Towards an Optimal Flow: Density-of-States-Informed Replica-Exchange Simulations

Thomas Vogel[*] and Danny Perez[†]
Theoretical Division (T-1), Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

(Received 13 May 2015; revised manuscript received 13 July 2015; published 5 November 2015)

Replica exchange (RE) is one of the most popular enhanced-sampling simulations technique in use today. Despite widespread successes, RE simulations can sometimes fail to converge in practical amounts of time, e.g., when sampling around phase transitions, or when a few hard-to-find configurations dominate the statistical averages. We introduce a generalized RE scheme, density-of-states-informed RE, that addresses some of these challenges. The key feature of our approach is to inform the simulation with readily available, but commonly unused, information on the density of states of the system as the RE simulation proceeds. This enables two improvements, namely, the introduction of resampling moves that actively move the system towards equilibrium and the continual adaptation of the optimal temperature set. As a consequence of these two innovations, we show that the configuration flow in temperature space is optimized and that the overall convergence of RE simulations can be dramatically accelerated.

DOI: 10.1103/PhysRevLett.115.190602 PACS numbers: 05.10.-a, 02.70.-c, 64.70.D-

Sampling the phase space of Hamiltonians to estimate thermodynamic properties is one of the fundamental problems in statistical physics. However, direct approaches often dramatically fail due to the presence of large free-energy barriers between different regions of phase space. While many methods have been proposed to address this challenge, few have had an impact as significant as the replica-exchange (RE) method [1–4]. Indeed, since its introduction, RE has established itself as a workhorse in atomistic and coarse-grained simulations and is currently used to investigate a large variety of systems in many areas, ranging from statistical physics [5–7], over biology and chemistry [8–10], to solid-state physics and materials science [11–13].

RE enhances the exploration of phase space by using a set of N individual simulations (often called "walkers") that evolve through Monte Carlo (MC) or molecular dynamics (MD) updates under different external parameters. For example, each walker might run at a different temperature T_i ($1 \leq i \leq N$), which is the situation we consider in the following (in this context, RE is often referred to as parallel tempering). After predefined time intervals τ_{RE}, the exchange of the current microstates between pairs of walkers is attempted and carried out with an appropriate probability, $W_{\mathrm{acc}}(U_i, U_j) = \min[1, e^{\Delta\beta\Delta U}]$ in the case of canonical walkers, where $\beta_i = (k_B T_i)^{-1}$ are the inverse temperatures of the heat baths and U_i the internal potential energies of the microstates. This exchange mechanism promotes configurational mixing by exposing replicas to external conditions (e.g., high temperatures) where free energy barriers can easily be overcome. It further provides a means for thermodynamic information to be transferred to conditions where the convergence of direct MC or MD simulations would require prohibitively long simulation times.

The exchange probabilities in RE strictly comply with detail balance, which ensures that the proper canonical distributions will be sampled at all temperatures. This property is extremely useful because samples taken at each temperature can be used without reweighting. It can, however, be restrictive, especially in the early stages of a simulation, when systems are far from equilibrium. This is related to the fact that conventional RE does not provide a natural mechanism for integrating and exploiting information that becomes available during the simulation.

In this Letter, we show how one such inexpensive and readily available source of information—namely, concurrent estimates of the density of states, $g(U)$—can significantly improve conventional RE. We leverage the (approximate) knowledge of $g(U)$ in two ways. First, we introduce a resampling operation akin to a Gibbs sampling move, which samples according to the $g(U)$-inferred canonical distribution over an ensemble of configurations previously visited by any replica, thereby explicitly steering the system toward equilibrium. Second, we use estimates of $g(U)$ to continuously improve the temperature set $\{T_i\}$. Because resampling breaks correlations along individual trajectories on each replica, the temperature set can be made optimal with respect to diffusion in temperature space [14].

The enabling factor of our approach is the concurrent estimation of the density of states $g(U)$. While it can be obtained by a number of techniques [15–18], we here rely on ideas from the adaptive biasing force (ABF) formalism [19] for MD simulations. The key is to frame the problem as the estimation of the free energy $F_\beta(U) = -k_B T \ln[g(U) \exp(-\beta U)]$ through its derivative s, which can be written in terms of microcanonical averages [19] as

$$s(U) = \frac{dF}{dU} = -\left\langle \frac{d}{dt}(w \cdot p) \right\rangle_U, \tag{1}$$

with p being the vector of momenta and $w = \nabla U/(\nabla U \cdot \nabla U)$. Here, the derivative with respect to time is understood to be a derivative along a microcanonical trajectory. In practice, $d(w \cdot p)/dt$ is measured periodically (say every 100 time steps) and is stored with its corresponding value of U. We then use binned averages to reconstruct $s(U)$. By integration, we recover $F_\beta(U)$, and hence $g(U)$ [20]. In an alternative representation, $s(U)$ can be used to define a microcanonical temperature observable: $T_m(U) = T_0/[1 - s(U)]$, where T_0 is the heat-bath temperature. By integrating the thermodynamic relation $1/T_m(U) = dS(U)/dU$, with $S(U) = k_B \ln g(U)$, one reaches the same result. Note that if the momenta are unavailable, e.g., when using MC dynamics, a configurational temperature $T_m(U)$ can be estimated based on structural data alone [21–23]. The cost of estimating $s(U)$ is negligible in practice. An advantage of this approach is that the validity of Eq. (1) does not depend on the sampling being carried out in any particular ensemble: the only requirement is that the dynamics yields equal probabilities of observing different configurations with the same U. To minimize the impact of initial—potentially far-from-equilibrium—states on the estimator at later times, we introduce a memory time (much larger than all other time scales) after which measurements are discarded.

While in the ABF method [19], $s(U)$ is used to create a multicanonical ensemble, we instead leverage it in our density-of-states-informed RE scheme (g-RE) in the following ways. While leaving the original RE mechanism untouched, we first introduce an additional, global resampling move (executed after time intervals τ_{resamp}) by which the microstate of a replica is resampled from the ensemble of configurations visited by *any* of its peers at *any* time in the past. In practice, this is enabled by a global configuration database populated by all walkers during the simulation. A configuration is selected from the database with a probability proportional to its estimated canonical weight $P_{\beta_i}(U) = g(U)\exp[-\beta_i U]$. This is akin to an approximate Gibbs sampling [28,29]. What is crucial here is that the $P_{\beta_i}(U)$ are inferred from *global* thermodynamic information, so that it can differ from the distribution locally sampled by the corresponding walker. In other words, this operation actively steers the distributions towards what is globally deemed equilibrium. It does so by allowing for the *replication* of thermodynamically relevant states, in contrast to conventional RE where configuration can only *diffuse* in configuration space. As will be shown below, resampling proves essential for convergence in the neighborhood of strong phase transitions and for the timely escape out of metastable states. One might, however, wonder whether g-RE produces correct statistics, as resampling does not *a priori* obey global balance when the configuration database is finite. In fact, the only requirement for correctness is that any two states with the same U have the same probability of eventually being observed during an arbitrary long simulation.

As discussed in the Supplemental Material [23], this remains the case when resampling is introduced, as long as the samplers on each replica (e.g., Langevin MD or Metropolis MC) are ergodic and canonical in and of themselves. In that limit, $g(U)$, as obtained through Eq. (1), will converge to the correct value. From the knowledge of $g(U)$ and from a sample of observed configurations, any canonical quantity can then be obtained. It is, however, important to note that resampling introduces correlations between the configurations stored in the database at a specific point in time and, hence, potentially also between replicas. Our approach exploits these correlations to share information between replicas. However, resampling too frequently will lead to statistical inefficiencies. The optimal choice of τ_{resamp} will be discussed in future works.

The availability of $g(U)$ also enables a second innovation: the continuous optimization of the temperature set $\{T_i\}$. The overall goal here is to minimize the round-trip time for replicas to wander between low and high temperatures. Much effort has been (and is still) dedicated to addressing this issue (see Refs. [14,30–36] for examples). From this body of work it emerges that performance is often characterized in terms of two key concepts: the average exchange acceptance probabilities $W_{\mathrm{acc}}(i, i+1)$ between pairs of neighboring temperatures and the flow ratio $f(i) = n_{\mathrm{up}}(i)/[n_{\mathrm{up}}(i) + n_{\mathrm{down}}(i)]$, i.e., the fraction of replicas that diffuse up in temperature for a given walker i (a replica is said to flow upwards if it has visited the minimum temperature more recently than the maximum temperature; cf. Refs. [31,32]). In the probability-centric view [14,33,35], the objective is to find the $\{T_i\}$ such that $W_{\mathrm{acc}}(i, i+1) = \mathrm{const}$, $\forall i$, the insight being that locally low acceptance probabilities would limit the free diffusion of replicas. In the flow-centric view [31,32], the optimal set $\{T_i\}$ is such that $f(i) = 1 - [(i-1)/(N-1)]$ (for $T_i < T_{i+1}$, $\forall i$), as this indicates an unimpeded flow of replicas. In contrast, the presence of a bottleneck would signal itself by comparatively flat regions separated by a sharp drop in $f(i)$. The two approaches have been contrasted by Nadler and Hansmann [14], who showed that, in the special case where the dynamics on each replica become completely uncorrelated between exchange attempts, the *optimal* choice is to make the exchange probabilities constant, as this minimizes the round-trip time from the lowest to the highest temperature. Furthermore, the flow ratio will also be linear in this case. In the current context, this limit can be approached by setting $\tau_{\mathrm{resamp}} = \tau_{\mathrm{RE}}$, as resampling decreases correlations (as long as τ_{RE} is not so short as to saturate the database with near-identical configurations). In g-RE, the optimal temperature set can easily be determined. Holding the minimal and maximal temperatures fixed, we apply a bisection scheme until $\{T_i\}$ converges to a situation where

$$W_{\mathrm{acc}}^{m}(P_{i+1}, P_i) = \int_{-\infty}^{\infty} P_{\beta_i}(U) \int_{-\infty}^{\infty} W_{\mathrm{acc}}(U, U') \times P_{\beta_{i+1}}(U')\, dU\, dU' = \mathrm{const}, \quad \forall i \quad (2)$$

(see the Supplemental Material [23] for more details). Here also, the $P_{\beta_i}(U)$'s are based on the current estimator of $g(U)$, and Eq. (2) can be evaluated numerically. This scheme is very fast in practice and does not require any preliminary calculation. We continuously readjust the temperatures at intervals τ_{adapt}, assisting convergence in situations where the walkers begin far from equilibrium. Note that adaption of the $\{T_i\}$ does not interfere with the averages necessary for the measurement of $s(U)$, as samples taken at different temperatures can be seamlessly integrated through Eq. (1).

We now demonstrate the performance of our method on a system of 500 silver atoms. The $N = 71$ walkers are canonical molecular dynamics runs with a time step of $\tau_{\mathrm{ts}} = 2$ fs, and the heat-bath temperatures are set by Langevin thermostats. Atoms interact via an embedded-atom potential [37]. The simulation cell is cubic and periodic boundary conditions are used in all directions. The minimum temperature is set to $T_{\mathrm{min}} = 100$ K and the maximal to $T_{\mathrm{max}} = 3500$ K. The particle density is fixed at $\rho = 0.0585$ Å^{-3}, which corresponds to the density that minimizes the energy of the fcc crystal; the fcc configuration is therefore the putative global energy minimum for this system and should dominate up to the vicinity of the melting point, given that the thermal concentration of defects is expected to be vanishingly small. However, all walkers are initialized from a quenched liquid (amorphous) configuration. This choice makes the system a very good prototype of a case where the thermodynamically relevant configurations are unknown *a priori* and are difficult to access. Indeed, from MD studies of metals (e.g., Ag [38–40], Cu [41], Ni [42], etc.), it is known that recovering the perfect crystalline state from the melt requires very slow cooling; otherwise, the system remains trapped in amorphous states (at fast cooling) or in a mixture of fcc and hcp regions separated by stacking defects (at moderate cooling rates). In addition, the presence of a first-order transition (melting) within the range of temperatures makes this an extremely challenging system to study. Finally, since the thermal concentration of crystal defects is expected to be vanishingly small in a system of that size away from the immediate vicinity of the melting point, the validity of the results is easy to assess.

As illustrated in Fig. 1, the coupling of the different replicas through the estimator of $g(U)$ and the configurational database directs the evolution of the system towards

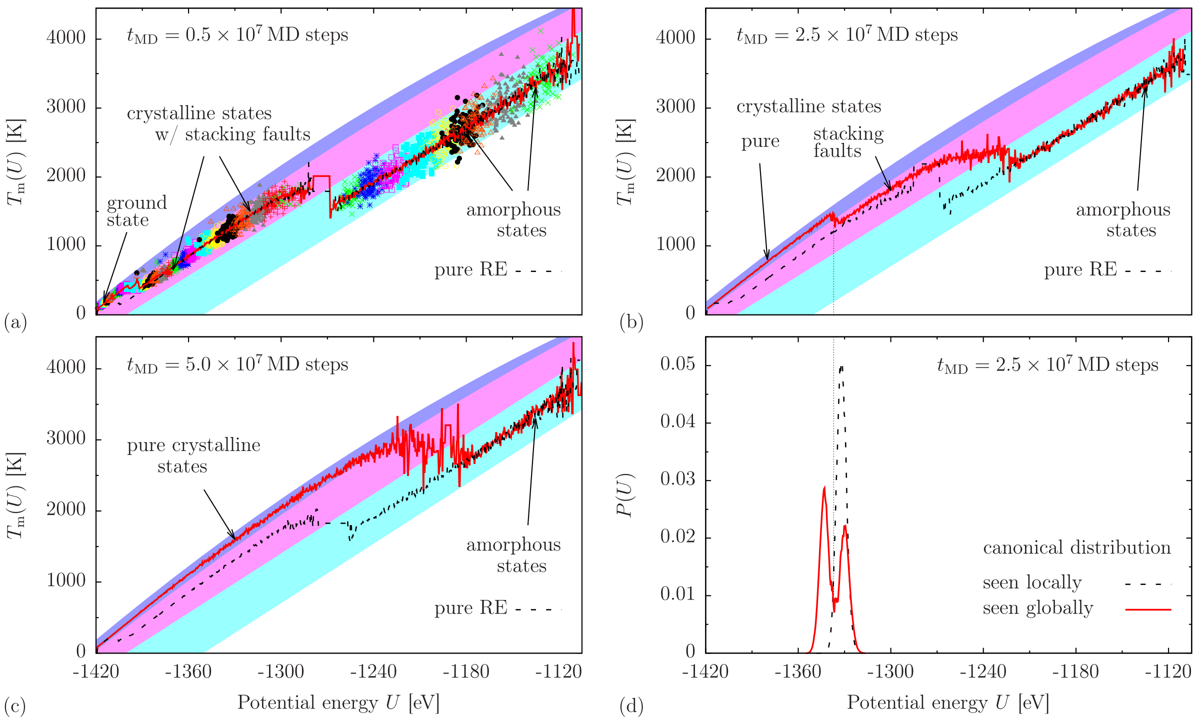

FIG. 1 (color online). Convergence of $T_m(U)$ at different simulation times (a) shortly after the start of the run and after the first pure crystalline states have been discovered; (b) as these states get replicated and move the spurious transition between pure and faulted states upwards in temperature; and (c) after convergence of the g-RE scheme. The dots in (a) show individual measurements; different colors correspond to different walkers (only a subset of data from every other walker is shown). The solid red line shows $T_m(U)$, the dashed black line the same obtained from an equivalent, conventional RE run (for comparison). The shaded bands in (a) to (c) are guides for the eyes and denote the types of configurations predominantly found in each of these regions, as indicated by the different labels. (d) shows the estimated canonical distribution $P_{\beta_i}(U)$, at the time corresponding to (b), as obtained from global data (the solid red line) compared to what would be inferred by an isolated walker trapped in faulted states (the dotted black line).

thermal equilibrium by enabling the replication of thermodynamically relevant, crystalline states, in contrast to conventional RE, where states are only *exchanged*. In that figure, we report the instantaneous estimators of the microcanonical temperature $T_m(U)$ at different times in the simulation. In our example, the perfect crystalline states (the upper, purple band) are essentially the only relevant ones below the melting point; they are, however, by construction not present in the early stages of the simulations. While, at lower temperatures, the system quickly leaves the amorphous region of phase space (the lower, pale blue band) and crystallizes, most of these crystalline configurations initially contain stacking faults (the middle, pink band) that take a very long time to anneal. Convergence of standard RE requires the perfect crystalline state to be *independently* found at least (and ideally much more than) $n_{<\mathrm{melt}}$ times (the number of walkers for which $T_i < T_{\mathrm{melt}}$). The rate at which this occurs hence controls the convergence speed; from standard RE we infer this rate to be of the order 10^8 s^{-1} and the convergence would require approximately 2.5×10^8 MD steps. In contrast, resampling enables the replication of the pure crystalline state as soon at it is found *once* by *any* replica. As illustrated in Fig. 1(d), this occurs because the global estimate of $g(U)$ eventually contains contributions from the crystalline region that are locally invisible to a walker trapped in faulted states. Computational gains follow because the probability of resampling a crystalline state significantly exceeds that of a replica independently finding it.

The same is true around the melting transition, which is initially biased to lower temperatures due to the initialization from a quenched liquid (note that the melting point at constant volume is much higher than the triple point). Before equilibration, a spurious transition between perfect crystals and faulted states [the sudden drop in $T_m(U)$ at low energies] is observed and the melting temperature is underestimated. This is illustrated in Figs. 1(a) and 1(b) (the red solid lines). A particular advantage of continuous temperature adaptation is that the temperature set remains nearly optimal at all times, even as the position of the (spurious or real) phase transitions evolve during convergence. With g-RE we observe convergence at $t_{\mathrm{MD}} \approx 4 \times 10^7$ MD steps, while conventional RE, even with temperature adaptation (the black dashed lines), still mainly samples faulted states. The temperature adaptation scheme is very robust and converges quickly, even for a poor choice of the initial set $\{T_i\}$ (see Ref. [23]).

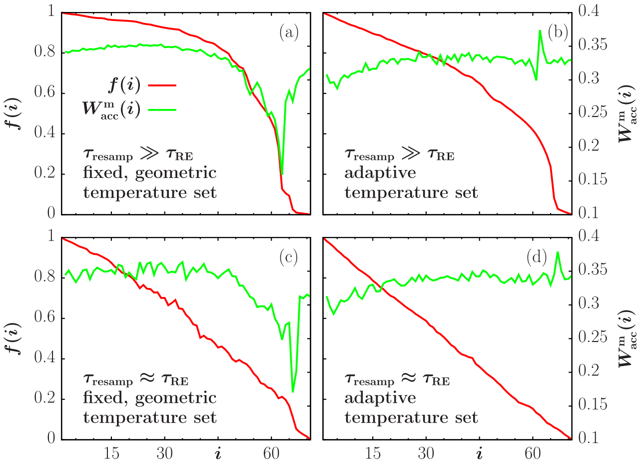

FIG. 2 (color online). Measured RE acceptance rates [$W^m_{\mathrm{acc}}(i, i+1)$; the green curves] and fraction of replicas diffusing from the lowest to the highest temperature [$f(i)$; the red curves] for the 500 Ag atoms system. (a,c) Fixed, geometric temperature set. (b,d) Adaptive temperature set. (a,b) Nonergodic sampler: $\tau_{\mathrm{resamp}} \gg \tau_{\mathrm{RE}}$. (c,d) Ergodic sampler: $\tau_{\mathrm{resamp}} \approx \tau_{\mathrm{RE}}$.

Finally, Fig. 2 shows how the introduction of global resampling at time scales comparable to the RE exchange time ($\tau_{\mathrm{resamp}} \approx \tau_{\mathrm{RE}} \sim 10^2 \tau_{\mathrm{ts}}$) and an adaptive temperature set results in an optimal flow of the replica through temperature space; i.e., *both* constant $W^m_{\mathrm{acc}}(i, i+1)$ *and* linear $f(i)$ are observed [Fig. 2(d)]. The measured round-trip times ($\tau^m_{\mathrm{rt}} \approx 14000\tau_{\mathrm{RE}}$) are in perfect agreement with measured round-trip times for a purely random exchange process with corresponding exchange probabilities, which here correspond to the optimal limit [14]. As shown in Figs. 2(a)–(c), these two conditions can only be obeyed when using both temperature adaptation and resampling. In this sense, our work integrates earlier efforts [31,32,35], where either constant exchange times, exchange rates, or the optimal flow had to be sacrificed.

In conclusion, we introduce a general scheme, g-RE, that can dramatically improve the efficiency of RE simulations. The method is based on the idea of informing RE simulations with estimators of the density of states $g(U)$ gathered on the fly. This allows for two key improvements: the introduction of a global resampling move that guides the system towards equilibrium and causes a dramatic reduction of correlation times of the sampling, and the on-the-fly determination of an optimal temperature set that simultaneously achieves constant exchange probabilities and linear flow ratio. We expect our method to be particularly useful for any system with dominant but hard to access states or around strong first-order transitions. So far, there have been two approaches to such problems: either to change the ensemble (see Refs. [33,43] for examples) or to introduce global MC trial moves. In the former case, one often chooses to work in the multicanonical ensemble where $P(U) = \mathrm{const}$; creating such an ensemble is in fact a common way to leverage the knowledge of $g(U)$. This approach, however, is not optimal for the system reported here because, even though that ensemble is free from gradients of $F(U)$, the energy landscape locally remains extremely rough, thereby severely limiting the diffusivity in U. Using diffusion in T space to promote mixing proved a more efficient alternative. Nonetheless, it still required the introduction of a global move to insure convergence. We here proposed a *generic* solution for constructing such a global move that does not require any *a priori* information about the system. We hence expect this approach to be useful for a wide range of systems and to also be applicable to RE schemes using other ensembles [9,12,44–47].

This work was supported by the U.S. Department of Energy through the Los Alamos National Laboratory (LANL)/LDRD Program and used computing resources provided by the Los Alamos National Laboratory Institutional Computing Program. Los Alamos National Laboratory is operated by Los Alamos National Security, LLC, for the National Nuclear Security administration of the U.S. DOE under Contract No. DE-AC52-06NA25396.

---

[*] Present address: Department of Physics, Stetson University, DeLand, FL 32723, USA.
tvogel@lanl.gov
[†] danny_perez@lanl.gov

[1] C. J. Geyer, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, edited by E. M. Keramidas (Interface Foundation, Fairfax Station, VA, 1991), p. 156.
[2] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992).
[3] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[4] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).
[5] M. Hasenbusch, A. Pelissetto, and E. Vicari, Phys. Rev. B **78**, 214205 (2008).
[6] B. Yucesoy, H. G. Katzgraber, and J. Machta, Phys. Rev. Lett. **109**, 177204 (2012).
[7] E. Bittner and W. Janke, Europhys. Lett. **74**, 195 (2006).
[8] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[9] J. Gross, T. Neuhaus, T. Vogel, and M. Bachmann, J. Chem. Phys. **138**, 074905 (2013).
[10] H.-H. G. Tsai, M. Reches, C.-J. Tsai, K. Gunasekaran, E. Gazit, and R. Nussinov, Proc. Natl. Acad. Sci. U.S.A. **102**, 8174 (2005).
[11] M. Falcioni and M. W. Deem, J. Chem. Phys. **110**, 1754 (1999).
[12] S. Auer and D. Frenkel, Nature (London) **409**, 1020 (2001).
[13] F. Chuang, C. Ciobanu, C. Predescu, C. Wang, and K. Ho, Surf. Sci. **578**, 183 (2005).
[14] W. Nadler and U. H. E. Hansmann, Phys. Rev. E **75**, 026109 (2007).
[15] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[16] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).
[17] J. Kim, J. E. Straub, and T. Keyes, Phys. Rev. Lett. **97**, 050601 (2006).
[18] C. Junghans, D. Perez, and T. Vogel, J. Chem. Theory Comput. **10**, 1843 (2014).
[19] E. Darve, D. Rodríguez-Gómez, and A. Pohorille, J. Chem. Phys. **128**, 144120 (2008).
[20] For brevity, we will use the same symbol for the exact $g(U)$ and its estimator in the following. It should, however, be understood that, in the practical implementation of g-RE, the exact $g(U)$ is not available, so that $g(U)$ there refers to the estimator based on Eq. (1) instead.
[21] H. H. Rugh, Phys. Rev. Lett. **78**, 772 (1997).
[22] B. D. Butler, G. Ayton, O. G. Jepps, and D. J. Evans, J. Chem. Phys. **109**, 6519 (1998).
[23] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.115.190602, which includes Refs. [24–27], for implementation details and a discussion of the convergence properties of the method.
[24] M. P. Allen and D. Quigley, Mol. Phys. **111**, 3442 (2013).
[25] C. Junghans, M. Bachmann, and W. Janke, Phys. Rev. Lett. **97**, 218103 (2006).
[26] S. Schnabel, D. T. Seaton, D. P. Landau, and M. Bachmann, Phys. Rev. E **84**, 011127 (2011).
[27] R. L. Burden and J. D. Faires, *Numerical Analysis*, 3rd ed. (Prindle, Weber & Schmidt Publishers, Boston, 1985).
[28] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, 2nd ed. (Springer-Verlag, New York, 2002).
[29] J. D. Chodera and M. R. Shirts, J. Chem. Phys. **135**, 194110 (2011).
[30] N. Rathore, M. Chopra, and J. J. de Pablo, J. Chem. Phys. **122**, 024111 (2005).
[31] H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, J. Stat. Mech.: Theory Exp. (**2006**) P03018.
[32] S. Trebst, M. Troyer, and U. H. E. Hansmann, J. Chem. Phys. **124**, 174903 (2006).
[33] T. Neuhaus, M. P. Magiera, and U. H. E. Hansmann, Phys. Rev. E **76**, 045701 (2007).
[34] A. Patriksson and D. van der Spoel, Phys. Chem. Chem. Phys. **10**, 2073 (2008).
[35] E. Bittner, A. Nußbaumer, and W. Janke, Phys. Rev. Lett. **101**, 130603 (2008).
[36] A. J. Ballard and D. J. Wales, J. Chem. Theory Comput. **10**, 5599 (2014).
[37] P. L. Williams, Y. Mishin, and J. C. Hamilton, Model. Simul. Mater. Sci. Eng. **14**, 817 (2006).
[38] S. Xiao and W. Hu, J. Chem. Phys. **125**, 014503 (2006).
[39] Z.-A. Tian, R.-S. Liu, H.-R. Liu, C.-X. Zheng, Z.-Y. Hou, and P. Peng, J. Non-Cryst. Solids **354**, 3705 (2008).
[40] Z. Jian, J. Chen, F. Chang, Z. Zeng, T. He, and W. Jie, Sci. China: Technol. Sci. **53**, 3203 (2010).
[41] C. S. Liu, J. Xia, Z. G. Zhu, and D. Y. Sun, J. Chem. Phys. **114**, 7506 (2001).
[42] F. Li, X. Liu, H. Hou, G. Chen, and G. Chen, Intermetallics **19**, 630 (2011).
[43] J. Kim, T. Keyes, and J. E. Straub, J. Chem. Phys. **132**, 224107 (2010).
[44] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **329**, 261 (2000).
[45] J. Kim, J. E. Straub, and T. Keyes, J. Phys. Chem. B **116**, 8646 (2012).
[46] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. Lett. **110**, 210603 (2013).
[47] T. Vogel, Y. W. Li, T. Wüst, and D. P. Landau, Phys. Rev. E **90**, 023302 (2014).