Parallel tempering and 3D spin glass models

T Papakonstantinou[1] **and A Malakis**[2]

[1,2] Department of Physics, Section of Solid State Physics, University of Athens, Panepistimiopolis, GR 15784 Zografos, Athens, Greece

E-mail: [1] tpapak@phys.uoa.gr; [2] amalakis@phys.uoa.gr

Abstract. We review parallel tempering schemes and examine their main ingredients for accuracy and efficiency. We discuss two selection methods of temperatures and some alternatives for the exchange of replicas, including all-pair exchange methods. We measure specific heat errors and round-trip efficiency using the two-dimensional (2D) Ising model, and also test the efficiency for the ground state production in 3D spin glass models. We find that the optimization of the GS problem is highly influenced by the choice of the temperature range of the PT process. Finally, we present numerical evidence concerning the universality aspects of an anisotropic case of the 3D spin-glass model.

1. Introduction

Parallel tempering (PT) or replica exchange is a Markov Chain Monte Carlo sampling method [1, 2, 3, 4, 5, 6, 7] that tends to become the method of choice in simulating complex systems where effective potentials have a complicated rugged landscape with many minima and maxima which become more pronounced with increasing system size [8, 9, 10, 11]. This behavior is encountered in spin glasses, where PT is widely used [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35]. Unlike Metropolis algorithm [36], where a single system is sampled in one temperature, in PT several replicas of the system evolve in their individual temperatures by local moves and exchange attempts between replicas are performed periodically (swap moves). This procedure gives PT the ability to roam more freely on the rugged energy landscape of complex systems. In PT, M replicas perform a number local moves, that is a number of MC steps not necessarily the same for each replica in their individual temperatures, and then an exchange (or swap) attempt between replicas is proposed. A successful implementation of PT sampling demands the selection of a suitable method for each part of the procedure and its fine-tuning. Initially one has to determine the temperature sequence of the replicas, then one has to decide which algorithm will be used for the local moves of the replicas, and how many local moves will be attempted for each replica before an exchange attempt. The algorithm for swapping replicas is also a basic ingredient of PT. Swapping algorithms between replicas which are adjacent in temperatures, is the most common practice. These algorithms will be denoted as $(NN)_x$, meaning nearest neighbor exchange according to some ordering specified by the sub-index x. Their alternatives are the all-pair exchange attempts (APE) methods and their recently proposed kinetic versions [37, 38].

In the present paper, we shall consider PT protocols with a variety of different ingredients. We test two methods for the temperature sequences. The widely used constant acceptance exchange (CAE) [6, 7, 39, 40], where the rate of successful swap attempts between replicas is

constant, and the constant entropy increase (CEI) method introduced by Sabo *et al.* [41], where as its name declares requires a constant increase in entropy between successive temperatures. Three different algorithms are used to perform local moves. The Metropolis algorithm, the n-fold or BKL algorithm [42, 43, 44, 45], and the cluster algorithm of Wolff (W) [42, 46, 47]. Some APE and $(NN)_x$ algorithms for the replicas' swapping procedure are implemented and tested. Numerical tests, involving measurements for the specific heat errors and round-trip efficiency, are carried out for the square Ising model. Comparative tests are presented for the ground state problem in 3D spin-glass models. In our final section, we consider an anisotropic bimodal 3D spin glass model and discuss its universality properties. This model is a particular case of a more general (anisotropic) bimodal spin glass model [48, 49], defined by the Hamiltonian

$$H = -\sum_{u} \sum_{\langle ij \rangle_u} J_{ij}^u s_i s_j, \tag{1}$$

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

$$P(J_{ij}^u) = p_u \delta(J_{ij}^u + J^u) + (1 - p_u)\delta(J_{ij}^u - J^u), \tag{2}$$

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 3D Edwards-Anderson bimodal (EAB) model [14, 12], corresponds to $J^z = J^{xy} = J(= 1)$ and $p_z = p_{xy}$. The anisotropic case considered in our studies is the model $p_z = 0; p_{xy} \leq \frac{1}{2}$ with $J^z = J^{xy} = J(= 1)$ [49].

2. Parallel Tempering (PT) Protocols

There is a rather large number of ideas that have been proposed in the last decade for the selection of temperatures in a PT protocol [37, 38, 40, 41, 50, 51, 52, 53, 54]. According to the approach followed by Katzgraber *et al.* [50], optimal temperature selection corresponds to the maximum rate of round trips between low and high temperatures in temperature space and can be obtained using a recursive readjustment of temperatures. Here we will consider two simpler but important methods. The CAE method, when used with appropriate number of sweeps between replica exchanges, was found to 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

$$R(1 \leftrightarrow 2) = \sum_{E_1, E_2} P_{T_1}(E_1) P_{T_2}(E_2) p(E_1, T_1 \leftrightarrow E_2, T_2), \tag{3}$$

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

$$p(E_1, T_1 \leftrightarrow E_2, T_2) = min[1, exp(\Delta\beta \Delta E)], \tag{4}$$

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 $R(1 \leftrightarrow 2) = r$ 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.

The second method requires a constant increase in entropy between successive temperatures [41]. Following Sabo *et al.* [41], we denote the M temperatures of the CEI sequence

by $(T_m; m = 1, ..., M)$ and the total increase in entropy from T_1 to T_M by ΔS. Then the adjacent temperatures are determined starting from the given T_1 successively from

$$\int_{T_m}^{T_{m+1}} dT \frac{C_u(T)}{T} = \frac{\Delta S}{(M-1)}, \tag{5}$$

where the specific heat at any temperature can be calculated also from preliminary MC runs (single replica or PT) and histogram methods [46, 42].

Using such a preliminary PT run, we determine both CAE and CEI T-sequences in a temperature range centered around the pseudo-critical temperature of the specific heat of the square Ising model with linear size $L = 50$ ($N = L^D = L^2$ is the number of lattice sites). This is carried out, in a unified implementation, by appropriate recursive schemes, and then by repeating a Metropolis run at these T-sequences we estimate their canonical correlation times [42]. This practice can be applied to a general system for which the DOS is not known, applying in the first preliminary run (especially in a spin glass system) a PT protocol in an *ad-hoc* reasonable set of temperatures. For the CAE selection, we start from $T_1 = 1.9200$ (we set $J/k_B = 1$) and proceed to find higher temperatures corresponding to $r = 0.5$, reaching after a total number of temperatures $M = 19$ the temperature $T_M = 2.6903$. We also apply the CEI procedure with $M = 19$ between the same $T_1 = 1.9200$ and a final temperature $T_M = 2.6975$. Thus, the two schemes are defined approximately in the same temperature range with the same number of replicas.

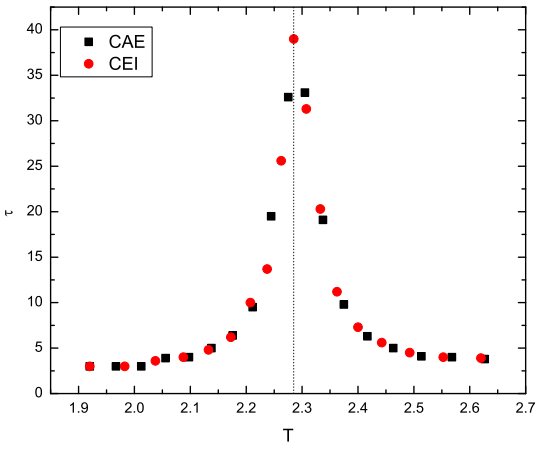

Figure 1. Temperature sequences for the CAE and CEI selection methods. Illustration of the corresponding canonical correlation times. The vertical dotted line indicates the specific heat maximum at $T_C^* = 2.285$ for $L = 50$.

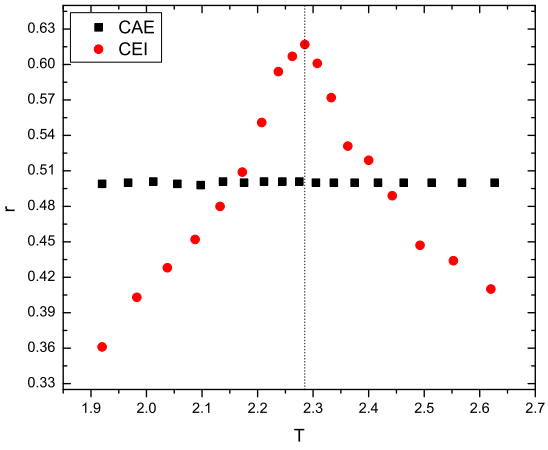

Figure 2. Illustration of the acceptance rates between adjacent replicas for the CAE and CEI selection methods. The vertical dotted line indicates again the specific heat maximum as in Figure 1.

Figure 1 illustrates the two T-sequences obtained according to the above description. In this illustration we display, for the CAE and CEI selection methods, the corresponding temperature sequences with their canonical correlation times, estimated from the discrete form of the energy autocorrelation function [42], in units of lattice sweeps (N Metropolis attempts). Figure 2 illustrates the acceptance rates between adjacent replicas for the two methods. The two methods produce temperature sequences that are more concentrated in the temperature range where the specific heat has a maximum (dotted lines in the figures), with the CEI method producing a more

dense set close to the maximum point, and this is reflected in the variation of the acceptance rates of the CEI method.

We now discuss alternatives for the ordering of the exchange attempts. Mixing local MC attempts with swap attempts is essential in the PT procedure. It is this feature that enables an ergodic walk in temperature space, transferring information between the highest and lowest temperatures. Most PT protocols use NN exchanges, but recently PT protocols using APE methods have also been implemented [37, 38]. In a NN exchange protocol only adjacent exchange attempts are proposed. There are exactly $M-1$ different NN proposals and these may be uniquely denoted by the lowest temperature index $i = 1, ..., M-1$. We have implemented four different algorithms, denoted as $(NN)_a$,$(NN)_b$,$(NN)_c$, and $(NN)_d$, for the ordering of the NN proposals. In $(NN)_a$, a random permutation (say: $j_1, ..., j_{M-1}$) is generated, from the set $i = 1, ..., M-1$, 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 such a swap cycle (SC) may be used in defining the unit of time of an elementary PT step. In $(NN)_b$, the lowest temperature index (i) is randomly chosen from the set $i = 1, ..., M-1$ and therefore multiple exchanges of the same pair are allowed in the swap cycle. In $(NN)_c$, the sequence of proposals is fully deterministic, consisting of two swap sub-cycles, starting from the odd proposals $i = 1, 3, ...$, in increasing order, and continuing with the even proposals $i = 2, 4, ...$ This is a most widespread sequential update algorithm, supposed to speed-up the diffusion process. Finally, in $(NN)_d$, the odd and even sub-sequences are randomly permuted before starting the odd and following with the even swap sub-cycles.

In PT protocols using APE methods, adjacent or not adjacent replicas may be exchanged $(x_i \leftrightarrow x_j)$. In these methods, the number of possible proposals $(A \to B)$ is $M(M-1)/2$, $A = \{x_1, x_2, ..., x_i, ..., x_j, ..., x_M\}$ and $B = \{x_1, x_2, ..., x_j, ..., x_i, ..., x_M\}$. 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. The generation of exchange attempts proceeds with equal probabilities, $P_{gen}(A \to B) = P_{gen}(B \to A) = 1/[M(M-1)/2]$, and the acceptance/rejection rule follows the Metropolis form, with an acceptance rate $P_{acc}(A \to B) = p(x_i \leftrightarrow x_j)$, given by Eq. (4). The swap attempts satisfy the detailed balance condition $P(A)P_{gen}(A \to B)P_{acc}(A \to B) = P(B)P_{gen}(B \to A)P_{acc}(B \to A)$, with $P(A) = \rho(x_1)\rho(x_2)...\rho(x_M)$, the product probability distribution, which becomes stationary with respect to the swap attempts. This is a simple APE method denoted here by APE_M, where the sub-index stands for the above Metropolis form of exchange acceptance rate. A different APE algorithm, denoted by APE_B, is the kinetic method of Brenner *et al.* [38]. The kinetic swapping procedure, involves explicitly only the $K = M(M-1)/2$ transitions of the form $A \to B \neq A$ $(j = 1, 2, ..., K)$ with selection probabilities $P_j = P_{acc}(A \to B)P_{gen}(A \to B)$, and $P_{gen}(A \to B) = 1/max\left\{\sum_{M\in\Phi} P_{acc}(A \to M), \sum_{L\in\Psi} P_{acc}(B \to L)\right\}$. Further details of such methods can be found in the original article [38] and in [55].

A PT step (PTS), is the elementary MC step used for the recording (measuring or averaging) process during an independent MC run. A PTS consists of one or several swap cycles of $M-1$ replica exchange attempts, and we shall be using the same definition, for simplicity, in the case of the APE method also. After each exchange attempt, all replicas perform a number of local moves (spin flips or cluster moves) at their current temperatures. The number of local moves $n(T_i)$, may vary with temperature and the total number of local moves (for any temperature in a SC) is $N_{local}(T) = (M-1)n(T)$. Without loss of generality, we will define the PTS to be just one SC. We may differentiate between various protocols by varying the number of local moves and we also are able to set $N_{local}(T) = f(T)\tau(T)N$, where $\tau(T)$ are the canonical correlation times and $f(T)$ are factors facilitating the adaption of the number of local moves $n(T)$. Note, that $N_{local}(T) = N$, corresponds to the standard lattice sweep. A first (disregarding) equilibration part of the simulation consist of t_{eq} PTSs, and then a large number of PTSs (t_{av}) is used for

recording and averaging. Local MC moves are carried out by the Metropolis algorithm, the n-fold way algorithm and the cluster Wolff algorithm. We also use large numbers of independent MC runs (N_r) to obtain accurate estimates of our measures. In order to compare PT schemes, we also adapt time parameters in a consistent way so as to get approximately the same CPU time.

3. Comparing PT protocols

3.1. Testing PT protocols using the 2D Ising model

We now present some tests on PT protocols, obtained by combining a variety of features mentioned above. We begin with tests carried out on the L=50 square ferromagnetic Ising system with periodic boundary conditions. For this model we have used the exact DOS [56] to calculate the values of the specific heat at the PT protocol temperatures and thus we determine the following error-measures

$$\epsilon(T_i) = [C_{exact}(T_i) - C_{PT}(T_i)]/C_{exact}(T_i), \tag{6}$$

$$\bar{\epsilon} = \sum_{i=1}^{M} \epsilon(T_i)/M, \tag{7}$$

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

$$\epsilon^* = \max(|\epsilon(T_i)|), \tag{9}$$

For $(NN)_x$ exchange algorithms, which use the Metropolis algorithm as a local algorithm, we found that the above specific heat errors are very small and are all of the same order, with no statistically significant difference between them [55]. Thus, hereafter we shall use the $(NN)_a$ PT protocol and vary the other ingredients of the schemes.

Fig. 3, illustrates the errors $\epsilon(T_i)$ of two NN and two APE protocols using both the CEI and CAE selections. For all protocols in Fig. 3 the Metropolis algorithm has been used for the local moves. The number of independent MC runs is $N_r = 200$, with $t_{eq} = 3N$, $t_{av} = 15N$ and $N_{local}(T) = (M-1)n(T) = N$, where $M = 19$ is the number of temperatures. The error-measures, $\bar{\epsilon}$, $\hat{\epsilon}$, and ϵ^*, as defined above in Eq. (7), Eq. (8) and Eq. (9) are given, in parenthesis, on the panels of our figures. As can be seen from this figure, the specific heat errors of the $(NN)_a$ protocols are much smaller than the corresponding errors of the APE protocols. The CAE and CEI selections give comparable accuracy. The APE_B protocol suffers from very large specific heat errors that are more pronounced close to the specific heat maximum. This erratic behavior, is a subtle reflection of the kinetic character of the APE_B method [38]. On the other hand, the APE_M protocol, which is a standard PT protocol also allowing distant (Metropolis) exchange attempts, shows reasonable error behavior.

Let us now discuss PT cases using different local algorithms. Fig. 4 illustrates specific heat error behavior of PT protocols, performing local moves with the Wolff cluster algorithm (W), and the one-spin flip n-fold way or BKL algorithm. Only the CAE selection method has been used, and in an obvious notation, the protocols are denoted by $[BKL, (NN)_a]$, $[BKL, APE_B]$, $[W, (NN)_a]$, and $[W, APE_B]$, corresponding to $(NN)_a$ and the APE_B methods. The number of independent MC runs is $N_r = 200$. The protocols using the BKL algorithm, as local algorithm, use $t_{eq} = 3N$, $t_{av} = 15N$, and a reduced number of local moves $N_{local}(T) = (M-1)n(T) = 0.216N$. This choice corresponds to $n(T) = 30$ BKL spin flips for each replica after a swap move and makes the time requirement of the BKL protocols approximately equivalent to the corresponding Metropolis protocols. For the PT protocols using the Wolff algorithm, we used $t_{eq} = 3N/50$, $t_{av} = N$, and $N_{local}(T) = (M-1)n(T) = 0.0216N$, which corresponds to $n(T) = 3$ Wolff cluster flips after each exchange attempt. The adjustments are now reflecting the small dynamical exponent of the Wolff algorithm, and produce approximately the same time requirements as the PT protocols (of Fig. 3) using Metropolis algorithm.

By comparing the error-measures of the APE_B protocol of Fig. 3, which uses the Metropolis algorithm for the local moves, with the error-measures of the present APE_B protocols of

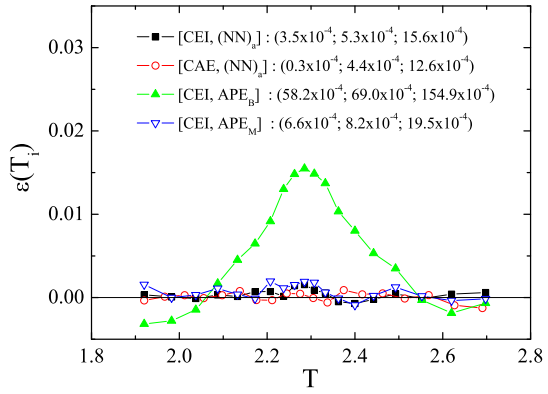

Figure 3. Specific heat errors for four different PT protocols using Metropolis algorithm for local moves. Error-measures, in parenthesis, as defined in Eq. (7), Eq. (8) and Eq. (9).

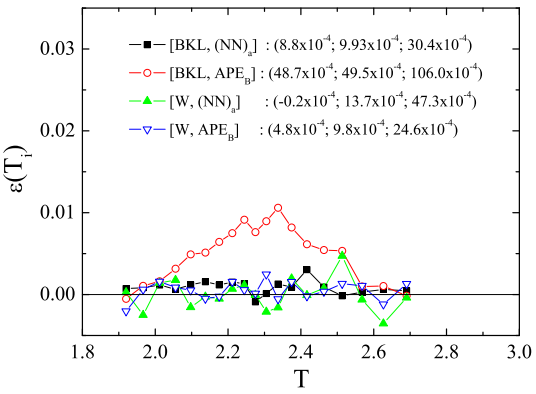

Figure 4. Behavior of PT protocols using n-fold way (BKL) algorithm and the Wolff cluster algorithm (W) for local moves. Only CAE selection of temperatures has been used.

Fig. 4, we observe clear improvement 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.

The efficiency of PT protocols is often characterized by measures indicating difficulties in the flow (bottleneck effects) as a replica moves from the high to low temperature and vise versa. Such a measure is the number of swap moves required to transfer any replica from the highest (lowest) temperature to the lowest (highest), and vise versa. Let the average number of exchange attempts required to transfer any replica from the highest (lowest) temperature to the lowest (highest) be denoted by u_j (d_j) for the corresponding transfer, averaged over the PTSs of a long independent run j ($j = 1, 2, ..., N_r$) and over the M different replicas of the PT protocol. Using these numbers we may define the average round-trip time $(u_j + d_j)$. Since, these quantities are strongly fluctuating (in the ensemble of N_r independent runs) it is more convenient to define and illustrate their behavior using running averages, defined bellow

$$[(u + d)]_k = \sum_{j=1}^{k} (u_j + d_j)/k, \tag{10}$$

where $k = 1, 2, ..., N_r$.

Figure 5 provides an illustration of the running averages of Eq. (10) for several PT protocols. All protocols in Figure 5 are using the Metropolis algorithm for local moves, $t_{eq} = 3N$, and $t_{av} = 15N$. The four protocols in the main panel use local time $N_{local}(T) = N$, while the three PT protocols in the inset use $N_{local}(T) = \tau(T)N$. In the notation, we have included also the numbers of independent runs $N_r = 200$ and $N_r = 20$ respectively as indicated in the legends (all protocols require approximately the same CPU time). By comparing the behavior in the panel and in the inset, we observe a strong influence of the number of local moves on the efficiency measure defined in Eq. (10). This is in accordance with the study of by Bittner *et al.* [40], verifying that a choice of the numbers of local moves, related to the canonical correlation times increases the efficiency of PT protocols. We also point out that the influence of temperature selection method is of minor importance, and that APE_B method yields a noticeable but not significant improvement.

Now, let us consider the behavior of the four PT protocols of Fig. 4, which use the Wolff cluster algorithm (W), and the BKL algorithm for local moves. Since, we have used different

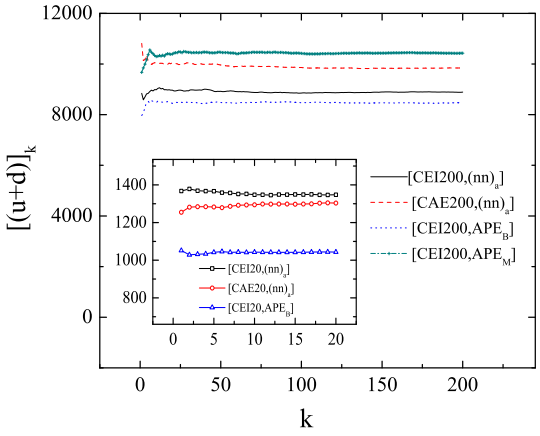

Figure 5. Running averages of round-trip time. Protocols in the main panel use $N_{local}(T) = N$, while in the inset $N_{local}(T) = \tau(T)N$.

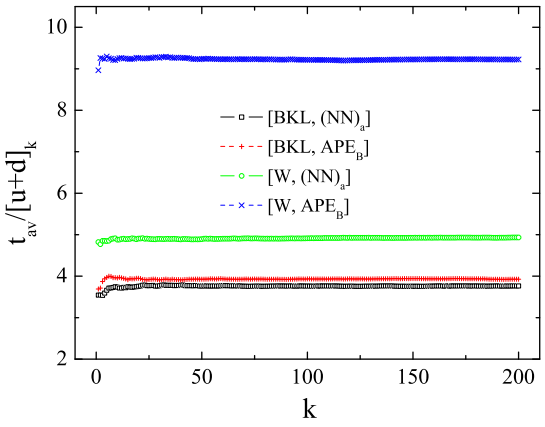

Figure 6. Running averages of $t_{av}/[u+d]_k$ ($\approx$ number of round-trips per replica).

t_{av}, we will now illustrate the behavior of running averages of the mean number of round trips per replica. The total number of swap moves in one independent run is $(M-1)t_{av}$, and thus $(M-1)t_{av}/(M[u+d]_k) \approx t_{av}/[u+d]_k$, measures approximately the mean number of round trips per replica. In Fig. 6, we have plotted the running averages of this quantity for the four PT protocols of Fig. 4, mentioned above. Comparing the two exchange methods we observe that, the APE_B method produces significant improvements in transfer efficiency. In the case of the $[W, APE_B]$ PT protocol, the improvement, compared with the $[W, (NN)_a]$ protocol, is very pronounced indicating also the strong influence of the local algorithm on the APE_B method. Of course, for each protocol the optimum in transfer efficiency will depend on a careful choice of the local times. The fast restoration of equilibrium that takes place after the Wolff local moves may be responsible for the improved behavior of the $[W, APE_B]$ protocol observed in both Fig. 4 and Fig. 6 . Thus, the tests on the Ising model have illustrated some of the merits and weaknesses of the selected PT schemes. However, the real power of PT methods, should be checked in rare-event problems in which the performance of conventional MC methods can become unreliable.

3.2. Testing PT protocols using the 3D EAB spin-glass model

Several studies have been carried out for determining true ground states (GS) of the (isotropic) 3D EAB model [11, 33, 34, 35, 57, 58]. Romá *et al.* [58] 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_{av} and N_r in our notation) for each realization of the disorder, called henceforth sample. The collapse example in Fig. 2b of their article, shows clearly that increasing the number of PTSs (t_{av}), has approximately the same effect with an analogous increase in the number of independent runs (N_r) for each sample. The production of a true GS depends essentially on the product $N_r t_{av}$. They have concentrated mainly on estimating the times necessary for generating true GS with a given probability. We have use this information in constructing the tests of this section, and also in undertaking the GS study for the anisotropic model in the next section.

We implement here two single spin-flip local algorithms, the Metropolis and the BKL algorithm, and try to observe the relative efficiency of the resulting protocols. We are interested in the dependence on the chosen local algorithm, and in the influence of the exchange method.

In order to illustrate the relative efficiency of the resulting PT protocols, we will first try to adjust the PT parameters in an approach requiring approximately the same CPU time for all protocols. This may be attempted by adjusting the local times of the protocols. However, as was pointed out in [55], the implementation of the APE_B method requires additional CPU time. This additional time comes from the involved recalculation of selection probabilities before each swap move. The relative time complexity of the method depends on the ratio between the local and averaging times, and increases considerably when we use a PT protocol with many temperatures. This obvious drawback of the method will be handled by restricting the possible remote exchanges only to significant exchanges (up to forth order), as was explained and implemented in [55]. Time differences, between $(NN)_a$ and APE_B exchange methods, remain considerable, even with the above restriction, in cases where the T-sequences consist of many temperatures. In such cases, we shall ignore the different time requirements of the exchange methods, and compare protocols with the same local time, because otherwise their adjustments may result in using inappropriately small local times for the APE_B method. We observe and compare their performances, albeit this practice favors the APE_B exchange method.

Each sample has its own peculiarities and these are reflected on the T-sequences obtained by a specified temperature selection method. In general, the dependence of T-sequence on the particular samples is very weak, and for our purposes it is more practical to use the same T-sequence for all samples. Thus, we average over T-sequences of a moderate number of $5-10$ samples, and then we use the averaged sequence for all samples included in our tests. Swap moves are attempted by $(NN)_a$, APE_M and APE_B methods. Numerical evidence for the performance of PT protocols, will be presented with the help of the sample averaged probability $[P]$ for producing true GS. This will be obtained from a relatively large set of N_s samples ($N_s = 100$), and from a relatively small set of N_s samples ($N_s = 5$). In the later case, as discussed in more detail bellow, the set is chosen to contain the hardest samples of the larger set. The sample averaged probability is defined by $[P] = \sum_{j=1}^{N_s} P_j/N_s$, where $P_j = n_j/N_r$ is the probability of reaching a GS for the jth sample. The probability P_j of reaching a true GS, is calculated from the n_j successful runs in a total number of independent N_r runs, carried out for each of the N_s samples.

Easy and hard samples exist with very different behavior [58], and in a large ensemble of samples the hardest samples have the smallest P_j. The sample averaged $[P]$, reflects the average performance of the PT protocols for the particular set of samples. If N_s is relatively large, the value of $[P]$ may be assumed to be a fair representative of efficiency measure for each protocol. However, we will be working in a regime of parameters, where for practically all relatively easy samples the probability P_j is ≈ 1.0. Note that, in order to verify that true GSs were found for each sample, we have used independent very long runs, with times longer than those derived by Ref. [58] for obtaining a true GS with probability ≥ 0.999. In the case of a small set of samples, $[P]$ is a less objective measure, but the choice of hardest samples magnifies the differences between the protocols, which are thus more easily observed.

In [55], we have presented comparative tests for the 3D EAB spin-glass, carried out for a cubic lattice with linear size $L = 6$. We found that the results for the CAE T-sequence were slightly better, but not significantly different, from the corresponding results for the CEI T-sequence. Furthermore, the superiority of the BKL algorithm, compared to the Metropolis algorithm, was firmly established. Also, we observed that the all-pair exchange APE_M method deteriorates the production process of GSs, while the APE_B method produced a noticeable improvement. Here, we shall use, in all our tests, a cubic lattice with linear size $L = 8$ and T-sequences only by the CAE selection method.

In our first test, we consider an ensemble of $N_s = 100$ samples and implement a temperature sequence corresponding to a constant acceptance rate $r = 0.1$, producing as explained above, an

averaged sequence of $M = 9$ temperatures in the range $T = 0.4 - 2.5$. The difference between BKL and Metropolis updates is taken into account by adjusting the local time by a factor of the order ≈ 4, in favor of the corresponding Metropolis PT protocols. For the simplest protocol $[M, (NN)_a]$, we set $N_{local}(T) = (M-1)n(T) = N(= L^3)$, which corresponds to $n(T) = 64$ Metropolis spin-flip attempts for all replicas after each swap move. Local time for the other protocols is determined by short preliminary runs. In a particular run with $t_{av} = 5N$ and $N_r = 600$, we used $n(T) = 64$, $n(T) = 18$, and $n(T) = 16$ for the $[M, (NN)_a]$, $[BKL, (NN)_a]$, and $[BKL, APE_B]$ protocols resulting in 1742, 1675 and 1806 CPU minutes respectively. Thus, in our first test we tried to compare PT protocols using approximately the same CPU time by varying t_{av}.

The behavior of the protocols $[M, (NN)_a]$, $[M, APE_B]$, $[BKL, (NN)_a]$, $[BKL, APE_B]$, and $[BKL, APE_M]$, by varying the running time parameter t_{av}, is illustrated in Fig. 7. APE_B method, in the $[BKL, APE_B]$ protocol, produces a marginal improvement in the large t_{av} range, when compared with the corresponding NN protocol $[BKL, (NN)_a]$. On the other hand APE_M method, in the $[BKL, APE_M]$ protocol, seems to worsen the production process of GSs, when compared with the $[BKL, (NN)_a]$ protocol. The superiority of the BKL local algorithm is very clearly revealed by comparing the corresponding PT protocols with those using a Metropolis local algorithm. Noteworthy is also that n-fold way updates have been also the basis of some previous searches for GSs of spin glasses [59, 60, 61]. The picture emerging from the above test is very similar with that discussed in [55]. Nevertheless, APE_B method was found, in our study of the 2D Ising model, to be sensitive to the local time used, as can be seen also in Fig. 5. APE_B may also be sensitive to the spacing of temperatures, in the temperature sequence used, since it is possible that denser sequences may facilitate remote exchanges. We consider these possibilities in the following tests.

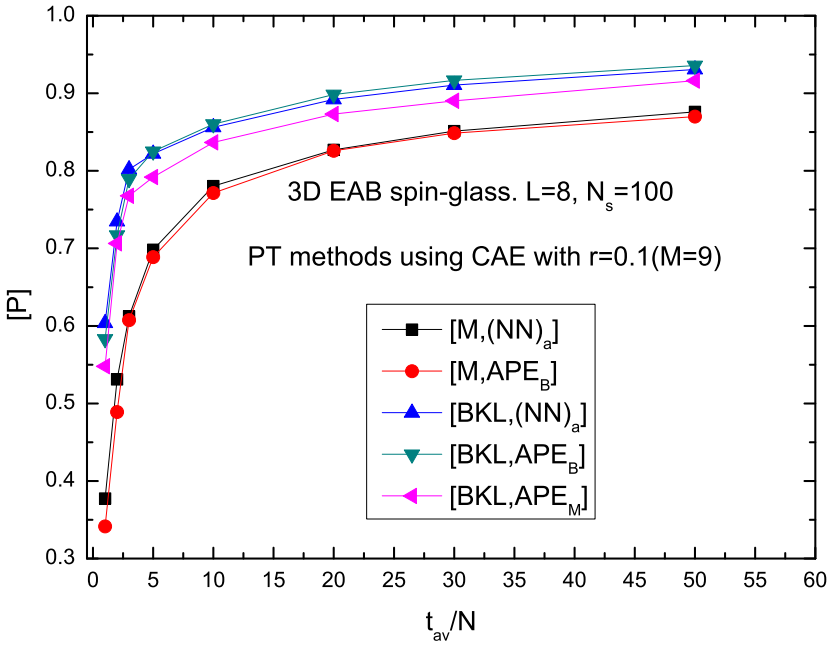

Figure 7. Performance of PT protocols, with local time adjusted to produce approximately the same CPU time.

We now ignore the additional CPU time of the APE_B method, and try to compare it with the $(NN)_a$ method, using exactly the same local time in the corresponding PT protocols. We consider the performance of only the two most efficient protocols $[BKL, (NN)_a]$, and

$[BKL, APE_B]$ protocols and also we implement a more dense T-sequence, the CAE T-sequence with $r = 0.5$, corresponding to $M = 21$ temperatures (approximately in the same temperature range). The dependence of the efficiency of APE_B method on local time is first inspected. We fix the number of PTSs to a value $t_{av} = 10N$, and vary the local time as follows: $N_{local}(T) \approx (0.24N), 5(0.24N), 10(0.24N), 20(0.24N)$. Since $M = 21$, the number of BKL spin-flips for all replicas after each swap attempt is correspondingly $n(T) = 6, 30, 61, 122$. In order to magnify the difference between the two exchange methods, we also restrict our tests only in the set of the five hardest realizations chosen from the original ensemble of $N_s = 100$ samples. Fig. 8, demonstrates that, as we increase local times APE_B method becomes more efficient from $(NN)_a$ method. This is, of course, in agreement with the behavior observed in Fig. 4. APE_B method appears more effective when the exchanging spin configurations are less correlated. However, the actual effectiveness of the APE_B for the GS-problem is still undisclosed, as we shall see bellow. From Fig. 8, APE_B method is already superior to $(NN)_a$ method when $N_{local}(T) = 5(0.24N)$ or $n(T) = 30$. We now fix local time in that value, and vary the running time: $t_{av} = 2N, 5N, 10N, 20N, 30N, 40N$. Fig. 9 shows the performance of the two exchange methods in this test, which we shall call test A. For the set of the five hardest realizations, APE_B method appears again to be more efficient in finding true GS, although we have to take into consideration the fact that the local time chosen favors the APE_B method.

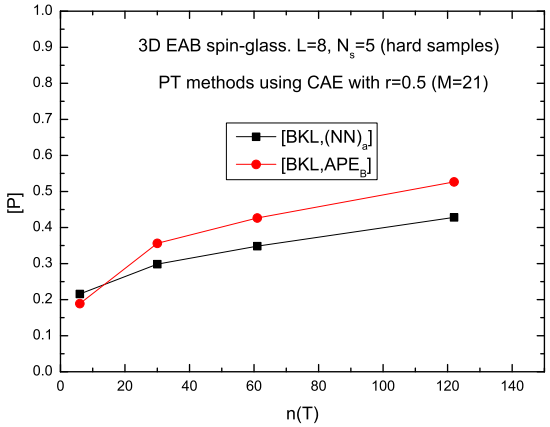

Figure 8. Performance of the exchange methods as we vary local time $n(T)$, with fixed $t_{av} = 10N$.

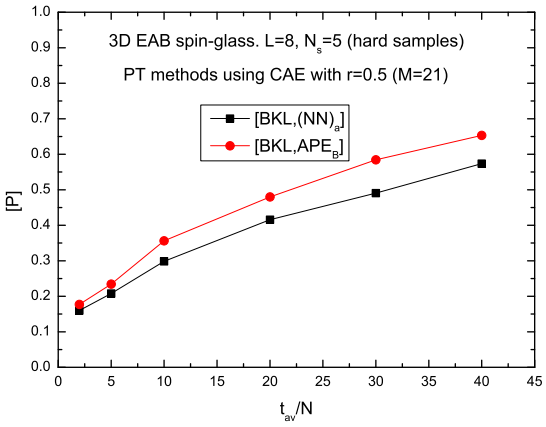

Figure 9. Performance of the exchange methods (test A), with fixed local time $n(T) = 30$.

In the next test, which we call test B, we use again the same CAE T-sequence ($M = 21$), but now we fix the local time to a lower level ($n(T) = 6$) and vary the averaging time: $t_{av} = 10N, 25N, 50N, 100N, 150N, 200N$. A further third test, test C, was undertaken, in which we consider only the lowest $M = 7$ temperatures of the $r = 0.5$ CAE T-sequence, in a T-range: $T = 0.4 - 0.967$. In test C, we use local time $n(T) = 20$ and vary the averaging time: $t_{av} = 10N, 25N, 50N, 100N, 150N, 200N$. Illustration in Fig. 10, contains the performance of the two exchange methods for the three tests in a new x-scale, which is simply the total number of BKL spin-flips divided by N, namely $M(M-1)n(T)t_{av}/N$. This is traditionally called total number of BKL Monte Carlo steps (MCS), in one independent run of the protocols. The new scale facilitates the comparison between points corresponding to the same value on the x-axis, since differences in CPU time are now due only to the additional CPU time of the APE_B method. Both exchange methods, $(NN)_a$ and APE_B, in tests B and C, produce comparable results for the probability $[P]$ of the selected five hardest samples, and therefore one may insist in using the simplest and faster $(NN)_a$ approach. Both tests B and C give higher values of $[P]$

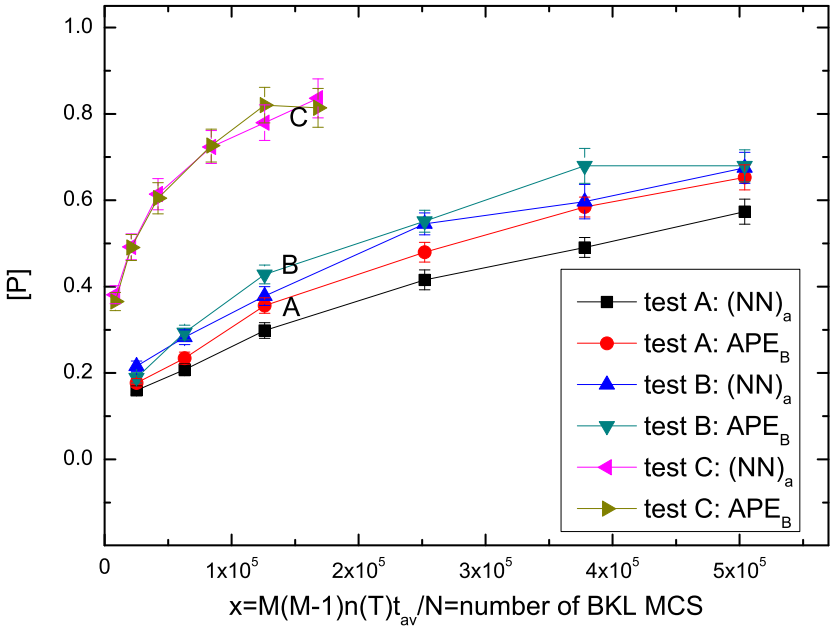

Figure 10. Performance of exchange methods for tests A, B and C. All protocols use BKL updates. Test C involves the lowest $M = 7$ temperatures of the $r = 0.5$ CAE T-sequence with $M = 21$ temperatures.

than those of test A, and the previous minor superiority of the APE_B method in Fig. 9 appears now as an insignificant incidence of the particular choice of local time. It is very interesting that C test, is by far the most efficient in producing true GS with much larger probabilities $[P]$.

To summarize, considering also the CPU time differences of the exchange methods, we observe that test A, in which the APE_B method appears to be a moderate winner, is not an optimum choice for the production of GS. In the other two tests, the $(NN)_a$ and APE_B exchange methods produce comparable results. The search for true GS depends on several parameters of the PT schemes, including the above mentioned ingredients, the number N_r of independent runs and, of course, the temperature range used. A comprehensive study and critical examination of all these factors appears necessary before optimality is claimed. Careful choice of the temperature range emerges now as one of the most influential features for the optimization of the GS-problem, which we believe that has not been recognized in previous studies.

4. Universality aspects of a 3D anisotropic spin-glass model

Recently, the spatially uniaxial anisotropic 3D EAB model $p_z = 0; p_{xy} \leq \frac{1}{2}$ with $J^z = J^{xy} = J(= 1)$, has been studied and its phase diagram has been estimated [49]. The phase diagrams for both the isotropic Edwards-Anderson model and the present anisotropic model are reproduced here in Fig. 11. The vertical solid line in this figure, indicates a ferromagnetic-paramagnetic (F-P) phase transition for $(p_z = 0; p_{xy} = 0.176)$. For this case, a detailed finite size scaling analysis (FSS) was carried out in [49], and Fig. 12, illustrates the behavior of effective exponents for the correlation length exponent at this point. The relevant estimation gives $\nu = 0.683(3)$, which is in excellent agreement with the estimate $\nu = 0.6837(53)$ of the extensive numerical investigations of Ballesteros *et al.* [62] for the site-diluted Ising model. This and similar findings for the other critical exponents have demonstrated very clearly the universality for the ferromagnetic-paramagnetic transitions, between the standard isotropic EAB model and the present anisotropic model [49].

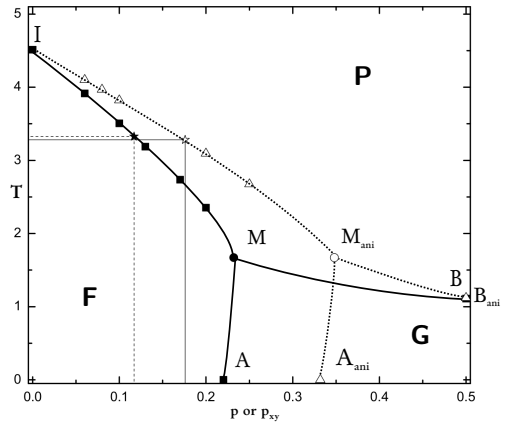

Figure 11. Phase diagrams for the isotropic Edwards-Anderson model (solid lines and full symbols) and the present anisotropic model $p_z = 0, p_{xy}$ (dash lines and open symbols) [49].

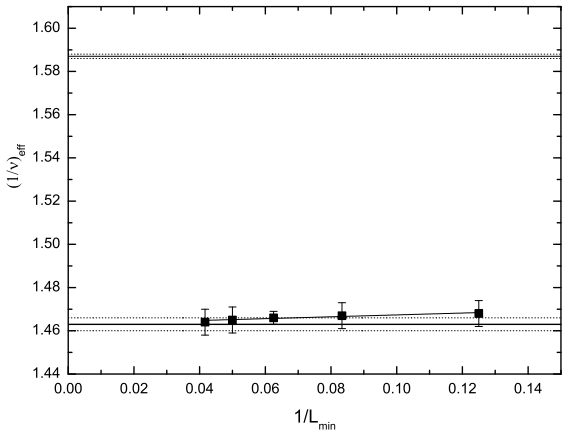

Figure 12. Effective exponents $(1/\nu)_{eff}$ for the anisotropic model, giving by linear extrapolation $1/\nu = 1.463(3)$ [49]. The estimation of the pure 3d Ising model [63] is also shown for comparison.

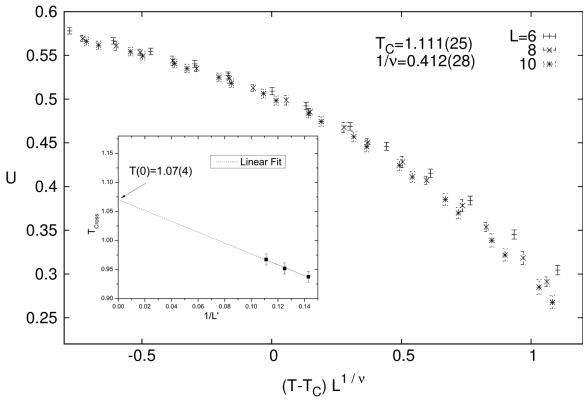

Figure 13. Two methods for the estimation of the phase diagram point B_{ani}. A linear extrapolation of crossings points of the Binder's fourth order cumulant, and a collapse attempt on these cumulants.

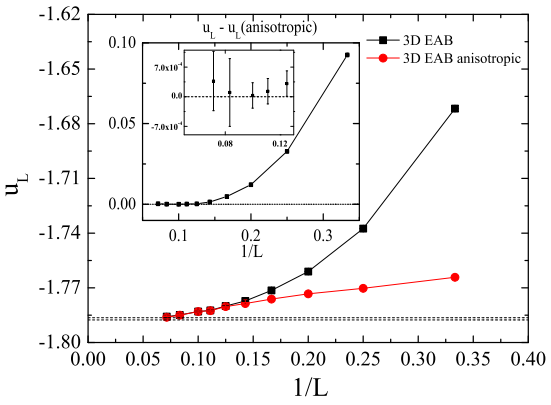

Figure 14. Finite-size behavior of GS energies per site for the 3D EAB model and the present anisotropic model. Their differences are smaller than the estimated errors.

The phase diagram point B_{ani} (open triangle on the phase diagram) corresponding to $(p_z = 0, p_{xy} = \frac{1}{2})$, appears to coincide with the isotopic phase diagram point B corresponding to $(p_z = p_{xy} = \frac{1}{2})$. Attempts to estimate the temperature of the B_{ani} point appear in Fig. 13. The numerical evidence presented is based on data for the Binder's fourth order cumulant $(U = 1 - \frac{1}{3}\left(\frac{[\langle q^4\rangle]}{[\langle q^2\rangle]^2}\right)$ where $\langle \dots \rangle$ and $[\dots]$ denote thermal and disorder average respectively and q the spin-glass overlap order parameter $(q = \frac{1}{N}\sum_{i=1}^{N} s_i^{\alpha} s_i^{\beta}$, where s_i denotes the spin of site i and "α" and "β" represent two replicas of the same disorder realization). The inset illustrates a linear extrapolation of the crossings points of Binder's fourth order cumulants and gives the estimate $T_{B_{ani}} = 1.07(4)$. In the main panel of Fig. 13, a data collapse for these cumulants provide the estimate $T_{B_{ani}} = 1.111(25)$, very close to the corresponding estimate

of the isotropic case $T_B = 1.109(10)$ [26]. The critical exponent $1/\nu$ in the panel of Fig. 13 corresponds to the phase diagram point B_{ani}. This gives for the correlation length exponent an estimate $\nu = 2.43(15)$, in fair agreement for the exponent estimate obtained for the isotropic case (phase diagram point B) $\nu = 2.39(5)$ given by Katzgraber et al. [32].

Finally, we consider the problem of ground states for the anisotropic model at $(p_z = 0, p_{xy} = 0.5)$. Following the approach of [58], we produced estimates for the finite-size GS energy per site for the anisotropic model. Our simulations cover the range of sizes $L = 3 - 14$, using the PT protocol $[CAE, BKL, (NN)_a]$. In all runs, we choose $N_r = 1$, use a short disregarding part $(t_{eq} = 2N)$ and vary the rest of the PT parameters in a way that the main running times (t_{av}) are comparable to those in Table B.3 of [58]. Fig. 14, presents 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 $L \geq 6$ are much smaller than the estimated errors. Therefore, the asymptotic limit of these GS energies practically coincide. The two dashed lines in the main panel indicate previous asymptotic estimations, namely $u_\infty = -1.7863(4)$ [64] and $-1.7876(3)$ [65].

5. 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 methods of selecting temperature sequences (CAE and CEI), produce results that are accurate and are almost equivalent in efficiency. Efficiency of PT protocols is greatly increased by using number of local moves related to the canonical correlation time. Wolff algorithm, when used for the local moves, greatly increases the efficiency of the PT protocols. In particular, APE_B method of Brenner *et al.* [38], when used with local Wolff updates, was found both accurate and very efficient. Using the problem of GS production of the 3D EAB model, we carried out a further comparative study of PT protocols. Various implementations of the APE methods were compared with protocols based on the $(NN)_a$ method, and a critical discussion on the routes for improving performance was carried out. We also found that, a careful choice of the temperature range used for the PT is one of the most influential features for the optimization of the GS-production.

Finally, we presented numerical evidence concerning some universality aspects of an anisotropic case of the 3D spin-glass model. The emerging picture, from the illustrated universality for the ferromagnetic-paramagnetic and spin-glass-paramagnetic transitions, between the standard isotropic EAB model and the anisotropic model, supports the view that the universality between the two models is a general feature covering the global phase diagram of Fig. 11. The finite temperature phase diagram points between spin-glass and paramagnetic phases appear to be very close, and we also presented numerical evidence that the asymptotic limit 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 also very close, and possibly coincide.

Acknowledgments

This work was supported by the special Account for Research of the University of Athens.

References

[1] Landau D P and Binder K 2000 *Monte Carlo Simulations in Statistical Physics* (Cambridge: Cambridge University Press)
[2] Swendsen R H and Wang J-S 1986 *Phys. Rev. Lett.* **57** 2607
[3] Hukushima K and Nemoto K 1996 *J. Phys. Soc. Jpn.* **65** 1604
[4] Marinari E, Parisi G and Ruiz-Lorenzo J J 1998 *Spin Glasses and Random Fields*, ed A P Young (Singapore: World Scientific) p 59
[5] Marinari E 1996 arXiv:cond-mat/9612010v1
[6] Earl D J and Deem M W 2005 *Phys. Chem. Chem. Phys* **7** 3910

[7] Katzgraber H G 2009 arXiv:0905.1629v3
[8] Fiore C E 2011 *J. Chem. Phys.* **135** 114107
[9] Fiore C E and da Luz M G E 2011 *Phys. Rev. Lett.* **107** 230601
[10] Fiore C E and da Luz M G E 2013 *J. Chem. Phys.* **138** 014105
[11] Alvarez Baños R et al. 2010 *J. Stat. Mech.* P06026
[12] Nishimori H 2001 *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (New York: Oxford University Press)
[13] Binder K and Kob W 2005 *Glassy Materials and Disordered Solids* (Singapore: World Scientific)
[14] Edwards S F and Anderson P W 1975 *J. Phys.* F **5** 965
[15] Nishimori H 1980 *J. Phys.* C **13** 4071
[16] Nishimori H 1986 *J. Phys. Soc. Japan* **55** 3305
[17] Ozeki Y and Nishimori H 1987 *J. Phy. Soc. Japan* **56** 3265
[18] Le Doussal P and Harris A B 1989 *Phys. Rev.* B **40** 9249
[19] Singh R R P 1991 *Phys. Rev. Lett.* **67** 899
[20] Ozeki Y and Ito N 1998 *J. Phys.* A **31** 5451
[21] Palassini M and Caracciolo S 1999 *Phys. Rev. Lett.* **82** 5128
[22] Kawashima N and Rieger H 2003 arXiv:cond-mat/0312432v2
[23] Jörg T 2006 *Phys. Rev.* B **73** 224431
[24] Hasenbusch M, Parisen Toldin F, Pelissetto A and Vicari E 2007 *Phys. Rev.* B **76** 184202
[25] Campbell I A, Hukushima K and Takayama H 2007 *Phys. Rev.* B **76** 134421
[26] Hasenbusch M, Pelissetto A and Vicari E 2008 *Phys. Rev.* B **78** 214205
[27] Ceccarelli G, Pelissetto A and Vicari E 2011 *Phys. Rev.* B **84** 134202
[28] Billoire A, Fernandez L A, Maiorano A, Marinari E, Martín-Mayor V and Yllanes D 2011 *J. Stat. Mech.* P10019
[29] Hukushima K, Takayama H and Yoshiro H. 1998 *J. Phys. Soc. Jpn.* **67** 12
[30] Ballesteros H G, Cruz A, Fernandez L A, Martín-Mayor V, Pech J, Ruiz-Lorenzo J J, Tarancon A, Tellez P, Ullod C L and Ungil C 2000 *Phys. Rev.* B **62** 14237
[31] Katzgraber H G, Palassini M and Young A P 2001 *Phys. Rev.* B **63** 184422
[32] Katzgraber H G, Körner M and Young A P *Phys. Rev.* B **73** 224432
[33] Moreno J J, Katzgraber H G and Hartmann A K 2003 *International Journal of Modern Physics* C **14** 285-302
[34] Katzgraber H G and Young A P 2003 *Phys. Rev.* B **67** 134410
[35] Katzgraber H G, Körner M, Liers F, Jünger M and Hartmann A K 2005 *Phys. Rev.* B **72** 094421
[36] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[37] Calvo F 2005 *J. Chem. Phys.* **123** 124106
[38] Brenner P, Sweet C R, VonHandorf D and Izaguirre J A 2007 *J. Chem. Phys.* **126** 074103
[39] Sugita Y and Okamoto Y 1999 *Chem. Phys. Lett.* **314** 141
[40] Bittner E, Nußbaumer A and Janke W 2008 *Phys. Rev. Lett.* **101** 130603
[41] Sabo D, Meuwly M, Freeman D L and Doll J D 2008 *J. Chem. Phys.* **128** 174109
[42] Newman M E J and Barkema G T 1999 *Monte Carlo Methods in Statistical Physics* (Oxford: Clarendon)
[43] Bortz A B, Kalos M H and Lebowitz J L 1975 *J. Comput. Phys.* **17** 10
[44] Schulz B J, Binder K and Muller M 2001 *Int. J. Mod. Phys.* C **13** 477
[45] Malakis A, Martinos S S, Hadjiagapiou I A and Peratzakis A S 2004 *Int. J. Mod. Phys.* C **15** 729
[46] Swendsen R H and Wang J-S 1987 *Phys. Rev. Lett.* **58** 86
[47] Wolff U 1989 *Phys. Rev. Lett.* **62** 361
[48] Güven C, Berker A N, Hinczewski M and Nishimori H 2008 *Phys. Rev.* E **77** 061110
[49] Papakonstantinou T and Malakis A 2013 *Phys. Rev.* E **87** 012132
[50] Katzgraber H G, Trebst S, Huse D A and Troyer M 2006 *J. Stat. Mech.: Theory Exp.* p03018
[51] Kofke D A 2002 *J. Chem. Phys.* **117** 6911; 2004 *J. Chem. Phys.* **120** 10852
[52] Kone A and Kofke D A 2005 *J. Chem. Phys.* **122** 206101
[53] Predescu C, Predescu M and Ciobanu C V 2004 *J. Chem. Phys.* **120** 4119
[54] Neirotti J P, Calvo F, Freeman D L and Doll J D 2000 *J. Chem. Phys.* **112** 10340
[55] Malakis A and Papakonstantinou T 2013 *Phys. Rev.* E **88** 013312
[56] Beale P D 1996 *Phys. Rev. Lett.* **76** 78
[57] Weigel M *Phys. Rev.* E **76** 066706
[58] Romá F, Risau-Gusman S, Ramìrez-Pastor A J, Nieto F and Vogel E 2009 *Physica* A **388** 2821
[59] Hartmann A K and Rieger H 2004 *New Optimization Algorithms in Physics* (Berlin: Wiley-VCH)
[60] Boettcher S and Percus A G 2001 *Phys. Rev. Lett.* **86**, 5211
[61] Dall J and Sibani P *Computer Physics Communications* **141** 260
[62] Ballesteros H G, Fernández L A, Martín-Mayor V, Muñoz Sudupe A, Parisi G and Ruiz-Lorenzo J J 1998

Phys. Rev. B **58** 2740
[63] Hasenbusch M *Phys. Rev.* B **82** 174433
[64] Pal K F 1996 *Physica* A **223** 283
[65] Hartmann A K 1997 *Europhys. Lett.* **40** 429