots$. Finally, in $(\mathrm{NN})_d$, the odd and even subsequences are randomly permuted before starting the odd and following with the even swap subcycles.

Next, we consider PT protocols using APE methods, in which adjacent or nonadjacent replicas may be exchanged $(x_i \leftrightarrow x_j)$. In these methods the number of possible proposals $(A \to B)$ is $M(M-1)/2$, $A=\{x_1,x_2,\ldots,x_i,\ldots,x_j,\ldots,x_M\}$ and $B=\{x_1,x_2,\ldots,x_j,\ldots,x_i,\ldots,x_M\}$. In the simplest case, for each proposal a pair (x_i,x_j) is randomly chosen from the set of all $M(M-1)/2$ different pairs. Thus, the generation of exchange attempts proceeds with equal probabilities, $P_{\mathrm{gen}}(A\to B)=P_{\mathrm{gen}}(B\to A)=1/[M(M-1)/2]$, and if the acceptance-rejection rule follows the Metropolis form, with an acceptance rate $p(x_i \leftrightarrow x_j)$, then the swap attempts satisfy the detailed balance condition. This simple APE method will be included in our tests and is denoted, in the following, as APE_M, whereas the method of Brenner *et al.* [44] will be denoted by APE_B. As pointed out in the Introduction, methods attempting remote exchange moves have been also tried in constructing PT protocols [42] and will be included in our tests in Sec. III. Remote exchanges may be thought of as replacing several adjacent swaps to a single move. However, the APE_M method suffers very small acceptance rates, and no significant increase in the sampling efficiency of the protocols should be expected.

Finally, we discuss the so-called kinetic MC methods and present details of the APE methods proposed by Calvo [43] and Brenner *et al.* [44]. In statistical physics, particularly in simulation studies of Ising models, the kinetic MC method is better known as the n-fold-way algorithm or BKL algorithm [1,46–48]. In describing this algorithm, we shall follow the original Ref. [46] and borrow from the terminology in Sec. 2.4 of Ref. [1]. In the traditional MC simulation (for instance the Metropolis algorithm) we use an acceptance-rejection rule in every MC attempt, and the system may stay in the same state for some time Δt. In a kinetic MC algorithm, we force the system to select a new state, but we also introduce a time step which corresponds to a varying length (stochastic variable), depending on how long we expect the system to remain in its current state before moving to a new one in the traditional MC method [1]. The averaging process for any observable becomes a time average and the values of the observable are weighted by the corresponding time steps (divided, at the end of the measuring process, by the total time), while the total time variable t is incremented by Δt. The selection of a new state assumes an appropriate set of probabilities and proceeds as follows: Let the current state of the system be μ and denote by $r_j \equiv r(\mu\to\nu)$ the acceptance rates (acceptance probabilities) for all possible K transitions $(j=1,2,\ldots,K)$ to new states from the current state. Draw a random number $0<R\leqslant 1$ and select the new state ν, corresponding to transition i, if $Z_{i-1}<RZ_K\leqslant Z_i$, where $Z_m=\sum_{j=1}^{m} r_j$. With the help of the cumulative function Z_m, the new state ν is selected with probability $P_i=r_i/Z_K$, proportional to the acceptance probability $r(\mu\to\nu)$. The P_j are the selection probabilities of the kinetic MC algorithm. The time intervals Δt have to be recalculated at each step. They can be obtained as average lifetimes [47], from the values of $r(\mu\to\nu)$, and $\Delta t \propto {Z_K}^{-1}$ [1,46–48]. Note also that, in an equivalent way, the selection of a new state i can proceed with the condition stated in terms of a cumulative function Q_m obtained from the selection probabilities. In this case, the condition reads as $Q_{i-1}<R\leqslant Q_i$, where $Q_m=\sum_{j=1}^{m} P_j$.

These are the main ingredients of the kinetic MC algorithm. In some cases, the set of all transitions can be classified into a small number of n classes (n-fold-way algorithm) and the method becomes very efficient, but in general the recalculation of all transition probabilities in each step is the obvious drawback of the method. In Ref. [46], the n-fold-way algorithm is described in detail for the square lattice Ising model [periodic boundary conditions (PBCs)] in a nonzero field, using the $n=10$ different classes, corresponding to the energy changes under a spin flip. For the zero-field square lattice (with N sites and PBCs) Ising model, $n=5$ classes cover the spin-flip energy changes $\Delta E_j=8-4(j-1)$, where the index numbering the classes is $j=5-z$ and $z=4,3,2,1,0$ is the number of nearest neighbor spins having the same sign as the spin to be flipped [46–48]. In this case, the statistical weight for the selection of a class is the sum of the acceptance rates of all spins in the class and takes the form $r_j^{cl}=N_jA_j$, where N_j are the current populations of spins $(\sum_j N_j=N)$ and $A_j=\min[1,\exp(-\beta\Delta E_j)]$ is the corresponding (Metropolis) acceptance rate of any spin in the class. The class selection proceeds as outlined above, using the cumulative function $Z_m=\sum_{j=1}^{m} r_j^{cl}$ and the average lifetimes are given by $\Delta t = N{Z_n}^{-1}$ [47]. The spin to be flipped is then chosen randomly from the selected class and the spin flip is enforced. This is a simple and efficient n-fold-way algorithm, a kinetic MC reorganization of the original Metropolis algorithm [46], with transitions that follow detailed balance. The method has been used successfully to simulate thermodynamic equilibrium of various types of Ising-like models, using, as mentioned earlier, an appropriate time averaging process. This version of an n-fold-way algorithm is implemented as a local algorithm in Sec. III for the square lattice Ising model and in Sec. IV for the 3D (cubic) spin-glass model ($n=7$).

The APE method proposed by Calvo [43] is an attempt to adjust the above ideas of the kinetic MC to the PT swapping procedure. Following the terminology of Brenner *et al.* [44], let Φ denote the set of all macrostates (replica configurations) $B=\{x_1,x_2,\ldots,x_j,\ldots,x_i,\ldots,x_{M-1},x_M\}$ reachable from the current macrostate $A=\{x_1,x_2,\ldots,x_i,\ldots,x_j,\ldots,x_{M-1},x_M\}$ by an adjacent or nonadjacent pair exchange $(x_i \leftrightarrow x_j)$. This set describes all possible transitions $A\to B$ to a new state in a kinetic MC scheme. Yet Calvo also includes the possibility of not performing any exchange (event $j=0$) to which the acceptance probability $P_{\mathrm{acc}}(A\to A)=1$ is attributed [43].

Thus, the PT swapping procedure of Calvo [43] involves, besides the $K = M(M-1)/2$ transitions to new states of the form $A \to B \neq A$ $(j = 1,2,\ldots,K)$, also the event $A \to A$ $(j = 0)$. This is an unconventional use of the kinetic MC method and the selection probabilities of the events are given by [43,44]

$$P_j = P_{\rm acc}(A \to C) \Big/ \left[1 + \sum_{M \in \Phi} P_{\rm acc}(A \to M) \right], \qquad (4)$$

where C is A or B. The event i is selected, and is enforced, from the set of $j = 0,1,2,\ldots,K$, using the cumulative function Q_m obtained from the selection probabilities. The condition is $Q_{i-1} < R \leqslant Q_i$, where $Q_m = \sum_{j=1}^{m} P_j$. As pointed out by Calvo [43], the attribution of a residence time (average lifetime) is inconvenient in PT schemes and has been replaced in his method by including the rejection $j = 0$ event. However, this is not a self-consistent use of the kinetic MC method for estimating thermodynamic equilibrium. Therefore, Calvo [43] suggests that the exchange moves (enforced without the use of time weights) could be considered as extra moves that are not directly involved in the averaging process. In a subsequent paper, Brenner *et al.* [44] pointed out a further inconsistency of the method. The set of macrostates Φ reachable from A is not, in general, the same as the set of macrostates Ψ reachable from B. Thus, the generation probability [inverse of the denominator in Eq. (4)] $P_{\rm gen}(A \to B) = [1 + \sum_{M \in \Phi} P_{\rm acc}(A \to M)]^{-1}$ for the transition $A \to B$, will in general be different from the generation probability $P_{\rm gen}(B \to A) = [1 + \sum_{L \in \Psi} P_{\rm acc}(B \to L)]^{-1}$ for the transition $B \to A$. As a consequence the detailed balance condition $P(A)P_{\rm gen}(A \to B)P_{\rm acc}(A \to B) = P(B)P_{\rm gen}(B \to A)P_{\rm acc}(B \to A)$ is not met.

In order to overcome this inconsistency, Brenner *et al.* [44] proposed a revision of the above method. The generation probabilities are now replaced by [44]

$$P_{\rm gen}(A \to B) = 1 \Big/ \max \left\{ \sum_{M \in \Phi} P_{\rm acc}(A \to M), \sum_{L \in \Psi} P_{\rm acc}(B \to L) \right\} \qquad (5)$$

and the event generation probabilities satisfy now the desired relation $P_{\rm gen}(A \to B) = P_{\rm gen}(B \to A)$ [44]. This condition makes the scheme consistent with detailed balance, in contrast with the method originally proposed by Calvo [43]. The PT swapping procedure of Brenner *et al.* [44] involves explicitly only the $K = M(M-1)/2$ transitions of the form $A \to B \neq A$ $(i = 1,2,\ldots,K)$ and the selection probabilities for the kinetic MC method are given by

$$P_j = P_{\rm acc}(A \to B)P_{\rm gen}(A \to B), \qquad (6)$$

where $P_{\rm acc}(A \to B)$ is the corresponding PT acceptance rate $p(x_i \leftrightarrow x_j)$, as specified in Eq. (2), and $P_{\rm gen}(A \to B)$ is given by Eq. (5). The selection of the state (macrostate) from the set of $j = 1,2,\ldots,K$ states proceeds again, using the cumulative function Q_m obtained from the selection probabilities. The condition is $Q_{i-1} < R \leqslant Q_i$, where $Q_m = \sum_{j=1}^{m} P_j$. The algorithm of Brenner *et al.* [44] does not include explicitly, in the set of selection probabilities, the possibility of not performing any exchange. However, this may happen in the direction of the less favorable Boltzmann exchange [44], since $P_{\rm gen}$ is a maximum and thus in some cases $Q_K = \sum_i P_i < 1$. In such cases, no new state is selected from the condition $Q_{i-1} < R \leqslant Q_i$, where $Q_m = \sum_{j=1}^{m} P_j$ if $R > Q_K$. But again, the method proposed by Brenner *et al.* [44] does not use any time weights (residence times or average lifetimes) for the MC averaging process, as is usually done in a kinetic MC algorithm when it is used for the estimation of thermodynamic equilibrium properties. This is a common problem of the above methods, and it may be crucial for the estimation of any observable in thermodynamic equilibrium. The accuracy of the resulting PT protocols has not been tested before and it cannot be guessed beforehand.

We will examine this last APE method in combination with several local algorithms, including the Metropolis, the n-fold way, and the cluster Wolff algorithm. As mentioned earlier, we shall use the notation APE_B to refer to the described all-pair exchange method of Brenner *et al.* [44] and the notation APE_M for the simple APE method using for the swap probabilities the acceptance-rejection rule of Metropolis form given in Eq. (2). The APE method proposed by Calvo [43] will not be included in the presentation of our tests, and we mention here only that, as we have verified, this method is slightly inferior to the APE method of Brenner *et al.* [44]. A further obvious drawback of the Brenner *et al.* [44] method is the costly recalculation of all generation probabilities in each exchange step for the application of Eqs. (5) and (6). Note that, before selecting a particular swap event in the kinetic MC method of Brenner *et al.* [44], all proposals $A \to B$ are considered and for each of them both sums in the denominator of Eq. (5) are calculated from the known acceptance rates. For this reason, the APE_B PT protocol needs considerable more CPU time. This additional time can be reduced by observing that the APE_B protocol does not create distant exchanges with significant probability. More details are provided in the discussions of Sec. III.

C. The elementary parallel tempering step and the local algorithms

We now define a PT step (PTS) as the elementary MC step used for the recording (measuring or averaging) process during an independent MC run. A PTS may consist of one or several swap cycles of $M-1$ replica exchange proposals, and we use this definition also in the case of the APE method, so that our unit is the same for all protocols to be tested. After each exchange attempt of the swap cycle, all replicas attempt a number of local moves (spin flips or cluster moves) at their respective temperatures. The number of these local moves is in general chosen to depend on the temperature and is denoted by $n(T_i)$. The total number of local moves at any temperature of the PT protocol (T), in a swap cycle, is $N_{\rm local}(T) = (M-1)n(T)$. The swap cycle is thus the above described mixture of standard MC and replica exchange attempts. Without loss of generality, we define the PTS to be just one such swap cycle, and we differentiate between various protocols by varying the number of local moves. It is convenient to set also $N_{\rm local}(T) = f(T)\tau(T)N$, where $N = L^D$ is the number of lattice sites, $\tau(T)$ are the canonical correlation times, and $f(T)$ are factors that facilitate the adaption of the numbers of local moves $n(T)$ in a style depending on the application and/or the behavior

of the system. $N_{\rm local}(T) = N$ corresponds to the usual lattice sweep.

The above defines our MC unit for the recording process. We note that no recording is attempted during a first disregarded or equilibration part of the simulation, and an adequate number of PTSs, $t_{\rm eq}$, is used for this part. Then we use a large number $t_{\rm av}$ of PTSs for the recording or averaging part of the simulation. For the local standard MC moves, the Metropolis algorithm, the n-fold-way algorithm, and the cluster Wolff algorithm will be implemented. Furthermore, a large number N_r of independent MC runs is used in our tests. This repetition makes more reliable the measures of accuracy and efficiency illustrated in our tests. It is also desirable to use approximately the same CPU time for the different PT schemes that are compared. In this way, we hope to obtain an objective assessment of their behavior. Since the local algorithms obey different time complexities, we will adapt the parameters $N_{\rm local}(T)$, $t_{\rm eq}$, $t_{\rm av}$, and N_r in a way that makes the algorithms (almost) equivalent in CPU time.

III. ERROR AND EFFICIENCY MEASURES: TESTS ON THE 2D ISING MODEL

In this section we present tests, carried out by using the $L = 50$ square ferromagnetic Ising system with PBCs. A variety of PT protocols will be implemented, obtained by combining features mentioned in the previous sections. Using the exact DOS [54], we calculate the values of the specific heat at the PT protocol temperatures and define the following error measures:

$$\epsilon(T_i) = [C_{\rm exact}(T_i) - C_{\rm PT}(T_i)]/C_{\rm exact}(T_i), \tag{7}$$

$$\bar{\epsilon} = \sum_{i=1}^{M} \epsilon(T_i)/M, \tag{8}$$

$$\hat{\epsilon} = \sum_{i=1}^{M} |\epsilon(T_i)|/M, \tag{9}$$

$$\epsilon^* = \max[|\epsilon(T_i)|], \tag{10}$$

Our first test concerns the NN PT protocols and in particular the four cases $(\mathrm{NN})_a$, $(\mathrm{NN})_b$, $(\mathrm{NN})_c$, and $(\mathrm{NN})_d$ described in Sec. II B. In Fig. 1, we illustrate their errors $\epsilon(T_i)$, and on the legend (in parentheses) we give the corresponding error measures, namely, $\bar{\epsilon}$, $\hat{\epsilon}$, and ϵ^*, as defined above. In each of the presented cases, we have used $N_r = 200$ independent MC runs with $t_{\rm eq} = 3N$, $t_{\rm av} = 15N$, and $N_{\rm local}(T) = N$ and the Metropolis algorithm for the local moves. It is evident from the excellent accuracy obtained in all cases that there is not any noticeable and statistically significant difference between the four choices; rather it appears that they all obey a good mixing of the exchange attempts in the long run. Therefore, hereafter we will use only the $(\mathrm{NN})_a$ PT protocol and vary the other ingredients of the schemes.

We now attempt to observe whether the two selections of temperature produce any significant difference. Also, we test the NN against the APE protocols. In Fig. 2, we illustrate the error behavior of two NN and two APE protocols using both the CEI and CAE selections as indicated in the legend. The details of the illustrated PT schemes $[N_r, t_{\rm eq}, t_{\rm av}, N_{\rm local}(T)]$, are the same as the ones specified above. The NN protocols

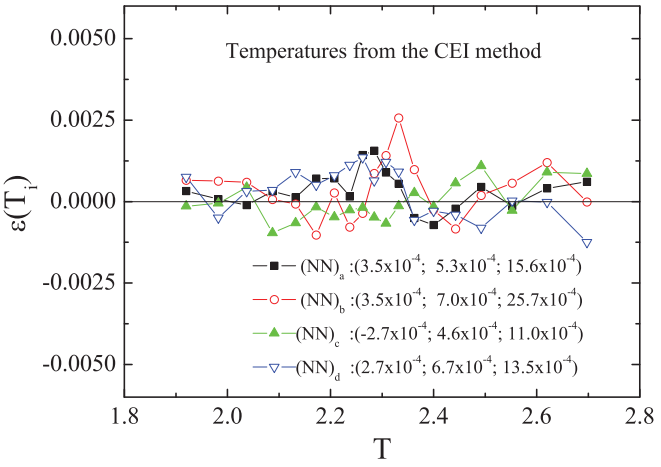

FIG. 1. (Color online) Specific heat errors for the nearest neighbor (NN) PT protocols. The CEI selection of temperatures, as given in Table I, has been used. The error measures, given in parentheses in the legend, are respectively $\bar{\epsilon}$, $\hat{\epsilon}$, and ϵ^*.

perform excellently when compared with the APE protocols. The CAE and CEI selections give comparable accuracy. The APE_B protocol suffers very large errors in the specific heat that are more pronounced close to the specific heat maximum. Apparently, this erratic behavior is a reflection of the problem mentioned earlier. The APE_B method does not use any time weights for the MC averaging process, and therefore is, in general, unreliable for the estimation of thermodynamic equilibrium properties. On the other hand, the APE_M protocol, which is a standard PT protocol attempting also distant (Metropolis) exchanges, shows a good error behavior. We should also point out here that the APE_B runs need almost double CPU time, due to the calculations needed for the application of Eqs. (5) and (6). However, this additional time can be greatly reduced by observing that the APE_B protocol

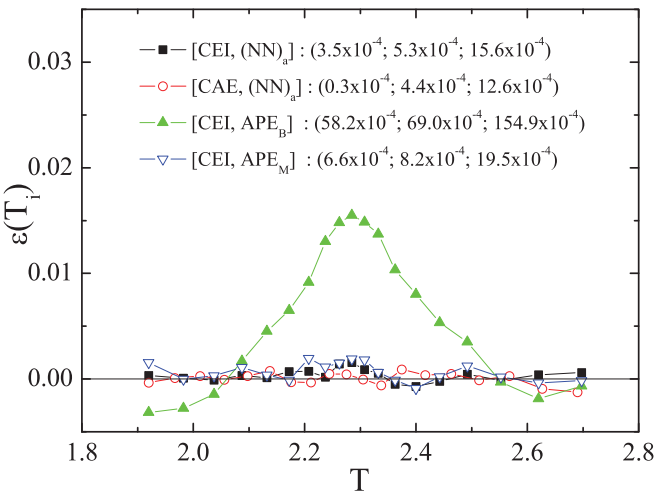

FIG. 2. (Color online) Specific heat errors for four different PT protocols. The first two use the nearest neighbor protocol $(\mathrm{NN})_a$, while the last two use the all-pair exchange protocols of Brenner *et al.* [44] and Metropolis (Sec. II B). The method for the selection of temperatures and the resulting error measures are indicated in the notation in the legend.

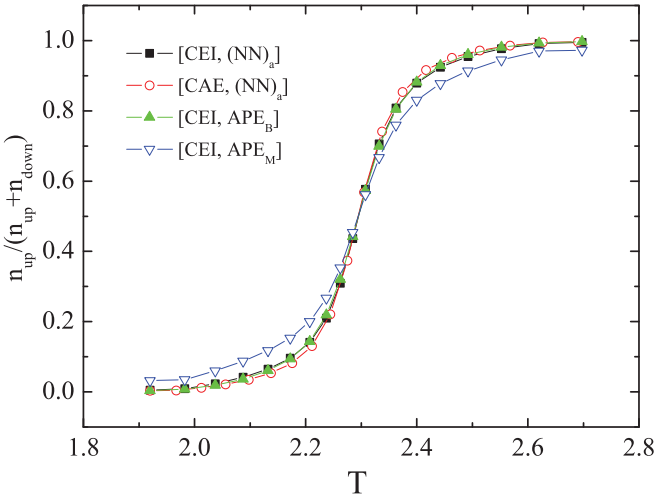

FIG. 3. (Color online) Diffusion behavior of the PT protocols. The diffusion fraction $n_{up}(T_i)/[n_{up}(T_i) + n_{down}(T_i)]$, as defined in the text, for the four PT protocols of Fig. 2.

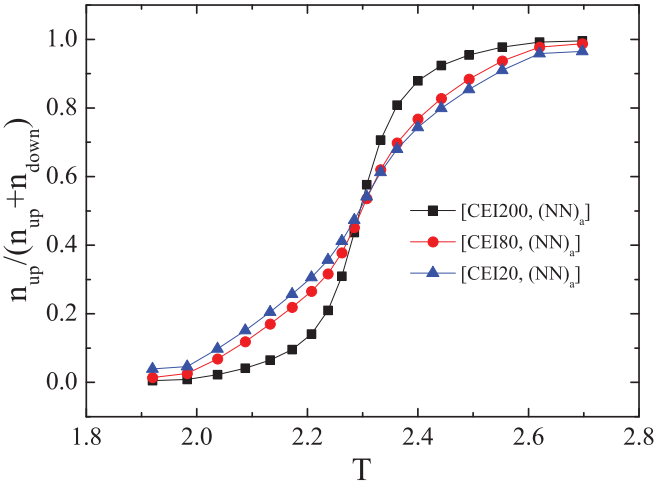

FIG. 4. (Color online) Changes induced in the behavior of the diffusion fraction by varying the numbers of local moves in accordance with the canonical correlation times, as proposed by Bittner *et al.* [40]. The three protocols use the CEI selection and correspond to $N_{\rm local}(T) = N$, $N_{\rm local}(T) = 0.25\tau(T)N$, and $N_{\rm local}(T) = \tau(T)N$. The corresponding numbers of independent MC runs are $N_r = 200$, 80, and 20 as indicated in the notation in the legend.

does not create distant exchanges with significant probability, and thus can be restricted to significant exchanges only.

In order to observe the diffusion behavior of the PT protocols, we define for each temperature T_i the fraction of replicas which have visited one of the two extremal temperatures most recently. In our implementations we assign labels *up* or *down* to each replica if its most recent visit to one of the two extremal temperatures is to T_M or T_1, respectively. The label of an *up* replica remains unchanged if the replica returns to T_M, but changes to *down* upon its first visit to T_1. Similarly, the label of a *down* replica remains unchanged if the replica returns to T_1, but changes to *up* upon its first visit to T_M. For each temperature T_i, we record histograms $n_{up}(T_i)$ and $n_{down}(T_i)$, which are incremented by 1 after a swap attempt involving an *up* or *down* replica at T_i respectively.

The diffusion fraction $n_{up}(T_i)/[n_{up}(T_i) + n_{down}(T_i)]$ is illustrated in Fig. 3, for the PT protocols of Fig. 2. Apparently, the same definition has been used in Ref. [40] and a similar one (with *up* replaced by *down* and vice versa) in Ref. [53]. All four cases in Fig. 3 show very similar diffusion behavior with a sharp decline close to the critical point (specific heat maximum). Thus, a significant increase of efficiency of kinetic MC methods [43,44] that attempt to enforce remote exchanges is not verified from the illustrated diffusion behavior. Figure 3 also illustrates the similarity in the diffusion behavior between the two selection methods. The larger concentration of replicas of the CEI method, close to the maximum specific heat point, does not produce any significant difference in the diffusion behavior of the protocol. We observe again comparable performance for the CAE and CEI selection rules. This observation, may be related to the finding of Ref. [55] (see Fig. 1, and the related discussion in Sec. V). Namely, moderate variations in the acceptance rate (r) do not produce significant changes in the efficiency of PT protocols, provided the acceptance rates remain in the range of optimum performance.

Our next test, illustrated in Fig. 4, verifies the important observation by Bittner *et al.* [40]. By varying the numbers of local moves, in accordance with the canonical correlation times, we observe strong changes induced in the behavior of the diffusion fraction. The three protocols illustrated use the CEI selection with $t_{\rm eq} = 3N$ and $t_{\rm av} = 15N$, and correspond to $N_{\rm local}(T) = N$, $N_{\rm local}(T) = 0.25\tau(T)N$ and $N_{\rm local}(T) = \tau(T)N$. The numbers of independent MC runs are $N_r = 200$, 80, and 20 as indicated in the notation in this figure ([CEI200,(NN)$_a$], [CEI80,(NN)$_a$], and [CEI20,(NN)$_a$], respectively), and thus in total the schemes require approximately the same CPU time (from Table I, the mean of the correlation times is approximately 10). Furthermore, in Fig. 5 we reproduce the [CEI200,(NN)$_a$] and [CEI20,(NN)$_a$] cases of the previous figure with the corresponding CAE protocols. This comparison supplements the crucial observation that the key ingredient, improving the efficiency of the protocols, is the proper choice of the number of local moves.

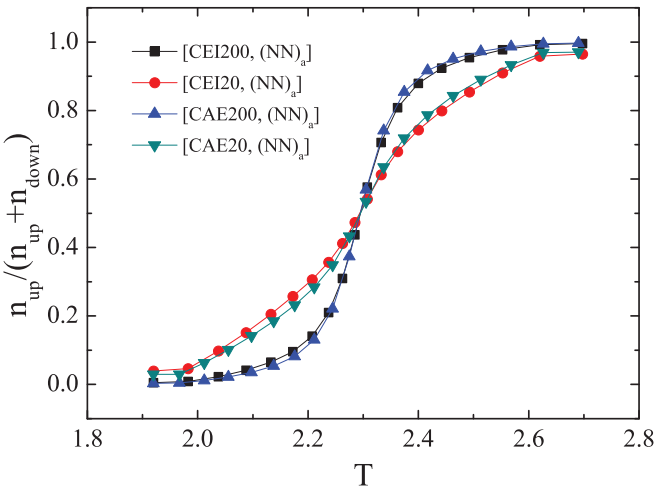

FIG. 5. (Color online) The first two CEI protocols are the same as those illustrated in Fig. 4, while the other two are the corresponding CAE protocols.

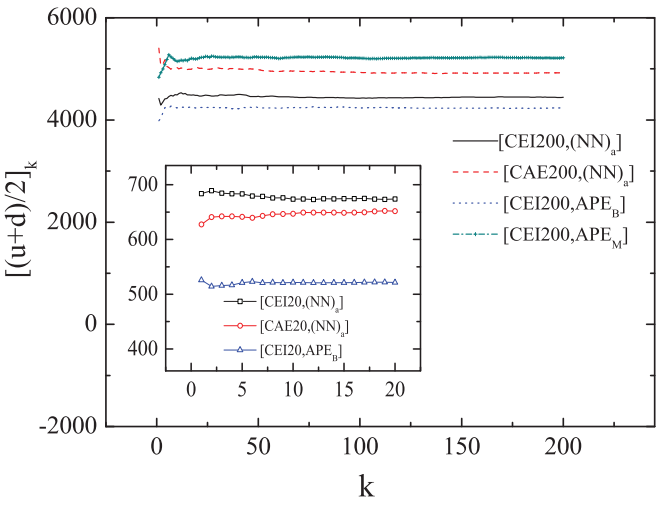

FIG. 6. (Color online) Illustration of the efficiency of the PT protocols using the measures defined in Eq. (11). The four protocols of the main panel use $N_{\rm local}(T) = N$, while the three PT protocols in the inset use $N_{\rm local}(T) = \tau(T)N$.

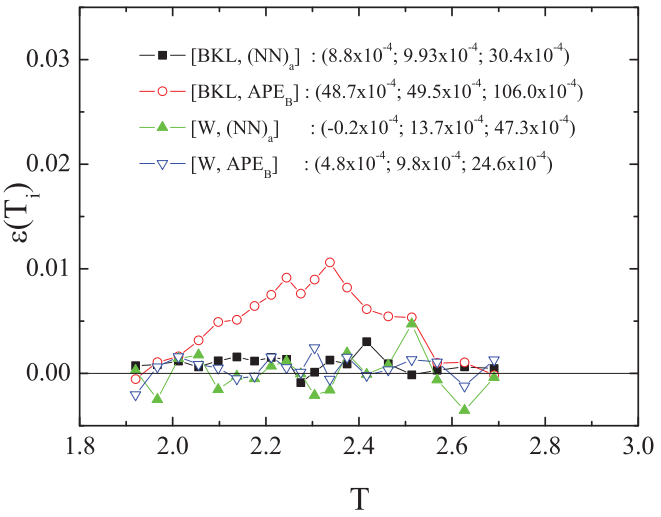

FIG. 7. (Color online) Specific heat errors for the PT protocols shown in the legend and discussed in more detail in the text. The CAE selection of temperatures, as given in Table I, has been used in all four cases.

An alternative way to measure the efficiency of the PT protocols is to observe global aspects of the statistics of the numbers of exchange attempts required to transfer any replica from the highest (lowest) temperature to the lowest (highest) [44]. Let u_j (d_j) denote the average numbers of exchange attempts required for the corresponding transfer, averaged over the PTSs of a long independent run j ($j = 1,2,\ldots,N_r$) and over the M different replicas of the protocol. These quantities are strongly fluctuating in the ensemble of N_r independent runs and it is more convenient to illustrate the behavior of the running averages of their combination

$$[(u+d)/2]_k = \sum_{j=1}^{k} [(u_j + d_j)/2]/k, \qquad (11)$$

where $k = 1,2,\ldots,N_r$, and we will also refer to the ratio of their running averages $[u]_k/[d]_k$.

Figure 6 provides a clear illustration of the efficiency of the PT protocols. We observe the strong influence of the numbers of local moves on the efficiency measure defined in Eq. (11). This verifies again that the choice of the number of local moves, related to the canonical correlation times, is the decisive ingredient of all PT protocols for increasing efficiency. Thus, our tests support the proposal made earlier by Bittner *et al.* [40] and also show that the influence of the selection method is of minor importance. Furthermore, the influence of the APE$_B$ method on the efficiency is also marginal, and we should keep in mind that this method suffers from larger specific heat errors, as illustrated in Fig. 2. The ratio of the corresponding running averages $[u]_k/[d]_k$ is not equal to 1, but depends on several details of the protocols and in particular on the selection of the highest and lowest temperatures. Although it appears that the efficiency is higher when this ratio is close to 1, the differences between the protocols using the same numbers of local moves are rather small.

The above findings raise questions regarding the importance of introducing cluster algorithms for the local moves in the PT schemes, since, as is well known, cluster algorithms have very small dynamical exponents. Therefore, it should be expected that the implementation of such algorithms, for the local moves, will increase the efficiency of the protocols. Naturally, we now consider schemes that use for the local moves two further alternatives, besides the Metropolis algorithm. These two alternatives are (i) the Wolff cluster algorithm (denoted in the figures as W), which is known for its small dynamical exponent, and (ii) the one-spin-flip algorithm known as the n-fold way or BKL algorithm (denoted in the figures as BKL) described in Sec. II B.

In order to simplify our presentation, we consider now only the CAE selection method and we omit the initials CAE from our notation. Thus, the protocols of interest are now denoted by [BKL,(NN)$_a$], [BKL,APE$_B$], [W,(NN)$_a$], and [W,APE$_B$], corresponding to NN exchange and the APE method of Brenner *et al.* [44]. Their behavior is illustrated in the following three figures.

In particular, Fig. 7 illustrates the specific heat error behavior of these four PT protocols. Comparing the two APE$_B$ protocols of Fig. 7 with the APE$_B$ protocols of Fig. 2 which use a local Metropolis algorithm, we observe clear improvements in the behavior of the illustrated error measures. The improvement is substantial in the case of the [W,APE$_B$] protocol and moderate in the case of the [BKL,APE$_B$] protocol. In order to appreciate these improvements we specify the rest of the details of the protocols. The number of independent MC runs is again $N_r = 200$ in all four cases. For the [BKL,(NN)$_a$] and [BKL,APE$_B$] protocols we use again $t_{\rm eq} = 3N$, $t_{\rm av} = 15N$, and $N_{\rm local}(T) = (M-1)n(T) = 0.216N$, which corresponds to $n(T) = 30$ BKL spin flips before each swap move. This choice makes the time requirement of the BKL protocols approximately equivalent to that for the corresponding Metropolis protocols. As mentioned earlier, the APE$_B$ PT protocol requires more time, due to the extra calculations needed for the application of Eqs. (5) and (6), before each swap move. Yet we can achieve approximately the same time requirements for this protocol, by restricting the

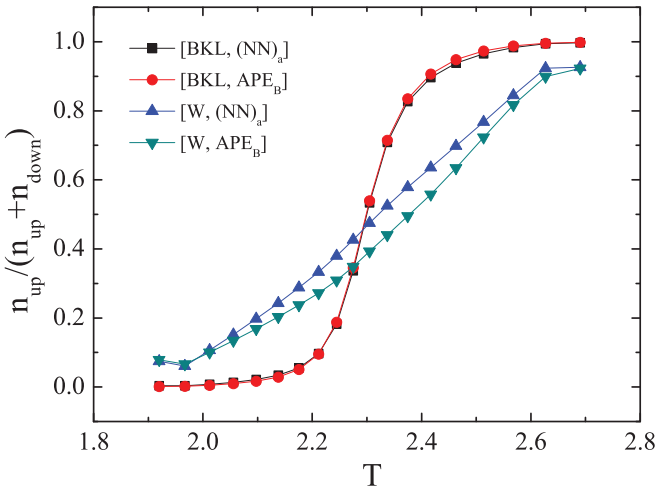

FIG. 8. (Color online) Illustration of the diffusion fraction for the PT protocols shown in the legend. Again, the CAE selection of temperatures, as given in Table I, has been used in all four cases.

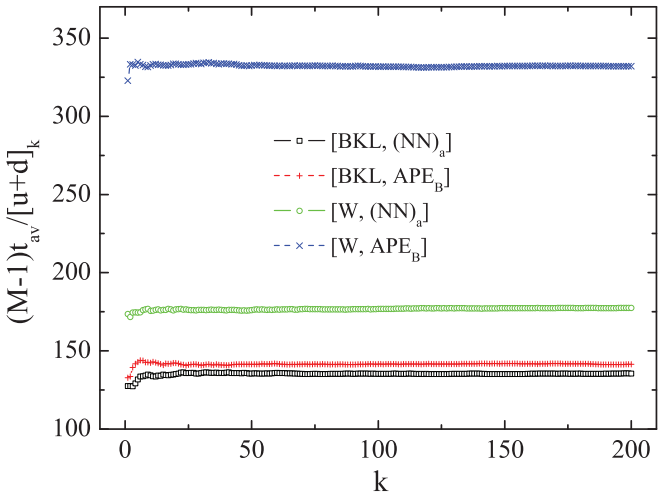

FIG. 9. (Color online) Running averages of the quantity $(M-1)t_{\rm av}/[u+d]_k$, which is the mean number of round trips of a replica during one independent run. The larger running averages of these numbers correspond to the most efficient protocol which is now definitely the [W,APE$_B$] PT scheme.

possible remote exchanges up to fourth order. We note that the sum of probabilities of orders 5, 6, 7, and 8 is only 0.074% and the higher orders never occur. The probabilities of the first four orders of exchanges are 71.01%, 23.17%, 5.03%, and 0.72% for first-, second-, third-, and fourth-neighbor exchanges, respectively. The error behavior and the efficiency of the protocol are very weakly influenced by this restriction, and thus it is more reasonable to use this restriction than to use small parameters for the APE$_B$ protocols. In the cases using the Wolff algorithm for the local moves, we have used $t_{\rm eq} = 3N/50$, $t_{\rm av} = N$, and $N_{\rm local}(T) = (M-1)n(T) = 0.0216N$, which corresponds now to $n(T) = 3$ Wolff cluster flips before each exchange attempt. The above adjustments for the Wolff PT protocols are essentially reflecting the small dynamical exponent of the Wolff algorithm, in an attempt to construct a protocol having approximately the same time requirements as the Metropolis PT protocols.

Figure 8 illustrates the behavior of the diffusion fraction of the four protocols. One can now observe that while the BKL cases behave very similarly to the Metropolis case shown in Fig. 3, the Wolff cases show a behavior resembling the Metropolis protocols using numbers of local moves analogous to the canonical correlation times of Figs. 4 and 5. In fact the changes induced in the behavior of the diffusion fraction are now more spectacular. Furthermore, Fig. 9 provides a demonstration that the combination of the APE$_B$ exchange method with a local Wolff algorithm is now the best choice for increasing the efficiency of the PT scheme. Since the numbers of PTSs used, $t_{\rm av}$, are different for the BKL and Wolff protocols we have plotted in Fig. 9 the running averages of the quantity $(M-1)t_{\rm av}/[u+d]_k$. This defines the mean number of round trips of a replica during one independent run and is larger for the most efficient protocol, which is now clearly the [W,APE$_B$] PT scheme.

Thus, taking also into account the observation that the implementation of the Wolff algorithm removes the problem with the large specific heat errors, we may declare here that, under some conditions, the APE$_B$ PT method may provide a respectable and efficient PT protocol. This appears as a justification of the suggestion of Calvo [43]: the exchange moves could be considered as extra moves that are not directly involved in the averaging process. An explanation for the generally large specific heat errors of the APE schemes may be sought in problems coming from the omission of appropriate residence times and possible strong disturbances caused by remote exchanges enforced by the APE exchange methods. Apparently, these seem to be cured by the very fast restoration of equilibrium that takes place after the Wolff local moves. On the other hand, the moderate improvement of the [BKL,APE$_B$] protocol, compared with the corresponding [M,APE$_B$] in Fig. 2, may be due to the use, at the local level, of appropriate average lifetimes, as is usually done in a BKL implementation [1,46–48]. The above observations provide clues for improving the all-pair exchange methods, such as those proposed by Calvo [43] and Brenner *et al.* [44]. Our tests on the Ising model illustrate some of the merits and weaknesses of these schemes. Of course, the real power of all PT methods should be checked in rare-event problems in which the performance of conventional MC methods can become unreliable.

IV. GROUND STATES OF 3D SPIN-GLASS MODELS

We proceed to test the efficiency of PT protocols for the production of ground states (GSs) in 3D spin-glass models. It is well known that finding the GS of a spin-glass system in $D = 3$ is an NP-complete problem [57,58] and there exist a large number of heuristic algorithms developed in order to tackle this outstanding problem [58–62]. In a recent paper, Roma *et al.* [37] have concluded that PT is comparable to the performance of the more powerful heuristics. In particular, they have concentrated on the estimation of the minimum number of PTSs needed to achieve a true GS with a given probability. In their study they considered the Edwards-Anderson model in 2D and 3D with bimodal and Gaussian

bond disorder. We will consider the 3D EAB model [10,12] and an anisotropic 3D EAB model.

We will first discuss, in Sec. IV A, the standard (isotropic) 3D EAB model and observe the relative performance of some of the presented PT protocols. Then, in Sec. IV B, we shall consider a variant of this model, with spatially uniaxial anisotropic exchange interactions. This model has been studied recently [63], and its phase diagram has been estimated. For this case we will provide results for its finite-size behavior of the GS energy, comparing our estimates with the isotropic case. The anisotropic spin-glass model [63,64], is defined by the Hamiltonian

$$H = -\sum_{u}\sum_{\langle ij\rangle_u} J_{ij}^{u} s_i s_j, \tag{12}$$

where the exchange interactions are uncorrelated quenched random variables, taking the values $\pm J^{xy}$ on the xy planes and the values $\pm J^{z}$ on the z axis. The bimodal distribution of J_{ij}^{u} takes the general form

$$P\left(J_{ij}^{u}\right) = p_u \delta\left(J_{ij}^{u} + J^{u}\right) + (1 - p_u)\delta\left(J_{ij}^{u} - J^{u}\right), \tag{13}$$

where u denotes the z axis ($u = z$) or the xy planes ($u = xy$), J^{u} denotes the corresponding exchange interaction strength, and p_u are the probabilities of two neighboring spins (ij) having antiferromagnetic interaction. The standard isotropic EAB model corresponds to $J^{z} = J^{xy} = J(=1)$ and $p_z = p_{xy}$. In the following we consider the production of GSs for the isotropic ($p_z = p_{xy} = 0.5$) and anisotropic ($p_z = 0, p_{xy} = 0.5$) cases.

A. Ground states of the 3D EAB model: Further tests of PT protocols

In this section, we consider the standard isotropic 3D EAB model on a cubic lattice of linear size $L = 6$. For this model, Roma *et al.* [37] have addressed the question of whether it is more efficient, in order to find a true GS, to use large running times or several independent runs ($t_{\rm av}$ and N_r in our notation) for each realization of the disorder, called henceforth the sample. We note that, due to some differences in defining the PTS, our notation is not identical with that in Ref. [37], but the analogies are obvious and we will also use N_s to denote the number of different samples. The collapse example in Fig. 2(b) of Ref. [37] shows clearly that increasing the number of PTSs ($t_{\rm av}$) has approximately the same effect as an analogous increase in the number of independent runs (N_r) for each sample. Accordingly the production of a true GS depends on the product $N_r t_{\rm av}$. Thus, in what follows we demonstrate the performance of some of the previous PT protocols, using $N_r = 1000$ and $N_r = 100$ for the number of independent runs, and adjusting $t_{\rm av}$ and the other parameters of the protocols in a way that enabled us to compare schemes requiring approximately the same CPU time. We will implement single-spin-flip local algorithms, since an efficient cluster algorithm for the production of GSs in the spin-glass problem is not available. Primarily, we wish here to observe the relative efficiency and dependence of the PT schemes on the local algorithms, on the exchange method used, and also on the selection method of temperatures for the PT process. These issues distinguish our comparative approach from that of Roma *et al.* [37], who concentrated mainly on estimating the running times necessary for generating true GSs with a given probability and not on differentiating among various PT recipes, using different local and swap moves. However, their estimated times have guided our tests, and also our study of the GS energy of the mentioned anisotropic variant of the EAB model, considered in the next section.

In Table II, we give the temperature sequences (T sequences) obtained by averaging the individual T sequences over 50 samples of the 3D EAB model for a lattice size $L = 6$. The individual T sequences were obtained using relatively short runs, by the histogram method outlined in Sec. II A. Sequences for two values of the acceptance rate r of the CAE method, together with the corresponding CEI method sequences, are shown. The T sequences displayed in Table II correspond to the values $r = 0.1$ and $r = 0.5$, involving $M = 5$ and $M = 11$ temperatures, respectively. Since each sample has its own T sequences, the illustrated averaged T sequences will be, in any case, only a rough approximation for each sample. Thus, short PT runs, using an *ad hoc* reasonable set of temperatures, were found to be adequate for each of the 50 samples used to construct the table. Of course, one

TABLE II. Temperature sequences obtained by averaging over $N_s = 50$ samples the individual T sequences. The individual T sequences are obtained by relatively short runs, as described in the text. Sequences for two values of the acceptance rate r of the CAE method, together with the corresponding CEI method sequences, are shown.

CAE($r = 0.1$): $M = 5$		CEI: $M = 5$		CAE($r = 0.5$): $M = 11$		CEI: $M = 11$	
T	r	T	r	T	r	T	r
0.5000	0.097	0.5000	0.033	0.5000	0.481	0.5000	0.211
0.8868	0.103	1.0200	0.130	0.6893	0.486	0.7800	0.382
1.2269	0.112	1.3550	0.222	0.8934	0.498	0.9600	0.511
1.5940	0.105	1.6750	0.144	0.9746	0.498	1.1150	0.489
2.0235		2.0350		1.1137	0.501	1.2550	0.537
				1.2553	0.516	1.3900	0.548
				1.4018	0.501	1.5250	0.519
				1.5560	0.507	1.6600	0.637
				1.7191	0.510	1.8000	0.622
				1.8973	0.498	1.9500	0.552
				2.0950		2.1100	

TABLE III. Sample averages ($[P]$) and minimum ($P_{\min}$) probabilities of generating a true GS for various PT protocols. In each case, the notation indicates the implementation of the Metropolis (M) or the BKL algorithm for the local moves, and the implementation of the $(\mathrm{NN})_a$ or the APE_B method for the PT exchanges. Finally, we have indicated the employed T sequences (CAE or CEI selection method) and give in the table the main parameters of the PT protocols.

CAE($r = 0.1$): $M = 5$	N_r	$n = N_{\mathrm{local}}/(M-1)$	t_{eq}	t_{av}	$[P]$	$P_{\min}$
[CAE,M,$(\mathrm{NN})_a$]	1000	54	108	216	0.8205	0.026
[CAE,M,APE_B]	1000	54	108	216	0.8302	0.024
[CAE,BKL,$(\mathrm{NN})_a$]	1000	16	108	216	0.9324	0.035
[CAE,BKL,APE_B]	1000	16	108	216	0.9398	0.039
CAE($r = 0.1$): $M = 5$	N_r	$n = N_{\mathrm{local}}/(M-1)$	t_{eq}	t_{av}	$[P]$	$P_{\min}$
[CAE,M,$(\mathrm{NN})_a$]	100	54	108	2160	0.9769	0.05
[CAE,M,APE_B]	100	54	108	2160	0.9787	0.04
[CAE,BKL,$(\mathrm{NN})_a$]	100	18	108	2160	0.9921	0.09
[CAE,BKL,APE_B]	100	18	108	2160	0.9929	0.11
CAE($r = 0.5$): $M = 11$	N_r	$n = N_{\mathrm{local}}/(M-1)$	t_{eq}	t_{av}	$[P]$	$P_{\min}$
[CAE,M,$(\mathrm{NN})_a$]	100	21	64	1271	0.9856	0.12
[CEI,M,$(\mathrm{NN})_a$]	100	21	64	1271	0.9818	0.10
[CAE,BKL,$(\mathrm{NN})_a$]	100	5	64	1271	0.9936	0.15
[CEI,BKL,$(\mathrm{NN})_a$]	100	5	64	1271	0.9917	0.14
[CEI,BKL,APE_B]	100	5	64	1271	0.9930	0.15

can improve this approximation by increasing the PT running times (t_{av}) for recording the histograms (of energy and specific heat), but this will produce nonsignificant changes in the final averaged T sequences. However, some defects of this short-time approximation can be observed in the fluctuation of the exchange rate in the CEI method in the second part of the table for the case $r = 0.5$.

We have used a set of $N_s = 1000$ samples, in order to test the performance of several PT protocols. The local algorithms used are the Metropolis (M) and the BKL (n-fold-way) algorithms. The swap moves are carried out by the $(\mathrm{NN})_a$ and the APE_B methods. In Table III we summarize these tests in three groups. The CAE T sequence of $r = 0.1$ with $M = 5$ temperatures has been used in the first and second group of four schemes, while both the CAE and CEI T sequences, corresponding to the case $r = 0.5$ and $M = 11$ temperatures, have been used in the third group of five schemes. In an obvious notation, in each case we give the abbreviation for the T sequence, then the acronym of the local algorithm, and finally the method of PT exchange. Thus, [CAE,M,$(\mathrm{NN})_a$] denotes the PT protocol based on the CAE sequence, using the Metropolis algorithm for the local moves and the $(\mathrm{NN})_a$ method for the swap moves, while [CEI,BKL,APE_B] denotes the PT protocol based on the CEI sequence, using the BKL algorithm for the local moves and the APE_B method for the swap moves.

For a particular sample, a large number of independent PT runs (N_r) is carried out and some of these runs (n_j) successfully find a true GS. Thus, $P_j = n_j/N_r$ is the probability of reaching a GS for the jth sample. This probability will depend on the details of the PT protocols, but is also strongly dependent on the particular sample. It is well known that easy and hard samples exist with very different behavior (see, for instance, Fig. 15 of Ref. [37]), and one expects that in a large ensemble of samples the hardest samples have the smallest P_j. The minimum of this probability $P_{\min}$, in a given ensemble, will therefore give us an indication of the performance of the PT protocol for the hard samples, while the sample average $[P] = \sum_{j=1}^{N_s} P_j/N_s$ reflects the average global performance of the PT protocol. The introduction of these probabilities for the production of true GSs enables us now to discuss the entries of Table III and compare the PT schemes.

For each sample j of the set $N_s = 1000$ samples, used for the construction of Table III, we denote the GS energy per site $u_{6,j}$, indicating also in the notation the lattice size ($L = 6$). The sample average (u_6) of the GS energy per site, for the particular set of samples, was found to be $u_6 = \sum_{j=1}^{N_s} u_{6,j}/N_s = -1.770\,26$. This appears as an exact result, for the chosen set of samples, since true GSs have been found for all samples (with probability almost 1). To verify that true GSs were found for each sample, we have used also runs longer than those in Table III by a factor of 4. Note here that according to Ref. [37], the probability of true GS production for running time $t = 2 \times 10^5$ is of order 0.999, and that the recording times ($N_r t_{\mathrm{av}}$) used for Table III are of the same order. Thus, using the longer runs and observing no difference in the above estimate u_6, we conclude that true GSs have been found for all samples.

As mentioned earlier, in each of the three groups of examples shown in Table III, care has been taken to adjust the PT parameters in a way that corresponds to the same CPU time within each group and reflects the different time requirements of the protocols involving local BKL or Metropolis moves. The CPU times of the second and third groups are approximately the same while the CPU time of the first group is larger by a factor 1.88, mainly because in the first group, the disregarded equilibration part (t_{eq}) is comparable to the recording part (t_{av}) of the protocols. We note here that, in all our implementations, the search for ground states and the rest of the recording processes are carried out only in the recording part (t_{av}) of the runs. The fine adjustments of parameters needed in order to achieve approximately the same CPU time in each group were fixed by short preliminary runs.

From Table III, we observe the superiority of the BKL algorithm compared to the Metropolis algorithm for the local moves. This is reflected, within each group, in the values of the global sample-averaged probability $[P]$, and it is more pronounced in the first group, where the averaged probability is not very close to 1. The superiority of the BKL algorithm is also reflected, within each group, in the values of the minimum probabilities $P_{\min}$ that concern the behavior of the hardest samples. The superiority observed here is not a surprise from an algorithmic point of view, since n-fold-way updates have been the basis for previous attempts in searches for GSs of spin glasses [58,59,65]. The all-pair exchange APE_B method improves rather marginally the production process of GSs. Finally, comparing the CAE and CEI cases in the third group, we notice that the results for the CAE T sequence are slightly better than the corresponding results for the CEI T sequence.

In the first and second groups of Table III only the results corresponding to the CAE T sequence are shown. These CAE results are again better than the corresponding results for the CEI T sequence (not included in the table). The requirement of an exchange rate $r = 0.1$ yields sequences with only $M = 5$ temperatures. In this case, the lowest-temperature exchange rate for the CEI method is too low (see Table II). This could be the source of the superiority of the CAE selection method. However, as illustrated in Table III, even in the case of an exchange rate $r = 0.5$, the CAE results are slightly better. Finally, comparing the probabilities of the first group with the other two groups, one can see the strong dependence of the sample-averaged probability $[P]$ on the running times t_{av}, as should be expected and as pointed out by Roma *et al.* [37].

The main conclusion of this section is the observation that, among several PT recipes tested, the PT protocol using for the local moves the BKL algorithm and a T sequence obtained by the CAE method is superior to the other tested protocols. This conclusion was verified by more tests not presented here for brevity. Furthermore, we have tried to observe the effect of using larger numbers of local moves on the probability measures $P_{\min}$ and $[P]$. In particular the run $[\mathrm{CAE,BKL,(NN)}_a]$, of the second group in Table III, was repeated using the choices $(n = 4 \times 18;\ t_{\mathrm{av}} = 2160)$, $(n = 2 \times 18;\ t_{\mathrm{av}} = 2 \times 2160)$, and $(n = 18;\ t_{\mathrm{av}} = 4 \times 2160)$. Note that these cases are almost equivalent in CPU time. The resulting probabilities were $([P] = 0.996;\ P_{\min} = 0.14)$, $([P] = 0.997;\ P_{\min} = 0.07)$, and $([P] = 0.997;\ P_{\min} = 0.18)$, respectively. It appears that at this level of precision the increase of the numbers of local moves is of minor importance for the production of true GSs of the spin-glass system, or to put it differently, it appears that the production of true GSs depends only on the product nt_{av}. At this point, one should also appreciate that the increase of efficiency by the increase of the numbers of local moves, indicated in previous sections, is compensated by the fact that one can increase the above probabilities by increasing also the number of independent runs (N_r) for each sample, as shown by Roma *et al.* [37].

Finally, we report here that using $[\mathrm{CAE,BKL,(NN)}_a]$ of the second group in Table III we have estimated the sample average of the GS energy per site, for a larger set of $N_s = 10\,000$ samples, to be $u_6 = \sum_{j=1}^{N_s} u_{6,j}/N_s = -1.7711(6)$. This is now comparable with the estimate $u_6 = -1.7714(3)$ given by Roma *et al.* [37] for a sample set with the same number of disorder realizations $(N_s = 10^4)$.

B. Ground state energy of the anisotropic 3D EAB model

We now consider the finite-size behavior of the GS energy of the anisotropic case $(p_z = 0, p_{xy} = 0.5)$. As mentioned earlier, the anisotropic model, $p_z = 0$, $p_{xy} \leqslant \frac{1}{2}$ with $J^z = J^{xy} = J(=1)$, has been studied recently by the present authors, and its phase diagram has been presented in some detail [63]. The irrelevance of the anisotropy for the ferromagnetic-paramagnetic transition was established, and further signs of universality concerning the other two kinds of transition, ferromagnetic–spin-glass and spin-glass–paramagnetic, were pointed out. Of particular interest, for our study here, is the observation [63] that the phase diagram points of the spin-glass–paramagnetic transition for the isotropic $(p_z = p_{xy} = 0.5)$ and the anisotropic $(p_z = 0,\ p_{xy} = 0.5)$ cases are very close or even coincide. However, such a coincidence is a prediction which, although appealing, goes well beyond the general universality question, and cannot be trusted before a formal proof is provided. We will now present results that indicate an analogous situation for the asymptotic limit of the GS energy of these two cases, increasing the interest in a possible equivalence in the asymptotic limit.

TABLE IV. PT parameters and GS energy per site of the 3D anisotropic EAB model. The last column is the difference of GS energies per site between the isotropic, from Table B.3 of Ref. [37] and the present anisotropic case. All runs were carried out by the $[\mathrm{CAE,BKL,(NN)}_a]$ protocol using an initial value $t_{\mathrm{eq}} = 2N$.

L	$(r;M)$	$n = N_{\mathrm{local}}/(M-1)$	N_s	t_{av}	$u_L(\mathrm{ani})$	$u_L - u_L(\mathrm{ani})$
3	(0.5;5)	2	5×10^5	2.7×10^2	−1.7642(3)	0.0925
4	(0.35;5)	4	10^5	6.4×10^2	−1.7703(3)	0.0328
5	(0.2;5)	9	10^5	1.25×10^3	−1.7733(3)	0.0122
6	(0.1;5)	216	2×10^4	3.24×10^3	−1.7762(3)	0.0048
7	(0.15;7)	17	6×10^3	3.43×10^4	−1.7786(3)	0.0014
8	(0.1;7)	102	10^4	4.2×10^4	−1.7803(3)	0.0003
9	(0.1;9)	27	2×10^3	1.2×10^6	−1.7825(3)	0.0001
10	(0.1;11)	120	2.5×10^3	1.2×10^6	−1.7830(2)	0.0000
12	(0.1;13)	43	10^2	10^7	−1.7850(8)	0.0001
14	(0.1;13)	68	10^2	1.4×10^7	−1.7862(8)	0.0004

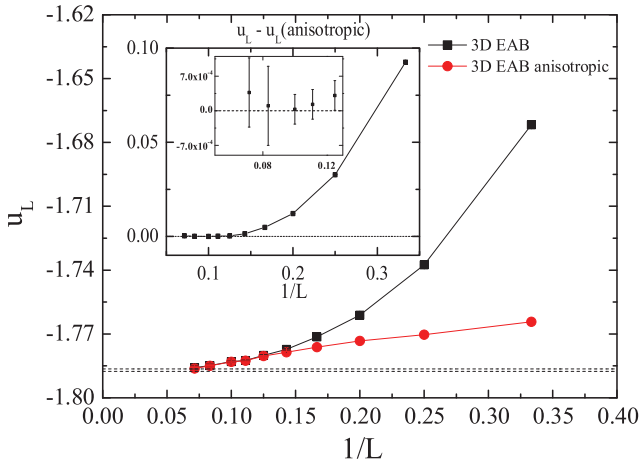

FIG. 10. (Color online) Finite-size behavior of GS energies per site for the 3D EAB model and the present anisotropic model. In the insets we show their difference, which as shown in the nested inset, is much smaller than the estimated errors for $L \geqslant 6$.

Following the approach of [37], we produce here estimates for the finite-size GS energy per site for the anisotropic model. Our simulations cover the range of sizes $L = 3$–14 and we are using the PT protocol [CAE,BKL,$(\mathrm{NN})_a$], which is a simple and efficient choice. In all runs, we choose $N_r = 1$, use a short disregarded part ($t_{\mathrm{eq}} = 2N$), and vary the rest of the PT parameters as shown in Table IV. In particular, the main running times (t_{av}) are almost comparable to those in Table B.3 of Ref. [37]. The temperature range used is approximately the range $T = 0.4$–2.0 and the CAE T sequences were obtained using the practice outlined in the previous section, with an exchange rate and corresponding number of temperatures indicated in the second column of Table IV. For some sizes ($L = 6$, 8, and 10), the numbers of local moves were varied in alternate runs, to test their effect on the estimated GS energies. Our estimates of GS energies for the anisotropic model are given in Table IV. Also, in the last column of the table, we give the differences of GS energies per site between the isotropic model (from Table B.3 of Ref. [37]) and the present anisotropic model.

In Fig. 10, we show the finite-size behavior of the GS energy per site for both isotropic and anisotropic 3D EAB models. In the insets, their differences are illustrated. In particular, it shown that their differences for $L \geqslant 6$ are much smaller than the estimated errors. Therefore, the asymptotic limits of these GS energies practically coincide. The two dashed lines in the main panel indicate previous asymptotic estimations, namely, $u_\infty = -1.7863(4)$ [60] and $-1.7876(3)$ [61].

V. SUMMARY AND CONCLUSIONS

We reviewed several PT schemes and examined their accuracy and efficiency. Our tests on the 2D Ising model suggest that the two different methods of selecting the temperature sequences (CAE and CEI) considered in this paper produce results that are accurate and they are almost equivalent in efficiency. The efficiency of PT protocols is greatly increased by using numbers of local moves related to the canonical correlation times, as proposed earlier by Bittner *et al.* [40]. Accordingly, we found that PT protocols using a Wolff algorithm for the local moves increase the efficiency of the schemes. In particular an all-pair exchange method, the APE_B of Brenner *et al.* [44], when used with local Wolff updates, has been found reasonably accurate and very efficient. However, it was also found that, in general, APE_B protocols may show an unreliable behavior in estimating thermodynamic equilibrium properties, such as the specific heat behavior illustrated in Fig. 2. As argued, this may be related to an improper implementation of the kinetic MC method, which avoids the use of time weights for the MC averaging process.

We also considered the problem of GS production in the 3D EAB model, and we demonstrated the performance and relative efficiency of several PT protocols. We found that PT protocols based on the CAE T sequences appear to be slightly better than those based on the corresponding CEI T sequences. In all our tests, the superiority of the PT protocols involving BKL (or n-fold-way) local updates was firmly established. Finally, we presented evidence that the asymptotic limits of the GS energy of the isotropic ($p_z = p_{xy} = 0.5$) and the anisotropic ($p_z = 0, p_{xy} = 0.5$) EAB models are very close, and possibly coincide. This seems relevant to an analogous interesting behavior found recently for the finite-temperature phase diagram points between spin-glass and paramagnetic phases [63].

ACKNOWLEDGMENT

T.P has been supported by a Ph.D grant of the Special Account of the University of Athens.

[1] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon, Oxford, 1999).
[2] D. P. Landau and K. Binder, *Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2000).
[3] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[4] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
[5] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[6] E. Marinari, G. Parisi, and J. J. Ruiz-Lorenzo, in *Spin Glasses and Random Fields*, edited by A. P. Young (World Scientific, Singapore, 1998), p.59.
[7] E. Marinari, arXiv:cond-mat/9612010v1.
[8] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).
[9] H. G. Katzgraber, arXiv:0905.1629v3.
[10] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, New York, 2001).
[11] K. Binder and W. Kob, *Glassy Materials and Disordered Solids* (World Scientific, Singapore, 2005).

[12] S. F. Edwards and P. W. Anderson, J. Phys. F **5**, 965 (1975).
[13] H. Nishimori, J. Phys. C **13**, 4071 (1980).
[14] H. Nishimori, J. Phys. Soc. Jpn. **55**, 3305 (1986).
[15] Y. Ozeki and H. Nishimori, J. Phys. Soc. Jpn. **56**, 3265 (1987).
[16] P. Le Doussal and A. B. Harris, Phys. Rev. B **40**, 9249 (1989).
[17] R. R. P. Singh, Phys. Rev. Lett. **67**, 899 (1991).
[18] Y. Ozeki and N. Ito, J. Phys. A **31**, 5451 (1998).
[19] M. Palassini and S. Caracciolo, Phys. Rev. Lett. **82**, 5128 (1999).
[20] N. Kawashima and H. Rieger, arXiv:cond-mat/0312432v2.
[21] T. Jörg, Phys. Rev. B **73**, 224431 (2006).
[22] M. Hasenbusch, F. Parisen Toldin, A. Pelissetto, and E. Vicari, Phys. Rev. B **76**, 184202 (2007).
[23] I. A. Campbell, K. Hukushima, and H. Takayama, Phys. Rev. B **76**, 134421 (2007).
[24] M. Hasenbusch, A. Pelissetto, and E. Vicari, Phys. Rev. B **78**, 214205 (2008).
[25] G. Ceccarelli, A. Pelissetto, and E. Vicari, Phys. Rev. B **84**, 134202 (2011).
[26] A. Billoire, L. A. Fernandez, A. Maiorano, E. Marinari, V. Martín-Mayor, and D. Yllanes, J. Stat. Mech.: Theory Exp. (2011) P10019.
[27] K. Hukushima, H. Takayama, and H. Yoshiro, J. Phys. Soc. Jpn. **67**, 12 (1998).
[28] H. G. Ballesteros, A. Cruz, L. A. Fernandez, V. Martin-Mayor, J. Pech, J. J. Ruiz-Lorenzo, A. Tarancon, P. Tellez, C. L. Ullod, and C. Ungil, Phys. Rev. B **62**, 14237 (2000).
[29] H. G. Katzgraber, M. Palassini, and A. P. Young, Phys. Rev. B **63**, 184422 (2001).
[30] H. G. Katzgraber, M. Körner, and A. P. Young, Phys. Rev. B **73**, 224432 (2006).
[31] J. Skolnick and A. Kolinski, Comput. Sci. Eng. **3**, 40 (2001).
[32] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[33] Y. Yamada, Y. Ueda, and Y. Kataoka, J. Comput. Chem. Jpn. **4**, 127 (2005).
[34] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[35] M. Weigel, Phys. Rev. E **76**, 066706 (2007).
[36] J. Machta, Phys. Rev. E **80**, 056706 (2009).
[37] F. Roma, S. Risau-Gusman, A. J. Ramirez-Pastor, F. Nieto, and E. E. Vogel, Physica A **388**, 2821 (2009).
[38] C. E. Fiore, J. Chem. Phys. **135**, 114107 (2011); C. E. Fiore and M. G. E. da Luz, *ibid*. **138**, 014105 (2013); Phys. Rev. Lett. **107**, 230601 (2011).
[39] R. Alvarez Baños, A. Cruz, L. A. Fernandez, J. M. Gil-Narvion, A. Gordillo-Guerrero, M. Guidetti, A. Maiorano, F. Mantovani, E. Marinari, V. Martin-Mayor, J. Monforte-Garcia, A. Muñoz Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, J. J. Ruiz-Lorenzo, S. F. Schifano, B. Seoane, A. Tarancon, R. Tripiccione, and D. Yllanes, J. Stat. Mech.: Theory Exp. (2010) P06026.
[40] E. Bittner, Andreas Nußbaumer, and W. Janke, Phys. Rev. Lett. **101**, 130603 (2008).
[41] D. Sabo, M. Meuwly, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **128**, 174109 (2008).
[42] J. P. Neirotti, F. Calvo, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **112**, 10340 (2000).
[43] F. Calvo, J. Chem. Phys. **123**, 124106 (2005).
[44] P. Brenner, C. R. Sweet, D. VonHandorf, and J. A. Izaguirre, J. Chem. Phys. **126**, 074103 (2007).
[45] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987); U. Wolff, *ibid*. **62**, 361 (1989).
[46] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. **17**, 10 (1975).
[47] B. J. Schulz, K. Binder, and M. Muller, Int. J. Mod. Phys. C **13**, 477 (2001).
[48] A. Malakis, S. S. Martinos, I. A. Hadjiagapiou, and A. S. Peratzakis, Int. J. Mod. Phys. C **15**, 729 (2004).
[49] N. Plattner, J. D. Doll, P. Dupuis, H. Wang, Y. Liu, and J. E. Gubernatis, J. Chem. Phys. **135**, 134111 (2011); **137**, 204112 (2012).
[50] D. A. Kofke, J. Chem. Phys. **117**, 6911 (2002); **120**, 10852 (2004).
[51] C. Predescu, M. Predescu, and C. V. Ciobanu, J. Chem. Phys. **120**, 4119 (2004).
[52] A. Kone and D. A. Kofke, J. Chem. Phys. **122**, 206101 (2005).
[53] H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, J. Stat. Mech.: Theory Exp. (2006) P03018.
[54] P. D. Beale, Phys. Rev. Lett. **76**, 78 (1996).
[55] C. Predescu, M. Predescu, and C. V. Ciobanu, J. Phys. Chem. B **105**, 4189 (2005).
[56] E. Bittner and W. Janke, Phys. Rev. E **84**, 036701 (2011).
[57] J. Houdayer and O. C. Martin, Phys. Rev. E **64**, 056704 (2001).
[58] A. K. Hartmann, and H. Rieger, *New Optimization Algorithms in Physics* (Wiley-VCH, Berlin, 2004).
[59] S. Boettcher and A. G. Percus, Phys. Rev. Lett. **86**, 5211 (2001).
[60] K. F. Pal, Physica A **223**, 283 (1996).
[61] A. K. Hartmann, Europhys. Lett. **40**, 429 (1997).
[62] J. J. Moreno, H. G. Katzgraber, and A. K. Hartmann, Int. J. Mod. Phys. C **14**, 285 (2003); H. G. Katzgraber and A. P. Young, Phys. Rev. B **67**, 134410 (2003); H. G. Katzgraber, M. Körner, F. Liers, M. Jünger, and A. K. Hartmann, *ibid*. **72**, 094421 (2005).
[63] T. Papakonstantinou and A. Malakis, Phys. Rev. E **87**, 012132 (2013).
[64] C. Güven, A. N. Berker, M. Hinczewski, and H. Nishimori, Phys. Rev. E **77**, 061110 (2008).
[65] J. Dall and P. Sibani, Comput. Phys. Commun. **141**, 260 (2001).