Improved configuration space sampling: Langevin dynamics with alternative mobility

Chau, C.D.; Sevink, G.J.A.; Fraaije, J.G.E.M.

Note: To cite this publication please use the final published version (if applicable).

Improved configuration space sampling: Langevin dynamics with alternative mobility

C. D. Chau, G. J. A. Sevink, and J. G. E. M. Fraaije

Improved configuration space sampling: Langevin dynamics with alternative mobility

C. D. Chau,[a] G. J. A. Sevink, and J. G. E. M. Fraaije
Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

We present a new and efficient method for determining optimal configurations of a large number (N) of interacting particles. We use a coarse-grained stochastic Langevin equation in the overdamped limit to describe the dynamics of this system and replace the standard mobility by an effective space dependent inverse Hessian correlation matrix. Due to the analogy of the drift term in the Langevin equation and the update scheme in Newton's method, we expect accelerated dynamics or improved convergence in the convex part of the potential energy surface Φ. The stochastic noise term, however, is not only essential for proper thermodynamic sampling but also allows the system to access transition states in the concave parts of Φ. We employ a Broyden–Fletcher–Goldfarb–Shannon method for updating the local mobility matrix. Quantitative analysis for one and two dimensional systems shows that the new method is indeed more efficient than standard methods with constant effective friction. Due to the construction, our effective mobility adapts high values/ low friction in configurations which are less optimal and low values/high friction in configurations that are more optimal. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2943313]

I. INTRODUCTION

Condensed phases, whether liquid, glassy, or crystalline, owe their existence and measurable properties to the interactions between the constituent particles. These interactions are comprised in a potential energy function $\Phi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ that depends on the location $\mathbf{r}_i$ for each of those (N) particles. Material-specific contributions to this potential energy function constitute a multidimensional ($3N$ for structureless particles) potential energy landscape. Various static and dynamic phenomena in condensed phases emerge as manifestations of the complex topography of this hypersurface Φ.[1]

An issue concerns the presence and determination of special points on the Φ hypersurface, in particular, the minima and saddle points. Minima correspond to stable particle configurations, such that any small distortion will result in a restoring force to the undistorted arrangement. The global minimum is related to the state of the system at zero Kelvin, provided that the system is cooled slowly enough to maintain thermal equilibrium. In general, several minima with a substantial variation in depth are arranged in a complex pattern throughout the configuration space; for a single-component system rather general arguments show that the total number of minima scales as $N \exp(\alpha N)$,[2] where $\alpha > 0$ depends on the chemical nature of the system considered. Each minimum is enclosed in its own basin, consisting of all points $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ in the direct vicinity of the minimum where $\Phi(\mathbf{R})$ is monotonic. Saddle points can be found on the boundary between contiguous basins and represent the transition states of the system.

Standard numerical optimization techniques for the de-termination of these special points on the multidimensional hypersurface have several drawbacks: (1) in practice, nonlinear optimization methods [for instance, Monte Carlo (MC) methods] are compute intensive due to rather poor convergence; (2) deterministic algorithms are not devised to sample multiple basins/minima and access them via transition states. Moreover, general numerical methods are not based on fundamental physical laws underlying the time evolution of the system and obscure the relation between the simulation pathway and the dynamical phenomena that one would like to capture. In molecular dynamics (MD), the pathway is prescribed by the classical Newtonian equations of motion, which incorporate forces specified by Φ. Although MD is commended as an exact method and, in general, provides an adequate description of the particle dynamics, the length of the pathway in configuration space is seriously limited by the restriction of the time integration step to very small values, caused by the presence of high frequency modes. For modeling of phenomena on a long time and/or large length scales, as well as the configurational sampling of large dense systems, coarse-grained approaches are a reasonable alternative to MD. Coarse-grained descriptions reduce the degrees of freedom in the system by employing time scale separation and take into account only the stochastic properties of the rapidly varying quantities. Stochastic dynamics techniques (of which Brownian dynamics is the most simple form[3]) use this approach to represent the presence of solvent by a stochastic and a frictional force in the Langevin equation for the solute. Moreover, one can argue that hydrodynamic modes become irrelevant for the dynamics of dense and strongly structured systems[4] and assume that the velocity distribution plays no relevant role. In particular, for dense systems, the rapid flow of momentum and energy due to particle-particle collisions in the Newtonian dynamics is given by the bath in

a)Electronic mail: c.chau@chem.leidenuniv.nl.

the Brownian dynamics. In the equilibrium limit, as a very long time average of the dynamic evolution, the Newtonian and the Brownian equations of motion should give the same result and be equal to those of the equilibrium statistical ensemble. Our aim here is to model the Brownian behavior of N particles interacting via a pairwise potential. This potential could be a Coulomb interaction, a Lennard-Jones-type interaction, or simply a hard sphere repulsion. We employ microscopic equations of motion and include information about the local curvature of the potential energy hypersurface Φ in the Langevin equation describing the Brownian dynamics. The configuration-space sampling path resulting from our method is unconditional and not necessary physically realistic. There are conceptual similarities to some efficient MC techniques.[5,6]

   This article is structured as follows. First we introduce our method and discuss the details and numerical implementation. Following, we compare the performance of this method for a few lower dimensional problems to standard methods. We discuss the properties and results obtained by the new method in details. The focus here is on the efficiency, the sampling path, and the numerically constructed sampling ensemble. The application of this method to systems of higher dimensionality is left for future work.

II. OUR METHOD WITH ALTERNATIVE MOBILITY TENSOR

   As mentioned in the Introduction, we consider the Brownian behavior of N particles, interacting via a pairwise potential, in the high friction limit. The dynamics of this system is described by the general position Langevin equation, written in Ito form as[7]

$$d\mathbf{x} = [-M(\mathbf{x}) \nabla \Phi(\mathbf{x}) + k_B T \nabla \cdot M(\mathbf{x})]dt$$
$$+ \sqrt{2k_B T} L(\mathbf{x}) dW(t). \tag{1}$$

The second term on the right-hand side of Eq. (1) is the compensation for the flux caused by the random force.[7] Here Φ is the potential energy, k_B is the Boltzmann constant, T the temperature, $M(\mathbf{x})$ is the space dependent mobility tensor with

$$M(\mathbf{x}) = L(\mathbf{x})L(\mathbf{x})^T, \tag{2}$$

and $W(t)$ is the multivariate Wiener process with $\langle dW_i(t)dW_j(t)\rangle = \delta_{ij}dt$. In the simulations, we use random numbers with Gaussian distribution for the Wiener process. For simplicity, we set $k_B = 1$ and render Eq. (1) in the dimensionless form, i.e., all variables and functions [e.g., the energy potential $\Phi(\mathbf{x})$] in Eq. (1) will be considered dimensionless. Further, we will write κ instead of $2k_B T = 2T$. The amplitude of the noise term is determined in agreement with the fluctuation-dissipation theorem. As a result, the corresponding Fokker–Planck equation gives rise to proper sampling, according to the Boltzmann distribution, see Appendix.

   In the standard approach, the mobility tensor is chosen as the identity I, giving rise to a scalar prefactor describing the friction due to the implicit solvent. Here we include space dependency via the second order derivative and con-

sider the mobility equal to the inverse Hessian H^{-1}. As a consequence, the deterministic part of Eq. (1) (for $T=0$),

$$d\mathbf{x} = -[H(\mathbf{x})]^{-1} \nabla \Phi(\mathbf{x})dt, \tag{3}$$

is similar to the update scheme in Newton's method,

$$\mathbf{x}_{k+1} - \mathbf{x}_k = \Delta\mathbf{x} = -\alpha H_k^{-1} \nabla \Phi(\mathbf{x}_k), \tag{4}$$

where dt is the (infinitesimal) time interval and α is an appropriate positive step size, often obtained by a line search method. In the following we use t as continuous variable and k as iteration variable. Newton's method is standard in unconstrained nonlinear optimization, i.e., methods that aim at minimizing a certain objective function $\Phi(\mathbf{x})$, $\Phi(\mathbf{x}):\mathbb{R}^N \rightarrow \mathbb{R}$. The methods make use of a *quadratic model* in which $\Phi(\mathbf{x})$ is approximated on the kth iteration by a Taylor series about $\mathbf{x}_k$,

$$\Phi(\mathbf{x_k} + \delta_k) \approx q_k(\delta_k) = \Phi(\mathbf{x}_k) + \delta_k^T \nabla \Phi(\mathbf{x}_k) + \tfrac{1}{2}\delta_k^T H(\mathbf{x}_k)\delta_k. \tag{5}$$

The displacement δ_k on iteration k follows as the minimizer of $q_k(\delta)$. A unique minimizer of $q_k(\delta)$ exists if and only if $H(\mathbf{x}_k)$ is positive definite, and Newton's method is only well defined in this case. Usually, $H(\mathbf{x}_k)$ is positive definite for small δ. It can then be proven that the sequence $\{\mathbf{x}_k\}$ converges and that the order of convergence is of second order (quadratic convergence). When $\mathbf{x}_k$ is remote from the local solution $\mathbf{x}^*$ Newton's method may not converge and may not be defined [when $H(\mathbf{x}_k)$ is not positive definite]. Prototype algorithms such as line search and trust region methods can be employed to avoid this problem.

   Especially for large N or when evaluation of $\Phi(\mathbf{x})$ is expensive, explicit calculation of the Hessian or inverse Hessian is too demanding, and Newton-type methods are a good alternative. These methods are based on approximating the exact Hessian $H(\mathbf{x}_k) = H_k$ in Eq. (5) by $G(\mathbf{x}_k) = G_k$ [giving rise to $B_k = G_k^{-1}$ approximate of H_k^{-1} in Eq. (4)]. The reduction of $\Phi(\mathbf{x})$, the descent property, is guaranteed if G_k (and therefore B_k) is positive definite. The rate of convergence depends on the second derivative information installed in B_k and ranges from linear to quadratic. The simplest positive definite choice is $B_k = G_k = I$, which does not contain any curvature information and gives rise to the well-known steepest descent (SD) method with slow linear convergence. An alternative choice for indefinite H_k^{-1} is $B_k = H_k^{-1} + D_k$, where D_k is a diagonal matrix with proper elements. The efficient quasi-Newton (QN) method builds up second derivative information by estimating the curvature along a sequence of search directions.[8] Each curvature estimate is installed in an approximate inverse Hessian B_k by applying a rank-one or a rank-two update. One of the most successful updates is the Broyden–Fletcher–Goldfarb–Shannon (BFGS) formula, which is a member of the wider Broyden class of rank-two updates. The QN method with the BFGS update formula is also known as the BFGS method.

   Here, we aim at developing a general and efficient scheme for equilibrating systems with many particles (large N), using the stochastic Langevin Eq. (1) in discretized form. In analogy with Newton-type methods, we consider a mobil-
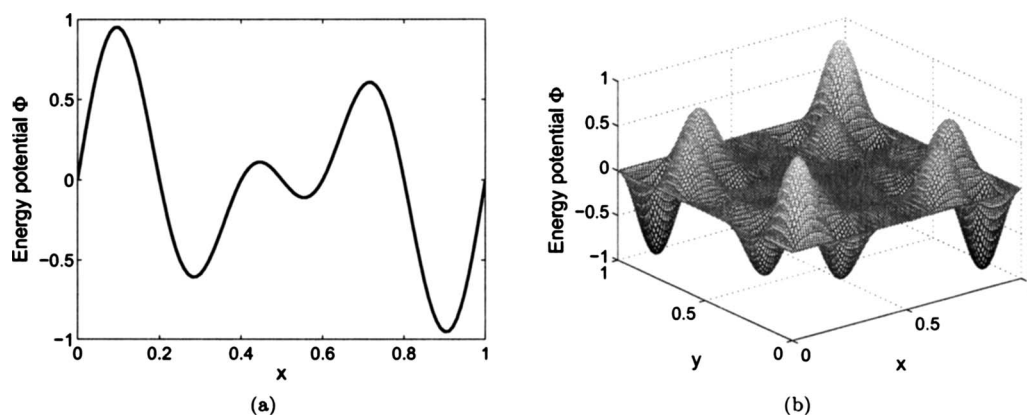
FIG. 1. The considered low dimensional sample cases. (a) 1D energy landscape; (b) two dimensional energy landscape.

ity matrix M_k that is a good approximation of the inverse Hessian H_k^{-1} to avoid direct calculation of this large $3N \times 3N$ matrix. In particular, we use the BFGS method to determine this approximate matrix $M_k = B_k$ (see the Appendix for details about the BFGS update scheme). For $T=0$, our update scheme is equal to that of the BFGS method. The BFGS method is known to yield a positive definite matrix B_k if the Wolfe conditions are satisfied,[9] which requires that the next step is in the descending direction and that the displacement is sufficiently large. Moreover, in that case it shows global convergence for convex functions $\Phi(\mathbf{x})$.[10] In contrast to the standard practice, we deal with potentials $\Phi(\mathbf{x})$ that are convex only in a relatively small subset of configuration space. Our method should be able to hop between different minimal states, that are separated by energy barriers, and not be too sensitive to the starting configuration $\mathbf{x}_0$. A similar problem is present in reaction path sampling: in some regions, especially in the vicinity of saddle points, the Hessian can be negative definite, and BFGS update becomes ill conditioned.[11] Recently, a number of alternative methods that do not necessarily always satisfy the descent condition, the symmetric rank-one formula (SRI), Powell-symmetric-Broyden method, and Bofill's formula, were developed and applied for reaction path sampling. Calculated pathways between reactant and product minima, via optimized transition states, were found to be accurate for a number of problems.[12] Nevertheless, we need to maintain the positive definiteness of the mobility matrix M_k at all times, since the Choleski decomposition in Eq. (2) requires this property.

As a consequence, the BFGS method needs to be adapted to handle the cases where the Wolfe conditions are violated. For computational reasons, we want to avoid any type of preconditioning of the matrix B_k which requires the explicit calculation of eigenvalues. A possible general remedy to avoid ill conditioning is to switch back to the SD method, i.e., restart the BFGS update from the initial guess $B_k = I$. A disadvantage of this restart is that it would disregard all curvature information that has been built up along the pathway in configuration space. Another solution is updating the Choleski factor L rather than $B = LL^T$ itself, which ensures a positive definite B.[13] However, no practical advantage can be expected for ill-conditioned problems,[14] apart from an assurance that B_{k+1} remains positive definite in the presence of round-off errors.

The key difference between Eq. (1) and standard BFGS methods is the stochastic term that is added at every step to ensure correct Boltzmann weights $[\sim\exp(-\Phi/k_BT)]$, which can be seen as an information-based correction to the BFGS update with varying amplitude and direction. Hence, our update is a combination of a coherent and stochastic contribution. The coherent contribution is always in the descending direction, by virtue of BFGS. The stochastic contribution allows for sampling in the vicinity of the coherent update, the extent of which is determined by thermodynamic properties, and on average gives rise to the important ascending updates and sampling of transition states. The efficiency of the coherent part, the BFGS method itself, depends on the appropriateness of the curvature estimate. We choose to conserve the quadratic information that is accumulated in the sequence of expanding subspaces in the convex part of Φ (where B_k is positive definite) by the BFGS method. We use a small (constant) step size and do not update B_k if the Wolfe conditions are violated, i.e., if the approximate curvature $\mathbf{y}_k^T\mathbf{s}_k \leq 0$, we take $B_{k+1} = B_k$. One can show that this choice does not affect the theoretical equilibrium distribution of states, which is the Boltzmann distribution. Since the pathway between different minima should pass through regions with relatively small Boltzmann weights, we expect fewer sampling in the part of the potential energy surface where B is not updated, and hence an increased efficiency of the update scheme in this region. In terms of physical properties, reduced sampling can be associated with accelerated kinetics and lower friction. We will illustrate the peculiarities of this choice by considering the one dimensional (1D) and two dimensional problems. Since this scheme ensures the positive definiteness of $M_k = B_k$ at all times, the decomposition in Eq. (2) can be obtained by the standard Choleski decomposition.

The calculation of the divergence of the mobility tensor in the second term of the right-hand side of Eq. (1) is compute expensive. To avoid direct calculation of the divergence term Hütter and Öttinger[15] proposed the following version of the stochastic differential equation (SDE):

$$d\mathbf{x} = [-M(\mathbf{x})\nabla\Phi(\mathbf{x})]dt + \tfrac{1}{2}[M(\mathbf{x}+d\mathbf{x})M(\mathbf{x})^{-1} + I]\sqrt{2k_BT}L(\mathbf{x})dW(t). \tag{6}$$

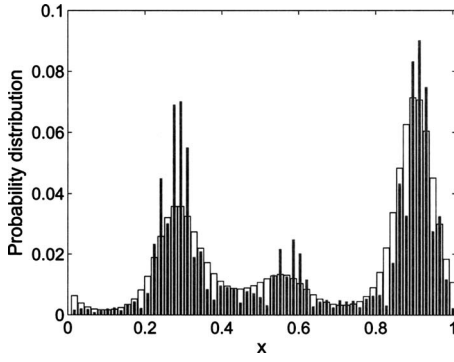Equation (6) suggests the use of the predictor-corrector

FIG. 2. Comparison of numerically calculated (full bars) and analytical (open bars) probability distributions. The numerical results are obtained using our sampling method with $\kappa=1$ and $\Delta t=0.01$.

method in the numerical evaluation. The corresponding numerical scheme becomes

$$\Delta \mathbf{x} = -\frac{1}{2}[M(\mathbf{x}+\Delta \mathbf{x}^p)\nabla\Phi(\mathbf{x}+\Delta \mathbf{x}^p)+M(\mathbf{x})\nabla\Phi(\mathbf{x})]\Delta t$$
$$+\frac{1}{2}[M(\mathbf{x}+\Delta \mathbf{x}^p)M^{-1}(\mathbf{x})+I]\sqrt{2k_BT}L(\mathbf{x})\Delta W_t, \qquad (7)$$

where ΔW_t is the Wiener increment with $\langle \Delta W_t\rangle=0$ and $\langle \Delta W_{t_i}\Delta W_{t_j}\rangle=\delta_{ij}I\Delta t$ and $\Delta \mathbf{x}^p$ is the predictor step,

$$\Delta \mathbf{x}^p = -M(\mathbf{x})\nabla\Phi(\mathbf{x})\Delta t+\sqrt{2k_BT}L(\mathbf{x})\Delta W_t. \qquad (8)$$

The updated mobility tensor obtained by the BFGS method, $M(\mathbf{x}_k)=B(\mathbf{x}_k)$, guarantees the existence of the inverse in Eq. (7). The integration scheme is weakly convergent to first order in the time step Δt. This method replaces the calculation of the divergence term and is clearly favored to the direct calculation [except in cases where $\nabla\cdot M(\mathbf{x})$ is given in closed form]. Details about the BFGS update, space, and time dependent, can be found in the Appendix.

III. RESULTS AND DISCUSSION

In order to compare our method to standard approaches for which the performance is known analytically, we restrict the application of the new method to 1D and two dimensional systems. In particular, these energy potentials are not related to any real physical system.
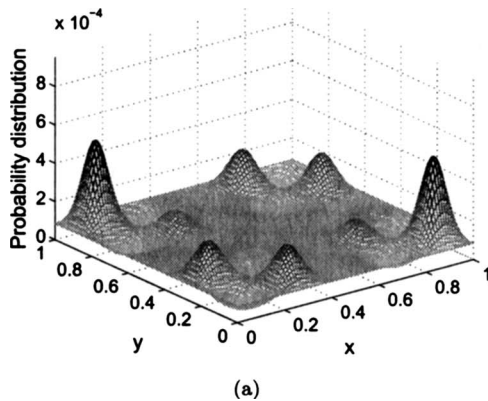
A. Simulated sampling distributions

First, we verify the thermodynamic accuracy, i.e., whether our method indeed samples according to the Boltzmann distribution $\mathcal{N}\exp(-\Phi/k_BT)$. The considered energy landscapes, containing multiple minima, are shown in Figs. 1(a) and 1(b) and are periodic in space (we use periodic boundary conditions) in order to obtain a more differentiated sampling. The simulated sampling distributions were calculated from the simulation results by a binning routine, with small bin width h_b or area h_b^2. To obtain Figs. 2 and 3(b), the number of samples obtained in each bin were divided by the total number of samples $K=10^6$.

In Fig. 2, the simulated distribution (for $\Delta t=0.01$ and $\kappa=1$) is compared to the expected distribution function. The expected distribution is the stationary solution of the Fokker–Planck equation. There is no driving force in equilibrium, hence there is zero flux. This is equal to the Boltzmann distribution for our choice of the drift and noise term.[16]

It can be observed that the simulated distribution matches the theoretical distribution reasonably well, qualitatively but also quantitatively. The oversampling of the maxima, and the global maximum of the distribution, in particular, is due to finite sampling and the peculiarities in the update scheme for B_k. On average, we also observe undersampling of states with lower Boltzmann weights. At a later stage, we discuss how our choice for the mobility $M_k=B_k$ contributes to these effects. For the two dimensional case, the theoretical distribution [Fig. 3(a)] and the simulated distribution [Fig. 3(b)] match rather well, but show even somewhat more pronounced over- and undersampling than in 1D. Overall, we conclude that the sampling is accurate in both cases.

For larger problems, i.e., large N, the efficiency of our new method is of importance. Here, we compare the mean first passage times (MFPTs) obtained by our mobility to that of Eq. (1) for M_k is constant. The latter choice gives rise to the standard Langevin equation for particle motion in the overdamped situation, with a constant friction coefficient. Instead of the 1D potential in Fig. 1(a) with multiple minima, we use a double-well potential. The advantage of this double-well potential is that an analytical expression for the MFPT is known.[17] We note that the 1D case is special, as any
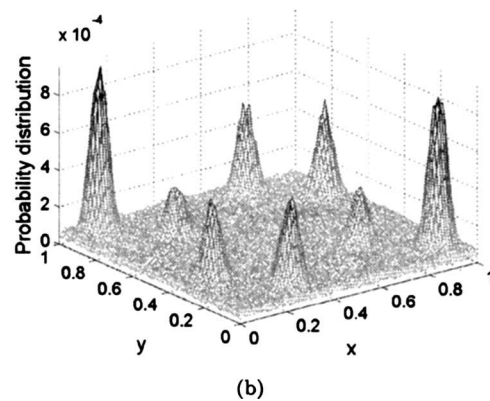

(a)

(b)

FIG. 3. Comparison between analytical distribution function and the calculated distribution function with $\kappa=1$ and $\Delta t=0.01$. (a) Analytical probability distribution; (b) calculated probability distribution.
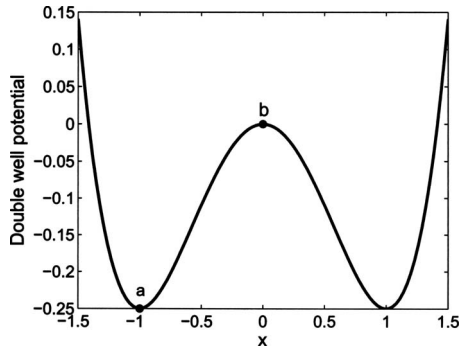
FIG. 4. The double-well potential considered; $\Phi(x)=-\frac{1}{2}x^2+\frac{1}{4}x^4$.
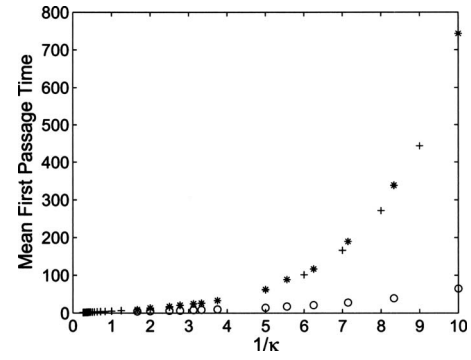

FIG. 5. MFPTs of the double-well potential, calculated for different noise amplitudes. The plusses (+) are the MFPTs calculated by numerically intergrating the analytical expresson (10), with constant mobility $M=H(x_{\min})^{-1}$. The MFPTs found by our scheme (1) using a constant $H(x_{\min})^{-1}$ (*) and space dependent B_k (○) are shown as well. The step size $\Delta t=0.01$.

sampling path between different (quasi) equilibrium states automatically requires crossing the potential energy barrier. For higher-dimensional energy surfaces, high energy barriers can be avoided in favor of lower energy transition states or saddle points, and the performance of the method crucially depends on the ability to find the optimal route.

B. MFPTs

We consider the mean escape time from a 1D well, see Fig. 4. By this we mean the first passage time from a to x, where $x \to b$. The MFPT $\langle t \rangle$, defined as the average time needed from a well to the top of the hill is a quantitative measure for the average transition time between the two different (quasi) equilibrium states. Numerically, we performed a large number of simulations (1000), all starting in point a in Fig. 4 using different noise seeds. The simulations were finalized upon passing b for the first time, i.e., for $k=\tilde{k}$ such that $x_k > b$. The simulation times $\tilde{k} \cdot \Delta k$ (in this article, k and t can be interchanged and both represent time) were later averaged to obtain the simulated $\langle t \rangle$.

As mentioned, the simulated MFPT can be directly compared to the MFPT for the standard Langevin equation, Eq. (1) with $M(\mathbf{x})=M$ is constant. We make our comparison scale invariant by choosing the constant mobility $M=H^{-1}(x_{\min})$, where $x_{\min}$ is the location of the starting point. The corresponding dimensionless position Langevin equation is

$$\dot{x} = -M \nabla \Phi(x) + \zeta(t), \tag{9}$$

and the analytical representation for the MFPT is given by the integral formula

$$\langle t \rangle = \frac{2}{\xi} \int_a^b dy \, \exp\left(\frac{2}{\kappa}\Phi(y)\right) \int_{-\infty}^y dz \, \exp\left(-\frac{2}{\kappa}\Phi(z)\right), \tag{10}$$

where ξ is the noise intensity of ζ. Again, a is the initial condition and b is the final state.[18] For the double-well potential shown in Fig. 4, defined by $\Phi(x)=-\frac{1}{2}x^2+\frac{1}{4}x^4$, we can numerically integrate Eq. (10) and derive theoretical values of the MFPT for varying κ. Alternatively, one could also directly use the well-known Kramers equation[18] to calculate theoretical MFPTs, but this explicit equation is only strictly valid in the limit of high friction. The theoretical values (10) and the simulated MFPTs, with $M(x_k)=B(x_k)$, are compared in Fig. 5 for varying κ. Additionally we have tested our

implementation by evaluating our scheme for constant mobility $M=M(x_{\min})$. We observe that the simulated MFPTs for this mobility are in very good agreement with the numerically integrated theoretical predictions of Eq. (10).

An important observation is that the simulated MFPTs for $M(x_k)=B(x_k)$ are significantly smaller than the MFPTs for the constant mobility ($M=M(x_{\min})$) for the whole κ range. The efficiency increases significantly for small κ (or large κ^{-1}, see Fig. 5). Since the contribution of the stochastic term is proportional to κ, it is a clear indication that the incorporation of our new mobility, $M_k=B_k \approx H_k^{-1}$, in the deterministic part of Eq. (1) allows for a much more efficient crossing of energy barriers. This finding is important, since one could think on beforehand that the effect of different mobilities is rather small in 1D, where M is a scalar variable. In particular, using $M=H^{-1}(x_{\min})$ or $M=B(x_k)$ in Eq. (1) gives rise to the same update for the deterministic part, apart from a space dependent prefactor. For higher-dimensional energy surfaces, the search direction itself will play an important role. We have earlier noticed that, apart from the prefactor, the deterministic part of Eq. (1) for $M=M(x_{\min})$ is analogous to the update in the standard SD method, independent of the dimensionality of the system. From transition state theory for chemical reactions it is well known[19] that the QN method is much better suited for locating saddle points or transition states than the SD method. Hence, we expect that especially for high-dimensional systems the choice $M_k=B_k$, related to the QN method, will lead to an improved sampling compared to $M_k=M(x_{\min})$, which is, related to the SD method. In conclusion, we expect a further reduction of the transition times (or MFPTs) for the new method and large N.

We have omitted error bars in Fig. 5 and calculated the standard deviation of the MFPTs of the simulations. We found that the ratio of standard deviation to the mean is around 1 for each of the mobility tensors. This indicates that the MFPTs are exponentially distributed.

C. The update scheme for mobility

Finally, we turn to our update scheme for $M(\mathbf{x}_k)=B_k$ and consider how our procedure to maintain positive definiteness affects the sampling and efficiency. In particular, we only use
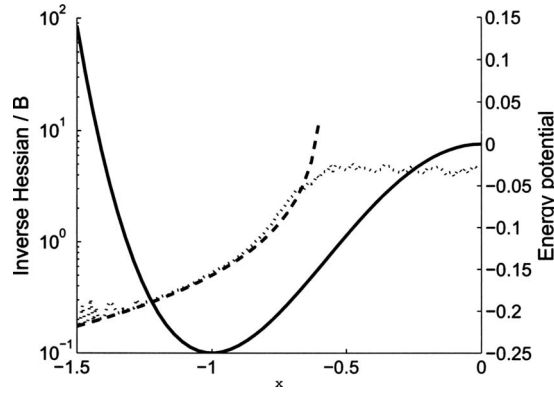
FIG. 6. The double-well potential (solid line), the analytical inverse Hessian (dashed line), and calculated average mobility (dotted line) for $\kappa=0.2$ and $\Delta t=0.01$.
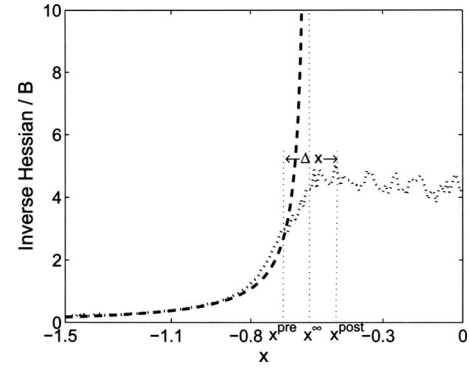

FIG. 7. Analytical inverse Hessian (dashed line) and the calculated average mobility (dotted line) with $\kappa=0.2$ and $\Delta t=0.01$. In addition we include a schematic view of x^{pre}, x^{post}, and the asymptote at $x=x^{\infty}$, used to estimate the mobility when the Hessian H is negative. The actual position of x^{pre} and x^{post} depends on the noise amplitude.

the BFGS update for B_{k+1} if the Wolfe conditions are satisfied, otherwise $B_{k+1}=B_k$. The particular potential surface, the 1D double-well potential $\Phi(x)$, $H^{-1}(x)$ (only for positive values) and the average B as a function of Cartesian coordinate x are combined in Fig. 6. Since we again consider the 1D potential, all functions have scalar values, and we use the values from the many simulations that were carried out for the calculation of the MFPT. Here and further on, all averages are calculated by a binning routine. For each bin, the average is calculated over all $B(x)$ within the bin.

In the convex region, where H (and thus H^{-1}) is positive, our update scheme clearly yields a good approximate for H^{-1}, as expected by virtue of the BFGS method. Only close to the initial value a the average B is less accurate. This is due to the initial guess $B_0=I$ and buildup of curvature information. In the concave region, $H^{-1}(x)$ is negative, and hence there is no update in this region, i.e., $M(x_{k+1})=B_{k+1}=B_k$. The mobility remains unchanged as long as x lies in the concave region. Only when the update crosses over to a convex region again, the mobility is updated. We refer to the boundary limit the region, where the BFGS update is accepted or rejected as the crossover point x^c. By virtue of our method the location of this crossover point depends on the noise amplitude and the temperature T. We determine the average $B(x^c)=B^c$ value both numerically and by a simple theoretical estimate. Consider $x_k=x^{\mathrm{pre}}$ in the convex region and $x_{k+1}=x^{\mathrm{post}}$ in the concave region, and x^{∞} the inflection point, i.e., the location where $H=0$ (see Fig. 7). To determine x^{pre} we consider the following equation:

$$(\Delta x)^2 = (x^{\mathrm{pre}} - x^{\mathrm{post}})^2 = 4(x^{\mathrm{pre}} - x^{\infty})^2, \tag{11}$$

where the latter equation is obtained by using $x^{\mathrm{pre}}=x^{\infty}-\frac{1}{2}\Delta x$.

Equation (11) can be rewritten as

$$4(x^{\mathrm{pre}} - x^{\infty})^2 = \kappa H^{-1}(x^{\mathrm{pre}})\Delta t, \tag{12}$$

which becomes a simple fourth power equation, since $H(x)$ is a quadratic function in this special case. Because $H(x)$ and x^{∞} are known, Eq. (12) can easily be solved analytically, which gives the theoretical value of x^{pre}. We compare the theoretical estimate $H^{-1}(x^{\mathrm{pre}})$ and the numerically obtained average B^c in Fig. 8. Since we consider the average B^c, we loose infor-

mation about the individual simulation pathways. In particular, some pathways cross the inflection point multiple times before reaching the top b, and as a result the location of the crossover point, and hence the value of B^c, varies in time. We try to include these details in the averaging procedure by distinguishing between *all* pathways and pathways that cross the inflection point only *once* (to be called first-time passing in the remainder). The dotted line in Fig. 8 shows the average B^c after first time passing only, and the dashed line shows the average B^c using all simulated pathways. As expected, the value of B^c after first-time crossing is smaller than the value of B^c using all pathways, and for all graphs the value of B^c increases with decreasing noise amplitude κ. This behavior can be explained. First of all, one should realize that the *average* distance h between the crossover and the inflection point is determined by the average stochastic contribution to the search direction, and therefore by the noise amplitude in the stochastic term in Eq. (1). The distance h will decrease for decreasing κ (or increasing κ^{-1}), and since H^{-1} is a monotonically increasing function in this interval, the constant value of B^c will increase (see the relevant figures). Pathways that cross back from the concave to the convex region are more likely to also sample the interval between the first-time crossover point and inflection point, and
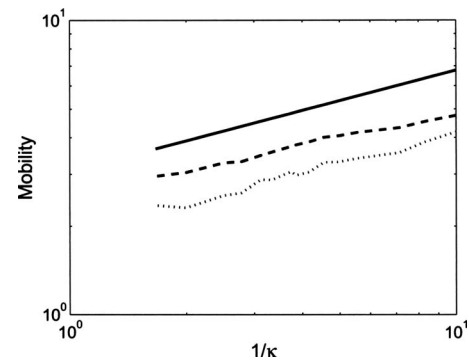

FIG. 8. Comparison of the analytically estimated mobility $H^{-1}(x^{\mathrm{pre}})$ (solid line) with from simulation obtained B^c for the mobility in the concave region of the double well; the average B^c after passing the inflection point (dashed line) and the average B^c after the first time passing the inflection point (dotted line), with $\Delta t=0.01$.
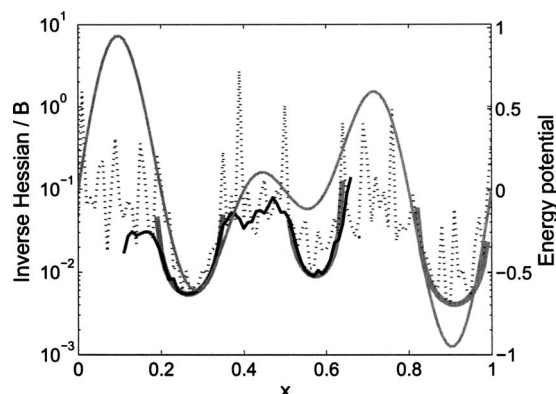
FIG. 9. The multistate energy potential (solid gray line), the analytical inverse Hessian (thick solid gray line), with the average mobility B_k used in simulations. The dots represent B_k with $\kappa=1$ (large noise) and solid black line B_k with $\kappa=0.2$ (lower noise).

subsequently the positive definite B^c is higher on average, following the same arguments. We observe that all graphs in Fig. 8 follow the same trend and are therefore equal, except for a scaling factor.

We see that our choice for the mobility update in the concave region results in a constant mobility B^c, with an average value that increases with decreasing κ. This is a convenient property, in general, since high mobility or equivalent low friction gives rise to an acceleration of the scheme and smaller residence times in this part of the energy potential landscape. Although the potential energy surface is different, this finding explains the oversampling of the maxima and undersampling of lower regions after a finite number of simulations steps, when compared to the theoretical Boltzmann distribution $\mathcal{N}\exp(-\Phi/k_BT)$ in Fig. 2 (for $\kappa=1$). For a decreased temperature or noise amplitude, the regions associated with low Boltzmann weights become even more pronounced, and standard MC methods experience a critical slowing down and very long transition times. Multicanonical ensemble methods have been developed to enhance the behavior in these regions, by manipulating the sampling distribution.[20] In our method, however, the distribution itself is left unchanged and does not have to be calculated explicitly. With decreasing temperature, the approach of the inflection points is slowed down (see Fig. 5), but crossing this point is followed by "tunneling" through these regions due to particular features of the update scheme.

To be complete, we have also considered the properties of our scheme (1) for the potential energy surface with multiple extrema [see Fig. 1(a)]. In Fig. 9, we compare the analytic $H^{-1}(x)$ for this potential energy surface and the average B for two different temperatures: $\kappa=0.2$ and 1. We observe that the average B associated with the lower temperature $\kappa=0.2$ is a very good approximate of the (analytic) inverse Hessian in the convex region and almost constant outside this region. This finding is very well in accordance with the double-well case. We also observe that the large energy barriers in the potential energy surface cannot be overcome for this relatively small κ. For higher κ, the scheme samples the whole interval $x \in [0,1]$, which can also be observed from the simulated distribution in Fig. 2. The average mobility B,

however, appears to be noisy. At first this may seem surprising, as the BFGS method is known to approximate the actual $H^{-1}(x)$ in one step for a 1D function $\Phi(x)$, if $\Phi(x)$ is a (nearly) quadratic function. Moreover, the (almost) constant value B^c in the concave part of the landscape is higher for $\kappa=1$ than for $\kappa=0.2$, which seems in contradiction with the conclusion drawn for the double-well potential. Looking at $B=B(x)$ in somewhat more detail, we may consider B a superposition of low and high frequency contributions. The shape of this low frequency contribution is very similar to that of the inverse Hessian in the convex region. In the concave region, it is a constant. Hence, we conclude that this contribution can be associated with the part of the pathway where the matrix B is frequently updated, and the BFGS scheme provides a good approximate B of the actual $H^{-1}(x)$. The high frequency contribution gives rise to a shift of the average B toward higher values, when compared to $H^{-1}(x)$. A similar shift in the direct vicinity of x^{pre} can be observed for the double-well potential in Fig. 7. Although these simulations were stopped when the pathway reached point b and here multiple extrema are sampled instead, the origin of this overestimation is the same. Due to the random displacements, the update can hop between convex and concave regions in the vicinity of the crossover point. Since B in the concave part is taken as the value of the *last* positive definite update B^{last}, an update that hops back to the convex region at step k uses B^{last} to update B_{k+1}. Consequently, the calculation of the average $B(x)$ in the vicinity of the crossover point includes a number of values B^{last} that may be inaccurate approximates of the inverse Hessian $H^{-1}(x)$. Moreover, the values associated with these last updates are likely to be higher (on average) than the actual inverse Hessians $H^{-1}(x)$ (see also the previous discussion for the double-well potential using only first-time passages and all pathways). Since the range of the random displacements scales with the noise amplitude, the affected range of $B(x)$ is also larger for larger κ. This explains the superposition of two contributions with a different character. This overestimation is observed in the whole range of $\Phi(x)$ and is due to the fact that the wells in Fig. 1(a) are not well separated. We conclude that only a part of the sampling pathway is affected and that the mobility in this part is artificially high. However, from the comparison in Fig. 2 we see that the Boltzmann distribution is well reproduced by the numerical sampling, so the effect is rather minimal.

For two dimensional potential energy surfaces, the simulated distribution in Fig. 3(b) has features that are similar to the ones observed in 1D (see the section about the simulated distribution of states), and we conclude that our choice of M_k is apparently also efficient for sampling transition pathways for $N=2$. We note that this is an important finding, as for $N>1$ pathways that cross energy barriers (maxima of the function Φ) can be avoided in favor of saddle points. The situation for higher dimensional systems ($N>2$) will be considered in future publications.

IV. CONCLUSION

We already discussed in detail the similarities between our displacement and the QN displacement. The drift term in Eq. (1) contains the approximate correlation matrix B multiplied with the minus gradient of the energy potential $\nabla\Phi$ and dt. This expression is exactly the same as the displacement in the QN method. By only taking into account $-B\nabla\Phi dt$ and the BFGS approach for B guarantees the descending property of the method. Mathematically, the drift term guarantees quadratic convergence to the local minimum. Physically, one immediately notices that $-\nabla\Phi$ is the force, since Φ is an energy potential. If B is the identity matrix I, the displacement contribution of the drift term is exactly biased in the direction of the force (corresponds to the mathematical SD method). For a general diagonal matrix, the biased direction is the force which magnitude is determined by the values of the diagonal. Using the approximate of the covariance matrix, the direction is determined by the correlation matrix B.

The stochastic term $\sqrt{2k_BTL(\mathbf{x})}dW_t$ contains the temperature and the decomposition of the covariance matrix. The noise term has been obtained mathematically by satisfying the fluctuation-dissipation theorem. The temperature in the noise term determines the shape of the distribution. If the temperature is low, the distribution function will have sharply peaked maxima. The MFPTs have already shown that it can take a very long time to cross an energy barrier if the temperature is low. In the limiting case energy barriers will never be crossed, for $T=0$ the method becomes the standard QN method and the method will be trapped in the nearest (local) minimum. If the temperature is high, the noise term will dominate. The distribution function will be flattened, which means that each state becomes almost equally likely to be visited. Due to the high temperature the system is able to cross energy barriers and hop from one minimum to another. Too high temperature makes it difficult to identify the minima since there is no explicitly a favored state. From the existing analytical expression for the MFPTs we already know that higher noise intensities, equivalent to the temperature in our case, will cause shorter MFPTs. We performed simulations to compare the numerical evaluation of the analytical expression for the MFPTs. Choosing the mobility as a constant gives us the same results as the analytical expression. Using our correlation matrix as mobility, the MFPTs are certainly shorter compared to the MFPTs obtained from a constant mobility $M=H^{-1}(x_{\min})$ and thus implies a better performance of our method, based on the the cycles needed to find the MFPTs. To be more accurate, we need to consider the arithmetical operations needed in each cycle. It is obvious that our method needs more arithmetical operations in each cycle because of the update for the approximate inverse Hessian $B_k(o(n^2))$ and the Cholesky decomposition $(o(n^3))$. In our 1D case, the MFPTs for $M=B_k$ are roughly one order lower than the MFPTs for $M=H^{-1}(x_{\min})$ for small κ. Obviously the arithmetical operations in our sample case are limited since $n=1$. For higher dimensions $n\gg1$, the arithmetical operations may start to dominate if the performance of our method over the standard method does not improve as well. Existing knowledge of the efficiency of the QN and SD methods indicates that this may actually be the case. However, future study will concentrate on the use of limited memory methods for the iterative updates.

In short, our method combines the quadratic convergence of the drift part with the properties of statistical thermodynamics in the noise part. The update of the mobility tensor in the 1D double-well case shows that the mobility is large in concave regions of the energy potential. This corresponds with faster crossing over energy barriers, i.e., less friction. In the convex regions, the mobility is lower, corresponding to high friction. Hence, due to the constructed mobility the method over sample regions which are more likely and under samples regions which are less likely. The temperature in our method has been kept constant during each of the simulations. One can imagine that the method can be improved by using the temperature as the tuning parameter of the system. For instance, the temperature may be increased if the sampling path is pinned to a certain minimum for a long time. An increased temperature can help the method to cross over a certain energy barrier. Adaptive temperatures are a standard procedure in "simulated annealing."[21]

Additionally, improvement of the numerical performance may come along by changing the time step Δt, that was considered constant throughout this study. This directly corresponds to changing the step size in a QN method. A very common method is using line search to determine the step size. Other improvements, with respect to the BFGS method used here, will be considered for larger dimensional systems, and are out of scope of the present study.

APPENDIX: DERIVATION OF THE GENERAL LANGEVIN EQUATION AND NUMERICAL IMPLEMENTATION

According to Gardiner[22] the many variable version of the Fokker–Planck equation (FPE), which describes the time evolution of the probability density function of a stochast $\mathbf{x}$, is given as

$$\frac{\partial p(\mathbf{x},t|\mathbf{x}_0,t_0)}{\partial t} = -\sum_{i=1}^{n} \partial_i(p(\mathbf{x},t|\mathbf{x}_0,t_0)a_i(\mathbf{x}))$$
$$+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \partial_i\partial_j(p(\mathbf{x},t|\mathbf{x}_0,t_0)D_{ij}(\mathbf{x})), \quad \text{(A1)}$$

which is related to the SDE,

$$d\mathbf{x} = \mathbf{a}(\mathbf{x})dt + B(\mathbf{x})d\mathbf{W} \quad \text{(A2)}$$

by $D(\mathbf{x})=B(\mathbf{x})B(\mathbf{x})^T$. The drift vector $\mathbf{a}(\mathbf{x})$ and noise matrix $B(\mathbf{x})$ are obtained by requiring the stationary solution of the FPE $p_s(\mathbf{x})$ to be the Boltzmann distribution $p_s(\mathbf{x})=N\exp[-\beta\phi(\mathbf{x})]$. Setting $\partial p(\mathbf{x},t|\mathbf{x}_0,t_0)/\partial t=0$ and substitute $p_s(\mathbf{x})$ for $p(\mathbf{x},t|\mathbf{x}_0,t_0)$ in the FPE give

$$p_s(\mathbf{x})a_i(\mathbf{x}) = \frac{1}{2}\sum_{j=1}^{n} \partial_j(p_s(\mathbf{x})D_{ij}(\mathbf{x})) \tag{A3}$$

$$= \frac{1}{2}\sum_{j=1}^{n} (D_{ij}(\mathbf{x})\partial_j p_s(\mathbf{x}) + p_s(\mathbf{x})\partial_j D_{ij}(\mathbf{x})) \tag{A4}$$

$$= \frac{1}{2}\sum_{j=1}^{n} (-\beta D_{ij}(\mathbf{x})\partial_j \phi(\mathbf{x}) p_s(\mathbf{x}) + p_s(\mathbf{x})\partial_j D_{ij}(\mathbf{x})), \tag{A5}$$

$$\rightarrow$$

$$a_i(\mathbf{x}) = \frac{1}{2}\sum_{j=1}^{n} (-\beta D_{ij}(\mathbf{x})\partial_j \phi(\mathbf{x}) + \partial_j D_{ij}(\mathbf{x})). \tag{A6}$$

This leads to the following SDE:

$$d\mathbf{x} = \frac{1}{2}[-\beta D(\mathbf{x})\nabla\phi(\mathbf{x}) + \nabla D(\mathbf{x})]dt + B(\mathbf{x})dW(t), \tag{A7}$$

where $B(\mathbf{x})B^T(\mathbf{x}) = D(\mathbf{x})$ and $\beta^{-1} = k_BT$. After defining $D(\mathbf{x}) = 2k_BTM(\mathbf{x}) = 2k_BTL(\mathbf{x})L^T(\mathbf{x})$ one obtains

$$d\mathbf{x} = [-M(\mathbf{x})\nabla\phi(\mathbf{x}) + k_BT\nabla\cdot M(\mathbf{x})]dt$$
$$+ \sqrt{2k_BT}L(\mathbf{x})dW(t), \tag{A8}$$

where the noise term satisfies the fluctuation-dissipation theorem. Equation (A8) is equivalent to the SDE proposed by Hütter and Öttinger,[15]

$$d\mathbf{x} = [-M(\mathbf{x})\nabla\Phi(\mathbf{x})]dt + \frac{1}{2}[M(\mathbf{x}+d\mathbf{x})M(\mathbf{x})^{-1}$$
$$+ I]\sqrt{2k_BT}L(\mathbf{x})dW(t). \tag{A9}$$

This can be easily proven by expanding $M(\mathbf{x}+d\mathbf{x})$ around $\mathbf{x}$ and obeying the rules $dWdt=0$ and $dWdW=dt$.

The discretized form of the SDE proposed by Hütter and Öttinger is given below. The update for $\mathbf{x}_k$ at simulation step k is given as

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{2}[M(\mathbf{x}_k + \Delta\mathbf{x}_k^p)\nabla\Phi(\mathbf{x}_k + \Delta\mathbf{x}_k^p)$$
$$+ M(\mathbf{x}_k)\nabla\Phi(\mathbf{x}_k)]\Delta t + \frac{1}{2}[M(\mathbf{x}_k + \Delta\mathbf{x}_k^p)M^{-1}(\mathbf{x}_k)$$
$$+ I]\sqrt{2k_BT}L(\mathbf{x}_k)\Delta W_t, \tag{A10}$$

with the corresponding predictor step

$$\Delta\mathbf{x}_k^p = -M(\mathbf{x}_k)\nabla\Phi(\mathbf{x}_k)\Delta t + \sqrt{2k_BT}L(\mathbf{x}_k)\Delta W_t. \tag{A11}$$

The approximate inverse Hessian $B(\mathbf{x}_k) = B_k$, which is taken as the mobility tensor $M(\mathbf{x}_k)$ is calculated using the BFGS update,

$$B_{k+1} = B_k - \frac{B_k\mathbf{y}_k\mathbf{y}_k^T B_k}{\mathbf{y}_k^T B_k\mathbf{y}_k} + \frac{\mathbf{s}_k\mathbf{s}_k^T}{\mathbf{y}_k^T\mathbf{s}_k}, \tag{A12}$$

where

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$$

and

$$\mathbf{y}_k = \nabla\Phi(\mathbf{x}_{k+1}) - \nabla\Phi(\mathbf{x}_k). \tag{A13}$$

[1] F. H. Stillinger, Science **267**, 1935 (1996).
[2] F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).
[3] W. F. van Gunsteren and H. J. C. Berendsen, Mol. Phys. **45**, 637 (1982).
[4] I. M. de Schepper and E. G. D. Cohen, J. Stat. Phys. **27**, 223 (1982).
[5] E. Leontidis and U. W. Suter, Mol. Phys. **83**, 489 (1994).
[6] S. Goldman, J. Comput. Phys. **62**, 441 (1986).
[7] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford Science, Oxford, 1986).
[8] L. Kaufman, SIAM J. Optim. **10**, 56 (1999).
[9] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimisation and Nonlinear Equations*, Siam Classics (SIAM, Philadelphia, PA, 1996).
[10] M. J. D. Powell, *Some Global Convergence Properties of a Variable-Metric Algorithm for Minimization without Exact Line Searches, Nonlinear Programming, SIAM-AMS Proceedings*, edited by R. W. Cottle and C. E. Lemke (SIAM, Philadelphia, 1976), vol. IX.
[11] J. M. Anglada and J. M. Bofill, J. Comput. Chem. **19**, 349 (1998).
[12] H. P. Hratchian and H. B. Schlegel, J. Chem. Theory Comput. **1**, 61 (2005).
[13] P. E. Gill and W. Murray, J. Inst. Math. Appl. **9**, 91 (1972).
[14] D. Goldfarb, Math. Comput. **30**, 796 (1976).
[15] M. Hütter and H. C. Öttinger, J. Chem. Soc., Faraday Trans. **94**, 1403 (1998).
[16] H. D. Vollmer and H. Risken, Z. Phys. B **34**, 313 (1979).
[17] C. W. Gardiner, J. Stat. Phys. **30**, 157 (1983).
[18] H. A. Kramers, Physica (Amsterdam) **7**, 284 (1940).
[19] C. J. Cerjan and W. H. Miller, J. Chem. Phys. **75**, 2800 (1981).
[20] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[21] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines* (Wiley, New York, NY, 1989).
[22] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1983).