Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations

Daniel K. Shenfeld,[1] Huafeng Xu,[1] Michael P. Eastwood,[1] Ron O. Dror,[1] and David E. Shaw[1,2,*]

[1]*D. E. Shaw Research, New York, New York 10036, USA*

[2]*Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032, USA*

In computational thermodynamics, a sequence of intermediate states is commonly introduced to connect two equilibrium states. We consider two cases where the choice of intermediate states is particularly important: minimizing statistical error in free-energy difference calculations and maximizing average acceptance probabilities in replica-exchange simulations. We derive bounds for these quantities in terms of the thermodynamic distance between the intermediates, and show that in both cases the intermediates should be chosen as equidistant points along a geodesic connecting the end states.

PACS number(s): 05.10.−a, 02.70.−c, 05.70.Ce

I. INTRODUCTION

Many computational thermodynamics applications involve the sampling of each of a sequence of ensembles lying between two equilibrium states. The choice of intermediate states can greatly affect the precision to which certain quantities can be derived from experimental measurements or computer simulations. In this work, we present results on how the intermediate states should be chosen in order to minimize the variance of computed free-energy differences (ΔF) and to (nearly) maximize average acceptance probabilities in replica-exchange (RE) simulations.

Estimation of ΔF between two equilibrium states is an important problem with many applications, including the prediction of binding affinities of pharmaceutical compounds [1]. The Bennett acceptance ratio method provides an asymptotically efficient estimator for ΔF that uses equilibrium samples at the two states [2,3], but converges slowly unless the two states overlap significantly in phase space. The same problem applies to other free-energy perturbation methods, thermodynamic integration techniques, and various other approaches to the estimation of free-energy differences. It is thus common to introduce a sequence of intermediate states in which adjacent states overlap extensively and thus have a small variance in their estimated ΔF. The intermediates should be chosen to minimize the total variance. Previous studies have suggested choosing intermediates so that the free-energy difference [4] or the entropy difference [5] between any two adjacent states is approximately equal; other heuristics have been proposed as well [6,7]. In general, however, these choices do not minimize the total variance. To obtain well-converged estimate of ΔF in practice, the intermediates are often determined by trial and error.

The choice of intermediate states is also important in RE simulations. RE enhances sampling by coupling the reference system to an "efficient sampler" [8], a replica of the system in a state where the sampling of the phase space is efficient, such as a state of high temperature ("temperature exchange") or a state where specific energy barriers are attenuated [9–12]. To allow more frequent exchanges between the efficient sampler and the reference system, intermediate replicas are introduced, with exchanges performed between adjacent replicas; we call such a sequence of replicas an *RE schedule*. For any given number of replicas, we wish to choose an RE schedule that maximizes the frequency of exchange between fixed end states. Extensive literature has been dedicated to the selection of intermediates in temperature exchange [13–17], but little is known in more general cases [18].

Suppose that each state of the system is characterized by a vector λ. The state of a gas, for example, might be represented by vector $\lambda = (T, P)$, where the coordinate T specifies the temperature of the system and the coordinate P specifies its pressure. In a computer simulation, the coordinates might include parameters of the Hamiltonian. Selecting the intermediate states involves two steps: choosing a path in coordinate space and assigning the intermediates to specific points along this path.

In this work we analyze the choice of intermediates using methods from information geometry. Prior work has described a Riemannian metric on the space of coordinates λ, called the *Fisher information metric* or *thermodynamic metric*, as described below. The associated Riemannian length is called the *thermodynamic length* [19]. It has been shown that the dissipation in a transition between two end states is minimized when following a geodesic—i.e., a path of minimal thermodynamic length—between them [20,21]. We show here that geodesics also play an important role in selecting the intermediate states used in free-energy calculations and RE simulations: the choice of equidistant intermediates along the geodesic connecting the end states both minimizes the variance in the calculated free energy (Theorem 1) and provides an "almost optimal" RE schedule (Theorem 2).

II. PRELIMINARIES AND NOTATION

Suppose that the equilibrium states are determined by a vector $\lambda = (\lambda^1, \ldots, \lambda^d)$, taking values in a convex set $D \subseteq \mathbb{R}^d$. Let h_λ be the Hamiltonian of the system at λ, multiplied by the inverse temperature $\beta = (k_B T)^{-1}$, where k_B is the Boltz-

---

*Author to whom correspondence should be addressed. David.Shaw@DEShawResearch.com

mann constant and T is the temperature. The equilibrium probability density at a point x in phase space is $p_\lambda(x) = Z_\lambda^{-1} \exp[-h_\lambda(x)]$, where $Z_\lambda = \int e^{-h_\lambda(x)} dx$ is the partition function. Let $f_\lambda = -\ln Z_\lambda$ be the free energy multiplied by β, and $\ell_\lambda(x) = \ln p_\lambda(x) = f_\lambda - h_\lambda(x)$ be the log probability. In this notation, β may be a component of λ. We will refer to h as the Hamiltonian and to f as the free energy of the system. We denote differentiation with respect to λ^i by ∂_i, and second derivatives by ∂_{ij}. We assume throughout that we can differentiate with respect to λ^i under the integral sign.

The *Fisher information matrix g* at λ is defined by

$$g(\lambda)_{ij} \equiv \mathrm{cov}_\lambda(\partial_i \ell_\lambda, \partial_j \ell_\lambda) = \langle \partial_i \ell_\lambda(x) \cdot \partial_j \ell_\lambda(x) \rangle_\lambda, \quad (1)$$

where angled brackets denote expectation with respect to p_λ (the second equality follows because $\langle \partial_i \ell_\lambda(x) \rangle_\lambda = 0$). Under a few mild regularity conditions ([22], 2.1), g is a well-defined positive-definite quadratic form varying smoothly with λ, endowing D with a Riemannian metric, called the *Fisher information metric*. The norm induced by g at $\lambda \in D$ is denoted $\|\cdot\|_\lambda$. The Riemannian length of a path $\gamma(t)$ in D ($0 \le t \le 1$) is

$$\mathcal{L}(\gamma) \equiv \int_0^1 \|\dot{\gamma}\|_\gamma dt = \int_0^1 \sqrt{\sum_{i,j} \dot{\gamma}^i g(\gamma)_{ij} \dot{\gamma}^j} dt, \quad (2)$$

where $\dot{\gamma} = d\gamma/dt$. The Riemannian distance between λ_a and λ_b is defined as the length of the connecting geodesic, and is denoted $\mathcal{L}(\lambda_a, \lambda_b)$.

The thermodynamic metric is closely related to the thermal fluctuations. This is most apparent when h_λ is linear in λ, i.e., $h_\lambda(x) = \Sigma_i \lambda^i X_i(x)$, where X_i are functions of the microstate x alone. In this case we say that the system is in *Gibbs ensemble*, and we have

$$g(\lambda)_{ij} = -\partial_{ij} f(\lambda) = \mathrm{cov}_\lambda(X_i, X_j), \quad (3)$$

emphasizing the connection between the metric and fluctuations. We define the energy coordinates by $U_i(\lambda) = \partial_i f_\lambda = \langle X_i \rangle_\lambda$; in the coordinates U, the metric is given by $g(U) = g(\lambda)^{-1}$ ([22], 3.4, 4.1).

As an example, consider n molecules of ideal gas in the isothermal-isobaric ensemble. Here, h is given by the enthalpy multiplied by β: $h(x) = \beta E(x) + \beta P V(x)$, so the system is in Gibbs ensemble with respect to the coordinates $(\beta, \beta P)$. The metric in these coordinates is

$$g(\beta, \beta P) = \begin{pmatrix} \mathrm{var}(E) & 0 \\ 0 & \mathrm{var}(V) \end{pmatrix} = \beta^{-2} \begin{pmatrix} C_V k_B^{-1} & 0 \\ 0 & n\kappa P^{-1} \end{pmatrix}, \quad (4)$$

where C_V is the heat capacity and κ is the isothermal compressibility. Geodesics for an ideal gas are illustrated in Fig. 1. The connection between the thermodynamic length and fluctuations is intuitive in the energy coordinates, where the metric is $g(\langle E \rangle, \langle V \rangle) = g(\beta, \beta P)^{-1} = \begin{pmatrix} \mathrm{var}(E)^{-1} & 0 \\ 0 & \mathrm{var}(V)^{-1} \end{pmatrix}$. The distance between two states differing in energy by $d\langle E \rangle$ and in volume by $d\langle V \rangle$ is $d\mathcal{L} = [d\langle E \rangle^2 / \mathrm{var}(E) + d\langle V \rangle^2 / \mathrm{var}(V)]^{1/2}$; the states are close if the fluctuations in energy and volume are large compared to the difference in their equilibrium values.
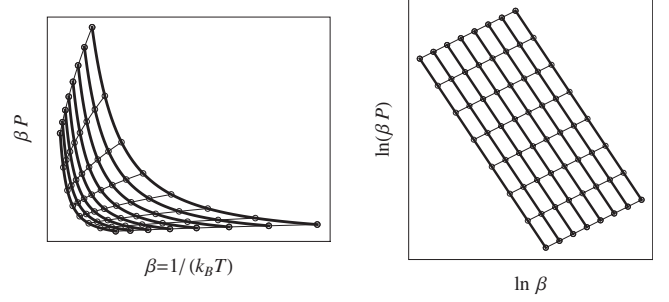

FIG. 1. Left: geodesics for an ideal gas in the $(\beta, \beta P)$ coordinates. Right: in the $[\ln \beta, \ln(\beta P)]$ coordinates, the metric is $g(\ln \beta, \ln(\beta P)) = \mathrm{diag}(C_V k_B^{-1}, n)$, and the geodesics are Euclidean straight lines. Quasistatic adiabatic expansions of an ideal gas follow geodesics (thick lines). Equidistant points on the geodesics are shown as circles.

III. MINIMIZING VARIANCE IN FREE-ENERGY CALCULATIONS

We study the estimation of the free-energy difference Δf between two states $\lambda_a, \lambda_b \in D$, based on samples drawn from a sequence of intermediate states, $\lambda_0, \lambda_1, \ldots, \lambda_k$, where $\lambda_0 = \lambda_a$ and $\lambda_k = \lambda_b$.

Theorem 1. Let $\Delta \hat{f}$ be an unbiased estimator for Δf based on a total of N observations. Then in the limit where the intermediates λ_i are close,

$$\mathrm{var}(\Delta \hat{f}) \ge \frac{\mathcal{L}(\lambda_a, \lambda_b)^2}{N}.$$

This bound can be achieved asymptotically as $N \to \infty$ by choosing the λ_i along the geodesic connecting λ_a and λ_b.

Our argument will also show that to minimize the variance, the amount of sampling for the estimation of $\Delta f_i = f_{\lambda_{i+1}} - f_{\lambda_i}$ should be proportional to $\mathcal{L}(\lambda_i, \lambda_{i+1})$. In particular, when sampling equally at each λ_i, the intermediates should be equidistant.

We first prove the theorem for a special case, which highlights that the thermodynamic length represents an inherent information-theoretical bound on the variance of a free-energy estimator. Suppose that the system is in Gibbs ensemble and $\Delta \hat{f}(x)$ is any unbiased estimator for Δf given an observation x drawn from λ_b. Let

$$\varphi_\lambda \equiv f_\lambda - f_a - \sum_i \partial_i f|_a \cdot (\lambda^i - \lambda_a^i). \quad (5)$$

Then we obtain an estimator for φ_b, namely: $\hat{\varphi}_b(x) = \Delta \hat{f}(x) - \Sigma_i \partial_i f|_a \cdot \Delta \lambda^i$, whose variance is the same as that of $\Delta \hat{f}$. Conversely, any unbiased estimator for φ_b gives rise to an unbiased estimator for Δf with the same variance. Note that $\partial_i \varphi|_b = \partial_i f|_b - \partial_i f|_a = \Delta U_i$.

By the Cramér-Rao lower bound [22], using $g(U) = g(\lambda)^{-1}$ and the invariance of the norm under coordinate transformations, we have

$$\mathrm{var}(\Delta\hat{f}) = \mathrm{var}(\hat{\varphi}_b) \geq \sum_{i,j} \partial_i \varphi|_b (g(\lambda_b)^{-1})_{ij} \partial_j \varphi|_b$$

$$= \sum_{i,j} \Delta U_i g(U^b)_{ij} \Delta U_j = \|\Delta U\|_{U^b}^2 = \|\Delta\lambda\|_{\lambda_b}^2. \quad (6)$$

Suppose now that we allow sampling at multiple intermediates and draw n_i observations at each λ_i. By Eq. (6), the variance for the estimation of Δf_i is bounded by $\|\Delta\lambda_i\|_{\lambda_i}^2 / n_i$. Minimizing subject to the constraint that $N = \Sigma_{i=1}^k n_i$ is a constant, we obtain that n_i should be proportional to $\|\Delta\lambda_i\|_{\lambda_i}$. In this case, assuming independent sampling at the various λ_i, we have the bound

$$\mathrm{var}(\Delta\hat{f}) = \sum_i \mathrm{var}(\Delta\hat{f}_i) \geq \frac{\left(\Sigma_i \|\Delta\lambda_i\|_{\lambda_i}\right)^2}{N}. \quad (7)$$

This bound is rigorous. If the λ_i are close enough and situated along a path γ, we obtain in the limit $k \rightarrow \infty$

$$\mathrm{var}(\Delta\hat{f}) \geq \frac{\mathcal{L}(\gamma)^2}{N}. \quad (8)$$

Equality can be achieved asymptotically as $N \rightarrow \infty$, e.g., by using a maximum-likelihood estimator (the needed regularity conditions hold by our hypotheses on g [22]). The variance is minimized when γ is a geodesic.

We now turn to the general case, where the system is not necessarily in Gibbs ensemble. Consider the estimation of Δf without intermediates, based on a maximum-likelihood estimator given N samples. We assume for simplicity a sampling protocol where the two states are sampled equally; similar results hold for different ratios as well. The Cramér-Rao lower bound gives again an estimate for the variance [3],

$$N \cdot \mathrm{var}(\Delta\hat{f}) \geq 2\langle(1 + \cosh\Delta\ell)^{-1}\rangle^{-1} - 4 \approx 4\langle 1 - \Delta^2\ell/4\rangle^{-1} - 4$$

$$\approx \langle\Delta^2\ell\rangle, \quad (9)$$

where the average is taken over all samples; the approximations are obtained by expanding to second order in $\Delta\ell$. In the limit of many samples, it follows from the protocol that $\langle\Delta^2\ell\rangle \rightarrow (\langle\Delta^2\ell\rangle_a + \langle\Delta^2\ell\rangle_b)/2$. Expanding in $\Delta\lambda$, the first-order term of $\langle\Delta\ell\rangle_a$ vanishes, so up to second order,

$$\langle\Delta^2\ell\rangle_a \approx \mathrm{var}(\Delta\ell)_a \approx \|\Delta\lambda\|_a^2, \quad (10)$$

where the second approximation follows directly from the definition of the metric. Combining this result with its counterpart for $\langle\Delta^2\ell\rangle_b$ in Eq. (9) we finally obtain

$$N \cdot \mathrm{var}(\Delta\hat{f}) \approx \frac{\|\Delta\lambda\|_a^2 + \|\Delta\lambda\|_b^2}{2}. \quad (11)$$

This approximation becomes arbitrarily good as $\Delta\lambda$ vanishes. The theorem now follows in the same way as it does from Eq. (6).

IV. MAXIMIZING THE TOTAL ACCEPTANCE IN RE SIMULATIONS

Let p_a, p_b be the distributions defined by two replicas whose Hamiltonians are parametrized by λ_a, λ_b. The Me-

tropolis acceptance probability for the exchange attempt $(x,y) \rightarrow (y,x)$ is $\min(1, \frac{p_a(y)p_b(x)}{p_a(x)p_b(y)})$, so the *average acceptance probability* is

$$\mathcal{A}(\lambda_a, \lambda_b) \equiv \int_x \int_y \min(p_a(x)p_b(y), p_a(y)p_b(x)) dx dy. \quad (12)$$

Given a sequence of replicas along a path γ, $\lambda_0 = \lambda_a, \lambda_1, \ldots, \lambda_k = \lambda_b$, we define the *total acceptance* by $\mathcal{A}(\gamma) = \Pi_i \mathcal{A}(\lambda_i, \lambda_{i+1})$. This is the probability that a state will propagate from λ_0 to λ_k in k consecutive attempts. $\mathcal{A}(\gamma)$ is not by itself an optimal measure for the efficiency of an RE schedule, as it does not take into account the number of replicas and the sampling quality. Other measures, such as the flux, the effective fraction, and the round-trip time [16,17,23], are more suitable for this end. Nevertheless, given k, maximizing $\mathcal{A}(\gamma)$ is equivalent to optimizing these other measures. We note that the sampling efficiency of RE is only accurately reflected by the above measures under the assumption that at least one replica is a perfect sampler [17].

Theorem 2. The total acceptance $\mathcal{A}(\gamma)$ is maximized when the λ_i are equidistant, and we have the bounds

$$-\frac{\mathcal{L}^2(\gamma)}{2k} - \frac{\mathcal{L}(\gamma)}{\sqrt{2}} \leq \ln\mathcal{A}(\gamma) \leq -\frac{\mathcal{L}^2(\gamma)}{2k}.$$

Once again, the bounds are maximized when γ is a geodesic. Even though sampling along a geodesic may not guarantee an optimal total acceptance, any other schedule cannot improve on it by a factor of more than $\exp(\mathcal{L}(\lambda_a, \lambda_b)/\sqrt{2})$.

For the proof, we examine the overlap integrals

$$J_q(\lambda_a, \lambda_b) \equiv \int 2^{-1/q}(p_a^q + p_b^q)^{1/q} dx \quad (q \leq 0). \quad (13)$$

J_q is nondecreasing in q. We shall use the case $q = -1$ and the limiting case $q \rightarrow -\infty$, given explicitly by

$$J_{-1} = \int \frac{2p_a p_b}{p_a + p_b} dx, \quad J_{-\infty} = \int \min(p_a, p_b) dx. \quad (14)$$

Using Jensen's inequality twice, we have:

$$\ln J_{-\infty} = \ln\langle\min(1, e^{\Delta\ell})\rangle_a \geq \langle\min(0, \Delta\ell)\rangle_a = \frac{1}{2}\langle\Delta\ell\rangle_a$$

$$-\frac{1}{2}\langle|\Delta\ell|\rangle_a \geq \frac{1}{2}\langle\Delta\ell\rangle_a - \frac{1}{2}\sigma(\Delta\ell)_a, \quad (15)$$

where $\sigma(\Delta\ell)$ is the standard deviation; by Eq. (10), $\sigma(\Delta\ell)_a \approx \|\Delta\lambda\|_a$. Up to second order, $\langle\Delta\ell\rangle_a \approx -\|\Delta\lambda\|_a^2/2$, as can be seen by noting that $\langle\Delta\ell\rangle_a$ is the negative of the Kullback-Leibler divergence from p_a to p_b ([22], 3.5). It follows that for λ_a, λ_b close enough,

$$\ln J_{-\infty} \geq -\frac{\|\Delta\lambda\|_a^2}{4} - \frac{\|\Delta\lambda\|_a}{2}. \quad (16)$$

For J_{-1}, we write $\ln J_{-1} = \ln\langle 2/(1 + e^{-\Delta\ell})\rangle_a$, expand to second order and use our previous approximations,

$$\ln J_{-1} \approx \frac{\langle \Delta \ell \rangle_a}{2} - \frac{\langle \Delta \ell \rangle_a^2}{8} \approx -\frac{\|\Delta \lambda\|_a^2}{4}. \tag{17}$$

Putting everything together, we obtain for λ_a, λ_b close enough and $\mathcal{L} = \mathcal{L}(\lambda_a, \lambda_b)$,

$$-\frac{\mathcal{L}^2}{4} - \frac{\mathcal{L}}{2} \leq \ln J_{-\infty} \leq \ln J_{-1} \approx -\frac{\mathcal{L}^2}{4}. \tag{18}$$

To make connection to the average acceptance probability $\mathcal{A}(\lambda_a, \lambda_b)$, note that we may regard $D \times D$ as a manifold of distributions, where (λ_a, λ_b) corresponds to the product density $p_a(x) p_b(y)$. Defining J_q on this manifold as before, it is clear from the definition (12) that

$$\mathcal{A}(\lambda_a, \lambda_b) = J_{-\infty}((\lambda_a, \lambda_b), (\lambda_b, \lambda_a)). \tag{19}$$

The independence of $p_a(x)$ and $p_b(y)$ implies that the Fisher metric on $D \times D$ is simply the product metric, given by $g(\lambda_a, \lambda_b) = \begin{pmatrix} g(\lambda_a) & 0 \\ 0 & g(\lambda_b) \end{pmatrix}$. If γ is a geodesic in $D \times D$, then by minimality the projections on each factor, γ_1, γ_2, are geodesics in D. It can be shown further that $\mathcal{L}_{D \times D}(\gamma) = (\mathcal{L}_D(\gamma_1)^2 + \mathcal{L}_D(\gamma_2)^2)^{1/2}$. It follows that

$$\mathcal{L}_{D \times D}((\lambda_a, \lambda_b), (\lambda_b, \lambda_a)) = \sqrt{2} \mathcal{L}_D(\lambda_a, \lambda_b).$$

Using Eq. (18) we obtain for $\mathcal{L} = \mathcal{L}_D(\lambda_a, \lambda_b)$,

$$-\frac{\mathcal{L}^2}{2} - \frac{\mathcal{L}}{\sqrt{2}} \leq \ln \mathcal{A}(\lambda_a, \lambda_b) \leq -\frac{\mathcal{L}^2}{2}. \tag{20}$$

The proof of theorem 2 now follows by applying these bounds to all acceptances $\mathcal{A}(\lambda_i, \lambda_{i+1})$ and maximizing the resultant bounds of $\mathcal{A}(\gamma)$ subject to the constraint that the length of γ is constant.

As an example, consider temperature exchange. The Hamiltonian depends linearly on a single parameter, the inverse temperature β. The metric in the coordinate $\ln \beta$ is $g(\ln \beta) = C_V(\beta) k_B^{-1}$, where $C_V(\beta)$ is the heat capacity. If $C_V(\beta)$ is constant, $\mathcal{L}(\beta_1, \beta_2)$ is proportional to $|\ln \beta_2 - \ln \beta_1|$. Hence, in order to obtain equidistant replicas, the temperatures should form a geometric progression, in accordance with previous results [13,14].

V. CONCLUSION

Previous studies have demonstrated examples in which well-chosen intermediates significantly improve the convergence of the estimated ΔF [5] and the RE sampling efficiency [15]. The present work shows how to select a nearly optimal set of intermediates in the general cases. We have shown that in the limit of a large number of intermediate states, choosing equidistant intermediates along the geodesic both minimizes the total variance in the estimate of ΔF and guarantees that the total acceptance probability in RE simulations differs by no more than a system-dependent constant factor from the optimum. We suggest the following method to obtain such intermediates. First, conduct short simulations on a lattice in coordinate space and estimate the thermodynamic metric at each lattice point. After refining the lattice as needed, construct the shortest path between the end states using the metric estimates and select equidistant points along it as the intermediate states. Further development of efficient methods to estimate the thermodynamic metric will be of interest for purposes of determining accurate geodesics.

ACKNOWLEDGMENTS

[1] P. Kollman, Chem. Rev. (Washington, D.C.) **93**, 2395 (1993).
[2] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[3] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).
[4] D. A. Pearlman and P. A. Kollman, J. Chem. Phys. **90**, 2460 (1989).
[5] N. Lu and D. A. Kofke, J. Chem. Phys. **111**, 4414 (1999).
[6] R. J. Radmer and P. A. Kollman, J. Comput. Chem. **18**, 902 (1997).
[7] D. Wu and D. A. Kofke, J. Chem. Phys. **123**, 084109 (2005).
[8] C. J. Geyer, in *Computing Science and Statistics Proceedings of the 23rd Symposium on the Interface* (American Statistical Association, New York, 1991), pp. 156–163.
[9] Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. **113**, 6042 (2000).
[10] H. Fukunishi, O. Watanabe, and S. Takada, J. Chem. Phys. **116**, 9058 (2002).
[11] R. Affentranger, I. Tavernelli, and E. E. D. Iorio, J. Chem. Theory Comput. **2**, 217 (2006).
[12] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).
[13] D. A. Kofke, J. Chem. Phys. **117**, 6911 (2002).
[14] C. Predescu, M. Predescu, and C. V. Ciobanu, J. Chem. Phys. **120**, 4119 (2004).
[15] A. Kone and D. A. Kofke, J. Chem. Phys. **122**, 206101 (2005).
[16] N. Rathore, M. Chopra, and J. J. de Pablo, J. Chem. Phys. **122**, 024111 (2005).
[17] C. Predescu, M. Predescu, and C. V. Ciobanu, J. Phys. Chem. B **109**, 4189 (2005).
[18] W. Nadler and U. H. E. Hansmann, Phys. Rev. E **76**, 065701(R) (2007).
[19] F. Weinhold, J. Chem. Phys. **63**, 2479 (1975).
[20] P. Salamon and R. S. Berry, Phys. Rev. Lett. **51**, 1127 (1983).
[21] G. E. Crooks, Phys. Rev. Lett. **99**, 100602 (2007).
[22] S.-I. Amari, *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics Vol. 28 (Springer-Verlag, Berlin, Heidelberg, 1985).
[23] H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, J. Stat. Mech.: Theory Exp. (2006) P03018.