# Micro-Macro accelerated Markov Chain Monte Carlo Methods with Applications in Molecular Dynamics

**Hannes Vandecasteele**

Supervisor:
Prof. dr. ir. G. Samaey

Dissertation presented in partial
fulfillment of the requirements for the
degree of Doctor of Engineering
Science (PhD): Computer Science

September 12, 2023

# Micro-Macro accelerated Markov Chain Monte Carlo Methods with Applications in Molecular Dynamics

**Hannes VANDECASTEELE**

Examination committee:
Prof. dr. ir. The Chairman, chair
Prof. dr. ir. G. Samaey, supervisor
Prof. dr. ir. Ronald Cools
Prof. dr. Christian Maes
Prof. dr. Ioannis Kevrekidis
  (Johns Hopkins University)
Prof. dr. Virginie Ehrlacher
  (Ecole Nationale des Ponts et Chaussées)

September 12, 2023

# Voorwoord

Voor zolang ik mij kan herinneren ben ik gefascineerd door wiskunde, programmeren en computers. Vanaf mijn 12 jaar heb ik leren programmeren in C++, Java en later Python, en mijn interesse voor wiskunde kwam niet veel later. Niemand had me op dat moment verteld dat het nét de ingredienten waren om succesvol een doctoraat in de toegepaste wiskunde voor te leggen. Van mijn promotor Giovanni kreeg ik de kans om te werken op nieuwe simulatie methodes voor moleculaire systemen. Tijdens de voorbije vijf jaar heb ik een nieuwe sampling mehode voorgesteld die het samplen versnelt van moleculaire processen met een tijd-schaal scheiding. Later hebben we ontdekt dat dit werk toepassingen heeft in het zogenaamde 'protein folding' probleem, het zoeken voor stabiele conformaties van hoog-dimensionale proteïnen.

Een voorwoord is de ideale plaats om terug te blikken op het doctoraat, mijn research en vooral het schrijven van deze tekst. Het is ook een plek om terug te gaan naar mijn initiële verwachtingen van dit doctoraat, mijn evolutie, en of deze verwachtingen al dan niet werden ingelost. Ik ben dit doctoraat begonnen na het vervolledigen van een scriptie bij mijn promoter, hopend dat het process van een doctoraat hetzelfde zou zijn, alleen diepgaander. Ik ben blij dat deze hoop inderdaad snel werd ingelost. Door de structuur van de onderzoeksgroep NUMA heb ik ook de kans gehad om met veel interessante mensen te praten over een grote varieteit aan onderwerpen. Dat is in andere universiteiten soms anders door de structuur van departementen en onderzoeksgroepen.

De universitaire omgeving heeft meer te bieden dan onderzoek alleen. Ik heb sterk genoten van lesgeven aan de studenten ingenieurswetenschappen. Het ogenblik dat een student een nieuw onderwerp begrijpt na soms uren uitleg is een speciaal moment. Het zijn die gebeurtenissen die mij drijven om beter les te geven.

De voorbije vijf jaar heb ik niet enkel in Leuven gespendeerd. Van mei tot oktober 2022 heb ik het geluk gehad om een buitenlands onderzoeksverblijf te

kunnen doen bij Yannis Kevrekidis aan Johns Hopkins University in Baltimore, Maryland. Tijdens dat verblijf heb ik met mijn collega, en ondertussen goede vriendin, Anastasia op reactie-pad methoden gewerkt. Het doel van deze methode is lokale minima in een potentiaal met elkaar te verbinden via een zadelpunt. Toepassingen zijn, onder andere, het vinden van stabiele configuraties van hoog-dimensionale proteïnen. Ik heb zo veel van die periode genoten, dat ik er binnenkort een postdoc zal starten.

In deze laatste paragraaf wil ik nog mijn promotor, collega's, vrienden, ouders en mijn vrouw, Julie bedanken voor hun steun gedurende de voorbije vijf jaar. Ze hielpen me er telkens terug bovenop als mijn code niet deed wat ik wou, of wanneer een bewijs niet werkte, de supercomputer mijn jobs niet wou starten, of wanneer ik het punt niet meer zag in het onderzoek en dit doctoraat. Ik wil ze ook bedanken om mijn toekomst in de goede richting te sturen wanneer ik het niet meer zag zitten. Verder wil ik ook nog Yannis, Anastasia en Nicholas bedanken om mijn verblijf in Baltimore aangenaam en interessant te maken, dat heeft meer voor mij gedaan dan jullie kunnen voorstellen. Ten slotte wil ik ook Wim Vranken bedanken voor de interessante discussies over de toepassingen en toekomst van mijn werk en voor de microscopische simulatie data van de proteïnen van het laatste hoofdstuk.

Tot slot wil ik het FWO bedanken voor het vertrouwen in mij en de financiering over de voorbije vijf jaar.

# Abstract

In many large molecular systems, there is a time-scale separation between the fast motion of microscopic degrees of freedom, and a small number of slow, macroscopic variables that determine the global conformation of the molecule. This time-scale separation presents a computational bottleneck for numerical sampling methods. On the one hand, one needs small proposal moves for stability of the microscopic dynamics, and on the other hand one needs long simulation times to fully explore the potential energy surface of the molecule.

In this thesis, we introduce the micro-macro Markov chain Monte Carlo (mM-MCMC) method that accelerates sampling of the Gibbs distribution in case of a time-scale separation. The mM-MCMC method works in five steps: 1) Compute the macroscopic reaction coordinate value of the current molecule; 2) Generate a new reaction coordinate proposal; 3) Accept or reject the macroscopic proposal based on an approximate macroscopic distribution; 4) Reconstruction to the microscopic level; 5) Accept or reject the microscopic sample based on the Gibbs distribution. We study two types of reconstruction, direct and indirect. Indirect reconstruction uses a biased simulation to pull the microscopic sample to the new reaction coordinate value, while direct reconstruction is simply drawing a molecule from a reconstruction distribution. The former is always applicable, but more costly. The computational gain steps from the fact that most proposals are rejected at the macroscopic level, at low computational cost, while microscopic states, once reconstructed, are always accepted. We analytically show convergence and discuss rate of convergence of the proposed algorithm, and numerically perform an efficiency study on three molecular examples, a three-atom molecule, butane and alanine-dipeptide.

The mM-MCMC method faces another computational bottleneck. One needs to know the free energy of the macroscopic variables to evaluate the microscopic acceptance probability. This free energy is expensive to compute so one needs to rely on a cheap, on-the-fly approximation that is statistically unbiased. The pseudo-marginal method is an ideal solution to this problem. The pseudo-

marginal method computes an approximation to the free energy by importance sampling where the samples are generated from indirect reconstruction. Only a few samples are needed to compute an unbiased estimate, making the estimate cheap. We illustrate the bias and variance, as well as the performance of the pseudo-marginal method on two molecular examples, butane and alkane.

In the final part of the thesis, we study backmapping as an alternative to indirect reconstruction for proteins. For such macromolecules, the macroscopic variables are typically given a linear combination of atoms, repeated over constituting units, such as the MARTINI coarse-grained force field. Indirect reconstruction suffers from the curse of dimensionality when the reaction coordinate is high dimensional. Backmapping solves this problem in two steps. First, it creates a microscopic molecule that is consistent with the macroscopic sample through a step called 'lifting'. Second, backmapping runs a time-dependent simulation where the microscopic forces slowly switch on, and the macroscopic forces slowly die down, to maintain stability. We present the mM-MCMC method with backmapping and show its performance on a single disulfide peptide.

# Beknopte Samenvatting

In veel grote moleculaire systemen is er een tijdschaalscheiding tussen de snelle beweging van microscopische vrijheidsgraden en een klein aantal trage, macroscopische variabelen die de globale conformatie van de molecule bepalen. Deze tijdschaalscheiding vormt een computationeel knelpunt voor numerieke bemonsteringsmethoden. Aan de ene kant heeft men een kleine tijdstap nodig om stabiliteit te bewaren van de microscopische dynamica, en aan de andere kant heeft men een lange simulatietijd nodig om het volledige potentiële energieoppervlak van de molecule te verkennen.

In dit proefschrift introduceren wij de micro-macro Markov keten Monte Carlo (mM-MCMC) methode die het bemonsteren van de Gibbs verdeling versnelt in geval van een tijdschaalscheiding. De mM-MCMC methode genereert een nieuw monster in vijf stappen: 1) Bereken de macroscopische reactiecoordinaat van de huidige molecule; 2) Genereer een nieuwe reactiecoordinaat waarde; 3) Aanvaard of verwerp deze waarde op basis van een benaderde macroscopische verdeling; 4) Reconstructie naar het microscopisch niveau; 5) Aanvaard of verwerp de microscopisch monster op basis van de Gibbs-verdeling. We bestuderen twee soorten reconstructie, directe en indirecte. Indirecte reconstructie gebruikt een vertekende simulatie om het microscopische monster naar de reactiecoordinaat waarde te trekken, terwijl directe reconstructie gewoon een molecule trekt uit een reconstructieverdeling. De eerste is altijd toepasbaar, maar kostbaarder. De computationele winst vloeit voort uit het feit dat de meeste voorstellen op macroscopisch niveau worden verworpen, tegen lage computationele kost, terwijl microscopische monsters, eenmaal gereconstrueerd, altijd worden aanvaard. We tonen analytisch de convergentie aan en bespreken de convergentiesnelheid van het voorgestelde algoritme, en voeren numeriek een efficiëntiestudie uit op drie moleculaire voorbeelden, een drie-atomig molecuul, butaan en alanine-dipeptide.

De mM-MCMC-methode heeft verder nog een computationeel knelpunt. Men moet de vrije energie van de reactiecoordinaat kennen om de microscopische aanvaardingskans te evalueren. De vrije energie is duur om te berekenen, men

moet vertrouwen op een goedkope, on-the-fly benadering die statistisch neutraal is. De pseudo-marginale methode is een ideale oplossing voor dit probleem. De pseudo-marginale methode berekent een benadering van de vrije energie door middel van *importance sampling*, waarbij de monsters worden verkregen uit indirecte reconstructie. Men heeft slechts enkele monsters nodig om een onvertekende schatting te berekenen, wat de schatting goedkoop maakt. Wij illustreren de vertekening en variantie, als de prestaties van de pseudo-marginale methode op twee moleculaire voorbeelden, butaan en alkaan.

In het laatste deel van het proefschrift bestuderen we backmapping als alternatief voor indirecte reconstructie voor proteïnen. Voor macromoleculen zoals proteïnen wordt de reactiecoordinaat typisch bepaald door een lineaire combinatie van atomen voor elk van de herhalende eenheden, zoals het MARTINI macroscopisch krachtveld. Indirecte reconstructie lijdt aan de 'vloek van de dimensionaliteit' (*curse of dimensionality*) wanneer de reactiecoordinaat van hoge dimensie is. Backmapping lost dit probleem in twee stappen op. Eerst wordt een microscopisch molecuul gecreëerd dat consistent is met het macroscopische monster door een stap die 'lifting' wordt genoemd. Ten tweede genereert backmapping een tijdsafhankelijke simulatie waarbij de microscopische krachten langzaam worden ingeschakeld en de macroscopische krachten worden afgeschakeld, om de stabiliteit te handhaven. Wij presenteren de mM-MCMC methode met backmapping en tonen de prestaties ervan op een enkel disulfide peptide.

# List of Abbvreviations

QM         Quantum Mechanics
DFT        Density Functional Theory
MD         Molecular Dynamics
PES         Potential Energy Surface

MH         Metropolis-Hastings
DB         Detailed Balance
MALA       Metropolis-Adjusted Langevin Algorithm
MCMC      Markov Chain Monte Carlo
mM-MCMC    Micro-Macro Markov Chain Monte Carlo
DA         Delayed Acceptance Monte Carlo
HMC        Hamiltonian Monte Carlo
MLMCMC    Multilevel Markov Chain Monte Carlo
GHMC      Generalized Hamiltonian Monte Carlo
KMC        Kinetic Monte Carlo
AdResS      Adaptive Resolution Scheme
HMM       Heterogeneous Multiscale Method

SDE        Stochastic Differential Equation
PDE        Partial Differential Equation
EM         Euler-Maruyama
TV         Total Variation Norm
MSE        Mean Squared Error
KDE        Kernel Density Estimation
QSD        Quasi-Stationary Distribution

| | |
|---|---|
| FEC | Free Energy Computations |
| ABP | Adaptive Biasing Potential |
| ABF | Adaptive Biasing Force |
| TI | Thermodynamic Integration |
| BI | Boltzmann Inversion |
| | |
| DNA | Deoxyribonucleic Acid |
| 6KN2 | A Single Difulfide Peptide |
| GLY | Glycine |
| PHE | Phenylalanine |
| ARG | Arginine |
| SER | Serine |
| PRO | Proline |
| CYS | Cysteine |

# List of Symbols

$V$    Potential Energy Function
$\nabla V$    Forces or Gradient of Potential Energy
$t$   Time
$D$    The Number of Atoms in the Molecule
$m$    Dimensionality of the Reaction Coordinate
$x$    $3D$-Dimensional Microscopic Positions Vector
$X_t$    Position Random Variable at Time $t$
$y$    $3D$-Dimensional Microscopic Positions Vector
$z$    $m$ - Dimensional Macroscopic Reaction Coordinate Value or Vector
$Z_t$    Random Reaction Coordinate Vector at Time $t$
$W_t$    Brownian Motion
$\delta t$    Microscopic Time Step Size
$\Delta t$    Macroscopic Time Step Size


$T$    Temperature
$k_b$    Boltzmann Constant
$\beta$    Inverse Temperature $(k_b T)^{-1}$
$\varepsilon$    Time-Scale Separation Parameter


$N$    Number of Sampling Steps
$K$    Number of Samples Generated by Indirect Reconstruction
$\mu$    Time-Invariant Gibbs Distribution
$\mu_{\text{ext}}$    Extended Gibbs Distribution of Microscopic and Macroscopic Variables
$\xi$    Reaction Coordinate
$Z_V$    Normalizing Constant of the Microscopic Gibbs Distribution
$G(x)$    Gramm Matrix


$A(z)$    Free Energy at Reaction Coordinate Value $z$

$\mu_0$    Invariant Distribution of Reaction Coordinate Values
$Z_A$    Normalization Constant of $\mu_0$
$\bar{\mu}_0$    Approximate Invariant Distribution of Reaction Coordinate Values
$\lambda$    Strength of Biasing Potential in Indirect Reconstruction
$Q_\lambda(z)$    Approximate Free Energy at Reaction Coordinate Value $z$


$q_0$    Macroscopic Proposal Distribution
$\alpha_{CG}$    Macroscopic Acceptance Probability
$\alpha_f$    Microscopic Acceptance Probability
$\nu_\lambda$    Indirect Reconstruction Distribution

# Contents

# Chapter 1

# Introduction

## 1.1 Context & Motivation

Numerous systems in chemistry and statistical mechanics consist of a large number of microscopic interacting particles [61]. The three-dimensional positions of these particles are collected into the state vector $x \in \mathbb{R}^{3D}$, with $D$ the number of particles. The dynamics of such systems is typically governed by a potential energy function $V(x)$ and Brownian Motion $W_t$, for instance through the overdamped Langevin equation

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dW_t. \tag{1.1}$$

Here, $X_t$ represents the time-dependent positions of a single realisation of the dynamics, and $\beta = (k_B T)^{-1}$ is the inverse temperature with $k_B$ the Boltzmann constant. Examples of these systems include proteins, polymers and DNA.

In the field of molecular dynamics, one often wants to sample the time-invariant Boltzmann-Gibbs distribution

$$d\mu(x) = Z_V^{-1} \exp\left(-\beta V(x)\right) dx, \tag{1.2}$$

which is the likelihood of finding the molecule in state $x$. In notation above, $Z_V$ is the normalization constant of $\mu$, and $dx$ is the Lebesgue measure. Sampling the Gibbs distribution is usually achieved via Markov chain Monte Carlo (MCMC) methods, in which the proposal moves, for instance based on the dynamics (1.1), are supplemented with an accept-reject criterion. Markov chain Monte Carlo methods were introduced by Metropolis in 1949 [74] and later generalized by Hastings [36].

Standard MCMC methods face two computational limitations when sampling the Boltzmann-Gibbs distribution. When the target distribution and the Langevin process (1.1) are high-dimensional, generating a proposal requires a considerable computational effort because the acceptance probability is typically low. It has been shown in the past that the number of samples required to converge to $\mu$ grows with $3D$ [22]. Specifically for the MALA algorithm, the number of samples required grows as $(3D)^{1/3}$ [29]. Recently, however, it has been suggested Crank-Nicholson proposals can increase this acceptance rate [22]. We do not investigate this claim further in this manuscript.

Second, there often exists a large time-scale separation between the fast dynamics of the full, high-dimensional (microscopic) variables, and the slow behaviour of some suitable low-dimensional (macroscopic) variables. Then for stability, simulating the microscopic dynamics (1.1) requires taking time steps with size on the order of the fastest mode in the system, limiting the size of proposal moves and slowing down exploration of the full state space. In particular, when the potential $V$ has multiple local minima, standard MCMC methods remain stuck in one minimum for a long time, almost never transitioning to other minima. This phenomenon is called metastability. There exist several techniques to overcome metastability in this context, for instance the parallel replica dynamics [94, 96, 57, 67], adaptive multilevel splitting [17] and kinetic Monte Carlo [95].

## 1.2  Overview of the Literature

For molecular dynamics simulations, a lot of effort has been put in obtaining coarse-grained descriptions of the system in terms of a small number of slow degrees of freedom that capture the essential macroscopic features of the system. We use a coarse-graining technique based on reaction coordinates. A reaction coordinate is a smooth function from the high-dimensional microscopic configuration space $\mathbb{R}^{3D}$ to a lower dimensional space $\mathbb{R}^m$, with $m \ll 3D$. We denote this function as

$$\xi : \mathbb{R}^{3D} \to \mathbb{R}^m, \ x \mapsto \xi(x) = z. \tag{1.3}$$

Based on the underlying evolution of the microscopic variables, one can approximate the exact dynamics of the reaction coordinate values [58] by an overdamped Langevin dynamics of the form

$$dZ_t = -\nabla A(Z_t)dt + \sqrt{2\beta^{-1}}dW_t, \tag{1.4}$$

with $Z_t = \xi(X_t)$. The function $A$ is the free energy of the reaction coordinate, defined by

$$A(z) = -\beta^{-1} \ln \int_{\Sigma(z)} \mu(x)\delta_{\xi(x)-z}(dx). \qquad (1.5)$$

The time-scale separation problem is far from new. The past decades have seen many creative methods to accelerate the sampling in case of a time-scale separation. We give an overview. An important technique is the kinetic Monte Carlo method [95], where the macroscopic variables are basins of attraction around local minima of the potential energy. Transitioning between basins is achieved through discrete transition probabilities. Another example is the adaptive resolution technique [81] which models certain regions with the most accurate microscopic dynamics, while other regions can be simulated accurately with a macroscopic model. Furthermore, effort has been put in obtaining a macroscopic Brownian Motion from the microscopic model in the context of molecular dynamics [31], or simply averaging out all microscopic degrees of freedom for homogenization [79, 97]. We also like to mention the recently developed Multilevel Markov Chain Monte Carlo method. There are two instantiations that reduce the variance of MCMC by coupling levels. Dodwell and co-authors [26] couple two levels by using the delayed-acceptance method [20]. That is, particle proposals are generated on the coarser level, and then only accepted or rejected on the finer level. A second approach to Multilevel MCMC is to use the same random numbers for proposals on both levels [44]. It is unclear which scheme can lead to the largest time/variance reduction. We note both MLMCMC methods use coarse and finer models on the same state space. Finally, in the Coupled-MCMC method [45, 47], the microscopic system is a fine lattice with an Ising-type potential energy function. Sampling the lattice system is accelerated by introducing a lower dimensional, macroscopic lattice. The Coupled-MCMC algorithm first generates a macroscopic proposal on the coarse lattice, followed by an accept-reject step. Second, a microscopic particle is generated from the corresponding macroscopic proposal, which is subjected to a final accept-reject step.

## 1.3 Contributions

In a first contribution, we introduced the micro-macro Markov chain Monte Carlo method (mM-MCMC) [92], inspired by the Coupled-MCMC method. The mM-MCMC method aims to accelerate the sampling of the invariant Gibbs distribution (1.2) of high-dimensional stochastic processes (1.1) in the presence of a time-scale separation. The objective of the mM-MCMC method

is to overcome metastability and enhance sampling over standard microscopic MCMC methods.

The mM-MCMC method consists of three steps to generate a new microscopic sample from the Gibbs distribution: i) *Restriction*, i.e., computation of the reaction coordinate value associated to the current microscopic sample; ii) *Macroscopic Proposal*, sampling a new value of the reaction coordinate based on some macroscopic dynamics such as (1.4); iii) *Reconstruction*, creation of a microscopic sample based on the sampled reaction coordinate value. This scheme contains two accept-reject steps. First, step ii) tests the new reaction coordinate value with respect to an approximate macroscopic distribution. We proceed to reconstruction if accepted, otherwise we propose a new macroscopic sample. Second, we perform a second accept-reject step after reconstruction to make sure the microscopic particle is a sample from the Gibbs distribution.

The advantage of using a macroscopic dynamics to generate proposals crucially depends on the quality of the proposal moves on the macroscopic level, for two reasons. First, one needs to ensure the fastest modes are not present at the reaction coordinate level, such that larger proposals are possible compared to the microscopic level, enhancing exploration of the state space. Second, the macroscopic dynamics should be constructed such that most rejected proposals are already rejected at the macroscopic level, i.e., without ever having to perform a (often costly) reconstruction. In particular, the acceptance rate of he reconstructed microscopic samples should be close to 1. If the macroscopic dynamics used to generate reaction coordinate proposals is not an accurate approximation of the exact time-dependent evolution of the reaction coordinate values, more proposals will only be rejected after reconstruction at the microscopic level. This scenario leads to wasted computational efforts.

In many large and complex molecules, the reconstruction step from the macroscopic to microscopic levels is convoluted. Indeed, the level set of of all microscopic particles that have a certain reaction coordinate value is usually highly curved, depending on the form of the reaction coordinate. Sampling on such manifold is expensive [90, Chapter 3]. To counter this computational issue, we propose two types of reconstruction: *direct* and *indirect* reconstruction. The mM-MCMC method with *direct* reconstruction always generates a microscopic sample on the level set of constant reaction coordinate value. This variant has limited use and is only viable when the reaction coordinate is simple. We nonetheless introduce the method in the broader theoretical framework. The indirect reconstruction scheme generates microscopic samples close to, but not on the level set. The main computational advantage over direct reconstruction is that we do not require an expensive manifold sampler. The mM-MCMC method with indirect reconstruction proposes samples by 'pulling' the previous microscopic sample towards the new macroscopic reaction coordinate value using

a biased simulation. We thus make no assumption on the form of the reaction coordinate, making the mM-MCMC method with indirect reconstruction more general.

It is possible to prove convergence of the mM-MCMC method. Indeed, both methods satisfy detailed balance, and irreducibility and aperiodicity and these properties are easy to verify. For mM-MCMC with direct reconstruction, we can also prove that the type of ergodicity (uniform or geometric) depends on the type of ergodicity of the macroscopic sampler. The assumptions required for this proof are stronger, but the author feels confident that geometric ergodicity remains true under relaxed conditions.

The mM-MCMC method with both types of reconstruction can obtain significant efficiency gains over traditional microscopic samplers [91]. Indeed, we show the numerical efficiency gain of mM-MCMC over the microscopic MALA method [84] on three molecules: a three-atom molecule, butane and alanine-dipeptide. On the three-atom molecule and butane, we test both the mM-MCMC method with direct and indirect reconstruction. Because alanine-dipeptide is a complex molecule, we only test mM-MCMC with indirect reconstruction because direct reconstruction is difficult to implement and is molecule-dependent.

The mM-MCMC method(s) does not work well in the form presented above, there is a computational problem. The indirect reconstruction method requires the free energy (1.5) of the (extended) reaction coordinate to evaluate the microscopic acceptance probability. Knowing the free energy of the reaction coordinate is crucial for obtaining unbiased samples from (1.2). Any bias on the free energy estimates results in an error on the Gibbs distribution, and we have no means of measuring this bias. The problem is that the free energy is only analytically available for simple and small molecules, and needs either be pre-computed, or computed on-the-fly after each reconstruction step.

Free energy computation methods (FEC's) take a central place in the molecular dynamics literature. We give a summary of the methods that we have implemented. The adaptive biasing potential / force (ABP/ABF) are a general class of schemes [25, 21, 65] that compute a local approximation to the respective free energy, or its gradient. Then subtracting the potential/force from the microscopic dynamics creates a flat energy landscape in the direction of the reaction coordinate. Another class of methods simulates a non-equilibrium dynamics between two reaction coordinate values. The free energy difference between these two values can then be computed by the Jarzynski theorem [64, 66]. A third method is thermodynamic integration. This method writes the free energy difference as an integral of the free energy derivative. One then approximates this derivative using a Monte Carlo estimator, together with a quadrature rule for integration. The Monte Carlo estimator is usually

constrained to the manifold of given reaction coordinate value. Sampling is typically done using the Rattle scheme [3] or generalized Hamiltonian Monte Carlo (GHMC) [66].

In this thesis, we use the pseudo-marginal approach [4, 33, 2, 89] to approximate the free energy at every iteration of the mM-MCMC algorithm. The main idea behind the pseudo-marginal method is to generate a bunch of microscopic samples from a constrained distribution on (or near) the level set of the given reaction coordinate value, and to compute a free energy estimate using importance sampling with respect to the microscopic Gibbs distribution. Key is that the constrained distribution can be arbitrary, although its choice largely affects the efficiency of the pseudo-marginal method. The resulting method can make a cheap and on-the-fly estimation at every mM-MCMC step. The only downside is an added variance on the mM-MCMC method because of the importance sampling.

We show that the indirect reconstruction distribution is a good choice for our importance distribution. The microscopic samples generated during the biased simulation can be seen as samples from the indirect reconstruction distribution. So, mM-MCMC with indirect reconstruction and pseudo-marginal approximation is therefore only a little bit slower than mM-MCMC with indirect reconstruction. Only the actual free energy estimation step is new. In a practical implementation however, we will first make a histogram approximation of the indirect reconstruction distribution. We explain the technical reason behind this additional computation in chapter 6.

In our numerical experiments, we will show that indirect reconstruction and the pseudo-marginal approximation methods work well together for low-dimensional reaction coordinates. But they do not for high-dimensional reaction coordinates. The typical situation in macro-molecules such as proteins, DNA and polymers is that the macroscopic description does grow linearly with the number of microscopic atoms. An example of such reaction coordinate is the MARTINI force field for proteins [71, 76]. The MARTINI force field computes the centre of mass of on average four atoms. We then obtain a bead model that repeats every amino-acid.

As the final contribution of this thesis, we explore backmapping [52, 31, 54, 53] as an alternative to indirect reconstruction in the case of high-dimensional reaction coordinates. In this regime, indirect reconstruction is not efficient anymore. Picture a ball around the proposed macroscopic sample This ball contains a large part of the total probability of the indirect reconstruction distribution. The problem is that this ball becomes smaller and smaller in comparison with the ambient space. As a consequence, indirect reconstruction spends a lot of time bringing the previous microscopic sample towards this

ball, instead of actually sampling the indirect reconstruction distribution. We therefore need an alternative that is in a sense dimension independent.

Backmapping is in many ways similar to indirect reconstruction, but there are a few key differences. It starts by initializing a microscopic molecule consistent with the new macroscopic sample. This step is often called 'lifting' [87], and is efficient due to the structure of the MARTINI reaction coordinate. Afterwards, backmapping runs a time-dependent process. Because lifting is out of equilibrium, the time-dependent process initializes with a sole macroscopic (MARTINI) dynamics. As the simulation progresses, the microscopic dynamics is slowly switched on, while the macroscopic dynamics is slowly switched off and dies down. Backmapping is a non-equilibrium process because the total potential is time-dependent. As a consequence, we cannot define a (equilibrium) microscopic acceptance probability, and there is no use anymore for free energy calculations. The main cost of the method is therefore the (offline) construction of the MARTINI force field, but we will assume that this force field is given. We end this thesis by showing the mM-MCMC sampling result on a single disulfide peptide (6kn2), which contains ten amino acids.

This manuscript presents the main storyline of the research I worked on over the past five years, but it does not contain all. First, I worked on a micro-macro acceleration method for SDEs [101, 91] during my master's thesis and the first year of my PhD, and this method has applications in accelerating the simulation of moment models for kinetic equations [50]. Second, I worked on reaction path continuation methods for finding transition states in molecular dynamics with Ioannis Kevrekidis, and this work culminated in a paper [7] that has been accepted for publication.

## 1.4   Outline of the Thesis

In chapter 2, we introduce the broad field of molecular dynamics. We start with essential mathematical notions such as Brownian Motion, the Fokker-Plank equation and SDE's in section 2.1, followed by an atomistic description of molecular dynamics in section 2.2. Afterwards, we derive the Langevin equations from Newtonian mechanics in section 2.3, and we end the chapter with a description of the macroscopic dynamics of reaction coordinates, section 2.4.

In chapter 3, we introduce Markov chain Monte Carlo methods. We start by describing the original algorithm by Metropolis and Hastings in section 3.1. Then, in section 3.2, we go over all requirements needed to prove ergodicity of any MCMC method. These are invariance (detailed balance), irreducibility and aperiodicity. We also prove the Metropolis-Hastings method satisfies these

requirements. Afterwards, we present three well-known MCMC methods in section 3.3, and we end the chapter by laying out a quantitative argument why standard MCMC methods are ill-suited for sampling systems with a time-scale separation in section 3.4.

In chapter 4 we introduce the mM-MCMC method. We explain mM-MCMC with direct reconstruction in section 4.1, and the mM-MCMC method with indirect reconstruction in section 4.2. We also prove stronger ergodicity of mM-MCMC with direct reconstruction in section 4.3.

Then, in chapter 5, we run the mM-MCMC methods on three molecular test cases. We first define the efficiency gain criterion in section 5.1. Afterwards, we implement the mM-MCMC method for: a three-atom molecule (section 5.2), butane (section 5.3), and alanine-dipeptide (section 5.4).

In chapter 6, we present the pseudo-marginal approximation method. In section 6.1, we go through three free energy computation methods in our context. Afterwards, in section 6.2, we explain the pseudo-marginal method in general terms. Next, we combine the pseudo-marginal and indirect reconstruction schemes in section 6.3. Finally, we show the performance of the mM-MCMC method with pseudo-marginal approximation in section 6.4.

Finally, we explain backmapping as an alternative to indirect reconstruction in chapter 7. We begin by introducing the MARTINI reaction coordinate and its macroscopic force field in section 7.1. Next, we describe the backmapping method as a time-dependent simulation in section 7.2. In section 7.3, we build the micro-macro MCMC method with backmapping. Finally, we show the added performance of the mM-MCMC method on a protein in section 7.4.

In chapter 8 we present general conclusions regarding the methods and results from this manuscript chapter by chapter. We also discuss where and how these methods can be successfully applied. We end the chapter by presenting a couple of ideas for future research.

# Chapter 2

# Introduction to Molecular Dynamics

The field of Molecular Dynamics (MD) broadly describes the motion of molecules and chemical reactions within a computational approach. In theory, quantum dynamics is fully able to explain the motion and chemical reactivity of molecules. A combination of Density Functional Theory (DFT) for nuclei and the Schrödinger equation for electrons are the basis for the Born-Oppenheimer and Car-Parrinello 'ab initio' approaches to molecular dynamics. We explain both approximations in section 2.2, after an introduction to SDEs and the Itô formalism in section 2.1.

In practice, the 'ab initio' approach to molecular dynamics is expensive for mid- to large molecules. Instead, we use an empiric interaction potential from a purely atomistic representation of the molecule. This potential energy includes all microscopic interactions such as vibrations of bonds and angles, as well as long-scale forces such as electrostatic and Vanderwaals interactions. We then integrate the resulting gradient dynamics over time. This is the (underdamped) Langevin approach [56]. We describe Langevin dynamics in section 2.3.

When the system under study is high-dimensional, simulating the full Langevin equation becomes computationally prohibitive. A macroscopic, low-dimensional representation of the molecular system that maintains the essential dynamics may speed-up the simulation. We do lose exactness because most non-essential microscopic components are disregarded. In this manuscript, we use the concept of reaction coordinates for coarse-graining. A reaction coordinate maps a microscopic molecule to its macroscopic representation, it is a local map. One

can then derive a Langevin-like effective dynamics for the macroscopic variables. This effective dynamics also has an invariant distribution, called the free energy'. We introduce reaction coordinates, the effective dynamics, and the free energy in section 2.4. The contents of this chapter do not appear in a separate paper, but are included to make the manuscript self contained.

## 2.1 Preliminaries

Stochastic Differential Equations (SDEs) are the language of molecular dynamics. They model collisions with and random motion of molecules in a fluid. In this section, we introduce SDE's and related concepts that we will use later in the manuscript. Our intention is not to go to the fundamentals of probability theory, but rather to establish important notions and concepts useful for molecular dynamics. The content of this section is largely taken from the books [14, 23].

Practically, an SDE is a mathematical expression of the form

$$dX_t = a(X_t)dt + b(X_t)dW(t). \tag{2.1}$$

Here, $X_t$ is an (unknown) time-dependent random variable. Further, $a(X)dt$ is the deterministic drift term with drift coefficient $a(X)$. Similarly, $b(X)dW(t)$ is the stochastic diffusion term, and $b(X)$ the diffusion coefficient. We assume that $a(X)$ is a vector and $b(X)$ a square matrix.

The remaining term, $dW(t)$, is Brownian motion. It is a fundamental stochastic process that is defined by four properties. These properties are

- W(0) = 0,

- W(t) is time-continuous with unit probability,

- The increments $W(t+s) - W(t)$ are normally distributed with zero mean and variance $s$, and

- The process $\{W(t)\}_{t \geq 0}$ has independent increments.

It can be shown that the only process that satisfies these requirements is $W(t) = \mathcal{N}(0, t)$. That is, Brownian motion is normally distributed with zero mean and variance $t$.

Finally, it should be noted that SDE (2.1) only makes sense when integrating over time. The notation used in (2.1) is only a proxy. In other words, the

solution $X(t)$ to the SDE at time $t$ is

$$X_t = X_0 + \int_0^t a(X_s)ds + \int_0^t b(X_s)dW(s), \qquad (2.2)$$

for a suitable initial condition $X_0$. The first term is a simple, deterministic integral in the sense of Lebesgue. It resembles the solution of an ordinary differential equation. The second integral in $dW(s)$ only makes sense in the context of an Itô integral. We refer the reader to [23, 14] for more information concerning SDEs and the Itô formalism.

### 2.1.1  The Fokker-Planck Equation

Instead of solving the SDE (2.1) by integrating for $X_t$ (2.2), we can solve for the probability distribution of $X_t$. Let $p(x,t)$ be the probability distribution of $X_t$ at time $t$. From (2.1), we can construct a partial differential equation that allows $p(x,t)$ as unique solution. This PDE is called the 'Fokker-Planck' equation [14] and reads

$$\frac{\partial}{\partial t}p(x,t) = -\sum_{i=1}^d \frac{\partial}{\partial x_i}\left(a_i(x,t)p(x,t)\right) + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j}\left(b_{ij}(x,t)p(x,t)\right). \quad (2.3)$$

The initial condition can by any regular probability distribution $p(x,0)$. When the initial condition $X_0$ to (2.1) is a fixed value, $p(x,0)$ is a delta function centred at $X_0$.

When the state space is high-dimensional, such as $\mathbb{R}^d$ with $d > 3$, solving the Fokker-Planck equation directly is prohibitively expensive. Monte-Carlo methods scale better with the number of dimensions. We will therefore not use the Fokker-Planck equation directly, only Monte Carlo approximations based on (2.2).

### 2.1.2  The Euler-Maruyama Method

To complete this section, we look at a numerical discretization of the SDE (2.1). Suppose we want to integrate the SDE up to time $T > 0$ with a time step of size $\Delta t$, such that $N\Delta t = T$. We can numerically solve the SDE by discretizing (2.2) using a first order method. Generally, we can approximate integrals in (2.2) using a forward scheme

$$\int_t^{t+\Delta t} f(s)ds \approx f(t)\Delta t,$$

for any function $f$.

Suppose $X_n$ is the solution to (2.2) at time $n\Delta t$. The Euler-Maruyama method computes $X_{n+1}$ at time $(n+1)\Delta t$ as

$$X_{n+1} = X_n + a(X_n)\Delta t + b(X_n)\left[W\left((n+1)\Delta t\right) - W\left(n\Delta t\right)\right]. \qquad (2.4)$$

The first part is the regular Euler method for ODEs, the second part requires sampling the Brownian motion at two points in time. Remembering the third property of Brownian motion, the increments are normally distributed, i.e.,

$$W\left((n+1)\Delta t\right) - W\left(n\Delta t\right) \sim \mathcal{N}(0, \Delta t).$$

Using this property, we can simplify (2.4) to

$$X_{n+1} = X_n + a(X_n)\Delta t + \sqrt{\Delta t}\, b(X_n)\, \eta_n, \ \eta_n \sim \mathcal{N}(0, 1). \qquad (2.5)$$

The random number $\eta_n$ are i.i.d. We use the square root of $\Delta t$ in front of the diffusion term because the standard deviation of the Brownian increments scales with $\sqrt{\Delta t}$.

It can be shown that the Euler-Maruyama converges weakly with order 1, and strongly with order 1/2 [98].

## 2.2 Atomistic Description of Molecular Dynamics

Molecules are made of atoms, and atoms of electrons and nuclei. Quantum Mechanics (QM) is the physics that describes the small. It governs the motion and properties of electrons and nuclei, specifically through the Schrödinger equation. The quantum mechanical approach should be able to fully describe the creation and breaking of bonds, interactions with the fluid and the minimum energy states. We explain the Born-Oppenheimer [90] and associated Car-Parrinello [15] ab initio approaches to quantum molecular dynamics in section 2.2.1

In practice, computational ab initio methods can only be used for a small number of atoms. Empirical formulas for the potential energy are used for larger system. These empirical interaction potentials are obtained by assuming a functional form, and bringing this form as close to the quantum mechanical description as possible by optimizing certain parameters. We explain this approach for fluids in section 2.2.2 and for large molecules in section 2.2.3

## 2.2.1 Ab Initio Quantum Potentials

Technically, the potential energy of interactions between nuclei and electrons should be determined from a quantum mechanical, 'ab initio' approach. Suppose $x_1, x_2, \ldots, x_N$ are the point positions of the nuclei, and that $Z_1, Z_2, \ldots, Z_n$ are their respective. The total, non-deficient, number of electrons is $M = Z_1 + Z_2 + \cdots + Z_N$. Using the Born-Oppenheimer assumption, the positions of the nuclei are fixed, and their potential energy is the total of Coulomb interactions between nuclei and the ground state energy of electrons. In symbols,

$$V(x_1, x_2, \ldots, x_N) = \sum_{1 \leq i \leq j \leq N} \frac{Z_i Z_j}{\|x_i - x_j\|} + V_{\text{ground}}(x_1, x_2, \ldots, x_N). \qquad (2.6)$$

The electron ground state energy is only a function of the positions of the nuclei, but is hard to compute. It is obtained by minimizing the electron Hamiltonian over all admissible wave functions, i.e.,

$$V_{\text{ground}}(x_1, x_2, \ldots, x_N) = \inf \ \{ \ \langle \psi | H_{x_1, \ldots x_N} | \psi \rangle \ | \ \|\psi\|_2 = 1 \ \},$$

where the electron Hamiltonian is

$$H_{x_1, \ldots x_N}(q_1, \ldots q_M) = -\frac{1}{2} \sum_{m=1}^{M} \Delta q_m - \sum_{m=1}^{M} \sum_{j=1}^{N} \frac{Z_i}{\|q_m - x_j\|} + \sum_{1 \leq n \leq m \leq M} \frac{1}{\|q_n - q_m\|}.$$

Here, $q_1, \ldots q_M$ are the electron positions, and $\Delta q_m$ is the fundamental Heisenberg uncertainty on those positions. Note that we have omitted many quantum effects for clarity of presentation, but these effects are important for actual computations.

The Born-Oppenheimer approach has two major problems [12, 42]. On the one hand, we need to integrate the equations of motion for the nuclei, which happens over long time scales. On the other hand, we need to compute the electron ground state at every time step. Computing this ground state is inherently expensive [12]. An alternative to avoid this computationally expensive problem is the Car-Parrinello approach. The Car-Parrinello way to molecular dynamics uses a fictitious dynamics that keeps the electrons close to their ground state. It overcomes the need for an expensive optimization routine. The fictitious dynamics relies on a small electron mass $\mu$ to minimize interactions with nuclei. The Car-Parrinello equations of motion then are

$$M_i \ddot{x}_i = -\nabla_{x_i} E\left[\{\Psi_k\}, \{x_k\}\right]$$

$$\mu \dot{\Psi}_i(x, t) = -\frac{\delta E}{\delta \Psi_i^*(x, t)} + \sum_{j=1}^{N} \lambda_{i,j} \Psi_j(x, t).$$

In the equations above, $\{\Psi_i\}$ are orthogonal orbital wave functions and the matrix $\lambda$ are the Lagrange multipliers. Additionally, $E$ is the Kohn-Sham energy density function from Density Functional Theory (DFT) [51]. We do not give further details.

### 2.2.2  Empiric Interaction Potentials for Fluids

In practical simulations, empiric potentials are necessary for larger molecular systems. These empiric interaction potentials can be calculated by assuming a simple form with parameters. These parameters are then optimized so that the interaction potential is close to (2.6), using ab initio computations. Functional parameters can also be determined to match known properties of the molecule or measurements.

In this section, we assume a fluid with $N$ interacting particles. The particles are identical and there are only pairwise interactions. Call $r_{i,j} = \|x_i - x_j\|$ the distance between particles $i$ and $j$. The total potential energy then is

$$V(x_1, x_2, \ldots, x_N) = \sum_{1 \le i < j \le N} V(r_{i,j}).$$

A very simple and much used example of an interaction potential is the Lennard-Jones potential [68],

$$V_{LJ}(r) = 4\varepsilon \left( \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right). \tag{2.7}$$

This potential depends on two parameters, the (negated) ground state energy $\varepsilon$ and a distance $\sigma$. The Lennard-Jones function models the Vanderwaals interactions between two atoms. There is a slight attraction when the atoms are far apart, and a strong repulsion when the atoms are closer to each other than the equilibrium distance of $2^{1/6}\sigma$. The Lennard-Jones model is often used to model mono-atomic gases, and especially the noble gases [90].

### 2.2.3  Empiric Interaction Potentials for Large Molecules

In all of organic chemistry, large molecules are built up from covalent bonds between atoms. More descriptive empiric potential functions are needed to model the creation and breaking of bonds in larger systems, as well as interactions involving three and four atoms [80, 90]. We explain these additional empiric potentials below.

1. Covalent bonds between atoms can be modelled by a quadratic or harmonic potential function

$$V_{\text{bond}}(x_i, x_{i+1}) = \frac{k_r}{2} (r_{i,i+1} - r_0)^2, \tag{2.8}$$

   to model the vibrations around the equilibrium bond length $r_0$. The larger $k_r$, the deeper the potential well around $r_0$ and the more energy required to break the covalent bond.

2. Three atoms connected by two bonds on a central atom can similarly be modelled by a three term harmonic potential concerning the angle $\theta$ between the atoms

$$V_{\text{angle}}(x_{i-1}, x_i, x_{i+1}) = \frac{k_\theta}{2} (\theta_i - \theta_0)^2$$

   where the equilibrium angle is $\theta_0$. The double-bond angle is

$$\theta_i = \arccos \left( \frac{x_{i-1} - x_i}{r_{i-1,i}} \cdot \frac{x_{i+1} - x_i}{r_{i+1,i}} \right)$$

3. Four connected atoms typically experience a four-body interaction potential through their torsion angle

$$V_{\text{torsion}}(x_i, x_{i+1}, x_{i+2}, x_{i+3}) = f(\tau_i).$$

   There are two kinds of torsion angles. A regular dihedral angle connects four atoms in a row and this torsion angle $\tau_i$ is the angle between the two planes determined by $x_i, x_{i+1}, x_{i+2}$ and $x_{i+1}, x_{i+2}, x_{i+3}$. An improper torsion angle is the 'non-planar' angle between three atoms connected to a central atom.

Besides interactions coming from covalent bonds, there are non-bonded interactions. Electrostatic forces are derived from the electrostatic potential (2.6), and Vanderwaals interactions are typically modelled by the Lennard-Jones empiric potential (2.7).

## 2.3   From Newton to Langevin

A frequently used model for the motion of molecules is the (underdamped) Langevin equation [56]. The Langevin equations are a phenomenological model, meaning they are mainly based on measurements. The Langevin model contains

both deterministic drift and stochastic diffusion terms. It is a stochastic differential equation for position and momentum of the molecule.

The (deterministic) drift term of the Langevin equations can be derived from Newton's second law [90]. We denote with $x \in \mathbb{R}^{3D}$ and $p \in \mathbb{R}^{3D}$ the respective positions and momenta of the molecule. Here, $D$ is the number of atoms, so $x$ and $p$ live in $3D$ dimensional real space. We also let $M$ be the diagonal mass matrix of all atoms. From Newton's third law, the force $F(x)$ at each point is the change in momentum, i.e.,

$$F(x) = \dot{p}.$$

In molecular dynamics, the force field $F$ is typically conservative. Then, there exists a potential energy function $V(x)$ so that

$$F(x) = -\nabla V(x).$$

The nabla operator $\nabla$ denotes the spatial gradient. Putting these equations together, we have

$$\dot{p} = -\nabla V(x).$$

The full position and momentum equations then read

$$\begin{cases} \dot{x} & = M^{-1}p, \\ \dot{p} & = -\nabla V(x), \end{cases} \tag{2.9}$$

with suitable initial conditions. Note that these equations perfectly coincide with (3.15) for Hamiltonian $H(x, p) = V(x) + \frac{1}{2}p^T M^{-1}p$.

Equations (2.9) are not the full Langevin model, however. In practice, we need to add a friction and Brownian term [61, 62]. The friction and Brownian terms both represent the same physical phenomenon, namely that of random collisions with the molecules in the surrounding fluid, like water. Together with these two terms, the full phenomenological Langevin model reads

$$\begin{cases} dX & = M^{-1}Pdt, \\ dP & = -\nabla V(X)dt - \gamma Pdt + \sqrt{2\gamma\beta^{-1}M}dW. \end{cases} \tag{2.10}$$

Here, $-\gamma P$ is the friction term with friction constant $\gamma$, while $\sqrt{2\gamma\beta^{-1}M}dW$ is the Brownian term. The random variable $dW$ is Brownian Motion given by the Wiener process. We use capital letters $X$ and $P$ to denote position and momentum because they are random variables. One can then numerically integrate (2.10) over many paths to understand the behaviour of the molecular system, and a lot of the molecular dynamics literature does just that [61].

There is an interesting limit case of (2.10) when the friction constant $\gamma$ increases to infinity. In this limit, the momentum equation is absorbed by the position equation. This reduced model is the overdamped Langevin dynamics, and reads

$$dX = -\nabla V(X)dt + \sqrt{2\beta^{-1}}dW. \qquad (2.11)$$

This SDE models a gradient descent towards a local minimum of $V$, together with small random nudges that help the molecule cross energy barriers. The precise combination of the two forces depends on their size. If $V$ is steep, it is unlikely that the molecule escapes a local minimum. Otherwise, if $V$ is almost flat, the molecule can explore the whole state space because equation (2.11) becomes pure Brownian Motion. The latter coincides with a high temperature, or small $\beta$. We will say more about this problem in the context of multiscale molecular dynamics in section 3.4.

A key property of the overdamped Langevin dynamics is that its unique invariant distribution is precisely the Gibbs distribution (1.2) [61]. For this reason, the overdamped Langevin dynamics is frequently used to generate samples of $\mu$. Indeed, we can expect that simulating (2.11) with a small time step will generate paths that follow $\mu$ closely after a long simulation time. This is precisely the idea behind the Metropolis Adjusted Langevin Algorithm (MALA) that we will introduce in section 3.3.1.

## 2.4 Reaction Coordinates, Free Energy and Effective Dynamics

When the molecule under study is high-dimensional, there are typically only a few degrees of freedom that determine the overall conformation of the molecule. The other atoms only describe the remaining, less important microscopic conformation. It is common in the molecular dynamics literature to coalesce these important variables into a reaction coordinate [90]. Reaction coordinates are a local means of coarse-graining, it only uses the current microscopic molecule. Examples are diverse. In small molecules, the reaction coordinates are often torsion angles between sub-groups. In large, mesoscale molecules such as proteins, DNA or polymers, the reaction coordinate is typically the centre of mass of each composing unit. An important example is MARTINI for proteins [71, 76, 70] see chapter 7.

Technically, a reaction coordinate is a twice continuously differentiable function $\xi$. The reaction coordinate function maps microscopic atoms $x$ onto the macroscopic

variables $z$,

$$\xi : \mathbb{R}^{3D} \to \mathbb{R}^m,$$

$$x \mapsto \xi(x) = z. \tag{2.12}$$

We will use the letter $z$ to indicate a reaction coordinate value throughout the manuscript. We also use the notation $\Sigma(z) = \{x \mid \xi(x) = z\}$ for the level set of $\xi$ at value $z$. The remainder of this section is largely based on [90].

## 2.4.1 The Free Energy

An important identity relating functions over the microscopic state space $\mathbb{R}^{3D}$ to level sets of $\xi$ is the co-area formula [90]. For any function $f \in L^1(\mathbb{R}^{3D})$ we have the relation

$$\int_{\mathbb{R}^{3D}} f(x)dx = \int_{\mathbb{R}^m} \int_{\Sigma(z)} f(x) \left(\det G(x)\right)^{-1/2} d\sigma_{\Sigma(z)}(x) \ dz. \tag{2.13}$$

In the equation above, $G(x)$ is the Gram matrix,

$$G(x) = \nabla \xi(x)^T \nabla \xi(x) \ \in \ \mathbb{R}^{m \times m}.$$

This matrix is positive definite, so $\left(\det G(x)\right)^{-1/2}$ is always positive. Additionally, $d\sigma_{\Sigma(z)}$ is the Lebesgue measure on $\Sigma(z)$ induced by the ambient space.

Starting from the microscopic time-invariant distribution (1.2), one can define the macroscopic invariant distribution of reaction coordinate values, $\mu_0(z)$. Plugging (1.2) into the co-area formula (2.13), we find

$$\int_{\mathbb{R}^{3D}} \mu(x)dx = 1 = \int_{\mathbb{R}^m} \int_{\Sigma(z)} \mu(x) \left(\det G(x)\right)^{-1/2} d\sigma_{\Sigma(z)} \ dz. \tag{2.14}$$

The integral over $\mathbb{R}^m$ integrates to one, meaning whatever is inside this integral must be the invariant distribution of reaction coordinate values. Therefore,

$$\mu_0(z) = \int_{\Sigma(z)} \mu(x) \left(\det G(x)\right)^{-1/2} d\sigma_{\Sigma(z)(x)}. \tag{2.15}$$

Note $\mu_0$ is essentially the integral of the microscopic Gibbs distribution, multiplied by an additional reaction coordinate factor. Because of this form, we can define a macroscopic potential energy. This 'free energy' is the negative log of $\mu_0$,

$$A(z) = -\frac{1}{\beta} \ln \left( \int_{\Sigma(z)} \mu(x) \left(\det G(x)\right)^{-1/2} d\sigma_{\Sigma(z)} \right), \tag{2.16}$$

and is defined up to an additive constant. With the free energy, the macroscopic distribution is of Gibbs form,

$$\mu_0(z) = Z_A^{-1} \exp\left(-\beta A(z)\right), \tag{2.17}$$

and $Z_A$ is the normalization constant. We will use this form of the macroscopic time-invariant distribution for the remainder of this manuscript.

### 2.4.2 Reconstruction Distribution

A second probability measure can be associated to the microscopic Gibbs distribution using the co-area formula. The 'reconstruction' distribution $\nu(\cdot|z)$ is the Gibbs distribution conditioned on a fixed reaction coordinate value $z$. It can be defined by the conditioning formula [58],

$$\int_{\mathbb{R}^{3D}} f(\xi(x))g(x) \, d\mu(x) = \int_{\mathbb{R}^m} f(z) \left( \int_{\Sigma(z)} g(x) d\nu(x|z) \right) d\mu_0(z). \tag{2.18}$$

That is, the unique distribution for which this formula holds for all measurable $f$ and $g$ is the reconstruction distribution $\nu(dx|z)$.

Using the co-area formula, we now prove that the reconstruction distribution has the form

$$\nu(\cdot|z) = \frac{\mu \left(\det G\right)^{-1/2} d\sigma_{\Sigma(z)}}{\int_{\Sigma(z)} \mu \left(\det G\right)^{-1/2} d\sigma_{\Sigma(z)}}. \tag{2.19}$$

Indeed, for any two measurable functions $f$ and $g$,

$$\int_{\mathbb{R}^{3D}} f(\xi(x))g(x)d\mu(x)$$

$$= \int_{\mathbb{R}^m} f(z) \left( \int_{\Sigma(z)} g\mu \left(\det G\right)^{-1/2} d\sigma_{\Sigma(z)} \right) \, dz$$

$$= \int_{\mathbb{R}^m} f(z) \left( \frac{\int_{\Sigma(z)} g\mu \left(\det G\right)^{-1/2} d\sigma_{\Sigma(z)}}{\int_{\Sigma(z)} \mu \left(\det G\right)^{-1/2} d\sigma_{\Sigma(z)}} \right) \int_{\Sigma(z)} \mu \left(\det G\right)^{-1/2} d\sigma_{\Sigma(z)} \, dz$$

$$= \int_{\mathbb{R}^m} f(z) \left( \frac{\int_{\Sigma(z)} g\mu \left(\det G\right)^{-1/2} d\sigma_{\Sigma(z)}}{\int_{\Sigma(z)} \mu \left(\det G\right)^{-1/2} d\sigma_{\Sigma(z)}} \right) d\mu_0(z).$$

The final equation aligns perfectly with (2.18). Indeed, the ratio in the big parenthesis must be an expectation of $g$ with respect to the conditional distribution, proving (2.19).

### 2.4.3   Effective Dynamics

From the microscopic molecular dynamics (6.6), we can write down an effective dynamics for the associated reaction coordinate values. There are two variants: a time dependent and time invariant effective dynamics. This section is largely based on the paper by Legoll and Lelièvre [58].

Suppose $X_t$ is the solution to the microscopic dynamics (6.6) at time $t$, and that $\mu_t$ is its distribution. From a simple application of Itô's lemma, we find

$$d\xi(X_t) = \left(-\nabla V \cdot \nabla \xi + \beta^{-1}\Delta\xi\right)(X_t)dt + \sqrt{2\beta^{-1}}\,\nabla\xi(X_t) \cdot dW(t).$$

In notation above, $\Delta\xi$ is the Laplacian of the reaction coordinate, the sum of all second derivatives. Clearly, this equation is not closed, it depends on $X_t$.

A simple closing procedure by [58, 35] is to consider a different SDE in $\xi$ only, with solution $z_t$,

$$dz_t = b(z_t, t)dt + \sqrt{2\beta^{-1}}\sigma(z_t, t)dW(t). \tag{2.20}$$

The Brownian motion $dW(t)$ is now one-dimensional. The coefficients of this SDE are expectations conditioned on $\xi(X_t) = z_t$,

$$b(z_t, t) = \mathbb{E}\left[\left(-\nabla V \cdot \nabla \xi + \beta^{-1}\Delta\xi\right)(X_t) \mid \xi(X_t) = z_t\right]$$

and

$$\sigma^2(z_t, t) = \mathbb{E}\left[\|\nabla\xi\|^2(X_t) \mid \xi(X_t) = z_t\right].$$

The coefficients $b(z_t, t)$ and $\sigma^2(z_t, t)$ still depend $t$ because the distribution $\mu_t$ of $X_t$ evolves in time. It can be shown that the invariant distribution of (2.20) is precisely $\mu_0$ (2.15).

The problem with the closure (2.20) is that the coefficients $b$ and $\sigma^2$ are expensive to compute. Indeed, we need to simulate the microscopic process (2.11) to know the distribution $\mu_t$, to only afterwards approximate these coefficients. Therefore, the time dependent closure is completely ineffective. A natural approach is to consider a time independent closure. That is, the time independent coefficients are

$$b(z) = \mathbb{E}\left[\left(-\nabla V \cdot \nabla \xi + \beta^{-1}\Delta\xi\right)(X) \mid \xi(X) = z\right]$$

$$= \int_{\Sigma(z)} \left(-\nabla V(x) \cdot \nabla\xi(x) + \beta^{-1}\Delta\xi(x)\right) d\nu(x|z)$$

and

$$\sigma^2(z) = \mathbb{E}\left[\|\nabla\xi\|^2(X) \mid \xi(X) = z\right]$$

$$= \int_{\Sigma(z)} \|\nabla\xi(x)\|^2 \, d\nu(x|z)$$

and the new coarse-grained SDE reads

$$dz = b(z)dt + \sqrt{2\beta^{-1}}\sigma(z)dW(t). \tag{2.21}$$

This SDE has $\mu_0$ (2.15) as its unique invariant distribution, retaining an important property from (2.20).

# Chapter 3

# Markov Chain Monte Carlo Methods

In this chapter, we are concerned with sampling from the Gibbs distribution $\mu$ (1.2) that emerges from molecular dynamics. One widely studied tool for sampling such distributions is Markov chain Monte Carlo (MCMC). Markov chain Monte Carlo methods fundamentally rely on the property of ergodicity to generate samples. Suppose one can construct a stochastic process $X(t)$ that has a unique invariant distribution $\mu(X)$. With ergodicity, one can interchange averages over the invariant distribution with averages over one sample path $\{x_n\}_{n=1}^{N}$, i.e.,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} g(x_n) = \int_{\mathbb{R}^d} g(x)d\mu(x),$$

for any $g \in L^1(\mathbb{R}^d, \mu)$. Therefore, Markov chain Monte Carlo (MCMC) methods reduce the problem of sampling $\mu$ to constructing an ergodic stochastic process that keeps $\mu$ uniquely invariant.

We start this chapter by introducing the original Metropolis-Hastings method in section 3.1. Afterwards, we go deeper into the theory of Markov chains in section 3.2. We particularly introduce the concepts that are used to prove invariance and ergodicity of Markov chains. Then, in section 3.3, we introduce three well-known MCMC algorithms, explain how they work, and go through advantages and disadvantages. Finally, we discuss the trade-off between exploration and exploitation and how it affects mixing in Markov chain Monte Carlo.

Finally, we turn to molecular systems with an inherent time-scale separation between the slow, macroscopic and fast, microscopic variables. These systems are the main object of this thesis. Sampling the Gibbs distribution of a molecular system with a time-scale separation is inherently hard. Indeed, on the one hand, we must generate proposals with small increments for stability of the fast modes, but on the other hand, we need many samples to fully explore the state space of the slow modes. The combination of both constraints makes for a slow and inefficient sampling algorithm. We further explain this challenge and put it into the context of 'Exploration vs. Exploitation' in section 3.4. The contents of this chapter do not appear in a separate paper, but are included to make the manuscript self contained.

## 3.1   Original Algorithm by Metropolis and Hastings

Metropolis [74, 73] introduced the first Markov chain Monte Carlo method in 1953. He originally used it to sample the kinetic Boltzmann distribution with symmetric proposal moves. Suppose $q(x'|x_n)$ is the proposal distribution that brings the current sample $x_n$ on the Markov chain to the next proposal move $x'$. If $q$ is symmetric, one can interchange $x_n$ and $x'$ and the probability remains the same, i.e., $q(x_n|x') = q(x'|x_n)$. An example of symmetric proposal moves is Brownian motion, see Example 1.

Metropolis noticed that if we accept $x'$ with probability

$$\alpha\left(x'|x_n\right) = \min\left\{1, \frac{\mu(x')}{\mu(x_n)}\right\} \tag{3.1}$$

the samples $\{x_n\}_{n=1}^{N}$ indeed sample the invariant distribution $\mu$. That is, we set $x_{n+1} = x'$ in case of acceptance, and keep $x_{n+1} = x_n$ if rejected. In code, we check acceptance by generating a random uniform number $u$ between 0 and 1. We accept $x'$ if $u \leq \alpha\left(x'|x_n\right)$, and we reject it otherwise.

The acceptance probability only depends on the energy difference between the old and new state. We can rewrite $\alpha\left(x'|x_n\right)$ as

$$\alpha\left(x'|x_n\right) = \min\left\{1, \exp\left(-\beta\left(V(x') - V(x_n)\right)\right)\right\}.$$

The Metropolis acceptance probability fundamentally compares the energies between the two samples. If we decrease in energy from $x_n$ to $x'$, $x'$ is always accepted. This result is reasonable as all physical systems tend to find a local state of minimal energy. However, if we increase in energy, we accept $x'$ with a certain probability different from one. Going to higher energy states helps exploring of the whole state space because there may be additional local minima.

**Example 1.** *Proof.* A well-known symmetric proposal move is Brownian Motion (BM) [14]. Brownian Motion is based on a pure-diffusion SDE

$$dX = dW,$$

where $dW$ is the Wiener process. We can then discretize the above process with time step $\Delta t$

$$x' = x_n + \sqrt{\Delta t}\eta_n,$$

where $\eta_n$ is a standard normally distributed sample, see section 2.1. The proposal distribution is then given by

$$q\left(x'|x_n\right) = \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{(x'-x_n)^2}{2\Delta t}\right). \tag{3.2}$$

$\square$

Hastings [36], in 1970, extended the work by Metropolis by allowing non-symmetrical proposal distributions. He noticed if the acceptance probability changes to

$$\alpha\left(x'|x_n\right) = \min\left\{1, \frac{\mu(x')\ q\left(x_n|x'\right)}{\mu(x_n)\ q\left(x'|x_n\right)}\right\}, \tag{3.3}$$

the particles $\{x_n\}_{n=1}^N$ still sample $\mu$ exactly. One can use any proposal distribution for the Markov chain, as long as $\alpha\left(x'|x_n\right)$ can be evaluated. The process of checking acceptance is the same as in the Metropolis case.

## 3.2   Invariance, Irreducibility and Aperiodicity

How can we prove that the Metropolis-Hastings method actually 'converges' to the distribution $\mu$? From the book by Meyn & Tweedie [75], a Markov chain converges to $\mu$ when it is positive and Harris recurrent. Verifying these sufficient conditions is hard and problem specific. Fortunately, one can boil these conditions down to requirements that are easier to check. These conditions are invariance (detailed balance), irreducibility and aperiodicity [83]. We go through each of these properties and show that the Metropolis-Hastings method satisfies them all.

### 3.2.1   Invariance & Detailed Balance

Invariance of $\mu$ under the Metropolis-Hastings algorithm means that the method will eventually converge to $\mu$ in distribution norm. That is, if we start from

a sample of $\mu$ and perform one Metropolis-Hastings step, the next sample on the Markov chain is still a sample from $\mu$. Symbolically, if $\mathcal{K}(x'|x)$ is the total probability of going from $x$ to $x'$ during one step, then

$$\int_{\mathbb{R}^d} \mathcal{K}(x'|x) \, \mu(x)dx = \mu(x'). \tag{3.4}$$

$\mathcal{K}$ is also called the transition distribution. For the Metropolis-Hastings algorithm, this transition distribution reads

$$\mathcal{K}^{MH}(x'|x) = \alpha(x'|x)q(x'|x) + \left(\int_{\mathbb{R}^d} \alpha(y|x)q(y|x)dy\right)\delta\left(x'-x\right). \tag{3.5}$$

That is, if we accept $x'$, the probability of going form $x$ to $x'$ is that of proposing $x'$ multiplied by the acceptance probability. Otherwise we keep $x$ and second term reflects that total probability of rejection.

Verifying invariance (3.4) may be difficult in practice because the transition kernel is typically hard to integrate. There is an alternative, but stronger condition called detailed balance (DB) [83]. It is sufficient for invariance, but not necessary. A Markov chain Monte Carlo method satisfies detailed balance if

$$\mathcal{K}(x'|x) \, \mu(x) = \mathcal{K}(x|x') \, \mu(x'). \tag{3.6}$$

That is, the probability of going from $x$ to $x'$ is the same as the reverse probability. We also call a Markov chain that fulfils detailed balance 'reversible'.

It is easy to prove that detailed balance implies invariance. Integrate (3.6) over $x$, so,

$$\int_{\mathbb{R}^d} \mathcal{K}(x'|x) \, \mu(x)dx = \int_{\mathbb{R}^d} \mathcal{K}(x|x') \, \mu(x')dx$$

$$= \mu(x') \int_{\mathbb{R}^d} \mathcal{K}(x|x')dx$$

$$= \mu(x'),$$

and we obtain (3.4). Finally, it only remains to prove that the Metropolis-Hastings method is reversible. For $x \neq x'$,

$$\mathcal{K}^{MH}(x'|x)\mu(x) = \alpha(x'|x)q(x'|x)\mu(x)$$

$$= \min\left\{1, \frac{\mu(x')q(x|x')}{\mu(x)q(x'|x)}\right\}q(x'|x)\mu(x)$$

$$= \min\{\mu(x)q(x'|x), \mu(x')q(x|x')\}$$

$$= \mathcal{K}^{MH}(x|x')\mu(x'),$$

because of the symmetry in the third equation.

### 3.2.2 Irreducibility

Irreducibility is fundamentally a property concerning the mixing of the MCMC method. If there is any subset of $\mathbb{R}^d$ with non-zero measure that the Markov chain never visits, the MCMC method is useless. In technical terms, let $A \subset \mathbb{R}^d$ so that $\mu(A) > 0$. A Markov chain with transition probability $\mathcal{K}$ is called ($\mu$-)irreducible when

$$\mathcal{K}(A|x) > 0, \quad \forall x \in \mathbb{R}^d. \tag{3.7}$$

This property says that we can reach any domain $A$ from any starting point $x$ with non-zero probability. An implication is that a Markov chain with the irreducibility property will visit $A$ infinitely often.

Let us verify irreducibility of Metropolis-Hastings. Take any $A \subset \mathbb{R}^d$ with $\mu(A) > 0$. Then,

$$\int_A \mathcal{K}^{MH}(x'|x)dx' \geq \int_{A\setminus\{x\}} \mathcal{K}^{MH}(x'|x)dx'$$

$$= \int_{A\setminus\{x\}} \alpha(x'|x)q(x'|x)dx'.$$

Because $\alpha$ and $q$ are strictly positive everywhere, the bottom integral must be greater than zero for any $x$, proving irreducibility.

### 3.2.3 Aperiodicity

Aperiodicity is the opposite of irreducibility, and requires that the probability of remaining in the same position occurs with non-zero probability. That is, a Markov chain is aperiodic when $\mathcal{K}(\{x\}|x) > 0$ for all $x \in \mathbb{R}^d$.

We prove this property of the Metropolis-Hastings method by means of a contradiction. Suppose that there exists an $x \in \mathbb{R}^d$ so that $\mathcal{K}^{MH}(\{x\}|x) = 0$, then

$$\int_{\mathbb{R}^d} \alpha(x'|x) \ q(x'|x)dx' = 1. \tag{3.8}$$

The above is never true, because the acceptance probability is continuous and smaller than 1 for some $x'$. Notable, these are the states $x'$ for which the energy is higher compared to $x$. This is a contradiction, therefore aperiodicity must hold for Metropolis-Hastings.

## 3.3   Three Well-Known MCMC Methods

So far, we have only discussed the original MCMC method by Metropolis and Hastings. In this section, we present three more well-known and broadly used MCMC methods. The first method, the Metropolis-Adjusted Langevin Algorithm (MALA) [84, 13, 10, 28], generates samples from a given distribution by simulating the Euler-Maruyama discretization of an SDE with additive noise. The second method, the Delayed-Acceptance scheme [20, 69, 99, 32], creates samples with a two-level construction. It is a general framework with many instantiations. First, samples are generated using an approximate invariant distribution, and after acceptance, the remaining samples are tested with respect to the true invariant distribution. This way, bad samples are rejected in the first accept/reject step, so that the total acceptance rate in the second step increases. The final method, Hamiltonian Monte Carlo (HMC) [27, 9, 10, 11, 18], sometimes called 'Hybrid Monte Caro', generates samples from Hamiltonian systems. HMC constructs samples using a symplectic splitting scheme, typically the leapfrog method [72]. We discuss each of these methods further below.

### 3.3.1   Metropolis-Adjusted Langevin Algorithm (MALA)

For a given probability distribution $\mu$, we can construct a stochastic differential equation with additive noise that keeps $\mu$ as the unique invariant distribution. This SDE has the form

$$dX = \nabla \log \mu(X)dt + \sqrt{2}dW, \tag{3.9}$$

where $X$ is a random variable, and $dW$ is the Wiener process. Note that the normalization constant of $\mu$ vanishes due to the gradient term. The invariant distribution of this SDE is

$$\exp\left(\log \mu\right) = \mu \tag{3.10}$$

By ergodicity, any path of (3.9) samples $\mu$ exactly.

We cannot simulate (3.9) directly, however, we need to discretize it first. The MALA algorithm uses an Euler-Maruyama discretization to generate samples, reading

$$x_{n+1} = x_n + \nabla \log \mu(x_n)\Delta t + \sqrt{2\Delta t}\, \eta_n, \quad \eta_n \sim \mathcal{N}(0,1). \tag{3.11}$$

The standard normally distributed increments $\eta_n$ are independent from each other. Due to the discrete time-stepping, we always make a discretization error while simulating (3.11). We can remove this error by a Metropolis-Hastings

accept/reject step, reading

$$\alpha(x_{n+1}|x_n) = \min\left\{1, \frac{\mu(x_{n+1}) \; q^{MALA}(x_n|x_{n+1})}{\mu(x_n) \; q^{MALA}(x_{n+1}|x_n)}\right\}. \tag{3.12}$$

On acceptance, we keep $x_{n+1}$, otherwise we set $x_{n+1} = x_n$. The MALA proposal distribution can be derived from (3.11), and has the form

$$q^{MALA}(x'|x) = \frac{1}{\sqrt{4\pi\Delta t}} \exp\left(-\frac{(x' - x - \Delta t\nabla\log\mu(x))^2}{4\Delta t}\right).$$

With this Metropolis-Hastings acceptance probability, the MALA method is guaranteed to converge to $\mu$ [84, 10, 28].

**Remark 1.** *In the special case that $\mu$ is a Gibbs distribution* (1.2)*, the SDE* (3.9) *reduces to*

$$dX = -\beta\nabla V(X)dt + \sqrt{2}dW,$$

*which is equivalent to*

$$dX = -\nabla V(X)dt + \sqrt{2\beta^{-1}}dW.$$

*This SDE is nothing but the overdamped Langevin equation as introduced in section 2.3.*

## 3.3.2  Delayed-Acceptance Methods

In many scenario's, the target distribution $\mu$ is expensive to evaluate. This situation occurs often in Bayesian inference, where the likelihood is intractable to compute. At every step of the Metropolis-Hastings algorithm, one needs to compute the acceptance probability, making the whole method impractical.

It is possible to decrease the total cost by using an approximate distribution, $\tilde{\mu}$, that is in some sense close to $\mu$. Generating samples from $\mu$ then happens in two stages. Given the current sample on the Markov chain, $x_n$, first generate a new sample using proposal distribution $q$

$$x' \sim q(\cdot|x_n).$$

Then accept or reject $x'$ with approximate distribution $\tilde{\mu}$. The acceptance probability is simply the Metropolis-Hastings probability, given by

$$\tilde{\alpha}(x'|x_n) = \min\left\{1, \frac{\tilde{\mu}(x') \; q(x_n|x')}{\tilde{\mu}(x_n) \; q(x'|x_n)}\right\}.$$

If $x'$ is rejected, keep the current sample $x_{n+1} = x_n$. Otherwise, $x'$ goes through a second accept/reject step, this time with the true distribution $\mu$. The second acceptance probability essentially compares the approximate distribution with $\mu$. It reads

$$\alpha(x'|x_n) = \min\left\{1, \frac{\mu(x')\ \tilde{\mu}(x_n)}{\mu(x_n)\ \tilde{\mu}(x')}\right\}. \tag{3.13}$$

Any error made during the first accept-reject step is removed in the second due to the form of the acceptance probability.

The key behind the delayed acceptance framework is that $\tilde{\mu}$ must be cheap to evaluate [20]. Bad samples are then largely rejected by $\tilde{\mu}$, avoiding superfluous evaluations of $\mu$. That way, the overall cost of the delayed acceptance is reduced compared to the standard Metropolis-Hastings method. The actual reduction depends on the application, as well as $\mu$, $\tilde{\mu}$ and $q$. Note that the name 'delayed acceptance' is not the best description of the method, 'early rejection' fits better.

It only remains to prove that the delayed-acceptance method converges to $\mu$. Irreducibility and aperiodicity depend on the form of $q$ and $\tilde{\mu}$, we thus only prove detailed balance. The transition distribution of delayed-acceptance is

$$\mathcal{K}^{DA}(x'|x) = \alpha(x'|x)\tilde{\alpha}(x'|x)q(x'|x) + \left(\int_{\mathbb{R}^d} \alpha(y|x)\tilde{\alpha}(y|x)q(y|x)dy\right)\delta(x'-x).$$

For $x' = x$, detailed balance is trivial. Suppose $x' \neq x$. Then,

$$\mathcal{K}^{DA}(x'|x)\ \mu(x) = \alpha(x'|x)\ \tilde{\alpha}(x'|x)\ q(x'|x)\ \mu(x)$$

$$= \min\left\{1, \frac{\mu(x')\ \tilde{\mu}(x)}{\mu(x)\ \tilde{\mu}(x')}\right\}\ \min\left\{1, \frac{\tilde{\mu}(x')\ q(x|x')}{\tilde{\mu}(x)\ q(x'|x)}\right\}\ q(x'|x)\ \mu(x)$$

$$= \min\left\{\mu(x)\tilde{\mu}(x'), \mu(x')\tilde{\mu}(x)\right\}\ \min\left\{\frac{q(x'|x)}{\tilde{\mu}(x')}, \frac{q(x|x')}{\tilde{\mu}(x)}\right\},$$

$$= \mathcal{K}^{DA}(x|x')\ \mu(x').$$

From the second to third equation, we pull $\tilde{\mu}(x')$ from $\tilde{\alpha}$ and the rightmost $\mu(x)$ into the left minimum, and pull $q(x'|x)$ on the right into the right minimum. The third equation is symmetric in $x$ and $x'$, and detailed balance follows.

## 3.3.3 Hamiltonian Monte Carlo (HMC)

Suppose that the target distribution $\mu$ has the general form

$$\mu(q, p) = Z^{-1} \exp\left(-\beta\ H(q, p)\right).$$

$H(q, p)$ is a Hamiltonian, $q$ the position and $p$ the momentum variable. The deterministic Hamiltonian equations of motion then are

$$\dot{q}_i = \frac{\partial H(q, p)}{\partial p_i}$$

$$\dot{q}_i = -\frac{\partial H(q, p)}{\partial q_i}.$$

These equations can be readily integrated by any symplectic method.

If the Hamiltonian is separable, the position and momentum term are decoupled. The Hamiltonian then has the usual form

$$H(q, p) = V(q) + \frac{1}{2} p^T M^{-1} p, \tag{3.14}$$

with $M$ the diagonal mass matrix. $V(q)$ is the potential energy, see (1.2), while the second term is the kinetic energy. The equations of motion further reduce to

$$\dot{q}_i = M^{-1} p_i$$
$$\dot{p}_i = -\nabla V(q)_i, \tag{3.15}$$

which is the form we will use here.

Hamiltonian Monte Carlo (HMC) uses $L$ leapfrog steps to integrate (3.15). Suppose $q_n$ is the current sample on the Markov chain. Let $q_n(0) = q_n$. First, we draw a new momentum variable $p_n(0) \sim \mathcal{N}(0, M)$. The new position and momentum vector after one leapfrog step with time step $\Delta t$ are

$$p_n \left( \frac{1}{2} \right) = p_n(0) - \frac{\Delta t}{2} \nabla V(q_n(0))$$

$$q_n(1) = x_n(0) + \Delta t p_n \left( \frac{1}{2} \right) \tag{3.16}$$

$$p_n(1) = p_n \left( \frac{1}{2} \right) - \frac{\Delta t}{2} \nabla V(q_n(1)).$$

Applying these equations $L$ times, we obtain $q_n(L)$ and $p_n(L)$. The samples $\{q_n(l), p_n(l)\}_{l=0}^{L}$ do however not preserve the Hamiltonian $H$, rather another Hamiltonian $\tilde{H}$ that is close to $H$ [10]. It is possible stay at the current energy level $H(q_n(0), p_n(0))$ by a Metropolis-Hastings accept-reject step,

$$\alpha(q_n(L), p_n(L) | q_n(0), p_n(0)) = \min \left\{ 1, \frac{\exp \left( -\beta H[q_n(L), p_n(L)] \right)}{\exp \left( -\beta H[q_n(0), p_n(0)] \right)} \right\}. \tag{3.17}$$

On acceptance, we set $q_{n+1} = q_n(L)$, otherwise $q_{n+1} = q_n$. Note that the proposal distribution of (3.16) does not appear in the acceptance probability because the leapfrog scheme is symmetric.

If exactly integrated, or non-exactly with accept-reject step, the leapfrog method will never change the initial energy level of the Hamiltonian. In order to fully explore all energy levels of $H$, we must draw a new momentum variable $p_n(0)$ at the start of each $L$ leapfrog sets. Only then can we do proper thermodynamic integration at the given temperature [27, 9].

The only question we have not addressed yet is how to choose $L$, the number of leapfrog steps. The value of $L$ may seriously affect the mixing and convergence rate of HMC. The solution presented by the authors of [39] is to choose the number of leapfrog steps adaptively. This No-U-Turn sampler is also implemented in the STAN language [34, 16].

## 3.4 Exploration versus Exploitation in Multiscale Molecular Dynamics

So far, we have introduced several Markov chain Monte Carlo methods without giving reference to the context in which we will use these methods. The ultimate goal is to sample Gibbs distributions with these MCMC methods using proposals from the Langevin model (6.6). In addition to a high dimensional target distribution, the existence of multiple time scales in the Langevin dynamics worsen the overall sampling result.

Let us illustrate this issue on a simple toy problem. Suppose we study a two-dimensional system with energy function

$$V(x, y) = \frac{1}{2} \left(x^2 - 1\right)^2 + \frac{1}{2\varepsilon} y^2. \tag{3.18}$$

In this system, $x$ is slow and follows a bimodal distribution centred at $-1$ and $1$. Further, $y$ is fast and follows a zero mean Gaussian distribution with a small variance $\varepsilon \ll 1$. From the constants in front of both terms, the time-scale separation is $1/\varepsilon \gg 1$. We therefore chose the reaction coordinate $\xi(x, y) = x$, and its free energy is

$$A(x) = \frac{1}{2} \left(x^2 - 1\right)^2. \tag{3.19}$$

We now demonstrate the computational problem associated to sampling the Gibbs distribution of (3.18) using any of the methods described in this chapter. We pick the MALA method because the analysis is easier. It can be shown
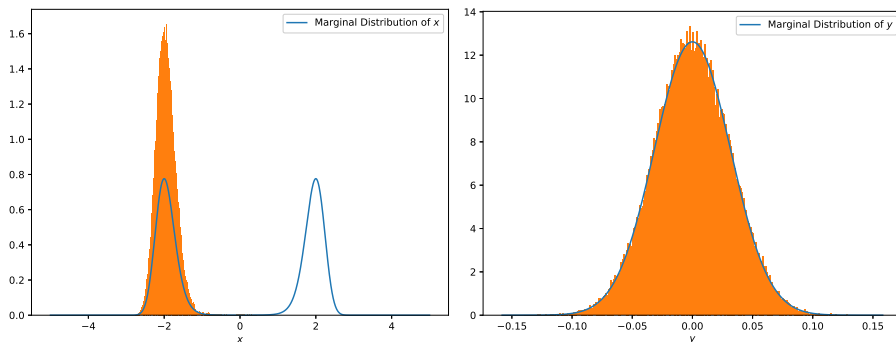
Figure 3.1: Histogram approximation of the MALA method to the marginal distributions (blue) of $x$ (left) and $y$ (right). The time scale parameter is $\varepsilon = 10^{-3}$ and the time step of the MALA method is $4\varepsilon$, right on the stability boundary. The averaged acceptance rate is 0.217. Clearly, the MALA method is unable to sample the bimodal distribution of $x$ accurately due to the time-scale separation.

that the optimal averaged acceptance rate for the MALA method is about 23%, although it depends on the dimensionality [82]. As a consequence, the proposal time step $\Delta t$ must be on the order of $\varepsilon$. Indeed, the maximal time step for stability of the Euler-Maruyama method is of order $\mathcal{O}(\varepsilon)$ [86]. For larger $\Delta t$, the actual averaged acceptance rate quickly decreases below the optimal rate.

This limit on the time step is not a problem for the fast mode, but for the slow mode. Indeed, if we want to completely sample the invariant distribution of $x$, we need to simulate the Langevin process over a time interval of size $\mathcal{O}(\varepsilon^{-1})$ with time steps of size $\mathcal{O}(\varepsilon)$. And even then, the total probability of crossing the energy barrier of (3.19) will be small because of the small time step. We illustrate this problem in figure 3.1.

Finally, we bring this computational problem into the context of 'Exploration versus Exploitation'. When the local minima of the potential energy function are unknown, we must explore the complete state space to find all stable conformations of the molecule. Exploration can take a lot of resources because of the infinite nature of the state space. On the other hand, when the Markov chain is in a local minimum, it is beneficial to sample the potential well thoroughly to discover saddle points, to compute transition probabilities and other information stored in the invariant distribution. This local exploration is known as exploitation. Since computational resources are always limited, one needs to make a trade-off between these two modes.

If a Markov chain spends too much time on exploitation, it cannot fully explore the complete state space. This is what happens in the multiscale system above. We design a new Markov chain Monte Carlo method that balances these two modes for multiscale systems in the next chapter.

# Chapter 4

# Micro-Macro Markov Chain Monte Carlo

The goal of this chapter is to construct a novel micro-macro Markov chain Monte Carlo (mM-MCMC) method that is able to overcome the time-scale separation barrier between the slow, macroscopic states, encoded in a reaction coordinate, and the remaining fast, microscopic variables. We explicitly design the algorithm to balance exploration and exploitation, something that traditional MCMC methods suffer from.

The micro-macro Markov chain Monte Carlo method is based on the 'Coupled Markov chain Monte Carlo' method [45, 47] that explores a similar strategy for lattice models. The micro-macro MCMC method extends the Coupled MCMC approach to continuous microscopic and macroscopic state spaces.

Given the current microscopic sample, the mM-MCMC method generates a new microscopic sample in five general steps: i) Restriction of the current microscopic sample to its reaction coordinate value; ii) Proposal of a new reaction coordinate sample on the macroscopic level with larger proposal moves; iii) Macroscopic accept/reject step so that the proposed reaction coordinate value follows an approximate macroscopic distribution; iv) Reconstruction by sampling a reconstruction distribution on the level set of the reaction coordinate; v) Microscopic accept/reject so that the reconstructed microscopic sample indeed samples (1.2).

Reconstruction, however, is an expensive step. Indeed, a level set of a reaction coordinate usually has a highly irregular shape, so that sampling

from a reconstruction distribution on such level sets requires an iterative procedure in itself, and can only be done in an analytic manner for simple examples. For this reason, we propose two variants of reconstruction, direct and indirect. With direct reconstruction, we simply generate a sample from a reconstruction distribution on a level set. This variant is simpler mathematically, but computationally expensive for high-dimensional problems. With indirect reconstruction, we run a biased stochastic simulation that pulls the previous microscopic sample towards the new reaction coordinate value at the macroscopic level. We loose simplicity, but gain generality.

This chapter is organised as follows. In section 4.1, we introduce the mM-MCMC method with direct reconstruction. We give details of all the steps, outline the complete algorithm, and prove convergence of the method. Then, section 4.2 is concerned with the mM-MCMC method with indirect reconstruction. We also give details of all the steps, together with the changes in relation to the former variant, and a proof of convergence. Finally, we state and prove an ergodicity result of mM-MCMC with direct reconstruction under certain conditions in section 4.3. The contents and results of this chapter have been published in [92], and the text below has been largely taken from said paper.

# 4.1 Micro-Macro Markov Chain Monte Carlo with Direct Reconstruction

The mM-MCMC method with direct reconstruction relies on the availability of two ingredients. First, we assume that we can sample from an approximate macroscopic distribution $\bar{\mu}_0(z)$ that is somewhat close to the free energy $\mu_0$. Second, we require a reconstruction distribution $\bar{\nu}(x|z)$ of microscopic samples conditioned upon a given reaction coordinate value. In principle, the choice of $\bar{\mu}_0$ and $\bar{\nu}$ is arbitrary for the mM-MCMC method with direct reconstruction to converge.

## 4.1.1 One Step of mM-MCMC with Direct Reconstruction

Suppose we start with a microscopic sample $x_n$ that samples (1.2). We generate a new microscopic sample $x_{n+1}$ in five steps, detailed below.

**Restriction**  Compute the reaction coordinate value of the current microscopic sample, i.e.,

$$z_n = \xi(x_n).$$

This step is cheap as it is a simple function evaluation.

**Macroscopic Proposal**   Once on the macroscopic level, we can generate a new reaction coordinate value $z'$ with larger proposal moves because the fast time scales have been removed. That is, we propose a new reaction coordinate value $z'$ using a macroscopic transition probability $q_0(z'|z_n)$, i.e.,

$$z' \sim q_0(\cdot|z_n)$$

This proposal distribution can, for instance, be based on a MALA scheme (3.11), a Hamiltonian Monte Carlo step (3.16), or a simple Brownian motion (section 2.1).

**Macroscopic Accept/Reject**   To ensure that $z'$ is indeed a sample from the approximate macroscopic distribution $\bar{\mu}_0$, we accept $z'$ with probability

$$\alpha_{CG}(z'|z_n) = \min \left\{ 1, \frac{\bar{\mu}_0(z') \, q_0(z_n|z')}{\bar{\mu}_0(z_n) \, q_0(z'|z_n)} \right\}, \tag{4.1}$$

which is the standard Metropolis-Hastings form for the acceptance rate. If $z'$ is accepted, we proceed to the reconstruction step. If not, we return to the first step and set $x_{n+1} = x_n$.

Using only the macroscopic steps, that we can define the macroscopic MCMC kernel $\mathcal{D}$ as

$$\mathcal{D}(z'|z_n) = \alpha_{CG}(z'|z_n) \, q_0(z'|z_n) + \left( 1 - \int_H \alpha_{CG}(y|z_n) \, q_0(y|z_n) \, dy \right) \delta(z' - z_n). \tag{4.2}$$

Note that $\bar{\mu}_0$ is indeed the time-invariant probability measure associated with $\mathcal{D}$.

**Direct Reconstruction**   If the reaction coordinate value $z'$ has been accepted, we construct a new microscopic sample $x'$ that is consistent with $z'$. That is, we impose that $\xi(x') = z'$. Direct reconstruction solves this by taking one sample from the given reconstruction distribution $\bar{\nu}(x'|z')$,

$$x' \sim \bar{\nu}(\cdot|z')$$

defined on the level set $\Sigma(z')$ of the reaction coordinate.

**Microscopic Accept/Reject** Finally, we need to make sure that $x'$ is indeed a sample of the microscopic Gibbs distribution. For random approximate macroscopic and reconstruction distributions, this will almost never be the case. Therefore, we need one final microscopic acceptance probability that takes these (approximate) distributions into account.

To compute this corresponding microscopic acceptance probability, we first define the complete proposal distribution of the mM-MCMC method with direct reconstruction. Starting from the previous microscopic sample $x_n$, the microscopic proposal distribution reads

$$q(x'|x_n) = \bar{\nu}(x'|\xi(x')) \, \mathcal{D}(\xi(x')|\xi(x_n)), \tag{4.3}$$

i.e., the probability of transitioning from $x_n$ to $x'$ is given by the probability of generating and accepting a reaction coordinate value $\xi(x')$ on the macroscopic level, multiplied by the probability of reconstructing the microscopic sample $x'$, given its reaction coordinate value.

Using the detailed balance condition on the macroscopic level,

$$\mathcal{D}(z'|z_n) \, \bar{\mu}_0(z_n) = \mathcal{D}(z_n|z') \, \bar{\mu}_0(z'),$$

the microscopic acceptance probability is given by

$$\alpha_F(x'|x_n) = \min\left\{1, \frac{\mu(x') \, q(x_n|x')}{\mu(x_n) \, q(x'|x_n)}\right\}$$

$$= \min\left\{1, \frac{\mu(x') \, \bar{\nu}(x|z_n) \, \mathcal{D}(z_n|z')}{\mu(x_n) \, \bar{\nu}(x'|z') \, \mathcal{D}(z'|z_n)}\right\}$$

$$= \min\left\{1, \frac{\mu(x') \, \bar{\mu}_0(z_n) \, \bar{\nu}(x_n|z_n)}{\mu(x_n) \, \bar{\mu}_0(z') \, \bar{\nu}(x'|z')}\right\}. \tag{4.4}$$

On acceptance, we set $x_{n+1} = x'$, otherwise, $x_{n+1} = x_n$.

**Remark 2.** *There is a special situation in which the microscopic acceptance probability is always 1. We call this situation 'exact reconstruction' and it holds when one can write*

$$\mu(x) = \bar{\nu}(x|\xi(x)) \, \bar{\mu}_0(\xi(x)). \tag{4.5}$$

*Because the microscopic distribution $\mu(x)$ can only be decomposed uniquely in the form $\mu(x) = \nu(x|\xi(x)) \, \mu_0(\xi(x))$, where $\mu_0$ and $\nu$ are defined in (2.15) and (2.19) respectively, we must have $\bar{\mu}_0(\xi(x)) = \mu_0(\xi(x))$ and $\bar{\nu}(x|\xi(x)) = \nu(x|\xi(x))$ for exact reconstruction. When (4.5) holds, it is easy to see that the reconstruction acceptance rate will always be 1, so that the mM-MCMC performs no unnecessary computational work during reconstruction. Hence, for*

*computational efficiency, it is beneficial to choose the approximate macroscopic distribution $\bar{\mu}_0$ and the reconstruction distribution $\bar{\nu}$ in such way that the exact reconstruction property (4.5) approximately holds.*

**Remark 3.** *The micro-macro MCMC method can be extended to allow for any kind of reaction coordinate / macroscopic variables. In that context, the kinetic Monte Carlo method [95] defines discrete macroscopic states as regions around the local minima in the potential energy $V(x)$. In this setting, the macroscopic variable $z$ represents one of the discrete states given by the local minima of $V(x)$, and the distribution $\bar{\mu}_0$ assigns a probability to each of these macroscopic states. Sampling these macroscopic states then consists of sampling discrete transition probabilities between the local minima using the macroscopic proposal distribution $q_0$. During reconstruction, we then sample a microscopic sample around the given local minimum. That is, the reconstruction distribution $\bar{\nu}$ is defined on the basin of attraction of the given local minimum, and is zero everywhere else. The formulation of Algorithm 1, including the acceptance probabilities remains unaltered in such a situation. A similar multilevel kinetic Monte Carlo method has already been applied in the context of discrete lattice systems [46]. That is, the macroscopic level consists of a coarser lattice and each macroscopic lattice point encompasses a number of microscopic lattice points. The main difference with the mM-MCMC method that we present here is that our macroscopic space is continuous instead of discrete. We therefore need another type of reconstruction that is able to generate a continuous microscopic sample from a discrete macroscopic variable.*

### 4.1.2 The Complete Algorithm

The complete algorithm of mM-MCMC with direct reconstruction is shown in algorithm 1, together with a visualization in figure 4.1.

### 4.1.3 Proof of Convergence

In this section, we prove that the mM-MCMC method with direct reconstruction indeed converges to the microscopic Gibbs distribution (1.2). Theorem 1 imposes mild conditions on the macroscopic transition distribution $q_0$, the approximate macroscopic invariant distribution $\bar{\mu}_0$ and the reconstruction distribution $\nu$.

Before formulating the convergence theorem, we define the transition kernel of the mM-MCMC method with direct reconstruction. This transition kernel is based on the microscopic proposal distribution (4.3). The probability of

---

**Algorithm 1** Micro-macro Markov chain Monte Carlo with direct reconstruction

---

Given a microscopic sample $x_n$, $n = 1, 2, \ldots$ .

(i) **Restriction**: compute the reaction coordinate value $z_n = \xi(x_n)$.

(ii) **Macroscopic Proposal**: Generate a reaction coordinate value via $z' \sim q_0(\cdot | z_n)$.

(iii) **Macroscopic Accept/Reject**

- Accept the reaction coordinate value with probability

$$\alpha_{CG}(z'|z_n) = \min\left\{ 1, \frac{\bar{\mu}_0(z')\; q_0(z_n|z')}{\bar{\mu}_0(z_n)\; q_0(z'|z_n)} \right\}$$

- On acceptance, proceed to step (iii), otherwise set $x_{n+1} = x_n$ and repeat step (ii).

(iv) **Direct Reconstruction**: Generate a microscopic sample $x' \sim \bar{\nu}(\cdot | z')$.

(v) **Microscopic Accept/Reject**

- Accept the microscopic sample with probability

$$\alpha_F(x'|x_n) = \min\left\{ 1, \frac{\mu(x')\; \bar{\mu}_0(z_n)\; \bar{\nu}(x_n|z_n)}{\mu(x_n)\; \bar{\mu}_0(z')\; \bar{\nu}(x'|z')} \right\}.$$

- On acceptance, set $x_{n+1} = x'$ and return to step (i) for the next microscopic sample. Otherwise, set $x_{n+1} = x_n$ and generate a new reaction coordinate value in step (ii).

---
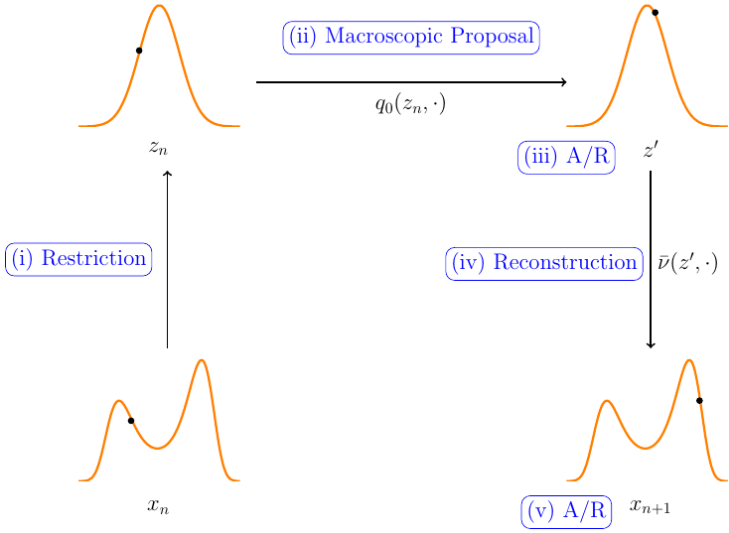
Figure 4.1: Visualization of one step of mM-MCMC with direct reconstruction.

transitioning from state $x$ to state $x'$ reads

$$
\mathcal{K}_{mM}(x'|x) = \begin{cases} \alpha_F(x'|x)\ \bar{\nu}(x'|\xi(x'))\ \alpha_{CG}(\xi(x')|\xi(x))\ q_0(\xi(x')|\xi(x)) & x' \neq x \\ \\ 1 - \displaystyle\int_{\mathbb{R}^d} \alpha_F(y|x)\ \bar{\nu}(y|\xi(y))\ \alpha_{CG}(\xi(y)|\xi(x))\ q_0(\xi(y)|\xi(x))dy & x' = x. \end{cases}
$$

Using the definition of the microscopic proposal probability $q(x, x')$, the full micro-macro transition kernel can be written as

$$
\mathcal{K}_{mM}(x'|x) = \alpha_F(x'|x)\ \bar{\nu}(x'|\xi(x'))\ \alpha_{CG}(\xi(x')|\xi(x))\ q_0(\xi(x')|\xi(x))
$$

$$
+ \left(1 - \int_{\mathbb{R}^d} \alpha_F(y|x)\ \bar{\nu}(y|\xi(y))\ \alpha_{CG}(\xi(y)|\xi(x))\ q_0(\xi(y)|\xi(x))dy\right)\delta(x' - x)
$$

$$
= \alpha_F(x'|x)\ q(x'|x) + \left(1 - \int_{\mathbb{R}^d} \alpha_F(y|x)\ q(y|x)dy\right)\delta(x' - x), \tag{4.6}
$$

We then have the following Theorem:

**Theorem 1.** *For every the macroscopic proposal $q_0$ and approximate macroscopic $\bar{\mu}_0$ distributions that are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$, and every reconstruction distribution $\bar{\nu}$ such that $q$ is absolutely continuous on $\mathbb{R}^d$, the following statements are true:*

(i) the transition probability kernel (4.6) *satisfies the detailed balance condition with the Gibbs distribution* $\mu$;

(ii) the Gibbs measure $\mu$ is a stationary distribution of $\mathcal{K}_{mM}$;

(iii) the chain $\{x_n\}$ is $\mu-$irreducible;

(iv) the chain $\{x_n\}$ is aperiodic.

Let us clarify the statement of theorem 1. The second statement (ii) implies that the Markov chain has $\mu$ as an invariant measure, which is naturally a consequence of the detailed balance condition (i). Detailed balance also implies that the stationary distribution is unique. The third and fourth statement ensure ergodicity of the Markov chain. Ergodicity means that the averages over one sample path, $N^{-1} \sum_{n=1}^{N} g(x_n)$ converge to averages over the stationary distribution $\int g \, d\mu$ almost surely as $N$ increases to infinity, for every $g \in L_1(\mathbb{R}^{3D}, \mu)$.

A proof similar to that of theorem 1 was already given in [47] in the specific context of stochastic models defined on lattice systems. Here, we extend that proof to molecular systems with reaction coordinates.

***Proof.***

(i) The case $x' = x$ is trivial. Take $x \neq x'$ and write for the transition kernel (4.6)

$$\mathcal{K}_{mM}(x'|x) \, \mu(x) = \alpha_F(x'|x) \, \bar{\nu}(x'|\xi(x')) \, \alpha_{CG}(\xi(x')|\xi(x)) \, q_0(\xi(x')|\xi(x)) \, \mu(x)$$

$$= \min \left\{ 1, \, \frac{\mu(x') \, \bar{\mu}_0(\xi(x)) \, \bar{\nu}(x|\xi(x))}{\mu(x) \, \bar{\mu}_0(\xi(x')) \, \bar{\nu}(x'|\xi(x'))} \right\} \bar{\nu}(x'|\xi(x'))$$

$$\times \min \left\{ 1, \, \frac{\bar{\mu}_0(\xi(x')) \, q_0(\xi(x)|\xi(x'))}{\bar{\mu}_0(\xi(x)) \, q_0(\xi(x')|\xi(x))} \right\} q_0(\xi(x')|\xi(x)) \, \mu(x)$$

$$= \min \left\{ \mu(x) \, \bar{\mu}_0(\xi(x')) \, \bar{\nu}(x'|\xi(x')), \, \mu(x') \, \bar{\mu}_0(\xi(x)) \, \bar{\nu}(x|\xi(x)) \right\}$$

$$\times \min \left\{ \frac{q_0(\xi(x')|\xi(x))}{\bar{\mu}_0(\xi(x'))}, \, \frac{q_0(\xi(x)|\xi(x'))}{\bar{\mu}_0(\xi(x))} \right\}$$

$$= \mathcal{K}_{mM}(x|x') \, \mu(x'),$$

since the third equality is symmetric in $x$ and $x'$.

(ii) Follows directly from (i).

(iii) To prove that the chain $\{x_n\}$ is $\mu$-irreducible, we need to show that $\mathcal{K}_{mM}(A|x) > 0$ for all $x \in \mathbb{R}^d$ and for all measurable sets $A \subset \mathbb{R}^{3D}$ with $\mu(A) > 0$. Note that

$$
\begin{aligned}
\mathcal{K}_{mM}(A|x) = \int_A \mathcal{K}_{mM}(x'|x)dx' &\geq \int_{A\setminus\{x\}} \mathcal{K}_{mM}(x'|x)dx' \\
&= \int_{A\setminus\{x\}} \alpha_F(x'|x) \, \bar{\nu}(x'|\xi(x')) \, \alpha_{CG}(\xi(x')|\xi(x)) \, q_0(\xi(x')|\xi(x))dx' \\
&= \int_{A\setminus\{x\}} \alpha_F(x'|x) \, q(x'|x)dx'.
\end{aligned}
$$
(4.7)

The proposal distribution $q$ is strictly positive by assumption. And because $A \subset \mathrm{supp}(\mu)$, the acceptance rate $\alpha_F$ is strictly positive as well, proving that $\mathcal{K}(A|x) > 0$.

(iv) For $\{x_n\}$ to be aperiodic, it is sufficient to show that there exists an $x \in \mathrm{supp}(\mu)$ such that $\mathcal{K}(x,\{x\}) > 0$, implying that $x_{n+1} = x_n$ can occur with positive probability [47]. We will prove this by contradiction. The transition kernel reads

$$
\mathcal{K}(\{x\}|x) = 1 - \int_{\mathbb{R}^d} \alpha_F(x'|x) \, \bar{\nu}(x'|\xi(x')) \, \alpha_{CG}(\xi(x')|\xi(x)) \, q_0(\xi(x')|\xi(x))dx'.
$$

If $\mathcal{K}(\{x\}, x) = 0$ for all $x \in \mathbb{R}^d$ then

$$
\int_{\mathbb{R}^d} \alpha_F(x, x') \, \bar{\nu}(\xi(x'), x') \, \alpha_{CG}(\xi(x), \xi(x')) \, q_0(\xi(x), \xi(x'))dx' = 1, \qquad (4.8)
$$

implying both acceptance probabilities $\alpha_F(x, x')$ and $\alpha_{CG}(\xi(x), \xi(x'))$ should be 1 almost everywhere. This implies that the proposal kernel $\bar{\nu}(x'|\xi(x')) \, q_0(\xi(x')|\xi(x))$ samples from the correct invariant distribution $\mu$ without rejections. Hence, there exists some $x \in \mathbb{R}^{3D}$ such that $\mathcal{K}(\{x\}|x) > 0$. □

## 4.2 Micro-Macro MCMC with Indirect Reconstruction

The reconstruction distribution (2.19) is in principle always available, but is usually hard to sample from. Indeed, sampling from the reconstruction distribution

$$
\nu(x|z) \propto \exp\left(-\beta V(x)\right) \, \left(\det G(x)\right)^{-1/2}, \quad x \in \Sigma(z)
$$

is often harder than sampling the target distribution $\mu$, because the domain $\Sigma(z)$ can have an irregular shape, depending on the form of $\xi$. In this section, we propose a more general *indirect reconstruction* scheme that generates a microscopic sample close to the sub-manifold $\Sigma(z)$, but only approximately corresponds to the desired reaction coordinate value. In that regard, we construct a biased stochastic process on which the indirect reconstruction distribution is based. The microscopic sample $x$, obtained by the indirect reconstruction scheme, does not generally lie on the given sub-manifold of constant reaction coordinate value $\Sigma(z)$, but close to it. The resulting indirect reconstruction scheme depends on couple of parameters that need to be chosen in advance.

We start this section by explaining the indirect reconstruction scheme in section 4.2.1, followed by an argument in section 4.2.2 on why we need to define the mM-MCMC method with indirect reconstruction on the extended state space $\mathbb{R}^{3D} \times \mathbb{R}^m$. Afterwards, we go through one step of the mM-MCMC method with indirect reconstruction in section 4.2.3, and we give a full overview of the algorithm in section 4.2.4. After that, we prove convergence of the resulting mM-MCMC algorithm to the microscopic Gibbs measure in section 4.2.5. Finally, we give heuristics for optimal values for the reconstruction parameters in section 4.2.6.

## 4.2.1 Indirect Reconstruction through Biased Simulation

Consider a reaction coordinate value $z$. To reconstruct a microscopic sample associated with $z$ via indirect reconstruction, we perform a time integration of the biased stochastic process

$$dX = -\nabla \left( V(X) + \frac{\lambda}{2} \left\| \xi(X) - z \right\|^2 \right) dt + \sqrt{2\beta^{-1}} dW, \qquad (4.9)$$

that pulls the reaction coordinate value of $X$ towards $z$. In our framework, $z$ has been sampled on the macroscopic level. This biased process has an invariant distribution of the form

$$\nu_\lambda(x; z) = \left( \frac{\lambda\beta}{2\pi} \right)^{n/2} \frac{1}{\mu_\lambda(z)} \exp\left( -\beta V(x) \right) \ \exp\left( -\frac{\lambda\beta}{2} \left\| \xi(x) - z \right\|^2 \right), \quad (4.10)$$

which we will also call the *indirect reconstruction distribution*. The value $\mu_\lambda(z)$ is a normalization constant depending on $z$ and $\lambda$. Note that we introduce the notation $\nu_\lambda(x; z)$ for the indirect reconstruction distribution because the microscopic sample $x \in \mathbb{R}^{3D}$ does not necessarily have reaction coordinate $z$; instead $z$ merely acts as a parameter in the invariant distribution of the process (4.10). The factor $\exp\left( -\beta V(x) \right)$ in (4.10) corresponds to the microscopic

Gibbs distribution (1.2), while the extra term $\frac{\lambda}{2} \|\xi(x) - z\|^2$ in the potential energy ensures that the measure $\nu_\lambda(x; z)$ will concentrate on values of $x$ for which $\xi(x)$ is close to $z$, provided $\lambda$ is large enough.

Before giving a numerical discretization of the biased stochastic process, we first derive an expression of the normalization constant $\mu_\lambda(z)$. Integrating (4.10) using the co-area formula gives

$$\int_{\mathbb{R}^{3D}} \left(\frac{\lambda\beta}{2\pi}\right)^{n/2} \frac{1}{\mu_\lambda(z)} \exp(-\beta V(x)) \exp\left(-\frac{\beta\lambda}{2} \|\xi(x) - z\|^2\right) \, dx$$

$$= \frac{Z_V}{Z_A} \int_{\mathbb{R}^m} \left(\frac{\lambda\beta}{2\pi}\right)^{n/2} \frac{\exp\left(-\frac{\lambda\beta}{2} \|u - z\|)^2\right)}{\mu_\lambda(z) \exp\left(\beta A(u)\right)} \times$$

$$\frac{Z_A}{Z_V} \int_{\Sigma(u)} \frac{\exp\left(-\beta(V(x) - A(u))\right)}{\|\nabla\xi(x)\|} d\sigma_{\Sigma(u)}(x) \, du.$$

The second integral over $\Sigma(u)$ is precisely the time-invariant reconstruction distribution $\nu(x|u)$ (2.19), which integrates to 1. We thus obtain

$$\int_{\mathbb{R}^{3D}} \nu_\lambda(x; z) dx = \frac{Z_V}{Z_A} \int_{\mathbb{R}^m} \left(\frac{\lambda\beta}{2\pi}\right)^{n/2} \frac{\exp\left(-\frac{\lambda\beta}{2} \|u - z\|^2\right) \exp\left(-\beta A(u)\right)}{\mu_\lambda(z')} du = 1.$$

Rewriting the expression above, the normalization constant $\mu_\lambda(z)$ reads

$$\mu_\lambda(z) = \frac{Z_V}{Z_A} \int_{\mathbb{R}^{3D}} \left(\frac{\lambda\beta}{2\pi}\right)^{n/2} \exp\left(-\frac{\lambda\beta}{2} \|u - z\|^2\right) \exp\left(-\beta A(u)\right) du. \quad (4.11)$$

The expression for $\mu_\lambda(z)$ is a convolution of a Gaussian form with the invariant distribution of the reaction coordinates $\mu_0(z)$. This convolution operation is a scaled Weierstrass transform [38] and can be viewed as a filter on $\mu_0$. It will become clear in section 4.2.3 that the value of $\mu_\lambda(z)$ needs to be known in advance to run the mM-MCMC method with indirect reconstruction. We will assume that $\mu_\lambda$ is analytically available for now and we refer approximations of $\mu_\lambda$ to chapter 6.

So how do we use the indirect reconstruction scheme? Starting with the previous microscopic sample obtained by the mM-MCMC method $x_n = x_{n,0}$, we generate a sequence of microscopic samples $x_{n,k}, k = 0, \ldots, K$ by time-stepping the biased dynamics (4.9), using the Euler-Maruyama discretization

$$x_{n,k+1} = x_{n,k} - \nabla V(x_{n,k})\delta t - \lambda(\xi(x_{n,k}) - z)\nabla\xi(x_{n,k})\delta t + \sqrt{2\beta^{-1}\delta t} \, \eta_{n,k}, \quad (4.12)$$

$\eta_{n,k} \sim \mathcal{N}(0,1)$, followed by an accept/reject step to ensure the microscopic samples $x_{n,k}$ indeed sample the indirect reconstruction distribution $\nu_\lambda(\cdot; z)$.

That is, we accept $x_{n,k+1}$ from $x_{n,k}$ with probability

$$\min\left\{1, \frac{\nu_\lambda(x_{n,k+1};z)\ q_{MALA}(x_{n,k}|x_{n,k+1})}{\nu_\lambda(x_{n,k};z)\ q_{MALA}(x_{n,k+1}|x_{n,k})}\right\},$$

where $q_{MALA}$ is the MALA transition distribution (3.11) based on the discretization (4.12). We expect the MALA method to work well on the biased dynamics as any problems with metastability or time-scale separations are alleviated due to the biasing potential. We perform $K$ time steps of the biased simulation process to overcome the burn-in period associated with the inconsistency of the initial condition $x_n$ with the desired reaction coordinate value $z$. After these $K$ steps, we propose $x_{n,K}$ to be the reconstructed microscopic sample.

## 4.2.2  The Need for an Extended State Space

A reconstructed microscopic sample $x$ does generally not have reaction coordinate value $z$, that is $x \notin \Sigma(z)$. As a consequence, there is a fundamental asymmetry between restriction and indirect reconstruction. Indeed, on the one hand, each microscopic sample corresponds to exactly one reaction coordinate value, but on the other hand, the reaction coordinate value of the reconstructed sample can by anything. In mathematical terms, indirect reconstruction is not the inverse of restriction. Therefore, if we simply plug in the indirect reconstruction distribution in the mM-MCMC method with direct reconstruction, i.e., if we use the reconstruction distribution $\bar{\nu} = \nu_\lambda$ in the microscopic acceptance rate (4.4), one can expect that the resulting method is no longer reversible. In this section, we prove this intuitive result.

Indeed, suppose we plug in the indirect reconstruction distribution $\nu_\lambda(\cdot; z)$ into the mM-MCMC method with direct reconstruction. The resulting expression for the transition kernel is then not only a function of the microscopic samples $x$ and $x'$, but also depends on the reaction coordinate value $z'$ sampled at the macroscopic level. Therefore, we must integrate $\mathcal{K}_{mM}$ over all values of $z'$ to obtain a closed expression for the mM-MCMC transition kernel with indirect reconstruction only in $x$ and $x'$. This integrated transition kernel with the indirect reconstruction method reads

$$\mathcal{K}(x'|x) = \begin{cases} \displaystyle\int_H \alpha_F(x'|x,z')\ \nu_\lambda(x';z')\ \alpha_{CG}(z'|\xi(x))\ q_0(z'|\xi(x))dz' & x' \neq x \\[4mm] \displaystyle 1 - \int_{\mathbb{R}^d}\int_H \alpha_F(y|x,z')\ \nu_\lambda(y;z')\ \alpha_{CG}(z'|\xi(x))\ q_0(z'|\xi(x))dz'\ dy, & x' = x, \end{cases}$$

$$(4.13)$$

where the microscopic acceptance rate $\alpha_F(x'|x, z')$ has the usual form, but needs the extra dependency on $z'$ since $\xi(x') \neq z'$, i.e.,

$$\alpha_F(x'|x, z') = \left\{ 1, \frac{\mu(x') \, \bar{\mu}_0(\xi(x)) \, \nu_\lambda(x; \xi(x))}{\mu(x) \, \bar{\mu}_0(z') \, \nu_\lambda(x'; z')} \right\}.$$

The transition kernel (4.13), however, fails to satisfy detailed balance. Indeed, if we write out the detailed balance condition for $x' \neq x$, we find

$$\mathcal{K}(x'|x) \, \mu(x) = \int_H \min \left\{ 1, \frac{\mu(x') \, \bar{\mu}_0(\xi(x)) \, \nu_\lambda(x; \xi(x))}{\mu(x) \, \bar{\mu}_0(z') \, \nu_\lambda(x'; z')} \right\} \, \nu_\lambda(x'; z')$$

$$\times \min \left\{ 1, \frac{\bar{\mu}_0(z') \, q_0(\xi(x)|z')}{\bar{\mu}_0(\xi(x)) \, q_0(z'|\xi(x))} \right\} \, q_0(z'|\xi(x))\mu(x)dz'$$

$$= \int_H \min \left\{ \mu(x) \, \bar{\mu}_0(z') \, \nu_\lambda(x'; z'), \mu(x') \, \bar{\mu}_0(\xi(x)) \, \nu_\lambda(x; \xi(x)) \right\}$$

$$\times \min \left\{ \frac{q_0(z'|\xi(x))}{\bar{\mu}_0(z')}, \frac{q_0(\xi(x)|z')}{\bar{\mu}_0(\xi(x))} \right\} \, dz'.$$

The above expression is not symmetric in $x$ and $x'$ because $\nu_\lambda(x'; z')$ is integrated over all possible $z'$, while $\nu_\lambda(x; \xi(x))$ is independent of $z'$. As a consequence, writing the mM-MCMC scheme with indirect reconstruction only in terms of the microscopic samples does not produce a reversible Markov chain. This above observation motivates the need to reformulate the mM-MCMC method on an extended state space that includes both the microscopic samples and the reaction coordinate values as decoupled variables.

## 4.2.3 mM-MCMC with Indirect Reconstruction on Extended State Space

Having introduced the indirect reconstruction scheme in Section 4.2.1, we go through the full mM-MCMC method with indirect reconstruction. As explained in the previous section 4.2.2 the mM-MCMC method with indirect reconstruction must be defined on the extended state space of both microscopic and macroscopic samples $\mathbb{R}^{3D} \times \mathbb{R}^m$.

Given the current combined microscopic/macroscopic sample $(x_n, z_n) \in \mathbb{R}^{3D} \times \mathbb{R}^m$, we generate a new sample $(x_{n+1}, z_{n+1})$ in four steps. These steps are the same as for mM-MCMC with direct reconstruction, except that the restriction step has been removed.

**Macroscopic Proposal**   In the first step, we propose a new reaction coordinate value $z'$ using the macroscopic proposal distribution $q_0(z'|z_n)$. That is

$$z' \sim q_0(\cdot|z_n)$$

These macroscopic proposal can, for instance, be based on a MALA scheme (3.11), a Hamiltonian Monte Carlo step (3.16), or a simple Brownian motion, exactly as the direct reconstruction algorithm.

**Macroscopic Accept/Reject**   To ensure that $z'$ is indeed a sample of the approximate macroscopic distribution $\bar{\mu}_0$, we accept $z'$ with probability $\alpha_{CG}$ (7.6). On acceptance, we proceed to the reconstruction step, otherwise, we immediately define $(x_{n+1}, z_{n+1}) = (x_n, z_n)$ and go back to the first step. Note that the macroscopic steps remain the same, since the reconstruction has no impact on the macroscopic level.

**Indirect Reconstructing**   After macroscopic acceptance, we reconstruct a new microscopic sample $x'$ from the proposed reaction coordinate value $z'$ using the indirect reconstruction step discussed in section 4.2.1. That is, we simulate $K$ MALA steps (4.12) with time step $\delta t$ and with initial condition $x_{n,0} = x_n$. Afterwards, we decide whether to accept $x_{n,K} = x'$ using an acceptance criterion.

**Microscopic Accept/Reject**   There is one final difference between the mM-MCMC variants, and that difference is in the microscopic acceptance probability. Indeed, if we use the same form of the microscopic acceptance probability as defined in the direct reconstruction algorithm (4.4) but plug in the indirect reconstruction distribution $\nu_\lambda$, the microscopic acceptance probability would read

$$\hat{\alpha}_F(x', z'|x, z) = \min \left\{ 1, \frac{\mu(x') \; \bar{\mu}_0(z) \; \exp\left(-\beta V(x)\right) \; \exp\left(-\frac{\beta\lambda}{2} \|\xi(x) - z\|^2\right) \; \mu_\lambda(z')}{\mu(x) \; \bar{\mu}_0(z') \; \exp\left(-\beta V(x')\right) \; \exp\left(-\frac{\beta\lambda}{2} \|\xi(x') - z'\|^2\right) \; \mu_\lambda(z)} \right\}.$$

One can see that when $\lambda$ is large, the ratio of the two sharp Gaussian factors could significantly lower the average microscopic acceptance rate when $\xi(x)$ lies farther from $z$ than $\xi(x')$ lies from $z'$. In such scenario, the mM-MCMC method with indirect reconstruction would become inefficient due to the many rejected reconstructed microscopic samples that are computationally expensive to obtain.

To overcome this computational issue, we use a modified microscopic acceptance probability where we remove the factors that contain $\lambda$. That is, we accept the

reconstructed microscopic sample $x'$ with the modified probability

$$
\alpha_F(x', z'|x, z) = \min\left\{1, \frac{\mu(x')\,\bar{\mu}_0(z)\,\exp\left(-\beta V(x)\right)\,\mu_\lambda(z')}{\mu(x)\,\bar{\mu}_0(z')\,\exp\left(-\beta V(x')\right)\,\mu_\lambda(z')}\right\}
$$

$$
= \min\left\{1, \frac{\exp(-\beta V(x'))\,\bar{\mu}_0(z)\,\exp\left(-\beta V(x)\right)\mu_\lambda(z')}{\exp(-\beta V(x))\,\bar{\mu}_0(z')\,\exp\left(-\beta V(x')\right)\,\mu_\lambda(z)}\right\}
$$

$$
= \min\left\{1, \frac{\bar{\mu}_0(z)\,\mu_\lambda(z')}{\bar{\mu}_0(z')\,\mu_\lambda(z)}\right\}. \tag{4.14}
$$

We will show in section 4.2.5 that the mM-MCMC method with this acceptance probability is indeed reversible with respect to $\mu$, and that only the marginal invariant distribution of the reaction coordinate values is changed. We will also show in the numerical experiments that the average microscopic acceptance rate is close the 1 as desired.

Finally, on acceptance we define $(x_{n+1}, z_{n+1}) = (x', z')$, otherwise we set $(x_{n+1}, z_{n+1}) = (x_n, z_n)$ and return to the macroscopic proposal step.

**Remark 4.** *If the approximate macroscopic distribution is the marginal distribution of the reaction coordinate, i.e., $\bar{\mu}_0 = \mu_0$, the expression for the microscopic acceptance probability converges to 1 as $\lambda$ increases to $\infty$. Indeed, we have the limit*

$$
\lim_{\lambda \to \infty} \mu_\lambda(z') = \frac{Z_V}{Z_A}\exp(-\beta A(z')) = Z_V \mu_0(z').
$$

*Therefore, if we drive $\lambda$ to infinite, the microscopic acceptance rate (4.14) becomes*

$$
\lim_{\lambda \to \infty} \min\left\{1, \frac{\mu_0(z)\,\mu_\lambda(z')}{\mu_0(z')\,\mu_\lambda(z)}\right\} = \min\left\{1, \lim_{\lambda \to \infty} \frac{\mu_0(z)\,\mu_\lambda(z')}{\mu_0(z')\,\mu_\lambda(z)}\right\}
$$

$$
= \min\left\{1, \frac{\mu_0(z)\,\mu_0(z')}{\mu_0(z')\,\mu_0(z)}\right\} = 1,
$$

*in point-wise sense. We can interchange the limit and the minimum in the first equality due to the fact that both functions inside the minimum are continuous in $z$ and $z'$, and so is the minimum of both functions.*

**Remark 5.** *One needs to know the normalization constant $\mu_\lambda$ of the indirect reconstruction distribution, to accurately evaluate the microscopic acceptance probability (4.14). As we showed in section 4.2.1, this normalization constant is close to the free energy of the chosen reaction coordinate, a quantity that can be extremely challenging to compute. However, the objective of the mM-MCMC*

*method (both with direct and indirect reconstruction) is not to compute the free energy, but to accelerate the MCMC sampling of the microscopic Gibbs measure using the free energy.*

**Remark 6.** *Given that we need to know the (approximate) free energy $\mu_\lambda$ for the mM-MCMC method anyway, one might ask the following question: Is it possible to bias the macroscopic reaction coordinate sample to achieve a Gaussian, or uniform distribution, and then use importance sampling to compute averages according to the Gibbs distribution? While such a method is theoretically possible, and worth exploring, there is one major practical difference with our method proposed above. If we assign an importance weight to each macroscopic sample instead of performing an accept/reject step, we would need to perform an indirect reconstruction step after each macroscopic proposal move, regardless of the value of the weight. Such a reconstruction step will be expensive for high-dimensional molecular systems. Our mM-MCMC method with indirect reconstruction reduces the overall runtime precisely by keeping the number of macroscopic accepted proposals, and thus reconstructions, low.*

### 4.2.4 Complete Algorithm

We outline the complete mM-MCMC algorithm with indirect reconstruction visually in figure 4.2, and algorithmically in algorithm 2.

### 4.2.5 Proof of Convergence

In this section, we show that the mM-MCMC scheme with indirect reconstruction has a unique time-invariant distribution and that the method is ergodic. First, we given an expression for the correct transition kernel of mM-MCMC with indirect reconstruction. This transition kernel, defined on the extended state space $\mathbb{R}^{3D} \times \mathbb{R}^m$, reads

$$
\mathcal{K}^*_{mM}(x', z'|x, z) = \begin{cases} \alpha_F(x', z'|x, z) \ \nu_\lambda(x'; z') \ \alpha_{CG}(z'|z) \ q_0(z'|z) \\[2ex] 1 - \displaystyle\int_{\mathbb{R}^{3D} \times \mathbb{R}^m} \alpha_F(y, u|x, z) \ \nu_\lambda(y; u) \ \alpha_{CG}(u|z) \ q_0(u|z) du \ dy. \end{cases}
$$
$$(4.15)$$

The first case holds when $(x', z') \neq (x, z)$, the second case when it doesn't.

We show in the proof of theorem 2 that there is a unique time-invariant distribution associated to the transition kernel $\mathcal{K}^*_{mM}$. We additionally show that the mM-MCMC method with indirect reconstruction converges to this distribution under the same conditions of theorem 1.

---

**Algorithm 2** The mM-MCMC method with indirect reconstruction.

---

Given parameters $\lambda > 0$, $\delta t > 0$, $K \in \mathbb{N}$, microscopic samples $x_n$, $n = 1, 2, \dots$ and reaction coordinate values $z_n$, $n = 1, 2, \dots$, with $z_0 = \xi(x_0)$.

(i) **Macroscopic Proposal**: Generate a reaction coordinate proposal $z' \sim q_0(\cdot | z_n)$.

(ii) **Macroscopic Accept/Reject**

- Accept the macroscopic proposal with probability

$$\alpha_{CG}(z'|z_n) = \min\left\{1, \frac{\bar{\mu}_0(z')\, q_0(z_n|z')}{\bar{\mu}_0(z_n)\, q_0(z'|z_n)}\right\}$$

- On acceptance, proceed to step (ii), otherwise set $(x_{n+1}, z_{n+1}) = (x_n, z_n)$ and repeat step (i).

(iii) **Indirect Reconstruction**:

- Perform $K$ MALA steps of the biased simulation (4.12) with step size $\delta t$ and initial condition $x_{n,0} = x_n$ and $k = 0, \dots, K-1$

$$x_{n,k+1} = x_{n,k} - \nabla\left(V(x_{n,k}) + \frac{\lambda}{2}\,\|\xi(x_{n,k}) - z'\|^2\right)\delta t + \sqrt{2\beta^{-1}\delta t}\,\eta_{n,k}$$

- Define the reconstructed microscopic sample $x' = x_{n,K}$.

(iv) **Microscopic Accept/Reject**

- Accept the microscopic sample with probability

$$\alpha_F(x', z'|x_n, z_n) = \min\left\{1, \frac{\bar{\mu}_0(z_n)}{\bar{\mu}_0(z')}\frac{\mu_\lambda(z')}{\mu_\lambda(z_n)}\right\}.$$

- On acceptance, set $(x_{n+1}, z_{n+1}) = (x', z')$ and return to step (i) for the next microscopic sample, otherwise set $(x_{n+1}, z_{n+1}) = (x_n, z_n)$ and generate a new reaction coordinate value in step (i).
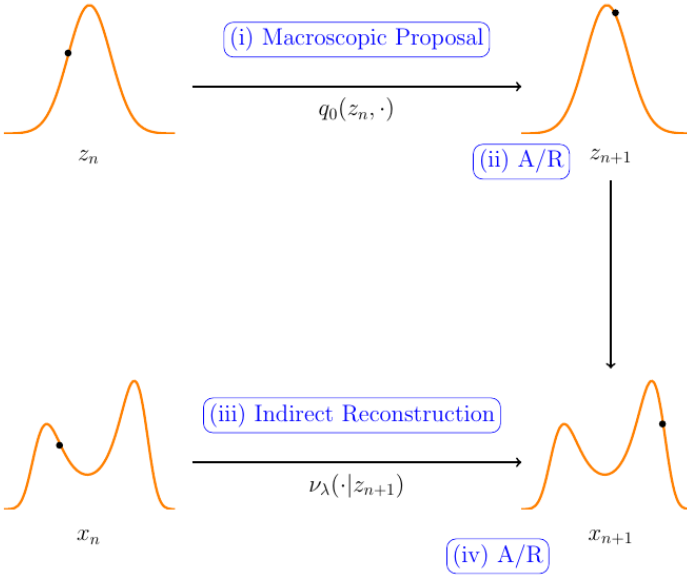
---

Figure 4.2: Visualization of one step of mM-MCMC with indirect reconstruction.

**Theorem 2.** *The mM-MCMC transition kernel $\mathcal{K}_{mM}^*$ has a unique stationary probability distribution of the form*

$$\mu_{ext}(x, z) = \sqrt{\frac{\beta\lambda}{2\pi}} \, \exp\left( -\frac{\beta\lambda}{2} \, (z - \xi(x))^2 \right) \, \mu(x). \qquad (4.16)$$

*Furthermore, under the same conditions of theorem 1,*

(i) *the transition probability kernel $\mathcal{K}_{mM}^*$ (4.15) satisfies the detailed balance condition with target measure $\mu_{ext}(x, z)$;*

(ii) *the chain $(x_n, z_n)$ is $\mu_{ext}$-irreducible;*

(iii) *the chain $(x_n, z_n)$ is aperiodic.*

Note that we can write (4.16) as

$$\mu_{\text{ext}}(x, z) = \mu(z|x)\mu(x),$$

to view $\mu(x)$ as the marginal distribution of the microscopic samples of $\mu_{\text{ext}}(x, z)$. This decomposition shows that the mM-MCMC method with indirect

reconstruction indeed samples from the microscopic Gibbs measure (1.2). The conditional distribution

$$\mu(z|x) = \sqrt{\frac{\beta\lambda}{2\pi}} \exp\left(-\frac{\beta\lambda}{2}\left(z - \xi(x)\right)^2\right)$$

is the time-invariant distribution of reaction coordinate values given a microscopic sample, and is given by a Gaussian distribution with mean $\xi(x)$ and variance $1/(\lambda\beta)$. The reaction coordinate value of $x$ is indeed close to the sampled value $z$ at the macroscopic level, which follows from the indirect reconstruction distribution. Therefore, in the limit of $\lambda$ increasing to infinity, we recover the mM-MCMC algorithm with direct reconstruction.

Additionally, we can decompose $\mu_{\text{ext}}(x, z) = \mu(z)\mu(x|z)$, i.e., as the marginal distribution of the reaction coordinate values at the macroscopic level multiplied by the conditional distribution of the microscopic samples, one can easily see that

$$\mu(z) = \mu_\lambda(z),$$

$$\mu(x|z) = \nu_\lambda(x; z)$$

so that the marginal distribution of the reaction coordinate values alone is not the free energy $\mu_0$ (2.15), but the filtered distribution $\mu_\lambda$ obtained by the Weierstrass transform (4.11).

*Proof.* We first show that the detailed balance condition holds because this is a sufficient condition for $\mu_{\text{ext}}$ to be a stationary distribution of $\mathcal{K}_{mM}^*$. The combination of statements (ii) and (iii) is then sufficient for $\mu_{\text{ext}}$ to be unique and to show that the mM-MCMC scheme with indirect reconstruction is ergodic and hence converges to $\mu_{\text{ext}}(x, z)$ in total variation norm.

(i) If we denote the total normalization constant $Z_V \sqrt{\frac{2\pi}{\beta\lambda}}$ of $\mu_{\text{ext}}(x, z)$ by $Z_\lambda$, the detailed balance condition reads

$$\mathcal{K}^*_{mM}(x', z'|x, z)\, \mu_{\text{ext}}(x, z) =$$

$$Z_\lambda^{-1} \min\left\{1, \frac{\bar{\mu}_0(z)\mu_\lambda(z')}{\bar{\mu}_0(z')\mu_\lambda(z)}\right\} \frac{\exp\left(-\beta V(x')\right)\exp\left(-\frac{\beta\lambda}{2}(\xi(x') - z')^2\right)}{\mu_\lambda(z')}$$

$$\times \min\left\{1, \frac{\bar{\mu}_0(z')q_0(z|z')}{\bar{\mu}_0(z)q_0(z'|z)}\right\} q_0(z'|z)\exp\left(-\beta V(x)\right)\exp\left(-\frac{\beta\lambda}{2}(\xi(x) - z)^2\right)$$

$$= Z_\lambda^{-1} \min\left\{\frac{\mu_\lambda(z')^{-1}}{\bar{\mu}_0(z)}, \frac{\mu_\lambda(z)^{-1}}{\bar{\mu}_0(z')}\right\} \exp\left(-\frac{\lambda\beta}{2}\left((\xi(x') - z')^2 + (\xi(x) - z)^2\right)\right)$$

$$\times \exp\left(-\beta V(x) - \beta V(x')\right) \min\left\{q_0(z'|z)\bar{\mu}_0(z), q_0(z|z')\bar{\mu}_0(z')\right\}$$

The above expression is indeed symmetric when interchanging $(x, z)$ and $(x', z')$ and hence detailed balance is satisfied. As a consequence, this fact implies the first claim of the theorem.

(ii) To prove $\mu_{\text{ext}}$-irreducibility, take a set measurable set $A \subset \mathbb{R}^{3D} \times \mathbb{R}^m$ with $\mu_{ext}(A) > 0$ and take an $(x, z) \in \mathbb{R}^{3D} \times \mathbb{R}^m$. We then have that

$$\mathcal{K}^*_{mM}(A|x, z) \geq \int_{A \setminus \{(x,z)\}} \alpha_F(x', z'|x, z)\, \nu_\lambda(x'; z')\, \alpha_{CG}(z'|z)\, q_0(z'|z)\, dx'\, dz'.$$

By assumption, $\nu_\lambda(x'; z'), q_0(z'|z), \alpha_{CG}(z'|z)$ and $\alpha_F(x', z'|x, z)$ are strictly positive everywhere, proving that $\mathcal{K}^*_{mM}(A|x, z) > 0$ for all $(x, z) \in \mathbb{R}^{3D} \times \mathbb{R}^m$.

(iii) Finally, we establish aperiodicity of the Markov chain by deriving a contradiction. Assume that for all $(x, z) \in \mathbb{R}^{3D} \times \mathbb{R}^m$, $\mathcal{K}^*_{mM}(\{(x, z)\}|x, z) = 0$. Then, the equality

$$\int_{\mathbb{R}^{3D}} \int_{\mathbb{R}^m} \alpha_F(x', z'|x, z)\, \nu_\lambda(x'; z')\, \alpha_{CG}(z'|z)\, q_0(z'|z)dz'\, dx' = 1,$$

implies the acceptance probabilities $\alpha_F(x', z'|x, z)$ and $\alpha_{CG}(z'|z)$ should be 1 everywhere because $\nu_\lambda(x'; z')$ and $q_0(z'|z)$ are strictly positive on the domain by assumption. This fact implies that the proposal distributions $\nu_\lambda(x'; z')$ and $q_0(z'|z)$ sample from the correct invariant distribution $\mu_{ext}$ without rejections. This is not the case, resulting in a contradiction. $\qquad\square$

## 4.2.6 Optimal Selection of Reconstruction Parameters

There is still some freedom regarding the choice of parameters in the indirect reconstruction scheme:

- What are optimal values for $\lambda$ and $\delta t$ that maximize the efficiency of indirect reconstruction?

- How many biased steps $K$ are required to overcome the burn-in effect during the biased simulation and to form a sample of (4.10)?

We give more context to these questions in this section, and propose heuristics for an efficient implementation.

**Optimal values of $\lambda$ and $\delta t$**   There are two limit cases for the choice of $\lambda$. In the limit of $\lambda$ decreasing to 0, the Gaussian factor in the reconstruction distribution (4.10) vanishes, and the biased simulation is then nothing more than the microscopic MALA method with invariant distribution $\mu$. We then make no use at all of the sampled reaction coordinate value $z$ at the macroscopic level, and we may then expect to obtain no efficiency gain. In this limit, we effectively annihilate the potential speed-up generated by going through the macroscopic level.

In the other limit of $\lambda \to \infty$, $\lambda$ is much larger than the stiffest modes of the potential energy $V$, we need many biased steps for the reaction coordinate to get close to the sampled value $z$ due to stiffness of the resulting biased dynamics. The microscopic time step $\delta t$ must be on the order of $1/\lambda$ due to the stability restrictions on the Euler-Maruyama scheme (4.12). One needs many biased time steps to achieve a thorough mixing in the microscopic state space. In this case, the indirect reconstruction scheme is also not efficient any more

This effect is also visible in another manner. When $\lambda$ increases to infinity, we have in the limit on $\mathbb{R}^{3D}$

$$\exp\left(-\beta V(x)\right) \sqrt{\frac{\lambda\beta}{2\pi}} \exp\left(-\frac{\lambda\beta}{2} \left\|\xi(x) - z\right\|^2\right) dx \to$$

$$\exp\left(-\beta V(x)\right) \delta_{\xi(x)-z}(dx') =$$

$$\exp\left(-\beta V(x)\right) \left\|\nabla\xi(x)\right\|^{-1} d\sigma_{\Sigma(z)}(x),$$

in point-wise sense because the Gaussian factor $\exp(-\frac{\lambda\beta}{2} \left\|\xi(x) - z\right\|^2)$ converges to $\delta_{\xi(x)-z}$. We are then effectively sampling the direct reconstruction distribution $\nu(\cdot\,|\,z)$ on the set $\Sigma(z)$, which we know to be inefficient in many situations.

Making a trade-off between these two limit effects, we propose to choose $\lambda$ close to the stiffest mode in the potential energy $V$. This choice should ensure a good mixing in all components of the system, while the microscopic samples $x_{n,k}$ will quickly have a reaction coordinate value close to $z$. In addition, we choose $\delta t$ on the order of $1/\lambda$ so that we quickly converge to $z$, while maintaining a stable biased simulation scheme. We will show numerically in chapter 5 that there is usually a range of values for $\lambda$ that give a maximal, or close to maximal, efficiency gain.

**Number of biased steps K** Deciding on the 'optimal' number of biased time steps $K$ is a theoretically hard problem, because it depends on the step size $\delta t$ and the strength of the basing potential $\lambda$. Indeed, for any finite number of time steps $K$, the microscopic sample obtained by indirect reconstruction is not in equilibrium with respect to the indirect reconstruction distribution $\nu_\lambda(\cdot; z)$. There is always a bias present because the initial condition $x_n$ can lie far from $\Sigma(z)$ and this bias may heavily depend on $\lambda$ and $\delta t$. However, we can make this bias arbitrarily small by making $K$ large enough. Usually, we observe numerically that 5 to 10 biased time-steps with accept/reject stage is a good trade-off between reducing the effect of burn-in in the biased simulation, and efficiency of the resulting method. We leave an investigation of the bias and the optimal parameters of indirect reconstruction for further work.

## 4.3   Proofs of Invariance and Ergodicity

In this final section, we relate the rate of convergence of mM-MCMC with direct reconstruction to the rate of convergence of the macroscopic sampler. In contrast to theorem 1, we assume that the image $H \subset \mathbb{R}^m$ of $\xi$ is compact and that the exact reconstruction property (4.5) holds. We are not aware whether a similar result holds when the exact reconstruction property does not hold or when $H$ is not compact. The proof relies on an expression for the mM-MCMC transition kernel that we derive in appendix A. The remainder of the proof is a straightforward calculation.

**Theorem 3.** *Assume that the exact reconstruction property* (4.5) *holds, that* $\nu(\cdot|z)$ *is bounded from above, and that the image of the reaction coordinate $H$ is compact. Then, there exist positive constants $\eta$ and $\kappa < 1$ such that*

$$\|\mathcal{K}_{mM}^n(\cdot|x) - \mu(\cdot)\|_{TV} \leq \|\mathcal{D}^n(\cdot|\xi(x)) - \mu_0(\cdot)\|_{TV} + \eta\,\kappa^n, \qquad (4.17)$$

*for all $x \in \mathbb{R}^d$. Here, $\mathcal{K}_{mM}^n(x'|x)$ is the probability* (4.6) *of reaching the microscopic sample $x'$ after $n$ mM-MCMC steps with initial value $x$. Similarly,*

$\mathcal{D}^n(\xi(x')|\xi(x))$ *is the probability* (4.2) *of reaching reaction coordinate value* $\xi(x')$ *after $n$ steps of the macroscopic MCMC sampler with initial value* $\xi(x)$.

**Proof.** In appendix A, we prove that we can write the $n-$th iteration of the mM-MCMC transition kernel with direct reconstruction as

$$\mathcal{K}^n_{mM}(x'|x) = \nu(x'|\xi(x')) \, \mathcal{D}^n(\xi(x')|\xi(x)) + \mathcal{C}(\xi(x))^n \left( \delta_{x'-x} - \nu(x'|\xi(x')) \, \delta_{\xi(x')-\xi(x)} \right),$$
(4.18)

with $C(\xi(x)) = 1 - \int_H \alpha_{CG}(z|\xi(x)) \, q_0(z|\xi(x)) dz$. This form for the $n$-th iteration of the micro-macro transition kernel has the advantage that it can be written as the product of the reconstruction distribution with the $n$-th iterate of the macroscopic invariant distribution plus another term. This form will come in handy later during the proof when we need to integrate this micro-macro transition kernel.

Subtracting the invariant measure $\mu(x') = \nu(x'|\xi(x')) \, \mu_0(\xi(x'))$ from this expression and defining the constant $\eta$ as

$$\sup_{x \in \mathbb{R}^d} \left\| \delta(\cdot - x) + \nu(\cdot|\xi(\cdot)) \, \delta_{\xi(\cdot)-\xi(x)} \right\|_{TV},$$

we can bound the total variation distance as

$$\|\mathcal{K}_{mM}(\cdot|x) - \mu(\cdot)\|_{TV} \leq \|\nu(\cdot|\xi(\cdot)) \left( \mathcal{D}^n(\cdot|\xi(x)) + \mu_0(\cdot) \right)\|_{TV}$$

$$+ \left\| \mathcal{C}(\xi(x))^n \left( \delta_{x'-x} - \nu(x'|\xi(x')) \, \delta_{\xi(x')-\xi(x)} \right) \right\|_{TV}$$

$$\leq \|\nu(\cdot|\xi(\cdot)) \left( \mathcal{D}^n(\xi(\cdot)|\xi(x)) - \mu_0(\cdot) \right)\|_{TV} + \mathcal{C}(\xi(x))^n \, \eta.$$
(4.19)

Note that the constant $\eta$ is finite because $\nu$ is bounded and the total variation distance of a delta function is 1.

We now bound each of the two terms in (4.19) independently. Using a property of the total variation distance [83, Prop. 3(a)], we rewrite the first term as

$$\left\|\nu(\cdot|\xi(\cdot))\Big(\mathcal{D}^n(\xi(\cdot)|\xi(x)) - \mu_0(\cdot)\Big)\right\|_{TV} =$$

$$\sup_{g:\ \mathbb{R}^d \to [0,1]} \left|\int_{\mathbb{R}^d} g(y)\ \nu(y|\xi(y))\Big(\mathcal{D}^n(\xi(y)|\xi(x)) - \mu_0(\xi(y))\Big)dy\right|$$

$$= \sup_{g:\ \mathbb{R}^d \to [0,1]} \left|\int_H \big(\mathcal{D}^n(z|\xi(x)) - \mu_0(z)\big)\int_{\Sigma(z)} \frac{g(x)\nu(x|z)}{\|\nabla\xi(x)\|}d\sigma_{\Sigma(z)}(x)dz\right|$$

$$= \sup_{g:\ \mathbb{R}^d \to [0,1]} \left|\int_H \mathbb{E}_\nu[g](z)\left(\mathcal{D}^n(z|\xi(x)) - \mu_0(z)\right)dz\right|$$

$$\leq \sup_{\tilde{g}:\ H \to [0,1]} \left|\int_H \tilde{g}(z)\left(\mathcal{D}^n(z|\xi(x)) - \mu_0\right)dz\right|$$

$$= \|\mathcal{D}^n(\cdot|\xi(x)) - \mu_0(\cdot)\|_{TV}.$$

The first equality stems from the definition of total variation norm [83], and the second equality is due to the co-area formula (2.13). Indeed, we have that the reconstruction distribution is given by $\nu(y|\xi(y)) = \mu(y)/\mu_0(\xi(y))$ inside the integral over $\mathbb{R}^d$ (2.19), and this expression becomes $\mu(y)/\mu_0(z)\|\nabla\xi(y)\|^{-1}$ inside the integral over $\Sigma(z)$. On the third line, we define the temporary variable

$$\mathbb{E}_\nu[g](z) = \int_{\Sigma(z)} \frac{g(x)\nu(x|z)}{\|\nabla\xi(x)\|}d\sigma_{\Sigma(z)}(x) \in [0,1],$$

and the inequality on the fourth line is because we take the supremum over a possibly larger class of bounded functions between 0 and 1.

To bound the the factor $\mathcal{C}(\xi(x))$ in the second term of (4.19), we use compactness of $H$. The functions $q_0$ and $\alpha_{CG}$ are strictly positive and hence the integral

$$\int_H \alpha_{CG}(y|\xi(x))\ q_0(y|\xi(x))dy$$

is strictly greater than 0 for every $\xi(x) \in H$. By compactness, the infimum of this integral for all $\xi(x)$ is hence also strictly positive, proving that the supremum of $\mathcal{C}(\xi(x))$ is strictly smaller than 1, i.e., $\sup_{x\in\mathbb{R}^d}\mathcal{C}(\xi(x)) = \kappa < 1$. Putting both bounds together, we conclude that

$$\|\mathcal{K}_{mM}^n(\cdot\mid x) - \mu(\cdot)\|_{TV} \leq \|\mathcal{D}^n(\xi(\cdot)|\xi(x)) - \mu_0(\cdot)\|_{TV} + \eta\ \kappa^n,$$

proving the theorem. □

A consequence of theorem 3 is that mM-MCMC with direct reconstruction inherits all ergodicity properties from the macroscopic MCMC sampler. For instance, when $\mathcal{D}$ is uniformly ergodic, that is, we can bound $\|\mathcal{D}^n(\cdot|z) - \mu_0(z)\|_{TV}$ by $C\rho^n$ with $C > 0$ and $\rho < 1$, then the mM-MCMC method is also uniformly ergodic since we can bound

$$\|\mathcal{K}_{mM}(\cdot|x) - \mu(\cdot)\|_{TV} \le (C + \eta)\max\{\rho, \kappa\}^n.$$

We can draw a similar conclusion for other types of ergodicity of the macroscopic sampler. We will numerically show in chapter 5, that the mM-MCMC method with exact reconstruction is more efficient than when the exact reconstruction property is not satisfied.

# Chapter 5

# Numerical Results

Measuring the efficiency gain of micro-macro Markov chain Monte Carlo is non-trivial. Many method and model parameters affect said efficiency. Just to name a couple: On the macroscopic level, the approximate macroscopic distributions determines the regions of high probability. If those regions do not line up with the microscopic regions of high probability, the efficiency is lower. Then there is reconstruction. With direct reconstruction, a good reconstruction distribution is one that is easily evaluated and results in a high microscopic acceptance probability. With indirect reconstruction, the strength of biasing potential, the time step and the number of time steps all have an effect on the mixing of the biased simulation, and thus its efficiency.

In this chapter, we carefully study how the efficiency gain depends on mM-MCMC method parameters. We first define the efficiency gain criterium in section 5.1. We will use this formulation for the remainder of the manuscript. Then, we study the mM-MCMC on three molecular examples. The three-atom molecule in section 5.2, butane in section 5.3, and alanine-dipeptide in section 5.4. All experiments and timings were performed on a Xeon Gold 6140 CPU with 2.3 GHz (Skylake). The contents and results of this chapter are included in our latest preprint [91], and the text below has been largely taken from this manuscript.

## 5.1 Efficiency Gain Measure

Consider a scalar function $F : \mathbb{R}^{3D} \to \mathbb{R}$ and suppose we are interested in the average of $F$ with respect to the Gibbs measure $\mu$,

$$\mathbb{E}_\mu[F] = \int_{\mathbb{R}^{3D}} F(x) \ d\mu(x) \tag{5.1}$$

If we sample the invariant measure $\mu$ using an MCMC method, we can estimate (5.1) by $\hat{F}^{(l)} = N^{-1} \sum_{n=1}^{N} F(x_n^{(l)})$ with an ensemble of microscopic samples $\{x_n^{(l)}\}_{n=1}^{N}$. To obtain an accurate estimate, we make multiple, independent estimates $\hat{F}^{(l)}$, with $l = 1, \ldots, L$. The mean squared error (MSE) of the sampler is then given by

$$\mathrm{MSE}[\hat{F}] = \frac{\sum_{l=1}^{L} \left( \mathbb{E}_\mu[F] - \hat{F}^{(l)} \right)^2}{L}. \tag{5.2}$$

For the following numerical experiments in this chapter, we are interested in reducing the mean squared error of the estimator $\hat{F}$ by use of the micro-macro Markov chain Monte Carlo method. Key here is that for good comparison, both the mM-MCMC and microscopic methods need to approximately have the same runtime. That is, the runtime of both methods must be accounted for. A good efficiency gain measure of mM-MCMC over the microscopic method is

$$\mathrm{Gain} = \frac{\mathrm{MSE}[\hat{F}]_\mathrm{micro}}{\mathrm{MSE}[\hat{F}]_\mathrm{mM}} \frac{T_\mathrm{micro}^N}{T_\mathrm{mM}^N}. \tag{5.3}$$

Here, $T_\mathrm{micro}^N$ is the measured execution time of the microscopic MCMC method for a fixed number of samples $N$, and $T_\mathrm{mM}^N$ is the measured execution time for the same number of samples $N$ of mM-MCMC.

How large can we expect this efficiency gain to be? In case of mM-MCMC with direct reconstruction, typically $T_\mathrm{mM}^N < T_\mathrm{micro}^N$ because most samples are rejected at the macroscopic level. Additionally, direct reconstruction is cheap. With indirect reconstruction, this picture changes. The biased simulation with generates $K$ microscopic samples, and defeats the runtime gain at the macroscopic level. From experiments, we can expect $T_\mathrm{mM}^N \approx 2T_\mathrm{micro}^N$. In both cases, the ratio $\frac{\mathrm{MSE}[\hat{F}]_\mathrm{micro}}{\mathrm{MSE}[\hat{F}]_\mathrm{mM}}$ determines most of the efficiency gain.

## 5.2 The Three-Atom Molecule

In this section, we consider the mM-MCMC algorithm on a simple, academic, three-atom molecule, as introduced in [59]. The three-atom molecule has a
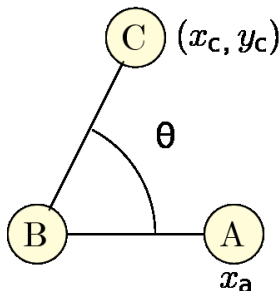
Figure 5.1: The three-atom molecule. Atom $A$ is constraint to the $x$-axis with $x$-coordinate $x_a$, atom $B$ is fixed at the origin of the plane and atom $C$ lies om the two-dimensional plane with Cartesian coordinates $(x_c, y_c)$.

central atom $B$, that we fix at the origin of the two-dimensional plane, and two outer atoms, $A$ and $C$. To fix the superfluous degrees of freedom, we constrain atom $A$ to the $x-$axis, while $C$ can move freely in the plane. The three-atom molecule is depicted on Figure 5.1. The potential energy for the three-atom system consists of three terms,

$$V(x_a, x_c, y_c) = \frac{1}{2\varepsilon}\,(x_a - 1)^2 + \frac{1}{2\varepsilon}\,(r_c - 1)^2 + \frac{208}{2}\left(\left(\theta - \frac{\pi}{2}\right)^2 - 0.3838^2\right)^2, \quad (5.4)$$

where $x_a$ is the $x-$coordinate of atom $A$ and $(x_c, y_c)$ are the Cartesian coordinates of atom $C$. The bond length $r_c$ between atoms $B$ and $C$ and the angle $\theta$ between atoms $A$, $B$ and $C$ are defined as

$$r_c = \sqrt{x_c^2 + y_c^2}$$

$$\theta = \arctan(y_c, x_c).$$

The first term in (5.4) describes the oscillatory potential energy of the bond between atoms $A$ and $B$, with equilibrium bond length 1. Similarly, the second term describes the oscillatory energy of the bond between atoms $B$ and $C$ with bond length $r_c$. Finally, the third term determines the potential energy of the angle $\theta$ between the two outer atoms, which has an interesting bimodal behavior. The distribution of $\theta$ has two peaks, one at $\frac{\pi}{2} - 0.3838$ and another at $\frac{\pi}{2} + 0.3838$.

The reaction coordinate that we consider in this section is the angle $\theta$, i.e.,

$$\xi(x_a, x_c, y_c) = \theta(x_a, x_c, y_c), \quad (5.5)$$

since this variable is the slow component of the three-atom molecule. Additionally, the angle $\theta$ is also independent of the time-scale separation, given by $\varepsilon$.

Finally, from the potential energy (5.4), it can be easily seen that the free energy of the reaction coordinate $\theta$ reads

$$A(\theta) = \frac{208}{2} \left( \left(\theta - \frac{\pi}{2}\right)^2 - 0.3838^2 \right)^2. \qquad (5.6)$$

**Outline of Experiments**  We perform experiments on the three-atom molecule with both mM-MCMC with direct reconstruction (section 5.2.1) and indirect reconstruction (section 5.2.2).

For mM-MCMC with direct reconstruction, we start with a visual inspection of the mM-MCMC sampling results in section 5.2.1, in conjunction with comparing with the microscopic MALA method. Then we compare direct reconstruction mM-MCMC with a macroscopic MCMC sampler. Thirdly, we study the efficiency gain of mM-MCMC with direct reconstruction as a function of $\varepsilon$. Afterwards, we show the efficiency gain for different approximate macroscopic and proposal distribution Finally, we do the same for the same approximate macroscopic and direct reconstruction distributions.

For mM-MCMC with indirect reconstruction, we first identify the optimal value for $\lambda$ in section 5.2.2. With this value, we visually inspect the sampling result of mM-MCMC and compare it to the microscopic sampler. Afterwards, we study how the efficiency gain scales with $\varepsilon$, followed by a comparison between mM-MCMC with direct and indirect reconstruction.

## 5.2.1   Experiments with Direct Reconstruction

### Visual inspection of the performance of mM-MCMC

The time-invariant distribution of the reaction coordinate

$$\mu_0(\theta) \propto \exp\left(-\beta A(\theta)\right), \quad A(\theta) = \frac{208}{2} \left( \left(\theta - \frac{\pi}{2}\right)^2 - 0.3838^2 \right)^2. \qquad (5.7)$$

is easy to evaluate and to sample from. Therefore, we choose to use this distribution as the approximate macroscopic distribution, i.e., $\bar{\mu}_0 = \mu_0$.

When an angle $\theta$ is given, for example after the macroscopic proposal step in the mM-MCMC method, we need to reconstruct the position $x_a$ of atom $A$, and the bond length $r_c$ between atoms $B$ and $C$. To maximize the efficiency gain of mM-MCMC, we choose the reconstruction distribution such that the microscopic acceptance rate is always 1. We thus choose as reconstruction distribution the exact reconstruction distribution (2.19), i.e., $\bar{\nu} = \nu$.

For the three-atom molecule, this reconstruction distribution takes the form

$$\nu(x_a, r_c|\ \theta) \propto \exp\left(-\frac{\beta}{2\varepsilon}\ (x_a - 1)^2\right) \exp\left(-\frac{\beta}{2\varepsilon}\ (r_c - 1)^2\right), \qquad (5.8)$$

which is defined on the sub-manifold $\Sigma(\theta) = \{(x_a, x_c, y_c) \in \mathbb{R}^3|\ x_c = r\cos(\theta), y_c = r\sin(\theta), r \geq 0\}$ of constant reaction coordinate value. While sampling the reconstruction distribution, we also need to make sure that $r_c$ is positive. First sample $r_c$ from the Gaussian distribution with mean 1 and variance $\varepsilon\beta^{-1}$ and reject the proposal whenever $r_c$ is negative. If the proposed value is positive, we accept.

**Experimental Setup**   On the macroscopic level, we use a MALA method based on an overdamped Langevin method to compute samples of $\theta$, with time step $\Delta t = 0.01$,

$$d\theta = -\nabla A(\theta)dt + \sqrt{2\beta^{-1}}dW, \qquad (5.9)$$

The initial condition of the Markov chain is $(x_a, x_c, y_c) = (1, 0, 1)$.

For this numerical experiment, we choose two values for the time-scale parameter, $\varepsilon = 10^{-4}$ and $\varepsilon = 10^{-6}$ and sample the three-atom molecule with $N = 10^6$ microscopic samples. For simplicity, we define the inverse temperature parameter as $\beta = 1$. The proposals of the microscopic MALA method are based on the Euler-Maruyama discretization of the overdamped Langevin dynamics (6.6) with microscopic time step size $\delta t = \varepsilon$. On Figure 5.2, we visually compare the histograms of $\theta$ obtained by MALA and mM-MCMC for both parameters.

**Numerical Results**   The sampling results of the microscopic MALA method are not good. The method remains stuck for a long time in either potential well. The mixing is better for larger $\varepsilon$ when the time-scale separation disappears. The mM-MCMC method is able to sample the distribution of $\theta$ accurately, regardless of the time-scale separation.

### Comparison with the macroscopic reaction coordinate sampler

In the second illustration, we compare the mM-MCMC method to MCMC samplers at the macroscopic level. That is, if the approximate macroscopic distribution $\bar{\mu}_0$ differs significantly from the exact marginal distribution of the reaction coordinate $\mu_0$ (5.6), any macroscopic MCMC sampler would not sample $\mu_0$ at all. However, the mM-MCMC method doesn't readily accept reaction coordinate values sampled at the macroscopic level, but performs
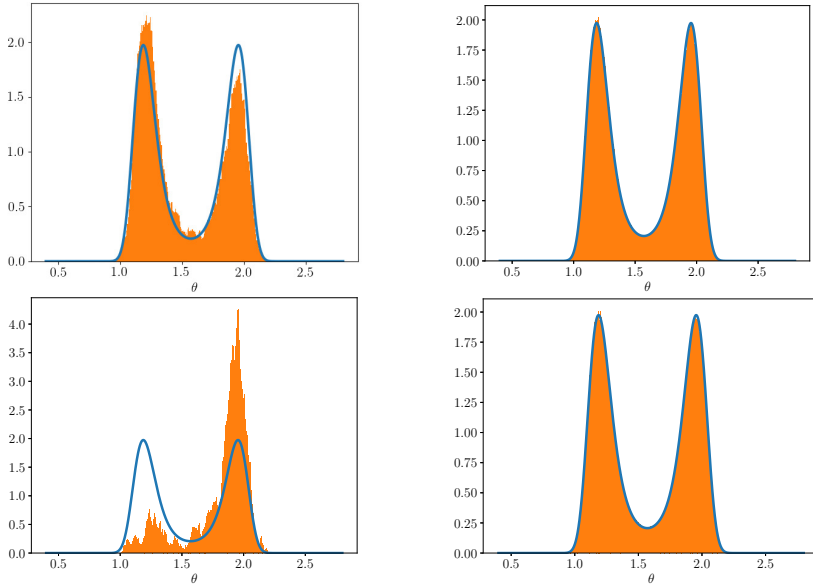
Figure 5.2: Visual representation of the histogram of $\theta$ of the microscopic MCMC method (MALA) (left) and mM-MCMC with direct reconstruction (right) on the three-atom molecule, with reaction coordinate $\theta$. The simulation parameters are $\varepsilon = 10^{-4}$ (top) and $\varepsilon = 10^{-6}$ (bottom) and the number of samples is $N = 10^6$. The MALA method remains stuck in the potential well of $\theta$ around $\pi/2 + 0.3838$, while the mM-MCMC method samples the distribution well.

a second accept/reject step at the microscopic level. These reconstructed microscopic samples are then samples from the microscopic Gibbs measure $\mu$. The reaction coordinate values generated by mM-MCMC should sample the marginal distribution $\mu_0$ exactly.

**Experimental setup** We verify this claim by defining two approximate free energy functions $\bar{A}$ that define two approximate macroscopic distributions $\bar{\mu}_0$. These approximate free energy functions are

$$\bar{A}^2(\theta) = \frac{208}{2} \left( \left( \theta - \frac{\pi}{2} \right)^2 - 0.4838^2 \right)^2, \tag{5.10}$$

$$\bar{A}^3(\theta) = \frac{208}{2} \left( \left( \theta - \frac{\pi}{2} \right)^2 - 0.3838^2 \right)^2 + \cos(\theta).$$
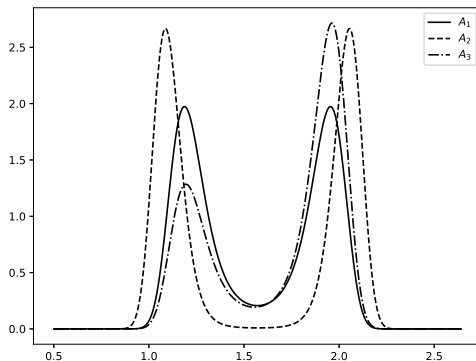
Figure 5.3: The three approximate free energy distributions as defined in equation (5.6). The full curve is the exact free energy of the three-atom molecule (5.4). The dashed curve is obtained by shifting the local maxima of the exact distribution by 0.1. The third dash-dotted curve is the distribution perturbed by the cosine function.

The first expression for the approximate free energy is obtained by perturbing the two peaks of the exact free energy (5.6) by a distance 0.1 of radians. The second formula, $\bar{A}^3$ perturbs the exact free energy expression by adding the cosine function to $A$, resulting in a large perturbation on the amplitude of the associated macroscopic invariant distribution. The effect of the cosine perturbation is that the height of the left peak in the macroscopic distribution of $\theta$ is decreased, while the height of the right peak is increased. The three Gibbs distributions associated to the correct free energy and the two approximate free energies defined above are shown in Figure 5.3.

In the experiments below, we investigate the sampling results of the mM-MCMC with each of the above defined approximate macroscopic distributions. On the macroscopic level, we use the Langevin dynamics with time step $\Delta t = 0.01$ to generate reaction coordinate values according to each of the approximate macroscopic distributions. For reconstruction, we use the time-invariant reconstruction distribution $\nu$ (2.19). The time-scale parameter is $\varepsilon = 10^{-6}$, and we use $N = 10^6$ sampling steps. The sampling results are shown in figure 5.4.

**Numerical results**  The mM-MCMC method is in both cases able to recover the marginal distribution of the reaction coordinate. A solely macroscopic sampler would return samples that follow the approximate macroscopic distribution, not
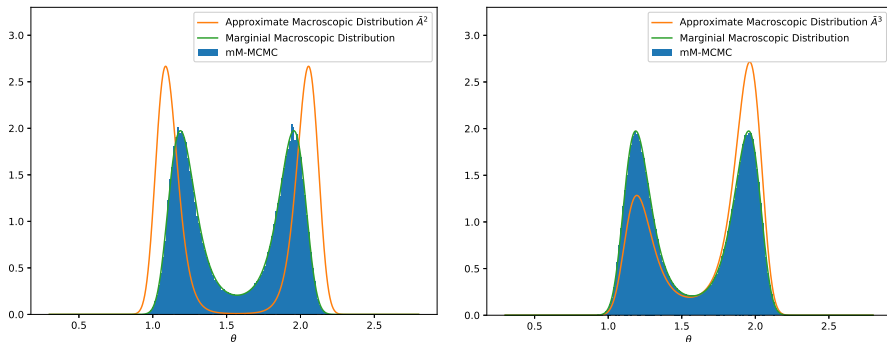
Figure 5.4: Histogram of the mM-MCMC method (blue) for two different approximate macroscopic distributions: $\bar{A}^2$ (left) and $\bar{A}^3$ (right). The orange line depicts the approximate macroscopic distribution used in each experiment, and the green line represents the marginal distribution of the reaction coordinate values. We can see that the mM-MCMC is able to overcome a bad approximate macroscopic distribution and samples the marginal distribution correctly.

the free energy. There is need for a correction on the microscopic level, and the mM-MCMC method is able to do so.

**Efficiency gain as a function of $\varepsilon$**

**Experimental setup**   In this section, we study the efficiency gain of mM-MCMC on the estimated mean and variance of the angle $\theta$, as a function of $\varepsilon$. We consider two choices for the approximate macroscopic distribution $\bar{\mu}_0$ and the reconstruction distribution $\bar{\nu}$. For the first choice, we take the exact free energy $\bar{A}^1$ (5.6) and the exact time-invariant reconstruction distribution $\bar{\nu}^1$ (5.12) such that the microscopic acceptance rate is 1. The other choice consists of approximate free energy $\bar{A}^2$ and reconstruction distribution $\bar{\nu}^2$ defined by

$$\bar{\nu}^2(x|\theta) \propto \exp\left(-\frac{(x_a - 1)^2}{4\varepsilon}\right) \exp\left(-\frac{(r_c - 1)^2}{4\varepsilon}\right).$$

For both choices, the macroscopic proposal moves are generated by a MALA sampler with corresponding (approximate) free energy function and time step $\Delta t = 0.01$. We measure the efficiency gain for four values of the time-scale separation, $\varepsilon = 10^{-i}$, $i = 3, \ldots, 6$ and with $N = 10^6$ sampling steps. The microscopic time step for the microscopic MALA algorithm is $\delta t = \varepsilon$, and the inverse temperature is $\beta = 1$.. The numerical results are shown in Figure 5.5.
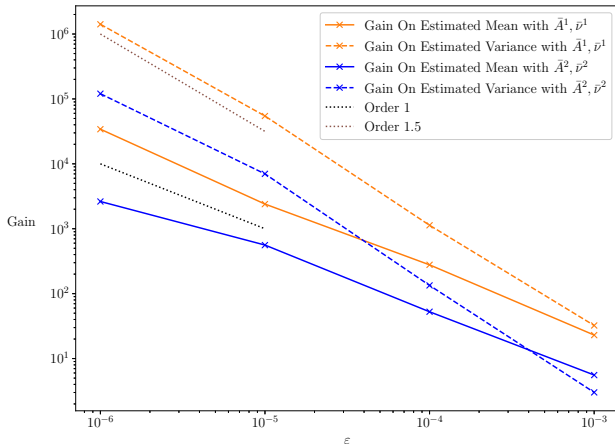
Figure 5.5: Efficiency gain of mM-MCMC over the microscopic MALA method on the estimated mean (solid line) and variance (dashed line) of the angle $\theta$ for two different parameter choices: $(\bar{A}^1, \bar{\nu}^1)$ (orange lines) and $(\bar{A}^2, \bar{\nu}^2)$ (blue lines). This efficiency gain is computed using (5.3), for $\varepsilon = 10^{-i}$, $i = 3, \ldots, 6$ and with $N = 10^6$ microscopic samples. For both parameter choices, the efficiency gain on the estimated mean increases linearly with decreasing $\varepsilon$, and the gain on the estimated variance increases faster, approximately $\varepsilon^{-1.5}$. The efficiency gain on the estimated mean and variance of $\theta$ increases slower when the exact reconstruction property does not hold, than when it does

**Numerical results**  In both experiments, the efficiency gain on the estimated mean and variance increases linearly or faster with decreasing $\varepsilon$. This result shows that mM-MCMC can improve the sampling of systems with a medium to large time-scale separation. Additionally, in case the exact reconstruction property (4.5) holds, the efficiency gain is larger than when the exact reconstruction property does not hold.. This numerical result is an illustration of Theorem 3, which states that the rate of convergence of mM-MCMC is identical to the rate of convergence of the macroscopic sampler when the exact reconstruction property holds.

Currently, we have no theoretical bounds on how the efficiency gain depends on the time-scale separation. We can only observe the numerical results in this thesis. A good point to start such an analysis would be study how the variance on the estimated mean of $\theta$ of the microscopic MALA method depends on the time-scale separation. We defer such an analysis to further research.

**Impact of $\bar{A}$ and $q_0$ on the efficiency gain**

With the knowledge that mM-MCMC can correct for an inaccurate approximate macroscopic distribution, we now numerically investigate the effect of various approximate macroscopic distributions $\bar{\mu}_0$ and of macroscopic transition distribution $q_0$ on the efficiency gain of mM-MCMC.

**Experimental setup**  For approximate macroscopic distributions, we use the exact free energy function (5.6) and the two approximate free energies defined in (5.10). In this experiments, we use two macroscopic proposal distribution $q_0$, based on an Euler-Maruyama discretization of the following two SDEs on the macroscopic level

$$q_0^1 : d\theta = -\nabla A_i(\theta)dt + \sqrt{2\beta^{-1}}dW$$

$$q_0^2 : d\theta = \sqrt{2\beta^{-1}}dW. \tag{5.11}$$

The first proposal distribution is the overdamped Langevin dynamics for each of the approximate free energy functions (MALA), while the latter equation is a simple Brownian motion in the reaction coordinate space.

In what follows, we run the mM-MCMC algorithm with each of these six combinations for approximate macroscopic invariant distribution $\bar{\mu}_0 \propto \exp\left(-\beta\bar{A}\right)$ and the macroscopic proposal distribution $q_0$ for $N = 10^6$ sampling steps and with a macroscopic time step $\Delta t = 0.01$. We also use the exact reconstruction distribution $\nu$ in each of these experiments. After each simulation, we compute the total macroscopic acceptance rate, the microscopic acceptance rate after reconstruction, the average runtime, the numerical MSE on the estimated mean of $\theta$, and the total efficiency gain of of mM-MCMC over the microscopic MALA algorithm, as explained in Section 5.1. The microscopic time step for the MALA method is $\delta t = \varepsilon$. The temperature parameter is the same, $\beta = 1$. For statistical good comparison, we average the results over 100 independent runs. The numerical efficiency gains are shown in Tables 5.1 and 5.2 for $\varepsilon = 10^{-4}$ and $\varepsilon = 10^{-6}$ respectively.

**Numerical results**  The numerical results in Tables 5.1 and 5.2 show a large efficiency gain of mM-MCMC over the microscopic MALA algorithm. The efficiency gain grows linearly with the time-scale separation $\sim \mathcal{O}(\varepsilon^{-1})$. For instance, a gain of a factor 8255 in Table 5.2 indicates that the mM-MCMC method needs 8255 times fewer sampling steps to obtain the same MSE on the estimated mean of $\theta$ than the microscopic MALA method, for the same runtime.

| Parameters | Macroscopic acceptance rate | Microscopic acceptance rate | Runtime gain | MSE gain | Efficiency gain |
|---|---|---|---|---|---|
| Langevin, $\bar{A}^1$ | 0.749932 | 1 | 2.45692 | 85.3266 | 209.64 |
| Langevin, $\bar{A}^2$ | 0.730384 | 0.432508 | 2.62306 | 28.4122 | 74.527 |
| Langevin, $\bar{A}^3$ | 0.749653 | 0.950238 | 1.95663 | 99.7186 | 195.112 |
| Brownian, $\bar{A}^1$ | 0.645188 | 1 | 3.03108 | 85.1586 | 258.122 |
| Brownian, $\bar{A}^2$ | 0.61375 | 0.597058 | 3.28045 | 35.6794 | 117.044 |
| Brownian, $\bar{A}^3$ | 0.645654 | 0.959794 | 2.72728 | 81.3405 | 221.838 |

Table 5.1: A summary of different statistics of the mM-MCMC method for six combinations of the (approximate) macroscopic invariant distribution and macroscopic proposal moves. The time-scale parameter is $\varepsilon = 10^{-4}$. We record the macroscopic and microscopic acceptance rates of mM-MCMC, the runtime and MSE gains of mM-MCMC over the microscopic MALA method, and the total efficiency gain.

| Parameters | Macroscopic acceptance rate | Microscopic acceptance rate | Runtime gain | MSE gain | Efficiency gain |
|---|---|---|---|---|---|
| Langevin, $\bar{A}^1$ | 0.749906 | 1 | 2.50343 | 3297.65 | 8255.44 |
| Langevin, $\bar{A}^2$ | 0.730538 | 0.43242 | 2.64621 | 933.64 | 2470.61 |
| Langevin, $\bar{A}^3$ | 0.749596 | 0.950308 | 1.99301 | 2461.08 | 4904.96 |
| Brownian, $\bar{A}^1$ | 0.645272 | 1 | 3.058 | 3223.1 | 9856.26 |
| Brownian, $\bar{A}^2$ | 0.613716 | 0.59717 | 3.31087 | 1274.01 | 4218.07 |
| Brownian, $\bar{A}^3$ | 0.645702 | 0.959848 | 2.77357 | 3229.06 | 8956.03 |

Table 5.2: A summary of different statistics of the mM-MCMC method for six combinations of the (approximate) macroscopic invariant distribution and macroscopic proposal moves. The time-scale parameter is $\varepsilon = 10^{-6}$. The columns are the same as in Table 5.1.

First, note that for both values of $\varepsilon$, the macroscopic acceptance rate is lower when using Brownian macroscopic proposals than when using Langevin dynamics proposals. This result makes intuitive sense, since the Brownian motion does not take into account the underlying macroscopic probability distribution, while the Langevin dynamics will automatically choose reaction coordinate values in regions of higher macroscopic probability. However, with a lower macroscopic acceptance rate comes a lower runtime as well, since we need to reconstruct fewer microscopic samples and hence fewer evaluations of the microscopic potential energy. This effect is indeed visible in the fourth column of both tables, where the runtime gain is higher for Brownian motion than that of the corresponding Langevin dynamics. Further, we note that the MSE gains of the Langevin and Brownian proposals are mostly the same. Hence, the total efficiency gain of mM-

MCMC is almost completely determined by the lower runtime with Brownian proposals, when the approximate distribution lies close to the true marginal distribution of the reaction coordinate. However, when there are significant differences between the approximate and exact macroscopic distributions, the choice of proposal moves does have a significant impact on the overall efficiency.

The choice of approximate free energy $\bar{A}_i$, however, has a larger impact on the efficiency of mM-MCMC with direct reconstruction. First of all, one can see that the macroscopic acceptance rate is not affected by the choice of macroscopic invariant distribution, only by the choice of macroscopic proposals $q_0$, as explained in the previous paragraph. However, the microscopic acceptance rate is significantly affected by the approximate macroscopic distribution $\bar{\mu}_0$. Since the microscopic acceptance probability is used to correct the microscopic samples from having the wrong macroscopic distribution, the more the approximate macroscopic distribution $\bar{\mu}_0$ deviates from the exact distribution $\mu_0$, the lower the microscopic acceptance rate will be. Indeed, the microscopic acceptance rate for the approximate free energy $\bar{A}^2$ is much lower than that for the correct free energy $A$, since the local minima of $\bar{A}^2$ are located at different positions. On the other hand, the approximate free energy $\bar{A}^3$ is closer to $A$ because only the aplitude of the peaks is different.

To conclude, we can split the total efficiency gain in two factors. The type of macroscopic proposals mostly determine the runtime gain, while the approximate macroscopic distribution mostly resolves the MSE and thus MSE gain.

### Impact of $\bar{A}$ and $\bar{\nu}$ on the efficiency gain

For the final experiment on the three-atom molecule with direct reconstruction, we investigate the effect of the choice of reconstruction distribution $\bar{\nu}$ on the efficiency gain of mM-MCMC over MALA, together with the same three approximate free energies $\bar{A}$ (5.10). For consistency of the numerical results, we use macroscopic proposal moves based on the overdamped Langevin dynamics (gradient descent) in $\theta$, with exact free energy $A$. The two reconstruction distributions that we consider in this section are

$$\bar{\nu}^1(x|\theta) = \nu(x|\theta)$$

$$\bar{\nu}^2(x|\theta) \propto \exp\left(-\frac{(x_a - 1)^2}{4\varepsilon}\right)\exp\left(-\frac{(r_c - 1)^2}{4\varepsilon}\right). \tag{5.12}$$

The first reconstruction distribution is the exact reconstruction distribution (5.8), while the second distribution is obtained by doubling the variance on $x_a$ and $r_c$ relative to $\nu$.

| Parameters | Macroscopic acceptance rate | Microscopic acceptance rate | Runtime gain | MSE gain | Efficiency gain |
|---|---|---|---|---|---|
| $A, \bar{\nu}^1$ | 0.749932 | 1 | 2.45692 | 85.3266 | 209.64 |
| $\bar{A}^1, \bar{\nu}^1$ | 0.730384 | 0.432508 | 2.62306 | 28.4122 | 74.527 |
| $\bar{A}^2, \bar{\nu}^1$ | 0.749653 | 0.950238 | 1.95663 | 99.7186 | 195.112 |
| $A, \bar{\nu}^2$ | 0.749891 | 0.476963 | 2.45479 | 37.5736 | 92.2354 |
| $\bar{A}^1, \bar{\nu}^2$ | 0.730482 | 0.266464 | 2.62922 | 17.2829 | 45.4405 |
| $\bar{A}^2, \bar{\nu}^2$ | 0.749654 | 0.474436 | 1.92062 | 50.257 | 96.5246 |

Table 5.3: A summary of different statistics of the mM-MCMC method for six combinations of the (approximate) macroscopic invariant distribution and reconstruction distribution. The time-scale parameter is $\varepsilon = 10^{-4}$. For each of these six combinations, we record the average acceptance rate at the macroscopic level, the average acceptance rate on the microscopic level, conditioned on all accepted macroscopic samples and the gain in runtime and variance on the estimated mean of $\theta$ of mM-MCMC over the microscopic MALA method. The final column record the total efficiency gain of mM-MCMC, which is the product of the two former columns.

**Experimental setup**  In this experiment, we again compute the macroscopic and microscopic acceptance rates, the gain in runtime for a fixed number of sampling steps, the gain in MSE of the estimated mean of $\theta$, and the total efficiency gain of mM-MCMC over the microscopic MALA algorithm for each of the six combinations of the approximate macroscopic distribution $\bar{\mu}_0$ and reconstruction distribution $\bar{\nu}$. We perform each experiment for $N = 10^6$ steps, the temperature parameter is $\beta = 1$, and the macroscopic time step is again $\Delta t = 0.01$. We use a microscopic time step $\delta t = \varepsilon$ for the MALA algorithm. For a good comparison, the numerical results are averaged over 100 independent runs. The numerical results are shown in Table 5.3 for $\varepsilon = 10^{-4}$, and in Table 5.4 for $\varepsilon = 10^{-6}$.

**Numerical results**  The choice of reconstruction distribution has a negligible effect on the macroscopic acceptance rate, but it does have a significant effect on the microscopic acceptance rate. For both values of $\varepsilon$, and for the three choices of $\bar{A}$, reconstruction distribution $\bar{\nu}^2$ results in a lower microscopic acceptance rate than $\bar{\nu}^1$. Since $\bar{\nu}^2$ is not the exact reconstruction distribution (5.8), the microscopic acceptance probability needs to correct for wrong microscopic samples, lowering the average microscopic acceptance rate. As we also noted in the previous experiment, a lower microscopic acceptance rate keeps the MSE on the estimated mean of $\theta$ high, and thus the total efficiency gain of mM-MCMC over the microscopic MALA algorithm will be smaller.

| Parameters | Macroscopic acceptance rate | Microscopic acceptance rate | Runtime gain | MSE gain | Efficiency gain |
|---|---|---|---|---|---|
| $A, \bar{\nu}^1$ | 0.749906 | 1 | 2.50343 | 3297.65 | 8255.44 |
| $\bar{A}^1, \bar{\nu}^1$ | 0.730538 | 0.43242 | 2.64621 | 933.64 | 2470.61 |
| $\bar{A}^2, \bar{\nu}^1$ | 0.749596 | 0.950308 | 1.99301 | 2461.08 | 4904.96 |
| $A, \bar{\nu}^2$ | 0.750029 | 0.47711 | 2.58486 | 1250.1 | 3231.34 |
| $\bar{A}^1, \bar{\nu}^2$ | 0.730455 | 0.266443 | 2.65114 | 547.826 | 1452.37 |
| $\bar{A}^2, \bar{\nu}^2$ | 0.749564 | 0.474496 | 1.93858 | 1396.05 | 2706.35 |

Table 5.4: A summary of different statistics of the mM-MCMC method for six combinations of the (approximate) macroscopic invariant distribution and reconstruction distribution. The time-scale parameter is $\varepsilon = 10^{-6}$. The columns are the same as in Table 5.3.

## 5.2.2 Experiments with Indirect Reconstruction

### Optimal Value of $\lambda$

Before we can use the mM-MCMC method with indirect reconstruction here, we need to determine an optimal value of $\lambda$. We claim in section 4.2.6 that a good value of $\lambda$ is the force constant corresponding to the fastest mode, in this case $\varepsilon^{-1}$. This experiment serves to verify this claim.

**Experimental setup** We compute the efficiency gain of mM-MCMC with indirect reconstruction over the microscopic MALA method for three values of the time-scale separation: $\varepsilon = 10^{-4}$, $10^{-5}$, and $10^{-6}$. For each $\varepsilon$, we select a set of $\lambda$ values, ranging from $10^{-1}\varepsilon^{-1}$ to $10^4\varepsilon^{-1}$. Indirect reconstruction parameters are $K = 5$, and $\delta t = \lambda^{-1}$. On the macroscopic level, the time step is $\Delta t = 0.01$, and proposals are generated by Langevin dynamics. The time step of the microscopic MALA method is $\varepsilon$ and we take $N = 10^6$ sampling steps. Moreover, we average the efficiency gain over 100 independent runs.

On Figure 5.6, we show the efficiency gain of mM-MCMC as a function of $\lambda \varepsilon$. Also, in Table 5.5, we gather some statistics of the mM-MCMC method for the estimated variance of $\theta$ for several values of $\lambda$ and a fixed value of $\varepsilon = 10^{-6}$.

**Numerical results** From figure 5.6, the efficiency gain increases when $\varepsilon$ decreases. The reason is simple: the higher the time-scale separation, the more gain can be made with mM-MCMC. Second, when $\lambda < 1/\varepsilon$ there is no efficiency gain at all. In this situation, the biased simulation is unable to pull the microscopic molecule towards the sampled reaction coordinate value.We
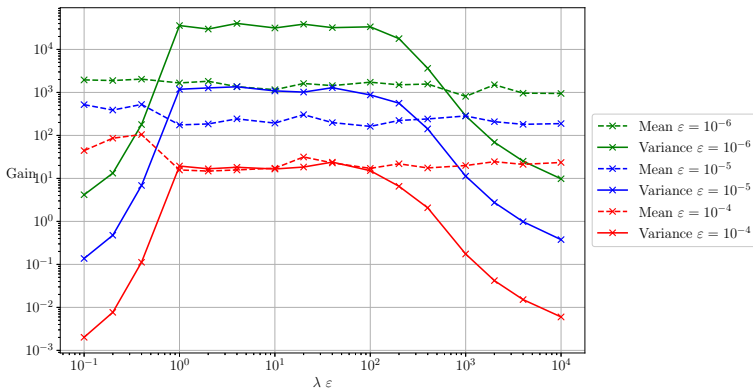
Figure 5.6: Efficiency gain of mM-MCMC on the estimated mean (full lines) and variance (dotted lines) of $\theta$, for different values of $\lambda\,\varepsilon$. When $\lambda$ is larger $1/\varepsilon$, there is a clear efficiency gain. This gain is almost constant for a big range $\lambda$ values. However, when $\lambda$ is too large, the gain starts to decrease. On the other hand, when $\lambda$ is smaller than $1/\varepsilon$, the gain also decreases.

| $\lambda \cdot \varepsilon$ | Macroscopic acceptance rate | Microscopic acceptance rate | Runtime gain | MSE gain | Efficiency gain |
|---|---|---|---|---|---|
| 0.1 | 0.749749 | 0.993677 | 0.218974 | 0.500816 | 0.109666 |
| 1 | 0. 0.74966 | 0.993468 | 0.233651 | 153735 | 35920.4 |
| 10 | 0.7496 | 0.993463 | 0.22258 | 141816 | 31565.4 |
| 100 | 0.749546 | 0.993421 | 0.219963 | 152517 | 33548.1 |
| 1000 | 0.749553 | 0.993469 | 0.218501 | 1303.67 | 284.853 |

Table 5.5: Several statistics of the performance of mM-MCMC with indirect reconstruction over the microscopic MALA method when estimating the variance of $\theta$, for several values of $\lambda$. The time-scale separation parameter is fixed at $\varepsilon = 10^{-6}$. The macroscopic and microscopic acceptance rates remain constant, as well as the execution time gain. Only the MSE gain, and efficiency gain vary in $\lambda$. When $\lambda < \varepsilon^{-1}$, the efficiency gain is low, however, from the point $\lambda > \varepsilon^{-1}$, there is a significant MSE reduction on the estimated variance of $\theta$, and therefore a significant increase in efficiency gain. Finally, when $\lambda$ is too large, relative to the fastest modes in the potential energy of the system, the total efficiency gain decreases again due to an increase of the MSE on the estimated variance of $\theta$ by mM-MCMC.

are then effectively ignoring the macroscopic MCMC step so that there is no variance reduction on moments of $\theta$. This effect is also visible in the first row of Table 5.5, where there is indeed no gain on the variance on the estimated variance of $\theta$.

On the other hand, when $\lambda$ is very large (larger than $10^2/\varepsilon$ in this case), the efficiency also starts to decrease. There are enough biased steps to pull the reaction coordinate value to the sampled value at the macroscopic level. In other words, there is poor mixing. This effect is especially visible on the efficiency gain of the estimate variance of $\theta$ in the bottom row of Table 5.5. More than 5 steps are needed to equilibrate the microscopic samples.

To conclude, there is a large range of $\lambda$ values that result in a big and constant efficiency gain (middle rows of Table 5.5). This range, between $1/\varepsilon$ and $100/\varepsilon$, is the same for every $\varepsilon$. It is a good idea to pick $\lambda$ on par with the fastest mode of the system. These results support our claim from section 4.2.6.

**Visual inspection of mM-MCMC with indirect reconstruction**

Now that we have a proper value for $\lambda$, we visually compare the sampling results of the mM-MCMC algorithm with indirect reconstruction and the microscopic MALA method for a fixed value of $\varepsilon = 10^{-6}$.

**Experimental setup**  We show the histograms of reaction coordinate $\theta$ of both methods in figure 5.7. For the mM-MCMC method, we use macroscopic proposals based on MALA with step size $\Delta t = 0.01$. The macroscopic invariant distribution $\bar{\mu}_0$ is the true free energy. The indirect reconstruction parameters are $K = 5$, $\lambda = \varepsilon^{-1}$ and the biased step size is $\delta t = \varepsilon$. This is also the step size of the microscopic MALA method.

**Numerical results**  The microscopic MALA method remains stuck and over-samples the right potential well of $\theta$. Only a small fraction of samples get to the left potential well due to the small time steps $\delta t = \varepsilon$. The mM-MCMC method with indirect reconstruction is able to sample the distribution of $\theta$ accurately due to the large macroscopic time steps. We can expect a large efficiency gain of mM-MCMC over the microscopic MALA method when estimating quantities of interest of $\theta$.
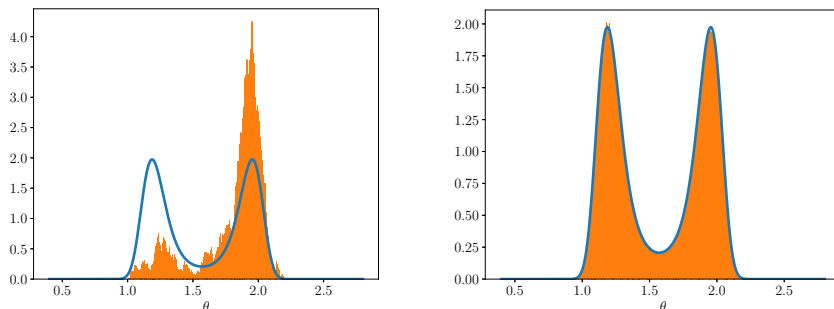
Figure 5.7: $\theta$-histogram of the microscopic MCMC method (left) and mM-MCMC (right) on the three-atom molecule, with reaction coordinate $\theta$. The simulation parameters are $\varepsilon = 10^{-6}$ and the number of samples is $N = 10^6$. The microscopic MCMC method remains stuck in the potential well of $\theta$ around 1.95, while the mM-MCMC method samples the distribution well.

### Efficiency gain as a function of time-scale separation

An import factor in assessing the mM-MCMC method with indirect reconstruction is how the efficiency gain increases with decreasing $\varepsilon$. We study this efficiency gain as a function of the time-scale separation in this section.

**Experimental setup**   We let the time-scale parameter $\varepsilon$ vary between $10^{-6}$ and $10^{-3}$. For every $\varepsilon$, we compute the efficiency gain on the estimated mean and variance of $\theta$, by averaging over 100 independent runs. Parameters are $\Delta t = 0.01$, $K = 5$, $\delta t = \varepsilon$, we use the free energy with MALA to generate reaction coordinate values, and the microscopic MALA time step is also $\varepsilon$.

The efficiency gains on the estimated mean and variance of $\theta$ are show on Figure 5.8. We also display the averaged macroscopic and microscopic acceptance rates, and the runtime and MSE gain on the estimated mean of $\theta$ in Table 5.6.

**Numerical results**   Figure 5.8 shows that the efficiency gain of mM-MCMC increases at least linearly with decreasing $\varepsilon$. For large values of $\varepsilon$, there is almost no efficiency gain at all. The reason is clear from Table 5.6. The runtime of mM-MCMC with indirect reconstruction is larger than that of the microscopic MALA method due to the computational overhead of indirect reconstruction. Hence, for large values of $\varepsilon$, this computational overhead is dominant over the reduction in MSE. However, when $\varepsilon$ is small, the reduction in MSE by mM-MCMC is dominant over the increase in computational time. Currently,

| $\varepsilon$ | Macroscopic acceptance rate | Microscopic acceptance rate | Runtime gain | MSE gain | Efficiency gain |
|---|---|---|---|---|---|
| $10^{-3}$ | 0.749975 | 0.993528 | 0.212558 | 10.3927 | 2.20905 |
| $10^{-4}$ | 0.74936 | 0.993588 | 0.211415 | 68.6401 | 14.5115 |
| $10^{-5}$ | 0.750197 | 0.993299 | 0.212561 | 920.651 | 195.695 |
| $10^{-6}$ | 0.749498 | 0.993405 | 0.237227 | 7041.69 | 1670.48 |

Table 5.6: A summary of different statistics of the mM-MCMC method when applied to estimating the mean of $\theta$, for multiple values of $\varepsilon$. The macroscopic acceptance rate remains constant when $\varepsilon$ decreases. The microscopic acceptance rate is almost 1, but the runtime of mM-MCMC with indirect reconstruction is larger than the runtime of microscopic MALA.. The total efficiency gain increases with $\varepsilon^{-1}$.
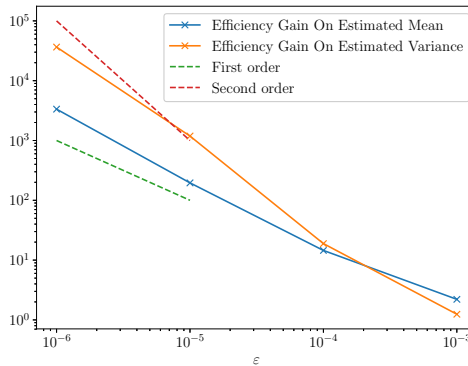


Figure 5.8: Efficiency gain of mM-MCMC over standard MCMC as a function of the time-scale separation. The efficiency gain for the mean of $\theta$ (blue) increases linearly with decreasing $\varepsilon$, while the efficiency gain for the variance of $\theta$ increases faster than linearly, although slower than quadratically.

we do not have analytic expressions for the gain as a function of the time-scale separation, but Figure 5.8 clearly shows the merit of mM-MCMC with indirect reconstruction for medium and large time-scale separations.

## Comparison of mM-MCMC with direct and indirect reconstruction

We have studied the mM-MCMC methods with direct and indirect reconstruction in much detail in previous experiments. In this final experiment on the three-atom molecule, we compare both variants. The direct reconstruction algorithm should be faster for a given number of sampling steps due to the computational

| $\varepsilon$ | Runtime gain | MSE gain | Efficiency gain |
|---|---|---|---|
| $10^{-3}$ | 10.0329 | 0.877206 | 8.80089 |
| $10^{-4}$ | 10.1001 | 0.680705 | 6.87522 |
| $10^{-5}$ | 10.1374 | 1.04204 | 10.5635 |
| $10^{-6}$ | 9.99052 | 1.57429 | 10.4853 |

Table 5.7: A summary of the runtime and MSE gains of mM-MCMC with direct reconstruction over mM-MCMC with indirect reconstruction. The efficiency gain of the direct reconstruction algorithm is almost the same for various values of $\varepsilon$ and this efficiency gain is completely due to the lower runtime of the direct reconstruction.

overhead of indirect reconstruction. The reduction in MSE on estimated quantities is expected be almost identical. That is, only if direct reconstruction can be applied efficiently.

**Experimental setup**    Figure 5.9 shows the efficiency gain of both mM-MCMC methods over the microscopic MALA method on the estimated mean of $\theta$ for various values of $\varepsilon$. In addition, we also show the efficiency gain of mM-MCMC with direct reconstruction algorithm over mM-MCMC with indirect reconstruction. We fix the macroscopic time step at $\Delta t = 0.01$ for both mM-MCMC variants. The numerical parameters for the indirect reconstruction method are $K = 5$, $\lambda = \varepsilon^{-1}$ and $\delta t = \varepsilon$, and the time step of the microscopic MALA method is also $\varepsilon$. The number of sampling steps is $N = 10^6$. We show the efficiency gain of both methods over the microscopic MALA method as a function of $\varepsilon$ in figure 5.9, and we also show the efficiency gain on the estimated mean of $\theta$ in Table 5.7 for both schemes.

**Numerical results**    Firstly, as we can see on figure 5.9, the efficiency gain of both mM-MCMC variants scales linearly with $\varepsilon^{-1}$. Secondly, the efficiency gain of mM-MCMC with indirect reconstruction is a constant factor 10 smaller than the efficiency gain of its direct reconstruction variant, independent of the time-scale separation. If we diagnose this effect more carefully in Table 5.7, the smaller efficiency gain is due to the higher runtime of mM-MCMC with indirect reconstruction. The MSE gain remains mostly constant.
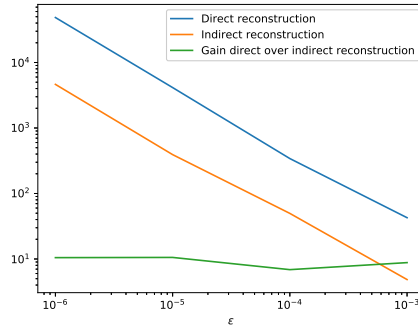
Figure 5.9: Efficiency gain of mM-MCMC with direct reconstruction (blue) and indirect reconstruction (orange) over the microscopic MALA method as a function of $\varepsilon$. The green curve measures the efficiency gain of the direct reconstruction algorithm over the indirect reconstruction method.
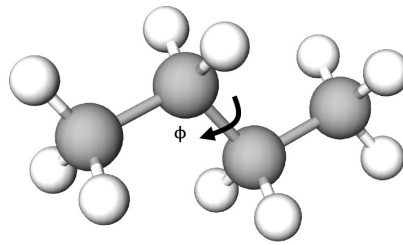


Figure 5.10: The butane molecule. The carbon atoms are gray and hydrogen is white.

## 5.3   Butane

We consider butane as our second test case. A typical conformation of butane is depicted on Figure 5.10 [88, 102]. For simplicity, we only simulate the carbon backbone of the molecule (the grey atoms) and remove the hydrogen atoms (white). Such lower-dimensional model is also called a 'united atom' description [88], and this model keeps the main features and difficulties of the molecule.

The potential energy of butane is

$$V(x) = \sum_{i=1}^{3} \frac{k_b}{2}(r_i(x) - r_0)^2 + \sum_{i=1}^{2} \frac{k_a}{2}(\theta(x) - \theta_0)^2 + V_\phi(\phi(x)), \ x \in \mathbb{R}^{12}. \ (5.13)$$

The first term is sum over the three carbon-carbon bond with length $r_i(x)$, and measure a locally quadratic potential around equilibrium bond length $r_0 = 1.53$Å, with force constant $k_b = 1.17 \cdot 10^6$. The second term likewise forms a quadratic potential of each carbon-carbon-carbon angle with equilibrium state $\theta_0 = 112 \deg$. The constant $k_a = 62500$ is more than an order of magnitude smaller than $k_b$. The final term is the energy of the torsion angle $\phi$, the torsion between the backbone carbon atoms. This function is

$$V_\phi(\phi) = c_1 \cos(\phi) + c_2 \cos(\phi)^2 + c_3 \cos(\phi)^3, \quad (5.14)$$

with constants $c_1 = 2037.82$, $c_2 = 158.52$ and $c_3 = -3227.7$. This energy function has three local minima, each of these represents a stable conformation of the molecule. Note that we do not include volume-exclusion forces into the molecule for simplicity and for a more thorough understanding of the numerical results of mM-MCMC.

The torsion angle determines how close the two outer $CH_3-$groups are to each other, and hence it determines the different global conformations the molecule can have. The most stable conformation is obtained when the two outer groups are as far away from each other as possible, as depicted on Figure 5.10. This situation coincides with $\phi = 0$. We therefore define our reaction coordinate to be the torsion angle

$$\xi(x) = \phi(x). \quad (5.15)$$

and the free energy then becomes

$$A(z) = V_\phi(z). \quad (5.16)$$

We prove that the free energy of $\phi(x)$ indeed has the above expression in Appendix B.

**Outline of this Section**  We carry out numerical experiments on butane with both mM-MCMC with direct reconstruction (section 5.3.1) and mM-MCMC with indirect reconstruction (section 5.3.2). The experiments and results are similar to those of the three-atom molecule.

For mM-MCMC with direct reconstruction, we start in section 5.3.1 with a visualization of the mM-MCMC sampling results with respect to the microscopic MALA method. Afterwards, compare the sampling results on

the reaction coordinate with a simple macroscopic sampler, in section. Finally, similarly to the three-atom molecule, we investigate the effect on efficiency of the approximate macroscopic distribution and the macroscopic proposal distribution, as well as, the effect of the approximate macroscopic distribution and reconstruction distribution in.

On the side of mM-MCMC with indirect reconstruction, we start by computing the optimal value of $\lambda$, followed by a visualization of mM-MCMC with indirect reconstruction with that $\lambda$.

## 5.3.1 Experiments with Direct Reconstruction

### Visual Inspection of mM-MCMC with Direct Reconstruction

In the experiment below, we plot the histogram approximation of the invariant distribution of the reaction coordinate $\phi$ of mM-MCMC, and compare it to the microscopic MALA approximation.

**Experimental setup** For mM-MCMC, we generate macroscopic reaction coordinate values using MALA proposals with time step $\Delta t = 5 \cdot 10^{-4}$, and with the exact free energy as approximate macroscopic distribution. Reconstruction is achieved by sampling the exact reconstruction distribution so that the microscopic acceptance rate is 1. For the microscopic MALA method, we use a time step of $\delta t = 10^{-6} \sim k_b^{-1}$. We start both Markov chains in the most stable conformation, i.e., $\phi = 0$. Both Markov chains run for $N = 10^6$ steps. The inverse temperature is $\beta = 10^{-2}$. The histogram of $\phi$ of both methods is represented in figure 5.11.

We measure the efficiency of mM-MCMC over the microscopic MALA by estimating the mean of $\phi$ with both methods. For a good comparison, we average the estimated means of $\phi$ over 100 independent runs. We also keep track of the macroscopic and microscopic acceptance rates of the mM-MCMC method, the runtime of both methods for a fixed number of samples, the MSE of the estimator of the mean of $\phi$ for both methods. These numerical quantities, as well as the total efficiency gain, are displayed in Table 5.8.

**Numerical results** The microscopic MALA algorithm is not able to sample the two outer lobes of the probability distribution of $\phi$ accurately due to the time-scale separation. As expected, the mM-MCMC scheme ìs able to precisely sample each of the three lobes in the distribution of $\phi$. Large (macroscopic) proposal steps implies a higher crossing frequency between lobes.
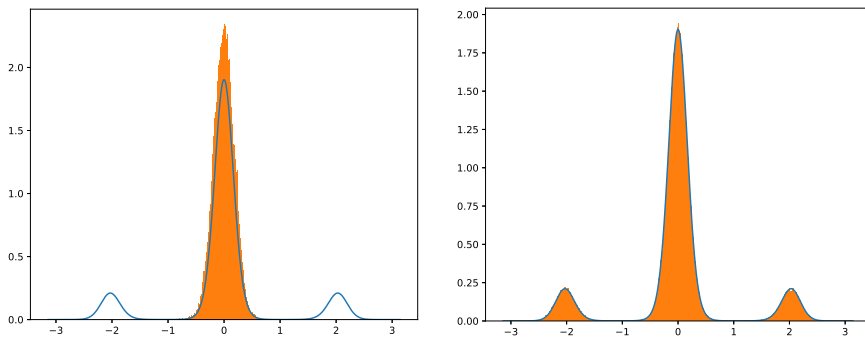
Figure 5.11: Histogram approximations to the invariant distribution of the torsion angle $\phi$ of the microscopic MALA method(left) and mM-MCMC with direct reconstruction (right).

| Method | Macroscopic acceptance rate | Microscopic acceptance rate | Runtime | MSE | Efficiency gain |
|---|---|---|---|---|---|
| MALA | / | 0.7365 | 5503 s | 0.0817 | 303.95 |
| mM-MCMC | 0.28919 | 1 | 279 s | 0.00564 | |

Table 5.8: Summary of all factors that influence the total efficiency of mM-MCMC with direct reconstruction over the microscopic MALA method on butane.

First of all, the microscopic acceptance rate of mM-MCMC is indeed 1, since the exact reconstruction property (4.5) is satisfied. Additionally, the averaged runtime of mM-MCMC with direct reconstruction is lower than for the MALA method method. We discussed this phenomenon in section 5.1. Furthermore, the MSE on the estimated mean of $\phi$ is also lower for the mM-MCMC, than the microscopic MALA scheme because of improved mixing.

Summing up, the efficiency gain of mM-MCMC over the microscopic MALA method is approximately a factor 300, bridging a large part of the time-scale separation. This efficiency gain means that mM-MCMC needs 300 times fewer samples than the microscopic MALA method, to obtain the same MSE on the estimated mean of $\phi$. That is, for the same runtime. Said differently, we obtain an MSE that is two or more orders of magnitude lower for the same computational cost.

## Comparison with Reaction Coordinate Sampler

Let us compare the mM-MCMC method with the macroscopic reaction coordinate sampler. The true invariant macroscopic distribution has three local maxima with different amplitudes, and far apart. We have already discovered that the mM-MCMC method can correct for large biases in the approximate macroscopic distribution in case of the three-atom molecule (section 5.2.1). The butane case is, however, more challenging. The local minima of $A(\phi)$ are further apart and two outer local minima have a smaller amplitude.

**Experimental setup**   We define two approximate free energy functions,

$$
\begin{aligned}
\bar{A}^1(\phi) &= c_0 + c_1 \cos(1.2\phi) + c_2 \cos(1.2\phi)^2 + c_3 \cos(1.2\phi)^3, \\
\bar{A}^2(\phi) &= c_0 + c_1 \cos(\phi) + c_2 \cos(\phi)^2 + c_3 \cos(\phi)^3 + 500 \cos(\phi - 1),
\end{aligned}
\tag{5.17}
$$

so that the corresponding approximate macroscopic distribution is $\bar{\mu}_0^i(dz) = \exp\left(-\beta \bar{A}^i(z)\right) dz$. The first approximate distribution contracts outer local maxima by a factor $1.2^{-1}$, while the second approximate distribution assigns a larger amplitude to left-most local maximum, while decreasing the amplitude of the other local maxima.

For each approximate macroscopic distribution, we run the mM-MCMC method with direct reconstruction, and plot the histogram approximation on the reaction coordinate. On the macroscopic level, we use a MALA method with time step $\Delta t = 5 \cdot 10^{-4}$ to generate proposals for both the macroscopic and mM-MCMC samplers. We also use the exact reconstruction distribution $\nu$ to generate microscopic samples for mM-MCMC. The number of sampling steps is $N = 10^6$. The numerical results are shown in Figure 5.12.

**Numerical results**   One can see that the mM-MCMC is able to overcome a bad approximate macroscopic distribution and samples the marginal distribution correctly. This conclusion is the same for the three-atom molecule. This correction effect is especially visible on the right figure. Indeed, the second free energy $\bar{A}^2$ gives completely different weights to each local minimum. The micro-macro MCMC corrects for this misrepresentation in the microscopic accept/reject step.

## Effect of $\bar{A}$ and $q_0$ on the efficiency gain

From the previous experiment, we know that the mM-MCMC can correct for bad approximate macroscopic distributions, and still sample the true free energy
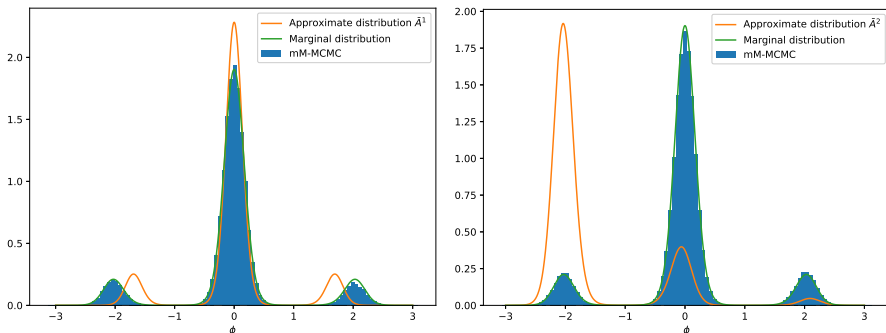
Figure 5.12: Histogram approximation to the invariant distribution of $\phi$ by the mM-MCMC method (blue) for two different approximate macroscopic distributions: $\bar{A}^1$ (left) and $\bar{A}^2$ (right). The orange line is the density each approximate macroscopic distribution used, while the green line is the true marginal distribution of the reaction coordinate values.

of $\phi$ accurately. In this section, we take this experiment one step further by comparing the actual efficiency gains for six combinations of the approximate macroscopic distribution $\bar{\mu}_0$ and the type proposals moves $q_0$. We measure the efficiency gain of mM-MCMC over the microscopic MALA method, where our quantity of interest is the variance of $\phi$.

**Experimental setup**   We fix the reconstruction distribution to exact reconstruction $\nu$ in this experiment. We consider three approximate macroscopic distributions in this experiment: the exact free energy (6.25) and the two approximate free energies defined above. The two macroscopic proposal distributions are defined by a discretization of the dynamics

$$q_0^1 : d\phi = -\nabla \bar{A}(\phi)dt + \sqrt{2\beta^{-1}}dW$$

$$q_0^2 : d\phi = \sqrt{2\beta^{-1}}dW,$$

where $\bar{A}$ is a shorthand for each of the three (approximate) macroscopic free energies used in this experiment. The first proposal distribution is a macroscopic MALA method (3.11), while the second proposal move is a scaled Brownian motion.

The macroscopic time step of the mM-MCMC method is $\Delta t = 5 \ 10^{-4}$, and that of the microscopic MALA method $\delta t = 10^{-6}$. Both methods use $N = 10^6$ sampling steps. For a good comparison, we average the estimated variances

| Parameters | Macroscopic acceptance rate | Microscopic acceptance rate | Runtime gain | MSE gain | Efficiency gain |
|---|---|---|---|---|---|
| $q_0^1$, $A$ | 0.289293 | 1 | 23.8795 | 1050.6 | 25087.9 |
| $q_0^2$, $A$ | 0.423083 | 1 | 19.4699 | 5156.4 | 100396 |
| $q_0^1$, $\bar{A}^1$ | 0.313243 | 0.297524 | 22.8156 | 0.085769 | 1.95688 |
| $q_0^2$, $\bar{A}^1$ | 0.407327 | 0.754024 | 19.6211 | 0.364752 | 7.15682 |
| $q_0^1$ $\bar{A}^2$ | 0.30793 | 0.762832 | 21.7298 | 1341.02 | 29140 |
| $q_0^2$ $\bar{A}^2$ | 0.432885 | 0.822621 | 18.3769 | 3279.37 | 60264.5 |

Table 5.9: The macroscopic and microscopic acceptance rates, the runtime and MSE gain and the total efficiency gain of mM-MCMC over the microscopic MALA method for six combinations of the approximate macroscopic distribution and type of macroscopic proposals.

of $\phi$ over 100 independent runs. We also keep track of the macroscopic and microscopic acceptance rates of the mM-MCMC method, the runtime of both methods for a given number of sampling steps, the MSE's on the estimated variances of $\phi$. These quantities are displayed in Table 5.9.

**Numerical results** Firstly, it is clear from the numerical results that approximate macroscopic distribution with $\bar{A}^1$ results in a poorer efficiency gain than $\bar{A}^2$. The reason is that the reaction coordinate samples spend most of their time in the left-most lobe with largest amplitude, while only few macroscopic samples should reside there according to $\mu_0$. Approximate macroscopic distribution with free energy $\bar{A}^2$ results in a larger efficiency gain because the outer lobes have the correct invariant probabilities.

Secondly, macroscopic proposal moves based on Brownian increments leads to a larger efficiency gain than those based on the macroscopic MALA method. Indeed, the MSE is factor 3 to 5 lower for $q_0^2$ compared to $q_0^1$, and the total efficiency gain is also higher for $q_0^2$. One reason is that gradient based methods take the potential energy into account, and typically remain in one lobe. Brownian motion is not aware of the potential, and can explore the state space without soft constraints.

To conclude, the closer the approximate macroscopic distribution is to the true marginal of the reaction coordinate, the higher the efficiency gain. Macroscopic proposal moves based on Brownian increments also result in a larger efficiency gain than proposal moves based on the MALA method.

| Parameters | Macroscopic acceptance rate | Microscopic acceptance rate | Runtime gain | MSE gain | Efficiency gain |
|---|---|---|---|---|---|
| $A$, $\bar{\nu}^1$ | 0.423083 | 1 | 19.4699 | 5156.4 | 100396 |
| $A$, $\bar{\nu}^2$ | 0.422962 | 0.154385 | 9.88807 | 0.90294 | 8.92833 |
| $\bar{A}^1$, $\bar{\nu}^1$ | 0.407327 | 0.754024 | 19.6211 | 0.364752 | 7.15682 |
| $\bar{A}^1$, $\bar{\nu}^2$ | 0.408502 | 0.130944 | 10.1383 | 1.09534 | 11.1049 |
| $\bar{A}^2$, $\bar{\nu}^1$ | 0.432885 | 0.822621 | 18.3769 | 3279.37 | 60264.5 |
| $\bar{A}^2$, $\bar{\nu}^2$ | 0.432576 | 0.14719 | 9.74243 | 2.5751 | 25.0878 |

Table 5.10: The macroscopic and microscopic acceptance rates, the runtime and variance gain and the total efficiency gain of the mM-MCMC method for six combinations of the approximate macroscopic distribution and the reconstruction distribution.

**Effect of $\bar{A}$, and $\bar{\nu}$ on the efficiency gain**

Finally, we investigate the role of the reconstruction distribution $\bar{\nu}$ on the efficiency gain of the mM-MCMC method. To this end, we choose two approximate reconstruction distributions and investigate the overall efficiency of mM-MCMC for six combinations of $(\bar{A}^i, \bar{\nu})$. The approximate free energies are the same as above, while our two choices for the reconstruction distribution read

$$\bar{\nu}^1(x|z) = \nu(x|z) = \mu(x)\mu_0(z)^{-1}$$

$$\bar{\nu}^2(x|z) = \mu(x)^{\frac{1}{2}} Z(z)^{-1}.$$

The first is the exact reconstruction distribution, and $\bar{\nu}^2$ is obtained by doubling the variances of each of the microscopic degrees of freedom of butane. Here, $Z(z)$ is the normalization constant of $\bar{\nu}^2(x|z)$.

**Experimental setup** For each of the above combinations, we compute the efficiency gain on the variance of $\phi$. The macroscopic time step for mM-MCMC is $\Delta t = 5 \cdot 10^{-4}$ and the microscopic time step for the MALA method is $\delta t = 10^{-6}$. We use $N = 10^6$ sampling steps for both method and we average each combination over 100 independent runs. The macroscopic and microscopic acceptance rates, the runtime and variance gains and the total efficiency gain of mM-MCMC over the microscopic MALA method are shown in Table 5.10.

**Numerical results** The choice of reconstruction distribution has a significant effect on the total efficiency gain of the mM-MCMC method. Indeed, for

approximate free energy functions $A$ and $\bar{A}^2$, the efficiency gain is much larger with $\bar{\nu}^1$.

## 5.3.2 Experiments with Indirect Reconstruction

### Optimal value of $\lambda$

Before we can run the mM-MCMC scheme with indirect reconstruction, we need to establish a good value of $\lambda$ first. From (5.13), $k_b$ determines the fastest modes. We therefore expect $\lambda$ to lie in a log-symmetric range around $k_b$.

**Experimental setup**  We define a range of values for $\lambda$ between $0.1kb$ and $100k_b$. For every value of $\lambda$, we compute the efficiency gain of the mM-MCMC method with indirect reconstruction over the microscopic MALA method. The quantities of interest are both the mean and variance of $\phi$. We average the efficiency gain over 100 independent runs to reduce the variance.

The mM-MCMC method uses Brownian increments with variance $\Delta t = 0.4$ on the macroscopic level, and the reconstruction parameters are $\delta t = 0.4/\lambda$, and $K = 10$. The microscopic sampler uses a microscopic time step of size $1.0k_b$ for stability. We run both methods for $N = 10^5$ sampling steps. The efficiency gains as a function of $\lambda$ are shown in Figure 5.13

**Numerical results**  For every value of $\lambda$, the mM-MCMC method with indirect reconstruction makes an efficiency gain. This gain is not necessarily maximal because we did not tune the reconstruction time step $\delta t$ at all. Firstly, the efficiency gain on the mean of $\phi$ remains constant for all $\lambda$. Secondly, the mM-MCMC method makes an efficiency gain on the variance of $\phi$ for all values of $\lambda$ smaller than $10k_b$. Any of these $\lambda$-values can be used. For $\lambda$ greater than $10k_b$, however, the efficiency gain drops significantly. This result is similar to that in section 5.2.2. Indeed, the best value for $\lambda$ is typically on the order of the fastest mode of the system.

### Visual Examination of mM-MCMC over Microscopic MALA

Finally, we compare mM-MCMC with indirect reconstruction to the microscopic MALA method. We specifically compare the histogram fits on the marginal distribution of $\phi$ by both methods. Afterwards, we also show the histograms
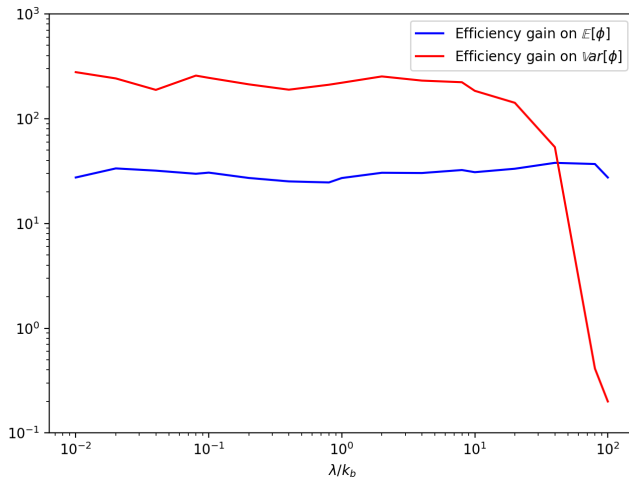
Figure 5.13: Efficiency gain of mM-MCMC over the microscopic MALA method as a function of $\lambda$. Blue: efficiency gain on the mean torsion angle. Red: efficiency gain on the variance of the torsion angle.

fits by mM-MCMC on the invariant distributions on both regular angles, $\theta_1$ and $\theta_2$.

**Experimental Setup**   For this experiments, we run both methods once and record the reaction coordinate values of all microscopic samples. For the mM-MCMC method, we additionally record both regular angles for each sample. The mM-MCMC parameters are $\Delta t = 0.001$, $\lambda = 4k_b$, $\delta = 0.8/\lambda$ and $K = 10$. The macroscopic proposal moves are based on Brownian motion, without gradient information. The time step of the microscopic MALA method is $0.8/k_b$. The initial condition lies in the central lobe of (6.25). Figure 5.14 shows the histogram approximations by both microscopic MALA (left) and mM-MCMC (rights) methods on the marginal distribution of $\phi$, while figure 5.15 contains the histogram approximations by the mM-MCMC method on both regular angles.

**Numerical Results**   The microscopic MALA method remains stuck in the central lobe of the marginal distribution of $\phi$, but the mM-MCMC method samples the full distribution correctly. This is of course due to the large macroscopic proposal moves. Additionally, the histograms on both angles, $\theta_1$ and $\theta_2$, by the mM-MCMC align well with their invariant distributions.
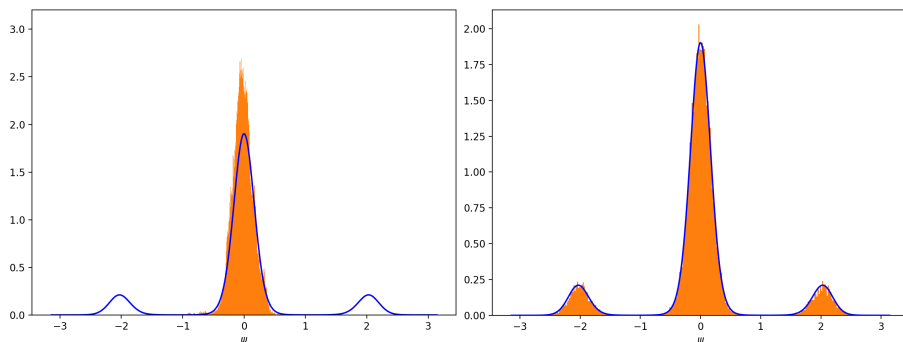
Figure 5.14: Histogram approximation (orange) to the invariant distribution (blue) of the torsion angle. Left is the result of the microscopic MALA method, right is that of mM-MCMC.
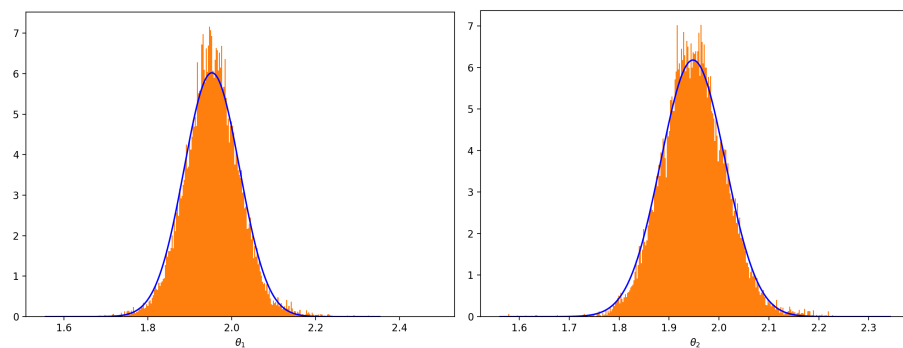


Figure 5.15: Histogram approximation (orange) to the invariant distribution (blue) of the two angles $\theta_1$ (left) and $\theta_2$ (right) of butane by the mM-MCMC method.

## 5.4 Alanine-Dipeptide

As final molecular example, we consider alanine-dipeptide, a much-used test case for numerical methods in molecular dynamics [60]. Alanine-dipeptide consists of one alanine amino acid, and two end groups. The organic R-group, or the side chain, is an $CH_3$ group, positioned on the central carbon atom of the backbone. A typical conformation is plotted on figure 5.16.

Alanine-dipeptide has two backbone torsion angles, $\phi$ and $\psi$. Here, $\psi$ is the slowest degree of freedom, and $\phi$ is an order of magnitude faster. The first torsion angle $\phi$ determines the position of the right subgroup of the molecule,
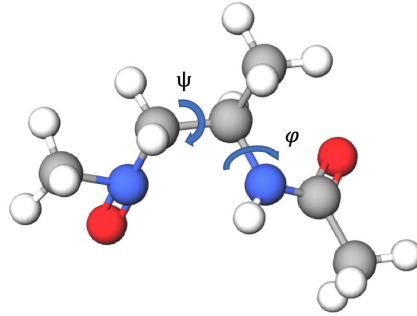
Figure 5.16: The alanine-dipeptide molecule. The carbon atoms are gray, hydrogen white, nitrogen blue and oxygen is red.

| Term | form | parameters |
|---|---|---|
| C-C Bond | $0.5\ k_{CC}\ (r - r_{CC})^2$ | $k_{CC} = 1.17 \cdot 10^6,\ r_{CC} = 1.515$ |
| C-N Bond | $0.5\ k_{CN}\ (r - r_{CN})^2$ | $k_{CN} = 1.147 \cdot 10^6,\ r_{CN} = 1.335$ |
| C-C-N Angle | $0.5\ k_{CCN}\ (\theta - \theta_{CCN})^2$ | $k_{CCN} = 2.68 \cdot 10^5,\ \theta_{CCN} = 113.9\,\mathrm{deg}$ |
| C-N-C Angle | $0.5\ k_{CNC}\ (\theta - \theta_{CNC})^2$ | $k_{CNC} = 1.84 \cdot 10^5,\ \theta_{CNC} = 117.6\,\mathrm{deg}$ |
| Torsion Angle $\phi$ | $k_\phi(1 + \cos(\phi + \pi))$ | $k_\phi = 3.98 \cdot 10^4$ |
| Torsion Angle $\psi$ | $k_\psi(1 + \cos(\psi + \pi))$ | $k_\psi = 2.93 \cdot 10^3$ |

Table 5.11: Terms with parameters in the potential energy of alanine-dipeptide.

starting from the carbon-nitrogen bond with the central carbon molecule. The other torsion angle $\psi$ then determines the position of the left subgroup, starting from the carbon-carbon bond with the central carbon atom.

The potential energy function of alanine-dipeptide contains of a term for each covalent bond, a term for each angle between two bonds that are connected by a common atom and a term for each of the two torsion angles. We do not consider non-bonded interactions. The terms and parameters of the potential energy are summarized in Table 5.11. Note that each potential energy term corresponding to a torsion angle has only one stable conformation. The parameters in the potential energy are taken from [37].

The reaction coordinate for the experiments below will be the torsion angle

$$\xi(x) = \psi(x),$$

because $\psi$ is the slowest degree of freedom. The free energy of $\psi$ can be derived identically to that of butane and is equal to

$$A(\psi) = k_\psi(1 + \cos(\psi + \pi)). \tag{5.18}$$

Note that this expression would be more complex if there were volume exclusion forces at play. We refer to appendix A for more details.

**Outline of this Section**   Alanine-dipeptide is more complex than the previous two molecules. Direct reconstruction is not feasible anymore because the reaction coordinate is a complex function of the atomic positions. We only experiment with mM-MCMC with indirect reconstruction. We first establish the optimal value for $\lambda$ in section 5.4.1. Afterwards, we compare the histogram approximations to the invariant distribution of the torsion angles by mM-MCMC to their true marginal distribution in 5.4.2. Finally, we also examine the efficiency gain on both dihedral angles in the same section.

## 5.4.1   Optimal value of $\lambda$

The fastest modes of the system are the carbon-carbon and carbon-nitrogen bonds, see table 5.11. We expect $\lambda$ to be on the order of $k_{CC} \approx k_{CN}$.

**Experimental setup**   In this experiment, we define 18 equidistant values of $\lambda$ between $10^6$ and $10^7$, and for each value, we compute the efficiency gain of mM-MCMC over the microscopic MALA method. For the mM-MCMC method, the macroscopic proposals are based on the MALA sampler in $\psi$, with a time step of size $\Delta t = 0.001$. For indirect reconstruction and we take $K = 8$ biased steps with time step $0.5\lambda^{-1}$. The time step for the microscopic MALA method is fixed at $10^{-7}$, while the temperature parameter is $\beta = 1/100$. We run both methods for $N = 10^6$ sampling steps. Figure 5.17 shows the efficiency gain on the estimated mean and variance of reaction coordinate $\psi$ as a function of $\lambda$.

**Numerical results**   The efficiency gain on the estimated mean of $\psi$ reaches a constant level when $\lambda \geq 2\ 10^6$. When $\lambda$ is smaller than $2\ 10^6$, the efficiency gain decreases fast to a level under 1. Indeed, we see that $\lambda = 2.5\ 10^6$ results in the largest efficiency gain on the estimated variance of $\psi$, consistent with the efficiency gain on the mean on $\psi$. This result is in agreement with section 4.2.6, because $\lambda$ is on the order of $k_{CC}$,. Therefore, we will use this optimal value $\lambda = 2.5\ 10^6$ in the subsequent numerical experiments on the alanine-dipeptide molecule.
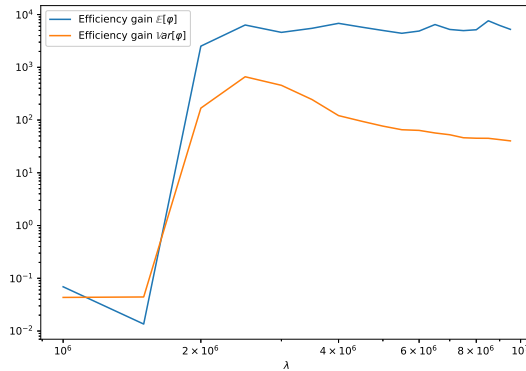
Figure 5.17: The efficiency gain of mM-MCMC with indirect reconstruction over the microscopic MALA method on the estimated mean (blue) and the estimated variance (orange) of $\psi$.

## 5.4.2 Visual inspection of mM-MCMC

Finally, we also show the efficiency gains of mM-MCMC with indirect reconstruction over the microscopic MALA method on alanine-dipeptide. Particularly, we compare the histogram fit of the microscopic MALA and the mM-MCMC methods to the true marginal distribution of both torsion angles.

**Experimental setup**   We run both methods once and store the microscopic samples. Afterwards, we compute the two torsion angles from each sample, and create histograms. The numerical parameters of the mM-MCMC method are $K = 8$, $\beta = 1/100$, $\lambda = 2.5 \cdot 10^6$, $\Delta t = 0.001$ and for biased simulation time step we take $0.5/\lambda$. The time step of the microscopic MALA method is $10^{-7}$. The histograms for $\phi$ and $\psi$ are shown on Figure 5.18.

**Numerical results**   Firstly, the microscopic MALA method (left) leads to a poor fit on the marginal distributions of torsion angles. Indeed, this is especially visible on $\psi$ (top left). However, the fit on the distribution of $\phi$ is more accurate, because $\phi$ has a faster dynamics, see table 5.11. On the other hand, the mM-MCMC method with indirect reconstruction samples both distributions more accurately due to the large time steps at the macroscopic level. The mM-MCMC makes large gains on $\psi$, especially. Additionally, the histogram fit on $\phi$ by the
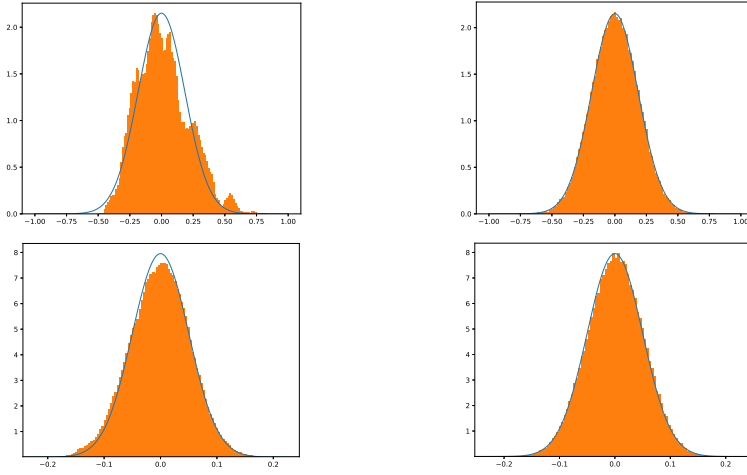
Figure 5.18: Histograms of $\psi$ (top) and $\phi$ (bottom), computed using the microscopic MALA method (left) and mM-MCMC (right).

mM-MCMC method is also a bit better than that of the microscopic MALA method.

# Chapter 6

# mM-MCMC with Pseudo Marginal Approximation to the Free Energy

In this chapter, we deal with the main computational bottleneck in the mM-MCMC method with indirect reconstruction. If we recall the microscopic acceptance probability (4.14),

$$\alpha_F(x', z'|x_n, z_n) = \left\{ 1, \frac{\mu_\lambda(z') \, \bar{\mu}_0(z_n)}{\mu_\lambda(z_n) \, \bar{\mu}_0(z')} \right\}, \qquad (6.1)$$

we need to compute the marginal distribution $\mu_\lambda(z)$ of $\mu_{\text{ext}}(x, z)$ to even evaluate (6.1). Recalling the definition (4.11), $\mu_\lambda$ is defined by an integral over the marginal invariant distribution of the reaction coordinate. We can never evaluate $\mu_\lambda$ exactly, nor can we pre-compute it because of the bias it would induce on the whole mM-MCMC method. The only way to compute the microscopic acceptance probability is to rely on an on-the-fly estimate $\tilde{\mu}_\lambda$ of $\mu_\lambda$ that maintains exactness of the mM-MCMC algorithm.

At this time, it is opportune to introduce the partition function $Q_\lambda(z)$ of $\mu_\lambda$

$$Q_\lambda(z) = -\frac{1}{\beta} \ln \mu_\lambda(z), \qquad (6.2)$$

so that the microscopic acceptance probability can be written as

$$\alpha_F(x', z'|x_n, z_n) = \min \left\{ 1, e^{-\beta \left( Q_\lambda(z') - Q_\lambda(z_n) \right)} \frac{\bar{\mu}_0(z_n)}{\bar{\mu}_0(z')} \right\}. \qquad (6.3)$$

From the definition (6.2), $Q_\lambda$ acts as the 'free energy' of the marginal distribution $\mu_\lambda$ of a new reaction coordinate on the extended state space $\mathbb{R}^{3D} \times \mathbb{R}^m$ given by

$$\xi_{\text{ext}} : \mathbb{R}^{3D} \times \mathbb{R}^m \to \mathbb{R}^m \tag{6.4}$$

$$(x, z) \mapsto z = \xi_{\text{ext}}(x, z). \tag{6.5}$$

It is this free energy we need to compute to evaluate the microscopic acceptance probability. In the remainder of this thesis, whenever we mention the 'free energy' or the 'marginal distribution', we refer to $Q_\lambda$ and $\mu_\lambda$ respectively.

In section 6.1, we discuss important free energy computation methods, and analyze them in the context of the mM-MCMC method. At the end of that section, we give general requirements for any FEC method for it can be used to estimate $\tilde{Q}\lambda$ in the mM-MCMC algorithm. Then, in section 6.2 we explain the pseudo-marginal method in general terms and discuss its applicability for computing an estimate $\tilde{Q}_\lambda$ in the microscopic acceptance probability. Afterwards, we propose and implement the necessary changes to the vanilla pseudo-marginal method to make it computationally efficient in section 6.3, before numerically analysing and showing the performance of the pseudo-marginal method in the mM-MCMC method with indirect reconstruction on two molecular examples: butane and alkane. The contents and results of this chapter have been published in a preprint [93], and the text below has been largely taken from said preprint.

## 6.1   Existing Free Energy Approximation Methods

Free energy computations (FEC) are central in the multiscale molecular dynamics literature. We give a couple of key references [90, 40, 19]. We do not need to approximate the free energy up to double precision in most applications. Indeed, only a rough free energy difference estimate

$$\tilde{Q}_\lambda(z') - \tilde{Q}_\lambda(z_n)$$

is enough to evaluate (6.3), as long the approximation is exact in expectation. In what follows we give a summary of the most important free energy computation schemes from the literature. We present each scheme in some detail, and discuss its applicability for evaluating $Q_\lambda$ in the context of the microscopic acceptance probability in the mM-MCMC algorithm. We end this section by presenting a list of properties that any FEC scheme should satisfy to estimate $Q_\lambda$ in the context of (6.3).

## 6.1.1 Three Free Energy Computation Schemes from the Literature

The three free energy computation schemes that we will present in this section are the biased potential / force (ABP/ ABF) methods, Thermodynamic Integration (TI), and a non-equilibrium stochastic scheme based on the famous Jarzynski relation [43]. We discuss where each of these methods work, and where they do not, and bring them in relation with (6.3).

### Adaptive Biasing Potential Force

The idea behind the Adaptive Biasing Potential (ABP) method is to build up knowledge of the free energy $Q_\lambda(z)$ over time, and then to subtract it from the microscopic process to 'flatten' the microscopic dynamics in the direction of the reaction coordinate. Suppose we are interested in the value of $\mu_\lambda$ or $Q_\lambda$ at the value $z^*$. The microscopic process in question is typically the overdamped Langevin dynamics that keeps $\mu_{\text{ext}}$ invariant

$$dq_t = \nabla \ln \left( \mu_{\text{ext}}(q_t) \right) dt + \sqrt{2\beta^{-1}} dW(t), \qquad (6.6)$$

with $q_t = (x_t, z_t)$.

The Adaptive Biased Potential approximation to the free energy $Q_\lambda(z')$ at $z^*$ is given by a mollification $Q_\lambda^\varepsilon(z^*)$

$$\exp \left( -\beta Q_\lambda^\varepsilon(z^*, t) \right) = Z_t^{-1} \left( 1 + \int_0^t \delta_\varepsilon(z_s - z^*) ds \right), \qquad (6.7)$$

where $\delta_\varepsilon$ is a Gaussian approximation to the delta function with variance $\varepsilon^2$, and $Z_t$ is the normalization constant so that $Z_t^{-1} \int_{\mathbb{R}^n} \exp \left( -\beta Q_\lambda^\varepsilon(z, t) \right) dz = 1$. Equation (6.7) aggregates the macroscopic samples $z_s$ and weighs them according to their distance with respect to $z^*$. Close sample have a higher weighing than farther samples.

The Adaptive Biased Force (ABF) method takes a different approach by computing the gradient $\nabla_{z^*} Q_\lambda^\varepsilon(z^*)$ instead,

$$\nabla_{z^*} Q_\lambda^\varepsilon(z^*, t) = -\beta^{-1} \frac{\int_0^t \nabla_{z^*} \delta_\varepsilon(z_s - z^*) ds}{1 + \int_0^t \delta_\varepsilon(z_s - z^*) ds}. \qquad (6.8)$$

We refer for more details to [65, 21, 25, 24] about the theory and implementation.

We have two remarks concerning the ABP/ABF framework. First, these methods work best when one wants to approximate the free energy at many

points at the same time. The Langevin dynamics (6.6) is independent of $z^*$, and the update procedure (6.7) can be implemented in parallel.

Second, both methods rely on the overdamped Langevin dynamics to generate microscopic samples that are generated according to (4.16). If we were to use the ABP method to calculate the free energy difference in the mM-MCMC algorithm, the indirect reconstruction and overdamped Langevin simulations would need to run simultaneously, adding overhead. We do not currently see use for the ABP/ABF methods for this reason.

**Thermodynamic Integration**

The method of thermodynamic integration is based on two observations [19, 41, 90]. The first observation is that the free energy difference between two states $z$ and $z'$ can be written in terms of the free energy derivative $dQ_\lambda/d\xi_{\text{ext}}$

$$Q_\lambda(z') - Q_\lambda(z) = \int_z^{z'} \frac{dQ_\lambda}{d\xi_{\text{ext}}} d\xi_{\text{ext}}, \qquad (6.9)$$

and that the free energy derivative, in turn, can be computed using a constrained microscopic statistical average

$$\frac{dQ_\lambda}{d\xi_{\text{ext}}}(z) = \frac{\int_{\mathbb{R}^{3D}} \frac{\partial V_{\text{ext}}(x,z)}{\partial z} \mu_{\text{ext}}(x,z) \ dx}{\int_{\mathbb{R}^{3D}} \mu_{\text{ext}}(x,z) \ dx} = \int_{\mathbb{R}^{3D}} \frac{\partial V_{\text{ext}}(x,z)}{\partial z} \mu_{\text{ext}}(x,z) \ dx. \qquad (6.10)$$

This is the second observation. In notation above, $V_{\text{ext}}$ is the potential energy of $\mu_{\text{ext}}$. In a practical implementation, one would choose $M$ collocation points $z_m$ between $z$ and $z'$, approximate the free energy derivative $\frac{dQ_\lambda}{d\xi_{\text{ext}}}(z_m)$ at each such point, and then integrate (6.9) using a quadrature rule.

We see two reasons why thermodynamic integration cannot be used efficiently to compute the free energy difference in the mM-MCMC algorithm. The first reason is about computation cost. For an accurate estimate of the free energy derivative, we need many particles over a large part of the microscopic state space $\mathbb{R}^{3D}$. This is too high a cost for TI to be used after every reconstruction step. The second reason is about accuracy. Even if one can compute the free energy derivative to any specified precision, there will always be a deterministic error on (6.9). Error control typically depends on a higher order quadrature rule, making TI too expensive

**Non-Equilibrium Integration**

The final FEC method we present is a non-equilibrium constraint dynamics together with the renowned Jarzynski formula [43]. The Jarzynski formula relates the free energy difference between two states $z$ and $z'$ to the irreversible work done by an ensemble of constraint paths joining $z$ with $z'$.

For a practical implementation, we first need a monotone differentiable schedule of reaction coordinate values $z(t)$ so that $z(0) = z$ and $z(T) = z'$ for some end time $T \geq 0$. The constrained microscopic Langevin dynamics then reads

$$\begin{cases} q_0 & \sim \nu_\lambda(\cdot|z) \\ dq_t & = -\nabla V_{\text{ext}}(q_t)dt + \sqrt{2\beta^{-1}}dW_t + \sum_{\alpha=1}^m \nabla\xi_{\text{ext},\alpha}(q_t)d\lambda_{\alpha,t} \\ z(t) & = \xi_{\text{ext}}(q_t). \end{cases} \quad (6.11)$$

In the equation above, $\lambda_{\alpha,t}$ are Lagrange multipliers associated to the $m$ constraints in the third equation. The mechanical work done by the constrained dynamics along one path $q_t, t = 0, \ldots, T$ can be computed by

$$\mathcal{W}_{0,T}(q_t) = \sum_{\alpha=1}^m \int_0^T f_\alpha(q_s)\dot{z}_\alpha(s)ds.$$

The function $f_\alpha$ is the local mean force of the reaction coordinate, see [64] for more details. Finally, the Jarzynski formula relates the average work by the constrained simulation to the free energy difference by

$$e^{-\beta\left(Q_\lambda(z') - Q_\lambda(z)\right)} = \mathbb{E}\left[e^{-\beta\mathcal{W}_{0,T}(q_t)}\right]. \quad (6.12)$$

We refer to [90, 66, 64, 85] for more details regarding discretizations of this scheme.

There are many similarities between the indirect reconstruction scheme and the constrained process. Both 'pull' the previous microscopic sample $x_n$ (or $q_0$) to the new reaction coordinate value, although indirect reconstruction does not enforce this constraint. The downside of the constrained / non-equilibrium process is many paths are needed to obtain a precise estimate of the free energy difference. We have also found the non-equilibrium scheme to be slow due to the many projections during the simulation of (6.11).

## 6.1.2 Discussion

We end this section by making a list of necessary requirements for FEC methods regarding their use for computing the free energy in the mM-MCMC algorithm. These requirements are the following:

- Speed: The FEC method should be fast and have little overhead because we compute the free energy after every reconstruction step.

- Work well with Indirect Reconstruction: Indirect reconstruction is expensive itself, so the FEC method ideally use the same reconstructed samples to generate an approximation.

- Exact in Expectation: The FEC estimator must be exact in estimation, or any deterministic error must be controllable in term of the parameters of the FEC method.

We present a different FEC method that fulfils each of these criteria in the next section.

## 6.2  The Pseudo-Marginal Method

Computing the marginal distribution $\mu_\lambda$ from a joint distribution $\mu_{\text{ext}}$ is a central problem in many applications [6]. A relatively easy and tractable method to evaluate $\mu_\lambda(z)$ would be to use a Markov chain Monte Carlo (MCMC) method that generates samples $\{(z_i, x_i)\}$ from the joint distribution $\mu_{\text{ext}}(z, x)$, and then single out the macroscopic particles $\{z_i\}$ to form an empirical density. However, it is well established that because of stiffness in the Langevin dynamics (1.1), such MCMC methods can result in strongly correlated samples, an undesirable property. On the other hand, if $\mu_\lambda(z)$ were analytically known or was easy to compute, one might simply generate samples from $\mu_\lambda(z)$ in a Markov chain Monte Carlo fashion. Indeed, if the Markov chain is at position $z$, the MCMC method generates a new sample $z'$ using some proposal distribution $q(z, z')$. We accept this proposal with probability

$$\alpha(z'|z) = \min\left\{1, \frac{\mu_\lambda(z')q(z|z')}{\mu_\lambda(z)q(z'|z)}\right\} \tag{6.13}$$

which could be evaluated analytically.

The pseudo-marginal method has been designed to combine the benefits of both approaches [4, 2, 89, 5]: possible computational and statistical efficiency gains by directly sampling from $\mu_\lambda(z)$ and ease of implementation by using auxiliary microscopic variables that sample from $\mu_{\text{ext}}(x, z)$. A particularly natural approach for estimating the intractable marginal density $\mu_\lambda(z)$ is importance sampling. That is, for some integer $K \geq 1$ and some conditional

distribution $q_z$, we consider the estimates

$$\tilde{\mu}_\lambda(z) = \frac{1}{K} \sum_{k=1}^{K} \frac{\mu_{\text{ext}}(x_k, z)}{q_z(x_k)}, \ x_k|z \sim q_z, \text{iid}, \tag{6.14}$$

$$\tilde{\mu}_\lambda(z') = \frac{1}{K} \sum_{k=1}^{K} \frac{\mu_{\text{ext}}(x'_k, z')}{q'_z(x'_k)}, \ x'_k|z' \sim q'_z, \text{iid} \tag{6.15}$$

These estimates are unbiased for $\mu_\lambda$ as long as all microscopic variables are independent. Let us denote the microscopic samples of (6.14) by $\mathcal{X} = \{x_1, ..., x_k\}$ and $\mathcal{X}' = \{x'_1, ..., x'_K\}$. These microscopic variables then have densities $q_z(\mathcal{X})$ and $q_{z'}(\mathcal{X}')$, respectively.

With the above estimates for the marginal distribution $\mu_\lambda(z)$ we can complete the Markov chain that we introduced at the top of the section. Specifically, if we plug $\tilde{\mu}_\lambda(z)$ and $\tilde{\mu}_\lambda(z')$ into acceptance probability (6.13), we obtain

$$\tilde{\alpha}(z'|z) = \min \left\{ 1, \frac{\left[ \frac{1}{K} \sum_{k=1}^{K} \frac{\mu_{\text{ext}}(x'_k, z')}{q'_z(x'_k)} \right] q(z|z')}{\left[ \frac{1}{K} \sum_{k=1}^{K} \frac{\mu_{\text{ext}}(x_k, z)}{q_z(x_k)} \right] q(z'|z)} \right\}.$$

However, there is no guarantee that we are still sampling from the marginal distribution $\mu_\lambda(z)$. A remarkable property [6] is that we can rewrite the above (approximate) acceptance rate as

$$\tilde{\alpha}(z'|z) = \min \left\{ 1, \frac{\left[ \frac{1}{K} \sum_{k=1}^{K} \mu_{\text{ext}}(x'_k, z') \Pi_{l=1, l \neq k}^{K} q_{z'}(x'_l) \right] q(z|z') q_z^K(\mathcal{X})}{\left[ \frac{1}{K} \sum_{k=1}^{K} \mu_{\text{ext}}(x_k, z) \Pi_{l=1, l \neq k}^{K} q_z(x_l) \right] q(z'|z) q_z'^K(\mathcal{X}')} \right\}. \tag{6.16}$$

This rearrangement of the (estimated) acceptance rate suggests that the pseudo-marginal method is a Markov chain on the larger state space $(\mathcal{X}, z)$, not just in $z$. This Markov chain has transition probability $q(z'|z) \, q_{z'}(\mathcal{X})$ and its invariant measure is in between the square brackets. This formula implies that the marginal distribution in $z$ of

$$\tilde{\mu_{\text{ext}}}(\mathcal{X}, z) = \frac{1}{K} \sum_{k=1}^{K} \mu_{\text{ext}}(x_k, z) \ \Pi_{l=1, l \neq k}^{K} q_z(x_l),$$

is still $\mu_\lambda(z)$. That is, one is still guaranteed to sample from the correct marginal distribution.

It is possible to further optimize the pseudo-marginal method by noting that we must only compute $\tilde{\mu}_\lambda(z')$ once. Indeed, on acceptance of $z'$ a new macroscopic

sample is generated and the exact same term now appears in the denominator of (6.16). This adaptation of the pseudo-marginal scheme is called the Grouped Independence Metropolis-Hastings method (GIMH) [6, 4], and is used in all our implementations of the method.

# 6.3 Adapting the Pseudo-Marginal Method to Reconstruction

Having established the pseudo-marginal method as an algorithm for sampling from and evaluating the marginal distribution, we now turn our attention to approximating the normalization constant $\mu_\lambda$ inside the microscopic acceptance probability (4.14). In section 6.3.1, we explain the main changes to the pseudo-marginal method to tailor it specifically to the mM-MCMC method. Then, in section 6.3.2, we present a faster alternative for practical implementations. Afterwards we illustrate the full mM-MCMC method with the pseudo-marginal approximation in section 6.3.3.

## 6.3.1 A Histogram Approximation to $q_z$

During indirect reconstruction, a sequence $\mathcal{X}_K = \{x_k\}_{k=1}^K$ of microscopic particles is sampled from the indirect reconstruction distribution $\nu_\lambda$, at least in the limit of large $K$. It therefore makes sense to choose $q_z = \nu_\lambda(\cdot|z)$ as our importance distribution in the pseudo-marginal method. The reconstructed microscopic samples are then used to estimate the free energy $\mu_\lambda$ using (6.14). There is thus an intimate connection between indirect reconstruction and pseudo-marginalization.

There is one more problem, however. If we look back to the formula for the indirect reconstruction distribution (4.10), the (filtered) free energy $\mu_\lambda$ is the normalization constant of said distribution. That is, we need to know $\mu_\lambda$ in advance to write an estimator for $\mu_\lambda$, which is nonsensical. We can overcome this issue by not using $\nu_\lambda$ itself as $q_z$, but rather an approximation to it that is explicitly normalized. A histogram is the natural tool for this.

When the microscopic state space is high dimensional, there is no guarantee that the histogram approximation to the indirect reconstruction distribution will be close to that distribution in any (pseudo-)norm. Fortunately, such accuracy is not needed for a proper implementation of the pseudo- marginal method, in fact, any qz is fine. However, because the histogram approximation is built on samples generated by indirect reconstruction, we can expect that

the histogram will have a similar shape compared to the indirect reconstruction distribution. For the same reason, the histogram implicitly depends on the reaction coordinate value z, and we will add it as a subscript to the histogram for clarity of exposition.

## Constructing the Histogram

To create a histogram $H_z[\mathcal{X}_K]$ that approximates $\nu_\lambda$, we discretize $\mathbb{R}^{3D}$ into equally sized square blocks or 'bins'. Suppose $h$ is the bin length, so that $h^{3D}$ is the total volume of one bin. Before actually constructing the histogram, let us introduce some notation. A particle $q \in \mathbb{R}^{3D}$ falls into bin $(l_1, l_2, \ldots, l_{3D-1}, l_{3D}) \in \mathbb{Z}^{3D}$ if

$$l_i h \leq q^{(i)} < (l_i + 1)h, \tag{6.17}$$

for all $i$ from 1 to $3D$. We use the notation $q^{(i)}$ to denote the $i$-th component of $q$. In equation (6.17), we implicitly assume that the origin is a corner of the 'central' bin.

Let $K_{(l_1, l_2, \ldots, l_{3D})}$ be the number of reconstructed samples that fall into bin $(l_1, l_2, \ldots, l_{3D})$, where

$$\sum_{l_1, l_2, \ldots, l_{3D} \in \mathbb{Z}} K_{(l_1, l_2, \ldots, l_{3D})} = K.$$

We then construct the histogram $H_z[\mathcal{X}_K](x)$ as

$$H_z[\mathcal{X}_K](x) = \sum_{l_1, l_2, \ldots, l_{3D} \in \mathbb{Z}} \mathcal{I}_{(l_1, l_2, \ldots, l_{3D})}(x) \frac{K_{(l_1, l_2, \ldots, l_{3D})}}{K}, \tag{6.18}$$

for each $x \in \mathbb{R}^{3D}$. In the notation above, $\mathcal{I}_{(l_1, l_2, \ldots, l_{3D})}(x)$ is the indicator function defined by

$$\mathcal{I}_{(l_1, l_2, \ldots, l_{3D})}(x) = \begin{cases} 1 & \text{if} \quad x \in \prod_{i=1}^{3D} \ [l_i h, (l_i + 1)h] \\ 0 & \text{else.} \end{cases}$$

Of course, we cannot sum (6.18) over $l_i$ from negative infinity to positive infinity. Most summands will be zero anyway due the nature of the indicator function. One can truncate this expression by only summing over all bins that contain at least one particle in practice. The histogram evidently remains the same.

**Calculating the Free Energy**

With the histogram definition of $q_z$, we can estimate $\mu_\lambda(z)$ by (6.14). To calculate an estimate, we must draw $K$ i.i.d. samples $\{y_k\}_{k=1}^K$ from $H_z[\mathcal{X}_K](x)$. Suppose we have condensed the active bins to a list $\{B_j, K_j\}_{j=1}^J$, where $B_j$ is the actual bin set and $K_j$ is its particle count.

We construct $y_k$ in two steps. First, draw a uniform number $u \sim U(0,1)$ between 0 and 1. The particle $y_k$ falls in the $j$'th bin if

$$\frac{\sum_{m=1}^{j-1} K_m}{K} \leq u < \frac{\sum_{m=1}^{j} K_m}{K}.$$

Second, draw a $3D$ dimensional sample from the uniform distribution on $B_j$. Carrying out this procedure for all samples $\{y_k\}_{k=1}^K$, the estimator for $\mu_\lambda(z)$ reads

$$\tilde{\mu}_\lambda(z) = \frac{1}{K} \sum_{k=1}^{K} \frac{\mu_{\text{ext}}(y_k, z)}{H_z[\mathcal{X}_K](y_k)}. \tag{6.19}$$

The histogram construction that we explained in this section is essentially inefficient. The reasons are twofold. In a practical implementation, we iterate over all reconstructed particles until we find a particle that has not been assigned a bin yet. Then, we loop over all remaining particles to calculate the bin count. This algorithm is essentially of $\mathcal{O}(K^2)$ complexity. There is no way to order the particles in advance to bring this complexity down.

Secondly, drawing samples from the histogram is non-obvious because we generally do not know which bins are active. Constructing the list of active bins adds a layer of complexity to code. We present a different approach to building the histogram that mitigates these inefficiencies in the next section.

## 6.3.2   A Tensorized Histogram for Faster Estimates

There are many ways to define a histogram that can be used t estimate $\mu_\lambda$. The essential property is that the histogram must be explicitly normalized. In this section we construct an alternative tensorized histogram that mitigates the computational bottlenecks from previous section. This approach rests on the fact that if $H_z^{(i)}[\mathcal{X}_K^{(i)}](x^{(i)})$ is a histogram in dimension $i$, then

$$H_z[\mathcal{X}_K](x) = \prod_{i=1}^{3D} H_z^{(i)}[\mathcal{X}_K^{(i)}](x^{(i)}) \tag{6.20}$$

is a histogram in $3D$ dimensional space. We use the decomposition $x = \left(x^{(1)}, \ldots, x^{(3D)}\right) \in \mathbb{R}^{3D}$.

### Constructing the Histogram

Constructing a one dimensional histogram is then similar to section 6.3.1, with the exception that we first sort the reconstructed particles $\mathcal{X}_K^{(i)}$ in ascending order. This brings the complexity of placing. particles in bins down to $\mathcal{O}(K \log K)$. In dimension $i$, let $K_l^{(i)}$ be the number of reconstructed particles $\mathcal{X}_K^{(i)}$ that fall into the interval $[lh, (l+1)h]$. Obviously,

$$\sum_{l=-\infty}^{\infty} K_l^{(i)} = K.$$

With these frequencies, the histogram in dimension $i$ reads

$$H_z^{(i)}[\mathcal{X}_K^{(i)}](x) = \frac{1}{K} \sum_{\in \mathbb{Z}} \mathcal{I}_{[lh,(l+1)h]}(x) K_l^{(i)}, \ x \in \mathbb{R} \tag{6.21}$$

with $\mathcal{I}_{[lh,(l+1)h]}(x)$ the indicator function on the interval $[lh, (l+1)h]$.

### Calculating the Free Energy

We again draw $K$ i.i.d. samples $\{y_k\}_{k=1}^{K}$ from $H_z[\mathcal{X}_K]$. This is an easy task because of the tensorized definition (6.20). Indeed, to generate a sample $y_k^{(i)}$ from the one-dimensional histogram $H_z^{(i)}[\mathcal{X}_K^{(i)}]$, draw a uniform number $u$ in $[0,1]$ and find the interval with number $m$ for which

$$\frac{\sum_{l=-\infty}^{m-1} K_l^{(i)}}{K} \leq u < \frac{\sum_{l=-\infty}^{m} K_l^{(i)}}{K}.$$

Finally, draw a one dimensional microscopic sample $y_k^{(i)}$ uniformly in $[mh, (m+1)h]$. The combined sample in all $3D$ dimensions, $y_k = (y_k^{(1)}, \dots, y_k^{(3d)})$, is then distributed according to (6.20). With these microscopic samples, an estimator for $\mu_\lambda$ finally reads

$$\tilde{\mu}_\lambda(z) = \frac{1}{K} \sum_{k=1}^{K} \frac{\mu_{\text{ext}}(y_k, z)}{H_z[\mathcal{X}_K](y_k)}. \tag{6.22}$$

**Remark 7.** *When sampling from $H_z[\mathcal{X}_K]$ is on the critical computational path in an application, this step may be dropped entirely. It is possible to write an estimator for $\mu_\lambda$ using only the reconstructed samples $\{x_k\}_{k=1}^{K}$ as follows*

$$\tilde{\mu}_\lambda(z) = \frac{1}{K} \sum_{k=1}^{K} \frac{\mu_{ext}(x_k, z)}{H_z[\mathcal{X}_K](x_k)}.$$

*That is, we use these microscopic samples to both construct ánd evaluate the histogram. This approach is not exact, but it can speed up computations if necessary.*

### Computing the Microscopic Acceptance Probability

Putting everything together from this section, our final pseudo-marginal approximation of the microscopic acceptance probability reads

$$\alpha_f(x', z'|x_n, z_n) = \min\left\{1, \frac{\tilde{\mu}_\lambda(z')\bar{\mu}_0(z_n)}{\tilde{\mu}_\lambda(z_n)\bar{\mu}_0(z')}\right\}. \tag{6.23}$$

We used the same notation for the microscopic and macroscopic particles as in chapter 4.

## 6.3.3 Complete Algorithm

The complete mM-MCMC method with pseudo-marginal approximation is illustrated in figure 6.1.

# 6.4 Numerical Results

Let us now review how the mM-MCMC method with pseudo-marginal approximation performs on molecular examples. We study the method on two molecules, butane and alkane. Butane is a simple molecule consisting of four chained carbon atoms, surrounded by hydrogen atoms. All chemical bonds are of covalent type. Alkane is a single chain that contains $D$ carbon atoms, and can be regarded as an extension of butane. The latter molecule allows us to study how the quality of pseudo-marginal approximation depends on the number of dimensions of the system.

Afterwards, we carry out the experiments on butane and alkane in sections 6.4.1 and 6.4.2 respectively.

## 6.4.1 Butane

The butane molecule is a single chain of four carbon atoms. The middle two atoms have two hydrogens and the other two have three. The Cartesian coordinates of these atoms are given by $x \in \mathbb{R}^12$. We are interested in the
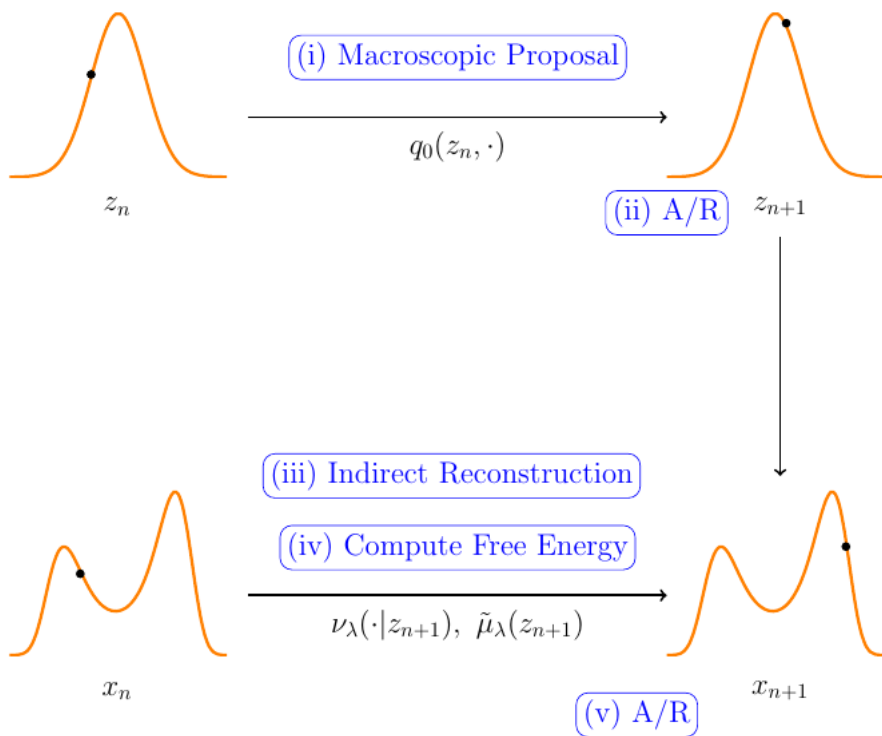
Figure 6.1: Illustration of the five steps in the mM-MCMC algorithm with pseudo-marginal approximation: macroscopic proposal, macroscopic accept/reject, indirect reconstruction, pseudo-marginal approximation of $\tilde{\mu}_\lambda$, and microscopic accept/reject.

torsion angle of the main chain, which will be our reaction coordinate. The potential energy surface (PES) for this molecule is

$$V(x) = \sum_{CH_x-CH_y} \frac{1}{2}k_b(r(x) - r_0)^2 + \sum_{CH_x-CH_y-CH_z} \frac{1}{2}k_a(\theta(x) - \theta_0)^2 + A(\tau(x)),$$

(6.24)

where the bond lengths $r(x)$, angles $\theta(x)$ and torsion angle $\tau(x)$ are all functions of the microscopic positions $x$. The first sum is taken over the three C-C bonds, the second over both C-C-C angles. The third term is the energy of the torsion angle. There are no electrostatic or Vanderwaals interactions [88]. The constants have values $k_b/k_B = 319225 K/\text{Å}^2$, $k_a/k_B = 62500 \ K, r_0 = 1.540\text{Å}$ and $\theta_0 = 114°$. Here, $k_B$ is the Boltzmann constant in appropriate units. The

free energy of the reaction coordinate is

$$A(z) = A(\tau(x)) = \sum_{i=0}^{3} c_i \cos(\tau(x))^i, \qquad (6.25)$$

$c_0/k_B = 1031.36\ K$, $c_1/k_B = 2037.82\ K$, $c_2/k_B = 158.52\ K$, $c_3/k_B = -3227.70\ K$.

We perform two experiments with mM-MCMC on butane. First, we show the histogram approximations by the mM-MCMC method with pseudo-marginal approximation to invariant probability distributions $\mu_\lambda$ and $\mu_0$. This experiment is in section 6.4.1. Afterwards, we plot all estimated values for $Q_\lambda$ as a function of $z$, obtained during a single sample run of the mM-MCMC method in section 6.4.1.

**Inspection of mM-MCMC with Pseudo-Marginal Approximation**

We run the mM-MCMC algorithm with pseudo-marginal approximation for a total of $N = 10^6$ sampling steps. On the macroscopic level, we use Brownian increments to sample reaction coordinate values with a time step $\Delta t = 0.001$. The biased simulation has parameters $\lambda = 2.0\ k_b$, time step $\delta t = 0.01/\lambda$ and number of steps $K = 15$. The bin size of the histogram is $\sqrt{\frac{1}{2\lambda}}$. The torsion angles obtained by mM-MCMC with pseudo-marginal approximation are demonstrated in figure 6.2. Left is the histogram of the reaction coordinate samples obtained on the macroscopic level, right those of the microscopic samples.

Both histograms follow $\mu_\lambda(z)$ and respectively $\mu_0(z)$ well, even though there is an added variance because of the pseudo-marginal approximation. The total macroscopic acceptance rate is 0.334, and the microscopic acceptance rate is 0.75. This number is smaller than the expected microscopic acceptance rate of 1.0 due to the added variance by the pseudo-marginal estimation. Nevertheless, the approximation is unbiased.

**Variance on Free Energy Estimates $\hat{Q}_\lambda(z)$ as a Function of Reaction Coordinate Value $z$**

The induced variance on $\mu_\lambda$ decreases with $K$ (6.14), but can vary with $z$. In this experiment, we run the mM-MCMC method with the same parameters as above, and record all values $\tilde{Q}_\lambda(z)$ seen during a sample run. We plot these values as well as the true free energy $Q_\lambda$ in figure 6.3.
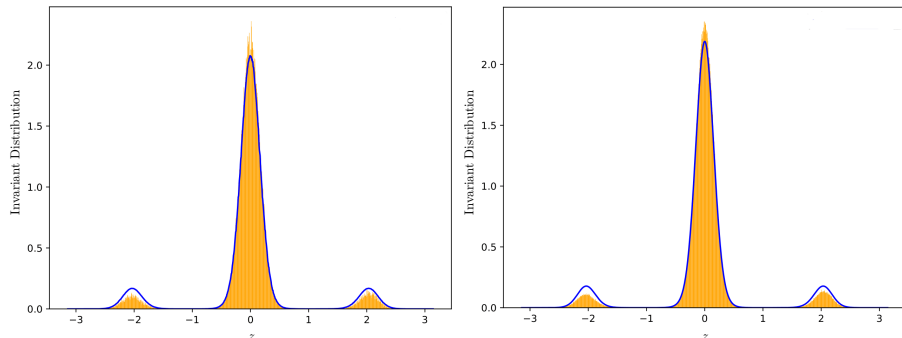
Figure 6.2: Histogram fit (orange) of a typical run of the mM-MCMC method with pseudo-marginal approximation on $\tilde{Q}_\lambda$ (blue, left) and the free energy (blue, right).

There are two things to note. First, there is a very large variance on $\tilde{Q}_\lambda$. For two reaction coordinate values close together, the pseudo-marginal method computes $\mu_\lambda$ independently. These values can vary significantly, there is no continuity. Second, the actual approximate values follow the free energy profile well, there is no bias depending on $z$.

## 6.4.2  Alkane

An alkane molecule consists of a single chain of $D$ carbon atoms, with each carbon having two bonded hydrogen atoms. The carbon atoms at the end of the chain have an extra bonded hydrogen atom. It can be regarded as an extension of butane. Our motive for choosing this molecule is that it allows us to study how properties of the mM-MCMC method with pseudo-marginal approximation vary with dimensionality.

With this objective in mind, we have designed four numerical experiments in this section. In the first experiment in section 6.4.2, we study the averaged microscopic acceptance rate as a function of $D$. We expect the microscopic acceptance rate to decrease as $D$ grows, but we do not know how fast. Then in section 6.4.2, we study the evolution of the averaged microscopic acceptance rate as function of the number of reconstructed particles $K$, for different values of $D$. Afterwards, we investigate the variance of the histogram estimator (6.22) for $\mu_\lambda$ as a function of $z$ and the number of reconstructed particles $K$. This experiment is in section 6.4.2. Finally, we combine everything by plotting the total efficiency gain on the expected mean and variance of $z$ as a function of $D$
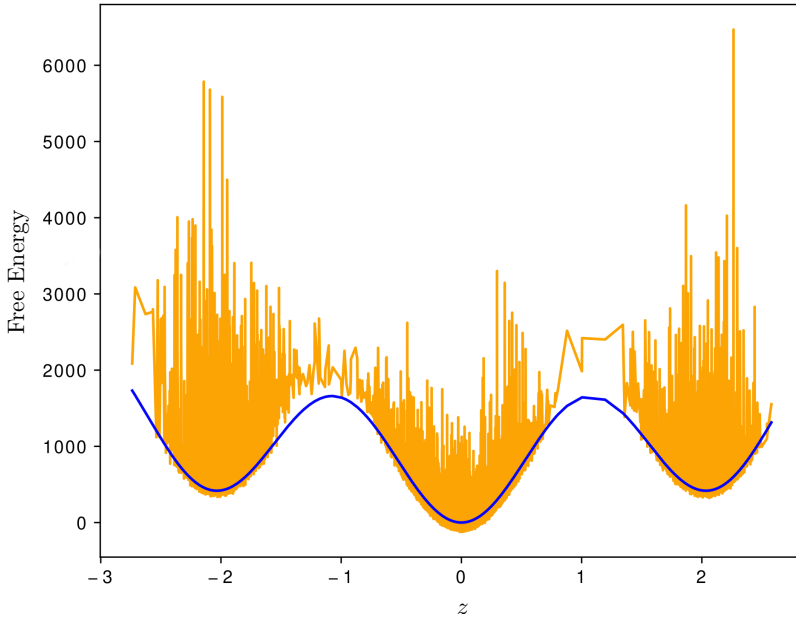
Figure 6.3: Value of $\tilde{Q}_\lambda(z)$ recorded during a mM-MCMC sample run (orange), as well as the exact free energy $Q_\lambda$ (blue).

and $K$ in section 6.4.2.

## Averaged Microscopic Acceptance Rate as a Function of $D$

In this experiment, we analyze how the microscopic acceptance rate (6.16) depends on $D$, the number of carbon atoms. The reaction coordinate is the torsion angle between the first four carbon atoms. For each value of $D$ between 4 and 45, we run the pseudo-marginal mM-MCMC algorithm for $10^5$ sampling steps. The macroscopic time step is $\Delta t = 0.001$, and number of indirect reconstruction steps $K = 20$ with step size $\delta t = 0.01 \; \lambda$. The constraining parameter is $\lambda = 2k_b$. The average microscopic acceptance rate as a function of $D$ is pictured in figure 6.4.

The acceptance rate decreases exponentially with $D$. This behaviour is typical for MCMC algorithms [22], where the acceptance rate decreases exponentially with the number of dimensions of the system. Afterwards, however, the averaged microscopic acceptance rate converges to an equilibrium value. This value is
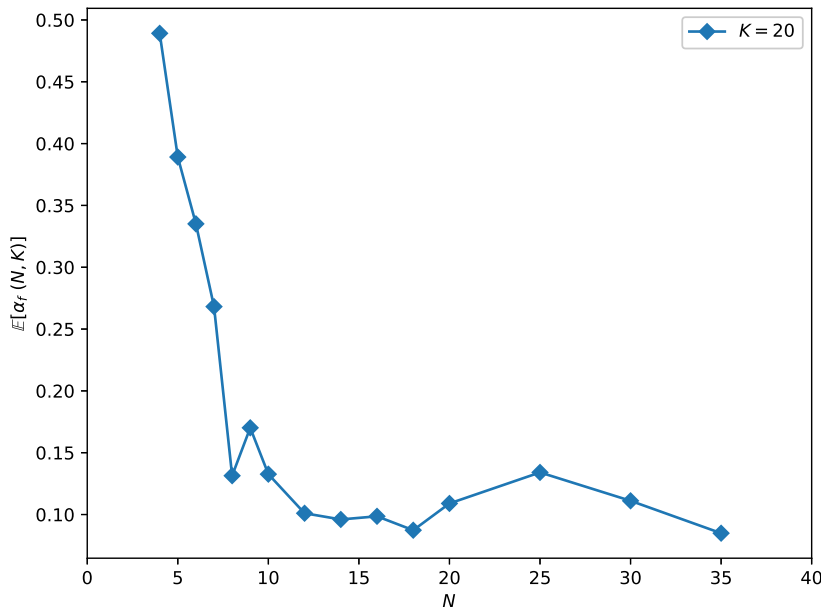
Figure 6.4: The averaged microscopic acceptance rate as a function of the number of carbon atoms (blue). Black lines indicate sampled values. The microscopic acceptance rate first decreases exponentially, then converges to an equilibrium value.

small, but can be tweaked by the number of reconstruction steps. We explore this phenomenon further in the next experiment.

### Averaged Microscopic Acceptance Rate as a Function of $K$

We saw in the previous experiment that the averaged microscopic acceptance rate first decreases, and then seemingly converges to some 'limit' acceptance rate. The number of reconstructed particles was, however, on the low side. This analysis forms the basis of the current experiment. Our question is how the averaged microscopic acceptance rate depends on $K$, the number of reconstructed particles. The setup of this experiment is as follows. We run the mM-MCMC method with pseudo-marginal approximation for the same values of $D$ as above, and for $K = 10, 20, 40, 100$ and $200$. The number of sampling steps is $10^5$. For each combination of parameters, we average the microscopic acceptance probability over 100 independent runs. These curves are being shown in figure 6.5.
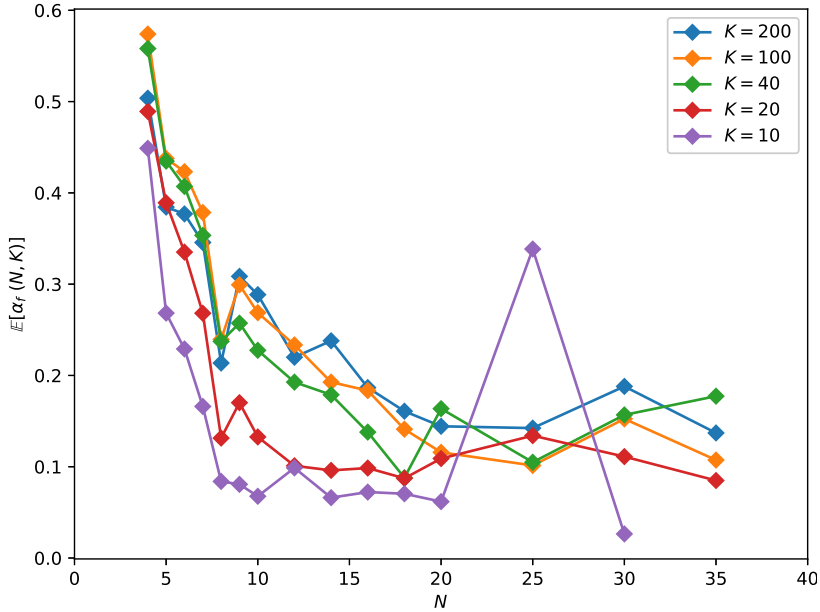
Figure 6.5: The averaged microscopic acceptance rate over 100 independent runs as a function of $D$. We show these curves for $K = 10$ (purple), $K = 20$ (red). $K = 40$ (green), $K = 100$ (orange), and $K = 200$ (blue).

First of all, we again see the exponential decay of the averaged microscopic acceptance rate as $D$ gets larger. We already reached this conclusion in the last experiment. Second, the 'limit' value increases with growing $K$. Indeed, the more reconstructed particles at each iteration of the mM-MCMC method, the lower the averaged variance on estimates of $\mu_\lambda$ will be. Then, according to (6.3) the (averaged) microscopic acceptance probability should increase because it is fundamentally limited by uncertainty on $\mu_\lambda$.

### Variance on $\mu_\lambda$ as a Function of $K$ and $z$

The analysis from the previous experiment relies on the fact that the variance on estimates of $\mu_\lambda(z)$ decreases with increasing $K$. We test this hypothesis here in this section. In this experiment, we plot the variance of statistical estimates of $\mu_\lambda$ as a function of $z$, and this for several values of $K$, the number of reconstructed particles. In a typical run of the mM-MCMC algorithm, there will never be two or more estimates at the same level of $z$ because the macroscopic proposals are everywhere between $-\pi$ and $\pi$. We need to aggregate estimates

close to the current value of $z$, and weigh them according to their distance with respect to the current abscissa. A Gaussian weight function is a natural approach.

Suppose $\{z_n, \tilde{Q}_\lambda(z_n)\}_{n=1}^N$ are the respective reaction coordinate values and estimated values for $Q_\lambda$ obtained during a run of the mM-MCMC method. Then, for a fixed reaction coordinate value $z$, the mean and variance of $\tilde{Q}_\lambda$ are given by

$$\mathbb{E}[\tilde{Q}_\lambda|z] = \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} \tilde{Q}_\lambda(u) \, d\tilde{Q}_\lambda(u) \, \delta_{u-z}(du) \tag{6.26}$$

$$\mathbb{V}[\tilde{Q}_\lambda|z] = \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} \left(\tilde{Q}_\lambda(u) - \mathbb{E}[\tilde{Q}_\lambda|z]\right)^2 \, d\tilde{Q}_\lambda(u) \, \delta_{u-z}(du) \tag{6.27}$$

As mentioned above, we then approximate the delta function by a sharp Gaussian factor centred at $z$ and with standard deviation $\varepsilon \ll 1$. The discrete, particle approximations to the integrals in (6.26) then read

$$m(z) = \frac{1}{N} \sum_{n=1}^N \tilde{Q}_\lambda(z_n) \, \frac{1}{\sqrt{2\pi\varepsilon^2}} \exp\left(-\frac{(z_n - z)^2}{2\varepsilon^2}\right) \tag{6.28}$$

$$\sigma^2(z) = \frac{1}{N} \sum_{n=1}^N \left(\tilde{Q}_\lambda(z_n) - m(z)\right)^2 \, \frac{1}{\sqrt{2\pi\varepsilon^2}} \exp\left(-\frac{(z_n - z)^2}{2\varepsilon^2}\right). \tag{6.29}$$

In the remainder of this section, we show two figures. The first figure depicts the aggregated mean $m(z)$ for $z$ between $-\pi$ and $\pi$, averaged over 100 independent runs of the mM-MCMC algorithm with $10^5$ samples. The other parameters are the same as in section 6.4.1. The number of carbon atoms is 8. Further parameters of the mM-MCMC method are $K = 20$, $\Delta t = 0.001$, $\lambda = 2k_b$, and $\delta t = 0.01/\lambda$. The standard deviation of the Gaussian weight function is $\varepsilon = 0.02$. This plot, together with the free energy (6.25) are shown in figure 6.6.

In the second figure, we show the variance on estimates $\tilde{\mu}_\lambda(z_n)$ for different values of $K$ These values are $10, 20, 40, 100, 200, 400, 1000$, and $2000$. All other parameters remain the same. These variance plots are shown in figure 6.7.

The results are clear. On figure 6.6 the averaged curve $m(z)$ follows $A(z) \approx Q_\lambda(z)$ closely, especially in regions where the density of reaction coordinate values are large. These regions are centred at $-2, 0$ and $2$. Only at the ends of the figure do the curves diverge because there are very few samples there to get an accurate estimate.

We see a similar pattern in figure 6.7. First of all, the variance on estimates $\tilde{Q}_\lambda$ does indeed decrease when $K$ increases. This decrease happens at a steady pace,
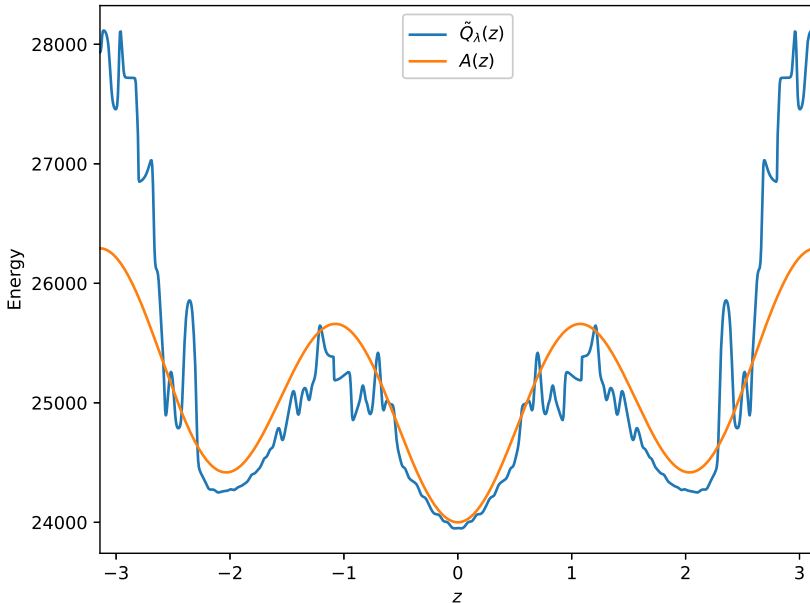
Figure 6.6: Average of estimates for $\tilde{Q}_\lambda(z)$ averaged over 100 independent runs (blue), and the free energy $A(z)$ (orange).

note the vertical axis uses a log-scale. This decrease confirms our statement from section 6.4.2. Second, the variance is lowest in regions of high density, and the variance is largest at both ends of the figure, especially for large $K$.

**Efficiency gain as a Function of $K$ and $D$**

The only factor we have not taken into account yet is the cost of reconstruction. We have shown in section 6.3.2 that computing $\mu_\lambda$ has a log-linear complexity in $K$. Although the total sampling accuracy grows with larger $K$, so does the total cost. It is unclear how the sampling accuracy grows with the number of reconstruction steps. We test this dependency here.

In this experiment, we plot the total efficiency gain of the mM-MCMC method with pseudo-marginal approximation over the microscopic MALA method 3.3.1 for several values of $D$ and $K$. We specifically measure the efficiency gain with respect to the estimated mean and variance of the reaction coordinate. In the
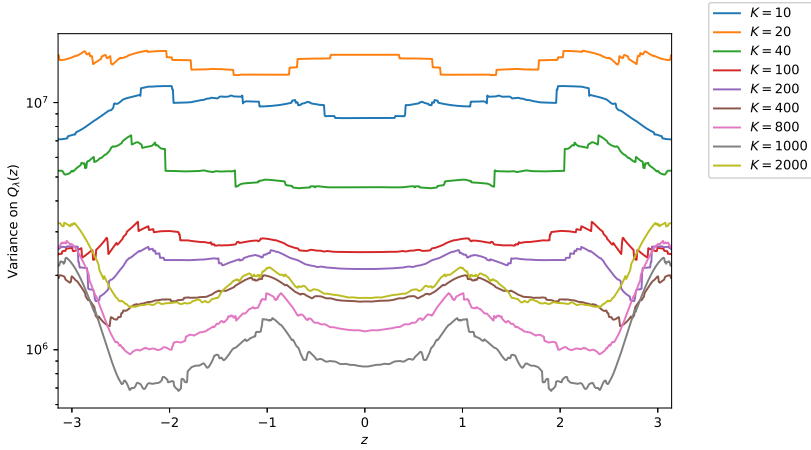
Figure 6.7: Variance of estimates for $\tilde{Q}_\lambda(z)$ averaged over 100 independent runs for several values of $K$, shown in the legend.

language of (5.1), the functions $F$ are respectively

$$\begin{cases} F_{\text{mean}}(z) & = z \\ F_{\text{variance}}(z) & = z^2. \end{cases}$$

The setup for this experiment is as follows. For each value of $D$, we run the mM-MCMC method for $10^5$ sampling steps. The macroscopic time step is $\Delta t = 0,001$, and $\lambda = 2k_b$. The number of reconstruction steps $K$ are $10, 20, 40, 100$ and $200$. The number of sampling steps of the microscopic MALA method is also $10^5$, and its time step is $\delta t = 0.01/\lambda$. We compute the efficiency gain over 100 independent runs. The results are shown in figure 6.8.

The results are unexpected. The mM-MCMC method is on average $\mathbb{E}[\alpha_{CG}]K \log K$ times slower per microscopic sample than the MALA method. For $K = 20$ and an averaged macroscopic acceptance rate of 0.3, it comes down to a factor 7.8 slower per sample. Still, the efficiency gain on $F_{\text{mean}}$ and $F_{\text{variance}}$ increase with increasing $K$, a very positive result. The reason can only be that mM-MCMC gets more and more accurate with increasing $K$, enough to offset the ratio of runtimes in (5.3). We hope the reader will be convinced of the benefit of the mM-MCMC method.

Let us now look closer at the details of figure 6.8. One can see that there are dips in the efficiency gain curves for specific values of $N$. These values are $N = 5, 8, 20,$ and $35$. For $N = 5, 8,$ and $35$, the dips only manifest in the efficiency gain on the expected mean, although for $N = 20$, both curves have
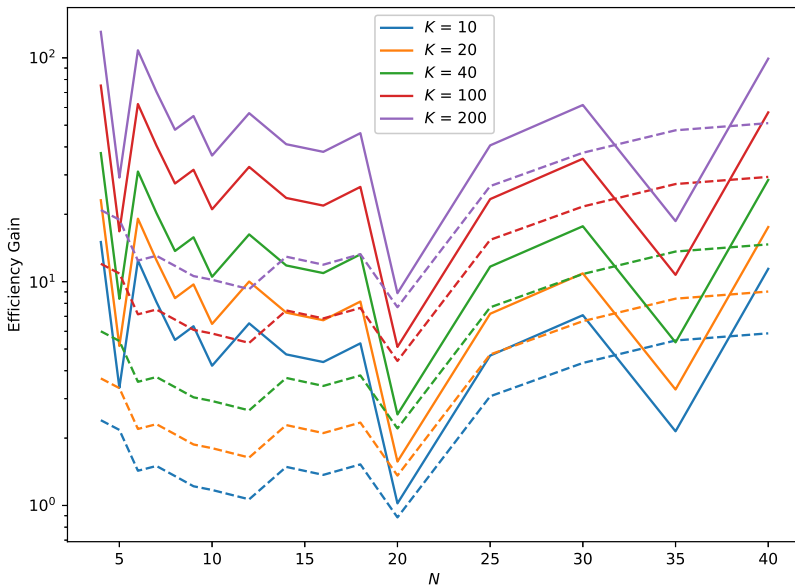
Figure 6.8: Efficiency gain of the mM-MCMC method with pseudo-marginal approximation over the microscopic MALA method on the mean (full) and variance (dashed) of the torsion angle for several values of $K$.

a sudden dip. We generated this figure by averaging the estimated efficiency gains over 100 independent runs. Each combination of $N$ and $K$ is also treated independently. We ran all experiments in parallel on the same type of CPU to make the timings comparable. Because the curves can be different due to the stochasticity in the pseudo-marginal method, we reran it one and received the same curve qualitatively, but with the same dips at the same locations. This leaves us with two possible explanations. Either the supercomputer was overloaded during the period it computed the efficiency gains for those specific values of $N$, and that twice, or there must be a theoretical explanation. The odds of the former happening are small, which leaves us only with the latter explanation. We will therefore do an in depth theoretical analysis of the efficiency gain as a function of $N$. This analysis, however, is not in the scope of this thesis.

# Chapter 7

# Backmapping as an Alternative to Reconstruction

We have observed in the previous chapters that the mM-MCMC method with indirect reconstruction works well when the reaction coordinate is low-dimensional. However, for molecules that are important practice, such as proteins, DNA and polymers, the reaction coordinate typically grows linearly with the number of dimensions of the microscopic system. In this chapter, we consider the case of one type of macromolecule, proteins.

The computational issues with high-dimensional reaction coordinates are two-fold. First, the number of soft constraints of the form $\frac{\lambda\beta}{2} \|\xi(x) - z'\|^2$ become hard to satisfy when one can only use a small number of biased steps ($K$). Especially when $\lambda$ is large, the ball $B(z', (\lambda\beta)^{-1})$ is small, and it becomes increasingly difficult for indirect reconstruction to remain in equilibrium with $\nu_\lambda(\cdot; z')$. Second, we also need a lot of reconstructed samples to build an accurate histogram approximation of $\nu_\lambda$ in the pseudo-marginal method. We leave this topic for further examination.

In this chapter, we study the backmapping method as an alternative to indirect reconstruction. Backmapping was introduced in [52, 53, 54], itself an adaptation of the Adaptive Resolution Scheme (AdResS) [81, 55, 31]. Backmapping is in many ways similar to indirect reconstruction, except for two crucial differences. The first difference concerns the initial condition. For indirect reconstruction, the initial condition is simply the previous microscopic sample. Backmapping creates an initial condition through a step called 'lifting' [48, 87]. Lifting computes a microscopic sample that is consistent with the macroscopic reaction

coordinate value. This step is cheap when the macroscopic reaction coordinate is a linear map of the microscopic positions, such as in the MARTINI [71, 76] force field for proteins. The second difference is a time-dependent versus biased simulation. Because lifting is out-of-equilibrium, backmapping slowly switches from an only macroscopic dynamics to an only microscopic dynamics. This switching is necessary to make sure the simulation remains finite equilibrates.

We now come to the first of two contributions, the mM-MCMC method with backmapping. We can simply replace indirect reconstruction by lifting and backmapping, and we can run the mM-MCMC algorithm as usual. But, because backmapping is a non-equilibrium process, we loose the guarantee of exact sampling. Exact sampling is hardly an issue for proteomic systems, because finding minimum energy states is more important than sampling every part of the state space accurately. For the second and final contribution, we run the mM-MCMC scheme with backmapping on a single disulfide peptide, also known as 6kn2. We specifically compare the Ramachandran torsion angles by the mM-MCMC method against reference Ramachandran plots obtained by a microscopic sampler. We show that the mM-MCMC method is able to find additional local minima of the potential energy, displaying clear improvements.

This chapter is organised as follows. In section 7.1 we introduce the MARTINI coarse grained force fields for proteins, as well as the lifting produce to find a microscopic molecule that is consistent with the new macroscopic sample. Then in section 7.2, we explain the backmapping method in full detail. Next, we go over the mM-MCMC method with backmapping in section 7.3, before concluding this chapter with numerical results of the mM-MCMC method applied to the 6kn2 protein and a comparison with a microscopic sampler in section 7.4. The contents and results in this chapter have not yet been written down, but we will write a preprint that contains this work soon.

## 7.1 The MARTINI Coarse-Grained Force Field for Proteins

Proteins are a fundamental agent in biochemistry. They are responsible for for almost any task in the cell. Proteins provide organization, regulate energy and maintenance. Physically, a protein is a long chain, built from small and repeating units. These units are called 'amino-aids'. An amino-acid consists of a small backbone chain and a remainder group connected to that chain. They are linked together with strong nitrogen-carbon bonds, also called a peptide bonds. We show a general amino-acid in figure 7.1.
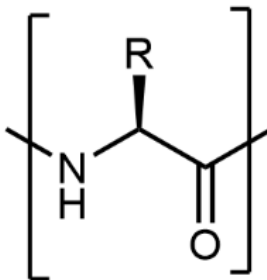
Figure 7.1: A single amino-acid. The main chain is $-NH - CR - CO-$, and the remainder group is denoted by $R$.

Because of the repeating structure in proteins, one can develop a coarse-grained potential separately for each amino acid, and glue them together afterwards to form a coarse-grained potential for the whole protein. One such approach is the MARTINI force field [76, 71]. The MARTINI reaction coordinate describes a united-atom model where the coarse atoms are obtained by the centres of masses of, on average four microscopic atoms. There is typically one large atom for the main chain, and other large atoms for the remainder group. We show a picture of the united atom model for each of the 21 amino-acids in appendix C. In our simulations, we had to infer the reaction coordinate ourselves through reverse engineering. The reaction coordinate alone is not enough, we need to know the free energy associated to the MARTINI reaction coordinate. A typical way to compute the free energy is through Boltzmann Inversion (BI) [77], but we assume the free energy is given. We do not give further details.

Similar to indirect reconstruction, we need to pick an initial condition for backmapping. In indirect reconstruction, this initial condition is the previous microscopic sample. The algorithm for choosing a proper initial condition for backmapping is called 'lifting', in reference to [87]. Call $z \in \mathbb{R}^m$ the MARTINI reaction coordinate value of $x \in \mathbb{R}^{3D}$. There is a convex relation between $x$ and $z$ given by

$$z = Mx, \tag{7.1}$$

where $M \in \mathbb{R}^{m \times 3D}$ is the normalized mass matrix. $M$ is a block-diagonal matrix, where each block corresponds to an amino-acid. To implement lifting, we first run a microscopic simulation of the protein up to a fixed end time. At the end of the simulation, we record the particular microscopic instance and its reaction coordinate value. Call these variables $x_{\text{frag}}$ and $z_{\text{frag}}$ respectively. We

can decompose these vectors per amino-acid

$$x_{\text{frag}} = \left( x_{\text{frag}}^1, x_{\text{frag}}^2, \ldots, x_{\text{frag}}^n \right) \tag{7.2}$$

$$z_{\text{frag}} = \left( z_{\text{frag}}^1, z_{\text{frag}}^2, \ldots, z_{\text{frag}}^n \right) \tag{7.3}$$

Now assume we sampled $z_{n+1}$ at the $n+1$ iteration on the macroscopic level. We compute an initial microscopic sample through lifting by considering each amino-acid independently. Unfortunately, the $M$ matrix is not square (and hence not invertible), we cannot simply compute

$$x_{n+1,0}^i - x_{\text{frag}}^i = M_i^{-1} \left( z_{n+1}^i - z_{\text{frag}}^i \right),$$

where $M_i$ is the block diagonal corresponding to the $i$-th amino-acid. Instead, we have to do lifting per united atom. Suppose $z^j$ is the $j$-th macroscopic atom, and suppose $x_{n+1,0}^j$ is the set of microscopic atoms that are mapped to $z_{n+1}^j$. We then compute

$$x_{n+1,0}^j - x_{\text{frag}}^j = z_{n+1}^j - z_{\text{frag}}^j. \tag{7.4}$$

because the MARTINI reaction coordinate takes convex combinations of the microscopic positions. We can now move on to backmapping.

## 7.2 Backmapping as an Alternative to Indirect Reconstruction

Let $x_{n+1,0}$ be the lifted microscopic sample after $n+1$ iterations. Because $x_{n+1,0}$ is in a sense arbitrary, it will not be in equilibrium with the Gibbs distribution. In fact, the forces $-\nabla V(x_{n+1,0})$ will be very large. That means that overdamped Langevin simulation

$$x_{n+1,k} = x_{n+1,k-1} - \delta t \nabla V(x_{n+1,k-1}) + \sqrt{2\delta t \beta^{-1}} \eta_k,$$

with $\eta_k \sim \mathcal{N}(0,1)$ will likely blow up to infinity for regular sized time steps. This behaviour is obviously not what we want for a stable execution of the mM-MCMC algorithm.

An improved approach is to use a mixing potential between the macroscopic and microscopic models, given by

$$V_t(x_{n+1,k}) = \lambda(t)^2 V(x_{n+1,k}) + \left( 1 - \lambda(t)^2 \right) A(M x_{n+1,k}).$$

In this mixing potential, $\lambda(t)$ is time-dependent function that evolves from 0 to 1. Initially, $\lambda(t)$ is zero, and only the MARTINI coarse-grained dynamics is

active. Because $x_{n+1,0}$ has been chosen consistent with $z_{n+1}$, the macroscopic force $-A'(Mx_{n+1,0})\nabla\xi(x_{n+1,0})$ will be small, and the simulation will be stable. As $\lambda(t)$ slowly increases, the microscopic dynamics becomes stronger, and macroscopic dynamics weaker, but the simulation will be stable. At the end, $\lambda(t)$ is 1, and only the microscopic dynamics remains, so the forces are small.

The full time-dependent backmapping dynamics then is

$$x_{n+1,k} = x_{n+1,k-1} - \delta t \nabla V_t(x_{n+1,k-1}) + \sqrt{2\delta t \beta^{-1}}\eta_k. \tag{7.5}$$

Time in this equation is defined by $t = k\delta t$, and $\lambda(t) = k/K$.

There is one caveat with backmapping as an alternative to indirect reconstruction. Backmapping (7.5) is a time-dependent process and is as a consequence out-of-equilibrium, while indirect reconstruction is not. We need to take this piece of information into account for the mM-MCMC with backmapping in the next section.

## 7.3 Micro-Macro Markov Chain Monte Carlo with Backmapping

Backmapping only replaces indirect reconstruction in the mM-MCMC algorithm, we can keep the other steps This section serves that purpose. As we saw in the previous section, backmapping creates a microscopic molecule that only depends on the corresponding macroscopic reaction coordinate value. Therefore, the mM-MCMC method with backmapping will look very similar to the mM-MCMC method with direct reconstruction. There is no need for an extended state space. We explain the mM-MCMC method with backmapping in section 7.3.1, and we show a schematic outline of this method in section 7.3.2.

### 7.3.1 Overview of mM-MCMC with backmapping

Suppose $x_n$ is the $n$-th sample on the Markov chain. We generate the next sample in four steps.

**Restriction** Compute the reaction coordinate value of the current microscopic sample, i.e.,

$$z_n = \xi(x_n).$$

For proteins, the reaction coordinate will be the MARTINI reduced model, based on centres of masses of a few connected atoms.

**Macroscopic Proposal**   Once on the macroscopic level, we can generate a new reaction coordinate value $z_{n+1}$ with larger proposal moves because the fast variables have been removed. That is, we propose a new reaction coordinate value $z_{n+1}$ using a macroscopic transition probability $q_0(\cdot|z_n)$, i.e.,

$$z_{n+1} \sim q_0(\cdot|z_n)$$

In the case of backmapping for proteins, the macroscopic proposal are based on Brownian increments, or an HMC time step.

**Macroscopic Accept/Reject**   To ensure that $z_{n+1}$ is indeed a sample of the MARTINI invariant distribution $\bar{\mu}_0$ of the reaction coordinate values, we accept $z_{n+1}$ with probability

$$\alpha_{CG}(z_{n+1}|z_n) = \min\left\{1, \frac{\bar{\mu}_0(z_{n+1})\ q_0(z_n|z_{n+1})}{\bar{\mu}_0(z_n)\ q_0(z_{n+1}|z_n)}\right\}, \tag{7.6}$$

which is the standard Metropolis-Hastings form for the acceptance rate. If $z'$ is accepted, we proceed to backmapping. If not, we return to the first step and set $x_{n+1} = x_n$.

**Backmapping**   If the reaction coordinate value $z_{n+1}$ has been accepted, we construct a new microscopic sample $x_{n+1,0}$ that is consistent with $z_{n+1}$ by fragmentation 7.1. This step in the literature is called 'lifting'. This lifting step is achieved by finding $x_0$ such that

$$z_{\text{frag}} - z_{n+1} = M\left(x_{\text{frag}} - x_{n+1,0}\right),$$

by means of an amino acid - by - amino acid construction.

After lifting, we generate a new microscopic sample by running the time-dependent simulation

$$x_{n+1,k+1} = x_{n+1,k} - \delta t\left(\lambda(k)^2 \nabla V(x_{n+1,k}) + (1-\lambda(k)^2)A'(\xi(xn+1,k))\nabla\xi(x_{n+1,k})\right)$$

$$+ \sqrt{2\delta t\beta^{-1}}\eta_k,$$

with $\lambda(k) = k/K$, and $\eta_k \sim \mathcal{N}(0,1)$. Afterwards, set $x_{n+1} = x_{n+1,K}$.

There is one major difference between mM-MCMC with backmapping and mM-MCMC with direct reconstruction.   That difference is the existence of a microscopic acceptance probability.   As we mentioned in section 7.2, backmapping is an out-of-equilibrium stochastic process. That means that we cannot define an equilibrium microscopic acceptance probability, nor can we prove that the Gibbs distribution is invariant under this algorithm. The upside is that the pseudo-marginal method is no longer necessary for backmapping.
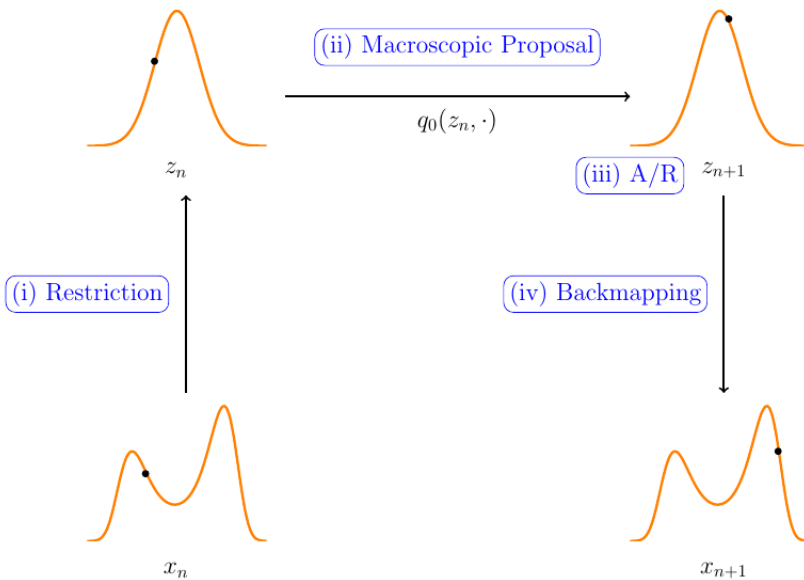
Figure 7.2: Schematic outline of the mM-MCMC algorithm with backmapping. The method consists of four steps: Restriction, Macroscopic Proposal, Macroscopic Accept/Reject, and Backmapping.

### 7.3.2 Complete Algorithm

A schematic version of the mM-MCMC method with backmapping is shown in figure 7.2.

## 7.4 Numerical Results

In this section, we apply the micro-macro Markov chain Monte Carlo method with backmapping to a proteomic system containing ten amino acids and 150 atoms. The goal of this section (and the chapter really) is to improve the exploration of the microscopic (atomistic) state space with the mM-MCMC method. Our reference simulation data are the results of a microscopic numerical integrator provided by Wim Vranken and his research group (`https://we.vub.ac.be/nl/wim-vranken`).

The 6kn2 protein consists of ten amino-acids on a single chain. These amino-acids are Glycine (GLY), Phenylalanine (PHE), Arginine (ARG), Serine (SER),
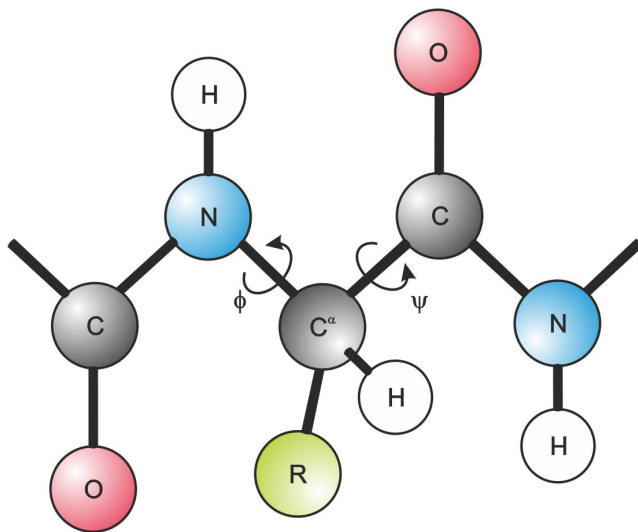
Figure 7.3: Visual illustration of the Ramachandran torsion angles $\phi(C, N, C_\alpha, C)$ and $\psi(N, C_\alpha, C, N)$. The first carbon atom in $\phi$ is the rightmost carbon atom from the previous amino-acid, and the second $N$ in $\psi$ comes from the next amino-acid. The illustration was taken from `https://www.peptideweb.com`.

Proline (PRO) and Cysteine (CYS). The specific order of amino-acids is GLY-PHE-ARG-SER-PRO-CYS-PRO-PRO-PHE-CYS. The terminal group is $NH_2$. The microscopic force field parameters are derived from the more general CHARMM36 force field, typically used for proteins. On the macroscopic level, we use the MARTINI reaction coordinate, where the free energy is derived from the microscopic force field. The exact nature of these calculations is outside the scope of this thesis.

In the following numerical experiment, we depict scatter plots of Ramachandran torsion angles obtained by a sample run of the mM-MCMC method with backmapping. For each amino-acid, the two Ramachandran torsion angles, are angles that describe bending of the main chain, independent from the remainder group. Figure 7.3 shows a visual representation of the Ramachandran torsion angles. In addition to a scatter plot of the sampled torsion angles, we also show the probability density function of these torsion angles obtained by Gaussian kernel density estimation (KDE).

The setup for this numerical experiment is as follows. We run the mM-MCMC method with backmapping for $N = 10^5$ samples. On the macroscopic level, we

use MARTINI proposals generated by Hamiltonian Monte Carlo, followed by an accept/reject step. The macroscopic step size is $\Delta t = 10^{-5}$. For reconstruction, we use the backmapping scheme with $K = 20$ steps and with a time step of size $\delta t = 6.0 \ 10^{-6}$. We compare the sampling results from the mM-MCMC method with a purely microscopic sampler. These microscopic proposals are also obtained using HMC. The numerical sampling results on the Ramachandran angles for amino-acids 2 to 9 are shown on figures 7.4 - 7.11. We do not include the first and last amino-acids because their Ramachandran plots only show 360-degree rotation of the terminal groups.
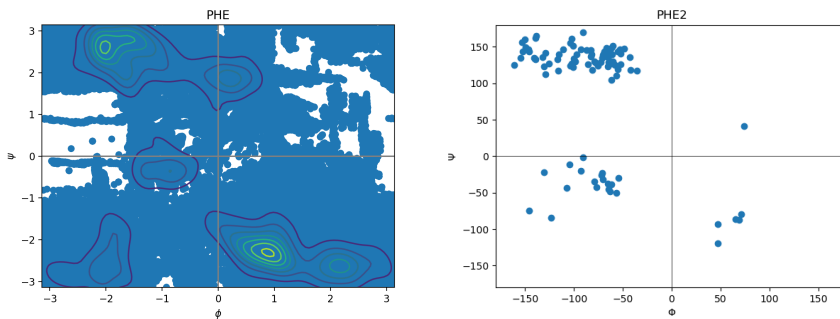


Figure 7.4: Scatter and probability density plots of the mM-MCMC method (left) and a scatter plot of the reference microscopic sampler (right) of phenylalanine.
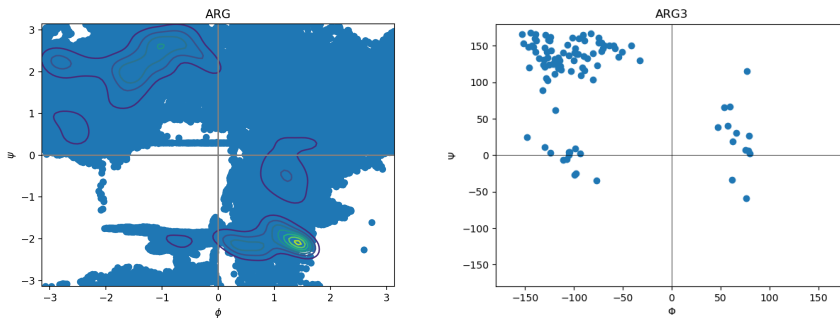


Figure 7.5: Scatter and probability density plots of the mM-MCMC method (left) and a scatter plot of the reference microscopic sampler (right) of arginine.

We can draw a few general conclusions from these Ramachandran figures. First, there is a major difference between the scatter plots obtained by the mM-MCMC method and those obtained by the microscopic integrator. The mM-MCMC
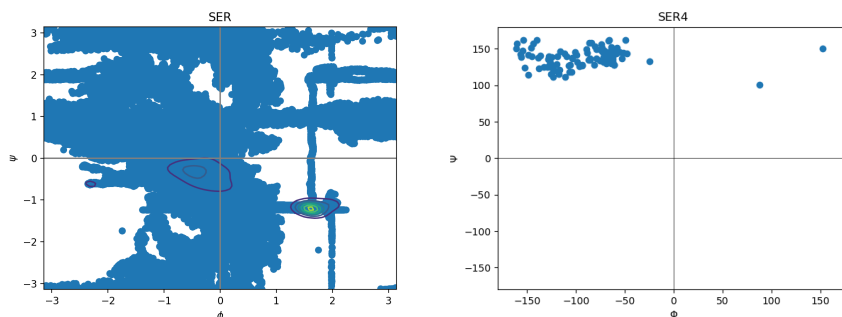
Figure 7.6: Scatter and probability density plots of the mM-MCMC method (left) and a scatter plot of the reference microscopic sampler (right) of serine.
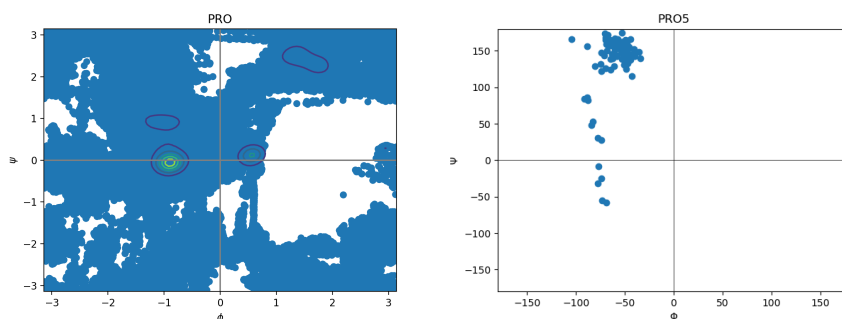


Figure 7.7: Scatter and probability density plots of the mM-MCMC method (left) and a scatter plot of the reference microscopic sampler (right) of proline.

method with backmapping is able to fully explore the state space of torsion angles, except for perhaps a few small regions. This sampling result is opposite to the microscopic sampler which only places samples in a few regions, and has trouble crossing from one region to another. Second, the regions of high probability by the mM-MCMC method do not necessarily match those found by the microscopic sampler.

If we study the Ramachandran diagrams more closely, we find that the mM-MCMC fit on amino-acids 2, 3, 6, 7, 8 and 9 are good, but the fits on amino-acids 4 and 5 lack accuracy compared to the microscopic reference samples. We see one reason why the density plot by the mM-MCMC method and the scatter plot by the microscopic sampler disagree. Clearly, the mM-MCMC method did find other local minima on figures 7.6and 7.7, local minima the microscopic
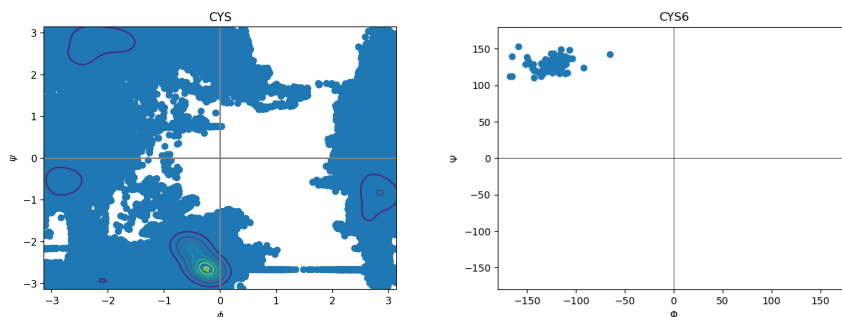
Figure 7.8: Scatter and probability density plots of the mM-MCMC method (left) and a scatter plot of the reference microscopic sampler (right) of cysteine.
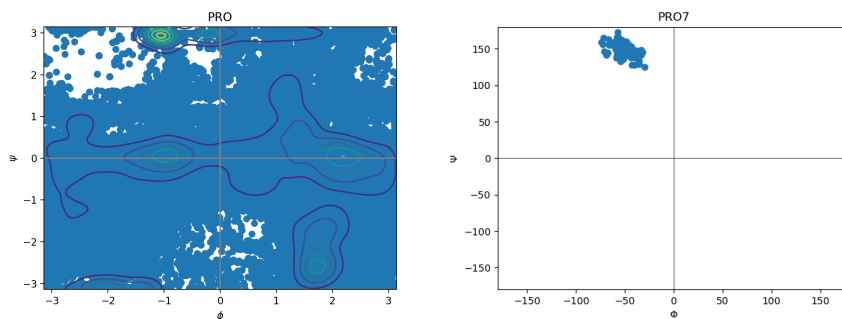


Figure 7.9: Scatter and probability density plots of the mM-MCMC method (left) and a scatter plot of the reference microscopic sampler (right) of proline.

sampler did not find. We therefore conjecture that there are too few samples in the regions of the local minima found by the microscopic sampler so that the mM-MCMC method could not converge too those local minima.

The above arguments summarize the mM-MCMC method with backmapping well. Indeed, this method improves the overall exploration of the microscopic state space (and for that matter the macroscopic state space as well), at the cost of correctness of its (empirical) invariant distribution. This effect is an example of the 'exploration vs. exploitation' trade-off, which we covered in section 3.4. The mM-MCMC method with backmapping is most useful for finding local minima in chemical systems, not for sampling the invariant distribution and for other tasks that require a high level of accuracy.
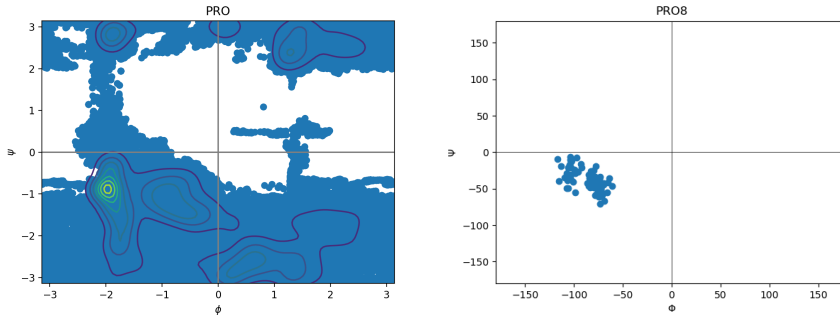
Figure 7.10: Scatter and probability density plots of the mM-MCMC method (left) and a scatter plot of the reference microscopic sampler (right) of proline.
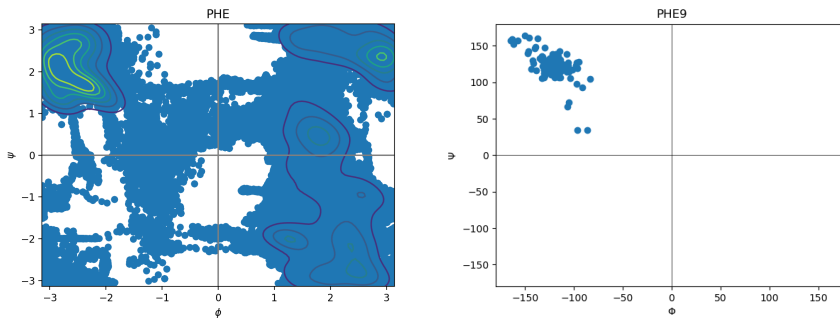


Figure 7.11: Scatter and probability density plots of the mM-MCMC method (left) and a scatter plot of the reference microscopic sampler (right) of phenylalanine.

# Chapter 8

# Conclusion

This chapter forms the conclusion of this thesis and almost five years of research. The goal is to contextualize the findings and results into the broader literature and the research community. First, we summarise this text and its major contributions in section 8.1, before turning to possible improvements and ideas for future research in section 8.2.

## 8.1 General Conclusions

It has been long known that many molecular systems exhibit behaviour on multiple time scales, and that such systems are particularly difficult to simulate over long periods of time. These multiple time scales often come from different components of the molecule. The vibrations of covalent bonds around an equilibrium bond length are usually fast, while the typical time scales of rotations of torsion angles, motion of groups of atoms and change in global conformation of the molecule are slow. The study of sampling the time-invariant distribution of such molecular systems was the topic of this PhD thesis.

Time-steppers for multiscale molecular dynamics have been around since the turn of the century. We particularly refer to the equation-free framework [48, 49] and the heterogeneous multiscale method (HMM) [30]. The equation-free method requires coarse or collective variables that capture that macroscopic behaviour. Fortunately, such variables are typically easy to identify for the systems we studied here. The equation-free method accelerates the simulation of the microscopic system by calculating the time derivative of the coarse

variables using a short microscopic simulation, followed by an extrapolation of these coarse variables using the estimated time derivative. Returning to the microscopic level is achieved by a lifting step. The heterogeneous multiscale method is a rehashing of the equation-free scheme that allows for an easier analysis.

Before this thesis, no Markov chain Monte Carlo methods akin to equation-free existed for sampling time-invariant distributions in molecular dynamics. Other methods were used to sample from these distributions, with low to moderate success. We found that the Hamiltonian Monte Carlo [9] method is able to bridge moderate time-scale separations up to a factor of 100. Other, improved multiscale MCMC methods are all based on the delayed-acceptance (DA) framework [32]. Both multilevel MCMC methods [26, 44] overcome the time-scale separation model by using simpler models on coarser levels, much like simulated annealing [8]. However, a series of models must be available. Then there is the coupled MCMC method [45], the main inspiration for this work. This method also uses the DA framework but the model on the coarse level is actually coarse, i.e., a model on a smaller lattice. The coarse lattice allows taking larger proposal moves because the fast time-scales are no longer present.

In chapter 4, we introduced the mM-MCMC method for sampling invariant distributions with a time-scale separation. The method uses five steps to accelerate sampling; restriction to the macroscopic level; generating a proposal on the macroscopic level; macroscopic accept/reject step; reconstruction back to the microscopic level; microscopic accept/reject step. We made the simplest assumption that helps to actually accelerate the sampling of these distributions: the existence of an approximate macroscopic distribution. Additionally, we proposed two types of reconstruction, direct and indirect. We mostly used indirect reconstruction in this thesis because it generalises well to complex molecules. We also proved convergence of both mM-MCMC variants, and even showed a stronger ergodicity result for mM-MCMC with direct reconstruction. Afterwards, in chapter 5, we illustrated that the mM-MCMC method can indeed efficiently sample from distributions with a time-scale separation, and we also numerically demonstrated its sensitivity with respect to parameter changes.

Because of our use of an approximate macroscopic distribution, the free energy of the reaction coordinate pops up in the microscopic acceptance rate. As we mentioned many times, this free energy is hard to compute to high precision, so in chapter 6, we searched for an alternative that is cheap to calculate and is statistically exact. There are many free energy computation methods in the literature, we listed the most promising ones for our application in the manuscript. These methods are the adaptive biasing potential/force methods, thermodynamic integration, and a non-equilibrium integration scheme based on the Jarzynski equality. None of these methods were found suitable because

they are either too slow for our application or they have a bias.

In the end, we settled on the pseudo-marginal method. This method computes a fast and unbiased estimate to the free energy by importance sampling. The importance samples are drawn from the indirect reconstruction distribution. We have adapted the pseudo-marginal method to work well in the mM-MCMC framework, so that the complete mM-MCMC method with pseudo-marginal approximation is still unbiased, except for an added variance.

In our experiments at the end of chapter 6, we showed numerically that the pseudo-marginal scheme is indeed unbiased but leads to a large variance on the free energy estimates. We also studied the effect of the pseudo-marginal scheme on the mM-MCMC method. If we plot the averaged microscopic acceptance rate as a function of the dimension of the microscopic system, we see that the acceptance rate first decreases exponentially before settling to a steady state value. This steady state value increases with the number of indirect reconstruction steps. In the same limit, we also see that the variance on free energy estimates decreases, and that the total efficiency gain of the mM-MCMC method over the microscopic sampler increases, demonstrating the practical use of pseudo-marginal reconstruction.

The pseudo-marginal method does, however, not scale well with high dimensional reaction coordinates. This is a big problem for the use of the mM-MCMC method to macromolecules like proteins where the reaction coordinate typically grows with the microscopic system. We used the MARTINI reaction coordinate for proteins in chapter 7. In that chapter, we studied backmapping as an alternative to indirect reconstruction and the pseudo-marginal scheme. Backmapping uses an out-of-equilibrium, time-dependent simulation for reconstruction. Initially, only macroscopic forces are active. As time increases on, the macroscopic forces phase out, and the microscopic forces kick in until they are the only forces active. The complete mM-MCMC method with backmapping does not necessarily converge to the invariant Gibbs distribution, but that doesn't make it useless. There is, for instance, no need for the pseudo-marginal approximation to the free energy because one cannot define an equilibrium microscopic acceptance rate. The mM-MCMC method actually makes a trade-off between exploration and precision, favouring the former.

At the end of chapter 7, we applied our mM-MCMC method with backmapping to a single disulfide peptide protein, or 6kn2 for short. We compared the sampling results to a microscopic reference solution. Our results were twofold. First, we saw that the mM-MCMC method was able to improve exploration of the microscopic state space compared to the reference. That is, the microscopic sampler only explored a few regions of the Ramachandran plot, while the mM-MCMC method almost explored the whole Ramachandran plot, except for a

few regions. Second, regions of high probability by the mM-MCMC method largely match the local minima found by the microscopic sampler, except for a few individual differences. We conclude that the mM-MCMC method with backmapping is suited for exploring the high-dimensional state space of proteins, and for discovering a collection of local minima of the potential.

## 8.2   Ideas for Future Research

A PhD is never a linear story. We present the research in this thesis as if every problem is immediately met with a solution, and as if that solution works well instantaneously without adaptations or errors. Obviously, the way we present the research does not capture of essence of the scientific process at all. At every step, we investigate multiple possible solutions to the problem, and the select the best alternative. In this section, we list tested and untested alternatives, as well as new ideas for future research per chapter.

**Chapter 4**   This chapter introduces the micro-macro Markov chain Monte Carlo method to overcome the time-scale separation problem in many molecular systems. Our solution is to introduce a macroscopic model with a continuous reaction coordinate, coupled to the microscopic level through reconstruction.

Other choices for the reaction coordinate are also possible. Drawing from the Kinetic Monte Carlo method (KMC) [95], one can define the reaction coordinate to be the local minima of the microscopic potential surface. The macroscopic model is then defined as the switching between these local minima, and reconstruction happens by drawing a sample from the quasi-stationary distribution or QSD [63]. Now, this method has the drawback of having to know the local minima in advance, although one can easily imagine an alternative where switching between local minima is alternated with the discovery of local minima.

Another possible extension to the mM-MCMC method is to use more than two models and levels. There are two options. One option is to use macroscopic systems with lower and lower dimensions as the level increases. The other option is to use microscopic models that become simpler and simpler, but maintain their dimensionality. The latter is the idea behind the MLMCMC method(s) [26, 44]. The former idea has not been tried yet, but the existence of a chain of macroscopic models depends largely on the application.

**Chapter 6**   This chapter presents the pseudo-marginal method as a solution to the free energy problem. The method is based on an importance sampling strategy that alleviates the need to either know the free energy analytically or to compute it in advance to high precision. One can simply replace the free energy by its pseudo-marginal approximation, and the mM-MCMC method remains unbiased.

If we go back to the root of the problem, we see that the free energy of the reaction coordinate is required to evaluate the microscopic acceptance probability. Without a macroscopic model, that is an approximate macroscopic distribution, the comparison with the free energy disappears as well. We therefore propose using the equation-free approach on the macroscopic level. The idea is to run a short burst of microscopic simulation to estimate the time derivative of the reaction coordinate. Proposing a new reaction coordinate value then happens by extrapolating the previous reaction coordinate value with the time derivative.

We did not use the pseudo-marginal method on systems with a high-dimensional reaction coordinate. We have implemented the mM-MCMC method with pseudo-marginal approximation on the 6kn2 protein in the past, but the results were underwhelming. There reason is that you need many more reconstructed samples to approximate a high-dimensional free energy. That number of samples was too many for our calculations. We therefore propose to find an adaptation to the pseudo-marginal method so that is also works for high-dimensional reaction coordinates.

Finally, once a new FEC method is proposed in the literature, one can try it to the free energy problem in the context of the mM-MCMC method.

**Chapter 7**   This chapter deals with the backmapping scheme as an alternative to indirect reconstruction and the pseudo-marginal approximation. Backmapping uses an out-of-equilibrium simulation to generate the reconstructed sample.

So far, we have only applied the mM-MCMC method with backmapping to one protein, 6kn2. As an extension, we propose to implement the mM-MCMC method on other proteins, and investigate if our results in this thesis generalise. In line with other proteins is implementing the mM-MCMC on other macromolecules, including polymers and DNA. Both these molecules are of a long chain of repeating units, hence a MARTINI-like reaction coordinate and force field must exist as well.

The final idea is about error estimates. We currently have no assessment of how close samples obtained by the mM-MCMC method are to the microscopic invariant distribution. We therefore propose to derive statistical error estimators that measure this error. A particular class of estimators, goal-oriented

estimators [78, 1] may prove useful. Goal-oriented estimators provide a-posteriori error estimates based on a quantity of interest, i.e., the microscopic invariant distribution, rather than the typical a-priori estimators based on parameters of the method and the problem.

As a final remark, we have only applied the mM-MCMC method on molecular systems without chemical reactions. But, chemical reactions can be described by a potential energy surface like any other system that we studied here. We propose to implement the mM-MCMC method on reacting systems as well to sample their invariant distribution, and learn from the reaction simultaneously.

Along those lines, we also propose to use the mM-MCMC method to sample from quantum molecular dynamics systems. The potential energy surface is derived from ab initio quantum mechanics by finding the electron ground state given the fixed positions of the nuclei. We foresee few difficulties applying the mM-MCMC method here.

# Appendix A

# Relation between $\mathcal{K}_{\mathbf{mM}}$ and $\mathcal{D}$ with exact reconstruction

Here, we prove statement (4.18) in the proof of Theorem 3.

We prove by induction that for any $n \in \mathbb{N}$ we have

$$\mathcal{K}^n_{mM}(x'|\, x) = \nu(x'|\xi(x'))\, \mathcal{D}^n(\xi(x')|\xi(x)) + \mathcal{C}(\xi(x))^n \left( \delta_{x'-x} - \nu(x'|\xi(x'))\delta_{\xi(x')-\xi(x)} \right),$$
$$\text{(A.1)}$$

with

$$\mathcal{C}(\xi(x)) = 1 - \int_H \alpha_{CG}(y|\xi(x))\, q_0(y|\xi(x))dy.$$

For $n = 1$, we can simply rewrite expression (4.6) using that $\alpha_F = 1$ by the exact reconstruction property.

Assume that $n > 1$ and that statement (A.1) holds for $n - 1$. Writing out the $n-$th composition and using the induction hypothesis yields

$$\mathcal{K}^n_{mM}(x'|x) = \int_{\mathbb{R}^d} \mathcal{K}_{mM}(x'|y)\, \mathcal{K}^{n-1}_{mM}(y|x)\, dy$$

$$= \int_{\mathbb{R}^d} \Big( \nu(x'|\xi(x'))\, \mathcal{D}(\xi(x')|\xi(y)) + \mathcal{C}(\xi(y))\, (\delta_{x'-y} - \nu(x'|\xi(x'))\, \delta_{\xi(x')-\xi(y)}) \Big)$$

$$\Big( \nu(y|\xi(y))\mathcal{D}^{n-1}(\xi(y)|\xi(x)) + \mathcal{C}(\xi(x))^{n-1}(\delta_{y-x} - \nu(y|\xi(y))\, \delta_{\xi(y)-\xi(x)}) \Big)dy$$

Splitting this integral formulation into different terms, we obtain

$$\mathcal{K}_{mM}^n(x'|x) = \nu(x'|\xi(x')) \; \mathcal{D}^n(\xi(x')|\xi(x))$$

$$+ \int_{\mathbb{R}^d} \nu(x'|\xi(x')) \; \mathcal{D}(\xi(x')|\xi(y)) \; \mathcal{C}(\xi(x))^{n-1} \; \left(\delta_{y-x} - \nu(y|\xi(y)) \; \delta_{\xi(y)-\xi(x)}\right) \; dy$$

$$+ \int_{\mathbb{R}^d} \nu(y|\xi(y)) \; \mathcal{D}^{n-1}(\xi(y)|\xi(x)) \; \mathcal{C}(\xi(y)) \; \left(\delta_{x'-y} - \nu(x'|\xi(x')) \; \delta_{\xi(x')-\xi(y)}\right) dy$$

$$+ \mathcal{C}(\xi(x))^{n-1} \int_{\mathbb{R}^d} \mathcal{C}(\xi(y)) \; \left(\delta_{x'-y} - \nu(x'|\xi(x')) \; \delta_{\xi(x')-\xi(y)}\right)$$

$$\left(\delta_{y-x} - \nu(y|\xi(y)) \; \delta_{\xi(y)-\xi(x)}\right) \; dy.$$

The first term of this expression is already in the desired form by the co-area formula. Writing out the second term using the definition of the $\delta$-function, we obtain

$$\int_{\mathbb{R}^d} \nu(x'|\xi(x')) \; \mathcal{D}(\xi(x')|\xi(y)) \; \mathcal{C}(\xi(x))^{n-1} \; \left(\delta_{y-x} - \nu(y|\xi(y)) \; \delta_{\xi(y)-\xi(x)}\right) \; dy$$

$$= \nu(x'|\xi(x')) \; \mathcal{C}(\xi(x))^{n-1}$$

$$\left(\int_{\mathbb{R}^d} \mathcal{D}(\xi(x')|\xi(y)) \; \delta_{y-x} - \int_{\mathbb{R}^d} \mathcal{D}(\xi(x')|\xi(y)) \; \nu(y|\xi(y)) \; \delta_{\xi(y)-\xi(x)} \; dy\right)$$

$$= \nu(x'|\xi(x')) \; \mathcal{C}(\xi(x))^{n-1}$$

$$\left(\mathcal{D}(\xi(x')|\xi(x)) - \int_H \mathcal{D}(\xi(x')|z) \; \delta_{z-\xi(x)} \int_{\Sigma(z)} \nu(y|z) \; \|\nabla\xi(y)\|^{-1} \; d\sigma_z(y)dz\right)$$

$$= 0.$$

This expression cancels because the probability density $\nu(y|z) \; \|\nabla\xi(y)\|^{-1}$ integrates to 1 on $\Sigma(z)$ by the co-area formula. Similarly, the third term

cancels as well and expanding the fourth term yields

$$\mathcal{C}(\xi(x))^{n-1} \int_{\mathbb{R}^d} \mathcal{C}(\xi(y))\ \delta_{x'-y}\ \delta_{y-x}\ dy$$

$$-\mathcal{C}(\xi(x))^{n-1} \int_{\mathbb{R}^d} \mathcal{C}(\xi(y))\ \delta_{x'-y}\ \nu(y|\xi(y))\ \delta_{\xi(y)-\xi(x)}\ dy$$

$$-\mathcal{C}(\xi(x))^{n-1} \int_{\mathbb{R}^d} \mathcal{C}(\xi(y))\ \nu(x'|\xi(x'))\ \delta_{\xi(x')-\xi(y)}\ \delta_{y-x}\ dy$$

$$+\mathcal{C}(\xi(x))^{n-1} \int_{\mathbb{R}^d} \mathcal{C}(\xi(y))\ \nu(x'|\xi(x'))\ \delta_{\xi(x')-\xi(y)}\ \nu(y|\xi(y))\ \delta_{\xi(y)-\xi(x)}\ dy$$

$$=\mathcal{C}(\xi(x))^n\ \delta_{x'-x} - \mathcal{C}(\xi(x))^n\ \nu(x'|\xi(x'))\ \delta_{\xi(x')-\xi(x)}$$

$$-\mathcal{C}(\xi(x))^n\ \nu(x'|\xi(x'))\ \delta_{\xi(x')-\xi(x)} + \mathcal{C}(\xi(x))^n\ \nu(x'|\xi(x'))\ \delta_{\xi(x')-\xi(x)}$$

$$=\mathcal{C}(\xi(x))^n\ \left(\delta_{x'-x} - \nu(x'|\xi(x'))\ \delta_{\xi(x')-\xi(x)}\right).$$

$$(A.2)$$

Putting these expressions together, we obtain (A.1) with iteration number $n$.

# Appendix B

# The Free Energy of the Torsion Angle of Butane

It is shown in [100, Section 1] that if $(z, \{q\}^{N-1})$ are rigid body coordinates of a molecule (a set of curvilinear coordinates), the free energy of $z$ is

$$A(z) = -\beta^{-1} \ln \int \left| J(z, \{q\}^{N-1}) \right| \exp(-\beta V(z, \{q\}^{N-1})) d\{q\}^{N-1} + C. \quad \text{(B.1)}$$

The expression $\left| J(z, \{q\}^{N-1}) \right|$ is the determinant of the Jacobian that describes the coordinate transformation from rigid body to Cartesian coordinates. We implicitly assume in Figure 5.10 that the left-most atom is fixed at the origin, the second atom lies on the positive $x$-axis with coordinates $(x_1, 0, 0)$ and the third atom lies in the $xy$ plane with coordinates $(x_2, y_2, 0)$. The coordinate of the fourth atom are $(x_3, y_3, z_3)$. The rigid body coordinates are then the three bond lengths $r_1, r_2$ and $r_3$, the two angles $\theta_1, \theta_2$ and the torsion angle $z$. That is, $\{q\}^{N-1} = (r_1, r_2, r_3, \theta_1, \theta_2)$. The Cartesian coordinates written as a function of the rigid body coordinates gives

$$
\begin{pmatrix} x_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \\ z_3 \end{pmatrix}
=
\begin{pmatrix} r_1 \\ r_1 \\ 0 \\ r_1 \\ 0 \\ 0 \end{pmatrix}
+
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sin(\theta_1) & -\cos(\theta_1) & 0 & 0 & 0 \\
0 & \cos(\theta_1) & \sin(\theta_1) & 0 & 0 & 0 \\
0 & 0 & 0 & \sin(\theta_1) & -\cos(\theta_1) & 0 \\
0 & 0 & 0 & \cos(\theta_1) & \sin(\theta_1) & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix} 0 \\ 0 \\ r_2 \\ r_3 \sin(\theta_2)\cos(z) \\ r_3 \cos(\theta_2) \\ r_3 \sin(\theta_2)\sin(z) \end{pmatrix}.
$$

The determinant of the Jacobian of the above coordinate transformation reads

$$\left| J(z, \{q\}^{N-1}) \right| = r_2 r_3^2 |\sin(\theta_2)|,$$

and is independent of $z$. Therefore, in case of butane, the only factor in (B.1) that depends on $z$ is the potential energy term corresponding to the torsion angle: $c_0 + c_1 \cos(z) + c_2 \cos(z)^2 + c_3 \cos(z)^3$. The remaining integral in (B.1) only serves as a normalization constant for the marginal distribution of reaction coordinate $\xi$, proving that (6.25) is indeed the free energy of the torsion angle.
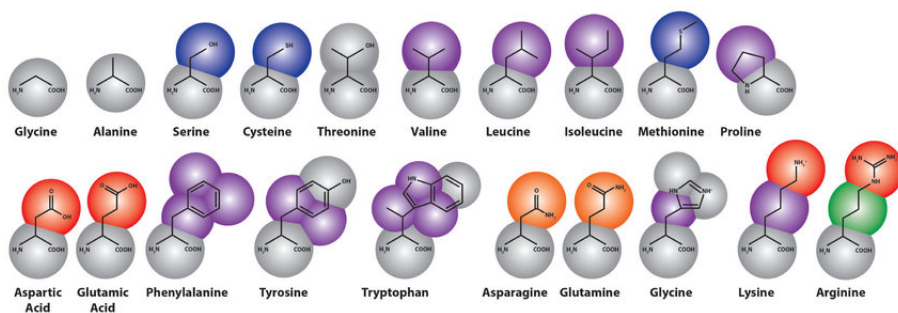
# Appendix C

# MARTINI Mapping of all Amino-Acids



Figure C.1: MARTINI mapping of all twenty amino-acids in the human body.

# Bibliography

[1] ABDELMALIK, M. R., AND VAN BRUMMELEN, H. Error estimation and adaptive moment hierarchies for goal-oriented approximations of the Boltzmann equation. *Computer Methods in Applied Mechanics and Engineering 325* (2017), 219–239.

[2] ALENLÖV, J., DOUCET, A., AND LINDSTEN, F. Pseudo-marginal Hamiltonian Monte Carlo. *arXiv preprint arXiv:1607.02516* (2016).

[3] ANDERSEN, H. C. RATTLE: A velocity version of the SHAKE algorithm for molecular dynamics calculations. *Journal of Computational Physics 52*, 1 (1983), 24–34.

[4] ANDRIEU, C., AND ROBERTS, G. O. The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics 37*, 2 (2009), 697–725.

[5] ANDRIEU, C., AND VIHOLA, M. Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Annals of Applied Probability 25*, 2 (2015), 1030–1077.

[6] BEAUMONT, M. A. Estimation of population growth or decline in genetically monitored populations. *Genetics 164*, 3 (2003), 1139–1160.

[7] BELLO-RIVAS, J. M., GEORGIOU, A., VANDECASTEELE, H., AND KEVREKIDIS, I. G. Gentlest ascent dynamics on manifolds defined by adaptively sampled point-clouds. *Accepted for Publication in The Journal of Physical Chemistry* (2023).

[8] BERTSIMAS, D., AND TSITSIKLIS, J. Simulated Annealing. *Statistical Science 8*, 1 (1993), 10–15.

[9] BETANCOURT, M. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434* (2017).

[10] BETANCOURT, M. The convergence of Markov chain Monte Carlo methods: from the Metropolis method to Hamiltonian Monte Carlo. *Annalen der Physik 531*, 3 (2019), 1700214.

[11] BETANCOURT, M., BYRNE, S., LIVINGSTONE, S., AND GIROLAMI, M. The geometric foundations of Hamiltonian Monte Carlo.

[12] BOERO, M., AND OSHIYAMA, A. Car–Parrinello molecular dynamics, 2015.

[13] BOU-RABEE, N., AND HAIRER, M. Nonasymptotic mixing of the MALA algorithm. *IMA Journal of Numerical Analysis 33*, 1 (2013), 80–110.

[14] BRÉMAUD, P. *Probability Theory and Stochastic Processes*. Springer, 2020.

[15] CAR, R., AND PARRINELLO, M. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters 55*, 22 (1985), 2471.

[16] CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P., AND RIDDELL, A. Stan: A probabilistic programming language. *Journal of Statistical Software 76*, 1 (2017).

[17] CÉROU, F., AND GUYADER, A. Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications 25*, 2 (2007), 417–443.

[18] CHEN, T., FOX, E., AND GUESTRIN, C. Stochastic gradient Hamiltonian Monte Carlo. In *International Conference on Machine Learning* (2014), PMLR, pp. 1683–1691.

[19] CHIPOT, C., AND POHORILLE, A. *Free energy calculations*, vol. 86. Springer, 2007.

[20] CHRISTEN, A. J., AND FOX, C. Markov chain monte carlo using an approximation. *Journal of Computational and Graphical Statistics 14*, 4 (2005), 795–810.

[21] COMER, J., GUMBART, J. C., HÉNIN, J., LELIÈVRE, T., POHORILLE, A., AND CHIPOT, C. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *The Journal of Physical Chemistry B 119*, 3 (2015), 1129–1151.

[22] COTTER, S. L., ROBERTS, G. O., STUART, A. M., AND WHITE, D. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science 28*, 3 (2013), 424–446.

[23] Cox, D. R. *The theory of Stochastic Processes.* Routledge, 2017.

[24] Darve, E., Rodríguez-Gómez, D., and Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations.

[25] Dickson, B. M., Legoll, F., Lelièvre, T., Stoltz, G., and Fleurat-Lessard, P. Free energy calculations: An efficient adaptive biasing potential method. *The Journal of Physical Chemistry B 114*, 17 (2010), 5823–5830.

[26] Dodwell, T. J., Ketelsen, C., Scheichl, R., and Teckentrup, A. L. Multilevel Markov chain Monte Carlo. *SIAM Review 61*, 3 (2019), 509–545.

[27] Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. Hybrid Monte Carlo. *Physics Letters B 195*, 2 (1987), 216–222.

[28] Durmus, A., and Moulines, E. On the geometric convergence for MALA under verifiable conditions. *arXiv preprint arXiv:2201.01951* (2022).

[29] Durmus, A., Roberts, G. O., Vilmart, G., and Zygalakis, K. C. Fast Langevin based algorithms for MCMC in high dimensions. *The Annals of Applied Probability 27*, 4 (2017), 2195–2237.

[30] E, W., Engquist, B., and Huang, Z. Heterogeneous multiscale method: a general methodology for multiscale modeling. *Physical Review B 67*, 9 (2003), 092101.

[31] Erban, R. From molecular dynamics to Brownian dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 470*, 2167 (2014), 20140036.

[32] Everitt, R. G., and Rowińska, P. A. Delayed acceptance abc-smc. *Journal of Computational and Graphical Statistics 30*, 1 (2021), 55–66.

[33] Filippone, M., and Girolami, M. Pseudo-marginal Bayesian inference for Gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 36*, 11 (2014), 2214–2226.

[34] Gelman, A., Lee, D., and Guo, J. Stan: A probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics 40*, 5 (2015), 530–543.

[35] Gyöngy, I. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields 71*, 4 (1986), 501–516.

[36] Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 1 (1970), 97–109.

[37] Head-Gordon, T., Stillinger, F. H., and Arrecis, J. A strategy for finding classes of minima on a hypersurface: implications for approaches to the protein folding problem. *Proceedings of the National Academy of Sciences 88*, 24 (1991), 11076–11080.

[38] Hirschman, I. I., and Widder, D. V. *The convolution transform.* Courier Corporation, 2012.

[39] Hoffman, M. D., and Gelman, A. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res. 15*, 1 (2014), 1593–1623.

[40] Homeyer, N., and Gohlke, H. Free energy calculations by the molecular mechanics Poisson- Boltzmann surface area method. *Molecular Informatics 31*, 2 (2012), 114–122.

[41] Hummer, G. Fast-growth thermodynamic integration: Error and efficiency analysis. *The Journal of Chemical Physics 114*, 17 (2001), 7330–7337.

[42] Hutter, J. Car–parrinello molecular dynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science 2*, 4 (2012), 604–612.

[43] Jarzynski, C. Nonequilibrium equality for free energy differences. *Physical Review Letters 78*, 14 (1997), 2690.

[44] Jasra, A., Law, K., and Xu, Y. Markov chain simulation for multilevel Monte Carlo. *arXiv preprint arXiv:1806.09754* (2018).

[45] Kalligiannaki, E., Katsoulakis, M. A., and Plecháč, P. Coupled coarse graining and Markov chain Monte Carlo for lattice systems. In *Numerical Analysis of Multiscale Computations: Proceedings of a Winter Workshop at the Banff International Research Station 2009* (2011), B. Engquist, O. Runborg, and Y.-H. Tsai, Eds., vol. 82, Springer, Berlin, Heidelberg, pp. 235–257.

[46] Kalligiannaki, E., Katsoulakis, M. A., and Plechác, P. Spatial two-level interacting particle simulations and information theory–based error quantification. *SIAM Journal on Scientific Computing 36*, 2 (2014), A634–A667.

[47] Kalligiannaki, E., Katsoulakis, M. A., Plecháč, P., and Vlachos, D. G. Multilevel coarse graining and nano-pattern discovery

in many particle stochastic systems. *Journal of Computational Physics 231*, 6 (2012), 2599–2620.

[48] KEVREKIDIS, I. G., GEAR, W. C., HYMAN, J. M., KEVREKIDID, P. G., RUNBORG, O., AND THEODOROPOULOS, C. Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis. *Communications in Mathematical Sciences 1*, 4 (2003), 715–762.

[49] KEVREKIDIS, I. G., AND SAMAEY, G. Equation-free multiscale computation: Algorithms and Applications. *Annual Review of Physical Chemistry 60* (2009), 321–344.

[50] KOELLERMEIER, J., AND VANDECASTEELE, H. Hierarchical micro-macro acceleration for moment models of kinetic equations. *Journal of Computational Physics* (2023), 112194.

[51] KOHN, W., AND SHAM, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review 140*, 4A (1965), A1133.

[52] KRAJNIAK, J., PANDIYAN, S., NIES, E., AND SAMAEY, G. Generic adaptive resolution method for reverse mapping of polymers from coarse-grained to atomistic descriptions. *Journal of chemical theory and computation 12*, 11 (2016), 5549–5562.

[53] KRAJNIAK, J., ZHANG, Z., PANDIYAN, S., NIES, E., AND SAMAEY, G. Coarse-grained molecular dynamics simulations of polymerization with forward and backward reactions. *Journal of Computational Chemistry 39*, 22 (2018), 1764–1778.

[54] KRAJNIAK, J., ZHANG, Z., PANDIYAN, S., NIES, E., AND SAMAEY, G. Reverse mapping method for complex polymer systems. *Journal of Computational Chemistry 39*, 11 (2018), 648–664.

[55] KREKELER, C., AGARWAL, A., JUNGHANS, C., PRAPROTNIK, M., AND DELLE SITE, L. Adaptive resolution molecular dynamics technique: Down to the essential. *The Journal of chemical physics 149*, 2 (2018), 024104.

[56] LANGEVIN, P. Sur la théorie du mouvement brownien. *Compt. Rendus 146* (1908), 530–533.

[57] LE BRIS, C., LELIÈVRE, T., LUSKIN, M., AND PEREZ, D. A mathematical formalization of the parallel replica dynamics. *Monte Carlo Methods and Applications 18*, 2 (2012), 119–146.

[58] LEGOLL, F., AND LELIÈVRE, T. Effective dynamics using conditional expectations. *Nonlinearity 23*, 9 (2010), 2131–2163.

[59] LEGOLL, F., AND LELIÈVRE, T. Some remarks on free energy and coarse-graining. In *Numerical Analysis of Multiscale Computations*. Springer, 2012, pp. 279–329.

[60] LEIMKUHLER, B., CHIPOT, C., ELBER, R., LAAKSONEN, A., MARK, A., SCHLICK, T., SCHÜTTE, C., AND SKEEL, R. *New algorithms for macromolecular simulation*, vol. 49. Springer Science & Business Media, 2006.

[61] LEIMKUHLER, B., AND MATTHEWS, C. *Molecular Dynamics*. Springer, 2016.

[62] LEIMKUHLER, B., NOORIZADEH, E., AND THEIL, F. A gentle stochastic thermostat for molecular dynamics. *Journal of Statistical Physics 135*, 2 (2009), 261–277.

[63] LELIÈVRE, T., RAMIL, M., AND REYGNER, J. Quasi-stationary distribution for the Langevin process in cylindrical domains, part I: existence, uniqueness and long-time convergence. *Stochastic Processes and their Applications 144* (2022), 173–201.

[64] LELIÈVRE, T., ROUSSET, M., AND STOLTZ, G. Computation of free energy differences through nonequilibrium stochastic dynamics: The reaction coordinate case. *Journal of Computational Physics 222*, 2 (2007), 624–643.

[65] LELIÈVRE, T., ROUSSET, M., AND STOLTZ, G. Long-time convergence of an adaptive biasing force method. *Nonlinearity 21*, 6 (2008), 1155.

[66] LELIEVRE, T., ROUSSET, M., AND STOLTZ, G. Langevin dynamics with constraints and computation of free energy differences. *Mathematics of computation 81*, 280 (2012), 2071–2125.

[67] LELIÈVRE, T., AND STOLTZ, G. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica 25* (2016), 681–880.

[68] LENNARD-JONES, J. E. The electronic structure of some diatomic molecules. *Transactions of the Faraday Society 25* (1929), 668–686.

[69] LYKKEGAARD, M. B., DODWELL, T. J., FOX, C., MINGAS, G., AND SCHEICHL, R. Multilevel delayed acceptance MCMC. *SIAM/ASA Journal on Uncertainty Quantification 11*, 1 (2023), 1–30.

[70] MARRINK, S. J., DE VRIES, A. H., AND MARK, A. E. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B 108*, 2 (2004), 750–760.

[71] Marrink, S. J., Risselada, J. H., Yefimov, S., Tieleman, P. D., and De Vries, A. H. The MARTINI force field: Coarse-grained model for biomolecular simulations. *The Journal of Physical Chemistry B 111*, 27 (2007), 7812–7824.

[72] McLachlan, R. I., and Quispel, R. G. Geometric integrators for ODEs. *Journal of Physics A: Mathematical and General 39*, 19 (2006), 5251.

[73] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics 21*, 6 (1953), 1087–1092.

[74] Metropolis, N., and Ulam, S. The Monte Carlo method. *Journal of the American Statistical Association 44*, 247 (1949), 335–341.

[75] Meyn, S. P., and Tweedie, R. L. *Markov chains and stochastic stability.* Springer Science & Business Media, 1993.

[76] Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, P. D., and Marrink, S.-J. The MARTINI coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation 4*, 5 (2008), 819–834.

[77] Moore, T. C., Iacovella, C. R., and McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of chemical physics 140*, 22 (2014), 06B606_1.

[78] Oden, J. T., and Prudhomme, S. Goal-oriented error estimation and adaptivity for the finite element method. *Computers & mathematics with applications 41*, 5-6 (2001), 735–756.

[79] Pavliotis, G., and Stuart, A. *Multiscale methods: averaging and homogenization.* Springer Science & Business Media, 2008.

[80] Ponder, J. W., and Case, D. A. Force fields for protein simulations. *Advances in Protein Chemistry 66* (2003), 27–85.

[81] Praprotnik, M., Delle Site, L., and Kremer, K. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *The Journal of Chemical Physics 123*, 22 (2005), 224106.

[82] Roberts, G. O., and Rosenthal, J. S. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science 16*, 4 (2001), 351–367.

[83] ROBERTS, G. O., AND ROSENTHAL, J. S. General state space Markov chains and MCMC algorithms. *Probability Surveys 1* (2004), 20–71.

[84] ROBERTS, G. O., AND TWEEDIE, R. L. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli 2*, 4 (1996), 341–363.

[85] ROUSSET, M., AND STOLTZ, G. Equilibrium sampling from nonequilibrium dynamics. *Journal of Statistical Physics 123* (2006), 1251–1272.

[86] SAITO, Y., AND MITSUI, T. Mean-square stability of numerical schemes for stochastic differential systems. *Vietnam J. Math 30* (2002), 551–560.

[87] SAMAEY, G., LELIEVRE, T., AND LEGAT, V. A numerical closure approach for kinetic models of polymeric fluids: exploring closure relations for fene dumbbells. *Computers & fluids 43*, 1 (2011), 119–133.

[88] SCHAPPALS, M., MECKLENFELD, A., KROGER, L., BOTAN, V., KOSTER, A., STEPHAN, S., GARCÍA, E. J., RUTKAI, G., RAABE, G., AND KLEIN, P. Round robin study: molecular simulation of thermodynamic properties from models with internal degrees of freedom. *Journal of Chemical Theory and Computation 13*, 9 (2017), 4270–4280.

[89] SHERLOCK, C., THIERY, A. H., ROBERTS, G. O., AND ROSENTHAL, J. S. On the efficiency of pseudo-marginal random walk Metropolis algorithms. *The Annals of Statistics 43*, 1 (2015), 238–275.

[90] STOLTZ, G., ROUSSET, M., AND LELIÈVRE, T. *Free energy computations: A mathematical perspective.* World Scientific, 2010.

[91] VANDECASTEELE, H., AND SAMAEY, G. Computational efficiency study of a micro-macro Markov chain Monte Carlo method for molecular dynamics. *arXiv preprint arXiv:2209.13056* (2022).

[92] VANDECASTEELE, H., AND SAMAEY, G. A micro-macro Markov chain Monte Carlo method for molecular dynamics using reaction coordinate proposals. *Accepted for Publication at SIAM SISC* (2023).

[93] VANDECASTEELE, H., AND SAMAEY, G. Pseudo-marginal approximation to the free energy in a micro-macro Markov chain Monte Carlo method. *arXiv preprint arXiv:2303.14422* (2023).

[94] VOTER, A. F. Parallel replica method for dynamics of infrequent events. *Physical Review B 57*, 22 (1998), 985–988.

[95] VOTER, A. F. Introduction to the kinetic Monte Carlo method. In *NATO Science Series* (2007), K. E. Sickafus, E. A. Kotomin, and B. P. Uberuaga, Eds., vol. 235, Springer, Dordrecht, pp. 1–23.

[96] VOTER, A. F., MONTALENTI, F., AND GERMANN, T. C. Extending the time scale in atomistic simulation of materials. *Annual Review of Materials Research 32*, 1 (2002), 321–346.

[97] WANG, H., HARTMANN, C., SCHÜTTE, C., AND DELLE SITE, L. Grand-canonical-like molecular-dynamics simulations by using an adaptive-resolution technique. *Physics Review X 3* (2013), 011018.

[98] WILKIE, J. Numerical methods for stochastic differential equations. *Physical Review E 70*, 1 (2004), 017701.

[99] WIQVIST, S., PICCHINI, U., AND FORMAN, J. L. Accelerating delayed-acceptance Markov chain Monte Carlo algorithms. *arXiv preprint arXiv:1806.05982* (2018).

[100] WONG, K.-Y., AND YORK, D. Exact relation between potential of mean force and free-energy profile. *Journal of Chemical Theory and Computation 8*, 11 (2012), 3998–4003.

[101] ZIELIŃSKI, P., VANDECASTEELE, H., AND SAMAEY, G. Convergence and stability of a micro–macro acceleration method: Linear slow–fast stochastic differential equations with additive noise. *Journal of Computational and Applied Mathematics 387* (2021), 112490.

[102] ZUCKERMAN, D. M., AND WOOLF, T. B. Transition events in butane simulations: Similarities across models. *The Journal of Chemical Physics 116*, 6 (2002), 2586–2591.

# Curriculum Vitae

## Personal Information

| | |
|---|---|
| Name | Hannes Vandecaseele |
| Email | Hannes.Vandecasteele@kuleuven.be |
| Date of Birth | 22 Januari 1995 |

## Grants Acquired

FWO Aspirant Mandate for Fundamental Research under Grant Number 1179820N.

## Education

- **PhD in Engineering Science - Computer Science**
  KU Leuven 2018 - 2023

- **Master of Science in Mathematical Engineering**
  KU Leuven 2016 - 2018

- **Master of Science in Computer Science & Electrical Engineering**
  KU Leuven 2013 - 2016

# International Research Visits

Research Visit in the Lab by Ioannis Kevrekidis, Department of Chemical Engineering, Johns Hopkins University, Baltimore. The visit was split up in two periods: May 9 - July 16 2022 and August 28 - October 8 2022.

# Teaching Contributions

- Advisor of 7 Thesis Students from the Master of Mathematical Engineering and the Master of Mathematics, KU Leuven 2018 - 2023

- Numerical Analysis for Second Year Engineering Students: 5 years, 8 hours per week during the first semester, KU Leuven 2018- 2022

- Analysis 3 for Second Year Engineering Students: 1 year, 5 hours per week during the first semester, KU Leuven 2018

- Project Mathematical Engineering for Master Students in Mathematical Engineering for 2 years, 7 weeks per year, KU Leuven 2022 - 2023

- Linear Algebra for Architecture Students, 1 Semester
  KU Leuven 2022

# List of Publications

## International Journal Papers

Bello-Rivas, J. M., Georgiou A., Vandecasteele H. and Kevrekidis, I. G., Gentlest Ascent Dynamics on Manifolds Defined by Adaptively Sampled Point-Clouds, *Accepted for Publication in the Journal for Physical Chemistry* (2023).

Vandecasteele, H. and Samaey, G., A Micro-Macro Markov Chain Monte Carlo Method for Molecular Dynamics using Reaction Coordinate Proposals, *SIAM Journal on Scientific Computing* 45.2 (2023), B107-B138.

Zieliński, P., Vandecasteele, H. and Samaey G., Convergence and Stability of a Micro-Macro Acceleration Method: Linear Slow-Fast Stochastic Differential Equations with Additive Noise, *Journal of Computational and Applied Mathematics* 387 (2021), 112490

Vandecasteele, H., Zieliński, P., and Samaey G., Efficiency of a Micro-Macro Acceleration Method for Scale-Separated Stochastic Differential Equations, *Multiscale Modeling & Simulation* 18 (3) (2020), 1272-1298

## International Journal Papers that are Submitted or in Preparation

Vandecasteele, H. and Samaey G., Pseudo-Marginal Approximation to the Free Energy in a Micro-Macro Markov Chain Monte Carlo Method, arXiv preprint arXiv:2303.14422

Koellermeier, J., and Vandecasteele H., Hierarchical Micro-Macro Acceleration for Moment Models of Kinetic Equations, arXiv preprint arXiv:2209.13356

Vandecasteele, H., and Samaey, G., Computational Efficiency Study of a Micro-Macro Markov Chain Monte Carlo Method for Molecular Dynamics, arXiv preprint arXiv:2209.13056

# International Conference Talks

Vandecasteele H., Zieliński, P., and Samaey G., Analysis of Efficient Simulation Methods for Stochastic Processes, *Conference on the Numerical Solution of Differential and Differential-Algebraic Equations, NUMDIFF-15*, Martin Luther University Halle-Wittenberg, Germany (2018)

Vandecasteele H., Zieliński, P., and Samaey G., A Micro-Macro Acceleration Method for Stiff Stochastic Differential Equations, *SIAM Conference on Computational Science and Engineering*, Spokane, Washington State, USA (2019)

Vandecasteele H. and Samaey G., A Micro-Macro Markov-Chain Monte Carlo Method for Molecular Dynamics, *International Conference on Scientific Computation and Differential Equations*, Innsbrück, Austria (2019)

Vandecasteele H. and Samaey G., A Micro-Macro Markov Chain Monte Carlo Method for Molecular Dynamics, *15th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, Linz, Austria (2022)

Vandecasteele H. and Samaey G., Pseudo-Marginal Approximation to the Free Energy in a Micro-Macro MCMC Method, *SIAM Conference on Computational Science and Engineering*, Amsterdam, The Netherlands (2023)

Vandecasteele H. and Samaey G., Pseudo-Marginal Approximation to the Free Energy in a Micro-Macro MCMC Method, *Copper Mountain Conference on Multigrid Methods*, Copper Mountain, Colorado, USA (2023)

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF COMPUTER SCIENCE
NUMA
Celestijnenlaan 200A box 2402
B-3001 Leuven
hannes.vandecasteele@kuleuven.be
http://www.wms.cs.kuleuven.be/groups/NUMA