Sigma method for the microcanonical entropy or density of states

Rasmus A. X. Persson[*]

Department of Chemistry and Molecular Biology, University of Gothenburg, SE-412 96 Gothenburg, Sweden

We introduce a simple improvement on the method to calculate equilibrium entropy differences between classical energy levels proposed by Davis [S. Davis, Phys. Rev. E **84**, 050101 (2011)]. We demonstrate that the modification is superior to the original whenever the energy levels are sufficiently closely spaced or whenever the microcanonical averaging needed in the method is carried out by importance sampling Monte Carlo. We also point out the necessary adjustments if Davis's method (improved or not) is to be used with molecular dynamics simulations.

Consider a system with configurational coordinates $\{\mathbf{r}_i\}$ and potential energy function $U(\{\mathbf{r}_i\})$. The Hamiltonian of the system is of the standard classical form, that is, separable in its coordinates and conjugate momenta, $\{\mathbf{p}_i\}$, so that it may be written thus:

$$H(\{\mathbf{r}_i\},\{\mathbf{p}_i\}) = U(\{\mathbf{r}_i\}) + K(\{\mathbf{p}_i\}), \tag{1}$$

where $K(\{\mathbf{p}_i\})$ is the kinetic energy of the system. In this Brief Report, we consider the calculation of the energy dependence of the microcanonical Boltzmann-Planck equilibrium entropy:

$$S(E) = k \ln \omega(E), \tag{2}$$

where k is Boltzmann's constant and

$$\omega(E) = C \int \{d\mathbf{r}_i\}\{d\mathbf{p}_i\}\delta(E - H(\{\mathbf{r}_i\},\{\mathbf{p}_i\})) \tag{3}$$

is the phase density, also known as the density of states, where C is a constant that assures $\omega(E)$ is dimensionless and δ is Dirac's δ function. Algorithms to evaluate $\omega(E)$ [and thus $S(E)$] abound in the literature [1–17]. They each have their advantages and drawbacks, and an exhaustive review of them all is not possible in this Brief Report. Here, we instead focus in particular on the recently proposed σ method by Davis [18]. We will recapitulate its derivation and offer an improvement on the original method. Our notation differs slightly from that of Davis [18].

For the moment, we consider a microcanonical ensemble whose only first integral of motion is the total energy, E. We will briefly consider the case with more first integrals of motion later. The Laplace principle of indifference assigns equal *a priori* probability to all phase space points on the energy shell $H(\{\mathbf{r}_i\},\{\mathbf{p}_i\}) = E$. In other words, the ensemble probability density is constant on this energy shell and zero everywhere else. We write this probability density as

$$W_E(\{\mathbf{r}_i\},\{\mathbf{p}_i\}) = \frac{C}{\omega(E)}\delta(E - H(\{\mathbf{r}_i\},\{\mathbf{p}_i\})). \tag{4}$$

If $K(\{\mathbf{p}_i\})$ is quadratic in each conjugate momentum coordinate and shows no complicated interdependencies (in this

equation $\{m_i\}$ are generalized masses),

$$K(\{\mathbf{p}_i\}) = \sum_i \frac{\mathbf{p}_i^2}{2m_i}, \tag{5}$$

the dependence on $\{\mathbf{p}_i\}$ can be integrated out, yielding [18–21]

$$\widehat{W}_E(\{\mathbf{r}_i\}) = C'_E(E - U(\{\mathbf{r}_i\}))^{n/2-1}\Theta(E - U(\{\mathbf{r}_i\})), \tag{6}$$

where

$$C'_E = \left(\int \{d\mathbf{r}_i\}(E - U(\{\mathbf{r}_i\}))^{n/2-1}\Theta(E - U(\{\mathbf{r}_i\}))\right)^{-1} \tag{7}$$

is a normalization constant that is inversely proportional to $\omega(E)$, Θ is the Heaviside step function, and n is the number of configurational degrees of freedom of the system (which for an unconstrained particle system is three times the number of particles in three dimensions). The quantity $\widehat{W}_E(\{\mathbf{r}_i\})$ is directly proportional to the density of kinetic energy states. The microcanonical average of a quantity $A(\{\mathbf{r}_i\})$ that does not depend explicitly on the momenta can now be expressed as

$$\langle A(\{\mathbf{r}_i\})\rangle_E = \int \{d\mathbf{r}_i\}\widehat{W}_E(\{\mathbf{r}_i\})A(\{\mathbf{r}_i\}). \tag{8}$$

The probability function in Eq. (6) can be used as the weighting factor in a microcanonical Markov chain Monte Carlo simulation [19–21] to calculate averages according to Eq. (8). In a molecular dynamics simulation, however, in which additional integrals of motion appear, the probability function of Eq. (6) is not the proper one, assumed ergodicity notwithstanding. In this case, the correct probability function is given by [22]

$$\widetilde{W}_E(\{\mathbf{r}_i\}) = \widetilde{C}_E\left(E - U(\{\mathbf{r}_i\}) - \frac{\mathbf{P}^2}{2M}\right)^{n/2-1}, \tag{9}$$

where $\mathbf{P}$ is the center-of-mass momentum, M is the total mass, $\widetilde{C}_E$ is a normalization constant, and n carries the same meaning as in Eq. (6) but does not correspond to the same numerical value, there being one degree of freedom less for each Cartesian component of the center-of-mass momentum.

At this point, Davis [18] introduces the quantity

$$\sigma_{E,E_r}(\{\mathbf{r}_i\}) = \frac{\Theta(E_r - U(\{\mathbf{r}_i\}))}{(E - U(\{\mathbf{r}_i\}))^{n/2-1}}, \tag{10}$$

[*]rasmusp@chem.gu.se

with the condition that $E_\mathrm{r} \leqslant E$ but otherwise arbitrary. When the microcanonical average of Eq. (10) is calculated using Eqs. (8) and (6), keeping in mind that $C'_E \propto 1/\omega(E)$, it is seen that that the entropy difference according to Eq. (2) between two energy levels E' and E'' is given by

$$\Delta_{E'}^{E''} S = k \ln \frac{\langle \sigma_{E',E_\mathrm{r}}(\{\mathbf{r}_i\})\rangle_{E'}}{\langle \sigma_{E'',E_\mathrm{r}}(\{\mathbf{r}_i\})\rangle_{E''}}, \quad (11)$$

which is the rational for introducing the σ function. Similar to the procedure in Ref. [18], let us introduce the quantity

$$\Sigma_{E'',E'}(\{\mathbf{r}_i\}) = \frac{(E' - U(\{\mathbf{r}_i\}))^{n/2-1}}{(E'' - U(\{\mathbf{r}_i\}))^{n/2-1}} \Theta(E' - U(\{\mathbf{r}_i\})), \quad (12)$$

with $E' \leqslant E''$. We may then write

$$\Delta_{E'}^{E''} S = -k \ln\langle \Sigma_{E'',E'}(\{\mathbf{r}_i\})\rangle_{E''}. \quad (13)$$

The proof of this equation follows directly from the substitution of Eq. (12) into Eq. (13), from which Eq. (7) can be identified, and, after using the inverse proportionality between C'_E and $\omega(E)$, this leads to Eq. (2) in difference form. Clearly, both Eqs. (11) and (13) may be used to calculate the entropy difference. Shortly, we will consider the question of which function is the most efficient from a computational perspective.

The above equations are to be used when the total energy is the only integral of motion in the mechanical system. For completeness, we note the form that the corresponding sigma functions must take when the averaging is done by molecular dynamics means, if the objective is to obtain the density of states. In this case, the σ function becomes

$$\widetilde{\sigma}_{E,E_\mathrm{r}}(\{\mathbf{r}_i\}) = \frac{\Theta\left(E_\mathrm{r} - U(\{\mathbf{r}_i\}) - \frac{\mathbf{P}^2}{2M}\right)}{\left(E - U(\{\mathbf{r}_i\}) - \frac{\mathbf{P}^2}{2M}\right)^{n/2-1}}, \quad (14)$$

and the Σ function is to be replaced by

$$\widetilde{\Sigma}_{E'',E'}(\{\mathbf{r}_i\}) = \frac{\left(E' - U(\{\mathbf{r}_i\}) - \frac{\mathbf{P}^2}{2M}\right)^{n/2-1}}{\left(E'' - U(\{\mathbf{r}_i\}) - \frac{\mathbf{P}^2}{2M}\right)^{n/2-1}}$$
$$\times \Theta\left(E' - U(\{\mathbf{r}_i\}) - \frac{\mathbf{P}^2}{2M}\right). \quad (15)$$

Once again, heed must be paid to the value of n so that the subtraction of the center-of-mass momentum degrees of freedom is accounted for. In every other respect, the equations for the entropy differences remain formally unchanged. Quite conceivably, one might want to extract the corresponding entropy of the system without these additional first integrals, in which case one may introduce the function

$$\widetilde{\Sigma}'_{E'',E'}(\{\mathbf{r}_i\}) = \frac{(E' - U(\{\mathbf{r}_i\}))^{n'/2-1}}{\left(E'' - U(\{\mathbf{r}_i\}) - \frac{\mathbf{P}^2}{2M}\right)^{n/2-1}} \Theta(E' - U(\{\mathbf{r}_i\}))$$
$$(16)$$

and calculate entropies by Eq. (13). Here n' exceeds n by the number of Cartesian components in the center-of-mass momentum. The corresponding form for Eq. (11) follows by analogy.

We now turn to an analysis of the relative computational merits of Eqs. (11) and (13). To obtain $S(E)$ as a (quasi-)continuous function of E, the calculation may be subdivided into N discrete energy segments over a predefined energy range. For instance, if the energy interval $[E', E'']$ is subdivided into N segments separated at energies $\{E_i\}$ such that $E_0 = E'$; $E_i < E_{i+1}$; $i = 0, 1, \ldots, N - 1$; and $E_N = E''$ then the total entropy difference is given as a sum over the individual entropy differences for each segment:

$$\Delta_{E'}^{E''} S = \sum_{i=0}^{N-1} \Delta_{E_i}^{E_{i+1}} S. \quad (17)$$

With Eq. (11) $N + 1$, and with Eq. (13) N averages are needed. This difference becomes negligible for large N, which often corresponds to the most interesting situations. In the limit $N \to \infty$, keeping E' and E'' fixed, this gives $S(E)$ as a continuous function of E in the interval $[E', E'']$. We now note that, as $N \to \infty$, $E_{i+1} - E_i \to 0$, and in this limit $\Sigma_{E_{i+1},E_i}(\{\mathbf{r}_j\}) \to 1$ for all $\{\mathbf{r}_j\}$ accessible in the microcanonical ensemble (that is, for all $\{\mathbf{r}_j\}$ such that $U(\{\mathbf{r}_j\}) \leqslant E_{i+1}$), and because the Σ function being averaged becomes identically unity the average $\langle \Sigma_{E_{i+1},E_i}(\{\mathbf{r}_j\})\rangle_{E_{i+1}}$ loses all statistical uncertainty. Before we continue, we note that this quality is not assured for the averages over the corresponding σ functions, as they do not enjoy the same guarantee, and is even less assured for their ratio.

In order to complete and strengthen the general argument, we should sum up and consider the uncertainties of all the N individual averages. Therefore, considering the rate by which the averages of the Σ functions approach unity (and lose their statistical uncertainty) is of importance. Rearranging Eq. (17), it is clear that

$$\ln\langle \Sigma_{E_{i+1},E_i}\rangle_{E_{i+1}} \sim \frac{\Delta_{E'}^{E''} S}{kN}. \quad (18)$$

In other words, the logarithm of each individual Σ average tends to zero inversely proportionally to N. In the numerical implementation, errors will accrue if the ratio on the right-hand side becomes of the order of the machine precision. To keep the notation as simple as possible, we temporarily restrict our attention to $n = 2$, in which case the formulas are drastically simplified. In this case, for instance, the uncertainty of the averages may be estimated from $\langle \Theta(E_i - U(\{\mathbf{r}_j\}))\rangle_{E_{i+1}}$. For N large enough, the converged value of $\langle \Theta(E_i - U(\{\mathbf{r}_j\}))\rangle_{E_{i+1}}$ will be very close to, but slightly less than, unity. With finite statistics, we estimate this value to be α_i. Because of our choice of $n = 2$, the average in question is composed only of terms being either unity or zero. If there are M_i terms equal to unity and m_i terms equal to zero sampled in the numerical averaging, then $\alpha_i = M_i/(M_i + m_i)$. In the "worst case scenario," the statistics of the ensemble averaging is so poor (because of $M_i + m_i$ being chosen too small) that α_i is virtually a nonuniform random number between zero and one. This value is quite likely different from the actual converged value, which we denote β_i. The average magnitude of this error is a measure of the uncertainty in the averaging. The variance of the relative error, for instance, can be formulated as [23]

$$\mathrm{var}\left(\frac{\alpha_i}{\beta_i}\right) \equiv \left\langle \frac{\alpha_i^2}{\beta_i^2}\right\rangle - 1 = \int \mathrm{d}\{\mathbf{r}_j\} \widehat{W}_{E_{i+1}}(\{\mathbf{r}_j\})$$
$$\times \frac{\Theta(E_i - U(\{\mathbf{r}_j\}))}{\beta_i^2} - 1. \quad (19)$$

Inserting the expression for $\widehat{W}_{E_{i+1}}$, we find that the integral on the right-hand side is

$$\int d\{\mathbf{r}_j\}\widehat{W}_{E_{i+1}}(\{\mathbf{r}_j\})\Theta(E_i - U(\{\mathbf{r}_j\}))$$

$$= \frac{\int d\{\mathbf{r}_j\}\Theta(E_i - U(\{\mathbf{r}_j\}))}{\int d\{\mathbf{r}'_j\}\Theta(E_i + \Delta_N E - U(\{\mathbf{r}'_j\}))}, \quad (20)$$

where we have introduced $\Delta_N E = (E'' - E')/N$.

We cannot hope to solve the integral in Eq. (20) in the general case, and like this obtain the variance as an explicit function of N. There are, however, some conclusions to be drawn from the general form of the right-hand side. In molecular systems, the accessible configuration space generally increases superlinearly with increasing potential energy. Hence, the integral in the denominator of Eq. (20) should increase superlinearly with increasing $\Delta_N E$. It follows immediately that the variance according to Eq. (19) should decrease superlinearly with decreasing $\Delta_N E \propto N^{-1}$ or, in other words,

$$\frac{\int d\{\mathbf{r}_j\}\Theta(E_i - U(\{\mathbf{r}_j\}))}{\int d\{\mathbf{r}'_j\}\Theta(E_i + \Delta_N E - U(\{\mathbf{r}'_j\}))} = 1 + O(N^{-a}), \quad (21)$$

where $a > 1$ is undetermined (but assuredly greater than unity). Thus, the total variance (given as N times the individual variance) will decrease to zero as $O(N^{1-a})$ when $N \to \infty$. It follows that for a sufficiently finely meshed energy grid the Σ function will always be computationally more efficient than the σ function, regardless of the complexity of the system, *as long as its accessible configuration space increases superlinearly with increasing potential energy*. The general argument, but with clumsier notation, can be carried through also with $n \neq 2$.

In the numerical implementation, the limit $N \to \infty$ may of course not be reached exactly and so a superior computational efficacy in the numerically very demanding $N \to \infty$ limit is not necessarily relevant in actual calculations. We must therefore also consider the relative efficacy of Eqs. (11) and (13) for finite energy differences.

As discussed by Davis in Ref. [18], in the case of Eq. (11), the constant E_r must be chosen so that both averages under the logarithm are calculated with enough statistics. Too small values of E_r restrict the statistics sampled, as not enough sampled configurations do then have potential energies $U(\{\mathbf{r}_i\}) \leqslant E_r$. At the same time E_r must be less than or equal to the smallest of the two energies for which the entropy difference is calculated, meaning that a too large energy gap will be detrimental to the statistics of the higher energy average. This essentially introduces an upper bound for the energy difference for which the entropy difference can be reliably calculated. As noted by Davis in Ref. [18], this upper bound will depend on the size of the system, because the fluctuations in potential energy become smaller the larger the system is (in the sense of the value of n). We note that a similar restriction (for the same reasons) applies to Eq. (13), in which E' takes the place of E_r.

To consider the question of convergence of the averages in more detail we restrict our attention somewhat and assume that the microcanonical statistics are sampled by importance sampling Metropolis Monte Carlo according to the probability function $\widehat{W}_E(\{\mathbf{r}_i\})$ [24]. In this case, a statistically good
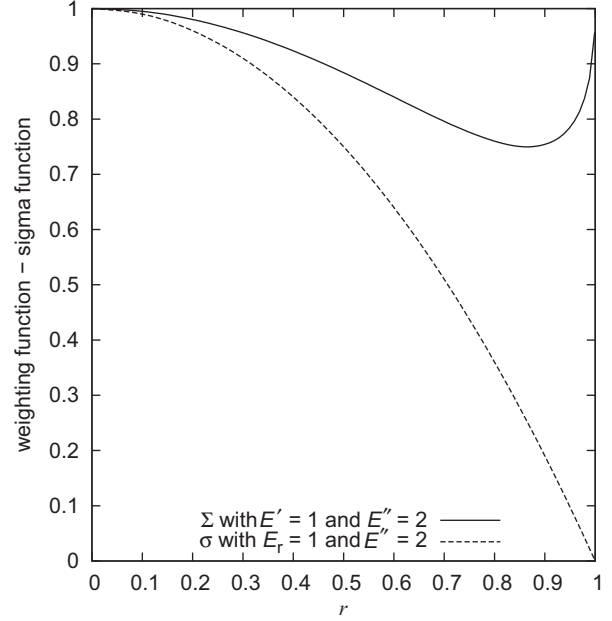

FIG. 1. Plot of the difference between either of the two sigma functions for specific and arbitrarily chosen $E'' = 2, E' = E_r = 1$ and $U(r) = r^2$ and the importance sampling function (with $n = 3$) used in microcanonical Monte Carlo simulations. In this plot, $C'_{E''}$ is arbitrarily set to $C'_{E''} = 1$.

estimate of the average of a function $A(\{\mathbf{r}_i\})$ is obtained if $A(\{\mathbf{r}_i\})$ contributes appreciably in regions where $\widehat{W}_E(\{\mathbf{r}_i\})$ is large and likewise contributes negligibly in regions where $\widehat{W}_E(\{\mathbf{r}_i\})$ is close to zero. The question thus reduces to which of the two sigma functions is most "similar" to $\widehat{W}_E(\{\mathbf{r}_i\})$, in the sense that they share the domains where they are both of appreciable magnitude. For instance, consider the ratios between the sigma functions and the Markov weighting function:

$$\frac{\Sigma_{E'',E'}(\{\mathbf{r}_i\})}{\widehat{W}_{E''}(\{\mathbf{r}_i\})} = \frac{(E' - U(\{\mathbf{r}_i\}))^{n/2-1}}{C'_{E''}(E'' - U(\{\mathbf{r}_i\}))^{n-2}}\frac{\Theta(E' - U(\{\mathbf{r}_i\}))}{\Theta(E'' - U(\{\mathbf{r}_i\}))}, \quad (22)$$

$$\frac{\sigma_{E'',E_r}(\{\mathbf{r}_i\})}{\widehat{W}_{E''}(\{\mathbf{r}_i\})} = \frac{\Theta(E_r - U(\{\mathbf{r}_i\}))}{C'_{E''}(E'' - U(\{\mathbf{r}_i\}))^{n-2}}, \quad (23)$$

$$\frac{\sigma_{E',E_r}(\{\mathbf{r}_i\})}{\widehat{W}_{E'}(\{\mathbf{r}_i\})} = \frac{\Theta(E_r - U(\{\mathbf{r}_i\}))}{C'_{E'}(E' - U(\{\mathbf{r}_i\}))^{n-2}}. \quad (24)$$

The less similar the two functions are, the less constant is their ratio. In the simplest case, the limiting case of an ideal gas, the $\{\mathbf{r}_i\}$ gradients vanish for all of these ratios and the relative qualities of the importance sampling of the averages are not distinguishable between the σ and Σ functions. When interactions are present, this is no longer the case. Whereas the resulting ratios of both Eqs. (23) and (24) consist of a more-or-less constant numerator and a monotonously and smoothly decreasing denominator as a function of $U(\{\mathbf{r}_j\})$, the ratio between the $\Sigma_{E'',E'}$ function and $\widehat{W}_{E''}$ presents a smoothly decreasing function of U for both numerator and denominator. Hence, it would seem that this ratio is more invariant with respect to changes in $\{\mathbf{r}_j\}$

[and hence $U(\{\mathbf{r}_j\})$] than the others. This becomes, once again, particularly pronounced when $E' \to E''$. A more-or-less constant difference can be used just as well as an indication of similarity, and this is what we consider in Fig. 1 in the case of a three-dimensional harmonic oscillator for which $U(r) = r^2$. This is a model potential for atomic crystals and makes for a reasonably relevant comparison. As anticipated, the difference $\Sigma_{E'',E'} - W_{E''}$ exhibits much less variation than $\sigma_{E'',E'} - W_{E''}$. Above $E = E' = E_r$, they become identical.

We do not offer any numerical experiments to illustrate the method. This has already been achieved by Davis [18] on a nontrivial system using the σ function. However, the numerical upper limitation on the size of tractable systems that Davis [18] points out is nonetheless important to recall. A similar limitation, although much less severe, is present already in the microcanonical sampling algorithm [19–21], as acceptance probabilities for a trial move taking the system from potential energy U' to U'' at total energy E are proportional to the ratio

$$\left(\frac{E - U''}{E - U'}\right)^{n/2-1} \Theta(E - U''),$$

which for large n values ought to become difficult for the computer architecture to resolve to sufficient accuracy, as the acceptance ratio takes on a more "step function"-like form. Severin *et al.* [19] initially introduced the sampling algorithm for sampling the internal degrees of freedom of single molecules. Obtaining the density of states of complicated polyatomics, needed for instance in statistical reaction rate theories, is thus a natural application of a method such as this. Nevertheless, Ray [21] reports comfortable simulations on up to 500 particles, using this microcanonical sampling. Such system sizes should be sufficient for many purposes in statistical mechanics.

In conclusion, we note one interesting formal property of the Σ averages: from a single microcanonical molecular dynamics (or Monte Carlo) run, *in principle* the entire $S(E)$ function is obtainable (up to an additive constant). This follows since the energy E' is arbitrary in Eq. (13), yet does not affect the dynamics. Nevertheless, it is clear from the limitations discussed above that good statistics would only be achieved in a narrow range below E''. However, for very small systems, this range might be quite broad.

[1] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).

[2] P. Labastie and R. L. Whetten, Phys. Rev. Lett. **65**, 1567 (1990).

[3] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).

[4] H.-P. Cheng, X. Li, R. L. Whetten, and R. S. Berry, Phys. Rev. A **46**, 791 (1992).

[5] R. Poteau, F. Spiegelmann, and P. Labastie, Z. Phys. D: At. Mol. Clu. **30**, 57 (1994).

[6] C. J. Geyer and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).

[7] F. Calvo and P. Labastie, Chem. Phys. Lett. **247**, 395 (1995).

[8] L. Ming, S. Nordholm, and H. W. Schranz, Chem. Phys. Lett. **248**, 228 (1996).

[9] J.-S. Wang, T. K. Tay, and R. H. Swendsen, Phys. Rev. Lett. **82**, 476 (1999).

[10] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[11] A. B. Adib, Phys. Rev. E **71**, 056128 (2005).

[12] F. Heilmann and K. H. Hoffmann, Europhys. Lett. **70**, 155 (2005).

[13] J. Skilling, Bayesian Anal. **1**, 833 (2006).

[14] L. B. Pártay, A. P. Bartók, and G. Csányi, J. Phys. Chem. B **114**, 10502 (2010).

[15] H. Do, J. D. Hirst, and R. J. Wheatley, J. Chem. Phys. **135**, 174105 (2011).

[16] R. A. X. Persson, Phys. Rev. E **86**, 066708 (2012).

[17] S. Davis, J. Phys.: Conf. Ser. **410**, 012161 (2013).

[18] S. Davis, Phys. Rev. E **84**, 050101 (2011).

[19] E. S. Severin, B. C. Freasier, N. D. Hamer, D. L. Jolly, and S. Nordholm, Chem. Phys. Lett. **57**, 117 (1978).

[20] H. W. Schranz, S. Nordholm, and G. Nyman, J. Chem. Phys. **94**, 1487 (1991).

[21] J. R. Ray, Phys. Rev. A **44**, 4061 (1991).

[22] J. R. Ray and H. Zhang, Phys. Rev. E **59**, 4781 (1999).

[23] For the sake of notational simplicity, we consider only the Monte Carlo probability distribution here. The argument is completely analogous in the molecular dynamics case.

[24] This analysis does hence not apply to the question of which sigma function is the most efficient in molecular dynamics.