List of symbols

$A(\xi)$	Standard free energy
$A^G(\xi)$	Geometric free energy
ABF	Adaptive-biasing Force method
CV	Collective Variable
CZAR	Corrected z-averaged restraint estimator
eABF	extended Adaptive-biasing Force method
FES	Free energy surface
H	Hamiltonian of a particle system
k	harmonic force constant
MC	Monte-Carlo simulation
MD	Molecular dynamics simulation
Mtd	Metadynamics
$P(\xi)$	Probability density under assumption of ergodicity
R	Ramp function for ABF methods
TI	Thermodynamic Integration
$U(\mathbf{x})$	Potential energy of a particle system
US	Umbrella sampling
v	inverse gradient
W	Gaussian height for Mtd and WTM
WHAM	Weighted Histogram Analysis Method
WTM	Well-Tempered Metadynamics
x	Cartesian coordinates of a particle system
Z	canonical partition function
σ	Variance of a Gaussian kernel
λ	Extended variable
τ_G	Update interval for Mtd potential
$\xi(\mathbf{x})$	Collective Variable
$\rho(\xi)$	Probability density

Contents

1 Theoretical Background

1.1 Statistical thermodynamics and Free Energy

The formulation of the problem of free energy estimation starts with the Hamiltonian

$$\mathbf{H}(\mathbf{x}, \mathbf{p}) = \sum_{i=0}^{N} \frac{p_i^2}{2m_i} + U(\mathbf{x}) \tag{1.1}$$

of an arbitrary chemical system, where $(\mathbf{x}, \mathbf{p})$ denotes a point in the phase space of the N-particle system of interest with atomic coordinates $\mathbf{x} = (x_1, ..., x_N)$, momenta $\mathbf{p} = (p_1, ..., p_N)$ and masses $(m_1, ..., m_N)$. $U(\mathbf{x})$ is the potential energy given by any quantum chemical method or force field. If the canonical (N,V,T)-ensemble of such a system is sampled, for example by means of Langevin dynamics or Monte Carlo simulations, the probability distribution $\rho(\mathbf{x})$ in the configuration space $\mathbf{x}$ follows the Boltzmann distribution

$$\rho(\mathbf{x}) = \frac{e^{-\beta U(\mathbf{x})}}{\int e^{-\beta U(\mathbf{x})} d\mathbf{x}} = Z^{-1} e^{-\beta U(\mathbf{x})} \tag{1.2}$$

were Z denotes the *canonical partition function* and $\beta = (k_B T)^{-1}$ the inverse temperature with Boltzmann constant k_B.[1] In this thesis we will focus on alchemical transformations, $A \longrightarrow B$, which are represented by a small set of collective variables (CV's)

$$\xi_1(\mathbf{x}) : \mathbb{R}^{3N} \to \mathbb{R}, ..., \xi_n(\mathbf{x}) : \mathbb{R}^{3N} \to \mathbb{R} \tag{1.3}$$

that are sufficient to distinguish between key states of interest. In this framework the probability distribution can be expressed by

$$\rho(\xi) = \int \delta[\xi'(\mathbf{x}) - \xi] \rho(\mathbf{x}) d\mathbf{x} = \langle \delta[\xi'(\mathbf{x}) - \xi] \rangle_\xi \tag{1.4}$$

where $\langle\rangle_\xi$ denotes the ξ-conditioned marginal distribution. The Helmholtz free energy A, i.e. potential of the canonical ensemble, is defined as

$$A(\xi) = -\beta^{-1} \ln \rho(\xi) \tag{1.5}$$

Note that the shape of the obtained free energy surface (FES) is not gauge invariant against the choice of CV function ξ. This is, however, no concern in the calculation of free energy differences, because the CV dependent feature is integrated out to obtain the probability that the system occupies state A or B.[2]

$$\Delta A_{A\to B} = -\beta^{-1} \ln \frac{\int_B \rho(\xi)d\xi}{\int_A \rho(\xi)d\xi} = -\beta^{-1} \ln \frac{p_B}{p_A} \tag{1.6}$$

However, in computer simulations the direct phase-space integrals used in equation 1.2 and 1.4 are impossible to calculate.[1] Instead the time average $P(\xi)$ is computed by monitoring ξ during the simulation.

$$P(\xi) = \lim_{t\to\infty} \frac{1}{t} \int_0^t \rho[\xi(t')]dt' \tag{1.7}$$

In practice, $P(\xi)$ is approximated as a histogram along the reaction coordinate with bins of fixed size. Assuming that the system at hand behaves *ergodic*, i.e. every point in phase space is visited during a infinitely long simulation, the ensemble average $\rho(\xi)$ converges to the time average $P(\xi)$ of one system. However, transitions of high energy barriers during chemical reactions are typically rare events, that might happen on timescales far beyond what is feasible to compute, especially on accurate quantum chemical level of theory. Therefore most processes, for example in biochemical context, behave *quasinonergodic*. The problem of calculating accurate free energy differences hence really consists in adequate sampling of regions with high free energy (transition states) along the reaction coordinate.

For this purpose several *enhanced sampling* algorithms were proposed.[3–8] In this work we will focus on CV-based approaches, that alter the potential energy $U(\mathbf{x})$ of the system in a way, that increases the time spent in important regions of configuration space during a simulation. One of the oldest and simplest methods to achieve the same is *Umbrella Sampling* (US),[6] which will be introduced in the following section. The main focus of this work lies on *Adaptive Biasing Methods*,[9–11] which aim for uniform sampling along the reaction coordinate by an history dependent biasing term (section 1.3).

1.2 Umbrella Sampling (US)

In its most common form US introduces a fixed harmonic bias potential of the form

$$U_i^B(\xi) = \frac{k_i}{2}(\xi(t) - \xi_i)^2 \tag{1.8}$$

where the index i indicates a given bias window along the reaction coordinate with force constant k_i.[6] From simulations in window i the biased probability distribution $P_i^B(\xi)$ is obtained

$$
\begin{aligned}
P_i^B(\xi) &= \frac{\int d\mathbf{x}\, e^{-\beta(U(\mathbf{x})+U_i^B(\xi(\mathbf{x})))}\delta[\xi'(\mathbf{x}) - \xi]}{\int d\mathbf{x}\, e^{-\beta(U(\mathbf{x})+U_i^B(\xi(\mathbf{x})))}} \\
&= \frac{e^{-\beta U_i^B(\xi(\mathbf{x}))}}{\int e^{-\beta(U(\mathbf{x})+U_i^B(\xi(\mathbf{x})))}d\mathbf{x}} \int d\mathbf{x}\, e^{-\beta U(\mathbf{x})}\delta[\xi'(\mathbf{x}) - \xi]
\end{aligned}
\tag{1.9}
$$

where Z_B denotes the partition function of the biased potential energy $U(\mathbf{x}) + U_i^B(\xi(\mathbf{x}))$. Using equation 1.2 and 1.4 the connection of P_i^B to the unbiased probability distribution P_i^0 can be derived.

$$
\begin{aligned}
P_i^0(\xi) &= P_i^B(\xi)e^{\beta U_i^B(\xi)}\frac{\int e^{-\beta(U(\mathbf{x})+U_i^B(\xi(\mathbf{x})))}d\mathbf{x}}{\int e^{-\beta U(\mathbf{x})}d\mathbf{x}} \\
&= P_i^B(\xi)e^{\beta U_i^B(\xi)}\frac{Z_B}{Z_0}
\end{aligned}
\tag{1.10}
$$

Inserting into equation 1.5 gives the unbiased free energy

$$
\begin{aligned}
A_i(\xi) &= -\beta^{-1}\ln P_i^B(\xi) - U_i^B(\xi) - \beta^{-1}\ln\frac{Z_B}{Z_0} \\
&= A_B(\xi) - U_i^B(\xi) - F_i
\end{aligned}
\tag{1.11}
$$

The last therm of equation (1.11) is the free energy difference between the unbiased and biased system. Since it is a constant it can be ignored in the calculation of free energy differences from a single window.

However, in order to enhance sampling along the reaction coordinate usually a stratification strategy is applied simulating multiple overlapping windows. In this case the global unbiased distribution of all combined windows is calculated with the *Weighted Histogram Analysis Method* (WHAM).[12] The goal of WHAM is to choose the weights $p_i(\xi)$ in order to minimize the statistical error of the

global probability distribution of S combined windows

$$P^U(\xi) = \sum_i^S p_i(\xi) P_i^0(\xi) \tag{1.12}$$

under the constraint $\sum p_i = 1$. This is achieved by the iterative solution of the self-consistent WHAM equations

$$p_i(\xi) = \frac{N_i e^{-\beta_i U_i^B(\xi) + \beta_i F_i}}{\sum_j^S N_j e^{-\beta_j U_j^B(\xi) + \beta_j F_j}} \tag{1.13}$$

$$e^{-\beta_i F_i} = \int P^U(\xi) e^{-\beta_i U_i^B(\xi)} d\xi \tag{1.14}$$

where F_i enters equation (1.13) and P^U enters equation (1.14). N_i is the number of samples collected in window i. The biggest advantage of this approach is that all windows can be simulated in parallel on different computing nodes and only have to be combined once to solve the WHAM equations. This way extensive sampling of the reaction coordinate can be obtained efficiently. However, the biasing potentials have to be chosen manually for each window, requiring some knowledge of the free energy surfaces at hand prior to the simulation. Additionally, up to 50% of the simulation time can be spent in overlapping regions of ascending windows.[10] In the next section we will look at different approaches, that eliminate both drawbacks by replacing equation 1.8 with an time dependent term, which evolves during the simulation to achieve uniform sampling of the reaction coordinate.

1.3 Adaptive Biasing Methods

Instead of dividing the reaction coordinate in several windows, with adaptive biasing methods the free energy can be estimated from one single simulation. For this purpose the systems dynamics are biased towards states corresponding to large values of the free energy along the transition coordinate via a history-dependent biasing potential. In contrast to other importance sampling strategies like umbrella sampling, this methods require less knowledge of the free energy surface at hand prior to the simulation, because the bias is not chosen manually. Instead, the biasing potential automatically adapts during the simulation to enable diffusive behavior along the transition coordinate.[9,10]

There are multiple adaptive biasing methods available, only differing in the construction of the bias. Potential based methods like metadynamics (MtD) disfavor already visited states by accumulating repulsive potentials along the CV (section 1.3.1), while adaptive biasing force (ABF) methods compensate the mean force along the reaction coordinate to obtain uniform sampling (section 1.3.2). Finally, as discussed in section 1.3.3, using an extended Lagrangian, the technical requirements of ABF can be lifted without loosing its rigorous convergence behavior.[11] Additionally, this framework enables the combination of repulsive Mtd-potentials with ABF (meta-eABF/WTM-eABF), to unite the benefits of both approaches.[13,14]

Although not necessary, stratification can increase the convergence of all adaptive biasing methods significantly. However, in contrast to US simulations, no WHAM and hence no overlap between ascending windows is required. For example in ABF simulations the global PMF can be recovered by simply joining the force estimates of all simulations.[10] Multiple-Walker strategies can also be trivially implemented, by letting more than one simulation contribute to the same biasing potential, as will be discussed in section 1.3.4.[15]

In principle adaptive biasing methods only rely on sampling of the canonical ensemble, which will be obtained using Langevin dynamics in this work. However, all methods can also be applied in combination with other MD or MC engines. A schematic procedure of adaptively biased Langevin dynamics is given in Algorithm 1.

Algorithm 1 Velocity Verlet integrator for adaptively biased Langevin dynamics with atomic masses $\mathbf{M}$, coordinates $\mathbf{x}(t)$, momenta $\mathbf{p}(t)$, potential $U(\mathbf{x}(t))$, forces $F(\mathbf{x}(t))$ and friction coefficient γ,

---

while $t < t_{end}$ **do**

$\mathbf{p}(t + \frac{1}{2}\Delta t) \leftarrow \mathbf{p}(t) + \frac{1}{2}\left(F(\mathbf{x}(t))dt - \gamma \mathbf{M}^{-1}\mathbf{p}(t)dt + \sqrt{2\gamma\beta^{-1}}dW_t\right)$
$\mathbf{x}(t + \Delta t) \leftarrow \mathbf{x}(t) + \frac{2}{2+\gamma dt}\mathbf{M}^{-1}\mathbf{p}(t + \frac{1}{2}\Delta t)dt$
/* Langevin dynamics

$F(\mathbf{x}(t + \Delta t)) \leftarrow -\nabla U(\mathbf{x}(t + \Delta t))$
/* get physical QM/MM forces

$\xi \leftarrow f(\mathbf{x}(t + \Delta t))$
/* Calculate reaction coordinate from Cartesian coordinates

if $\xi_{min} \leq \xi \leq \xi_{max}$ **then**

$F_B(\xi, t + \Delta t)) \leftarrow F_B(\xi, t)) + \Delta F_B(\xi, t + \Delta t))$
/* update history dependent bias force along reaction coordinate incrementally

$F(\mathbf{x}(t + \Delta t)) \leftarrow -\nabla U(\mathbf{x}(t + \Delta t)) + F_B(\xi, t + \Delta t))\nabla\xi$
/* Add biasing force to physical force

else

$F(\mathbf{x}(t + \Delta t)) \leftarrow F(\mathbf{x}(t + \Delta t)) + k\min(|\xi - \xi_{min}|, |\xi - \xi_{max}|)\nabla\xi$
/* confine system to range of interest with harmonic constraining force

end if

$\mathbf{p}(t + \Delta t) = \frac{2-\gamma dt}{2+\gamma dt}\mathbf{p}(t + \frac{1}{2}\Delta t) - \frac{1}{2}\left(F(\mathbf{x}(t + \Delta t))dt - \gamma \mathbf{M}^{-1}\mathbf{p}(t + \frac{1}{2}\Delta t))dt + \sqrt{2\gamma\beta^{-1}}dW\right)$
/* Langevin dynamics

end while
/* get unbiased free energy

1.3.1 (Well-Tempered) Metadynamics (MtD/WTM)

MetaD biases a systems dynamic towards undersampled regions along the reaction coordinate $\xi(\mathbf{x})$, by accumulating repulsive potentials in regions that have already been visited.[9] The bias potential is typically build by a superposition of repulsive Gaussian kernels and can be written

$$U_{MtD}(\xi, t) = \sum_{k < \frac{t}{\tau_G}} \tau_G \omega \exp\left(- \sum_{i=1}^{N_{dim}} \frac{1}{2\sigma_i^2}(\xi_i(\mathbf{x}) - \xi_i(\mathbf{x}, t_k))^2 \right), \qquad (1.15)$$

with deposition rate τ_G, Gaussian height $\omega = W/\tau_G$ and variance σ^2 as free input parameters. Over the course of a simulation the bias potential fills local minima along the reaction coordinate until the systems evolution finally resembles a Brownian motion along the flattened free energy surface. The converged bias potential provides an unbiased estimate of the underlying free energy surface

$$A(\xi) = -U_{MtD}(\xi, t \to \infty) + C \qquad (1.16)$$

To avoid oscillation of U_{MtD} around the correct free energy, Well-Tempered metadynamics (WTM)[8] introduces an additional scaling factor of the Gaussian height

$$\omega(\xi, t) = \frac{W}{\tau_G} \exp\left(-\frac{U_{WTM}(\xi, t)}{k_B \Delta T} \right) \qquad (1.17)$$

to ensure an decrease of ω over time and smooth convergence of the Well-Tempered bias potential $U_{WTM}(\xi, t)$. The WTM potential does not fully compensate the free energy barrier, but can be controlled by parameter ΔT. For $\Delta T \to 0$ the bias is zero and ordinary MD is recovered, whereas the limit $\Delta T \to \infty$ corresponds to normal MtD. Figure 1.1 shows a numerical example of the effect of U_{MtD} and U_{WTM} on the effective free energy energy surface that is sampled over the course of a simulation.

To again obtain the unbiased PMF from $U_{WTM}(\xi, t)$ it has to be scaled with an according correction factor to compensate the effect of ΔT:

$$A(\xi) = -\frac{T + \Delta T}{\Delta T} U_{WTM}(\xi, t) \qquad (1.18)$$

Despite its appealing theoretical and practical simplicity MtD and WTM both suffer from a rather big set of free parameters (W, τ_G, σ and ΔT). All of them have significant impact on the outcome of the simulation and have to be chosen carefully. If the bias potential builds up too fast the system is driven far away from equilibrium and the obtained free energy estimate is wrong. Small parameters on the other hand result in bad convergence behavior and inefficient sampling.[16] The next section

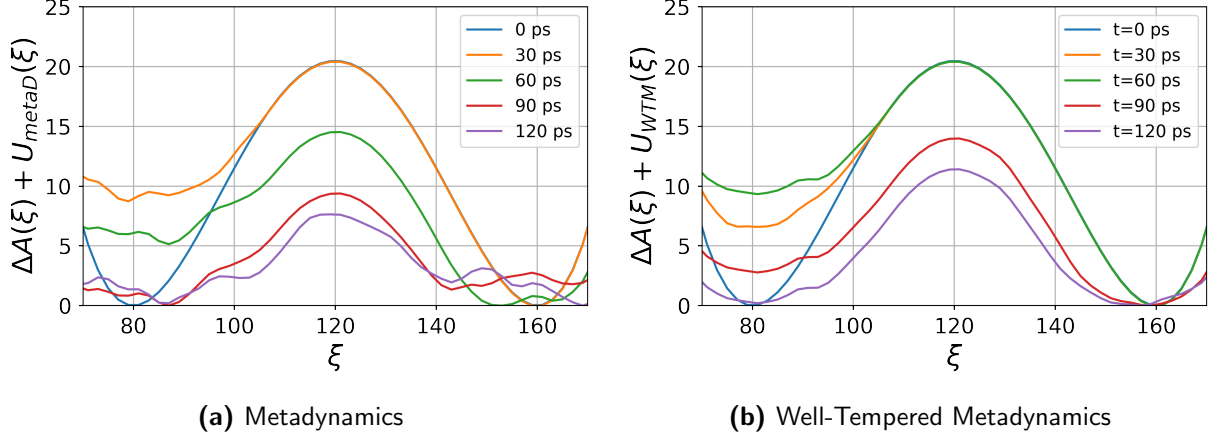will look into ABF, which is less dependent on free parameters and stands on a more rigorous mathematical framework.

(a) Metadynamics **(b)** Well-Tempered Metadynamics

Figure 1.1: Numerical example of the metaD and WTM algorithm for a particle in a 2D double well potential. The reaction coordinate ξ is the x-direction. Over time the bias potentials U_{MtD} and U_{WTM} build up and reduce the free energy barrier. U_{MtD} will ultimately completely remove the free energy barrier, whereas U_{WTM} converges to a certain limit, which can be controlled with parameter ΔT. Details are given in the Appendix.

1.3.2 Adaptive Biasing Force Method (ABF)

The intuition behind ABF is, that adding a force $\frac{\partial A(\mathbf{x})}{\partial \xi}\nabla \xi(\mathbf{x})$ that exactly compensates the average of the original force $-\nabla U(\mathbf{x})$ along a given coordinate would result in uniform sampling along this coordinate.[10] Historically, this idea emerged from thermodynamic integration (TI), were the free energy derivative is computed as the ensemble average of the instantaneous force, F, acting along the collective variable:

$$\frac{dA}{d\xi} = -\langle F \rangle_\xi \tag{1.19}$$

and the free energy is calculated as the integral over this force.[17,18] In practice, as one has no prior knowledge of the free energy derivative, ABF uses an on-the-fly estimate of the mean force acting along the reaction coordinate. For this purpose the transition coordinate ξ, connecting two end points, is divided in M equally spaced bins. The approximation of the bias force $\overline{F}(N,k)$ in bin k is than the average of collected force samples:[10]

$$\overline{F}(N,k) = \frac{1}{N^k}\sum_{\mu=1}^{N^k} F_\mu^k \tag{1.20}$$

At the beginning of the simulation $\overline{F}$ is scaled up by a linear ramp function $R(N^k)$ to prevent large fluctuations of the force estimate from driving the system away from equilibrium.

$$R(N^k) = \begin{cases} N_{full}, & N^k < N_{full} \\ 1, & N^k \geq N_{full} \end{cases} \tag{1.21}$$

The number of samples when the full biasing force is applied, N_{full}, and the bin size are the only free parameters. It can be proven mathematically, that equation 1.20 converges (under some constraints exponentially) to $-dA/d\xi$ for a sufficiently large number of force samples.[19] Figure 2.1 shows a numerical example for the flattening of the reaction coordinate during an ABF simulation.
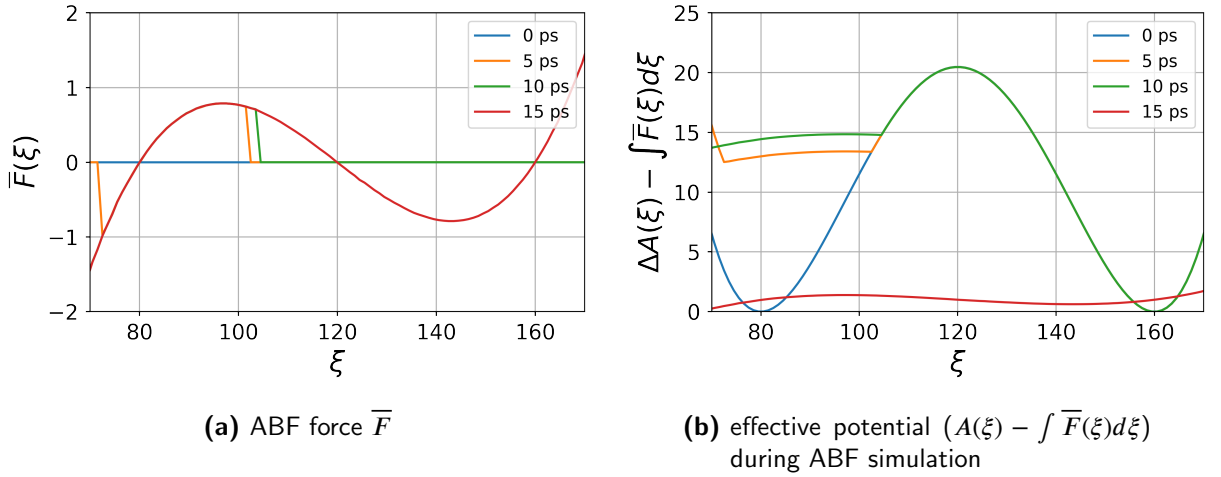

(a) ABF force $\overline{F}$

(b) effective potential $(A(\xi) - \int \overline{F}(\xi)d\xi)$ during ABF simulation

Figure 1.2: Numerical example of ABF algorithm for 2D double well potential. The reaction coordinate ξ is the x-direction. $\overline{F}(\xi)$ completely compensates the free energy barrier after 15 ps. From there on the system evolution resembles Brownian motion along the flattened reaction coordinate, which will ultimately lead to uniform sampling. Details are given in the Appendix.

For 1D reaction coordinates the unbiased free energy difference A can be trivially obtained from $\overline{F}$ with any numerical integration scheme (e.g. Simpson's rule). However, this is not true for multidimensional reaction coordinates, because $\overline{F}$ is no conservative force due to statistical errors. With simple integration the result would therefore depend on the integration path, which is physically incorrect. This problem can be avoided by using the *finite element method* (FEM).[20] For this purpose tent functions B_l are used to approximate $A(\xi)$

$$A(\xi) = \sum_l \alpha_l B_l(\xi) \tag{1.22}$$

where the coefficients α_l are obtained by minimizing

$$RMSD(\vec{\alpha}) = \sqrt{\frac{1}{M}\sum_k^M\left(\left(\sum_l \alpha_l \nabla B_l(\xi_k)\right) - \langle F(\xi_k)\rangle\right)^2} \tag{1.23}$$

with some numerical minimization routine (e.g. BFGS algorithm[21]) for force estimates of each dimension. An 1D example of the approximation of an arbitrary function by tent functions is given in 1.3a. To increase the smoothness of the reconstructed free energy surface in practice four control points per bin and dimension are chosen between data points, as shown in figure 1.3b.

(a) Example of 1D tent functions $B_l(\xi)$ (colored) approximating an arbitrary free energy surface $f(\xi)$ (blue)

(b) Illustration of numerical grid used for FEM integration of ABF-forces in 2D. Blue are original data points and orange control points obtained by linear interpolation. One 2D tent function is centered on each blue point.
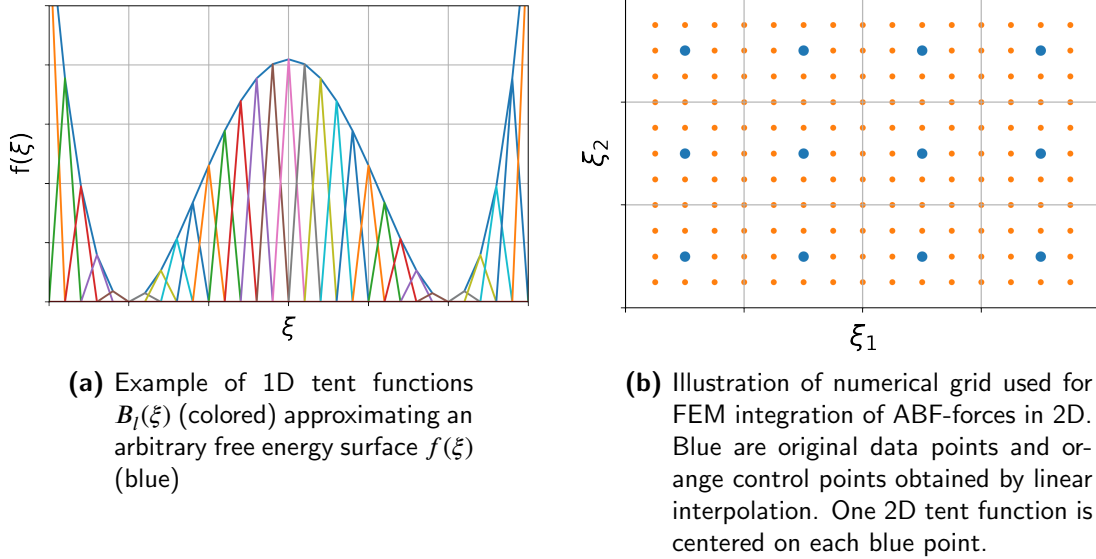
Figure 1.3: Illustration of the FEM integration

The last missing ingredient for the ABF method is an explicit expression for the instantaneous force samples F_ξ. Carter et al.[22] gave a first general expression:

$$F(\xi, \mathbf{q}) = -\frac{\partial U(\xi, \mathbf{q})}{\partial \xi} + \beta^{-1}\frac{\partial \ln|J(\xi, \mathbf{q})|}{\partial \xi} \tag{1.24}$$

which depends implicitly on a vector field $\partial x_i/\partial \xi$, hereafter referred to as "inverse gradient" and on an Jacobian correction term purely geometric in origin. The inverse gradient can be thought of as direction along which an infinitesimal change in ξ is propagated in Cartesian coordinates, the complementary coordinates $\mathbf{q}$ being kept constant. A major drawback of this formalism is the requirement of an full coordinate transform from Cartesian coordinates ($\mathbf{x}$) to generalized coordinates (ξ, $\mathbf{q}$).

This requirement could be lifted by den Otter,[5] who put forward the idea that the change in ξ can be propagated along an arbitrary vector field $\mathbf{v}_i$ ($\mathbb{R}^{3N} \to \mathbb{R}^{3N}$), provided it satisfies some orthonormality conditions. Extended to multidimensional reaction coordinates $\xi = (\xi_i)$ and in presence of a set of constraints $\sigma_k(\mathbf{x}) = 0$ these read:[7]

$$\mathbf{v}_i \cdot \nabla \xi_j = \delta_{ij} \tag{1.25}$$

$$\mathbf{v}_i \cdot \nabla \sigma_k = 0 \tag{1.26}$$

If all reaction coordinates ξ_i are orthogonal to one another and to all constraints, $\mathbf{v}_i = \nabla \xi_j / |\nabla \xi_j|^2$ is always a valid option, but not necessarily the best. Otherwise conditions 1.25 and 1.26 can be fulfilled by orthogonalization:[7]

$$\mathbf{v}_i(\mathbf{x}) = \frac{Q^i \nabla \xi_i(\mathbf{x})}{|Q^i \nabla \xi_i(\mathbf{x})|} \tag{1.27}$$

with projector Q^i given by the orthonormal basis $\{\hat{n}|_j^i(\mathbf{x})\}_{j \neq i}$ in the subspace spanned by $\{\nabla \xi_j(\mathbf{x})|\}_{j \neq i} \cup \{\nabla \sigma_j(\mathbf{x})|\}_{j=1,\dots,M}$:

$$Q^i = \mathbf{1} - \sum_{j \neq i} \hat{n}_j^i(\mathbf{x}) \otimes \hat{n}_j^i(\mathbf{x}) \tag{1.28}$$

Replacing the inverse gradient by vectorfield $\mathbf{v}_i$, expression 1.24 finally reduces to:

$$F(\xi_i, \mathbf{x}) = -\nabla U(\mathbf{x}) \cdot \mathbf{v}_i(\mathbf{x}) + \beta^{-1} \nabla \cdot \mathbf{v}_i(\mathbf{x}) \tag{1.29}$$

but still involves the calculation of second derivatives in the form of the divergence of vector fields $\mathbf{v}_i$.[10] Analytic expressions for distances, projected distances, bend angles and torsion angles, used in the present work, are given in the appendix. However, for complicated CV's like torsion angles orthogonalization via equation 1.27 becomes exceedingly tedious and impractical, significantly limiting the applicability of ABF for multidimensional reaction coordinates.

1.3.3 extended Lagrangian based methods (eABF, meta-eABF, WTM-eABF)

To circumvent the technical requirements of ABF Lesage et al.[11] proposed an more flexible approach termed extended ABF (eABF), where additional coordinates λ_i with mass m_i, which are coupled to reaction coordinates ξ_i with harmonic potentials, are introduced. The extended system $(\mathbf{x}, \lambda)$ evolves according to Langevin dynamics in the potential

$$U_{ext}(\mathbf{x}, \lambda) = U(\mathbf{x}) + \sum_i^n \frac{(\xi_i(\mathbf{x}) - \lambda_i)^2}{2\sigma_i^2}. \tag{1.30}$$

where the mass of the fictitious particle m_i and the variance between extended coordinate and physical coordinate $\sigma_i^2 = (\beta k_i)^{-1}$ are free parameters. The key intuition behind eABF is, that in the tight coupling (low σ, high k) limit efficient sampling of λ will result in efficient sampling of ξ. Therefore, to obtain uniform sampling along ξ it is sufficient to bias the dynamics of λ. The inverse gradient is chosen as zero for all physical coordinates $\mathbf{x}$ and 1 for λ. This way constraints 1.25 and 1.26 are always satisfied, which is especially useful for calculations involving a set of non-orthogonal reaction coordinates. In practice the Langevin integrator for the extended system is implemented separately from the physical MD engine and can be coupled to any collective variable of interest without additional effort.

Sampling the extended system in thermal equilibrium gives the following Boltzmann distribution in λ:

$$p_\lambda(\lambda) \propto \int e^{-\beta U_{ext}(\mathbf{x}, \lambda)} d\mathbf{x} \tag{1.31}$$

The bias on λ is the running average over the spring force in λ-bin k:

$$\overline{F}(\lambda_i, k) = \frac{\partial A^k(\lambda_i)}{\partial \lambda_i} = \frac{1}{N^k \sigma_i^2} \sum_{\mu=1}^{N^k} (\lambda_{i,\mu}^k - \xi_{i,\mu}^k) \tag{1.32}$$

Note that the instantaneous force estimate in each step no longer depends on the all atomic coordinates $\mathbf{x}$ like in normal ABF (cf. equation 1.29), but only on $(\lambda_i - \xi_i)$. This has the advantageous side effect, that the noise in $\overline{F}$ can be reduced significantly and faster convergence of $\overline{F}$ can be expected, if σ^2 is well chosen.[11] Specifically, the mean square fluctuation of the spring force is proportional to $(\beta\sigma)^{-1}$, which shows that the variance of $\overline{F}$ is smaller for higher values of σ. For small values of N^k the linear ramp function $R(N, k)$ given by equation 1.21 is used nonetheless. Overall, with the

time dependent bias converging towards $A(\lambda_i)$, the biased Boltzmann distribution is given by:

$$\tilde{p}_\lambda(\lambda_i) \propto \int e^{-\beta U_{ext}(\mathbf{x}, \lambda_i) + A(\lambda_i)} d\mathbf{x} \tag{1.33}$$

We now need to find a robust estimator to recover the unbiased free energy of the physical system $A(\xi_i)$ from biased dynamics in the extended system. The naive approach is to approximate $A(\xi_i)$ with $A(\lambda_i)$, which can be obtained by simply integrating the bias force $\overline{F}(\lambda_i)$ like in normal ABF. However, this will only produce unbiased results in the tight coupling limit for $\sigma \to 0$, where standard ABF is recovered.

An asymptotically unbiased estimator of the free energy can be derived by correcting the free energy gradient obtained from the eABF-biased distribution $\tilde{p}(\xi)$ with the average biasing force on ξ

$$\frac{\partial A(\xi_i)}{\partial \xi_i} = -\beta^{-1} \frac{\partial \ln \tilde{p}(\xi_i)}{\partial \xi_i} + k(\langle \lambda_i \rangle_{\xi_i} - \xi_i) \tag{1.34}$$

which is called *Corrected z-averaged restraint* (CZAR) and can be calculated on-the-fly during the simulation at little extra cost by introducing two additional accumulators for $\tilde{p}(z)$ and $k(\langle \lambda_i \rangle_z - z_i)$ which are only joined at output times.[11] Figure 1.4 gives a numerical example for the application of eABF and figure 1.5 shows a schematic overview of the full eABF/CZAR algorithm.
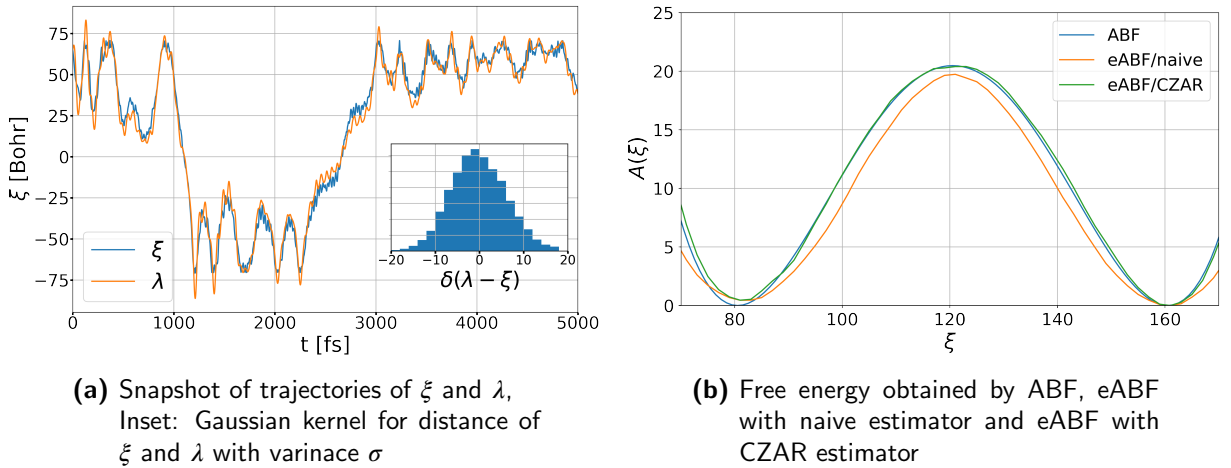

(a) Snapshot of trajectories of ξ and λ, Inset: Gaussian kernel for distance of ξ and λ with varinace σ

(b) Free energy obtained by ABF, eABF with naive estimator and eABF with CZAR estimator

Figure 1.4: Numerical examples for eABF algorithm with $m_\lambda = 20$ a.u. and $\sigma = 7$ Bohr. The reaction coordinate ξ is the x-direction.

Figure 1.5: Flowchart of the eABF algorithm. $k_{\lambda,\xi}$ denote λ or ξ conditioned bin indices in bins with size $\Delta\xi$. $\tilde{\rho}_\lambda$ and $\tilde{\rho}_\xi$ are λ- or ξ- conditioned histograms and R is the ABF ramp function. F_B is the sum bias force samples. Outputs are written in fixed intervals.
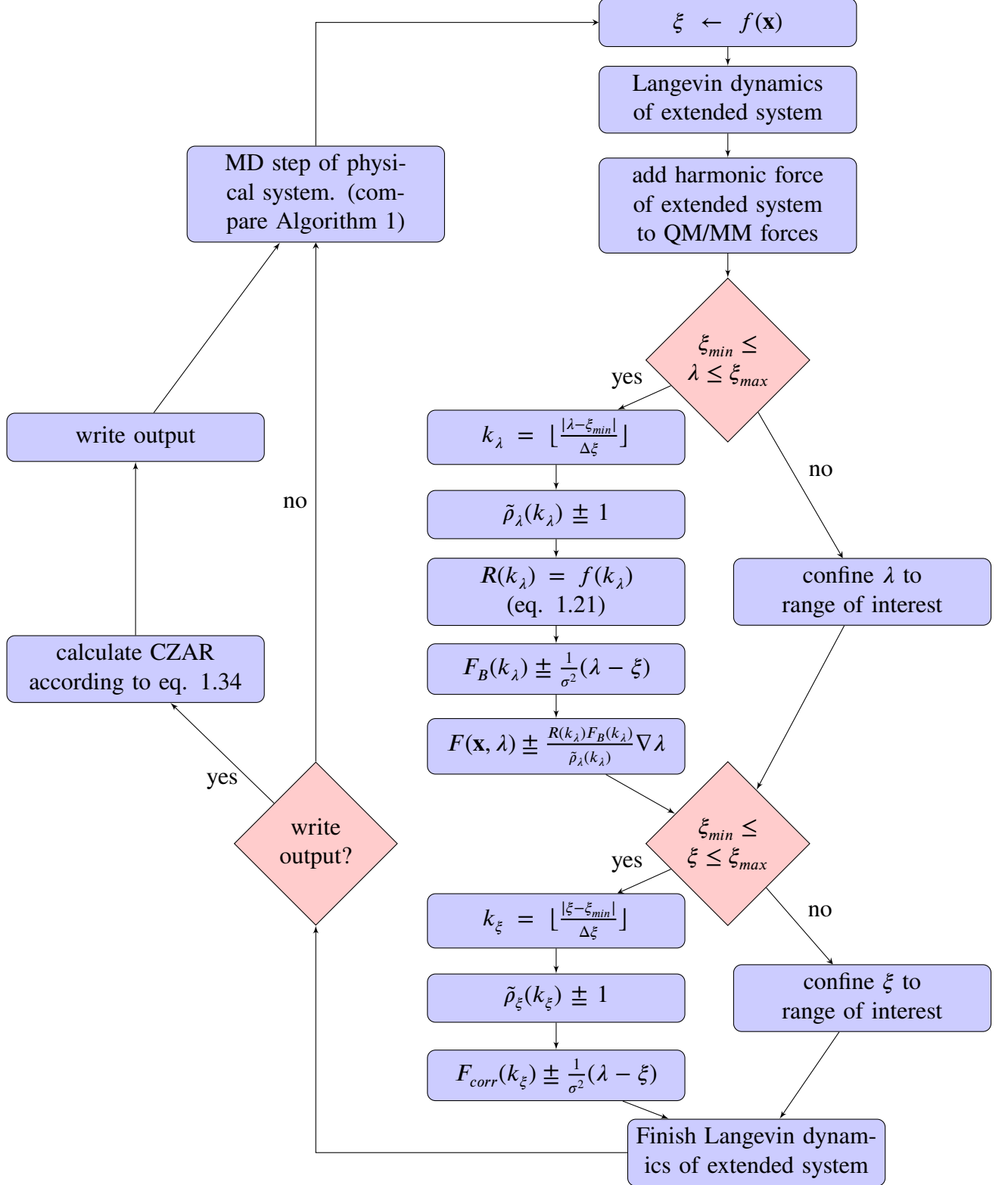
The CZAR estimator is able to recover the unbiased free energy derivative only from time trajectory (z_i, λ_i) of the extended system, independent of the bias on λ_i, as long as the simulation is in thermal equilibrium. It is therefore possible to bias the extended variable with any arbitrary force, for example metadynamics (eMtd, eWTM). One particular useful choice for $\overline{F}(\lambda_i, k)$ is to combine repulsive MtD or WTM potentials with the eABF bias, an idea the authors figuratively described with 'Shaving Barriers, and Flooding Valleys'.[13] This way the inherent difficulty for ABF methods to get accurate force estimates in high free energy regions at the beginning of the simulation can be solved, by pushing the system towards this regions. The full MtD-eABF and WTM-eABF bias terms read

$$F_B^{MtD-eABF}(\lambda_i, t) = -\frac{\langle \xi_i(\mathbf{x}) - \lambda_i \rangle_{\lambda_i}}{\beta \sigma_i^2} + \frac{\partial}{\partial \lambda_i} U_B^{MtD}(\lambda_i, t) \tag{1.35}$$

$$F_B^{WTM-eABF}(\lambda_i, t) = -\frac{\langle \xi_i(\mathbf{x}) - \lambda_i \rangle_{\lambda_i}}{\beta \sigma_i^2} + \frac{\partial}{\partial \lambda_i} U_B^{WTM}(\lambda_i, t) \tag{1.36}$$

with U_B^{MtD} and U_B^{WTM} given by equations 1.15 and 1.17. The relative contributions of the eABF and MtD/WTM term will fluctuate over the course of a simulation, but the sum of both will eventually converge. However, accurate evaluation of F_B is not even required, because it only enhances sampling of the reaction coordinate and the true PMF is obtained independently from CZAR.

1.3.4 Multiple-Walker (MW) strategy for adaptive biasing

By using the previously described methods the required simulation time to sample the reaction coordinate of alchemical transitions can be reduced from multiple ns to few ps, reducing the required MD steps by roughly a factor of 1000. However, instead of using only one single MD simulation, i.e. *walker*, one can think of letting multiple walkers contribute to the same bias, to exploit the full potential of modern HPC clusters. For this purpose all accumulators f required for the respective method, such as histograms, bias forces or potentials, are synced between otherwise independent walkers in fixed time intervals. Global and local accumulators evolve according to the following update formula

$$f_l(t_{sync}) \leftarrow f_g + \left(f_l(t_{sync}) - f_l(t_{sync} - 1) \right) \tag{1.37}$$

$$f_g \leftarrow f_l(t_{sync}) \tag{1.38}$$

with time of synchronization t_{sync}. Global contributions f_g are stored in a single buffer, which can only be accessed by one walker simultaneously to avoid inconsistencies.

1.4 Calculating free energy barriers and reaction rate constants

All methods described so far enable the efficient calculation of some CV-based FES, as defined in equation 1.5. As already noted in section 1.1, this formulation is not invariant under the transformation of CV, e.g. $A(\xi(\mathbf{x})) \neq A(\zeta(\mathbf{x}))$, which should not come as a surprise as $\exp(-\beta A(\xi))/Z$ and $\exp(-\beta A(\zeta)/Z$ gives the marginal distribution along ξ or ζ, respectively. Critical points like local minima or saddle points of $A(\xi)$ have therefore no coordinate independent meaning and can not be used to obtain kinetic information, like free energy barriers $\Delta^{\ddagger} A_{A \to B}$ or reaction rate constants $k_{A \to B}$.[2]

However, there also exists a different free energy definition, invariant under CV transformation, which will hereafter be referred to as geometric free energy, $A^G(\xi)$.[23] For this purpose the delta function of equation 1.4 is replaced by a proper surface measure

$$\rho^G(\xi) = \int e^{-\beta U(\mathbf{x})} d\xi(\mathbf{x}) \tag{1.39}$$

and the geometric free energy is defined analogous to the normal free energy (equation 1.5)

$$A^G(\xi) = -\beta^{-1} \ln \rho^G(\xi) \tag{1.40}$$

Instead of the marginal distribution of the collective variable $\rho^G(\xi)$ is the probability density of the FES. The statistical ensembles related to both free energy definitions can be augmented into one another by the Fixman correction term $U \mapsto U + \beta^{-1} \ln |\nabla \xi|$ which relates both free energies through the ξ-conditioned average of gradients using

$$A^G(\xi) = A(\xi) - \beta^{-1} \ln \langle \lambda | \nabla \xi | \rangle_{\xi} \tag{1.41}$$

where the length scale λ is introduced to keep the exponential term dimensionless. Because all methods described in section 1.3 already involve calculation of the gradient this correction can be obtained with almost no additional effort.[2] For multidimensional CVs $\langle \lambda | \nabla \xi | \rangle_{\xi}$ is replaced by $\langle \lambda^n \det d \rangle_{\xi}$ with

$$d_{ij}^2 = \nabla \xi_i \cdot \nabla \xi_j \tag{1.42}$$

Figure 1.6 gives an numerical example of the relation of A^G and A in an ABF simulation. Using A_G free energy barriers can be calculated that are independent of the choice of CV. However, simply taking the difference $A_{TS}^G - A_{min}$ would still contain the partition function inside a volume element around the transition state.[2] It therefore entails a dependence on the unit system of the simulation

which finally has to be removed to obtain the free energy barrier according to

$$\Delta^{\ddagger} A_{A \to B} = A_{TS}^{G} + \frac{n}{\beta} \ln \frac{\lambda}{h} \sqrt{\frac{2\pi m}{\beta}} - A_{min} \tag{1.43}$$

where n is the number of dimensions of the CV and h is the Planck constant. The mass factor $\sqrt{m}$ can be dropped by using mass scaled coordinates $\mathbf{q} = \sqrt{m}\mathbf{x}$. λ than has the unit of $\mathbf{q}$. The reaction rate constant $k_{A \to B}$ can be obtained from the Eyring equation

$$k_{A \to B} = \frac{\kappa}{h\beta} e^{-\beta \Delta^{\ddagger} A_{A \to B}} \tag{1.44}$$

where κ compensates the lowering of the reaction rate by recrossing of the transition state. However, in most chemical systems transport over the transition state can be assumed to be ballistic ($\kappa = 1$).[2] Figure 1.6 gives an example of the connection of standard and geometric free energy.
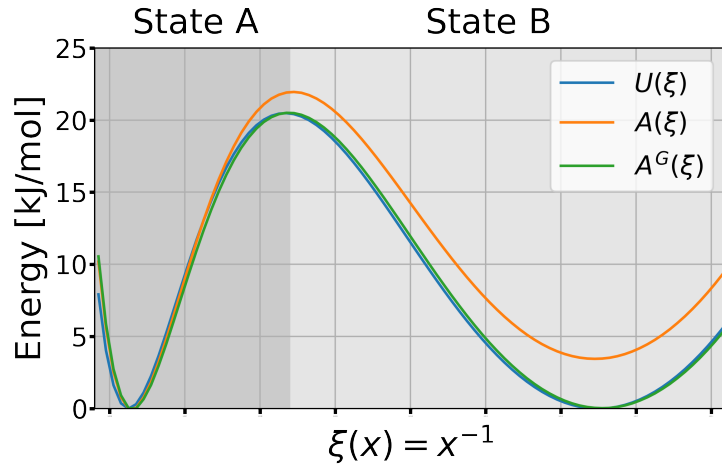
Figure 1.6: Standard and geometric free energy of an 2D double well potential sampled along $\xi(x) = 1/x$ using the ABF method.
Both minima are equally probable, which is correctly described by the standard free energy difference $\Delta A_{A \to B} = -\beta^{-1} \ln \frac{\int_{B} \rho(\xi)d\xi}{\int_{A} \rho(\xi)d\xi} = 0$.
Correcting $A(\xi)$ according to equation 1.41 gives the geometric free energy A^{G}. In this example the geometric free energy surface is equal to the underlying potential energy surface $U(\xi)$. This is only true if the collective variable is adiabatically separable from the remaining degrees of freedom, in this case the y-direction.[23] Details are given in the Appendix.

2 Appendix

2.1 Definition of Collective Variables

In the present work distances, projected distances, angles and torsion angles between groups of atoms were used as collective variables for US, MtD, ABF, eABF or meta-eABF simulations. Below analytic expressions for the calculation of all CVs, their gradients, inverse gradients and divergence of inverse gradients are given.

The center of mass r^C of a group of N atoms is calculated according to

$$r^C(\mathbf{x}) = \frac{1}{M^C} \sum_{i=0}^{N} m_i r_i \tag{2.1}$$

with mass of each atom m_i and Cartesian coordinates $r_i = (x_i, y_i, z_i)$. M^C is the mass of the group. All following collective variables are calculated between centers of mass. The index C will be dropped for convenience. To obtain derivatives of single atoms the gradient its corresponding group is weighted with m_i/M. The distances between group i and j is given by

$$r_{ij} = r_j - r_i \tag{2.2}$$

and the absolute distance is $d_{ij} = \|r_{ij}\|$. The derivatives with respect to r_i and r_j are defined by

$$\frac{\partial d_{ij}}{\partial r_j} = \frac{r_{ij}}{\|r_{ij}\|} = \hat{r}_{ij} \tag{2.3}$$

$$\frac{\partial d_{ij}}{\partial r_i} = -\hat{r}_{ij} \tag{2.4}$$

where the hat denotes unit vectors. The projection of distance $\|r_{ij}\|$ onto another axis r_{jk} is defined by

$$d_{ij}^p = r_{ij} \cdot \hat{r}_{jk} \tag{2.5}$$

and the derivative of the projection is defined by

$$\frac{\partial d_{ij}^p}{\partial r_i} = -\hat{r}_{jk} \tag{2.6}$$

$$\frac{\partial d_{ij}^p}{\partial r_j} = \hat{r}_{jk} \tag{2.7}$$

The bond angle θ between three centers (r_i, r_j, r_k) is defined by the dot product of bond distances r_{ji} and r_{jk}:

$$\cos \theta = \hat{r}_{ji} \cdot \hat{r}_{jk} \tag{2.8}$$

The gradient of θ is then given by

$$\frac{\partial \theta}{\partial r_i} = \frac{(r_{ji} \times (r_{ji} \times r_{jk}))}{\|r_{ij}\| \|(r_{ji} \times (r_{ji} \times r_{jk}))\|} \tag{2.9}$$

$$\frac{\partial \theta}{\partial r_k} = \frac{(r_{kj} \times (r_{ji} \times r_{ji}))}{\|r_{kj}\| \|(r_{kj} \times (r_{ji} \times r_{ji}))\|} \tag{2.10}$$

$$\frac{\partial \theta}{\partial r_j} = -\frac{\partial \theta}{\partial r_i} - \frac{\partial \theta}{\partial r_k} \tag{2.11}$$

The torsion angle ϕ between four Cartesian coordinates (r_i, r_j, r_k, r_l) is given by:

$$\tan \phi = \frac{(\hat{r}_{ij} \times n_1) \cdot n_2}{n_1 \cdot n_2} \tag{2.12}$$

with

$$n_1 = r_{jk} - (r_{ij} \cdot \hat{r}_{jk})\hat{r}_{jk} \tag{2.13}$$

$$n_2 = r_{kl} - (r_{kl} \cdot \hat{r}_{jk})\hat{r}_{jk} \tag{2.14}$$

The derivative of ϕ in Cartesian coordinates is given by:

$$\frac{\partial \phi}{\partial r_i} = -\frac{\hat{r}_{ji} \times \hat{r}_{jk}}{\|r_{ij}\| \sin^2 \theta_{ijk}} \tag{2.15}$$

$$\frac{\partial \phi}{\partial r_l} = -\frac{\hat{r}_{kl} \times \hat{r}_{jk}}{\|r_{kl}\| \sin^2 \theta_{jkl}} \tag{2.16}$$

$$\frac{\partial \phi}{\partial r_j} = \left(\frac{\|r_{ij}\| \cos \theta_{ijk}}{\|r_{jk}\|} - 1\right)\frac{\partial \phi}{\partial r_i} - \left(\frac{\|r_{kl}\| \cos \theta_{jkl}}{\|r_{jk}\|}\right)\frac{\partial \phi}{\partial r_l} \tag{2.17}$$

$$\frac{\partial \phi}{\partial r_k} = -\frac{\partial \phi}{\partial r_i} - \frac{\partial \phi}{\partial r_l} - \frac{\partial \phi}{\partial r_j} \tag{2.18}$$

where $\partial \phi / \partial r_j$ and $\partial \phi / \partial r_k$ are defined to fulfill the conservation of (angular) momentum.

Always choosing as inverse gradient $\mathbf{v}(\xi) = \nabla \xi / \|\nabla \xi\|^2$ the divergence of vector fields $\mathbf{v}_i(\xi)$ are given by

$$\nabla \cdot \mathbf{v}(d) = \frac{2}{d}, \tag{2.19}$$

$$\nabla \cdot \mathbf{v}(d^p) = 0, \tag{2.20}$$

$$\nabla \cdot \mathbf{v}(\theta) = \cot \theta, \tag{2.21}$$

$$\nabla \cdot \mathbf{v}(\phi) = 0, \tag{2.22}$$

for distances, projected distances, angles and dihedrals, respectively.

2.2 Numerical examples

To test adaptive biasing methods the 2D test potentials U_1 and U_2 were used. For numerical tests of adaptive biasing methods a particle with a mass of 10 a.u. is simulated according to Algorithm 1 with a step size of 5 fs and at a temperature of 300 K.

$$U_1(x, y) = a(x - c)^2(x + d)^2 + by^2 \tag{2.23}$$

$$U_2(x, y) = -\ln\left(e^{-a(x-c)^2 - b(y-d)^2} + e^{-a(x+c)^2 - b(y+d)^2}\right) \tag{2.24}$$

Table 2.1: Free Parameters for definition of U_1 and U_2.

	a	b	c	d
U_1	$8.0e-6$ kJ/mol	0.5 kJ/mol	80 Bohr	160 Bohr
U_2	0.005 kJ/mol	0.040 kJ/mol	40 Bohr	20 Bohr

(a) U_1 **(b)** U_2
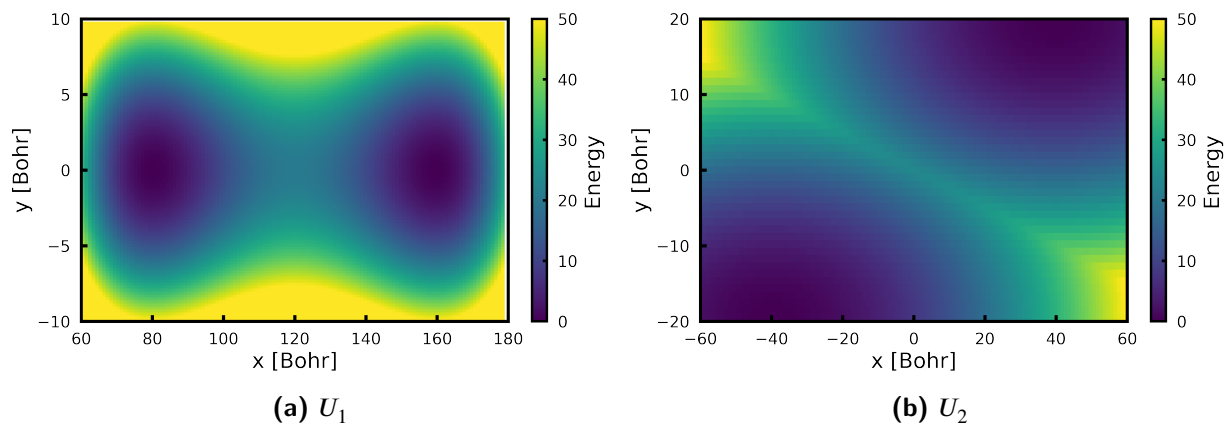
Figure 2.1: Potentials for test calculations

List of Tables

List of Figures

Bibliography

[1] C. Chipot, A. Pohorille, *Free energy calculations*, Springer, **2007**.

[2] K. M. Bal, S. Fukuhara, Y. Shibuta, E. C. Neyts, *The Journal of Chemical Physics* **2020**, *153*, 114118.

[3] W. Jiang, B. Roux, *Journal of chemical theory and computation* **2010**, *6*, 2559–2565.

[4] Y. Sugita, Y. Okamoto, *Chemical physics letters* **1999**, *314*, 141–151.

[5] W. K. den Otter, *The Journal of Chemical Physics* **2000**, *112*, 7283–7292.

[6] J. Kästner, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 932–942.

[7] G. Ciccotti, R. Kapral, E. Vanden-Eijnden, *ChemPhysChem* **2005**, *6*, 1809–1814.

[8] A. Barducci, G. Bussi, M. Parrinello, *Physical review letters* **2008**, *100*, 020603.

[9] A. Barducci, M. Bonomi, M. Parrinello, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 826–843.

[10] J. Comer, J. C. Gumbart, J. Henin, T. Lelievre, A. Pohorille, C. Chipot, *The Journal of Physical Chemistry B* **2015**, *119*, 1129–1151.

[11] A. Lesage, T. Lelievre, G. Stoltz, J. Henin, *The Journal of Physical Chemistry B* **2017**, *121*, 3676–3685.

[12] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *Journal of computational chemistry* **1992**, *13*, 1011–1021.

[13] H. Fu, H. Zhang, H. Chen, X. Shao, C. Chipot, W. Cai, *The journal of physical chemistry letters* **2018**, *9*, 4738–4745.

[14] H. Fu, X. Shao, W. Cai, C. Chipot, *Accounts of Chemical Research* **2019**, *52*, 3254–3264.

[15] K. Minoukadeh, C. Chipot, T. Lelievre, *Journal of Chemical Theory and Computation* **2010**, *6*, 1008–1017.

[16] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, M. Parrinello, *The journal of physical chemistry B* **2005**, *109*, 6714–6721.

[17] J. G. Kirkwood, *The Journal of chemical physics* **1935**, *3*, 300–313.

[18] R. W. Zwanzig, *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.

[19] H. Alrachid, T. Lelièvre, *The SMAI journal of computational mathematics* **2015**, *1*, 55–82.

[20] E. Darve, D. Rodriguez-Gomez, A. Pohorille, *The Journal of chemical physics* **2008**, *128*, 144120.

[21] J. Nocedal, S. Wright, *Numerical optimization*, Springer Science & Business Media, **2006**.

[22] E. Carter, G. Ciccotti, J. T. Hynes, R. Kapral, *Chemical Physics Letters* **1989**, *156*, 472–477.

[23] C. Hartmann, J. C. Latorre, G. Ciccotti, *The European Physical Journal Special Topics* **2011**, *200*, 73–89.