PROCEEDINGS OF SCIENCE

Tune the autocorrelation time and unleash the full power of the parallel tempering algorithm*

Elmar Bittner†, Andreas Nußbaumer, and Wolfhard Janke
Institut für Theoretische Physik, Universität Leipzig,
Postfach 100 920, D-04009 Leipzig, Germany
E-mail: elmar.bittner@itp.uni-leipzig.de

We introduce a new update scheme to systematically improve the efficiency of parallel tempering Monte Carlo simulations by taking into account the temperature dependence of autocorrelation times. In contrast to previous attempts the temperatures are not dynamically adjusted but chosen in such a way that the acceptance rate for proposed exchanges of all adjacent replica is about 50%. We show that by adapting the number of sweeps between the parallel tempering moves to the canonical autocorrelation time, the average round-trip time of a replica between the lowest and the highest temperatures is significantly decreased and, therefore, the efficiency of the parallel tempering algorithm is considerably improved. We illustrate the new algorithm with results for the two-dimensional Ising model and propose a toy model to find the optimal parameter set for the parallel tempering routine within no time.

The XXVII International Symposium on Lattice Field Theory
July 26-31, 2009
Peking University, Beijing, China

*Work supported by the Deutsche Forschungsgemeinschaft (DFG) under grant Nos. JA483/22-1 and JA483/23-1, the EU RTN-Network 'ENRAGE': *Random Geometry and Random Matrices: From Quantum Gravity to Econophysics* under grant No. MRTN-CT-2004-005616, and by the computer-time grant No. hlz10 of the John von Neumann Institute for Computing (NIC), Forschungszentrum Jülich. EB thanks the DAAD for a travel grant.

†Speaker.

PoS(LAT2009)027

The parallel tempering (PT), or replica exchange, simulation technique [1, 2, 3, 4] provides an efficient method to investigate a broad temperature range in a very effective way in only one simulation. PT and its extensions are used in many disciplines, e.g. biomolecules [5], bioinformatics [6], zeolite structure solution [7], classical and quantum frustrated spin systems [8], spin glasses [3, 4, 9] and QCD [10]. Also, the use of PT in interdisciplinary fields spanning physics, chemistry, biology, engineering and material sciences rapidly increases.

In a PT simulation, one generates many replica of Monte Carlo (MC) Markov chains or molecular dynamics (MD) trajectories at different temperatures in parallel. At regular intervals an attempt is made to exchange the configurations of different, usually adjacent replica, which is accepted with probability

$$P_{\mathrm{PT}}(E_1,\beta_1 \to E_2,\beta_2) = \min[1,\exp(\Delta\beta\Delta E)], \tag{1}$$

where $\Delta\beta = \beta_2 - \beta_1$ is the difference between the inverse temperatures of the two replica and $\Delta E = E_2 - E_1$ their energy difference. The acceptance probability is the smaller the larger the temperature difference or the system size gets. For PT simulations to be most efficient, each replica should spend the same amount of time at each temperature. To this end, several strategies have been proposed in the last years [11], but an efficient selection of optimal temperature intervals is still an open problem.

We employ the concept of a constant acceptance rate between adjacent replica, which can be calculated from

$$A(1 \to 2) = \sum_{E_1,E_2} P_{\beta_1}(E_1) P_{\beta_2}(E_2) P_{\mathrm{PT}}(E_1,\beta_1 \to E_2,\beta_2), \tag{2}$$

where $P_{\beta_i}(E_i)$ is the probability for replica i at β_i to have the energy E_i. Using this formula we can calculate, starting from β_1, a set of inverse temperatures which satisfy $A(i \to i+1) = \text{const}$. For systems with a diverging specific heat one obtains a high density of replica around the critical temperature, i.e., the difference between the inverse temperature of two adjacent replica is small. For high values of β, i.e. low temperatures, the difference between energy distributions at different temperatures becomes small and therefore $\Delta\beta$ increases. Furthermore, for small β values, ΔE decreases and the spacing between the replica grows.

As a basic example, we shall consider MC simulations of the two-dimensional (2D) Ising model where the density of states and hence (2) can be calculated exactly [12]. For all reasonably chosen rates $A(1 \to 2)$, the replica flow from high to low temperatures and vice versa turns out to be very slow, at least when a local update scheme, e.g. the Metropolis algorithm, is used for each of the replica. The replica flow through the temperature space shows a significant drop around the critical temperature. In Fig. 1 we show as an example for an acceptance rate of 50% the fraction of replica which have visited most recently the smallest β value and wander "up" in the inverse temperature space.

We want to remove the unwanted behavior at β_c, while keeping the temperatures fixed at their initial values. Looking at the trajectory of an arbitrarily chosen replica in temperature space shown in the upper plot of Fig. 2, we see a clear block structure, where the border of the blocks coincides with the critical temperature. Such a block structure is related to a bottleneck in the flow through the temperature space, or, in other words, for a replica starting from a high temperature it is hard to overcome this bottleneck and move to the low-temperature region. A plausible explanation of this

observation is as follows: toward the critical temperature the autocorrelation time increases due to critical slowing down and therefore two exchanged replica stay in phase space close to each other. It is hence more likely that these two replica exchange again.

To verify this idea, we use a toy model based on the bivariate Gaussian process with parameter $0 \leq \rho < 1$ [13],

$$e_i = \rho e_{i-1} + \sqrt{1-\rho^2}\, e'_i \,, \quad i \geq 1 \,, \tag{3}$$

where $e_0 = e'_0$, and the e'_i are *independent* Gaussian random variables satisfying $\langle e'_i \rangle = 0$ and $\langle e'_i e'_j \rangle - \langle e'_i \rangle^2 = \delta_{ij}$. Iterating this recursion it follows that the autocorrelation function is

$$A(k) = \langle e_0 e_k \rangle = \rho^k \equiv e^{-k/\tau_{\text{exp}}} \,, \tag{4}$$

where $\tau_{\text{exp}} = -1/\ln\rho$ is the exponential autocorrelation time. It can be shown that with increasing τ_{exp} the mean step size decreases, i.e., $\langle |e_{i+1} - e_i| \rangle = \frac{2}{\sqrt{\pi}}\sqrt{1-\rho}$, such that the system moves slower through the one-dimensional phase space, and this is what we are interested in.

The visual appearance of uncorrelated and correlated data with $\tau_{\text{exp}} = 100$ is depicted in the left plot of Fig. 3 where in each case the first 1000 consecutive measurements according to rule (3) are shown. Despite the quite distinct temporal evolutions, histogramming the total time series of 1 000 000 measurements leads to the same Gaussian distribution within error bars, as it should, cf. the right plot of Fig. 3.

Using the stochastic process (3), we are able to approximate for any realistic model the movement in energy space during a PT simulation. From the energy distribution of initial canonical simulations we obtain for each of the replica at β_i the mean and variance which, after a trivial shift and rescaling, can be reproduced with (3). Next we exploit the freedom in the model to adjust τ_{exp} for each temperature which allows us to investigate the dependence of the flow through

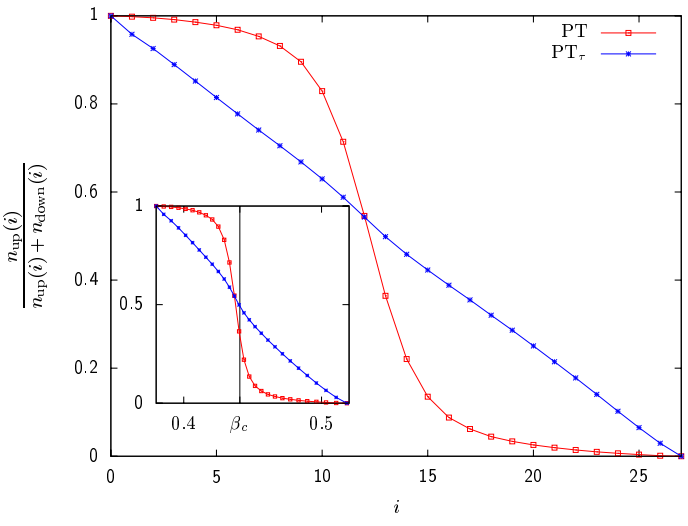

Figure 1: Fraction of replica which wander from the smallest β to the largest as a function of the replica index i for the 2D Ising model ($L = 80$). The simulations without optimization (PT) exhibit a sharp decline close to β_c, as one can see in the inset. Taking the canonical correlation times τ_{can} into account (PT$_\tau$), the fraction decreases, for the same set of temperatures, almost linearly.

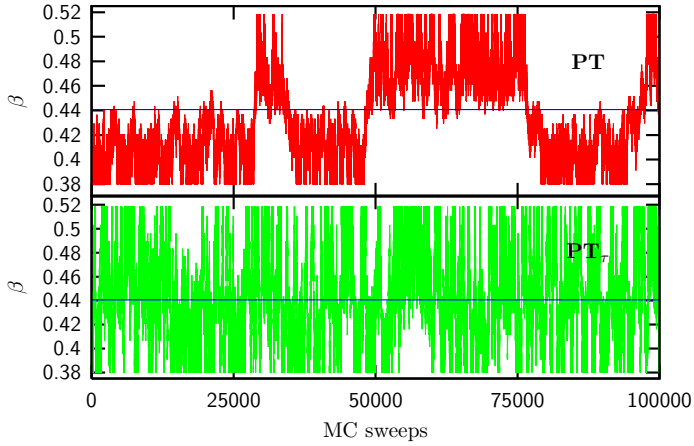

Figure 2: Time series of an arbitrarily chosen replica on its way through inverse temperature space of the 2D Ising model ($L = 80$). The upper plot shows the result of a PT simulation and the lower that of a PT$_\tau$ simulation with $N_{\text{local}} = \tau_{\text{can}}$. The horizontal lines indicate β_c, the infinite volume critical point. The blocks in the time series are a signal of the increasing autocorrelation times due to critical slowing down.

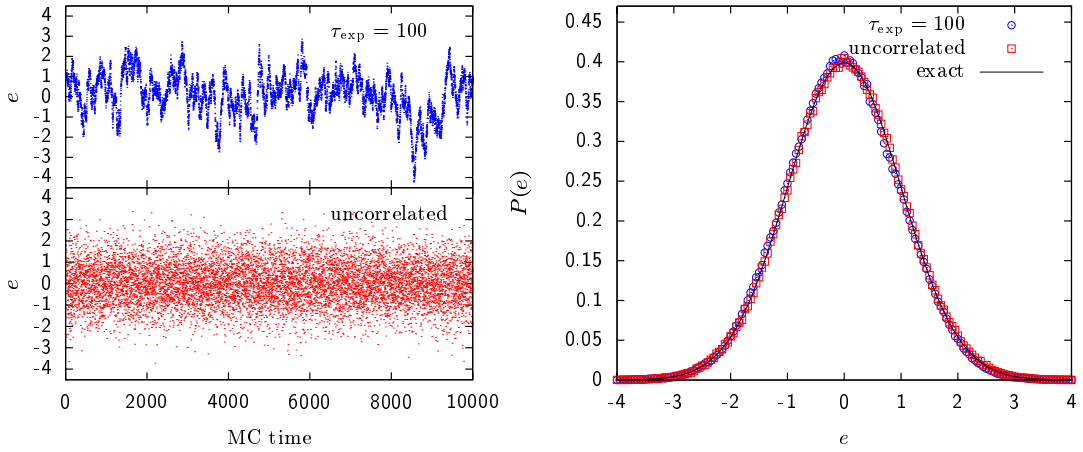

Figure 3: (left) Time series of the bivariate Gaussian process (3) with $\tau_{\text{exp}} = 100$ (upper plot) and for the uncorrelated case (lower plot). (right) Both time evolutions with a total of 1 000 000 consecutive measurements lead to the same Gaussian histogram.

temperature space on the autocorrelation times. In general, simulations near a second-order phase transition are affected by critical slowing down, i.e., an increasing autocorrelation time $\tau_{\text{can}} \sim \xi^z$, where ξ denotes the (spatial) correlation length and z is the dynamical critical exponent. To take this into account, we set τ_{exp} to the canonical autocorrelation time τ_{can} of the energy measured in the independent simulations. Together with the mean and variance this specifies the parameters of the replicated process (3).

By fitting to 2D Ising model MC data, our first finding comes from a comparison of the autocorrelation times for iterations of (3) with and without the PT routine. As expected, the autocorrelation times for the PT simulation are much smaller. The flow through temperature space looks similar as for the 2D Ising model depicted in Fig. 1. We also find a pronounced decline around the pseudo-critical point β_c. The reason for this behavior is, as already anticipated above, the slowed

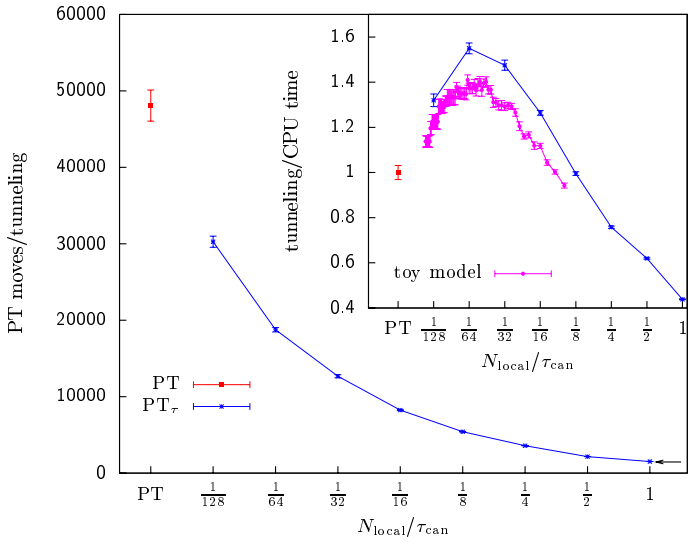

Figure 4: PT moves per tunneling as a function of $N_{\rm local}$ for the 2D Ising model ($L = 80$), approaching for $N_{\rm local} \approx \tau_{\rm can}$ the unbiased random walk limit. The inset shows the actually needed computing time in units of the total run time of the standard PT simulation.

down dynamics near β_c. That means, after two adjacent replica in the vicinity of β_c have been exchanged, they will stay close to each other and changing them back to the original state is more likely than an exchange with another replica. If the dynamics is even slower (by simply tuning $\tau_{\rm exp}$ larger) a complete trapping can be observed and the replica do not move from low to high temperatures at all.

By systematically varying the inputted autocorrelation times, our toy model suggests that an easy way to cure this problem is to increase the number of local updates between the PT exchanges proportional to the autocorrelation time of the initial (non PT) simulation for a given temperature. This general strategy will be now first tested for the 2D Ising model. For system sizes up to $L = 80$ we use the exact energy distributions [12] to calculate a set of inverse temperatures $\{\beta_i\}$ with an acceptance rate $A(i \to i+1) = 0.5$ starting from $\beta_1 = 0.38$. To cover almost the same temperature interval for different system sizes L the number of replica N has to increase with L [14]. For this set of inverse temperatures we perform short independent Metropolis MC simulations to estimate the canonical autocorrelation times $\tau_{\rm can}(\beta)$ of the energy, together with the mean and width of the energy distribution. In the actual simulations we then use the usual PT update scheme with only one important modification, namely, we choose the number of sweeps $N_{\rm local}(\beta)$ between the attempts to exchange the configurations proportional to $\tau_{\rm can}(\beta)$ ($N_{\rm local}(\beta) = 1$ for standard PT). The larger the number of sweeps between the exchange attempts, the smaller the correlation between adjacent replica. Therefore, one has to find a compromise between accuracy and computer time, which can be easily achieved by using our toy model (which runs orders of magnitude faster than the actual simulations). To illustrate this we include in Fig. 4 a comparison for different choices of $N_{\rm local}$.

The main plot of Fig. 4 shows the number of PT moves necessary for an arbitrarily chosen replica to move from the lowest to the highest temperature and back again. In the following such a round trip will be called tunneling. We clearly see that with an increasing number of sweeps per replica the tunneling time converges to the value of an unbiased random walk (indicated by

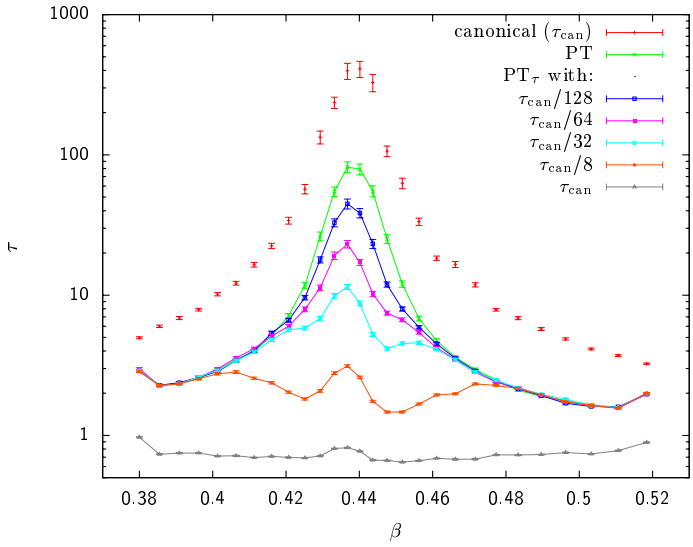

Figure 5: Autocorrelation times as function of inverse temperature for the canonical simulations, the parallel tempering (PT) update scheme, and five different runs of our improved parallel tempering update (PT$_\tau$) scheme for the 2D Ising model ($L = 80$).

the arrow in the lower right corner) consisting of two legs of length $(N-1)$. The limit for one round trip is hence given by $2(N-1)^2$. If we choose $N_{\text{local}}(\beta) = \tau_{\text{can}}(\beta)$, the correlation between adjacent replica is negligible and each replica performs a random walk through temperature space (see lower plot in Fig. 2). Furthermore, the sweeps needed for a tunneling event are close to the theoretical value, as is also reflected in the inset of Fig. 1, where we show that the fraction of replica moving "up" in the inverse temperature is an almost linear function of β. In the inset of Fig. 4, we compare the computational cost of our improved PT (denoted by PT$_\tau$) with that for standard PT. If one increases N_{local}, the ratio of tunnelings per CPU time decreases, i.e., above a certain threshold value of N_{local} the computational effort of PT$_\tau$ increases faster than the improvement of the tunneling speed.

To compare our improved PT$_\tau$ with other methods one should not only look at the computational cost but also at the accuracy that is achieved for the same amount of measurements. An easy way to check this is to measure the autocorrelation time τ. In Fig. 5, we show the autocorrelation times of the 2D Ising model with $L = 80$ for standard PT and our PT$_\tau$ algorithm with different choices of $N_{\text{local}}(\beta)$. The improvement gained by using PT instead of simulating each temperature independently is almost one order of magnitude in the region around the critical point. Taking in PT$_\tau$ the local autocorrelation times $\tau_{\text{can}}(\beta)$ into account we can decrease τ systematically. If we use $N_{\text{local}}(\beta) = \tau_{\text{can}}(\beta)$, then the autocorrelation times are smaller than unity for all temperatures and the resulting time series are nearly uncorrelated, but the computational costs are clearly too high to make this choice useful. Therefore, we suggest to use our toy model to find the optimal tunneling/CPU time ratio for each problem individually.

To summarize, we discovered a remarkable block building structure in PT simulations, revealed the mechanism behind it and showed how to cure this problem by taking into account the temperature dependence of autocorrelation times [15]. This demonstrates how easily the quality of

PT simulation data can be improved both in MC and MD studies. Furthermore, we proposed a toy model to find the optimal parameter set for the PT routine with almost no computational effort.

References

[1] R.H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).

[2] C.J. Geyer, in: *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, ed. E.M. Keramidas (Interface Foundation, Fairfax Station, VA, 1991), p. 156.

[3] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[4] E. Marinari, G. Parisi, and J.J. Ruiz-Lorenzo, in: *Spin Glasses and Random Fields*, ed. A.P. Young (World Scientific, Singapore, 1998), p. 59.

[5] U.H.E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997); Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999); C.Y. Lin, C.K. Hu, and U.H. Hansmann, Proteins **53**, 436 (2003); Y. Yamada, Y. Ueda, and Y. Kataoka, J. Comput. Chem. Jpn. **4**, 127 (2005); A. Schug, T. Herges, A. Verma, and W. Wenzel, J. Phys.: Condens. Matter **17**, S1641 (2005); T. Bataille, N. Mahé, E. Le Fur, J.-Y. Pivan, and D. Louër, Z. Kristallogr. Suppl. **23**, 9 (2006); H.-H. (Gavin) Tsai, M. Reches, C.-J. Tsai, K. Gunasekaran, E. Gazit, and R. Nussinov, Proc. Natl. Acad. Sci. U.S.A. **102**, 8174 (2005).

[6] M. Habeck, M. Nilges, and W. Rieping, Phys. Rev. Lett. **94**, 018105 (2005).

[7] M. Falcioni and M.W. Deem, J. Chem. Phys. **110**, 1754 (1999).

[8] A. de Candia and A. Coniglio, Phys. Rev. E **65**, 016132 (2001); R.G. Melko, J. Phys.: Condens. Matter **19**, 145203 (2007).

[9] E. Marinari, in: *Advances in Computer Simulation*, eds. J. Kertész and I. Kondor (Springer-Verlag, Berlin, 1998), p. 50; K. Hukushima, H. Takayama, and H. Yoshino, J. Phys. Soc. Jpn. **67**, 12 (1998): H.G. Katzgraber, M. Palassini, and A.P. Young, Phys. Rev. B **63**, 184422 (2001); E. Bittner and W. Janke, Europhys. Lett. **74**, 195 (2006).

[10] B. Joó, B. Pendleton, St.M. Pickles, Z. Sroczynski, A.C. Irving, and J.C. Sexton, Phys. Rev. D **59**, 114501 (1999); E.-M. Ilgenfritz, W. Kerler, M. Müller-Preussker, and H. Stüben, Phys. Rev. D **65**, 094506 (2002); G. Burgio, M. Fuhrmann, W. Kerler, and M. Müller-Preussker, Phys. Rev. D **75**, 014504 (2007).

[11] D.A. Kofke, J. Chem. Phys. **117**, 6911 (2002); Erratum: **120**, 10852 (2004); C. Predescu, M. Predescu, and C. Ciabanu, J. Phys. Chem. B **109**, 4189 (2005); N. Rathore, M. Chopra, and J.J. de Pablo, J. Chem. Phys. **122**, 024111 (2005); A. Kone and D.A. Kofke, J. Chem. Phys. **122**, 206101 (2005); H.G. Katzgraber, S. Trebst, D.A. Huse, and M. Troyer, J. Stat. Mech. (2006) P03018; S. Trebst, M. Troyer, and U.H.E. Hansmann, J. Chem. Phys. **124**, 174903 (2006); D. Gront and A. Kolinski, J. Phys.: Condens. Matter **19**, 036225 (2007).

[12] P.D. Beale, Phys. Rev. Lett. **76**, 78 (1996).

[13] W. Janke, *Introduction to simulation techniques*, in: *Ageing and the Glass Transition*, eds. M. Henkel, M. Pleimling, and R. Sanctuary, Lect. Notes Phys. **716** (Springer, Berlin, 2007), pp. 207–260.

[14] E. Bittner and W. Janke, in preparation; W. Janke and E. Bittner, PoS (LAT2009) 042.

[15] E. Bittner, A. Nußbaumer, and W. Janke, Phys. Rev. Lett. **101**, 130603 (2008).