

Dynamics and optimal number of replicas in parallel tempering simulations

Walter Nadler^{1,*} and Ulrich H. E. Hansmann^{1,2,†}

¹*Department of Physics, Michigan Technological University, Houghton, Michigan, USA*

²*John-von-Neumann Institute for Computing, Forschungszentrum Jülich, D-52425 Jülich, Germany*

(Received 22 June 2007; published 4 December 2007)

We study the dynamics of parallel tempering simulations, also known as the replica exchange technique, which has become the method of choice for simulation of proteins and other complex systems. Recent results for the optimal choice of the control parameter discretization allow a treatment independent of the system in question. By analyzing mean first passage times across the control parameter space, we find an expression for the optimal number of replicas in simulations covering a given temperature range. Our results suggest a particular protocol to optimize the number of replicas in actual simulations.

DOI: [10.1103/PhysRevE.76.065701](https://doi.org/10.1103/PhysRevE.76.065701)

PACS number(s): 05.10.Ln, 02.70.Tt, 02.70.Uu

The effective simulation of proteins, glasses, and similar complex systems has remained one of the defining challenges in computational physics. The main problem in such simulations is slow relaxation due to barriers and bottlenecks. Parallel tempering—also known as the replica exchange method—promised a way out of this dilemma [1–3]. Here, canonical or generalized-ensemble simulations [4] are performed in parallel at different values of a control parameter, most often the temperature. At certain times the current conformations of replicas at neighboring control parameter values are exchanged according to a generalized Metropolis rule [5]. An individual replica performs a random walk in control parameter space, allowing it to enter and escape local free energy minima. As a consequence, the state space is explored more evenly, especially, e.g., at low temperatures.

Replica exchange simulations are usually performed on massively parallel machines, with one or more computing nodes dedicated to performing the simulation of a replica at a particular control parameter value. In order to optimize the use of resources, an important question is whether an optimal choice for the number of replicas—or, equivalently, control parameter values [6]—exists, and how it might be determined. To our knowledge, no systematic investigation of this particular problem has been done to date.

A likely reason is that this question is closely connected to the metadynamics of parallel tempering, and a full understanding of this method for complex systems—exhibiting broken ergodicity [7]—is still missing [11]. It is also not a well-posed problem. Apart from the number of replicas, the main adjustable parameter is the *distribution* of control parameter values. The optimal number of replicas will depend strongly on the strategy used for the temperature discretization. Usually, a constant discretization is employed. However, often bottlenecks exist in some control parameter regions, and speeding up the equilibration of the system is possible only by using a finer discretization in these regions. In order to investigate the question of the optimal number of replicas systematically, a method of discretization has to be employed that also complies with some optimality criteria.

Major advances have been made recently in that direction. Instead of concentrating on stationary distributions that arise from sampling, Trebst *et al.* [8–10] have investigated the *flow* across control parameter space and have provided an iterative scheme for adjusting the discretization to optimize the flow distribution. Subsequently, we have shown that optimizing the flow is equivalent to minimizing the total first passage time to cross control parameter space [11]. In this Rapid Communication, we use those previous results as a basis for investigating the optimal number of replicas. We restrict ourselves to the situation of optimized flow and determine what number of control parameter values—identical to the number of replicas—minimizes the first passage time of a single replica.

In the following, we will consider parallel tempering with $N+1$ replicas. Hence, we will assume $N+1$ different control parameter values $\beta_0 < \beta_1 < \dots < \beta_N$, i.e., we have N control parameter intervals $[\beta_n, \beta_{n+1}]$. We will also use the conventions $\beta_0 = \beta_{\min}$ and $\beta_N = \beta_{\max}$. For simplicity we will call the control parameter a (inverse) temperature in the rest of this paper.

The time evolution of the probability $P(n, t)$ that an individual replica is at temperature β_n at time t can be approximated by a master equation [12] in discrete time [11],

$$\begin{aligned} P(n, t+1) = & P(n, t) \times [1 - W(\beta_n \rightarrow \beta_{n-1}) - W(\beta_n \rightarrow \beta_{n+1})] \\ & + P(n-1, t)W(\beta_{n-1} \rightarrow \beta_n) \\ & + P(n+1, t)W(\beta_{n+1} \rightarrow \beta_n), \end{aligned} \quad (1)$$

where $W(\beta \rightarrow \beta')$ are transition probabilities between neighboring temperatures. Of course, these probabilities depend on those temperatures, and the master equation for replica exchange is characterized by symmetric transition probabilities,

$$W(\beta \rightarrow \beta') = W(\beta' \rightarrow \beta) \equiv W(\beta, \beta') = W(\beta', \beta). \quad (2)$$

The relation between stationary flow J and first passage time τ in one-dimensional stochastic systems has been investigated in the context of channel flow in Ref. [13]. The two quantities are related via

*wnadler@mtu.edu

†hansmann@mtu.edu; u.hansmann@fz-juelich.de

$$J = C/\tau, \quad (3)$$

where C is a measure for the capacity of the channel. The mean first passage time for a single replica to cross the system defined by Eq. (1) in both directions is given by [11,14]

$$\tau = (N+1) \sum_{i=0}^{N-1} \frac{1}{W(\beta_i, \beta_{i+1})}, \quad (4)$$

while the channel capacity is simply the number of temperature values $C=N+1$. It was shown in Ref. [11] that—for a particular number of control parameter values—the current is maximized (and therefore the first passage time minimized) if the flow distribution is linear in the temperature number. This criterion allows an optimization of the temperature distribution [10,11]. We will assume in the following that such an optimized distribution of temperatures has been obtained. In this case, the effective transition probabilities in Eq. (1) are constant across the chain of temperatures [11],

$$W_{\text{opt}}(\beta_i, \beta_{i+1}) = \text{const} \quad \text{for } i = 0, \dots, N-1, \quad (5)$$

and this property will be essential in the analysis below.

With increasing number of temperatures, i.e., finer discretization, the transition probabilities approach their maximum $W \rightarrow W_0$. Its numerical value depends on the particular implementation of the replica exchange algorithm and the choice of the exchange time scale. In this limit, the mean first passage time shows the asymptotic behavior

$$\tau \propto N^2, \quad (6)$$

i.e., it grows quadratically with number of replicas. However, for smaller values of N the transition probabilities begin to decrease, leading eventually again to an increase in τ for small N . We are interested in the value of N where τ is minimal.

This value will depend on the change of the transition probabilities with the control parameter interval $[\beta, \beta']$. For the case of temperature as the control parameter, this question has been investigated in depth [11,17–20]. It has been found that the transition probability can be effectively approximated by [11,19–21]

$$W(\beta, \beta') \approx W_0 f\left(\frac{|\beta - \beta'|}{b}\right), \quad (7)$$

with $f(x)$ a monotonically decreasing function obeying $f(0) = 1$. The important quantity here is $b > 0$, denoting the scale of inverse temperatures over which the transition probability decreases. This scale is usually inverse to the widths of the thermal equilibrium energy distributions at β and β' . It decreases monotonically with system size and with the extensive heat capacity. In particular, it will be small near phase transitions. More details can be found in Refs. [11,17–20].

In the following, we assume that for a particular system the functional form in Eq. (7) is the same over the full temperature range. The only dependence of $W(\beta_i, \beta_{i+1})$ on the inverse temperature interval $[\beta_i, \beta_{i+1}]$ is through the corresponding scale parameter that we denote by $b_{i,i+1}$. Under this assumption the requirement that—for the optimal temperature distribution—all effective transition probabilities are

constant, Eq. (5), is equivalent to the condition that all individual arguments are identical. Hence,

$$|\beta_{i+1} - \beta_i|/b_{i,i+1} = \text{const} \equiv r \quad (8)$$

holds. Introducing the average scale \bar{b} ,

$$\bar{b} = \frac{1}{N} \sum_{i=0}^{N-1} b_{i,i+1}, \quad (9)$$

it can be seen that the constant r has the property

$$rN\bar{b} = \beta_{\text{max}} - \beta_{\text{min}}. \quad (10)$$

This property allows us to introduce the renormalized number of replicas:

$$\nu = \frac{\bar{b}}{\beta_{\text{max}} - \beta_{\text{min}}} N, \quad (11)$$

which, in turn, allows us to cast the mean first passage time for a replica to cross the system into the parameter-free form

$$\tau(\nu) \propto \nu^2 / f\left(\frac{1}{\nu}\right). \quad (12)$$

Minimizing ν for a particular functional form $f(x)$ of the transition probability decrease will give us finally the optimal number of replicas.

We analyze Eq. (12) using the following functional forms for $f(x)$:

$$f(x) = \begin{cases} \exp(-x) & \text{(a),} \\ \text{erfc}(\sqrt{\pi/4}x) & \text{(b),} \\ 1-x & \text{(c).} \end{cases} \quad (13)$$

In order to ensure they are comparable, we chose $f(0)=1 = f'(0)$, i.e., the initial slope is identical for all three forms. Form (b) has actually been derived for temperature intervals [11,19–21] and exhibits an $\exp(-x^2)/x$ tail. It is the one most likely to occur in an actual situation. We have included forms (a) and (c) as worst case scenarios since they cover a wide range of behavior around (b): (a) exhibits a simple exponential decrease, much slower than (b), while (c) exhibits a much faster, linear decrease; note that the latter is valid only for $x < 1$. Figure 1 shows a graphical comparison of the three functions.

Minimizing Eq. (12) gives the optimal value for ν ,

$$\nu_{\text{opt}} = \begin{cases} 1/2 & \text{(a),} \\ 1.052\,67 & \text{(b),} \\ 3/2 & \text{(c).} \end{cases} \quad (14)$$

These values are all of order 1, despite the wide range of functional behavior of the transition probability decrease they describe. Rewriting Eq. (11), we obtain our final result for the optimal number of replicas:

$$N_{\text{opt}} = \nu_{\text{opt}}(\beta_{\text{max}} - \beta_{\text{min}})/\bar{b}. \quad (15)$$

The ratio of the full temperature range of the simulation to the average scale \bar{b} is the main determining quantity in that

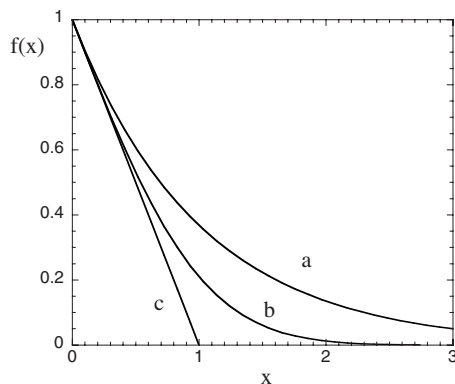


FIG. 1. Different functional forms for the decay of the transition probability with control parameter distance; compare Eqs. (7) and (13).

equation. In particular, it controls the order of magnitude for N_{opt} . Particular functional forms for the distance dependence of the transition probabilities appear to have less influence since ν_{opt} is of $O(1)$ only. Although the range of values given in Eq. (14) for ν_{opt} is still covered by a factor of 3, this is a worst case scenario, and actual values for realistic functional forms will be closer to (14) case (b).

The influence of the different functions, Eq. (13), is more important for the form of the minimum of the mean first passage time. Figure 2 shows how the mean first passage time changes when the number of replicas N deviates from the optimal value N_{opt} . In order to enable comparison we use the renormalized replica number, Eq. (11), as variable here. While the minimum is pronounced for all functional forms, it is steepest for form (c), i.e., the fastest decreasing probability function, and most shallow for form (a).

Before we discuss the consequences of Eq. (15) for protocols to optimize the number of replicas, we need to address some subtleties of the above derivation that we skipped over in favor of a compact derivation.

(i) Equation (1) is an effective description of the long-time properties of the random walk of replicas in parallel tempering simulations. The transition probabilities for such a

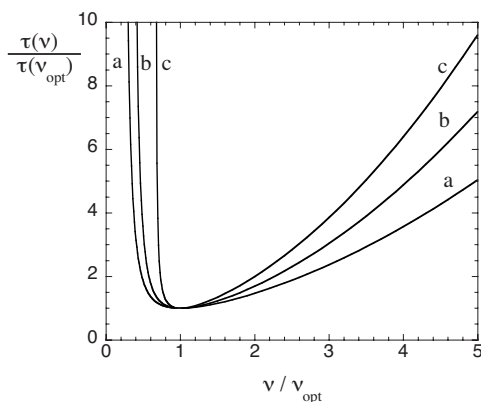


FIG. 2. Dependence of relative mean first passage time, Eq. (12), on deviations from the optimal number of replicas for the different functions of Eq. (13).

long-time description differ from observed acceptance rates, which are, on the replica exchange time scale, short-time properties. In particular, it has been noted that for optimized flow the observed acceptance rates are *not* constant [10]. This is due to broken ergodicity [7] at particular control parameter values, which gives rise to a hierarchical, treelike structure for the random walk of replicas [11]. Observed flow and acceptance rates are just projections onto the one-dimensional chain of control parameter values of the more complicated flow processes on the tree. However, the possibility of flow optimization shows that for the long-time transition probabilities the property (5) holds nevertheless. Such discrepancies between short-time and long-time properties of stochastic processes are well known [15,16]. By using Eq. (7) we implicitly assume that those effective transition probabilities exhibit the same qualitative behavior with control parameter difference as was derived for the short-time transition probabilities. We feel that this is justified since our qualitative results for N_{opt} , Eq. (15), are independent of the particular functional form.

(ii) J and τ differ by a factor of $N+1$; see Eq. (3). The mean first passage time is usually a good estimate for the lowest eigenvalue of the equation system (1), i.e., it determines the time scale of equilibration [15,16]. Since we are interested in fast equilibration, τ is a more adequate quantity than J to use for comparing systems with different numbers of replicas and to optimize with respect to N .

(iii) We have omitted a discussion of the replica exchange time scale. Depending on the frequency of replica exchange moves, the time scale of Eq. (1) may differ by a factor of N . However, our result (15) is stable with respect to changes in the power of N . Although the particular numerical values change, the value of ν_{opt} remains an $O(1)$ constant if the exponent in Eq. (12) changes from 2 to 1 ($\nu_{\text{opt}}=1, 1.6671, 2$) or 3 ($\nu_{\text{opt}}=1/3, 0.820\,024, 4/3$).

What are the consequences of our results, particularly of Eq. (15), for protocols to optimize the number of replicas? The main result of Eq. (15) is that it identifies, separates, and quantifies the contributions of various properties of the simulation system to the optimal number of replicas. In particular, it exhibits the quantitative hierarchy of the individual contributions. The above analysis also shows the importance of how transition probabilities change with the control parameter interval. To our surprise, for the inverse temperature as the control parameter, this complex contribution could be summarized formally into the averaged scale \bar{b} , Eq. (9). Taking into account the dependence of the scales $b_{i,i+1}$ on the extensive properties of a system [11,19,20] suggests that N_{opt} scales with system size V as $N_{\text{opt}} \propto \sqrt{V}$. Corrections due to anomalous scaling in the critical region might occur, though.

We note that the determination of \bar{b} in actual simulations is by no means simple. Since it is defined for the situation of optimal control parameter spacing for a particular number of replicas, such an optimization would have to be performed beforehand. Also, since it describes the behavior of the *effective* transition probabilities (see the above discussion), it would have to be determined from the flow distribution together with the actual first passage time, upon slightly varying the discretization.

Instead, our analysis suggests that the direct approach to optimizing the number of replicas is the most promising one. Comparing first passage times of replicas to cross the simulation system for the optimized discretization is readily possible for different values of N . Figure 2, in particular form (b), can then be used as a guideline to extrapolate to N_{opt} .

In summary, we have studied the dynamics of parallel tempering simulations. By analyzing these dynamics, we have determined the main factors influencing the optimal number of replicas in such simulations and their quantitative hierarchy. Since the evaluation of the essential term \bar{b} , the

average scale of transition probability decrease, may need costly computations, we propose to base optimizing the number of replicas on the generic behavior of a replica's first passage time to cross the simulation system given in Fig. 2. The technique of replica exchange has become the method of choice for the simulation of proteins and other complex systems. The above results add to its understanding, and we believe they will also advance its practical use.

It is a pleasure to thank P. Grassberger and H. Katzgraber for discussions. This research was supported by NSF Grant No. CHE-0313618.

-
- [1] C. J. Geyer and A. Thompson, *J. Am. Stat. Assoc.* **90**, 909 (1995).
 - [2] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
 - [3] U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
 - [4] U. H. E. Hansmann and Y. Okamoto, *Phys. Rev. E* **56**, 2228 (1997).
 - [5] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
 - [6] We will not treat the case where several replicas are simulated at the same control parameter value [see, e.g., M. Nania, C. Czaplewski, and H. A. Scheraga, *J. Chem. Theory Comput.* **2**, 513 (2006), and references therein], since that situation is still less understood.
 - [7] R. Palmer, *Adv. Phys.* **31**, 669 (1982); D. L. Stein and C. M. Newman, *Phys. Rev. E* **51**, 5228 (1995).
 - [8] S. Trebst, D. A. Huse, and M. Troyer, *Phys. Rev. E* **70**, 046701 (2004).
 - [9] H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, *J. Stat. Mech.: Theory Exp.* (2006) P03018.
 - [10] S. Trebst, M. Troyer, and U. H. E. Hansmann, *J. Chem. Phys.* **124**, 174903 (2006).
 - [11] W. Nadler and U. H. E. Hansmann, *Phys. Rev. E*, **75** 026109 (2007).
 - [12] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer, Berlin, 1985).
 - [13] W. R. Bauer and W. Nadler, *J. Chem. Phys.* **122**, 244904 (2005); *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11446 (2006).
 - [14] Compare Eqs. (45) and (46) in Ref. [11]; due to a typographical error the term $(N+1)$ is missing in (46); however the subsequent discussion is unaffected by that.
 - [15] K. Schulten, Z. Schulten, and A. Szabo, *J. Chem. Phys.* **74**, 4426 (1981).
 - [16] W. Nadler and K. Schulten, *J. Chem. Phys.* **82**, 151 (1985); *Z. Phys. B: Condens. Matter* **59**, 53 (1985); *J. Chem. Phys.* **84**, 4015 (1986).
 - [17] C. Predescu, M. Predescu, and C. Ciabanu, *J. Chem. Phys.* **120**, 4119 (2004).
 - [18] D. A. Kofke, *J. Chem. Phys.* **117**, 6911 (2002); **120**, 10852(E) (2004).
 - [19] D. A. Kofke, *J. Chem. Phys.* **121**, 1167 (2004).
 - [20] A. Kone and D. A. Kofke, *J. Chem. Phys.* **122**, 206101 (2005).
 - [21] Note that in Refs. [19,20] the argument used is the ratio $B = \beta'/\beta$; however, since it occurs only in the form $(1-B)/(1+B)$, it is equivalent to the argument in Eq. (7).