Non-Reversible Markov-Chain Monte Carlo: Theoretical Foundations and Applications in Molecular Simulation

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Philipp Höllmer
aus
Hamburg

Bonn, August 2023

Summary

The overarching objective of this doctoral thesis is the development of a rigorous paradigm for molecular simulation based on non-reversible Markov-chain Monte Carlo (MCMC) algorithms. The numerical exploration of the thermodynamic equilibrium of complex molecular systems, such as proteins in water, is of enormous importance to numerous fields including physics, chemistry, biology, and engineering. Based on the atomic hypothesis that all matter consists of atoms, molecular-mechanics models describe complex molecular systems by classical atomic sites that interact through empirical potential-energy functions. This enables efficient computer simulations of large-scale molecular systems. As a benchmark for different computational methods, dipolar water models are of particular interest because they appear in many molecular simulations as an explicit aqueous solution and because they typically contain the performance-limiting long-range interactions.

This doctoral thesis discusses theoretical foundations of non-reversible Markov chains in the context of the event-chain Monte Carlo (ECMC) algorithm for an ultimate application to molecular simulation. Non-reversible direction sweeps in an analytically tractable simplified model of a dipole profoundly modify the Markov-chain trajectory and introduce persistent rotations in both directions. The comparison of mixing times indicates that introducing direction sweeps into ECMC can yield faster rotation dynamics of the dipole. For a collection of dipoles, the rotation dynamics are characterized through the integrated autocorrelation time of the polarization. A large-scale numerical benchmark considers thousands of parameter sets for different ECMC variants and reveals remarkable speed differences among them. Escape times from almost locally stable hard-disk configurations are proposed as a model for the analysis of local MCMC algorithms. A scaling theory for the escape times of various ECMC variants separates them into two entirely different classes. The significant speedup of some ECMC variants is rooted in their lack of an intrinsic scale and their event-driven nature.

Motivated by the previous systematic evaluations of the manifold of ECMC variants, this doctoral thesis generalizes the Newtonian ECMC variant to molecular systems. This generalization preserves the fundamental properties of ECMC which enable rigorous molecular simulations in the canonical ensemble. The Boltzmann distribution is strictly sampled by realizing a non-equilibrium system with steady-state probability flows. Long-range interactions are treated without approximations. This doctoral thesis is accompanied by an implementation of generalized Newtonian ECMC in the JeLLyFysh application. Simulations of N long-range-interacting water molecules confirm the expected $O(N \log N)$ computational complexity. This matches the complexity of state-of-the-art molecular simulation with the widely-used molecular-dynamics (MD) method. However, MD treats long-range interactions inaccurately. JeLLyFysh reaches a break-even point with respect to a long-developed standard MD code below machine precision. This proves the competitive efficiency of ECMC or, more generally, of non-reversible MCMC algorithms in a rigorous paradigm for molecular simulation that is free of approximations and unphysical artifacts. It thus promises to become a gold standard for critical applications.

List of Publications

The following publications are part of this doctoral thesis:

[P1] Liang Qin, **Philipp Höllmer**, and Werner Krauth,
Direction-sweep Markov chains,
Journal of Physics A: Mathematical and Theoretical **55** (2022) 105003,
URL: https://doi.org/10.1088/1751-8121/ac508a.
The accepted manuscript of this article is reproduced in Chapter 3.

[P2] **Philipp Höllmer**, Anthony C. Maggs, and Werner Krauth,
Hard-disk dipoles and non-reversible Markov chains,
The Journal of Chemical Physics **156** (2022) 084108,
URL: https://doi.org/10.1063/5.0080101.
The published version of this article is reproduced in Chapter 4.

[P3] **Philipp Höllmer**, Nicolas Noirault, Botao Li, Anthony C. Maggs, and Werner Krauth,
Sparse Hard-Disk Packings and Local Markov Chains,
Journal of Statistical Physics **187** (2022) 31,
URL: https://doi.org/10.1007/s10955-022-02908-4.
The published version of this article is reproduced in Chapter 5.

Moreover, the following preprint is included in this doctoral thesis:

[P4] **Philipp Höllmer**, Anthony C. Maggs, and Werner Krauth,
Molecular simulation from modern statistics: Continuous-time, continuous-space, exact,
arXiv:2305.02979v2 [physics.chem-ph] (2023),
URL: https://doi.org/10.48550/arXiv.2305.02979.
The manuscript submitted for publication is reproduced in Chapter 6 and rephrased for the context of this doctoral thesis in Section 6.0.

In addition, I contributed to the following publication during the time of my doctoral studies:

[P5] Botao Li, Yoshihiko Nishikawa, **Philipp Höllmer**, Louis Carillo, Anthony C. Maggs, and Werner Krauth,
Hard-disk computer simulations—a historic perspective,
The Journal of Chemical Physics **157** (2022) 234111,
URL: https://doi.org/10.1063/5.0126437.

The following publications were produced before the start of my doctoral studies:

[P6] **Philipp Höllmer**, Liang Qin, Michael F. Faulkner, Anthony C. Maggs, and Werner Krauth,
JeLLyFysh-Version1.0 — a Python application for all-atom event-chain Monte Carlo,
Computer Physics Communications **253** (2020) 107168,
URL: https://doi.org/10.1016/j.cpc.2020.107168.

[P7] **Philipp Höllmer**, Jean-Sébastien Bernier, Corinna Kollath, Christian Baals, Bodhaditya Santra, and Herwig Ott,
Talbot effect in the presence of interactions,
Physical Review A **100** (2019) 063613,
URL: https://doi.org/10.1103/PhysRevA.100.063613.

I am the main developer of the following open-source applications:

[P8] BIGBORO: An arbitrary-precision Python/Go software package for Böröczky packings with event-chain Monte Carlo computations. Version 1.0 was released in connection with Ref. [P3] during the time of my doctoral studies.
URL: https://github.com/jellyfysh/BigBoro.

[P9] JeLLyFysh: A general-purpose Python application for all-atom event-chain Monte Carlo. Version 1.0 was released in connection with Ref. [P6] before the start of my doctoral studies. Versions 1.1 and 2.0 were released in connection with Ref. [P4] during the time of my doctoral studies.
URL: https://github.com/jellyfysh/JeLLyFysh/tree/Version2.0-alpha.

In addition, I am a co-developer of the following open-source application:

[P10] HISTORICDISKS: A synopsis of pressure data, sampling algorithms, and pressure estimators for the hard-disk system of statistical physics. Version 1.0 was released in connection with Ref. [P5] during the time of my doctoral studies.
URL: https://github.com/jellyfysh/HistoricDisks.

Contents

Introduction

As much as this doctoral thesis is a documentation of novel scientific insights and conclusions, it is also a documentation of a scientific methodology. It provides an example of systematically breaking down a difficult scientific problem into a series of smaller problems. An uncertain solution on a long time scale is approached by a series of solutions on shorter time scales. Even though the secondary problems often turn out to be interesting on their own, one always keeps the primary problem in mind. Any new scientific insight is fed back into an *evolving* solution of the overarching objective. This doctoral thesis also bears witness to the fact that problems which appear to have an obvious solution sometimes (or maybe even often) turn out to be much more difficult than expected. This should be understood as an opportunity because it paves the way to new scientific knowledge.

The overarching objective of this doctoral thesis is the development of a new rigorous paradigm for molecular simulations. These are of great importance to many physical sciences and can be motivated by the atomic hypothesis that all matter consists of atoms, which was described as the greatest insight of science by Nobel Laureate R. P. Feynman (see Section 1.1). Especially for simulations of molecular systems with long-range Coulomb interactions such as water, the prevalent computational method of choice is molecular dynamics. It numerically integrates the classical equations of motion of the atoms in discretized time steps. Over the last decades, powerful molecular-dynamics software packages have been developed, yet it was understood that its approximations and unphysical artifacts have to be controlled algorithmically (see Section 1.2). Molecular dynamics is the natural choice to obtain dynamical properties. However, it is frequently only employed as a means of sampling from the Boltzmann distribution of the canonical ensemble by coupling the system to a thermostat that mimics the effect of a thermal reservoir. Markov-chain Monte Carlo algorithms are designed from the beginning to rigorously sample a target probability distribution and, hence, can explore the thermodynamic equilibrium without thermostats. In order to achieve that purpose in the most efficient way, they are furthermore not artificially restricted to mirror the real microscopic dynamical behavior of the atoms (see Section 1.3). This makes the Markov-chain Monte Carlo sampling technique, in principle, a promising alternative to the thermostatted molecular-dynamics approach to obtain thermodynamic averages in molecular simulations.

Unfortunately, designing general and efficient Markov-chain Monte Carlo algorithms is a complex endeavor, especially in molecular systems. Historically, most developments focused on local reversible Markov-chain Monte Carlo algorithms whose behavior is inspired by the microscopic time-reversibility of a system in thermodynamic equilibrium. Their slow diffusive dynamics and their subpar computational complexity in long-range-interacting systems were not competitive, which resulted in a dominance of molecular dynamics in molecular simulation (see Section 1.3.1). Nowadays, systematic non-reversible Markov-chain Monte Carlo algorithms are available and show the potential to overcome the drawbacks of their traditional reversible counterparts. The ballistic dynamics of the local non-reversible event-chain Monte Carlo algorithm, in particular, led to spectacular speedups in the hard-disk system of statistical physics. Moreover, it was theoretically proven that it can treat long-range interactions in molecular systems rigorously exact with a computational complexity that is on par with molecular dynamics. This motivated the implementation of the first version of the general-purpose open-source event-chain Monte Carlo application JᴇLLʏFʏsʜ by the author and collaborators shortly before the studies for this doctoral thesis begun. Alas, large-scale simulations of long-range-interacting water systems in JᴇLLʏFʏsʜ-Version1.0 as a first practical benchmark revealed that the individual molecules remained dynamically arrested for prohibitively long times (see Section 1.3.2).

At the beginning of the studies for this doctoral thesis, the straight event-chain Monte Carlo variant that was implemented in JᴇLLʏFʏsʜ-Version1.0 clearly appeared as the foundation of a novel paradigm for molecular simulation that is rigorously free of approximations, artifacts, and thermostats, but yet remains efficient. However, because of its prohibitively slow dynamics in the water system, the establishment of such a paradigm turned out to be much more difficult than expected. Eventually enabling a competitive non-reversible Markov-chain Monte Carlo approach to molecular simulation thus became the overarching objective that inspired the first three publications of this doctoral thesis [P1, P2, P3]. They emphasize that event-chain Monte Carlo should be fundamentally understood as an entire family of local non-reversible Markov-chain Monte Carlo algorithms where each variant may have widely different behaviors in different applications. By using both analytical and numerical approaches, they systematically evaluate the manifold of possible event-chain Monte Carlo variants that were proposed over the years by different research groups. More generally, irrespective of their initial motivation, they also advocate specific models for the analysis and benchmarking of any local Markov-chain Monte Carlo algorithm, and reveal the fundamental changes that are introduced by non-reversibility. As the evolving solution to the overarching objective, one particular variant of event-chain Monte Carlo emerged as the most promising candidate for molecular simulation (see Section 1.4). This doctoral thesis thus culminates in the fourth research paper [P4] which first generalizes that particular variant from the peculiar hard-sphere system, for which it was initially proposed, to the general interaction potentials of molecular systems. The development of a major update of JᴇLLʏFʏsʜ allows to show that event-chain Monte Carlo is finally competitive to a long-developed molecular-dynamics application at high precision in a long-range-interacting water system (see Section 1.4). This lays the foundation of a paradigm for molecular simulation that is based on general non-reversible Markov-chain Monte Carlo algorithms. Since it is free of approximations and artifacts, it promises to grow into a gold standard for critical applications. As the extent of non-reversible Markov-chain Monte Carlo methods remains largely unexplored in their young scientific history, the novel paradigm provides great potential for further improvement in the future.

1.1 Molecular Simulation—From Feynman to Molecular Mechanics

It is rather fortunate for posterity that the 1961–1963 introductory physics lectures at the California Institute of Technology by R. P. Feynman were tape-recorded. The edited transcript—which "required the close attention of a professional physicist for from ten to twenty hours per lecture" [1, Foreword]—resulted in the timeless three-volume textbook *The Feynman Lectures on Physics* by R. P. Feynman and fellow physicists R. B. Leighton and M. L. Sand [1–3] that presents Feynman's physics about "simplicity, beauty, unity and analogy" with "enthusiasm and insight" [4, p. 30].

It is rather fortunate for this doctoral thesis that *The Feynman Lectures on Physics* lays the foundation for molecular simulation in the very first lecture titled "Atoms in Motion" [1, Section 1-2]:

> If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generations of creatures, what statement would contain the most information in the fewest words? I believe it is the *atomic hypothesis* (or the atomic *fact*, or whatever you wish to call it) that *all things are made of atoms—little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another*. In that one sentence, you will see, there is an *enormous* amount of information about the world, if just a little imagination and thinking are applied.

> —R. P. Feynman, R. B. Leighton, and M. L. Sands.[1]

The atomic hypothesis that all matter consists of fundamental constituents was already contemplated by philosophers in ancient Greek and India [5]. In the early nineteenth century, the atomic theory of chemist J. Dalton grew beyond the older empiricism and formulated fixed rules according to which chemical compounds are composed of discrete units of weights called atoms (derived from the Greek word ἄτομος (atomos) meaning "uncuttable") [6]. We know today that Dalton's atoms like hydrogen and oxygen are, in fact, "cuttable" into the subatomic elementary particles of the standard model of particle physics (see, e.g., Ref. [7]). However, as assessed by Nobel Laureate P. W. Anderson, the *reductionist* hypothesis that everything can be reduced to this (almost) fundamental theory does by no means imply a *constructionist* hypothesis, that is, the ability to reconstruct all sciences, from physics to chemistry to biology to psychology, by applying these fundamental laws [8]. In that sense, we can envisage atoms as the elementary entities, and understand properties and behaviors of chemical compounds with the help of just-as-fundamental laws for the atoms.

Feynman applies his "little imagination and thinking" to atoms that attract each other at little distances but repel when they become too close, and that can form chemical (covalent) bonds as, for example, in a water molecule where an oxygen has two hydrogens tied to it. Heat is represented as the jiggling motion of the atoms that increases with temperature. These statements alone allow one to understand why a drop of water does not fall apart at standard conditions—because the attraction between the molecules wins—but why the water molecules do fly apart upon heating—because the

---

[1] At the time of writing, it is not yet possible to include audio in the printed version of this doctoral thesis, and an audiobook version is currently not in planning. Thus, if the reader would like to hear R. P. Feynman himself say the words that led to the quoted paragraph, we refer to the tape recordings of his 1961–1963 introductory physics lectures available at https://www.feynmanlectures.caltech.edu/flptapes.html. The relevant part starts at 9′12″ of recording #1.

jiggling motion wins. These statements also lead to a kinematic definition of the pressure and a crystalline structure of solids [1, Section 1-2]. Ultimately, one may even understand atomic processes such as salt dissolving in water [1, Section 1-3] and chemical reactions [1, Section 1-4].

Consequently following the line of thought of interacting, jiggling atoms from *The Feynman Lectures of Physics* leads to the realm of classical *molecular simulation*, whose ongoing importance to numerous fields, ranging from physics to chemistry to biology to engineering, cannot be overemphasized (see, e.g., Refs [9–15]). To name but a few examples, molecular simulations predict protein-folding reactions [16] and elucidate the atomic-level behavior during such a conformational change [17]. They are also an essential part of modern-day drug discovery by illuminating the processes of drug–target recognition and binding [18, 19]. They rely on *molecular-mechanics models* that approximate a molecular system with a fixed covalent structure (ranging from small chemical systems to large biomolecules) as an assemblage of classical point-like sites which interact through an empirical potential-energy function U [20, 21]. The physical behavior of the interaction sites, which do not have to correspond to physical atoms (see, e.g., Ref. [14, Section 2.5]), may then be treated efficiently with computational methods, where the two corner stones are the molecular-dynamics method and the Markov-chain Monte Carlo sampling technique (see Sections 1.2 and 1.3).

The functional form of the potential energy in a molecular-mechanics model is usually motivated by physical intuition and insight. For instance, the London dispersion interaction between two charge-neutral atoms with a symmetric electron distribution leads to an attraction at little distances, while the Pauli exclusion principle leads to a strong repulsion when the atoms get too close. Both effects may be well approximated by a Lennard-Jones potential between two interaction sites in a molecular-mechanics model [21, 22]. As another example, the vibration of the length of a covalent bond between two atoms around an equilibrium length is usually considered in a harmonic approximation. Such approximations in a molecular-mechanics model generally result in an overall potential-energy function $U = \sum_M U_M$ that is a sum of *factor potentials U_M* of *factors M* which only depend on a small number of interaction sites. The factor potentials usually depend on constant parameters like the equilibrium length of the harmonic-bond potential, or the depth and location of the potential well in the Lennard-Jones potential. These are empirically chosen by comparing the predicted values of physical observables in the molecular-mechanics model with experimental data, detailed *ab initio* quantum-mechanical calculations, or both [20, 21]. Naturally, this approach assumes that the empirical parameters are transferable from relatively simple systems, in which they were fine-tuned, to more complex systems, in which they are eventually used to make predictions. As it turns out, this assumption is usually valid and when it is not, the discrepancy may actually give an opportunity to improve the understanding of the problematic system [21].

A molecule of particular interest to molecular simulation is water because of its ubiquity in our environment. Many molecular simulations, especially of biological systems, include explicit water molecules to model an aqueous solution. Since the seminal proposal for a classical model of water by J. D. Bernal and R. H. Fowler in 1933 [23], numerous water models with widely varying complexities attempt to reproduce the experimentally well-studied bulk properties of water with varying success (see, e.g., Refs [24–27]). The most simple models of a water molecule use three interaction sites which makes them computationally feasible and thus widely used. The flexible simple point-charge
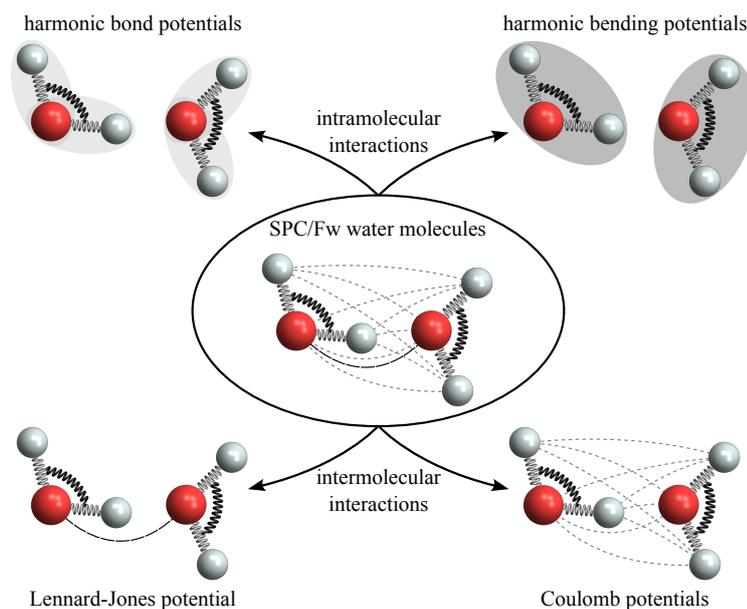
Figure 1.1: The SPC/Fw molecular-mechanics model of water uses three charged interaction sites that are centered on the red oxygen (O) and white hydrogen (H) nuclei [28]. Harmonic bond and bending potentials, indicated as white and black springs, allow for fluctuations of the O−H bond lengths and H−O−H opening angles around empirical equilibrium values. Lennard-Jones and Coulomb potentials, indicated as black and gray lines, yield long-range interactions between different water molecules. Every potential only depends on a small number of interaction sites.

water model SPC/Fw [28] centers three interaction sites on the atomic oxygen (O) and hydrogen (H) nuclei. Each water molecule has three internal degrees of freedom, two O−H bond lengths and one H−O−H opening angle, that may fluctuate around their equilibrium value in intramolecular harmonic potentials. The intermolecular interactions between two water molecules consist of a Lennard-Jones interaction that is centered on the two oxygen sites, and Coulomb potentials that consider the electrostatic interactions between all pairs of charged interaction sites (see Fig. 1.1). In comparison to three-site water models that use rigid O−H bonds and H−O−H opening angles (such as the TIP3P [29] and SPC/E [30] models), the SPC/Fw water model reflects, e.g., dielectric properties of bulk water better [28].

1.2 Molecular Dynamics and Its Problems

While a molecular-mechanics model heavily reduces the complexity of the treatment of a molecular system by replacing a quantum-mechanical description by a much more simple, but approximate, classical one, an analytical treatment is generally impossible (with one notable exception given by works of F. H. Westheimer from the 1940s [31, 32]). As a product of today's computer age, molecular simulations use computational methods to numerically explore the structure, dynamics, and thermodynamic equilibrium of complex molecular systems. The transition from quantum-mechanical

ab initio computations to classical ones allows for order-of-magnitudes faster and larger simulations. Furthermore, as a more philosophical advantage, classical models are conceptually easier and provide us, as humans, the opportunity to understand molecular systems in the same way as the physical objects of our everyday life [20, 21]. (It must be noted, however, that classical molecular simulation can clearly not be used for processes where electronic effects such as breaking of covalent bonds are involved.)

Especially for long-range-interacting molecular systems that contain electrostatic interactions between charged interaction sites (as in the SPC/Fw water model), the most common computational method of choice is *molecular dynamics* (MD). Here, the interaction sites follow the familiar laws of classical, Newtonian mechanics. The negative gradients of the (time-independent) potential-energy function U of the molecular-mechanics model yield the forces at the positions of the interaction sites (which is why the potential-energy function is often called a force field). Newton's equations of motion are numerically integrated in discrete time steps. This approach mimics the dynamical evolution of the simulated system in physical time. It gives access to dynamical properties and, for systems that obey the ergodic hypothesis, it samples the microcanonical (NVE) ensemble [9–12]. A wide range of thermostats and barostats may be introduced in MD simulations. They aim to control the temperature or pressure and thus sample the canonical (NVT) or isothermal–isobaric (NPT) ensembles instead (see, e.g., Refs [9–12, 33, 34]).

Formally, widely-used symplectic numerical integrators with a small-enough choice of the finite time step can guarantee long-time stability of an MD simulation which makes it a well-behaved computational method. In the microcanonical ensemble, for instance, the total energy, which should be exactly conserved under Newton's equations of motion, can be shown to fluctuate within a narrow band around the true value with an exponentially small energy drift over an exponentially long period of time (see Refs [35–37] for the original explicit proofs; see Refs [38, Chapter 5] and [39, Chapter 5] for extensive reviews). These results technically consider an infinite precision in the force computation. In many practical MD simulations of molecular systems, however, the strict mathematical results *break down* because simplifications in the force computation become too inaccurate. This can then introduce a systematic drift of the total energy (and even unphysical artifacts, as discussed below) [38, Chapter 11].

For long-range interacting molecular systems as the SPC/Fw water model under the usual periodic boundary conditions, such inaccurate simplifications typically regard the long-range intermolecular interactions to reduce computational cost. Systematic drift of the total energy can be introduced because the quickly decaying Lennard-Jones potential is cut off in a non-analytic way [40–43], or because the cutoff is practically implemented with infrequent neighbor-list updates so that interactions are occasionally missed [43, 44]. The slowly decaying Coulomb potential (which should not be simply cut off [45, 46]) is usually treated by an Ewald summation which divides it into two short-range contributions in real and reciprocal space [12, 47, 48]. The two involved cutoffs and an additional discretization of reciprocal space in fast mesh-based Ewald methods [49–51] lead to a limited accuracy of the force computation that is typically selected to be much worse than allowed by machine precision because the computational cost grows with increasing accuracy. This can also lead to a systematic drift in the total energy [43, 49, 51–53]. As a concrete example of these effects, an MD simulation of $N = 1000$ SPC/Fw water molecules at a standard density of $\rho \approx 1\,\mathrm{g\,cm^{-3}}$ in the microcanonical

ensemble with the LAMMPS software [54, 55] that treats the long-range intermolecular interactions with these inaccurate (but computationally fast) simplifications shows a non-negligible systematic drift of the total energy (see Fig. 1.2a and b).

Most MD simulations are not carried out in the microcanonical ensemble because most common experimental setups are not isolated from their surroundings and, hence, the condition of a constant total energy is not valid. Simulations in other statistical ensembles are necessary for different thermodynamic control parameters. One very important example is the canonical ensemble that controls the temperature instead of the total energy. In MD, this is achieved by including a thermostat which perturbs the trajectories and mimics the effect of a coupled thermal reservoir. At the same time, the thermostat can "bleed out" any extra energy resulting from inaccurate force computations, which would have shown up as a systematic energy drift in the microcanonical ensemble. This can create a steady-state system with a constant temperature that, however, does not necessarily obey the desired canonical distribution [44, 57]. Another MD simulation of $N = 1000$ SPC/Fw water molecules at standard density with exactly the same treatment of long-range intermolecular interactions as the microcanonical simulation with a noticeable systematic drift of total energy (see Fig. 1.2a and b again), but now also including a thermostat to achieve a canonical simulation at room temperature $T \approx 300$ K, shows that the inaccuracy of the simulation is masked by the thermostat as both the total energy and temperature show much larger fluctuations (see Fig. 1.2c and d).

A certain small amount of inaccuracy in MD simulations of molecular systems is usually accepted (see, e.g., Ref. [53, Section III C 2] for a short discussion about simulations of water models). This does not lessen its power and effectiveness for molecular simulation, which was proven time and time again in a great range of successful applications (see, e.g., Refs [9–19] again). Moreover, systematic analyses and tests for the accuracy of MD simulations are available [40, 44, 53, 57]. Still, there are many examples of unphysical artifacts in MD simulations that had to be understood and overcome in a voluminous literature over the last decades (see, e.g., Ref. [46] for a review). One severe example is the observation of a unidirectional flow of water through a carbon nanotube without any supplied energy [46, Section 2]. This violates the second law of thermodynamics which physicists are generally not fond of.[2] Indeed, the unidirectional flow vanished as soon as the parameters and protocols of the MD simulation were properly chosen. Explicit examples for reasons of unphysical artifacts (that, in this case, go beyond a systematic drift of a conserved quantity) include, again, non-careful truncations of long-range interactions [45, 46, 59], discretization of reciprocal space in fast mesh-based Ewald methods that may lead to force aliasing [49], and too infrequent updates of neighbor lists [46, 60]. Thermostats—which were called "necessary evils" in Ref. [46] and of which some do not even correctly reproduce the canonical ensemble [61, 62]—were another big source of unphysical artifacts as the "flying ice cube effect" [63] and the "hot solvent–cold solute problem" [64, 65].

In MD, the underlying microscopic equations of motion of the system are numerically integrated. This can be viewed as a strength because it allows to access dynamical properties. However, it also introduces a *time-scale problem* [15, 66]. For a well-behaved simulation with long-time stability, the time step has to be chosen small enough to ensure that the fastest motion is integrated accurately. For instance, vibrations of the O−H bond lengths or the H−O−H opening angle in the SPC/Fw water

---

[2] To put it in Homer Simpson's words [58]: "In this house we obey the laws of thermodynamics."
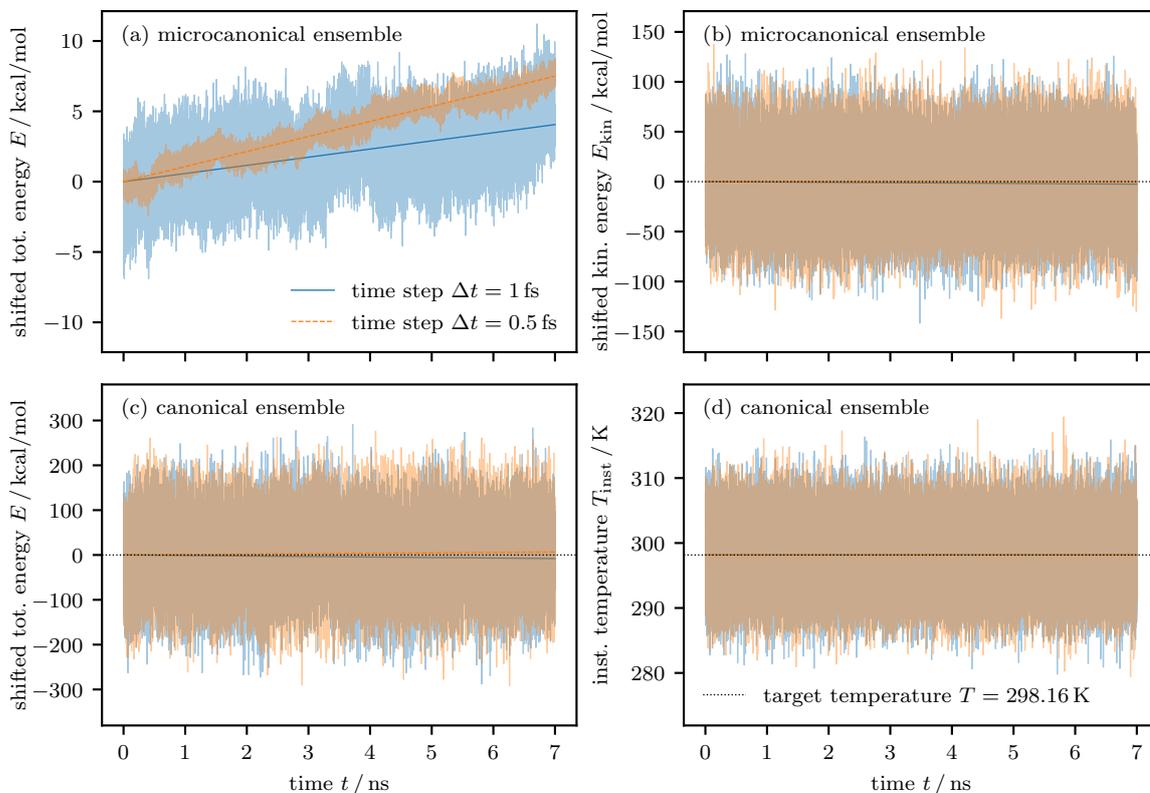
Figure 1.2: MD simulations of $N = 1000$ SPC/Fw water molecules show a non-negligible drift of the total energy E in the microcanonical ensemble that is masked by the thermostat in the canonical ensemble. All plots show a linear fit (solid and dashed lines) to the instantaneous fluctuations (light background) of the quantity of interest. The plots (a), (b), and (c) are shifted by the negative fitted intercept. (a): The fitted drift of the total energy in the microcanonical ensemble increases from $\Delta E = (0.579 \pm 0.003)\,\mathrm{kcal\,mol^{-1}\,ns^{-1}}$ for the time step $\Delta t = 1\,\mathrm{fs}$ to $\Delta E = (1.071 \pm 0.001)\,\mathrm{kcal\,mol^{-1}\,ns^{-1}}$ for $\Delta t = 0.5\,\mathrm{fs}$ which shows that an inaccurate force computation is responsible for the drift [34]. (b): The drift of the total energy quickly becomes non-negligible compared to the fluctuations $\sigma(E_{\mathrm{kin}}) \approx 30\,\mathrm{kcal\,mol^{-1}}$ of the kinetic energy E_{kin} around its average $\langle E_{\mathrm{kin}} \rangle \approx 2\,670\,\mathrm{kcal\,mol^{-1}}$. (c): Any drift of the total energy is masked by the fluctuations that are introduced by the thermostat in the canonical ensemble. (d): The time-average of the instantaneous temperature T_{inst} that follows from the instantaneous kinetic energy quickly approaches the thermostat's target temperature T despite the inaccurate force computation.

All simulations were carried out with the LAMMPS software [54, 55]. They started from the same initial velocities, and from the same initial configuration with total energy $E \approx -7\,521\,\mathrm{kcal\,mol^{-1}}$. This configuration was obtained by equilibrating a configuration from the software package PLAYMOL [56] in an MD simulation of LAMMPS in the isothermal–isobaric ensemble at target temperature $T = 298.16\,\mathrm{K}$ and pressure $P = 1\,\mathrm{atm}$ with time step $\Delta t = 1\,\mathrm{fs}$ for a total time $t = 1\,\mathrm{ns}$. This resulted in a configuration at standard density $\rho \approx 1\,\mathrm{g\,cm^{-3}}$ (using the hydrogen mass $m_{\mathrm{H}} = 1.0079\,\mathrm{Da}$ and the oxygen mass $m_{\mathrm{O}} = 15.9994001\,\mathrm{Da}$) in a periodic cubic box of side length $L \approx 31\,\text{Å}$. The Coulomb potential was treated by a fast mesh-based Ewald method with a target accuracy of 10^{-5} while the Lennard-Jones potential was cut off beyond $9.8\,\text{Å}$.

model occur over time scales of a few femtoseconds. In contrast, conformational changes of molecules or proteins in an aqueous solution may occur on macroscopic time scales from milliseconds to hours. Sampling the conformations in thermodynamic equilibrium and accessing the transition dynamics in a reasonable amount of computation time is then impossible. One possible approach to the time-scale problem introduces a systematic bias into the simulation in order to enforce certain motions such as conformational changes (see, e.g., Refs [67, 68]). Such enhanced-sampling techniques (which are not restricted to MD but can also enhance Markov-chain Monte Carlo simulations) cannot fully replace unbiased molecular simulations because of their requirement for *a priori* information. Instead, they can complement the unbiased molecular simulations that are considered in this doctoral thesis. Other approaches to the time-scale problem consist of constraining bond lengths and opening angles to fixed values, or to rely on multiple time-scale integration algorithms [15]. Both approaches can lead to inaccuracies and unphysical artifacts on their own, and hence motivated the development of more advanced algorithmic controls (see, e.g., Refs [69, 70]).

MD is nowadays the prevalent computational method of choice in molecular simulation, especially if long-range interactions like the Coulomb interaction are present (for reasons that are clarified in the next Section 1.3). This is always the correct approach if dynamical properties are of interest because its numerical integration of the equations of motion provides a window into the microscopical behavior of the individual interaction sites in the molecular-mechanics model. However, as soon as, for instance, a thermostat attempts to mimic the effect of a thermal reservoir in the simulation, the trajectories are perturbed and thus also any dynamical properties. Although it was empirically found that some thermostats only have little effect on particular dynamical properties [71], it is generally preferable to sample a set of configurations from the thermodynamic equilibrium with the thermostat turned on, and use these as initial conditions for dynamical trajectories with the thermostat turned off to infer dynamical properties (see, e.g., Refs [12, 34]). Here it becomes evident that MD simulations entangle two objectives: first, generating dynamical trajectories—for which it was designed—and second, sampling from statistical ensembles and determining equilibrium properties—for which it was not designed but is still used to a large extent. *The great effort to find and understand artifacts and approximations in MD, as well as to develop algorithmic controls for them, will always be necessary for the first objective. In contrast, computational methods that are based on modern non-reversible Markov-chain Monte algorithms can be specifically designed for the second objective and allow for a rigorously exact and efficient approach without any thermostats. The establishment of such an alternative paradigm for molecular simulation is the overarching objective of this doctoral thesis.*

1.3 Markov-Chain Monte Carlo

The alternative paradigm to the thermostatted MD approach for molecular systems in thermodynamic equilibrium is based on the *Markov-chain Monte Carlo* (MCMC) sampling technique. Since its introduction in 1953 in statistical physics [72], it has become an active research topic in mathematics (see, e.g., Refs [73, 74]) and has been used in an astonishing array of disciplines across all natural sciences and beyond (see, e.g., Refs [75–77]). MCMC is designed from the beginning to rigorously sample a given probability distribution. This doctoral thesis is in particular concerned with the

Boltzmann probability distribution π of the canonical ensemble (see, e.g., Refs [74, 78] in addition to the general Refs [9–12] on molecular simulation):

$$\pi(\mathbf{x}) = \frac{1}{Z} \, e^{-\beta U(\mathbf{x})}. \tag{1.1}$$

Here, $\beta = 1/(k_B T)$ is the thermodynamic beta that depends on the Boltzmann constant k_B and temperature T of the canonical ensemble, Z is the generally unknown canonical partition function that normalizes π, and U is the empirical potential-energy function of the molecular-mechanics model that depends on the positions $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ of the N interaction sites.

MD starts from an initial configuration and simulates a dynamical trajectory in physical time. MCMC simulates a memoryless stochastic process instead. It constructs a Markov chain, that is, a sequence of random variables $(X_0, X_1, X_2, \ldots)$ on Ω where for all discrete Monte Carlo times $t \in \mathbb{N}_0$ only the current state of X_t determines how the next state X_{t+1} is sampled. The previous states are unimportant because the Markov chain "lost its memory." Consider, for ease of notation, a finite configuration space Ω with configurations $x \in \Omega$. Although the configuration space of the interaction-site positions $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ in molecular systems is typically continuous (i.e., uncountably infinite), finite Markov chains already exhibit the essential features [73, Section 3.4].[3] (One may even argue that, on a computer with a finite amount of memory, every continuous configuration space becomes finite in its floating-point representation.) The characteristic memoryless behavior of a Markov chain is encoded in its time-independent transition matrix P whose entries $P(x, y)$ give the conditional probability to move from the configuration x to the configuration y. The current state $X_t = x_t$ of the Markov chain thus determines the row $P(x_t, \cdot)$ that stores the set of probabilities for all configurations $y \in \Omega$ to become the next state of the random variable X_{t+1}. MCMC algorithms then sample the next state $X_{t+1} = x_{t+1}$ according to its probability in a Monte Carlo move. This is possible without ever storing the entire transition matrix which makes the algorithms practicable in the first place.

The desired property of a Markov chain for molecular simulation is that the asymptotic proportion of time that it spends in any configuration $x \in \Omega$ is determined by the Boltzmann distribution $\pi(x)$. The time average over the sequence of states of the simulation can then be used to estimate thermodynamic averages. This can only be achieved if the transition matrix P satisfies the necessary *global-balance condition* that connects its entries to π [74]:

$$\pi(y) = \sum_{x \in \Omega} \pi(x) \, P(x, y) \quad \text{for all } y \in \Omega. \tag{1.2}$$

Additional conditions of irreducibility and aperiodicity on the Markov chain [73, 74] are properly introduced in Section 2.1 but can usually be established beyond doubt in practical applications [80]. Such a Markov chain exhibits four constructive properties [73, 74, 81]: First, the Markov chain converges to the target distribution π from any initial configuration. Second, the distribution π

---

[3] This approach complies with the common paraphrase of a quotation of Nobel Laureate A. Einstein [79]: "Everything should be made as simple as possible, but not simpler." With this, note that the (almost) obligatory A. Einstein quote in a doctoral thesis in physics was successfully included.

remains stationary from then on. Third, the time average over the Markov chain converges to the thermodynamic average. Fourth, the time average becomes normally distributed in a similar manner as the sample average in the classical central limit theorem. The fourth point is essential to understand the *errors* that arise from a Monte Carlo estimation in the third point.

Two important figures of merit of an MCMC algorithm are its *mixing time* and its *integrated autocorrelation time* (see Section 2.2 for their precise definitions). The former determines the time after which it can be guaranteed that the Markov chain samples the target distribution π. The latter gives a measure on the correlation between subsequent states in the Markov chain once π is reached, and determines the time after which an effectively independent sample is obtained. The art of designing MCMC algorithms lies in finding the transition matrix that minimizes the mixing and integrated autocorrelation times. In contrast to MD which is strictly restricted to physical Newtonian dynamics and is subject to the inherent time-scale problem, the only restriction for MCMC algorithms is the global-balance condition in Eq. (1.2). In principle, this offers the great opportunity to find non-physical dynamics that mix and decorrelate much more efficient than MD.

1.3.1 Reversible Markov-Chain Monte Carlo and Its Problems

The first MCMC algorithm was the Metropolis algorithm from 1953 [72]. It is successfully used in a wide variety of settings until today based on its beautifully simple though flexible mathematical recipe (explaining why it is ranked among the top ten algorithms of the twentieth century [82, 83]). Its transition matrix does indeed satisfy the necessary global-balance condition in Eq. (1.2) but it does so by satisfying the more restrictive *detailed-balance condition*:

$$\pi(x)\, P(x, y) = \pi(y)\, P(y, x) \quad \text{for all } x, y \in \Omega. \tag{1.3}$$

By using that the transition matrix has to be a stochastic matrix because of the interpretation of its entries as conditional probabilities (i.e., $\sum_{y \in \Omega} P(x, y) = 1$), one can directly show that any P that satisfies the sufficient Eq. (1.3) also satisfies the necessary Eq. (1.2). The Metropolis algorithm is furthermore of the acceptance–rejection type in which a Monte Carlo move, given the current configuration $X_t = x_t$, consists of two steps. First, a proposed configuration x'_{t+1} is sampled from the corresponding row of a symmetric proposal distribution $g(x_t, \cdot)$, where $g(x, y)$ gives the conditional *a priori* probability to propose the configuration y if the current state of the Markov chain is x. Second, the proposed configuration x'_{t+1} is either accepted in which case $X_{t+1} = x'_{t+1}$, or it is rejected in which case $X_{t+1} = x_t$. The acceptance probability is given by the Metropolis filter

$$p^{\text{Met}}(x_t, x'_{t+1}) = \min\left[1, \frac{\pi(x'_{t+1})}{\pi(x_t)}\right] = \min\left[1, e^{-\beta\left[U(x'_{t+1}) - U(x_t)\right]}\right]. \tag{1.4}$$

A rather important point about the Metropolis algorithm, and about MCMC algorithms in general, is that it only ever depends on relative probabilities of different configurations: The Metropolis filter solely contains the ratio $\pi(x'_{t+1})/\pi(x_t)$. Consequently, the canonical partition function Z in Eq. (1.1) need not be known in order to sample the Boltzmann distribution (and if it is known, a numerical treatment is not necessary anyway). Instead, only the difference in potential energy is important. If the

proposed configuration reduces the potential energy, it is always accepted in the Monte Carlo move; if it increases the potential energy, the acceptance probability decreases exponentially. The entries of the transition matrix P^{Met} of the Metropolis algorithm are given by

$$P^{\text{Met}}(x, y) = \begin{cases} g(x, y) \, p^{\text{Met}}(x, y) & \text{if } x \neq y, \\ 1 - \sum_{z \in \Omega, z \neq x} g(x, z) \, p^{\text{Met}}(x, z) & \text{if } x = y. \end{cases} \tag{1.5}$$

It is clearly a stochastic matrix and one can show that it satisfies the detailed-balance condition in Eq. (1.3) by distinguishing the cases $x \neq y$ or $x = y$, and $\pi(x) > \pi(y)$ or $\pi(x) \leq \pi(y)$.

Any MCMC algorithm that satisfies the detailed-balance condition in Eq. (1.3) is called *reversible*. It demands that every probability flow $\mathcal{F}(x, y) = \pi(x) \, P(x, y)$ from the configuration x to the configuration y is equal to the reverse probability flow $\mathcal{F}(y, x)$. The net probability flow vanishes between all configurations $x, y \in \Omega$: $\mathcal{F}(x, y) - \mathcal{F}(y, x) = 0$. This is inspired by the behavior of a physical system in thermodynamic equilibrium where every elementary process is in equilibrium with the reverse process because of the time-reversibility of the microscopic equations of motion (as first noted by L. Boltzmann in his work on the H-theorem; see, e.g., Ref. [84, Chapter VI]). Ever since the introduction of the reversible Metropolis algorithm in 1953, most MCMC algorithms remained reversible because the detailed-balance condition only considers pairs of configurations and is thus much easier to satisfy when a transition matrix is systematically designed. Two important examples are the reversible Metropolis–Hastings algorithm that modifies the Metropolis filter in Eq. (1.4) to allow for non-symmetric proposal distributions $g(x, y) \neq g(y, x)$ [85], and the reversible heat-bath algorithm [86–88]. The global-balance condition, in contrast, considers global probability flows into any configuration from the full set of configurations which makes a systematic treatment more complex.

The widely-used reversible Metropolis–Hastings algorithm becomes especially powerful when the *a priori* choice of the proposal distribution for the Monte Carlo moves can be systematically customized for the current configuration. In the best case, this allows to accept large, global changes of the configuration in a single Monte Carlo move and the configuration space is explored rapidly. The most famous examples for this are non-local cluster moves in the Swendsen–Wang [89] and Wolff [90] algorithms for spin systems. However, such insightful global Monte Carlo moves are not available for most problems and they are, in particular, missing for molecular systems. Alternatively, one then relies on *local moves* that, e.g., displace one of the interaction sites. Small sizes of the displacement ensure that the potential-energy change remains small and that, therefore, the acceptance probability of the local move remains reasonable. Such local Monte Carlo moves, however, lead to large correlations between subsequent configurations in the Markov chain because they are mostly the same. Coupled with the detailed-balance condition, this results in time-reversible *diffusive* dynamics [80, 91, 92] that explore the configuration space slowly. Although the Boltzmann distribution π is still rigorously sampled after the mixing time, a lot of Monte Carlo moves are required to converge and to obtain effectively independent samples.

The slow diffusive dynamics of local reversible MCMC algorithms is the first reason why MD is often preferred for molecular simulations, even when just the thermodynamic equilibrium is of interest. Although local Monte Carlo moves can be optimized beyond displacing a single interaction site for

the specific molecular system at hand (for example, by proposing explicit rotations of SPC/Fw water molecules), and although such optimized Monte Carlo move sets can, in certain cases, allow to cross potential-energy barriers between minima that would trap MD simulations on prohibitively long time scales [93], gradient-based techniques as MD are generally assumed to be "vastly superior" [94] for an efficient exploration of configuration space. Furthermore, choosing an appropriate set of Monte Carlo moves in practice may be challenging and requires a great amount of knowledge and experience in the specific molecular system. The Newtonian dynamics of MD, on the other hand, are automatically determined by the force field of the molecular-mechanics model.

Another reason for the dominance of MD over local reversible MCMC algorithms in molecular simulations is their respective treatment of long-range interactions between N interaction sites. In both cases, a naïve implementation would require $O(N^2)$ operations to modify all N positions. In MD, an Ewald summation can reduce the computational complexity to $O(N^{3/2})$ [12, 47, 48]. Fast mesh-based Ewald methods even achieve $O(N \log N)$ [49–51] and have thus become the prevalent methods of choice. Albeit the complexity of the fast multipole method [95] just grows as $O(N)$, its prohibitive prefactor currently makes it dominant only at rather large N [96, 97]. These complexities can be achieved in MD because a single force computation modifies the positions of all interaction sites. In contrast, local reversible MCMC algorithms have to compute the potential-energy change to attempt a move of a *single* or very few interaction sites. This prevents an efficient integration of the fast mesh-based Ewald methods and leaves an Ewald summation that can only achieve an $O(N^{3/2})$ complexity for a full sweep of N attempted Monte Carlo moves [98]. Only very recently, a variant of the fast multipole method accomplished an $O(N \log N)$ scaling [99] but a direct comparison to MD is missing.

Given the extensive discussion on the inaccurate treatment of long-range interactions in MD in Section 1.2, it should be further noted that local reversible MCMC algorithms are by no means better in that respect. The computation of the potential-energy change in a Monte Carlo move is a global computation that is again simplified to reduce computational cost. The Lennard-Jones potential is cut off and the Coulomb potential is treated by an Ewald summation with two cutoffs in real and reciprocal space. This leads to a limited accuracy that is typically orders-or-magnitude worse than allowed by the available machine precision because the proportionality constant of the $O(N^{3/2})$ complexity grows with increasing accuracy. Such an inaccurate treatment of long-range interactions certainly leads to a systematic error (see, e.g., Ref. [100]). However, in contrast to MD, the simulation still remains stable. In order to further emphasize the stability: MCMC algorithms that only consider potential-energy changes do not have to rely on carefully designed initial configurations. In MD, initial configurations from packing optimizations prevent excessive forces that would, in turn, necessitate prohibitively small time steps for a stable simulation [38, 101]. Local reversible MCMC, on the other hand, does not know about forces and could, in principle, relax any configuration (which could, however, require a very long time). Moreover, constraints on bond lengths or opening angles can be easily considered in the set of Monte Carlo moves.

Even though local reversible MCMC removes most reasons for unphysical artifacts and approximations, and even though it can be shown to sample the Boltzmann distribution in the infinite-time limit, MD is nowadays the most common choice to simulate the thermodynamic equilibrium of classical molecular

systems, especially when long-range interactions are present. A recent review that considered, among other things, best practices for MD simulations even called MCMC the "often-forgotten relative" [102]. As discussed, this is rooted in two problems: First, slow diffusive dynamics of local reversible MCMC algorithms may only be overcome in certain systems by a challenging choice of an appropriate Monte Carlo move set. Second, the computational complexity of the treatment of long-range interactions was historically much more favorable in MD. This resulted in a broad range of powerful computer packages for classical molecular simulation based on MD (see, e.g., Refs [54, 103–106]). They were developed and optimized for decades, and are therefore fast, reliant, and easily accessible (which was already relevant for this doctoral thesis because the MD simulations for Fig. 1.2 could be done efficiently in LAMMPS [54, 55] without programming anything). In a self-amplifying effect, a great deal of research was centered on improving the efficiency, accuracy, and stability of MD which resulted in state-of-the-art applications which in turn attracted more research on MD. In contrast, software packages for MCMC are sparse and tend to be developed by small groups [94] (see, e.g., Refs [107, 108] for examples of software that is available). *However, although MD evolved into an efficient and widely-distributed computational approach for molecular simulation, only MCMC can serve as a gold standard in general because it can guarantee an unbiased sampling of the Boltzmann distribution. MCMC simulations can thus identify any artifacts and approximations that are still hidden in MD and which are not yet controlled algorithmically. To achieve this, the slow diffusive dynamics, and the inefficient and inaccurate treatment of long-range interactions in local reversible MCMC algorithms have to be overcome.*

1.3.2 Non-Reversible Markov-Chain Monte Carlo to the Rescue (or not?)

The slow diffusive dynamics of local reversible MCMC algorithms follows from the detailed-balance condition in Eq. (1.3) that mirrors the behavior of a physical system in thermodynamic equilibrium. However, there is no reason why equilibrium dynamics should be the most efficient strategy for fast decorrelation and mixing in Markov-chain sampling. In fact, everyday experience tells us that this is definitely *not* the case as a simple experiment inspired by Ref. [109] shows: The writing of this doctoral thesis involved many cups of coffee. Although they did not contain any sugar in general, two of them did *for the sake of science* (and this experiment). The process of mixing sugar in a cup of coffee can be compared to a Markov chain. The configuration space of the sugar is the entire cup of coffee, the stationary target distribution is a uniform sugar concentration, and the initial distribution of sugar is usually concentrated at the bottom of the cup. In the first cup, equilibrium detailed-balance dynamics (that is, simply waiting) were used to mix the sugar into the coffee. The diffusive dynamics resulted in an enormous mixing time and, hence, in a coffee that was not only sugary but also cold. In the second cup, a non-equilibrium external flow was introduced into the system by stirring the coffee with a spoon. Amazingly, the mixing time was significantly reduced and the uniform-sugary coffee could be "enjoyed" while still hot.[4]

The detailed-balance condition and its diffusive dynamics in local reversible Markov chains are not necessary. *Non-reversible* Markov chains break the detailed-balance condition and only satisfy the

---

[4] The reader is invited to replicate this experiment in order to verify its results. Alternatively, accelerating mixing fluids by stirring is an active research topic that supports the results (see, e.g., Ref. [110] and references therein).

necessary global-balance condition in Eq. (1.2). Under the additional conditions of irreducibility and aperiodicity that are usually valid in practical applications (see Section 2.1), they converge to the target distribution π but may include non-vanishing net probability flows $\mathcal{F}(x, y) - \mathcal{F}(y, x)$ between any configurations $x, y \in \Omega$. After convergence, they further sample the distribution π in a non-equilibrium random process. Because they are harder to conceptualize, systematic non-reversible Markov-chain schemes remained unavailable until the *lifting framework* [111, 112] was introduced around the turn of the millennium (see Section 2.3 for a theoretical introduction). It starts from a reversible Markov chain and enlarges its configuration space Ω to a lifted configuration space $\widehat{\Omega}$ so that it contains some or all of the Monte Carlo moves and explicit transition probabilities between them in a lifted transition matrix $\widehat{P}$ [80, 113]. A non-reversible lifting may then result, in the best case, in a diffusive-to-ballistic speedup compared to the original (collapsed) reversible Markov chain [80, 112].

For a motivating example of the lifting framework, consider the reversible nearest-neighbor random walk on a one-dimensional lattice of length $n \geq 2$, $\Omega = \{1, 2, \ldots, n\}$. One possible extended lifted configuration space $\widehat{\Omega} = \Omega \times \mathcal{D}$ contains the set of possible directions $\mathcal{D} = \{+1, -1\}$ of the movement of the random walker as lifting variables. Given a lifted configuration $(x, d) \in \widehat{\Omega}$ (where $x \in \Omega$ is not at one of the ends of the lattice), the entries of $\widehat{P}$ are chosen so that the random walker moves to $(x + d, d)$ with probability $1 - 1/n$, or to $(x + d, -d)$ with probability $1/n$. This implies that the random walker moves *deterministically* from x to $x + d$, just the change of the lifting variable is stochastic (see Fig. 1.3 for a visual representation of the transfer matrices of the reversible nearest-neighbor random walk and its non-reversible lifted version that also illustrates what should happen at the ends of Ω). The reversible random walk takes $O(n^2)$ Monte Carlo moves to converge to a uniform distribution on Ω because it takes $O(k^2)$ Monte Carlo moves to travel a distance of $O(k)$ with its diffusive motion. In contrast, it can be shown that the non-reversible lifted random walk only takes $O(n)$ Monte Carlo moves to converge to a uniform stationary distribution on $\widehat{\Omega}$ (and thus also on Ω) [111]. Such a diffusive-to-ballistic speedup where the the mixing time of the original reversible Markov chain is reduced to its square root in the non-reversible lifted Markov chain is, in fact, the largest possible improvement that can be reached within the lifting framework [80, 111, 112]. Rigorous results in tractable test problems (either mathematically proven or obtained numerically) highlight that there are many inequivalent liftings to a lifted configuration space $\widehat{\Omega}$ and many inequivalent choices for the lifted transition matrix $\widehat{P}$ on $\widehat{\Omega}$. Finding the correct choices is essential but then often allows to come close to the optimal diffusive-to-ballistic speedup (see Ref. [80, Sections 3 and 4] and references therein; see also Ref. [114] for a very recent result from this year).

Systematic schemes that use the lifting framework to generate non-reversible Markov chains with considerable speedups in applications of practical relevance are nowadays available [80, 109, 113, 115–117]. (The lifting framework is, however, not the only possibility to generate non-reversible Markov chains with one alternative approach being introduced in Ref. [118].) As most recent examples from only last year, lifted directed-worm algorithms [119] and lifted self-avoiding walks [120] were shown to be much more efficient than their underlying reversible algorithms. This doctoral thesis is in particular concerned with the local *event-chain Monte Carlo* (ECMC) algorithm [80, 115, 121]. This is a non-reversible lifted MCMC algorithm for continuous configuration spaces that implements a rejection-free continuous-time Markov chain in an event-driven manner. Different variants of ECMC have been advantageously applied to, e.g., soft disks [122], hard spheres [123,
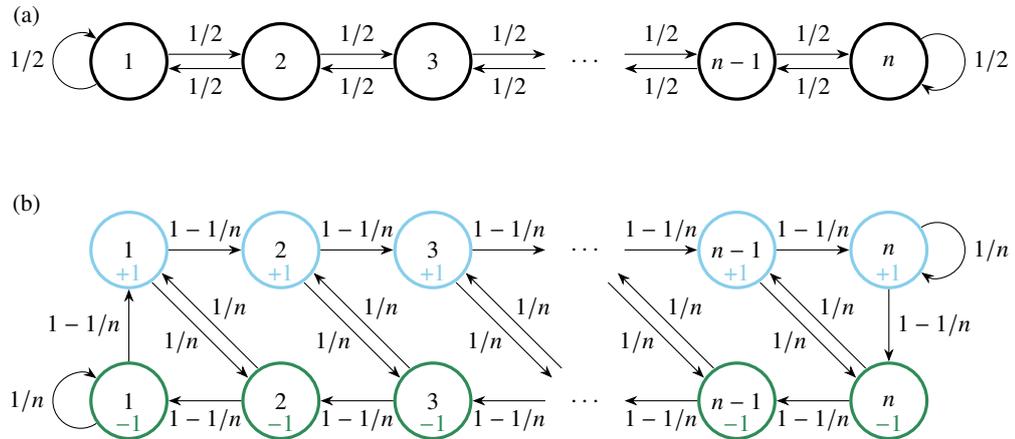
Figure 1.3: In the nearest-neighbor random walk a one-dimensional lattice $\Omega = \{1, 2, 3, \ldots, n-1, n\}$, the random walker starts on a random lattice point and the target distribution is uniform on Ω. (a): The transition matrix of the reversible random walk, whose non-zero entries are indicated as edges of the path graph, connects neighbored lattice sites. (b): The lifting framework introduces two copies "+1" and "−1" of every configuration to create a lifted configuration space $\widehat{\Omega} = \Omega \times \{+1, -1\}$. The non-zero entries of the lifted transition matrix of the non-reversible lifted random walk are indicated as edges of the path graph and depend on $1/n$. The non-reversible lifted Markov chain converges to a stationary uniform distribution on $\widehat{\Omega}$ [111].

124], continuous spin systems [125–128], dense soft-matter systems such as polymers [129–132], field-theoretical models [133], polyhedral particles [134], and, in a kinetic Monte Carlo adaption, to active matter [135].[5] ECMC can also be interpreted as piecewise-deterministic Markov processes [136, 137] that attract much attention in applied probability, statistics, and machine learning because of their superior mixing and correlation times (see, e.g., Refs [138–141]).

The very first proof for a spectacular speedup of the ECMC algorithm compared to local reversible MCMC algorithms was obtained in the hard-disk model [142, 143], which subsequently stimulated the broad range of successful applications mentioned in the previous paragraph. The hard-disk model is the simplest of all particle models in statistical physics and was among the first applications of the Metropolis algorithm [72] and (event-driven) MD [144]. The latter simulations led to the surprising insight that, although the hard-disk model lacks any attractive interactions that pulls the disks together, it shows a phase transition from a disordered fluid to a solid. The precise melting scenario, however, remained controversial for another fifty years until the superior mixing and correlation times of ECMC allowed to finally settle the debate in 2011 [142]. In Ref. [P5], we discuss a series of historic hard-disk simulations and argue that the phase-transition controversy resulted from the limited predictive power of simulations below the mixing-time scale. Only beyond the mixing time, which depends on the sampling method, the initial configuration is forgotten and the target distribution is sampled with well-understood statistical errors (see Section 2.2). Historically, ECMC was the first sampling method to reach the regime beyond the mixing time in a reasonable amount of computation time, which allowed for the first definite statements about the phase transition (that were afterwards confirmed

---

[5] In this list, the results of this doctoral thesis are left out in order maintain narrative tension.

by a massively-parallel Metropolis algorithm on 1536 cores, and ~ 7-times longer event-driven MD simulations [143]).

Most of the essential ideas of ECMC are already apparent in the hard-disk model. The potential energy $U(\mathbf{x})$ of a hard-disk configuration $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ of N two-dimensional hard disks of radius σ in a periodic box is given by

$$U(\mathbf{x}) = \sum_{i=1}^{N} \sum_{j=1}^{i-1} U_{ij}^{\text{disk}}(|\mathbf{r}_{ij}|), \tag{1.6}$$

where $\mathbf{r}_{ij} = \mathbf{x}_j - \mathbf{x}_i$ is the shortest separation vector from disk i to disk j that is possibly corrected for periodic boundary conditions, and U_{ij}^{disk} is the repulsive distance-dependent pair potential between them:

$$U_{ij}^{\text{disk}}(r) = \begin{cases} 0 & \text{if } r \geq 2\sigma, \\ \infty & \text{if } r < 2\sigma. \end{cases} \tag{1.7}$$

The hard-disk pair potential in Eq. (1.7) is infinite when the disks overlap or zero otherwise. In consequence, the Boltzmann probability $\pi(\mathbf{x})$ has the same finite value for every configuration $\mathbf{x}$ that does not contain any overlaps and vanishes otherwise. The most efficient *straight* ECMC variant for the hard-disk system extends its configuration space Ω to the lifted configuration space $\widehat{\Omega} = \Omega \times \mathcal{V} \times \mathcal{N}$, where $\mathcal{N} = \{1, 2, \ldots, N\}$ is the set of disk indices, and $\mathcal{V} = \{(1, 0)^T, (0, 1)^T\}$ is a set of velocities. Given a lifted configuration $(\mathbf{x}, \mathbf{v}_a, i) \in \widehat{\Omega}$, the "active" disk i moves in continuous Monte Carlo time with velocity $\mathbf{v}_a$ until it collides with another target disk j in an event at time t_{ev} and configuration $\mathbf{x}' = (\mathbf{x}_1, \ldots, \mathbf{x}_i + t_{\text{ev}} \mathbf{v}_a, \ldots, \mathbf{x}_N)$. The event time t_{ev} is the minimum of the set of (possibly infinite) collision times of the active disk i with all other disks k (and can, in practice, be efficiently computed with cell-occupancy systems by only considering a local neighborhood of disk i). At t_{ev}, a continued movement of i with $\mathbf{v}_a$ would introduce an overlap into the configuration and must be rejected. This rejection is replaced by a lifting move that changes the lifted configuration from $(\mathbf{x}', \mathbf{v}_a, i)$ to $(\mathbf{x}', \mathbf{v}_a, j)$ so that the target disk j continuous to move with exactly the same velocity (see Fig. 1.4). With resamplings of the velocity $\mathbf{v}_a \in \mathcal{V}$ in periodic intervals of the chain time τ_{chain}, these non-equilibrium dynamics sample the equal-weight Boltzmann distribution of the hard-disk model. In between resamplings of the velocity $\mathbf{v}_a$, it remains fixed so that a chain of disks moves in the same direction. The "velocity" does not have a kinematic meaning as in MD.

After the straight ECMC variant proved its power in the hard-disk system, it was generalized to smooth interaction potentials so that it can be used for sampling the canonical ensemble of three-dimensional molecular systems with their molecular-mechanics potential-energy functions $U = \sum_M U_M$ [121, 145, 146]. The generalization exploits the translational invariance of the factor potentials U_M, and every factor M may depend on an arbitrary number of interaction sites. Similar to the hard-disk system with periodic boundary conditions, the lifted configuration space is given by $\widehat{\Omega} = \Omega \times \mathcal{V} \times \mathcal{N}$ where the velocity set $\mathcal{V}$ now contains three velocities that are aligned with the coordinate axes, respectively. As before, given the lifted configuration $(\mathbf{x}, \mathbf{v}_a, i) \in \widehat{\Omega}$, the active interaction site moves with its velocity $\mathbf{v}_a$ in a deterministic straight-line trajectory until it is interrupted in an event at time t_{ev}. Here, a

$$t = 0 \qquad t = 0.19\,\tau_{\text{chain}} \qquad t = 0.37\,\tau_{\text{chain}} \qquad t = 0.42\,\tau_{\text{chain}} \qquad t = \tau_{\text{chain}}$$
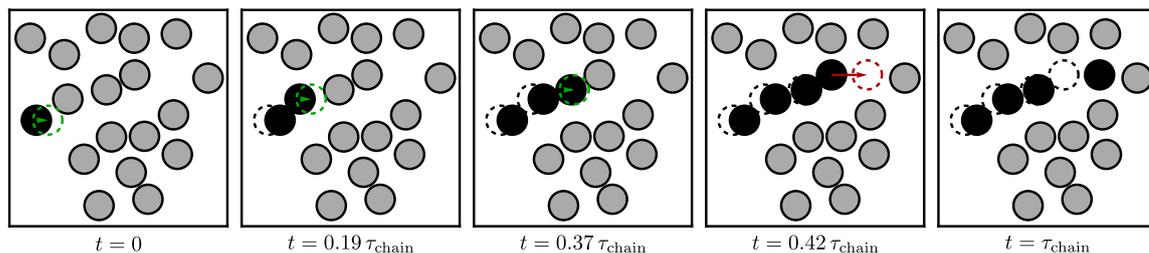
Figure 1.4: The straight ECMC algorithm for the hard-disk system continuously moves the highlighted active disk with constant velocity until it collides with another target disk in an event. Afterwards, the target disk continues to move with exactly the same velocity. At the chain time τ_{chain}, the currently active disk stops and ends the event chain. With velocity resamplings after each event chain, this non-equilibrium dynamics exactly samples the thermodynamic equilibrium distribution.

key ingredient for the generalized computation of t_{ev} is expressing the Boltzmann weight $\exp(-\beta U)$ as a factorized product $\prod_M \exp(-\beta U_M)$ of statistically independent factors M. Every factor M *stochastically* generates a candidate event time when it would interrupt the movement of the active interaction site. The factor M_{ev} with the minimum time actually triggers an event. This generalizes the comparison of the *deterministic* collision times of the hard-disk pair-potential factors. At the computed event time t_{ev}, the previously active interaction stops and another interaction site of M_{ev} becomes active with the same velocity instead. The events of the non-reversible piecewise-deterministic Markov process can be computed, as before, in an event-driven manner without discretizing time (see Chapter 2.4 for a theoretical introduction of ECMC).

Because of the factorization of the Boltzmann distribution and the statistical independence of the factors, ECMC never evaluates the total potential energy U but it still allows to sample the stationary Boltzmann distribution. Furthermore, the factorization fundamentally changes the treatment of long-range interactions. In MD, the force on every interaction site depends on the position of all other interaction sites. The computational cost of the tedious force-computation is reduced by introducing cutoffs and discretizing reciprocal space in fast mesh-based Ewald methods so that the result is only correct up to a limited accuracy that is typically much worse than allowed by machine precision. (Moreover, for convenience, MD computer packages such as Lammps usually allow to specify a target accuracy instead of the intrinsic parameters of the fast mesh-based Ewald methods such as the cutoffs and discretization size. The intrinsic parameters are then derived from the target accuracy with analytic error estimates for homogeneously-random charge distributions that should rather be carefully supplemented by additional force-accuracy tests in strongly inhomogeneous systems [96, 147].) Likewise, the computation of the potential-energy change under a local move of an interaction site in local reversible MCMC algorithms is an inaccurate global computation that considers the positions of all interaction sites. In ECMC, in contrast, every statistically independent factor only considers the long-range interactions between a small number of interaction sites that stays constant with increasing system size. For the SPC/Fw water model, for instance, there could be one factor for every Lennard-Jones potential between any two oxygens, and one molecular factor for the Coulomb potentials between any two molecules (as in the visual representation of the intermolecular interactions in Fig. 1.1). The long-range interaction between this small number of interaction sites can be computed

with small computational cost without any approximations in a high-accuracy Ewald summation (where the intrinsic parameters can be tuned to reach machine precision without any assumptions on the global charge distribution) [146].

Although the number of interaction sites in the factors remains constant with increasing system size, the number of factors does certainly not. If a single interaction site moves in a system of N SPC/Fw water molecules, $O(N)$ intermolecular factors [and $O(1)$ intramolecular ones] are changing in time and should stochastically generate candidate event times. The cell-veto algorithm [148] allows to bundle $O(N)$ long-range factors together and generates a single candidate event time for its set of factors. If that candidate event becomes a real event because it is the minimum of all candidate event times, the cell-veto algorithm further samples a single event-triggering factor among its bundled set of factors. All of this can be achieved with a constant computational complexity $O(1)$ that is independent of the system size. If the Coulomb and Lennard-Jones interactions in the SPC/Fw water model are each treated with the cell-veto algorithm, every event can be processed in constant computer time while the long-range interactions are considered without any approximations (and this is, in fact, numerically confirmed for large N in the last research paper of this doctoral thesis; see Chapter 6 and Ref. [P4]). Additionally, theory predicts that the mean-free path between events of the molecular Coulomb factors decreases as $O(1/\log N)$, while it remains constant for the Lennard-Jones and intramolecular factors [146]. Therefore, every interaction site is advanced by a constant $O(1)$ distance with a computational complexity of $O(N \log N)$. This matches the complexity of fast mesh-based Ewald methods in MD and the recent variant of the fast multipole method in the reversible Metropolis algorithm. ECMC, however, relies on a high-accuracy Ewald summation from the beginning and comes without a slowdown as the accuracy is increased.

Apparently, non-reversible ECMC algorithms (or more general, non-reversible piecewise-deterministic Markov processes) are optimal candidates to become the gold standard of molecular simulations in thermodynamic equilibrium. As MCMC algorithms, they allow to rigorously sample the Boltzmann distribution without thermostats. Moreover, they may solve both problems of local reversible MCMC algorithms in long-range-interacting molecular systems. The slow diffusive dynamics is, in the best case, replaced by fast ballistic dynamics, and the inefficient and inaccurate treatment of long-range interactions is replaced by an efficient and exact treatment. This motivated the first version of general-purpose open-source ECMC application JɛLLʏFʏsʜ, which we implemented shortly before the work for this doctoral thesis begun [P6]. The first project of this doctoral thesis was then supposed to consider the SPC/Fw water model as a proof of concept. Water molecules appear in many biomolecular simulations as an explicit aqueous solution and already contain the long-range interactions whose treatment is the limiting factor in molecular simulations with any computational method. Simulating large-scale water systems in JɛLLʏFʏsʜ is therefore an essential first step for its establishment for molecular simulations. Specifically, we chose the integrated autocorrelation time of the electric polarization [48] (or total electric dipole moment) as a benchmark. It is coupled to the rotation dynamics of the individual water molecules and determines the dielectric properties of the system. Because the straight ECMC variant moves all interaction sites in the same direction in between resamplings of the velocity, the mode of the polarization is expected equilibrate slowly. Unfortunately, it did indeed equilibrate slowly but much slower than expected. *In the straight ECMC*

*variant of JᴇLLʏFʏsʜ-Version*1.0*, the simulated water molecules resisted rotation and the polarization remained dynamically arrested for prohibitively long computation times* (see Ref. [149, Chapter 6] for a doctoral thesis of a past collaborator that contains these unfortunate results). The study of large systems was not possible and defeat had to be accepted.

1.4 Evolving Solution of this Doctoral Thesis

The four research papers of this doctoral thesis [P1, P2, P3, P4] document how we developed a solution for the problem of the dynamical arrest in straight ECMC simulations of the thermodynamic equilibrium of long-range-interacting SPC/Fw water molecules in a periodic cubic box. JᴇLLʏFʏsʜ-Version1.0 always aligned the three-dimensional velocity of the active interaction site with the positive coordinate axes, that is, it chose the finite velocity space $\mathcal{V} = \{(1,0,0)^T, (0,1,0)^T, (0,0,1)^T\}$. The analogous choice is superior in the periodic two-dimensional hard-disk system [115, 142]. (In Fig. 1.4, such an ECMC simulation only considers event chains going right and up. Chains going left and down are not necessary in periodic boundary conditions.) The finite velocity space $\mathcal{V}$ was also essential for an efficient treatment of the long-range interactions because, at that time, the cell-veto algorithm was only formulated for finite $\mathcal{V}$ (see Section 2.4.6). As a side effect, the alignment of the velocities with the coordinate axes reflects basic symmetry axes of the cubic simulation box which allowed for a much easier, and thus more performant, implementation in JᴇLLʏFʏsʜ-Version1.0.

In order to analyze and overcome the dynamical arrest of the rotation dynamics of straight ECMC, we considered simpler models instead of the complex SPC/Fw water model in the first three research papers of this doctoral thesis. On the one hand, this allowed for analytical approaches. On the other hand, this largely reduced the required implementation efforts and computing requirements. Concerning the overarching objective of this doctoral thesis—the introduction of a competitive, rigorous paradigm for molecular simulation that is based on exact non-reversible MCMC algorithms— the following insights evolved: First, analyzing a single, analytically tractable tethered hard-disk dipole, which serves as a simplified two-dimensional model of an extended dipole with flexible bond length that resembles the SPC/Fw water molecule, suggested that the particular straight ECMC variant of JᴇLLʏFʏsʜ-Version1.0 was particularly ill-suited for rotating the dipole (that is, for decorrelating the polarization). Larger finite velocity spaces $\mathcal{V}$, or even a continuous velocity space, are preferable (see Chapter 3 and Ref. [P1]).

Second, even though the precise choice of the proposal distribution for a Monte Carlo move of a single interaction site does typically not influence the performance of local reversible MCMC algorithms by a lot, the analogous case is *not* true for non-reversible MCMC algorithms. For an increasing number of tethered hard-disk dipoles at increasing densities, we considered the decorrelation of the polarization, which again characterizes the rotation dynamics of the individual dipoles, in the original straight ECMC variant and in reflective [115], forward [150], and Newtonian ECMC [124], which each change both the active interaction site and the velocity in an event (see Fig. 1.5). All ECMC variants similarly move a single interaction site with a constant velocity between events. Dipole rotations are thus pieced together from subsequent displacements of single disks. Currently, there is no insight on how explicit continuous-time rotations of dipoles or molecules could be generally included
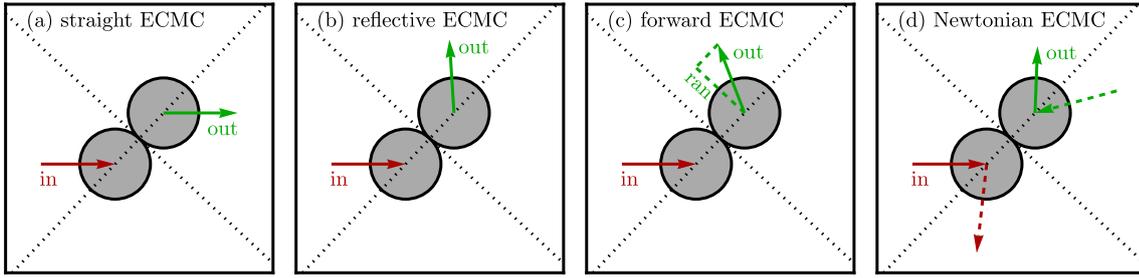
Figure 1.5: Events of ECMC in the two-dimensional hard-disk system correspond to collisions of the active disk with another target disk. They interrupt the motion of the incoming active disk ("in") and the outgoing target disk starts to move instead ("out"). (a): In straight ECMC [115], the outgoing target disk starts to move with exactly the same velocity as the incoming active disk. (b): Reflective ECMC [115] reflects the incoming velocity by the line connecting the active and target disk to obtain the velocity of the outgoing target disk. (c): The outgoing velocity of forward ECMC [150] is located, in comparison to reflective ECMC, in the same quadrant of the coordinate system based on axes parallel and orthogonal to the line connecting the active and target disk. Forward ECMC introduces an event-based randomness where the orthogonal component of the outgoing velocity is randomly sampled while the parallel component is chosen to keep the absolute value of the velocity constant. (d): Newtonian ECMC [124] assigns a hypothetical velocity to every disk in the system but only one of them actually moves at any time. An event changes both velocities of the active and target disks according to an elastic Newtonian collision between two disks of equal mass, and the target disk continuous to move with its outgoing velocity.

in the framework of ECMC, especially in the presence of long-range interactions in the ultimate application to molecular systems. Adding reversible rotation moves based on the Metropolis algorithm is, in principle, possible [134] but would require an inefficient and inaccurate computation of the potential-energy change in molecular systems which is thus also not an option for the overarching objective of this doctoral thesis. Fortunately, the comparison between the different ECMC variants reveal an up to an order-of-magnitude spread in performance between the worst variant, straight ECMC, and by far the best one, Newtonian ECMC (see Chapter 4 and Ref. [P2]). Even though Newtonian ECMC only relies on straight-line motions of individual interaction sites exactly as straight ECMC, it appears to allow for much faster rotation dynamics.

The dependence of the performance of ECMC on details of the algorithm was even more extreme in escape times from tightly confined hard-disk configurations (see Chapter 5 and Ref. [P3]). A scaling theory suggests, as the third insight, that finding the optimal choice of an intrinsic scale in straight ECMC (by fixing the chain time τ_{chain} after which the velocity is resampled; see Fig. 1.4) does not only require laborious fine-tuning. The optimal choice may also strongly differ in different parts of the configuration space. For the rotation dynamics of hard-disk dipoles and the escape dynamics from tightly confined hard-disk configurations, reflective, forward, and Newtonian ECMC become optimal when an intrinsic scale is missing (i.e., $\tau_{\text{chain}} \to \infty$). All of the first three research papers of this doctoral thesis also led to scientific results that go beyond the overarching objective. These are highlighted in more detail in the respective prologue of the chapters.

The first three research papers of this doctoral thesis systematically evaluated the manifold of possible choices within the framework of ECMC, which was previously not fully acknowledged for

molecular simulations. Newtonian ECMC, in particular, emerged as the most promising ECMC variant for molecular simulation because of its superior performance for the decorrelation of the polarization in a dense system of tethered hard-disk dipoles, and because it becomes most effective without any intrinsic parameter that needs to be fine-tuned. The fourth research paper of this doctoral thesis first generalizes the Newtonian ECMC variant from the peculiar hard-sphere system, for which it was proposed [124], to the general interaction potentials that appear in molecular simulations. It also lifts the restriction of the cell-veto algorithm to finite velocity spaces $\mathcal{V}$ so that it can be used for generalized Newtonian ECMC with its continuous velocity space $\mathcal{V}$. It is complemented by a majorly updated JeLLyFysh-Version2.0 that implements these theoretical ideas [P9]. We used JeLLyFysh-Version2.0 for large-scale simulations of the long-range-interacting SPC/Fw water model. Finally, the electric polarization was not dynamically arrested anymore and its integrated autocorrelation times could be determined. Here, the computationally demanding long-range Coulomb interaction is treated by the cell-veto algorithm which results in an observed constant computer time for every event. At the same time, it is treated without any approximations and the Boltzmann distribution is rigorously sampled. The performance of MD decreases as the desired accuracy for the Coulomb interaction is increased. At some desired accuracy, JeLLyFysh-Version2.0 can thus reach a break-even point with MD. By comparing it to the long-developed MD software Lammps, we were ultimately able to show that JeLLyFysh-Version2.0 reaches this break-even point for the long-range-interacting water system already well below machine precision (see Chapter 6 and Ref. [P4]). For such an exact treatment of the long-range interactions, one should thus, in fact, already now prefer JeLLyFysh-Version2.0.

The fact that generalized Newtonian ECMC in JeLLyFysh-Version2.0 becomes competitive to MD in Lammps at high precision in a long-range-interacting water system lays the foundation for a rigorous paradigm for molecular simulations that is based on general non-reversible MCMC algorithms. In that paradigm, the Boltzmann distribution is rigorously sampled without approximations, unphysical artifacts, and thermostats. Therefore, it can serve as a gold standard in critical applications where it identifies artifacts and approximations of MD that are not yet controlled algorithmically. Moreover, with its great potential for improvements, this paradigm may even become faster than MD in exploring the thermodynamic equilibrium of molecular systems (see Chapter 7).

Before the four research papers of this doctoral thesis are reproduced in Chapters 3–6 and an overall conclusion is drawn in Chapter 7, it is essential to introduce the theoretical foundations underlying non-reversible Markov chains in the context of ECMC in Chapter 2.

Theoretical Foundations of Non-Reversible Markov Chains

This chapter lays the theoretical foundation of non-reversible Markov chains in the context of the event-chain Monte Carlo algorithm that is necessary for the reproduced research papers in the following Chapters 3–6. It starts with a short introduction into general Markov-chain theory in Section 2.1. Afterwards, Section 2.2 establishes the most important figures of merits to measure the performance of Markov-chain Monte Carlo algorithms and also discusses necessary error-analysis techniques. Section 2.3 introduces lifted Markov chains before Section 2.4 presents the theory underlying the event-chain Monte Carlo algorithm.

2.1 Primer on Markov Chains

Consider, for concreteness, a finite configuration space Ω with a finite number of configurations $|\Omega|$ (a comment on Markov-chain theory in general configuration spaces that yields virtually unchanged results is given in Section 2.1.9). Markov-chain Monte Carlo (MCMC) algorithms are formulated for a target probability distribution π on Ω such as the Boltzmann distribution of the canonical ensemble [see Eq. (1.1)]. The aim of MCMC simulations is to generate a Markov chain in order to estimate expectations $\mathbb{E}_{\pi}(\mathcal{A})$ of bounded real-valued functions $\mathcal{A}$ on Ω:

$$\mathbb{E}_{\pi}(\mathcal{A}) = \sum_{x \in \Omega} \mathcal{A}(x)\, \pi(x). \tag{2.1}$$

Here, $\mathcal{A}(x)$ yields the physical observable of interest in the configuration $x \in \Omega$ as, for instance, the energy, and $\mathcal{A}$ should be bounded so that $\mathbb{E}_{\pi}(\mathcal{A}) < \infty$ remains finite. If π is the Boltzmann distribution, Eq. (2.1) yields thermodynamic averages. Naturally, MCMC algorithms are only relevant when the configuration space is too large to carry out the summation in Eq. (2.1) directly (as, e.g., in the famous Ising model of statistical physics where the configuration space grows exponentially fast as 2^L with the number of spins L). In continuous problems, the sum in Eq. (2.1) is replaced

by an integral over the continuous configuration space. For molecular systems, one obtains a very complicated high-dimensional integral whose direct integration (either analytic or numerical) is infeasible. Therefore, computational methods as, e.g., molecular dynamics or, in the case of this doctoral thesis, MCMC becomes necessary. Before one can understand how MCMC algorithms can be used to estimate the expectations $\mathbb{E}_\pi(\mathcal{A})$ for a given target distribution π, Markov chains and their properties have to be properly introduced.

2.1.1 Markov Property

A discrete-time (and time-homogeneous) Markov chain on Ω is a sequence of random variables $(X_0, X_1, X_2, \dots)$ where the probability $\mathbf{P}(X_0 = x_0, X_1 = x_1, X_2 = x_2, \dots, X_t = x_t)$ of any specific realized of configurations $\{x_0, x_1, x_2, \dots, x_t\}$ for any time $t \in \mathbb{N}_0$ can be expressed as [74]

$$\mathbf{P}(X_0 = x_0, X_1 = x_1, X_2 = x_2, \dots, X_t = x_t) = \pi^{\{0\}}(x_0)\, P(x_0, x_1)\, P(x_1, x_2)\, \dots\, P(x_{t-1}, x_t). \quad (2.2)$$

Here, $\pi^{\{0\}}$ is the probability distribution of the initial configuration as a row vector of size $1 \times |\Omega|$ (which could, e.g., just contain a single non-zero entry equal to one if a specific initial configuration $x_0 \in \Omega$ is of interest), and P is a time-independent transition matrix P of size $|\Omega| \times |\Omega|$ whose non-negative entries $P(x, y)$ give the conditional probability of moving from any configuration $x \in \Omega$ to any $y \in \Omega$ in the next step of the random process. This interpretation of the transition motivates an equivalent expression of Eq. (2.2),

$$\mathbf{P}(X_{t+1} = x_{t+1} | X_t = x_t, \dots, X_0 = x_0) = P(x_t, x_{t+1}), \quad (2.3)$$

that relates the entries of the transition matrix with the conditional probability of observing the configuration $X_{t+1} = x_{t+1} \in \Omega$ under the condition of a specific realization of previous configurations. Equation (2.3) is often called the *Markov property* and incorporates the "loss of memory" of Markov chains [74]. Because of the interpretation of its entries as conditional probabilities, the transition matrix P has to be stochastic, that is, $P(x, y) \geq 0$ for all $x, y \in \Omega$ and [74]

$$\sum_{y \in \Omega} P(x, y) = 1 \quad \text{for all } x \in \Omega. \quad (2.4)$$

2.1.2 Stationary Distribution and Global-Balance Condition

The transition matrix does not only connect two specific configurations but also the probability distributions $\pi^{\{t\}}$ and $\pi^{\{t+1\}}$ of configurations at succeeding time steps as $\pi^{\{t+1\}} = \pi^{\{t\}} P$ (where a matrix product is used). For any initial distribution $\pi^{\{0\}}$, this yields $\pi^{\{t\}} = \pi^{\{0\}} P^t$. A *stationary distribution* π of the Markov chain is a fixed point satisfying $\pi = \pi P$ which is often expressed element wise as [74]

$$\pi(y) = \sum_{x \in \Omega} \pi(x)\, P(x, y) \quad \text{for all } y \in \Omega. \quad (2.5)$$

In MCMC algorithms, the transition matrix is chosen so that the given target distribution π coincides with the stationary distribution of the Markov chain. Equation (2.5), which reproduces Eq. (1.2) in the

introduction, is then known as the *global-balance condition* for the transition matrix P. As already mentioned in Section 1.3.1 of the introduction, *reversible* Markov chains satisfy the more restrictive detailed-balance condition

$$\pi(x)\, P(x, y) = \pi(y)\, P(y, x) \quad \text{for all } x, y \in \Omega, \tag{2.6}$$

which implies that the global-balance condition is satisfied [74]. *Non-reversible* Markov chains break the detailed-balance condition and only satisfy Eq. (2.5).

Since the initial distribution $\pi^{\{0\}}$ is almost certainly not equal to the stationary distribution π (because this would require another way to sample π), one has to mathematically ensure that for a given Markov chain on Ω with transfer matrix P and initial distribution $\pi^{\{0\}}$, first, a unique π even exists and, second, that $\pi^{\{0\}}$ converges to π. This imposes further conditions on the Markov chain.

2.1.3 Irreducibility, Aperiodicity, and Positive Recurrence

A Markov chain is *irreducible* if, for any two configurations $x, y \in \Omega$, there exists a time t such that $P^t(x, y) > 0$ (where t may depend on x and y) [74]. This expresses the possibility to get from any configuration x to any other configuration y in the random process.

The *period* of a configuration $x \in \Omega$ is the greatest common divisor of the set $\mathcal{T}(x) = \{t \geq 1 : P^t(x, x) > 0\}$, that is, the set of times t at which it is possible for the chain to return to the starting configuration x [74]. In an irreducible Markov chain, all configurations share the same period [74, Lemma 1.6] and the chain is called *aperiodic* if their period is one. Aperiodicity prevents that the configuration space Ω can be partitioned into several classes [74, Exercise 1.6]. Consider, as an example of such a partitioning, the simple random walk on a discrete circle with an even number n of sites. The configuration space is given by $\Omega = \{1, 2, 3, \ldots, n - 1, n\}$. The Markov chain considers a random walker that starts on one site and follows a transition matrix defined by the procedure that, at each time step, the walker moves either to the left or the right with probability $1/2$ (where sites 1 and n are connected by periodic boundary conditions). This chain has a period of 2 and the configuration space can be divided into two distinct classes of even and odd sites. The Markov chain only makes transitions between configurations in complementary classes but never within a class. Such a partitioning of the configuration space clearly prevents a general convergence to a stationary distribution π (although it can still exist and even be reached if, for example, the Markov chain already starts with $\pi^{\{0\}} = \pi$).

Suppose a Markov chain starts at $X_0 = x \in \Omega$ (that is, $\pi^{\{0\}}(y) = \delta_{x,y}$ for all $y \in \Omega$). The *first return time* is defined as $\tau_x^+ = \min\{t \geq 1 : X_t = x\}$, i.e., the first time at which the chain revisits the configuration x after $X_0 = x$. A configuration $x \in \Omega$ is *positive recurrent* if the expectation of the return time remains finite: $\mathbb{E}(\tau_x^+) < \infty$ [74]. If a single configuration of an irreducible Markov chain is positive recurrent, all configurations are positive recurrent and, hence, the entire chain is classified as such [74, Propositions 21.11]. Any irreducible Markov chain on a finite configuration space Ω is positive recurrent [74, Lemma 1.13]. This property is of importance for (countably and uncountably) infinite configuration spaces (see Section 2.1.9).

2.1.4 Existence of a Unique Stationary Distribution and Convergence

For finite configuration spaces Ω, it can be shown that a unique stationary distribution π exists for irreducible Markov chains [74, Corollary 1.17]. In order for the Markov chain to converge to π, it must also be aperiodic [74, Theorem 4.9]. Here, the *total variation distance* is one possible distance measure between the evolving distribution $\pi^{\{t\}} = \pi^{\{0\}} P^t$ of the Markov chain and the stationary distribution π. For any two probability distributions μ and ν on Ω, it is defined as [74]

$$||\mu - \nu||_{\text{TVD}} := \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|, \tag{2.7}$$

where the probability of an event $A \subseteq \Omega$ under the given probability distribution is given by

$$\mu(A) := \sum_{x \in A} \mu(x). \tag{2.8}$$

Equation (2.7) uses an intuitive probabilistic definition of the distance between two probability distributions by considering the single event to which they assign the most different probability. The total variation distance can be shown to reduce to [74, Proposition 4.2]

$$||\mu - \nu||_{\text{TVD}} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \tag{2.9}$$

Since Eq. (2.9) only considers a sum over the configuration space Ω and, compared to Eq. (2.7), not the full power set of Ω (which grows exponentially fast as $2^{|\Omega|}$), it becomes more approachable. This enables a numerical computation in the first research paper of this doctoral thesis for a comparably small (discretized version of a continuous) configuration space (see Chapter 3).

The definition of the total variation distance allows us to formulate a *convergence theorem*. Let $\mathcal{P}$ denote the collection of all probability distributions on Ω. For an irreducible and aperiodic Markov chain with transition matrix P, there exist finite constants $C > 0$ and $\alpha \in (0, 1)$ such that [74, Theorem 4.9 and Exercise 4.1]

$$\sup_{\pi^{\{0\}} \in \mathcal{P}} ||\pi^{\{0\}} P^t - \pi||_{\text{TVD}} \leq C\alpha^t. \tag{2.10}$$

This expresses that the total variation distance between the evolving distribution $\pi^{\{t\}}$ of the Markov chain and its stationary distribution π remains below an exponential bound for *any* initial distribution $\pi^{\{0\}}$. Consequently, the distance between $\pi^{\{t\}}$ and π becomes negligible at some point in time t (see Section 2.2). In practice, most Markov chains start from a specific configuration $X_0 = x \in \Omega$. This implies $\pi^{\{0\}}(y) = \delta_{x,y}$ for all $y \in \Omega$ and thus $\pi^{\{t\}} = P^t(x, \cdot)$ where $P^t(x, \cdot)$ is the x-th row of P. It can be shown that only these particular initial probability distributions are of importance in the convergence theorem because the left-hand side of Eq. (2.11) can be rewritten to only consider all initial configurations $X_0 = x \in \Omega$ [74, Exercise 4.1] which yields

$$\max_{x \in \Omega} ||P^t(x, \cdot) - \pi||_{\text{TVD}} \leq C\alpha^t. \tag{2.11}$$

This is important in the first research paper of this doctoral thesis because Eq. (2.11) is, again, more approachable from a numerical point of view than Eq. (2.10) (see Chapter 3).

2.1.5 Ergodic Theorem

The ergodic theorem allows to replace the configuration-space average in Eq. (2.1) by a time average over the Markov chain $(X_0, X_1, \ldots, X_t)$ of length $n = t + 1 \in \mathbb{N}$ that is introduced as

$$\overline{\mathcal{A}}_n := \frac{1}{n} \sum_{s=0}^{n-1} \mathcal{A}(X_s). \tag{2.12}$$

For an irreducible Markov chain with stationary distribution π starting from any initial distribution $\pi^{\{0\}}$, one can deduce from the strong law of large numbers that the time average $\overline{\mathcal{A}}_n$ converges almost surely to the space average $\mathbb{E}_\pi(\mathcal{A})$ with increasing n [74, Theorem C.1]:

$$\mathbf{P}\left(\lim_{n \to \infty} \overline{\mathcal{A}}_n = \mathbb{E}_\pi(\mathcal{A})\right) = 1. \tag{2.13}$$

Equation (2.13) is the *ergodic theorem* (or strong law of large numbers) for Markov chains. An equivalent statement is that the asymptotic proportion of time that the Markov chain spends in the configuration x is $\pi(x)$. This allows to estimate intractable configuration-space averages with an MCMC simulation.

2.1.6 Central Limit Theorem

The central limit theorem for Markov chains is an even stronger statement than the ergodic theorem and is reminiscent of the classic central limit theorem for independent and identically distributed random variables. Consider an irreducible Markov chain $(X_0, X_1, \ldots X_t)$ of length $n = t + 1 \in \mathbb{N}$ that starts from its stationary distribution $\pi^{\{0\}} = \pi$ (for example, because it is also aperiodic and already converged to π beforehand). In stationarity, one knows that all X_t are identically distributed according to $\pi^{\{t\}} = \pi$. Thus, $\mathbb{E}_{\pi^{\{t\}}}[\mathcal{A}(X_t)] = \mathbb{E}_\pi(\mathcal{A})$ and the expectation of the time average over the Markov chain is just the configuration-space average:

$$\mathbb{E}_\pi(\overline{\mathcal{A}}_n) = \frac{1}{n} \sum_{s=0}^{n-1} \mathbb{E}_\pi[\mathcal{A}(X_s)] = \mathbb{E}_\pi(\mathcal{A}). \tag{2.14}$$

Here, the linearity of the expectation $\mathbb{E}_\pi$ was used. The ergodic theorem is, of course, a much stronger property than Eq. (2.14) because it does not rely on a stationary initial distribution and because it ensures that $\overline{\mathcal{A}}_n$ itself converges to $\mathbb{E}_\pi(\mathcal{A})$.

The variance of the time average $\sigma_\pi^2(\overline{\mathcal{A}}_n)$ contains correlation terms because the random variables X_t of the Markov chain are not independent. Therefore, the covariance is introduced as

$$\begin{aligned} \mathrm{cov}_\pi\left[\mathcal{A}(X_s), \mathcal{A}(X_r)\right] &:= \mathbb{E}_\pi\left[(\mathcal{A}(X_s) - \mathbb{E}_\pi[\mathcal{A}(X_s)])(\mathcal{A}(X_r) - \mathbb{E}_\pi[\mathcal{A}(X_r)])\right] \\ &= \mathbb{E}_\pi\left[\mathcal{A}(X_s)\mathcal{A}(X_r)\right] - \mathbb{E}_\pi[\mathcal{A}(X_s)]\mathbb{E}_\pi[\mathcal{A}(X_r)]. \end{aligned} \tag{2.15}$$

Because of stationarity, $\mathrm{cov}_\pi\left[\mathcal{A}(X_s), \mathcal{A}(X_{s+k})\right]$ does not depend on s but only on the time difference k. Hence, one introduces $\gamma_{\pi,k}(\mathcal{A}) := \mathrm{cov}_\pi\left[\mathcal{A}(X_s), \mathcal{A}(X_{s+k})\right]$. This allows to compute the variance of the time average $\overline{\mathcal{A}}_n$ multiplied by $\sqrt{n}$ [151]:

$$
\begin{aligned}
\sigma_\pi^2(\sqrt{n}\,\overline{\mathcal{A}}_n) &= n\,\sigma_\pi^2(\overline{\mathcal{A}}_n) \\
&= n\,\mathbb{E}_\pi(\overline{\mathcal{A}}_n^2) - n\left[\mathbb{E}_\pi(\overline{\mathcal{A}}_n)\right]^2 \\
&= \frac{1}{n}\sum_{s=0}^{n-1}\sum_{r=0}^{n-1}\mathrm{cov}_\pi\left[\mathcal{A}(X_s), \mathcal{A}(X_r)\right] \\
&= \gamma_{\pi,0}(\mathcal{A}) + 2\sum_{k=1}^{n-1}\frac{n-k}{n}\,\gamma_{\pi,k}(\mathcal{A}).
\end{aligned}
\tag{2.16}
$$

Equations (2.14) and (2.16) essentially compute the first and second moments of the distribution of $\overline{\mathcal{A}}_n$. The central limit theorem makes a statement for the entire distribution.

For an irreducible Markov chain starting from its stationary distribution so that $\pi^{\{0\}} = \pi$, the *central limit theorem* for Markov chains shows that the fluctuations of the time average $\overline{\mathcal{A}}_n$ around its limit $\mathbb{E}_\pi(\mathcal{A})$ when multiplied by $\sqrt{n}$ converge to a normal distribution $\mathcal{N}[0, \sigma_{\pi,\mathrm{CLT}}^2(\mathcal{A})]$ with mean 0 and variance $\sigma_{\pi,\mathrm{CLT}}^2(\mathcal{A})$ [81, Theorem 23 and Remark after Proposition 30]:

$$
\sqrt{n}\left(\overline{\mathcal{A}}_n - \mathbb{E}_\pi(\mathcal{A})\right) = \frac{1}{\sqrt{n}}\sum_{s=0}^{n-1}\left[\mathcal{A}(X_s) - \mathbb{E}_\pi(\mathcal{A})\right] \xrightarrow{d} \mathcal{N}[0, \sigma_{\pi,\mathrm{CLT}}^2(\mathcal{A})] \quad \text{for } n \to \infty.
\tag{2.17}
$$

This result is similar to the one of the classical central limit theorem that assumes that the random variables are independent and identically distributed. In a stationary Markov chain, they are only identically distributed but the Markov property still allows to derive Eq. (2.17). For large enough n, the distribution of $\overline{\mathcal{A}}_n$ still converges to a normal distribution with mean $\mathbb{E}_\pi(\mathcal{A})$ and variance $\sigma_{\pi,\mathrm{CLT}}^2(\mathcal{A})/n$. This implies that the standard error of the estimate $\overline{\mathcal{A}}_n$ decreases as $1/\sqrt{n}$. The variance $\sigma_{\pi,\mathrm{CLT}}^2(\mathcal{A})$ of the asymptotic normal distribution $\mathcal{N}[0, \sigma_{\pi,\mathrm{CLT}}^2(\mathcal{A})]$ is given by the limit of $\sigma_\pi^2(\sqrt{n}\,\overline{\mathcal{A}}_n)$ in Eq. (2.16) [151]:

$$
\sigma_{\pi,\mathrm{CLT}}^2(\mathcal{A}) = \lim_{n\to\infty}\sigma_\pi^2(\sqrt{n}\,\overline{\mathcal{A}}_n) = \gamma_{\pi,0}(\mathcal{A}) + 2\sum_{k=1}^{\infty}\gamma_{\pi,k}(\mathcal{A}).
\tag{2.18}
$$

For the preceding discussion of the central limit theorem for irreducible Markov chains, it was assumed that it already starts from its stationary distribution $\pi^{\{0\}} = \pi$ to connect the variance of the asymptotic normal distribution in Eq. (2.18) to the limit of the variance of the time average over the stationary Markov chain in Eq. (2.16). However, initial stationarity is not necessary and one can show that the asymptotic convergence in distribution of the central limit theorem (as the ergodic theorem in Section 2.1.5) holds for any initial distribution $\pi^{\{0\}}$ because it becomes irrelevant in the $n \to \infty$ limit [81, Proposition 29].

2.1.7 Necessity of Aperiodicity

The existence of a unique stationary distribution π and, based on this, the ergodic and central limit theorems for Markov chains only require their irreducibility. The convergence theorem, in contrast, also requires aperiodicity because a periodic Markov chain partitions the configuration space into distinct classes (see Section 2.1.3). In this case, a general statement about its *rate* of convergence to stationarity as in the convergence theorem is not possible because the evolving distribution $\pi^{\{t\}}$ may actually never converge to π. For the asymptotic results of the ergodic and central limit theorems, in contrast, one can, informally speaking, average over the distinct partitions of configuration space in the $n \to \infty$ limit [81, Corollary 6]. If one is only interested in estimating configuration-space averages by time averages over the Markov chain, aperiodicity is therefore not important (see also Ref. [75, Section 4.3]). Because aperiodicity is easy to establish in most practical applications [80], this doctoral thesis nevertheless considers irreducible and aperiodic MCMC algorithms so that the convergence theorem is valid.

2.1.8 Continuous-Time Markov Chains

A discrete-time Markov chain on configuration space Ω with transition matrix P and initial distribution $\pi^{\{0\}}$ considers a sequence of random variables $(X_0, X_1, X_2, \ldots, X_t)$ at times $t \in \mathbb{N}_0$. Continuous-time Markov chains use times $t \in \mathbb{R}_{\geq 0}$ instead. One may construct a continuous-time Markov chain from a discrete-time transfer matrix P by making transitions at a constant rate, or in other words, at time intervals that are independent and identically distributed exponential variables [74, Chapter 20]. The number of transition times per unit time is then a Poisson-distributed random variable which is why the transition times are also said to follow a Poisson process. The stationary distribution of an embedded irreducible discrete-time Markov chain can be connected to the stationary distribution of an irreducible continuous-time version by considering the average time between transitions [152, Theorem 4.29]. It can further be shown that the irreducible continuous-time Markov chain converges to its stationary distribution, and that a time-average over the Markov chain, practically implemented by taking samples in periodic time intervals, yields the corresponding configuration-space average in an ergodic theorem [152, Section 4.4.3]. This doctoral thesis considers the non-reversible event-chain Monte Carlo algorithm which generates piecewise-deterministic Markov processes in continuous time but they are constructed from an infinitesimal limit of a discrete-time Markov chain.

2.1.9 General Configuration Spaces

Previously, Markov chains on finite configuration spaces Ω were considered. For countably infinite configuration spaces Ω, one can still think of P as a transition matrix and of the evolving distributions $\pi^{\{t\}}$ for any time $t \in \mathbb{N}_0$ as a row vector. Both P and $\pi^{\{t\}}$ now contain countably many entries. The definitions of irreducibility, aperiodicity, and positive recurrence are unchanged (see Section 2.1.3) but irreducibility does not imply positive recurrence anymore (e.g., the unbiased random walk on $\mathbb{Z}$ starting at 0 is irreducible but not positive recurrent because the expectation value of the return time is infinite [74, Example 21.10]). A Markov chain now has to be irreducible and positive recurrent in order show the existence of a unique stationary distribution π [74, Theorem 21.13]. The reverse is

also true, that is, if there exists a unique stationary distribution π of an irreducible Markov chain, then it is also positive recurrent. Since positive recurrence can be difficult to prove and MCMC algorithms are designed so that the target distribution becomes the *known* stationary distribution, the existence of a unique stationary distribution π is often understood as another prerequisite instead [73, 75, 81]. The general formulation of Markov chains that applies to continuous, uncountably infinite configuration spaces is based on measure theory (see, e.g., Refs [73] or [75, Chapter 4]). The definitions of irreducibility and aperiodicity have to be properly generalized, and positive recurrence is replaced by an even stronger Harris recurrence [73].

The ergodic theorem (or strong law of large numbers) in Eq. (2.13) can be shown for irreducible and Harris recurrent Markov chains with stationary distribution π for any real-valued functions $\mathcal{A}$ such that $\mathbb{E}_\pi(|\mathcal{A}|)$ exists and is finite [75, Theorem 4.3 and Section 4.4]. Furthermore, any irreducible, Harris recurrent, and aperiodic Markov chain with stationary distribution π converges [75, Theorem 4.4 and Section 4.4]:

$$\lim_{t\to\infty} ||P^t(x, \cdot) - \pi||_{\text{TVD}} = 0 \quad \text{for all } x \in \Omega. \tag{2.19}$$

This equation does not make any statement about about the rate of convergence as the convergence theorem for finite Markov chains in Eq. (2.11). General irreducible, aperiodic, and positive Harris recurrent Markov chains with stationary distribution π may, but do not have to, satisfy Eq. (2.11) and if they do, so they are called uniformly ergodic [75, Definition 4.6]. A weaker bound on the convergence rate appears in geometrically ergodic Markov chains [75, Definition 4.5]. These stronger conditions are necessary for a generalized central limit theorem for Markov chains. Equation (2.17) holds for uniformly ergodic Markov chains with stationary distribution π for real-valued functions $\mathcal{A}$ such that $\mathbb{E}_\pi(\mathcal{A}^2) < \infty$ which is required so that $\sigma^2_{\pi,CLT}(\mathcal{A}) < \infty$ [75, Theorem 4.6]. It also holds for geometrically ergodic Markov chains with stationary distribution π if $\mathbb{E}_\pi(|\mathcal{A}|^{2+\varepsilon})$ for some $\varepsilon > 0$ [75, Theorem 4.6]. Here, as in the finite case, the condition on aperiodicity (that is hidden in the uniformly or geometrically ergodicity) is not strictly necessary [81, Proposition 30]. Actually proving that a given MCMC algorithm is geometrically or even uniformly ergodic can be quite difficult [75, Section 4.5]. For the non-reversible piecewise-deterministic Markov processes of this doctoral thesis, proofs under certain conditions on the target distribution are available [153–156]. In practice, one generally assumes that the central limit theorem holds which is justified by the good performance that one observes in the simulations. Alternatively, one may consider the easier-to-satisfy irreducibility and aperiodicity conditions of finite Markov chains on a discretized version of the continuous configuration space.

2.2 Figures of Merits of Markov-Chain Monte Carlo Algorithms

Two important figures of merits of MCMC algorithms with the stationary target distribution π that satisfy the conditions for the convergence, ergodic, and central limit theorems are the mixing time and the integrated autocorrelation time. The former determines the time after which the generated Markov chain is sufficiently close to the stationary distribution π (see Section 2.2.1). The latter determines how many Monte-Carlo steps are required to get an effectively independent sample in the Markov chain once it samples the stationary distribution π which paves the way to a rigorous error analysis (see Section 2.2.2). Both quantities appear repeatedly throughout this doctoral thesis.

2.2.1 Mixing Time

The *mixing time* t_{mix} measures the time that is required until the Markov chain can be guaranteed to be close to its stationary distribution π, independent of the initial distribution $\pi^{\{0\}}$. Let $d(t)$ denote the total variation distance between the evolving distribution $\pi^{\{t\}}$ from the most unfavorable initial distribution $\pi^{\{0\}}$ and π [74, Section 4.4]:

$$d(t) := \sup_{\pi^{\{0\}} \in \mathcal{P}} ||\pi^{\{0\}} P^t - \pi||_{\text{TVD}} = \max_{x \in \Omega} ||P^t(x, \cdot) - \pi||_{\text{TVD}}. \qquad (2.20)$$

This regards, again, a finite configuration space Ω and, as in the the rewriting of the convergence theorem in Eqs (2.10) and (2.11), the second equality only considers specific configurations $X_0 = x \in \Omega$ as the initial distribution of the Markov chain [74, Exercise 4.1]. The mixing time is defined as [74, Section 4.5]

$$t_{\text{mix}}(\varepsilon) := \min \{t : d(t) \leq \varepsilon\} \qquad (2.21)$$

and

$$t_{\text{mix}} := t_{\text{mix}}(1/4). \qquad (2.22)$$

The chosen value $\varepsilon = 1/4$ is arbitrary but should satisfy $0 < \varepsilon < 1/2$ to ensure that the entire configuration space is explored [80, Section 2.1.2]. Beyond the mixing time t_{mix}, one can show that the distance $d(t)$ is bounded by an exponential decay, that is, $d(\ell t_{\text{mix}}) \leq 2^{-\ell}$ where $\ell \in \mathbb{N}$ [74, Eq. (4.33)]. Therefore, the mixing time determines the point in time when a first sample from the stationary distribution π can be safely extracted from a Markov chain even when it was started from the most unfavorable initial configuration.

Upper bounds on the mixing time are necessary to be confident that a practical MCMC simulation yields rigorous results because they can only run for a finite number n of time steps. The ergodic and central limit theorems in principle require the unpractical $n \to \infty$ limit. If one knows, however, that the simulated number of time steps n is well above the mixing time, one can be sure that the time average over the Markov chain yields a well-behaved estimate of the configuration-space average with an $O(1/\sqrt{n})$ standard error (by using the techniques of the next Section 2.2.2). The $O(1/n)$ bias from the initial part of the trajectory where the Markov chain was not yet in equilibrium, in contrast, becomes insignificant or is, in practice, considered to be a "burn-in" part of the simulation and thus not used for any time averages [76, 157].

Rigorous results for mixing times are only available for tractable test problems (see, e.g., the discussion about mixing times of lifted Markov chains in Section 1.3.2). In the first research paper of this doctoral thesis, we consider a small enough (discretized version of a continuous) configuration space so that we can numerically compute $d(t)$ in Eq. (2.20) by considering all possible initial configurations which then yields the mixing time t_{mix} (see Chapter 3). In the third research paper of this doctoral thesis, we propose an especially bad initial configuration for the hard-disk system from which it takes local MCMC algorithms long times to converge to their stationary distribution. This initial configuration may be used to bound mixing times and to reflect the relative merits of the sampling algorithms (see Chapter 5). In the more complex systems of the second and fourth research papers of this doctoral thesis (see Chapters 4 and 6),

we make no attempt on estimating the mixing time and assume that the initial configuration is already a representative sample of the stationary distribution. The first part of any simulation is still considered to be a "burn-in" period where no data is collected and this period is chosen much larger than the estimated integrated autocorrelation times. In Ref. [P5], we consider two initial configurations of the hard-disk system that widely differ in an observable that is supposed to decorrelate most slowly. We then assume that the simulations reach stationarity once their estimated observables coalesce. In a different approach, we suppose that the simulation reaches the mixing-time scale once the slowest observable explored a wide part of its possible range.

2.2.2 Integrated Autocorrelation Time and Error Analysis

Once the Markov chain is stationary, it yields identically distributed but correlated random variables. For uncorrelated (that is, independent and identically distributed) random variables, the covariance at different times would vanish so that $\gamma_{\pi,k}(\mathcal{A}) = 0$ for all $k > 0$. This would imply that the variance of the asymptotic normal distribution in the central limit theorem for Markov chains in Eq. (2.18) becomes $\sigma^2_{\pi,\text{CLT}} = \gamma_{\pi,0}(\mathcal{A})$. This recovers the classic central limit theorem because $\gamma_{\pi,0}$ is just the variance $\sigma^2_\pi[\mathcal{A}(X_i)]$ of the uncorrelated random variables. The correlation in the random variables of a Markov chain leads to the fact that variance $\sigma^2_{\pi,\text{CLT}}(\mathcal{A})$ differs from $\gamma_{\pi,0}$. The *integrated autocorrelation time* $\tau_{\text{int}}(\mathcal{A})$ is thus introduced as (see, e.g., Refs [81, 151, 157])

$$\tau_{\text{int}}(\mathcal{A}) := \frac{\sigma^2_{\pi,\text{CLT}}(\mathcal{A})}{\gamma_{\pi,0}(\mathcal{A})} = 1 + 2 \sum_{k=1}^{\infty} \frac{\gamma_{\pi,k}(\mathcal{A})}{\gamma_{\pi,0}(\mathcal{A})}. \tag{2.23}$$

By comparison with the classic central limit theorem for uncorrelated random variables, one may say that the number of "effectively independent samples" in n correlated samples is roughly $n/\tau_{\text{int}}(\mathcal{A})$ [157]. This implies that $\tau_{\text{int}}(\mathcal{A})$ is a direct measure on how quickly the algorithm can decorrelate the function $\mathcal{A}$ once it samples the stationary distribution π. The definition of the integrated autocorrelation time based on the covariances $\gamma_{\pi,k}(\mathcal{A})$ on the right-hand side of Eq. (2.23) applies to any set of correlated data of a stationary process (that may have also been, for instance, generated from a molecular-dynamics simulation in thermodynamic equilibrium) [158, Section 5.3]. It is then still connected to the standard error of the sample mean but it does make no assumptions of the asymptotic distribution of $\overline{\mathcal{A}}_n$ as the central limit theorem for Markov chains. (It can be shown, however, that $\overline{\mathcal{A}}_n$ approaches a normal distribution in typical sufficiently "well-behaved" cases [158, Section 5.3.5].)

In order to assess the accuracy of the time average $\overline{\mathcal{A}}_n$ over a Markov chain with the help of the central limit theorem, one has to estimate $\sigma^2_{\pi,\text{CLT}}(\mathcal{A})$. In the following, different estimation techniques, which are used in the research papers of this doctoral thesis, are presented. Since $\sigma^2_{\pi,\text{CLT}}(\mathcal{A})$ is closely connected to $\tau_{\text{int}}(\mathcal{A})$, the same techniques may be used to estimate the integrated autocorrelation time, and thus to measure the efficiency of an MCMC algorithm once it samples its target stationary distribution.

2.2.2.1 Window Estimators

Given the definition of $\gamma_{\pi,k}(\mathcal{A})$ in terms of the covariance, a first idea may be to estimate $\gamma_{\pi,k}(\mathcal{A})$ from the Markov chain $(X_0, X_1, X_2, \ldots, X_t)$ of length $n = t + 1 \in \mathbb{N}$ by [157, 158]

$$\overline{\gamma}_{n,k} := \frac{1}{n-k} \sum_{s=0}^{n-k-1} \left[\mathcal{A}(X_s) - \overline{\mathcal{A}}_n \right] \left[\mathcal{A}(X_{s+k}) - \overline{\mathcal{A}}_n \right] \quad \text{for } k \in \{0, 1, \ldots, n-1\}. \tag{2.24}$$

This estimator is biased because $\mathbb{E}_\pi(\overline{\gamma}_{n,k}) \neq \gamma_{\pi,k}(\mathcal{A})$, however, the bias is only $O(1/n)$ [158, Section 5.3.3]. Very often, the divisor $n - k$ is replaced by n which increases the bias but also leads to advantages if one sums the estimates $\overline{\gamma}_{n,k}$ over k to obtain an estimate for $\sigma^2_{\pi,\mathrm{CLT}}(\mathcal{A})$ or $\tau_{\mathrm{int}}(\mathcal{A})$ [75, 151, 158]. However, independent of the choice of divisor, these estimators are inconsistent because they have a variance that does not go to zero as $n \to \infty$. This problem can be solved by using a *window estimator* instead, that is, for the example of estimating $\sigma^2_{\pi,\mathrm{CLT}}(\mathcal{A})$ [75, 151, 157]:

$$\overline{\sigma^2}_{n,\mathrm{CLT}} := \overline{\gamma}_{n,0} + 2 \sum_{k=1}^{m} \overline{\gamma}_{n,k}. \tag{2.25}$$

Here, $m \ll n$ is a cutoff that introduces an additional bias but decreases the variance. Good empirical choices that balance the trade-off between bias and variance are known and the window-estimator appears to work well if a sufficient amount of data is available, say, $n \gtrsim 1000\,\overline{\tau}_{n,\mathrm{int}}$ where $\overline{\tau}_{n,\mathrm{int}}$ is the estimate of the integrated autocorrelation time [157]. This window-estimator procedure is implemented in a general manner in the Python EMCEE package [159] and is used in the second research paper of this doctoral thesis to obtain integrated autocorrelation times (see Chapter 4).

2.2.2.2 Blocking Method

The window-estimator procedure relies on biased estimators and it cannot make statements about the errors of its estimates [157]. The *blocking method* yields rigorous estimates by a renormalization-group technique and also gives information about the quality of its estimates [160]. It relies on the fact that correlated samples in a Markov chain at large time differences are, in fact, uncorrelated. Therefore, if a sufficient number b of consecutive correlated samples are bunched into a single bunch average, the set of bunch averages is uncorrelated if b was chosen large enough. Because of this, their standard error can now be rigorously estimated without any bias. Since the bunching procedure does not change the standard error, this also yields an estimate for the initial correlated samples.

The blocking method systematically adapts the bunch sizes. It starts by (wrongly) assuming that all correlated samples are independent and estimates the resulting *apparent variance* $\overline{\sigma^2}_{n,\mathrm{CLT}}/n$ of $\overline{\mathcal{A}}_n$ by estimating $\overline{\sigma^2}_{n,\mathrm{CLT}} = \overline{\gamma}_{n,0}$ (where the biased estimator in Eq. (2.24) can be made unbiased for $k = 0$ under the assumption of uncorrelated samples by replacing the divisor n by $n - 1$ [78, Section 1.3.4]). Then, the set of samples is cut in half by substituting two adjacent samples with their average. The apparent variance is again estimated under the assumption that the reduced set of samples is uncorrelated. This procedure is repeated until the set of samples was reduced to two entries. The series of apparent variances approaches a constant value within fluctuations because

the sets of samples become increasingly uncorrelated, and this constant value is the final estimate on the real variance $\overline{\sigma^2}_{n,\text{CLT}}/n$ of $\overline{\mathcal{A}}_n$. Dividing the estimate of $\overline{\sigma^2}_{n,\text{CLT}}$ by the estimate $\overline{\gamma}_{n,0}$ also yields an estimate on the integrated autocorrelation time [see Eq. (2.23)]. In the blocking method, the set of apparent variances have a known error and if they do not reach a constant value within fluctuations, one knows that the original number of correlated samples n was not large enough. A general automated procedure for extracting the approached constant value of the apparent variances is, unfortunately, difficult to implement. However, because the blocking method is more rigorous and fail-safe than the window-estimator procedure, we check that the two methods give similar results in the second research paper of this doctoral thesis (see Chapter 4).

2.2.2.3 Exponential Autocorrelation Time

Often, $\gamma_{\pi,k}(\mathcal{A})$ has a simple exponential decay, i.e., $\gamma_{\pi,k}(\mathcal{A}) = \gamma_{\pi,0}(\mathcal{A}) \exp(-k/\tau_0)$ with $\tau_0 \gg 1$. The exponential autocorrelation time τ_0 is connected to the integrated autocorrelation time $\tau_{\text{int}}(\mathcal{A})$ by using a geometric series and a series expansion:

$$\tau_{\text{int}}(\mathcal{A}) = 1 + 2 \sum_{k=1}^{\infty} e^{-k/\tau_0} = 1 + \frac{2}{\exp(1/\tau_0) - 1} = \coth\left(\frac{1}{2\tau_0}\right) = 2\tau_0 + O\left(\frac{1}{\tau_0}\right). \tag{2.26}$$

In practice, an exponential function is fitted to the estimates $\overline{\gamma}_{n,k}$ in Eq. (2.24) to obtain τ_0 and τ_{int} *if an exponential decay is indeed observed*. This approach is considered in the fourth research paper of this doctoral thesis (see Section 6). Note that, more generally, the correlations of any function $\mathcal{A}$ decay exponentially at large times $t \to \infty$. For reversible Markov chains, this decay can be related to spectral measures of its transition matrix [74, 80].

2.2.2.4 Bootstrapping

Up to now, elaborate estimates for the standard error on the time average $\overline{\mathcal{A}}_n$ of correlated data were considered. For uncorrelated data, the classic central limit theorem yields a simpler estimate on the standard error based on the variance of the data. However, even for uncorrelated data, different approaches are necessary to estimate the standard errors and confidence intervals of other statistics than the average. For that purpose, *bootstrapping* is a powerful and automatic resampling technique [161]. It repeatedly resamples the single uncorrelated data set to create many bootstrap data sets and then carries out a rigorous error analysis on them. Adaptions of bootstrapping are also available for correlated data (see, e.g., Ref. [162] and references therein). In the third and fourth research papers of this doctoral thesis, we use bootstrapping to estimate the standard error of a median of uncorrelated data (see Chapters 5 and 6).

2.3 Lifted Markov Chains

The appendix of the first research paper of this doctoral thesis contains a collection of basic notions about lifted Markov chains [80, 111, 112] (see Appendix 3.A.1). Since the introduction of the

theoretical background of the event-chain Monte Carlo algorithm in the next section relies on the lifting framework, that collection is briefly summarized for convenience in the following.

As in Section 2.1, consider a finite configuration space Ω. A lifted Markov chain consists of a lifted configuration space $\widehat{\Omega}$, lifted stationary distribution $\hat{\pi}$, and lifted transition matrix $\widehat{P}$. Suppose that $\widehat{\Omega} = \Omega \times \mathcal{L}$, that is, every configuration $x \in \Omega$ is lifted into $|\mathcal{L}|$ copies in $\widehat{\Omega}$ that are written as $(x, i) \in \widehat{\Omega}$ where $i \in \mathcal{L}$. The necessary global-balance condition for the lifted Markov chain becomes

$$\hat{\pi}(y, j) = \sum_{(x,i) \in \widehat{\Omega}} \hat{\pi}(x, i) \, \widehat{P}[(x, i), (y, j)] \quad \text{for all } (y, j) \in \widehat{\Omega}. \tag{2.27}$$

The aim is to use a lifted Markov chain in place of a connected collapsed Markov chain that is typically reversible. As motivated in Section 1.3.2, this may, in the best case, lead to a diffusive-to-ballistic speedup if a non-reversible lifting is used. Let π be the stationary distribution of the irreducible collapsed Markov chain and P its transition matrix. The connection between the lifted and collapsed Markov chains is enforced by the two conditions

$$\pi(x) = \sum_{i \in \mathcal{L}} \hat{\pi}(x, i) \quad \text{for all } x \in \Omega, \tag{2.28}$$

and

$$\pi(x) \, P(x, y) = \sum_{i,j \in \mathcal{L}} \hat{\pi}(x, i) \, \widehat{P}[(x, i), (y, j)] \quad \text{for all } x, y \in \Omega. \tag{2.29}$$

The first condition enforces that the combined probabilities of all lifted copies of $x \in \Omega$ is equal to its desired stationary probability. The second condition ensures that the probability flows between all lifted copies of two configurations $x, y \in \Omega$ is the same as the probability flow between them in the collapsed Markov chain. With these conditions, one can sample the collapsed stationary distribution π with a lifted MCMC algorithm on $\widehat{\Omega}$ that can be constructed to be non-reversible in a systematic manner.

As discussed in Section 1.3.2, the lifting framework allows for a systematic construction of a non-reversible Markov chain from an underlying reversible one. A non-reversible lifting of a reversible Markov chain can reduce the mixing time of the collapsed chain to at most its square root in a diffusive-to-ballistic speedup (as, for example, in the nearest-neighbor random walk in Fig. 1.3) [111, 112]. The possible improvement of a non-reversible lifted Markov chain depends on the details of the target distribution on the configuration space Ω. Remarkably, there is always a single bottleneck that partitions Ω into two pieces (say, a high potential-energy barrier that separates distinct regions of Ω with comparable potential-energy minima) [74, Section 7.2]. The probability flow across that bottleneck, the conductance, is left unchanged by a lifting and bounds the mixing time from below and above [80, Section 2.1.3]. The lower "ballistic" bound results from the necessity to cross the bottleneck and it differs from the upper "diffusive" bound by a square-root operation. The diffusive-to-ballistic speedup can only be reaped if the reversible collapsed Markov chain is close to the upper bound. If the collapsed chain is, in contrast, limited by the crossings of the bottleneck and not by the exploration of the respective configuration-space partitions, the non-reversible lifted Markov chain will be similarly restricted by the conductance (see, e.g., the example in Ref. [80, Section 3.2.1]).

2.4 Foundations of Non-Reversible Event-Chain Monte Carlo

All research papers of this doctoral thesis consider the non-reversible event-chain Monte Carlo (ECMC) algorithm that samples the canonical (*NVT*) ensemble of continuous configuration spaces in a rejection-free continuous-time Markov chain in an event-driven manner. It expresses the Boltzmann weight $\exp(-\beta U)$ as a factorized product $\prod_M \exp(-\beta U_M)$ over statistically independent factors M. Here, for the example of molecular systems such as N SPC/Fw water molecules, the factor potentials U_M construct the molecular-mechanics potential-energy function $U = \sum_M U_M$ (see Section 2.4.1). The factorization enables a rigorously exact treatment of long-range interactions without any approximations because a tedious interaction between all interaction sites is replaced by a set of manageable interactions between a small number of interaction sites. ECMC is a systematic non-reversible lifting of the reversible factorized Metropolis algorithm (see Section 2.4.2). A piecewise-deterministic motion of a single interaction site is interrupted by an event which results from a competition between stochastic inhomogeneous Poisson processes for every factor M that each generate a candidate event time (see Section 2.4.3). The winning factor that realizes the event determines the initial conditions for the next deterministic piece in the non-reversible dynamics in a lifting move. Periodic resamplings of its lifting variables, allow ECMC to become irreducible and to sample the thermodynamic equilibrium in its trajectory (see Section 2.4.4). Thinning (see Section 2.4.5) is an essential ingredient for the cell-veto algorithm (see Section 2.4.6) that bundles the competition between $O(N)$ inhomogeneous Poisson processes in long-range-interacting systems such as N SPC/Fw water molecules into a single candidate event time that can be generated in constant computational time. This paves the way to a treatment of long-range interactions that is exact without any approximation yet remains efficient. The open-source application JᴇʟʟʏFʏsʜ implements these essential ideas of ECMC in a general manner [P6]. It may also move rigid molecules instead of single interaction sites. This is generally possible in ECMC [P6] although the following theoretical introduction only considers the case of a single moving interaction site for concreteness.

ECMC is an entire family of non-reversible local MCMC algorithms. This encompasses, on the one hand, the choice of different factor sets for the factorization of the Boltzmann distribution (see Section 2.4.1). On the other hand, this doctoral thesis highlights that, even for a given set of factors, the different straight [115], reflective [115], forward [150], and Newtonian ECMC [124] variants that consider different lifting moves in their events may have widely different behaviors (see Fig. 1.5). Before this doctoral thesis, only the straight ECMC variant [115] was considered for molecular systems [P6, 146]. The original reason for this is that straight ECMC was found to be superior in the hard-disk system [115]. The subsequent generalization to arbitrary factor potentials [121, 145] and the introduction of the cell-veto algorithm [148] therefore only concentrated on straight ECMC, and so did the first applications to molecular systems. Reference [146] gives an extensive introduction into the straight ECMC variant for general factor potentials. Except for the introduction of the cell-veto algorithm that (for now, see below) only considers straight ECMC (see Section 2.4.6), the following theoretical introduction attempts to be more general and considers both straight and reflective ECMC. This is in particular interesting for the discussion of the lifting moves for general translationally invariant factor potentials that may depend on an arbitrary number of interaction sites (see Section 2.4.7). To the author's knowledge, these general lifting moves were only formulated

for straight ECMC before and this doctoral thesis therefore provides the first explicit treatment of reflective ECMC.

Forward and Newtonian ECMC are *not* considered in the following theoretical introduction of ECMC. In this doctoral thesis, forward ECMC, although introduced more generally [150], is only applied to the hard-disk system and closely related hard-disk dipoles where it can be established more straightforward (see the second and third research papers in Chapters 4 and 5, and Fig. 1.5). Newtonian ECMC, which was initially proposed in the hard-sphere system [124], is generalized to smooth factor potentials that depend on an arbitrary number of interaction sites in the fourth research paper of this doctoral thesis (see Chapter 6). The same research paper also introduces how the cell-veto algorithm can be used for general ECMC variants. These generalizations are essential for a competitive performance of (Newtonian) ECMC in molecular systems and are thus part of JeLLyFysh-Version2.0 that we implemented for the fourth research paper of this doctoral thesis (see Chapter 6) [P6, P9].

2.4.1 Factorization of the SPC/Fw Water Model

The potential energy U of a molecular-mechanics model can be typically expressed as a sum over factor potentials U_M, i.e., $U(\mathbf{x}) = \sum_M U_M(\mathbf{x}_M)$ (see Section 1.1). Here, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_i})$ contains all positions of N_i three-dimensional interaction sites in a periodic simulation box, whereas the factor configuration $\mathbf{x}_M$ only contains a small subset of interaction-site positions $\mathbf{x}_i$. The introduction of the factor potentials U_M leads to a *factorization* of the Boltzmann distribution:

$$\pi(\mathbf{x}) = \frac{1}{Z}\, \mathrm{e}^{-\beta U(\mathbf{x})} = \prod_M \frac{1}{Z_M}\, \mathrm{e}^{-\beta U(\mathbf{x}_M)} = \prod_M \pi_M(\mathbf{x}_M). \tag{2.30}$$

ECMC relies on the factorized Metropolis filter that treats every of the factor distributions $\pi_M(\mathbf{x}_M)$ statistically independent (see Section 2.4.2). As usual in MCMC algorithms, this is possible without knowing their normalization factor Z_M.

Consider, for an example of the factorization of a molecular system, the SPC/Fw water model with $N = N_i/3$ water molecules in a periodic cubic box of side length L [28]. For simplicity, denote the positions of the two hydrogens in the ith molecule by $\mathbf{x}^i_{\mathrm{H},1}$ and $\mathbf{x}^i_{\mathrm{H},2}$, and the position of the oxygen by $\mathbf{x}^i_{\mathrm{O}}$. The potential energy is given by (see Fig. 1.1)

$$\begin{aligned} U(\mathbf{x}) = &\sum_{i=1}^{N} \left[U^{i,1}_{\mathrm{bond}}\left(\mathbf{x}^i_{\mathrm{H},1}, \mathbf{x}^i_{\mathrm{O}}\right) + U^{i,2}_{\mathrm{bond}}\left(\mathbf{x}^i_{\mathrm{H},2}, \mathbf{x}^i_{\mathrm{O}}\right) + U^i_{\mathrm{bend}}\left(\mathbf{x}^i_{\mathrm{H},1}, \mathbf{x}^i_{\mathrm{O}}, \mathbf{x}^i_{\mathrm{H},2}\right) \right] \\ &+ \sum_{i=1}^{N}\sum_{j=1}^{i-1} \left[U^{ij}_{\mathrm{LJ}}\left(\mathbf{x}^i_{\mathrm{O}}, \mathbf{x}^j_{\mathrm{O}}\right) + U^{ij}_{\mathrm{C}}\left(\mathbf{x}^i_{\mathrm{H},1}, \mathbf{x}^i_{\mathrm{O}}, \mathbf{x}^i_{\mathrm{H},2}, \mathbf{x}^j_{\mathrm{H},1}, \mathbf{x}^j_{\mathrm{O}}, \mathbf{x}^j_{\mathrm{H},2}\right) \right]. \end{aligned} \tag{2.31}$$

All of the factor potentials are translationally invariant because they only depend on the separations between the interaction sites of their factor configuration. Let $\mathbf{r}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_j - \mathbf{x}_i$ denote the shortest separation vector from $\mathbf{x}_i$ to $\mathbf{x}_j$ that is possibly corrected for periodic boundary conditions. The $2N$ intramolecular bond factor potentials yield fluctuations of the O−H bond lengths around their

equilibrium length r^0_{OH} and are given by

$$U^{i,k}_{\mathrm{bond}}\left(\mathbf{x}^i_{\mathrm{H},k},\mathbf{x}^i_{\mathrm{O}}\right) := \frac{k_b}{2}\left[\left|\mathbf{r}\left(\mathbf{x}^i_{\mathrm{H},k},\mathbf{x}^i_{\mathrm{O}}\right)\right| - r^0_{\mathrm{OH}}\right]^2 \quad \text{for } k \in \{1,2\}. \tag{2.32}$$

With $\theta(\mathbf{x}_i,\mathbf{x}_j)$ as the angle between the vectors $\mathbf{x}_i$ and $\mathbf{x}_j$, the N intramolecular bending factor potentials lead to fluctuations of the H−O−H opening angles around the equilibrium angle $\theta^0_{\angle\mathrm{HOH}}$ and are given by

$$U^i_{\mathrm{bend}}\left(\mathbf{x}^i_{\mathrm{H},1},\mathbf{x}^i_{\mathrm{O}},\mathbf{x}^i_{\mathrm{H},2}\right) := \frac{k_a}{2}\left\{\theta\left[\mathbf{r}\left(\mathbf{x}^i_{\mathrm{O}},\mathbf{x}^i_{\mathrm{H},1}\right),\mathbf{r}\left(\mathbf{x}^i_{\mathrm{O}},\mathbf{x}^i_{\mathrm{H},2}\right)\right] - \theta^0_{\angle\mathrm{HOH}}\right\}^2. \tag{2.33}$$

The intermolecular factors also have to consider the periodic boundary conditions. Every interaction site i at position $\mathbf{x}_i$ does not only interact with the closest periodic image of the interaction site j at position $\mathbf{x}_j = \mathbf{x}_i + \mathbf{r}(\mathbf{x}_i,\mathbf{x}_j)$ but with *every* periodic image of j at positions $\mathbf{x}_i + \mathbf{r}(\mathbf{x}_i,\mathbf{x}_j) + \mathbf{m}L$, where $\mathbf{m} \in \mathbb{Z}^3$. The $N(N-1)/2$ Lennard-Jones factor potentials are thus given by the infinite sum

$$U^{ij}_{\mathrm{LJ}}\left(\mathbf{x}^i_{\mathrm{O}},\mathbf{x}^j_{\mathrm{O}}\right) := \sum_{\mathbf{m}\in\mathbb{Z}^3} 4\varepsilon_{\mathrm{LJ}}\left[\left(\frac{\sigma_{\mathrm{LJ}}}{\left|\mathbf{r}\left(\mathbf{x}^i_{\mathrm{O}},\mathbf{x}^j_{\mathrm{O}}\right)+\mathbf{m}L\right|}\right)^{12} - \left(\frac{\sigma_{\mathrm{LJ}}}{\left|\mathbf{r}\left(\mathbf{x}^i_{\mathrm{O}},\mathbf{x}^j_{\mathrm{O}}\right)+\mathbf{m}L\right|}\right)^6\right]. \tag{2.34}$$

The $N(N-1)/2$ intermolecular Coulomb factor potentials are themselves a sum of two-body Coulomb potentials between any two interaction sites in different water molecules:

$$U^{ij}_{\mathrm{C}}\left(\mathbf{x}^i_{\mathrm{H},1},\mathbf{x}^i_{\mathrm{O}},\mathbf{x}^i_{\mathrm{H},2},\mathbf{x}^j_{\mathrm{H},1},\mathbf{x}^j_{\mathrm{O}},\mathbf{x}^j_{\mathrm{H},2}\right) := \sum_{a_i}\sum_{a_j} U_{\mathrm{C}}\left[\mathbf{r}\left(\mathbf{x}^i_{a_i},\mathbf{x}^j_{a_j}\right),q_{a_i},q_{a_j}\right], \tag{2.35}$$

where $a_i, a_j \in \{(\mathrm{H},1),(\mathrm{O}),(\mathrm{H},2)\}$ sum over the possible types of interaction sites in the two water molecules and q_{a_i} are their charges in the molecular-mechanics model. The two-body Coulomb potentials again contain an infinite sum over periodic images

$$U_{\mathrm{C}}\left(\mathbf{r}_{ij},q_i,q_j\right) := \sum_{\mathbf{m}\in\mathbb{Z}^3} k_e \frac{q_i q_j}{\left|\mathbf{r}_{ij}+\mathbf{m}L\right|}, \tag{2.36}$$

with the Coulomb constant k_e. In practice, the infinite sums in Eqs (2.34) and (2.36) are either cut off or treated by an Ewald summation (see, e.g., Ref. [12, Appendix B]). The various parameters k_b, r^0_{OH}, k_a, $\theta^0_{\angle\mathrm{HOH}}$, $\varepsilon_{\mathrm{LJ}}$, σ_{LJ}, $q_{\mathrm{H},1} = q_{\mathrm{H},2}$, and q_{O} of the SPC/Fw water model are chosen empirically and can be found in Ref. [28].

The factor potentials of the SPC/Fw water model in Eqs (2.32), (2.33), (2.34), and (2.35) are one possible choice. One could also, for instance, treat every of the nine two-body Coulomb potentials in Eq. (2.35) as their own factor potential. The correct choice of factors is essential and influences the dynamics and computational complexity of ECMC. For N SPC/Fw water molecules, only the specified choice of factors in this section yields an optimal $O(N\log N)$ scaling to displace every interaction site by a constant distance [146].

2.4.2 Factorized Metropolis Filter

The Metropolis filter gives the acceptance probability of a proposed move from $\mathbf{x}$ to $\mathbf{x}'$ in the Metropolis algorithm as [72]

$$
\begin{aligned}
p_\pi^{\mathrm{Met}}(\mathbf{x}, \mathbf{x}') &:= \min\left[1, \frac{\pi(\mathbf{x}')}{\pi(\mathbf{x})}\right] = \min\left[1, \mathrm{e}^{-\beta[U(\mathbf{x}')-U(\mathbf{x})]}\right] \\
&= \min\left[1, \prod_M \mathrm{e}^{-\beta[U_M(\mathbf{x}'_M)-U(\mathbf{x}_M)]}\right].
\end{aligned}
\tag{2.37}
$$

As discussed in Section 1.3.1, it appears in the transition matrix of the Metropolis algorithm in Eq. (1.5) that can be shown to satisfy the detailed-balance condition. In a similar manner, it can be shown that the Metropolis filter already satisfies the detailed-balance condition:

$$
\pi(\mathbf{x})\, p_\pi^{\mathrm{Met}}(\mathbf{x}, \mathbf{x}') = \pi(\mathbf{x}')\, p_\pi^{\mathrm{Met}}(\mathbf{x}', \mathbf{x}).
\tag{2.38}
$$

The Metropolis filter may be understood as drawing a Boolean random variable that yields either an acceptance or a rejection of the proposed configuration $\mathbf{x}'$. This acceptance–rejection decision requires a global computation of the change of all factor potentials $\sum_M [U_M(\mathbf{x}'_M) - U_M(\mathbf{x}_M)]$. The *factorized Metropolis filter* instead considers a consensus between independent Metropolis decisions for every factor [146]. Here, every factor either accepts or rejects the proposed factor configuration $\mathbf{x}'_M$ based on the change of the factor potential $U_M(\mathbf{x}'_M) - U_M(\mathbf{x}_M)$. Only when all factors agree in the consensus, the proposed configuration $\mathbf{x}'$ is accepted. The acceptance probability of the factorized Metropolis filter is thus given by [121]

$$
\begin{aligned}
p_\pi^{\mathrm{Fact}}(\mathbf{x}, \mathbf{x}') &:= \prod_M p_{\pi_M}^{\mathrm{Met}}(\mathbf{x}_M, \mathbf{x}'_M) = \prod_M \min\left[1, \frac{\pi_M(\mathbf{x}'_M)}{\pi_M(\mathbf{x}_M)}\right] \\
&= \prod_M \min\left[1, \mathrm{e}^{-\beta[U_M(\mathbf{x}'_M)-U_M(\mathbf{x}_M)]}\right]
\end{aligned}
\tag{2.39}
$$

Because of its close connection to the original Metropolis filter, the factorized Metropolis filter also satisfies the detailed-balance condition:

$$
\begin{aligned}
\pi(\mathbf{x})\, p_\pi^{\mathrm{Fact}}(\mathbf{x}, \mathbf{x}') &= \prod_M \pi_M(\mathbf{x}_M)\, p_{\pi_M}^{\mathrm{Met}}(\mathbf{x}_M, \mathbf{x}'_M) \\
&= \prod_M \pi_M(\mathbf{x}'_M)\, p_{\pi_M}^{\mathrm{Met}}(\mathbf{x}'_M, \mathbf{x}_M) = \pi(\mathbf{x}')\, p_\pi^{\mathrm{Fact}}(\mathbf{x}', \mathbf{x})
\end{aligned}
\tag{2.40}
$$

Therefore, one can simply use the factorized Metropolis filter in the Metropolis algorithm which would, however, give smaller acceptance probabilities because changes in factor potentials cannot compensate each other. At the same time, sampling the outcome of the consensus process between the factors under a local move has a lower computational complexity than sampling the outcome of the original Metropolis decision in long-range-interacting systems (using similar ideas that were developed beforehand in the context of ECMC and are presented in the following) [163]. Moreover, the

factorization allows for an exact treatment of long-range interactions in ECMC. For instance, instead of considering the Coulomb potential between all water molecules at once with limited accuracy (as in molecular dynamics or the Metropolis algorithm), ECMC only explicitly considers the Coulomb factor potential between two water molecules in Eq. (2.35) that can be efficiently computed to machine precision with constant computational complexity with an Ewald summation.

2.4.3 Competing Inhomogeneous Poisson Processes

ECMC is a non-reversible lifting of the factorized Metropolis algorithm in the continuous-time limit and can also be interpreted as a piecewise-deterministic Markov process [136, 137]. For periodic three-dimensional molecular systems, straight and reflective ECMC consider the lifted configuration space $\widehat{\Omega} = \Omega \times \mathcal{V} \times \mathcal{N}$, where $\mathcal{N} = \{1, 2, \ldots, N_i\}$ is the set of interaction-site indices and $\mathcal{V}$ is a (finite or infinite) set of three-dimensional velocities of absolute value 1 (other choices of the absolute value lead to trivial rescalings of the continuous Monte Carlo time). Reflective ECMC uses the unit two-sphere $\mathcal{V} = S^2$ as its continuous velocity space. Straight ECMC may also consider $\mathcal{V} = S^2$ but also allows for finite velocity sets $\mathcal{V}$. For instance, it may choose $\mathcal{V} = \{(1, 0, 0)^T, (0, 1, 0)^T, (0, 0, 1)^T\}$, which is inspired by the powerful application to the hard-disk system that similarly aligned the velocity with all coordinate axes [115, 142]. In any case, the lifted Markov chain targets the lifted stationary distribution $\hat{\pi}(\mathbf{x}, \mathbf{v}, i) = \pi(\mathbf{x}) \times \mu_{\mathcal{V}}(\mathbf{v}) \times \mu_{\mathcal{N}}(i)$ that separates into the Boltzmann distribution $\pi(\mathbf{x})$, and the uniform distributions $\mu_{\mathcal{V}}(\mathbf{v})$ on $\mathcal{V}$ and $\mu_{\mathcal{N}}(i)$ on $\mathcal{N}$. Hence, the two conditions in Eqs (2.28) and (2.29) that connect the lifted ECMC algorithm to the collapsed factorized Metropolis algorithm are trivially satisfied.

Given a lifted configuration $(\mathbf{x}, \mathbf{v}_a, i) \in \widehat{\Omega}$ at time $t = t_0$, the active interaction site moves with its velocity $\mathbf{v}_a$ in a deterministic straight-line trajectory which leads to the time-dependent configuration $\mathbf{x}(t) = (\mathbf{x}_1, \ldots, \mathbf{x}_i + (t-t_0)\mathbf{v}_a, \ldots, \mathbf{x}_{N_i})$. At every continuous Monte-Carlo time $t > t_0$, the infinitesimal change of position is accepted or rejected by the consensus of the factorized Metropolis filter. An infinitesimal change of the position implies an infinitesimal change of the factor potentials. In this limit, the rejection probability $1 - p^{\text{fact}}$ becomes a time-dependent event rate $\lambda(t) \geq 0$ that can be interpreted as a (unnormalized) probability density in time to interrupt the piecewise-deterministic motion of the active interaction site [121, 137, 146]:

$$\lambda(t) = \sum_M \lambda_M(t) = \sum_M \beta \, \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_i} U_M[\mathbf{x}_M(t)] \cdot \mathbf{v}_a\right]. \tag{2.41}$$

The *factor event rates* $\lambda_M(t)$ can only be non-zero for factors M that contain the active interaction site i, and for times t where the motion increases the factor potential U_M. This results from the factorized Metropolis filter where a factor M always accepts changes of the the factor configuration that do not increase the factor potential. If it does increase, the acceptance probability decreases with the change of the factor potential. This translates into the dependence of the factor event rate $\lambda_M(t)$ on the rate of change $\boldsymbol{\nabla}_{\mathbf{x}_i} U_M[\mathbf{x}_M(t)] \cdot \mathbf{v}_i$ of the factor potential under the motion of the active interaction site.

The factorized Metropolis filter was viewed as a consensus decision between factor-based acceptance–rejection decisions. Similarly, one can understand Eq. (2.41) as a competition of the factor event rates $\lambda_M(t)$ that each attempt to interrupt the straight-line trajectory of the active interaction site. Instead of

advancing the time in discrete time steps [which would involve an approximation because Eq. (2.41) only becomes exact in the limit of infinitesimal time steps], the competition between the event rates can be treated rigorously exact in an event-driven manner [116, 121]. Every factor M generates a candidate event time $t_{\text{ev},M}$ in an *inhomogeneous Poisson process* based on its factor event rate $\lambda_M(t)$ (see, e.g., Ref. [164, Chapter VI]). Let $T_{\text{ev,M}}$ be the random variable for the candidate event time of the factor M and let $\mathbf{P}_M(T_{\text{ev},M} > t)$ be the probability that $T_{\text{ev, M}}$ is larger than $t \geq t_0$, that is, the probability that there is no candidate event up to time t. By definition, the probability for a candidate event of the factor M in the infinitesimal time interval $[t, t + dt]$ is given by $\lambda_M(t)\, dt$. Thus,

$$\mathbf{P}_M(T_{\text{ev},M} > t + dt) = \mathbf{P}_M(T_{\text{ev},M} > t) \left[1 - \lambda_M(t)\, dt \right] . \tag{2.42}$$

With $\mathbf{P}_M(T_{\text{ev},M} > t + dt) - \mathbf{P}_M(T_{\text{ev},M} > t) = d\mathbf{P}_M(T_{\text{ev},M} > t)$, one gets

$$\frac{d\mathbf{P}_M(T_{\text{ev},M} > t)}{dt} = -\lambda_M(t)\, \mathbf{P}_M(T_{\text{ev},M} > t). \tag{2.43}$$

This yields

$$\mathbf{P}_M(T_{\text{ev},M} > t) = \exp\left[-\int_{t_0}^{t} \lambda_M(t)\, dt \right] . \tag{2.44}$$

The cumulative distribution function $F_{\text{ev},M}(t)$ of the random variable $T_{\text{ev},M}$ is given by the probability $\mathbf{P}_M(T_{\text{ev},M} \leq t)$ that there is a candidate event up to time t. Using Eq. (2.44), it is therefore given by $F_{T_{\text{ev, M}}}(t) = \mathbf{P}_M(T_{\text{ev},M} \leq t) = 1 - \mathbf{P}_M(T_{\text{ev},M} > t)$. With that, one can use inverse transform sampling for the random variable $T_{\text{ev},M}$ in order to obtain a *candidate event time* $t_{\text{ev},M}$ with the correct probability distribution (see, e.g., Ref. [164, Section II.2]):

$$\text{ran}_M(0, 1) = \exp\left[-\int_{t_0}^{t_{\text{ev},M}} \lambda_M(t)\, dt \right] , \tag{2.45}$$

where $\text{ran}_M(0, 1)$ is a uniformly distributed random number between 0 and 1. At the candidate event time $t_{\text{ev},M}$, an exponentially distributed random number with mean $1/\beta$ is equal to the cumulative increments of the factor potential U_M under the motion of the active interaction site since it started moving at time t_0.

Given the lifted configuration $(\mathbf{x}(t_0), \mathbf{v}_a, i) \in \widehat{\Omega}$, every factor M containing the active interaction site i yields an independent candidate event time $t_{\text{ev},M} \geq t_0$ via Eq. (2.45). This is the point in time where the factor M would break consensus and reject a continued movement of the active interaction site. (For the hard-disk system, this is the point in time of a collision.) The set of candidate event times that were obtained by independent inhomogeneous Poisson processes compete against one another and the active interaction site can only move up to the unique minimum candidate event time $t_{\text{ev}} = \min_M t_{\text{ev},M}$. At the realized *event time* t_{ev}, the active interaction site stops at the configuration $\mathbf{x}(t_{\text{ev}})$ and a lifting move of the event-triggering factor $M = \arg\min_M t_{\text{ev},M}$ changes the lifted configuration from $(\mathbf{x}(t_{\text{ev}}), \mathbf{v}_a, i) \in \widehat{\Omega}$ to $(\mathbf{x}(t_{\text{ev}}), \mathbf{v}'_a, j) \in \widehat{\Omega}$ (see Section 2.4.7). This determines the initial conditions for the next leg of the piecewise-deterministic Markov process starting at $t_0 = t_{\text{ev}}$. Lifting moves in non-reversible ECMC replace rejections of the underlying factorized Metropolis algorithm.

2.4.4 Resampling of Lifting Variables and Sampling

Events of the straight ECMC variant transfer the velocity $\mathbf{v}_a$ of the previously active interaction site i to another interaction site j of the event-triggering factor M (see Section 2.4.7.1). In order to generate an irreducible Markov chain, it thus has to resample the lifting variables (i.e., the velocity and the active interaction site) from their respective uniform distributions $\mu(\mathcal{V})$ and $\mu(\mathcal{N})$ in periodic intervals of the chain time τ_{chain}. Moreover, a possible finite set of velocities $\mathcal{V}$ should contain at least three linear independent velocities in periodic three-dimensional simulation boxes. Although reflective ECMC changes both the active interaction site and the velocity in an event (see Section 2.4.7.2), it can still fail to be irreducible in simple systems if the lifting variables are not resampled after the chain time τ_{chain} (see, e.g., the second research paper of this doctoral thesis in Chapter 4). This was, in fact, one of the motivations of the introduction of forward ECMC that randomizes the velocity update in an event [150]. The chain time τ_{chain} is an intrinsic parameter of straight and reflective ECMC that has to be tuned to minimize mixing and autocorrelation times. As the second and third research papers of this doctoral thesis show, straight ECMC is the only variant that requires a non-trivial choice (see Chapters 4 and 5).

Since ECMC generates a continuous-time Markov chain, one could, in principle, consider the time average of the function $\mathcal{A}$ along the entire trajectory to estimate its thermodynamic average via the ergodic theorem (see, e.g., Refs [139, 140]). This results in a series of integrals over the function values of $\mathcal{A}$ along the piecewise-deterministic legs between events which, in principle, just requires to store the event times t_{ev} together with the configurations $\mathbf{x}(t_{\mathrm{ev}})$ during the simulation. In the hard-disk system, for example, this allows to estimate the pressure without any extrapolations as we highlighted in Ref. [P5]. However, since these integrals are typically intractable in other systems, one may instead just discretize the continuous-time trajectory in fixed time intervals and then compute the average of the function $\mathcal{A}$ over this set of configurations [which is similar to the time average over a discrete-time Markov chain in Eq. (2.12)]. This can also heavily reduce the amount of generated data if long ECMC simulations are necessary by choosing a large sampling time interval.

2.4.5 Thinning

In order to determine the next event time $t_{\mathrm{ev}} = \min_M t_{\mathrm{ev},M}$ that interrupts the motion of the active interaction site i with velocity $\mathbf{v}_a$, every factor M has to compute its candidate event time $t_{\mathrm{ev},M}$ by solving Eq. (2.45). This requires the knowledge of the minima and maxima of the factor potential along the path of the active interaction site. In the SPC/Fw water model, for example, the bond factor potentials in Eq. (2.32) and cut-off Lennard-Jones factor potentials in Eq. (2.34) that only need to consider the $\mathbf{m} = (0, 0, 0)$ term can be treated in this way. The bending and Coulomb factor potentials in Eqs (2.33) and (2.36), however, require a different procedure. *Thinning* replaces the event rate $\lambda_M(t)$ in Eq. (2.45) by an upper bound $\tilde{\lambda}_M^+(t)$ that satisfies $\tilde{\lambda}_M^+(t) \geq \lambda_M(t)$ for all t [165] and can be understood as resulting from a bounding potential instead of the original factor potential [P6]. Naturally, $\tilde{\lambda}_M^+(t)$ should be chosen so that the inversion-problem becomes solvable. The simplest option would be to choose a constant bound $\tilde{\lambda}_M^+(t)$. Increasing the real event rate $\lambda_M(t)$ to $\tilde{\lambda}_M^+(t)$ "fattens" the inhomogeneous Poisson process and yields, on average, more candidate event times $t_{\mathrm{ev},M}$ for the factor M. When the candidate event time $t_{\mathrm{ev, M}}$ of the factor $M = \arg\min_{M'} t_{\mathrm{ev},M'}$ becomes an actual event, superfluous events from

the fattened Poisson process have to be thinned out. This is done by accepting the event with probability $\lambda_M(t_{\text{ev},M})/\tilde{\lambda}_M^+(t_{\text{ev},M}) \leq 1$. Only if the event is accepted, the lifting move in the factor M is carried out. Otherwise, the interaction site i continues its motion. In this case, only the factor M has to compute a new candidate event time $t_{\text{ev},M}$ with the event rate bound $\tilde{\lambda}_M^+(t)$. All candidate event times of other factors M' involving i remain unchanged. This is implemented in JₑLLᵧFYSH-Version2.0 for the fourth research paper of this doctoral thesis (see Chapter 6) [P6, P9]. JₑLLᵧFYSH-Version2.0 considers various thinning procedures that, for example, automatically obtain a good short-time estimate on a constant event rate bound $\tilde{\lambda}_M^+$ (which corresponds to adaptive thinning; see, e.g., Ref. [140, Section 2.3.2]).

2.4.6 Cell-Veto Algorithm

If a single interaction site moves among N long-range-interacting SPC/Fw water molecules, $O(1)$ intramolecular factors are changing in time whose candidate event times can be computed in $O(1)$ computational complexity. For the intermolecular factors, in contrast, $O(N)$ candidate event times have to be computed and sorted to obtain the realized event time. The *cell-veto algorithm* allows to bundle the intermolecular factors together and, hence, to obtain an event with an $O(1)$ computational complexity [148]. Since the overall event rate in the SPC/Fw water model that is factorized as in Section 2.4.1 increases as $O(\log N)$, the cell-veto algorithm then allows to achieve an $O(N \log N)$ scaling to displace every interaction site by a constant distance [146]. This matches the scaling of fast mesh-based Ewald methods in molecular dynamics. The combination of ECMC and the cell-veto algorithm, however, treats the long-range interactions rigorously exact. The cell-veto algorithm was initially formulated for a finite velocity space $\mathcal{V}$ and could thus only be used in the straight ECMC variant (see, e.g., Refs [P6, 146]). This formulation is presented in the current section. The fourth research paper of this doctoral thesis shows how the cell-veto algorithm can be used for continuous velocity spaces as they appear in other ECMC variants (see Section 6).

Consider, for instance, the Coulomb factors between two water molecules in Eq. (2.36). A cell-occupancy system maps the centers of the water molecules to a discrete set of cells $\{\mathcal{D}_i\}$ and vice versa. For every pair of cells $\mathcal{D}_a$ and $\mathcal{D}_b$, one then *assumes* that there is a single molecule present in them with interaction-site positions $\mathbf{x}_{\text{HOH}}^a = (\mathbf{x}_{\text{H},1}^a, \mathbf{x}_{\text{O}}^a, \mathbf{x}_{\text{H},2}^a) \in \mathcal{D}_a$ and $\mathbf{x}_{\text{HOH}}^b = (\mathbf{x}_{\text{H},1}^b, \mathbf{x}_{\text{O}}^b, \mathbf{x}_{\text{H},2}^b) \in \mathcal{D}_b$, respectively. Here, with a little abuse of notation, $\mathbf{x}_{\text{HOH}}^a \in \mathcal{D}_a$ means that the center position of the set $(\mathbf{x}_{\text{H},1}^a, \mathbf{x}_{\text{O}}^a, \mathbf{x}_{\text{H},2}^a)$ is part of $\mathcal{D}_a$. The factor event rate of the corresponding Coulomb factor $M = \text{C}^{ab}$ can then be bounded by a *cell bound* $\tilde{\lambda}_{\text{C}^{ab}}^+(\mathbf{x}_a, \mathbf{v}_a)$ for every possible active interaction site a in the first water molecule, $\mathbf{x}_a \in \mathbf{x}_{\text{HOH}}^a$, and every possible velocity $\mathbf{v}_a \in \mathcal{V}$. The cell bound is computed *before the simulation starts* by formally considering

$$\tilde{\lambda}_{\text{C}^{ab}}^+(\mathbf{x}_a, \mathbf{v}_a) \geq \max_{\mathbf{x}_{\text{HOH}}^a \in \mathcal{D}_a, \mathbf{x}_{\text{HOH}}^b \in \mathcal{D}_b} \beta \max \left[0, \boldsymbol{\nabla}_{\mathbf{x}_a} U_{\text{C}}^{ab}(\mathbf{x}_{\text{HOH}}^a, \mathbf{x}_{\text{HOH}}^b) \cdot \mathbf{v}_a \right]. \qquad (2.46)$$

In practice, it is not necessary to consider every possible interaction-site configuration of a water molecule that would lead its center falling into the corresponding cell. For example, O−H bond lengths that span the entire cell never appear in an ECMC simulation that runs for a finite simulation time. One could instead vary the center positions and orientations of the two water molecules with "reasonable" bond lengths and opening angles in their cells to obtain the cell bound. Also, nearby cells

are usually excluded from the cell-veto algorithm because their estimated cell bounds would either be very large or even diverge (which could, in principle, be fixed in a more complex estimation procedure that does not consider problematic configurations).

During the actual simulation, given the active interaction site a at position $\mathbf{x}_a \in \mathbf{x}_{\text{HOH}}^a \in \mathcal{D}_a$ and its velocity $\mathbf{v}_a$, the cell-veto algorithm considers the entire set of Coulomb factors C^a that contains all Coulomb factors C^{ab} with hypothetical water molecules in all non-excluded cells $\mathcal{D}_b$ at once. Similar to the thinning procedure, it fattens the inhomogeneous Poisson process and computes a candidate event time $t_{\text{ev},\mathrm{C}^a}$ with the precomputed total cell-event rate $\tilde{\lambda}_{\mathrm{C}^a}^+(\mathbf{x}_a, \mathbf{v}_a) = \sum_b \tilde{\lambda}_{\mathrm{C}^{ab}}^+(\mathbf{x}_a, \mathbf{v}_a)$. If the candidate event time becomes an actual event time because $\mathrm{C}^a = \arg\min_M t_{\text{ev},M}$, three confirmation steps are necessary. First, if the event time would move the interaction site a outside of its cell $\mathcal{D}_a$, it is only placed on the boundary to the next cell and a new candidate event time is computed based on the cell bounds of that next cell (in the JᴇLLʏFʏsʜ application, this is solved by cell-boundary pseudo factors that yield candidate event times whenever the active interaction site crosses a cell boundary [P6]). Second, Walker's algorithm [166] samples a factor C^{ab} from the set of factors C^a with probability $\propto \tilde{\lambda}_{\mathrm{C}^{ab}}^+(\mathbf{x}_a, \mathbf{v}_a)$ which is then deemed responsible for the candidate event time. This sampling from a discrete probability distribution only requires constant computational complexity. Second, the event is accepted in a thinning procedure by comparing the event-rate bound $\tilde{\lambda}_{\mathrm{C}^{ab}}^+(\mathbf{x}_a, \mathbf{v}_a)$ to the actual event rate of the molecular C^{ab} factor (see Section 2.4.5). Only then, the actual positions of the water molecules in the cells $\mathcal{D}_a$ and $\mathcal{D}_b$ are considered and only then, the Coulomb event rate has to be computed with a high-accuracy Ewald summation. This thinning procedure makes the treatment of the long-range interactions in the cell-veto algorithm rigorously exact although the cell bounds are used most of the time. If the cell $\mathcal{D}_b$ does not contain any water molecule, the event is trivially rejected.

The cell-veto algorithm treats all intermolecular Coulomb factors C^{ab} between molecules in non-excluded cells $\mathcal{D}_a$ and $\mathcal{D}_b$ rigorously exact at once. With the help of precomputed cell bounds, it achieves an $O(1)$ computational complexity, despite the $O(N)$ increase of the number of involved factors. Coulomb factors between the water molecule containing the active interaction site, on the one hand, and water molecules in nearby excluded cells or surplus water molecules of cells with multiple occupants, on the other hand, are treated directly by a thinning procedure. Since the number of excluded and surplus Coulomb factors stays constant with increasing N [P4], one obtains an overall $O(1)$ computational complexity to compute the next event of the long-range Coulomb interaction. The cell-veto algorithm can be applied to any long-range interaction by changing the estimation procedure for the precomputed cell bounds. In the SPC/Fw water model, it can also be used for the Lennard-Jones factors (with or without a cutoff). The actual number of cell bounds that have to be precomputed can be heavily reduced by explicitly considering periodic boundary conditions and the translational invariance of the factor potentials [P4, P6, 146]. JᴇLLʏFʏsʜ-Version2.0 implements the cell-veto algorithm in a general manner and provides various estimation procedures for cell bounds [P6, P9]. It is thus well equipped for efficient and exact molecular simulations.

2.4.7 Lifting Moves

Lifting moves replace rejections of the collapsed factorized Metropolis algorithm in the lifted ECMC algorithm. They are necessary so that the necessary global-balance condition for the lifted Markov

chain in Eq. (2.27) is satisfied. Let $\mathcal{F}[(\mathbf{x}', \mathbf{v}'_a, j), (\mathbf{x}, \mathbf{v}_a, i)] := \hat{\pi}(\mathbf{x}', \mathbf{v}'_a, j) \, \widehat{P}[(\mathbf{x}', \mathbf{v}'_a, j), (\mathbf{x}, \mathbf{v}_a, i)]$ be the probability flow from the lifted configuration $(\mathbf{x}', \mathbf{v}'_a, j) \in \widehat{\Omega}$ into $(\mathbf{x}, \mathbf{v}_a, i) \in \widehat{\Omega}$. The global-balance condition demands that the total probability flow $\mathcal{F}(\mathbf{x}, \mathbf{v}_a, i)$ into any configuration $(\mathbf{x}, \mathbf{v}_a, i) \in \Omega$ is equal to its probability $\hat{\pi}(\mathbf{x}, \mathbf{v}_a, i)$:

$$\hat{\pi}(\mathbf{x}, \mathbf{v}_a, i) = \mathcal{F}(\mathbf{x}, \mathbf{v}_a, i) := \sum_{(\mathbf{x}', \mathbf{v}'_a, j) \in \widehat{\Omega}} \mathcal{F}[(\mathbf{x}', \mathbf{v}'_a, j), (\mathbf{x}, \mathbf{v}_a, i)], \tag{2.47}$$

where it is again assumed that all spaces in $\widehat{\Omega} = \Omega \times \mathcal{V} \times \mathcal{N}$ are finite. If the continuous velocity space $\mathcal{V} = S^2$ is considered instead, the corresponding sum has to replaced by an integral. Given the introduction of the piecewise-deterministic Markov process in the ECMC algorithm in Section 2.4.3, $\mathcal{F}(\mathbf{x}, \mathbf{v}_a, i)$ consists of two parts:

$$\mathcal{F}(\mathbf{x}, \mathbf{v}_a, i) = \mathcal{F}^{\text{phys}}(\mathbf{x}, \mathbf{v}_a, i) + \mathcal{F}^{\text{lift}}(\mathbf{x}, \mathbf{v}_a, i). \tag{2.48}$$

The physical flow $\mathcal{F}^{\text{phys}}(\mathbf{x}, \mathbf{v}_a, i)$ originates from the motion of the active interaction site i that was not interrupted by an event. It results from the unique lifted configuration $(\mathbf{x}', \mathbf{v}_a, i)$ where $\mathbf{x}' = (\mathbf{x}_1, \ldots, \mathbf{x}_i - \mathbf{v}_a \, dt, \ldots, \mathbf{x}_{N_i})$. The probability that the infinitesimal time step dt was accepted is given by the infinitesimal limit of the factorized Metropolis filter that is connected to the event rate in Eq. (2.41) (which gave the rejection probability in the infinitesimal limit):

$$\begin{aligned} \mathcal{F}^{\text{phys}}(\mathbf{x}, \mathbf{v}_a, i) &= \hat{\pi}(\mathbf{x}', \mathbf{v}_a, i) \left\{ 1 - \sum_M \beta \, \max \left[0, \boldsymbol{\nabla}_{\mathbf{x}'_i} U_M(\mathbf{x}'_M) \cdot \mathbf{v}_a \right] \right\} \\ &= \hat{\pi}(\mathbf{x}, \mathbf{v}_a, i) \left\{ 1 - \sum_M \beta \, \max \left[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a \right] \right\} \\ &=: \hat{\pi}(\mathbf{x}, \mathbf{v}_a, i) + \sum_M \mathcal{F}^{\text{phys}}_M(\mathbf{x}, \mathbf{v}_a, i). \end{aligned} \tag{2.49}$$

The second line used the detailed-balance property of the factorized Metropolis filter that holds in its infinitesimal limit [see Eq. (2.40)]. The lifting flow $\mathcal{F}^{\text{lift}}(\mathbf{x}, \mathbf{v}_a, i)$ results from lifted configurations $(\mathbf{x}, \mathbf{v}'_a, j)$ where the motion of the active interaction site j was interrupted by the factor M and a lifting move changed the lifted configuration to $(\mathbf{x}, \mathbf{v}_a, i)$ with probability $p_M[(\mathbf{v}'_a, j), (\mathbf{v}_a, i)]$. Therefore,

$$\begin{aligned} \mathcal{F}^{\text{lift}}(\mathbf{x}, \mathbf{v}_a, i) &= \sum_M \sum_{\mathbf{v}'_a \in \mathcal{V}} \sum_{j \in \mathcal{N}} \hat{\pi}(\mathbf{x}, \mathbf{v}'_a, j) \, \beta \, \max \left[0, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}'_a \right] \, p_M[(\mathbf{v}'_a, j), (\mathbf{v}_a, i)] \\ &=: \sum_M \mathcal{F}^{\text{lift}}_M(\mathbf{x}, \mathbf{v}_a, i). \end{aligned} \tag{2.50}$$

In order to satisfy the lifted global-balance condition, one requires

$$\sum_M \mathcal{F}^{\text{phys}}_M(\mathbf{x}, \mathbf{v}_a, i) = -\sum_M \mathcal{F}^{\text{lift}}_M(\mathbf{x}, \mathbf{v}_a, i). \tag{2.51}$$

In ECMC, this is solved by demanding

$$\mathcal{F}_M^{\text{phys}}(\mathbf{x}, \mathbf{v}_a, i) = -\mathcal{F}_M^{\text{lift}}(\mathbf{x}, \mathbf{v}_a, i), \tag{2.52}$$

which, by using the uniform distributions of $\mathbf{v}_a$ on $\mathcal{V}$ and i on $\mathcal{N}$, becomes

$$\max\left[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a\right] = \sum_{\mathbf{v}_a' \in \mathcal{V}} \sum_{j \in \mathcal{N}} \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a'\right] p_M\left[(\mathbf{v}_a', j), (\mathbf{v}_a, i)\right]. \tag{2.53}$$

Consider now the continuous velocity space on the unit two-sphere $\mathcal{V} = S^2$. A trivial solution of Eq. (2.53) would be to simply invert the velocity of the active interaction site j in an event, that is, $p_M^{\text{inv}}\left[(\mathbf{v}_a', j), (\mathbf{v}_a, i)\right] = \delta_{i,j}\,\delta^{(3)}(\mathbf{v}_a + \mathbf{v}_a')$ (which is also possible in a finite velocity space $\mathcal{V}$ containing both $\mathbf{v}_a$ and $-\mathbf{v}_a$). Another solution reflects the velocity with respect to the equipotential surface with normal vector $\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)$:

$$
\begin{aligned}
p_M^{\text{ref}}\left[(\mathbf{v}_a', j), (\mathbf{v}_a, i)\right] &= \delta_{i,j}\,\delta^{(3)}\left(\mathbf{v}_a - \mathbf{v}_a' + 2\frac{\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a'}{|\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)|^2}\,\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)\right) \\
&= \delta_{i,j}\,\delta^{(3)}\left(\mathbf{v}_a' - \mathbf{v}_a + 2\frac{\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a}{|\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)|^2}\,\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)\right).
\end{aligned}
\tag{2.54}
$$

Here, it was used that applying the reflection twice yields the original velocity. Both of these solutions are typically inefficient in a molecular system because they keep the same interaction site active but may be necessary, e.g., for potentials that depend on a single interaction site as confining potentials or hard walls [132]. Different ECMC variants consider more elaborate lifting moves to satisfy Eq. (2.53) while also changing the active interaction site. In the following, general lifting moves for the straight and reflective ECMC variants are presented.

2.4.7.1 Straight Event-Chain Monte Carlo

The straight ECMC variant can consider a finite or infinite set of velocities. It is valid for events from factors M with translationally invariant factor potentials U_M as they appear, for instance, in the SPC/Fw water model (see Section 2.4.1). It was one of the first ECMC variants and initially developed for the hard-disk system [115]. Afterwards, it was first generalized to smooth factor potentials that may depend on the separation between two interaction sites [121], and then to translationally invariant factor potentials that may depend on an arbitrary number of interaction sites [145]. In an event of the factor M, it just transfers the velocity $\mathbf{v}_a'$ of the previously active interaction site j to another interaction site i of M. Considering a finite velocity space $\mathcal{V}$, the lifting-move probabilities are given by

$$p_M^{\text{straight}}\left[(\mathbf{v}_a', j), (\mathbf{v}_a, i)\right] = \delta_{\mathbf{v}_a, \mathbf{v}_a'} \frac{\max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]}{\sum_{k \in \mathcal{N}} \max[0, \boldsymbol{\nabla}_{\mathbf{x}_k} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]}, \tag{2.55}$$

which trivially satisfy the necessary Eq. (2.53) for the lifted global-balance condition. They are also properly normalized:

$$\sum_{\mathbf{v}_a \in \mathcal{V}} \sum_{i \in \mathcal{N}} p_M^{\text{straight}}[(\mathbf{v}'_a, j), (\mathbf{v}_a, i)] = \frac{\sum_{i \in \mathcal{N}} \max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}'_a]}{\sum_{k \in \mathcal{N}} \max[0, \boldsymbol{\nabla}_{\mathbf{x}_k} U_M(\mathbf{x}_M) \cdot \mathbf{v}'_a]} = 1. \qquad (2.56)$$

Here, it was used that the translational invariance of the factor potential U_M implies $\sum_{i \in \mathcal{N}} \boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) = \mathbf{0}$ and thus

$$\sum_{i \in \mathcal{N}} \boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a = 0 \quad \text{for all } \mathbf{v}_a \in \mathcal{V}. \qquad (2.57)$$

Collecting positive and negative terms yields

$$\sum_{i \in \mathcal{N}} \max[0, \boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a] = \sum_{i \in \mathcal{N}} \max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a] \quad \text{for all } \mathbf{v}_a \in \mathcal{V}. \qquad (2.58)$$

In Eq. (2.55), the probability of the interaction site i becoming active in an event by the factor M is proportional to the negative rate of change $\max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]$ that its motion would induce. The active interaction site always changes in an event. For factors that depend on more than three interaction sites, other lifting schemes are available where the probability of i becoming active depends on the previously active interaction site j [146]. For the Coulomb factor between two SPC/Fw water molecules in Eq. (2.35), for instance, an "inside-first" lifting scheme keeps the active interaction site preferentially within the same water molecule, while an "outside-first" lifting scheme preferentially changes to an active interaction site in the other molecule.

2.4.7.2 Reflective Event-Chain Monte Carlo

The reflective ECMC variant considers the continuous set of velocities on the unit two-sphere $\mathcal{V} = S^2$. It was proposed together with the straight ECMC variant for the hard-disk system where it was found to show an inferior performance to decorrelate the global orientational order parameter [115]. As this doctoral thesis shows, however, reflective ECMC yields favorable rotation dynamics in hard-disk dipoles (see Chapter 4) and entirely different escape dynamics from relaxed hard-disk packings (see Chapter 5). As straight ECMC, reflective ECMC is valid for events from factors M with translationally invariant factor potentials U_M. Consider an event of the factor M where the interaction site j was previously active with velocity $\mathbf{v}'_a$. Reflective ECMC obtains a new velocity $\mathbf{v}_a$ by first reflecting $\mathbf{v}'_a$ with respect to the equipotential surface with normal vector $\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)$, and then inverting the reflected vector (see Fig. 1.5 for the example of the hard-disk system):

$$\mathbf{v}_a = -\mathbf{v}'_a + 2 \frac{\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}'_a}{|\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)|^2} \boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M). \qquad (2.59)$$

Since the factor M interrupted the motion of the previously active interaction site j, the corresponding event rate has to be greater than zero which implies $\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}'_a > 0$. Because of the inversion of the reflected velocity, the new velocity leaves the event rate invariant as $\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a = \boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}'_a$.

The lifting-move probabilities are given by

$$
\begin{aligned}
p_M^{\text{reflective}}[(\mathbf{v}_a', j), (\mathbf{v}_a, i)] &= \delta^{(3)}\left(\mathbf{v}_a + \mathbf{v}_a' - 2\frac{\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a'}{|\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)|^2} \boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)\right) \\
&\quad \times \frac{\max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]}{\sum_{k\in\mathcal{N}} \max[0, -\boldsymbol{\nabla}_{\mathbf{x}_k} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]}, \\
&= \delta^{(3)}\left(\mathbf{v}_a' + \mathbf{v}_a - 2\frac{\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a}{|\boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)|^2} \boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M)\right) \\
&\quad \times \frac{\max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]}{\sum_{k\in\mathcal{N}} \max[0, -\boldsymbol{\nabla}_{\mathbf{x}_k} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]},
\end{aligned}
\tag{2.60}
$$

where one uses that applying the reflection in Eq. (2.59) twice yields the original velocity. The lifting-move probabilities are properly normalized

$$
\sum_{i\in\mathcal{N}} \int_{\mathcal{V}} \mathrm{d}^3\mathbf{v}_a \, p_M^{\text{reflective}}[(\mathbf{v}_a', j), (\mathbf{v}_a, i)] = \frac{\sum_{i\in\mathcal{N}} \max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a(\mathbf{v}_a')]}{\sum_{k\in\mathcal{N}} \max[0, -\boldsymbol{\nabla}_{\mathbf{x}_k} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a(\mathbf{v}_a')]} = 1, \tag{2.61}
$$

where $\mathbf{v}_a(\mathbf{v}_a')$ is short for the right-hand side of Eq. (2.59) that only depends on $\mathbf{v}_a'$. By using Eq. (2.58) for the translationally invariant factor potential U_M, one can show that the necessary Eq. (2.53) for the lifted global-balance condition is also satisfied:

$$
\begin{aligned}
\sum_{j\in\mathcal{N}} \int_{\mathcal{V}} & \mathrm{d}^3\mathbf{v}_a' \, \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a'\right] \, p_M^{\text{reflective}}[(\mathbf{v}_a', j), (\mathbf{v}_a, i)] \\
&= \frac{\sum_{j\in\mathcal{N}} \max[0, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]}{\sum_{k\in\mathcal{N}} \max[0, -\boldsymbol{\nabla}_{\mathbf{x}_k} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]} \max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a] \\
&= \max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a].
\end{aligned}
\tag{2.62}
$$

In an event of the factor M, reflective ECMC reflects the velocity from $\mathbf{v}_a'$ to $\mathbf{v}_a$ by using Eq. (2.59). As in straight ECMC, the probability of the interaction site i becoming active is then proportional to the negative rate of change $\max[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M(\mathbf{x}_M) \cdot \mathbf{v}_a]$ that its motion would induce. The active interaction site always changes in an event.

Research Paper 1—Direction-Sweep Markov Chains

This chapter considers the publication

Section 3.0 provides a summary of the publication and contextualizes it within this doctoral thesis. Furthermore, the author's contributions are clearly indicated. The remaining part of this chapter reproduces the accepted manuscript of the publication itself. All co-authors of the publication agree with its use for this doctoral thesis and with the identification of the author's contributions.

3.0 Prologue

For sampling the canonical (*NVT*) ensemble of three-dimensional molecular systems in a periodic cubic box, JᴇʟʟʏFʏsʜ-Version1.0 implemented the straight event-chain Monte Carlo (ECMC) variant that always aligns the tree-dimensional velocity of the active interaction site with the positive coordinate axes (see Section 1.4). The finite velocity space is therefore given by $\mathcal{V} = \{(1, 0, 0)^T, (0, 1, 0)^T, (0, 0, 1)^T\}$ which, in a three-dimensional cubic box, contains the necessary three linear independent velocities to ensure an irreducible algorithm (see Section 2.4.4). The velocity $\mathbf{v}_a \in \mathcal{V}$ remains constant in an event chain, i.e., in between random resamplings from $\mathcal{V}$ in periodic time intervals of the chain time τ_{chain}. Only the active interaction site changes (see Fig. 1.4).

Although individual SPC/Fw water molecules resist rotation in large-scale simulations and, hence, the electric polarization remained dynamically arrested for prohibitively long computation times (see Section 1.3.2), small numbers of SPC/Fw water molecules do rotate within straight ECMC as implemented in JᴇʟʟʏFʏsʜ-Version1.0 [P6, 146]. The choice of a new velocity $\mathbf{v}_a$ from the finite velocity space $\mathcal{V}$ after each event chain need not be random in straight ECMC. In fact, choosing

a periodic cyclic sequence of the three possible velocities rotates a single SPC/Fw water molecule faster than a random sequence [146]. This idea is generalized to *direction sweeps* in the reproduced publication in this chapter. The direction of a local Monte Carlo move is understood as a lifting variable in the lifting framework (see Section 2.3) and can thus be applied to various local Markov-chain Monte Carlo (MCMC) algorithms such as the Metropolis or ECMC algorithms. For the example of straight ECMC in a two-dimensional system, a direction sweep deterministically rotates the velocity by an angle increment $\Delta\phi$ after each event chain. Similarly, the Metropolis algorithm deterministically rotates the direction of its Monte Carlo moves that propose a new position of a random interaction site along the current direction in a symmetric distribution about the original position.

We study direction sweeps for a single tethered hard-disk dipole that serves as a simplified two-dimensional model of an extended dipole with flexible bond length and resembles the complex SPC/Fw water molecule. For a certain reversible MCMC algorithm (that is a limit of the local Metropolis and ECMC algorithms), we show analytically that slow non-reversible direction sweeps with angle increments $\Delta\phi \to 0$ profoundly change the rotation dynamics of the tethered hard-disk dipole (see Fig. 3.3). Specifically, we show that the dipole balances between persistent rotations in both directions instead of showing a diffusive motion in its rotation dynamics. The persistent rotations of the dipole are closely connected to the dynamics of an impact parameter that, depending of the direction of the rotation, is described by a Langevin equation [167] or a non-linear differential equation.

For this simple system, we also numerically compute the mixing time (see Section 2.2.1). This reveals that picking the directions only along the coordinate axes as in JELLYFYSH-Version1.0, be it periodically or randomly, leads by far to the largest mixing time. A slow direction sweep, or even a completely random choice of the direction, much decreases mixing times and rotates the tethered hard-disk dipole considerably faster. In the context of the overarching objective of this doctoral thesis—the introduction of a competitive, rigorous paradigm for molecular simulations that is based on non-reversible MCMC algorithms—this is a first hint that the particular straight ECMC variant of JELLYFYSH-Version1.0 was particularly ill-suited for decorrelating the polarization of SPC/Fw water molecules. In order to overcome the observed dynamical arrest of the polarization, one should thus refrain from restricting the velocity of the interaction site to a small number of possible velocities.

In a more general context, the direction sweeps are reminiscent of the sequential sweeps in the Monte Carlo moves of, for instance, the local Metropolis algorithm: Instead of proposing a move for a random "active" interaction site of the system at each Monte Carlo time step, a move is attempted for each site in succession. Once the active site is fixed, its move is sampled with the reversible Metropolis algorithm. Such sequential schemes are in wide use because they are somewhat faster [168–172] and were, in fact, already used in the seminal introduction of the Metropolis algorithm in 1953 [72]. Yet, they are clearly non-reversible because a sequential sweep in the reverse order never occurs [173]. In other words, they apparently introduce a memory (or analogously a time-dependency) into the Markov chain because it remembers the direction of the sweep. This would break the defining property of a Markov chain and render its rigorous theoretical foundation useless. The lifting framework resolves that issue. It is used in the appendix of the reproduced publication in this chapter to prove the correctness of site-sweep versions of local reversible MCMC algorithms as long as the site-sweep version is still irreducible (see Appendix 3.A.2; another proof without the lifting framework can be found in Ref. [173]).

Direction sweeps are just as general as sequential site sweeps. From an application point-of-view, the necessary optimization of the angle increment $\Delta\phi$ is a downside of the method. Also, as the dimension of the system is increased, even more intrinsic parameter would emerge. Still, the essential lesson of avoiding small sets of possible directions (as in JeLLyFysh-V1.0) is a valuable insight. Also, in comparison to a completely random choice of the direction, any value of $\Delta\phi$ that leads to a large set of possible directions already shows slightly reduced mixing times (see Fig. 3.6). Moreover, from the perspective of a theorist, the simple tethered hard-disk dipole model allows for an analytical proof that non-reversibility profoundly changes basic properties of local MCMC algorithms, and for an implementation of a numerical protocol for the computation of mixing times which, even though they are the primary characteristic of a Markov chain, is impossible in most practical applications.

Contributions of the Author

The author contributed in an essential manner to the entire research project underlying this research paper. This includes, in particular, essential contributions to the development of the tethered hard-disk dipole model, the development of direction sweeps, the analytical proofs, the software development for the numerical protocol for the computation of mixing times and the data collection, the creation of the figures, and the writing of the research paper itself. Following the Contributor Role Taxonomy (CRediT) [174], the author contributions are as follows:

- Conceptualization (ideas; formulation or evolution of overarching research goals and aims).

- Methodology (development or design of methodology; creation of models).

- Software (programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components).

- Validation (verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs).

- Formal analysis (application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data).

- Investigation (conducting a research and investigation process, specifically performing the experiments, or data/evidence collection).

- Writing—Original Draft [preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)].

- Writing—Review & Editing (preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision—including pre- or post-publication stages).

- Visualization (preparation, creation and/or presentation of the published work, specifically visualization/data presentation).

Direction-Sweep Markov Chains

Abstract   We discuss a non-reversible, lifted Markov-chain Monte Carlo (MCMC) algorithm for particle systems in which the direction of proposed displacements is changed deterministically. This algorithm sweeps through directions analogously to the popular MCMC sweep methods for particle or spin indices. Direction-sweep MCMC can be applied to a wide range of reversible or non-reversible Markov chains, such as the Metropolis algorithm or the event-chain Monte Carlo algorithm. For a single two-dimensional tethered hard-disk dipole, we consider direction-sweep MCMC in the limit where restricted equilibrium is reached among the accessible configurations for a fixed direction before incrementing it. We show rigorously that direction-sweep MCMC leaves the stationary probability distribution unchanged and that it profoundly modifies the Markov-chain trajectory. Long excursions, with persistent rotation in one direction, alternate with long sequences of rapid zigzags resulting in persistent rotation in the opposite direction in the limit of small direction increments. The mapping to a Langevin equation then yields the exact scaling of excursions while the zigzags are described through a non-linear differential equation that is solved exactly. We show that the direction-sweep algorithm can have shorter mixing times than the algorithms with random updates of directions. We point out possible applications of direction-sweep MCMC in polymer physics and in molecular simulation.

3.1 Introduction

Since its introduction in 1953, the Markov-chain Monte Carlo (MCMC) method [72] has developed into an essential tool in science and engineering, and into a prominent mathematical research discipline [74]. MCMC is concerned with the sampling of a probability distribution π, for example a Boltzmann distribution in equilibrium statistical physics. MCMC's trademark properties are randomness and absence of memory: Samples j at Monte-Carlo time step $t + 1$ are produced from samples i at time step t with independent probabilities contained in a transition matrix $P = (P_{i,j})$. The stationary distribution π is reached in the limit $t \rightarrow \infty$. It satisfies the global-balance condition $\sum_i \pi_i P_{i,j} = \pi_j$. Most MCMC algorithms are reversible. They satisfy the detailed-balance condition $\pi_i P_{i,j} = \pi_j P_{j,i}$ that implies the less restrictive global balance by summing over i. In physical terms, a reversible Markov chain implements an equilibrium dynamics that approaches the Boltzmann distribution, with the detailed-balance condition expressing the vanishing of all flows. In contrast, a non-reversible Markov chain implements out-of-equilibrium dynamics with a steady state (imposed by the global-balance condition) that again coincides with the Boltzmann distribution π.

Under the necessary conditions of irreducibility and aperiodicity [74], MCMC algorithms often allow for sequential variants that seemingly introduce memory effects to the move (the move at time $t + 1$ may depend on the move at time t). For systems of N particles, one such variant was pioneered in the original 1953 reference: Instead of attempting at time t a move of a randomly chosen particle ". . . we move each of the particles in succession . . . " [72, p. 22], that is, attempt a move of particle $i + 1$ (modulo N) at time $t + 1$ after an attempt of i at time t. In particle systems with central potentials, this non-reversible particle-sweep Monte Carlo algorithm is somewhat faster (apparently by a constant factor) than the reversible random-choice variant [169–172]. In the Ising model and related systems, sequential updates of spin $i + 1$ after spin i, etc., ("spin sweeps") also break detailed balance yet satisfy global balance. Spin-sweep algorithms are again faster, by a constant factor, than the detailed-balance MCMC that flip spins in random order [168].

Lifting [111] allows one to formulate such a partly deterministic algorithm as a Markov chain with a time-independent transition matrix, and to expose its close connection with the "collapsed" Markov chain in which moves are proposed randomly. The lifted transition matrix acts on lifted (extended) samples. In the above example of particle-sweep or spin-sweep algorithms, the so-called "lifted sample" comprises the active-particle index in addition to all the particle coordinates. The particle- and spin-sweep algorithms are precursors of non-reversible Markov chains that have been much studied in the recent past [109, 118, 175–177]. One powerful non-reversible MCMC method is the event-chain Monte Carlo (ECMC) algorithm [80, 115, 121].

In more than one dimension, naturally, particles must move into different spatial directions. Rather than to sample the direction of the proposed move at time t randomly, one may define it as a lifting variable, and modify it deterministically without influencing the stationary distribution. In the same way in which, in a particle-sweep algorithm, particles i, $i + 1$, and then $i + 2$, etc., move in succession, we can force all moves to go forward or backward in direction ϕ, and then (for a two-dimensional case) in direction $\phi + \Delta\phi$, then $\phi + 2\Delta\phi$, etc. This "direction sweep" appears at first sight as deterministic, but the lifting framework again allows it to be formulated as a memory-less Markov process by combining the configuration and the direction of proposed moves into a lifted sample.

In the family of ECMC algorithms, direction lifting can be applied to straight ECMC that uses the same direction for every move in an event chain (in contrast to, e.g., reflective [115], Newtonian [124] and forward ECMC [150] that change the direction at each event). Very slow changes of the direction after each event chain were studied for straight ECMC simulations of hard-sphere systems, where they were not found to improve the convergence properties [178].

In the present paper, we discuss direction-sweep MCMC for a tethered hard-disk dipole as a simplified two-dimensional model of an extended flexible dipole consisting of two atoms, that resembles a flexible water molecule in the context of molecular simulation (see Refs [P6, 146]). The variable dipole length allows the entire dipole to rotate through local MCMC moves of both atoms along slowly changing directions. We obtain exact results for direction sweeps in the limit where a given direction remains fixed until "restricted equilibrium" is reached among the accessible configurations by the application of any local MCMC algorithm (whose moves can be constructed with finite acceptance probability from a sequence of infinitesimal moves). Here, restricted equilibrium is understood as the stationary conditional probability density on the line of accessible configurations which is induced by the full equilibrium distribution π. We show analytically that slow direction sweeps (that is, small direction-

angle increments) yield long-lived rotations of the dipole by a cumulative ("rolled-out") mean rotation angle that diverges as the inverse angle increment in the limit of infinitely slow sweeps. This motion is relayed by a counter-rotation that proceeds in a fast sequence of steps. Both motions exactly balance, and the expectation of the net rotation vanishes. Numerically, we show that direction sweeps can lower mixing times compared to MCMC with random choice of the direction. We find that only picking directions among the x- and y-axes is by far the most unfavorable case, although it was previously used in applications of straight ECMC to dipolar systems [P6, 146]. Our results, for dipoles, differ from those for hard-sphere systems and probably, more generally, from those for simple liquids [178].

Our simple dipole model serves as an analytically tractable test bed for molecules such as the simple-point-charge-with-flexible-water (SPC/Fw) water model [28]. The model is also very closely related to the flexible polymer models that are being intensely studied using ECMC [131, 132]. Direction lifting is insensitive to the dipole's structure and interactions. It remains valid for N-particle systems, even though in that case the dynamics can no longer be analyzed analytically [P2].

The paper is organized as follows: Section 3.2 introduces the dipole model. In Section 3.3, we introduce direction-sweep MCMC that reaches restricted equilibrium in a single step, and prove that it converges towards the stationary distribution π. Section 3.4 studies the trajectory of the dipole orientation. Mixing times are determined and compared in Section 3.5. Section 3.6 provides a summary of our main results. Appendix 3.A provides additional context on lifted Markov chains and particle-sweep algorithms.

3.2 Dipole Model

We study a two-dimensional flexible dipole consisting of two atoms 1 and 2 with an interaction that only enforces a minimum length r and a maximum length R. Specifically, the two atoms are at positions $\mathbf{x}_1$ and $\mathbf{x}_2$ in a two-dimensional homogeneous domain. The flat interaction

$$U(\rho) = \begin{cases} 0 & \text{if } \rho \in [r, R], \\ \infty & \text{otherwise,} \end{cases} \tag{3.1}$$

only depends on the dipole length $\rho = |\mathbf{x}_2 - \mathbf{x}_1|$. The two-dimensional dipole thus consists of two tethered hard disks with a contact distance r and tether length R. The model can be generalized to three spatial dimensions, and also made more realistic, for example through the SPC/Fw water model [28]. The single dipole of Eq. (3.1) is to be envisaged as part of a more complex many-dipole system with hard-sphere atomic pair interactions (see Fig. 3.1a) that is studied elsewhere [P2].

Any local MCMC move proposes a displacement of $\mathbf{x}_1$ or $\mathbf{x}_2$ from its present position along a line $\mathbf{l}$ of angle ϕ with the x-axis. If the reached configuration has a dipole length with infinite U, the move is rejected. Such single-atom moves induce translations and rotations of the entire dipole. We do not consider explicit global MCMC translations or global rotations because, in the application to ECMC that we have in mind, they are difficult to implement. Because of homogeneity, uniform translations of the dipole decouple from its rotations. We may thus set $\mathbf{x}_1 = (0, 0)$ and consider the relative position in polar coordinates: $\mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1 = (\rho, \theta)$, with the dipole angle θ. In
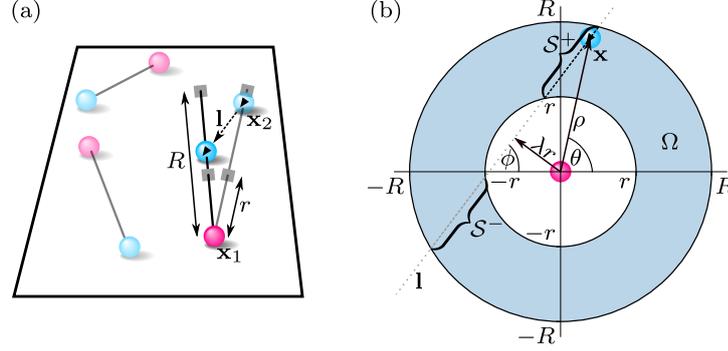
Figure 3.1: Two-dimensional dipole of length $\rho \in [r, R]$. (a): Local MCMC move of a dipole atom at position $\mathbf{x}_2$ along the line $\mathbf{l}$. (b): The same move in the ring representation $\mathbf{x}_2 - \mathbf{x}_1 = (\rho, \theta) \in \Omega$, with the dipole angle θ and the sample space Ω. The line $\mathbf{l}$ has angle ϕ with the x-axis and distance λr to the origin. For impact parameters $|\lambda| < 1$, the line $\mathbf{l}$ forms two segments $\mathcal{S}^-, \mathcal{S}^+ \subset \mathbf{l} \cap \Omega$ in the sample space.

equilibrium, $\mathbf{x}$ is uniformly distributed on the sample space Ω from which translations have been factored out, namely the two-dimensional ring of inner radius r and outer radius R centered at $(0, 0)$ (see Fig. 3.1b). The uniform Euclidean distribution on Ω translates into a dipole-length distribution $\pi(\rho) = 2\rho/[r^2(\eta^2 - 1)]$ for $\rho \in [r, R]$ with the tether ratio $\eta = R/r$, and a dipole-angle distribution for θ that is uniform in $(-\pi, \pi]$.

The sampling of the dipole $(\mathbf{x}_1, \mathbf{x}_2)$ may be tracked in the ring representation (for $\mathbf{x} \in \Omega$), because any move of $\mathbf{x}_2$ corresponds to the identical move for $\mathbf{x}$, and any move of $\mathbf{x}_1$ yields an inverse move for $\mathbf{x}$. In both cases, the move is on a line $\mathbf{l} \ni \mathbf{x}$ of angle ϕ with the x-axis. We thus parameterize a direction of a local MCMC move by ϕ. The trajectory of the dipole angle under slow direction sweeps is closely connected to the trajectory of the impact parameter

$$\lambda = \frac{\sin(\theta - \phi)\rho}{r}, \tag{3.2}$$

which denotes the signed distance (in units of r) of $\mathbf{l}$ to the origin in a local MCMC move. In the reference frame where $\phi = 0$, i.e., where $\mathbf{l}$ runs parallel to the x-axis, λ is positive for $\mathbf{x}$ in the upper half plane and negative in the lower half plane. If $|\lambda| > 1$, $\mathbf{l} \cap \Omega$ forms a single segment $\mathcal{S}$ that contains all accessible configurations. If $|\lambda| < 1$, $\mathbf{l} \cap \Omega$ forms two such segments, namely $\mathcal{S}^-$ (to the left for $\phi = 0$) and $\mathcal{S}^+$ (to the right) (see Fig. 3.1b again). In realistic applications like, e.g., dense dipole systems, accepted local MCMC moves with $|\lambda| < 1$ that jump from $\mathcal{S}^-$ to $\mathcal{S}^+$ are highly unlikely. We thus only consider local MCMC moves that remain within their respective segment ($\mathcal{S}, \mathcal{S}^-$ or $\mathcal{S}^+$) or, in other words, local MCMC moves that can be constructed from infinitesimal legal moves.

3.3 Direction-Sweep MCMC

Repeated local MCMC moves along one direction ϕ, using, e.g., ECMC or the Metropolis algorithm with small step sizes, tend towards a restricted equilibrium among the accessible configurations in $\mathcal{S}, \mathcal{S}^-$ or $\mathcal{S}^+$. In other words, the distribution of the position on the segment approaches the
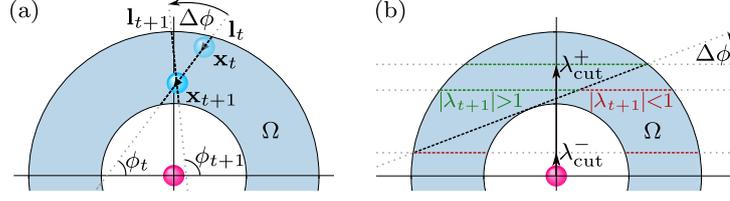
Figure 3.2: Direction-sweep MCMC with subsequent directions $\phi_{t+1} - \phi_t$ differing by the angle increment $\Delta\phi$. (a): The position $\mathbf{x}_{t+1}$ is a direct uniform sample on the line $\mathbf{l}_t$ that contains the previous $\mathbf{x}_t$ and has angle ϕ_t with the x-axis. The line $\mathbf{l}_{t+1}$ of the subsequent move and $\mathbf{l}_t$ intersect at $\mathbf{x}_{t+1}$. (b): A large impact parameter ($|\lambda_t| > \lambda_{\text{cut}}^+$) always leads to $|\lambda_{t+1}| > 1 \quad \forall \mathbf{x}_{t+1} \in \mathbf{l}_t$. Likewise, a small impact parameter ($|\lambda_t| < \lambda_{\text{cut}}^-$) leads to $|\lambda_{t+1}| < 1 \quad \forall \mathbf{x}_{t+1} \in \mathbf{l}_t$. Otherwise, if $\lambda_{\text{cut}}^- < |\lambda_t| < \lambda_{\text{cut}}^+$, the value of $\mathbf{x}_{t+1}$ determines whether $|\lambda_{t+1}| > 1$ or $|\lambda_{t+1}| < 1$.

stationary conditional probability density that is induced by the equilibrium distribution π. (The segment is to be produced in the limit of successively narrower strips in order to properly define the conditional probability density [179, Chapter 15].) For the single two-dimensional dipole, the stationary distribution on the segment is uniform, and the restricted-equilibrium limit can be reached in a single step by sampling the final position of the dipole in the segment that contains the starting position $\mathbf{x}$. This allows us to focus on the effects introduced by the choice of directions. In the following, one unit of MCMC time corresponds to one fixed direction. We therefore obtain the next position $\mathbf{x}_{t+1}$ as a direct uniform sample [78] in $\mathcal{S}$ for $|\lambda_t| > 1$, and in $\mathcal{S}^-$ or in $\mathcal{S}^+$ (depending on the starting position $\mathbf{x}_t$) for $|\lambda_t| < 1$. The direction is incremented by a constant value after each time step, and the line $\mathbf{l}_{t+1}$ thus goes through $\mathbf{x}_{t+1}$ with the new angle $\phi_{t+1} = \phi_t + \Delta\phi$ with the angle increment $\Delta\phi$ (for concreteness, we suppose $\Delta\phi > 0$) (see Fig. 3.2a).

Because of the π-periodicity of the directions, a given value of the angle increment $\Delta\phi$ and the starting direction ϕ_0 imply a set of directions $\mathcal{D} = \{\phi_t\}$ that contains all possible directions of a simulation. We only consider finite direction sets $\mathcal{D}$ for simplicity. Direction-sweep MCMC is a non-reversible lifting of reversible local MCMC with a lifted sample space $\widehat{\Omega} = \Omega \times \mathcal{D}$. The lifted stationary distribution $\hat{\pi}_{(\mathbf{x},\phi)}$ incorporates the lifting variable ϕ with $\sum_\phi \hat{\pi}_{(\mathbf{x},\phi)} = \pi_{\mathbf{x}}$ (see Appendix 3.A.1 for definitions of lifted Markov chains). We will show below that $\hat{\pi}_{(\mathbf{x},\phi)}$ is proportional to $\pi_{\mathbf{x}}$. In order to converge towards the stationary distribution π, the direction-sweep MCMC algorithm must satisfy the global-balance condition:

$$\hat{\pi}_{(\mathbf{x}_{t+1},\phi_{t+1})} = \sum_{\mathbf{x}_t,\phi_t} \hat{\pi}_{(\mathbf{x}_t,\phi_t)} \widehat{P}_{(\mathbf{x}_t,\phi_t),(\mathbf{x}_{t+1},\phi_{t+1})}, \qquad (3.3)$$

where $\widehat{P}_{(\mathbf{x},\phi),(\mathbf{x}',\phi')}$ denotes the conditional transition probability to the lifted sample $(\mathbf{x}',\phi')$, given that the lifted Markov chain is already at $(\mathbf{x},\phi)$ with, in our case, $\phi' = \phi + \Delta\phi$. For the collapsed Markov chain, ϕ' is a random element in $\mathcal{D}$.

As explained, any move from t to $t+1$ is composed of two parts, that we describe by two lifted transition matrices, namely $\widehat{P}^{(1)}$ (that takes a half step from t to $t+1/2$) and $\widehat{P}^{(2)}$ (half step from $t+1/2$ to $t+1$). They combine as $\widehat{P} = \widehat{P}^{(1)} \widehat{P}^{(2)}$. In the first half step, the lifting variable is fixed ($\phi_{t+1/2} = \phi_t = \phi$) while a new position $\mathbf{x}_{t+1/2}$ is directly sampled among the accessible

configurations given $\mathbf{x}_t$ and ϕ_t. This restricted MCMC algorithm satisfies global balance for the given direction ϕ by construction, that is, $\hat{\pi}_{(\mathbf{x}_{t+1/2},\phi)} = \sum_{\mathbf{x}_t} \hat{\pi}_{(\mathbf{x}_t,\phi)} \widehat{P}^{(1)}_{(\mathbf{x}_t,\phi),(\mathbf{x}_{t+1/2},\phi)}$. This yields specifically $\hat{\pi}_{(\mathbf{x}_{t+1/2},\phi)} = \hat{\pi}_{(\mathbf{x}_t,\phi)}$ for the simple dipole model with its uniform π. In the second half step, the position $\mathbf{x}$ remains fixed and only the direction is incremented by $\Delta\phi$:

$$\hat{\pi}_{(\mathbf{x},\phi_{t+1})} = \sum_{\phi_{t+1/2}} \hat{\pi}_{(\mathbf{x},\phi_{t+1/2})} \widehat{P}^{(2)}_{(\mathbf{x},\phi_{t+1/2}),(\mathbf{x},\phi_{t+1})} = \hat{\pi}_{(\mathbf{x},\phi_{t+1}-\Delta\phi)}. \tag{3.4}$$

For a finite direction set, and taking into account the periodicity of directions, this establishes that the lifted stationary distribution $\hat{\pi}_{(\mathbf{x},\phi)}$ is independent of ϕ. The two half steps together establish the validity of the global-balance condition in Eq. (3.3). For the two-dimensional dipole model, direction-sweep MCMC is aperiodic and irreducible (that is, ergodic) for any choice of two or more directions. The irreducibility of the lifted Markov chain in the lifted sample space $\widehat{\Omega}$ follows from the fact that any two positions $\mathbf{x}$ and $\mathbf{x}'$ in the ring Ω can be joined by lines with the angles available in the set $\mathcal{D}$. Moreover, $\mathbf{x}$ and $\mathbf{x}'$ may coincide for a given move, so that for any given sample $\mathbf{x} \in \Omega$, all the corresponding lifted samples $(\mathbf{x},\phi) \in \widehat{\Omega}$ can be realized. Aperiodicity follows again from the possible coincidence of $\mathbf{x}$ and $\mathbf{x}'$ in a direction-sweep MCMC move. The irreducibility in the lifted sample space is sometimes more difficult to verify [P2]. The independence of the stationary distribution $\hat{\pi}_{(\mathbf{x},\phi)}$ with respect to the lifting variable is a general property of lifted MCMC [80] (see also Appendix 3.A.2 for the simple example of particle-sweep algorithms).

For small angle increments $\Delta\phi$, direction-sweep MCMC features two cutoff impact parameters λ^+_{cut} and λ^-_{cut}:

$$\lambda^\pm_{\text{cut}} = \cos(\Delta\phi) \pm \sqrt{(\eta^2 - 1)\sin^2(\Delta\phi)}. \tag{3.5}$$

Here, λ^+_{cut} exists for $\cos(\Delta\phi) > 1/\eta$, and λ^-_{cut} for $\cos(\Delta\phi) > (\eta^2 - 2)/\eta^2$. For a fixed value of η, $\lambda^\pm_{\text{cut}}$ both exist in the limit $\Delta\phi \to 1$ and approach 1 as $\lambda^\pm_{\text{cut}} \to 1^\pm$. The interval $[\lambda^-_{\text{cut}}, \lambda^+_{\text{cut}}]$ is a separation layer for the impact parameter λ because $|\lambda_t| > \lambda^+_{\text{cut}}$ implies $|\lambda_{t+1}| > 1$, whereas $|\lambda_t| < \lambda^-_{\text{cut}}$ implies $|\lambda_{t+1}| < 1$ (see Fig. 3.2b).

3.4 Equilibrium Properties

For small angle increments $\Delta\phi > 0$, direction-sweep MCMC simulations of the single two-dimensional dipole yield trajectories of $\tilde{\theta}_t$, the rolled-out dipole angle (not wrapped back into a 2π interval), with alternating positive and negative rotations. The positive rotations fluctuate around their average of $\Delta\phi$ per time step, so that the trajectory of $(\tilde{\theta}_t - \tilde{\theta}_0)\Delta\phi$ vs. $(\tilde{\phi}_t - \tilde{\phi}_0)\Delta\phi$ has an average unit slope ($\tilde{\phi}_t$ is the rolled-out direction: $\tilde{\phi}_t - \tilde{\phi}_0 = t\Delta\phi$). The negative rotations exhibit, in contrast, intermittent sharp decreasing steps and constant plateaus (see Fig. 3.3).

We will show that the trajectory of the rolled-out dipole angle $\tilde{\theta}_t$ depends on that of the impact parameter λ_t (see Sections 3.4.1 and 3.4.2). Therefore, we first treat the trajectory of λ_t that may be described through drift and diffusion terms. The current value of the impact parameter λ_t is a function of the position $\mathbf{x}_t$ and direction ϕ_t (for fixed system parameters). Direction-sweep MCMC algorithm then uniformly samples $\mathbf{x}_{t+1}$ on the segment that also contains $\mathbf{x}_t$. The subsequent increment of the
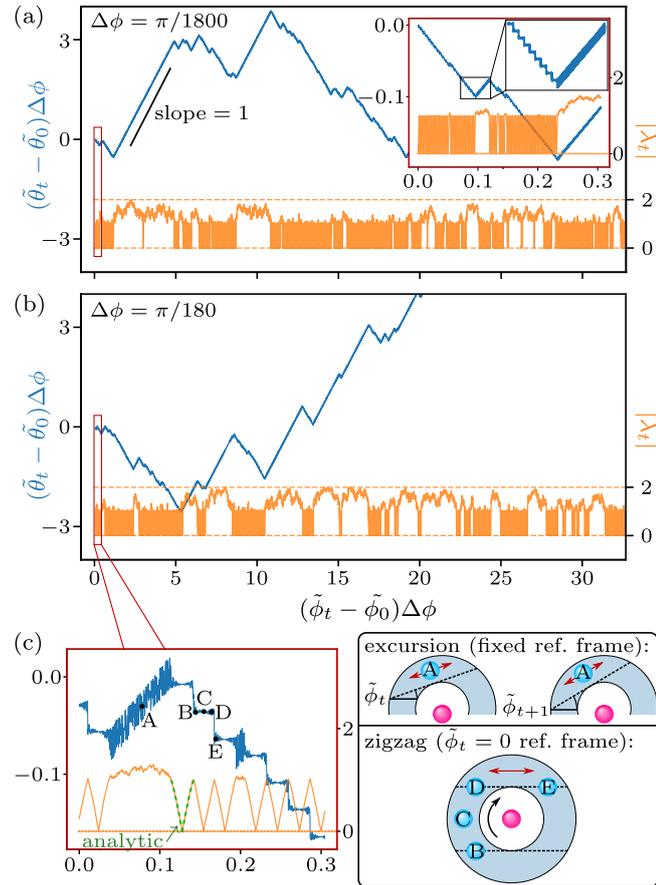
Figure 3.3: Direction-sweep MCMC trajectories of the rescaled rolled-out dipole angle $(\tilde{\theta}_t - \tilde{\theta}_0)\Delta\phi$ (blue, upper) and of the impact parameter $|\lambda_t|$ (yellow, lower) vs. the rescaled change of the rolled-out direction $(\tilde{\phi}_t - \tilde{\phi}_0)\Delta\phi$ for the tether ratio $\eta = 2$. (a): Angle increment $\Delta\phi = \pi/1800$. The inset shows the initial trajectories. (b): $\Delta\phi = \pi/180$. (c): Initial trajectories for $\Delta\phi = \pi/180$. Points A picture the trajectory during an "excursion" (see Section 3.4.1). During a "zigzag", the trajectory leads from point B over C to D in the reference frame with fixed direction $\tilde{\phi}_t = 0$ (see Section 3.4.2). After a zigzag, the trajectory either jumps from D to E into a new zigzag or switches over to an excursion (see Section 3.4.3).

direction by $\Delta\phi$ linearly maps $\mathbf{x}_{t+1}$ onto the impact parameter λ_{t+1}. The position $\mathbf{x}_{t+1}$, given $\mathbf{x}_t$, is a random variable, and so is λ_{t+1}. Its conditional expectation is

$$\mathbb{E}\left(\lambda_{t+1}|\lambda_t\right) = \begin{cases} \lambda_t \cos \Delta\phi \mp A_t \sin \Delta\phi & \text{if } |\lambda_t| < 1; \mathbf{x}_t \in \mathcal{S}^{\pm}, \\ \lambda_t \cos \Delta\phi & \text{if } |\lambda_t| > 1, \end{cases} \tag{3.6}$$

where $A_t = (B_t + C_t)/2$ with $B_t = \sqrt{\eta^2 - \lambda_t^2}$ and $C_t = \sqrt{1 - \lambda_t^2}$. The variance of λ_{t+1} is

$$\sigma^2 \left(\lambda_{t+1}|\lambda_t\right) = \begin{cases} \frac{1}{12} \sin^2 \left(\Delta\phi\right) \left(B_t - C_t\right)^2 & \text{if } |\lambda_t| < 1, \\ \frac{1}{3} \sin^2 \left(\Delta\phi\right) B_t^2 & \text{if } |\lambda_t| > 1. \end{cases} \tag{3.7}$$

Equations (3.6) and (3.7) are the expectation and variance of a uniform distribution of λ_{t+1}.

For small angle increments $\Delta\phi > 0$, the trajectories of the impact parameter of direction-sweep MCMC are exactly equivalent to those of a Gaussian process:

$$\lambda_{t+1} = \lambda_t + \left[\mathbb{E}\left(\lambda_{t+1}|\lambda_t\right) - \lambda_t\right] + \sqrt{\sigma^2 \left(\lambda_{t+1}|\lambda_t\right)} \, w_t, \tag{3.8}$$

where w_t samples a standard normal distribution. For small $\Delta\phi$, the fluctuation term $\sqrt{\sigma^2(\lambda_{t+1}|\lambda_t)}$ is proportional to $\Delta\phi$. The drift $\mathbb{E}(\lambda_{t+1}|\lambda_t) - \lambda_t$ is proportional to $(\Delta\phi)^2$ for $|\lambda| > 1$ and proportional to $\Delta\phi$ for $|\lambda| < 1$. This leads to distinctive dynamics for $|\lambda| > 1$ and for $|\lambda| < 1$,

3.4.1 Excursions ($|\lambda| > 1$)

For impact parameters $|\lambda_t| > 1$, Eq. (3.8) becomes

$$\lambda_{t+1} = \lambda_t - \frac{\lambda_t}{2}(\Delta\phi)^2 + \sqrt{\frac{1}{3}B_t^2(\Delta\phi)^2} \, w_t \tag{3.9}$$

in the limit of small angle increments $\Delta\phi$. This equation agrees with the discrete-time Langevin equation

$$\lambda_{t+1} = \lambda_t + D^{(1)}(\lambda_t, t_t) \, \tau + \sqrt{2D^{(2)}(\lambda_t, t_t) \, \tau} \, w_t, \tag{3.10}$$

where the discrete times t_t are separated by the time step τ, and $D^{(1)}$ and $D^{(2)}$ are the Kramers–Moyal expansion coefficients of the time-dependent probability distribution of λ that correspond to drift and diffusion, respectively (see [167, eq. (3.138)]). As long as $|\lambda_t| > 1$, the impact parameter thus performs an "excursion," a random walk in the quantity $\tau = (\Delta\phi)^2$. Such an excursion corresponds to a total number of time steps that scales as $\sim \text{const}/(\Delta\phi)^2$. As each time step increases the direction by $\Delta\phi$, this increases the rolled-out direction $\tilde{\phi}_t$ during the excursion by an amount that diverges as $\sim \text{const}/\Delta\phi$. The excursions with $|\lambda_t| > 1$ for different small angle increments $\Delta\phi$ therefore become similar if the trajectory λ_t of the impact parameter is recorded as a function of $(\tilde{\phi}_t - \tilde{\phi}_0)\Delta\phi$ (see Fig. 3.3a and b).

During the excursion, the position $\mathbf{x}_t$ follows on average the rotation of the direction in steps of the angle increment $\Delta\phi$ (see points A in Fig. 3.3c). Thus, the difference in the rolled-out dipole angle $\tilde{\theta}_t$

during an excursion diverges as $\sim \mathrm{const}/\Delta\phi$. For $\Delta\phi > 0$, this is evidenced by the unit slope of the increasing parts of the trajectories of $(\tilde{\theta}_t - \tilde{\theta}_0)\Delta\phi$ as a function of $(\tilde{\phi}_t - \tilde{\phi}_0)\Delta\phi$ (see Fig. 3.3a and b, again).

3.4.2 Zigzags ($|\lambda| < 1$)

For impact parameters $|\lambda_t| < 1$ and small angle increments $\Delta\phi$, the fluctuations of the impact parameter in Eq. (3.7) are negligible in comparison to the drift in Eq. (3.6). The Gaussian process of Eq. (3.8) then approaches its deterministic limit:

$$\Delta\lambda = \lambda_{t+1} - \lambda_t = \mp \frac{\sqrt{\eta^2 - \lambda_t^2} + \sqrt{1 - \lambda_t^2}}{2}\,\Delta\phi, \qquad (3.11)$$

where the negative sign on the right-hand side is for $\mathbf{x}_t \in \mathcal{S}^+$, and the positive sign for $\mathbf{x}_t \in \mathcal{S}^-$. For small $\Delta\phi$, this becomes a non-linear differential equation whose solution (up to an integration constant) is

$$\phi(\lambda) = \pm \frac{\arcsin(\lambda) - \eta^2 \arcsin(\lambda/\eta) + \lambda[C(\lambda) - B(\lambda)]}{\eta^2 - 1}, \qquad (3.12)$$

where $B(\lambda) = \sqrt{\eta^2 - \lambda^2}$ and $C(\lambda) = \sqrt{1 - \lambda^2}$. The numerical trajectories (for $|\lambda| < 1$) reproduce the exact solution of Eq. (3.12) (see green curve in Fig. 3.3c).

For small angle increments $\Delta\phi > 0$, if $\mathbf{x}$ enters the segment $\mathcal{S}^+$ with unit impact parameter $\lambda = 1$, λ will decrease at every step until $\lambda = -1$. In $\mathcal{S}^-$, the impact parameter likewise increases from $\lambda = -1$ to $\lambda = 1$. This trapped deterministic motion creates one "zigzag" in the trajectory of $|\lambda|$. The total change of the rolled-out direction $\Delta\tilde{\phi}_{\mathrm{ZZ}}$ during such a zigzag is given by

$$\Delta\tilde{\phi}_{\mathrm{ZZ}} = \frac{2\eta^2 \mathrm{arccsc}(\eta) + 2\sqrt{\eta^2 - 1} - \pi}{\eta^2 - 1}, \qquad (3.13)$$

and is always smaller than π. In a reference frame with fixed direction $\phi_t = 0$ (which is rotated by $-\Delta\phi$ at every time step), the position $\mathbf{x}_t$ performs a negative rotation (see the trajectory from point B over C to D in Fig. 3.3c), and thus follows the rotation of the system. Therefore, the rolled-out dipole angle $\tilde{\theta}_t$ remains roughly constant, leading to a plateau in the non-rotating reference frame. For $\lambda_t = 0$, the dipole angle $\tilde{\theta}_t$ is independent of the precise position $\mathbf{x}_t$ on its segment. At the center of each plateau, the fluctuations of $\tilde{\theta}_t$ thus vanish even at finite $\Delta\phi$ (see point C in Fig. 3.3c).

Since the total change of the rolled-out direction $\Delta\tilde{\phi}_{\mathrm{ZZ}}$ in Eq. (3.13) is independent of $\Delta\phi$, the zigzags in the trajectories of $|\lambda_t|$ for different (small) values of $\Delta\phi$ are similar if plotted as a function of $\tilde{\phi}_t - \tilde{\phi}_0$, but ever steeper as a function of $(\tilde{\phi}_t - \tilde{\phi}_0)\Delta\phi$ (see Fig. 3.3a and b).

3.4.3 Interplay of Excursions and Zigzags

After one rapid motion from $\lambda = 1$ to $\lambda = -1$ or *vice versa*, the trajectory may switch segments to continue at $|\lambda| < 1$ (see the jump from point D to E in Fig. 3.3c), adding one more leg to the negative-rotation zigzag of λ_t (or two legs to the zigzag of $|\lambda_t|$ in Fig. 3.3c). The trajectory may also
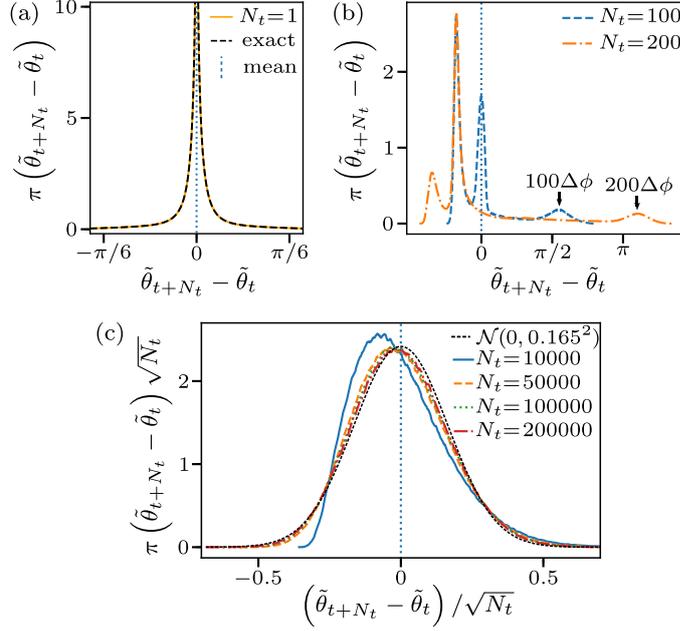
Figure 3.4: Distributions $\pi(\tilde{\theta}_{t+N_t} - \tilde{\theta}_t)$ of the change of the rolled-out dipole angle $\tilde{\theta}$ in N_t steps of direction-sweep MCMC for the tether ratio $\eta = 1.1$ and angle increment $\Delta\phi = \pi/180$. The mean of the distributions always vanishes. (a): Symmetric distribution for $N_t = 1$. (b): Highly asymmetric distributions for moderate N_t. (c): Rescaled large-N_t distributions compared to a symmetric Gaussian distribution.

switch to an excursion with $|\lambda| > 1$, that is, to a positive rotation of the rolled-out dipole angle $\tilde{\theta}$. The time (number of steps) of an excursion scales as $\sim \text{const}/(\Delta\phi)^2$ whereas the time of one zigzag is shorter by factor of $\Delta\phi$ as it scales as $\sim \text{const}/\Delta\phi$. Nevertheless, positive and negative rotations balance, and the expectations $\mathbb{E}(\tilde{\theta}_{t+N_t} - \tilde{\theta}_t)$ are zero for all numbers of steps N_t. This holds for a single move ($N_t = 1$) because $\pi(\tilde{\theta}_{t+1} - \tilde{\theta}_t)$ is symmetric in consequence of the detailed balance of the move from t to $t + 1$ (see Fig. 3.4a). For $N_t > 1$, $\tilde{\theta}_{t+N_t} - \tilde{\theta}_t = \tilde{\theta}_{t+N_t} - \tilde{\theta}_{t+N_t-1} + \cdots + \tilde{\theta}_{t+1} - \tilde{\theta}_t$ yields

$$\mathbb{E}\left(\tilde{\theta}_{t+N_t} - \tilde{\theta}_t\right) = \mathbb{E}\left(\tilde{\theta}_{t+N_t} - \tilde{\theta}_{t+N_t-1}\right) + \ldots + \mathbb{E}\left(\tilde{\theta}_{t+2} - \tilde{\theta}_{t+1}\right) + \mathbb{E}\left(\tilde{\theta}_{t+1} - \tilde{\theta}_t\right) = 0, \tag{3.14}$$

because the expectation of a sum of (possibly dependent) random variables equals the sum of expectations. The distribution $\pi(\tilde{\theta}_{t+N_t} - \tilde{\theta}_t)$ of the change of the rolled-out dipole angle in N_t steps, although it is of zero expectation, can be highly asymmetric (see Fig. 3.4b). For a moderate number of steps $N_t \lesssim \text{const}/(\Delta\phi)^2$, the distribution peaks for large $\tilde{\theta}$ that corresponds to trajectories that remain on long excursions. For $N_t \gg \text{const}/(\Delta\phi)^2$, the distribution approaches a Gaussian and becomes again symmetric because the large number of steps allows excursions and zigzags to compensate in a single trajectory (see Fig. 3.4c). The vanishing of $\mathbb{E}(\tilde{\theta}_{t+N_t} - \tilde{\theta}_t)$ implies that there are $O(1/\Delta\phi)$ zigzags for each excursion. Microscopically, this can be understood through the existence of the cutoff value λ_{cut}^+ [see Eq. (3.5) and Fig. 3.2b]. If $|\lambda_t| > \lambda_{\text{cut}}^+$, the next value of the impact parameter $|\lambda_{t+1}| > 1$; in contrast, a current value of the impact parameter $1 < |\lambda_t| < \lambda_{\text{cut}}^+$ may either produce $|\lambda_{t+1}| > 1$ or $|\lambda_{t+1}| < 1$. In the latter case, $\mathbf{x}$ gets trapped in its corresponding segment.

3.5 Approach to Equilibrium

The trajectory of the dipole in its sample space $\Omega = \{(\rho, \theta)\}$ is characterized by persistent negative and positive rotations. In order to quantify the approach to equilibrium of direction-sweep MCMC, we consider mixing times [74, 180]

$$t_{\text{mix}}(\varepsilon) = \min\{t : d(t) \leq \varepsilon\}, \tag{3.15}$$

where $d(t)$ is the total variation distance (TVD) between the stationary distribution π and the probability distribution $P^t(\bar{\mathbf{x}}_0, \cdot)$ at time t obtained by starting from the most unfavorable lifted initial configuration $\bar{\mathbf{x}}_0$:

$$d(t) = \max_{\bar{\mathbf{x}}_0} ||P^t(\bar{\mathbf{x}}_0, \cdot) - \pi||_{\text{TVD}} \tag{3.16}$$

$$= \max_{\bar{\mathbf{x}}_0} \frac{1}{2} \int |P^t(\bar{\mathbf{x}}_0, \bar{\mathbf{x}}) - \pi(\bar{\mathbf{x}})| \, d\bar{\mathbf{x}}. \tag{3.17}$$

The time $t_{\text{mix}} = t_{\text{mix}}(1/4)$ is defined as the mixing time. For $\varepsilon < 1/4$, $t_{\text{mix}}(\varepsilon)$ is bounded through t_{mix},

$$t_{\text{mix}}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{\text{mix}}, \tag{3.18}$$

showing that the mixing process is exponential [74]. In this section, we study mixing times rather than the more common equilibrium autocorrelation times because they are better adapted to the highly correlated trajectories discussed in Section 3.4. In addition, the tethered dipole model is sufficiently simple to allow for this more rigorous analysis.

We checked numerically for the dipole that the same $\bar{\mathbf{x}}_0$ maximizes the TVD for ε in the neighborhood of $1/4$ and determine t_{mix} *via* the time $t_{\text{mix}}(\bar{\mathbf{x}}_0)$:

$$t_{\text{mix}}(\bar{\mathbf{x}}_0) = \min \left\{ t : ||P_\theta^t(\bar{\mathbf{x}}_0, \cdot) - \pi_\theta||_{\text{TVD}} \leq 1/4 \right\}. \tag{3.19}$$

The maximum of $t_{\text{mix}}(\bar{\mathbf{x}}_0)$ over the initial configurations $\bar{\mathbf{x}}_0$ then yields t_{mix} [by doing so, we effectively interchanged the "min" in Eq. (3.15) with the "max" in Eq. (3.17)]. Due to the rotational invariance of the ring system, $t_{\text{mix}}(\bar{\mathbf{x}}_0)$ only depends on the angle difference $\theta_0 - \phi_0$. We thus set the initial direction $\phi_0 = 0$ and consider $t_{\text{mix}}(\rho_0, \theta_0)$, which we determine numerically by running 100 000 simulations that all start from $\bar{\mathbf{x}}_0 = (\rho_0, \theta_0)$. At each time step t, we use these simulations to determine $P_\theta^t(\bar{\mathbf{x}}_0, \cdot)$ and its TVD with π. The evaluation of the TVD requires in our case the evaluation of a two-dimensional integral over Ω. However, since the dipole length ρ relaxes very quickly, we approximate it by the one-dimensional integral over the dipole angle θ.

3.5.1 Identifying Unfavorable Initial Configurations

For small angle increments $\Delta\phi > 0$, two unfavorable initial configurations stand out. First, trajectories with an initial impact parameter $\lambda_0 = 1$ and $\bar{\mathbf{x}}_0 \in \mathcal{S}^+$ (or $\lambda_0 = -1$ and $\bar{\mathbf{x}}_0 \in \mathcal{S}^-$) always start with a deterministic zigzag until $\lambda_t = -1$ (or $\lambda_t = 1$). At the time t after this first zigzag, the probability

distribution $P_\theta^t(\bar{\mathbf{x}}_0, \cdot)$ is therefore strongly peaked, because every simulation had a similar zigzag-trajectory, and produces a large TVD in Eq. (3.17). Thereafter, different trajectories either continue with more zigzags or else with an excursion, which then flattens $P_\theta^t(\bar{\mathbf{x}}_0, \cdot)$. Second, a trajectory from the maximum $|\lambda_0| = \eta$ starts with an excursion and $P_\theta^t(\bar{\mathbf{x}}_0, \cdot)$ thus peaks at $\theta_t = t\Delta\phi$ because all simulations advance the dipole angle on average by $\Delta\phi$ per time step. Only once the random walk in $|\lambda_t|$ reaches the cutoff impact parameter λ_{cut}^+, different simulations may either continue on the excursion or change into a zigzag, and the distribution $P_\theta^t(\bar{\mathbf{x}}_0, \cdot)$ starts to flatten. The most unfavorable initial configuration among these two depends on the tether ratio η. For $\eta \to 1$, we find that starting the trajectory with a zigzag is the most unfavorable initial state (see Fig. 3.5a, the yellow regions correspond to $\lambda_0 \approx \pm 1$), whereas starting the trajectory with an excursion is most unfavorable for larger η (see Fig. 3.5b, the yellow regions correspond to $\lambda_0 \approx \pm\eta$). This can be understood by the fact that the number of time steps in a zigzag increases as $\eta \to 1$ [see Eq. (3.13)], whereas the difference $\eta - \lambda_{\text{cut}}^+$ that the impact parameter has to overcome in the initial excursion decreases [see Eq. (3.5)].

For large angle increments $\Delta\phi$, the random walk of λ_t is no longer described in terms of excursions and zigzags, and the initial configuration does not strongly influence $t_{\text{mix}}(\bar{\mathbf{x}}_0)$, except for values of $\Delta\phi$ that correspond to small direction sets $\mathcal{D}$. Then $t_{\text{mix}}(\rho_0, \theta_0)$ is roughly periodic in θ_0 (see the periodic stripes in Fig. 3.5c and d). For $\Delta\phi = \pi/2$, that is, directions ϕ_t drawn from the direction set $\mathcal{D} = \{0, \pi/2\}$, this may be due to the fact that with an initial dipole angle $\theta_0 \in \{-\pi/2, 0, \pi/2, \pi\}$ one of the two alternating directions hardly modifies θ_t during the initial part of the trajectory.

3.5.2 Mixing Time

We now systematically study t_{mix} for direction-sweep MCMC as a function of the direction set $\mathcal{D}$. We compare it with the reversible version that samples ϕ_{t+1} randomly from $\mathcal{D}$ (random discrete MCMC), and also with reversible MCMC with continuous directions $\Delta\phi = \text{ran}(0, \pi)$ (random continuous MCMC). Both versions satisfy detailed balance.

Several properties stand out (see Fig. 3.6a). First, the mixing time is very sensitive to the size of $\mathcal{D}$ regardless of whether its elements are accessed sequentially or randomly. For a thin ring (tether ratio $\eta \to 1$), the mixing time t_{mix} shows characteristic peaks for small direction sets $\mathcal{D}$. The height of these peaks (for not too large set sizes $|\mathcal{D}|$) is proportional to $1/|\mathcal{D}|$. This yields a particularly large mixing time for $\Delta\phi = \pi/2$ where $|\mathcal{D}| = 2$.

Second, we find that sweeping through the elements of $\mathcal{D}$ is generically better than randomly sampling the direction from $\mathcal{D}$, except for $\Delta\phi \to 0$ where the sweeps are too slow and the mixing time diverges. For small $|\mathcal{D}|$, this benefit of direction-sweep MCMC is easily understood by the non-vanishing probability of repeated (redundant) moves in the same direction that only appear in random discrete MCMC.

For all considered values of the tether ratio η, we find that direction-sweep MCMC with an appropriate angle increment $\Delta\phi$ is faster than random continuous MCMC and, in particular, as direction-sweep MCMC with $\Delta\phi = \pi/2$ (see Fig. 3.6b). The speedup compared to the choice $\Delta\phi = \pi/2$ is large, and it appears to diverge as $\eta \to 1$. This may render the non-reversible scheme especially promising for dipolar particles in ECMC where up to now $\Delta\phi = \pi/2$ was always chosen. We confirm that the smallest mixing time in direction-sweep MCMC is indeed reached for small $\Delta\phi$,
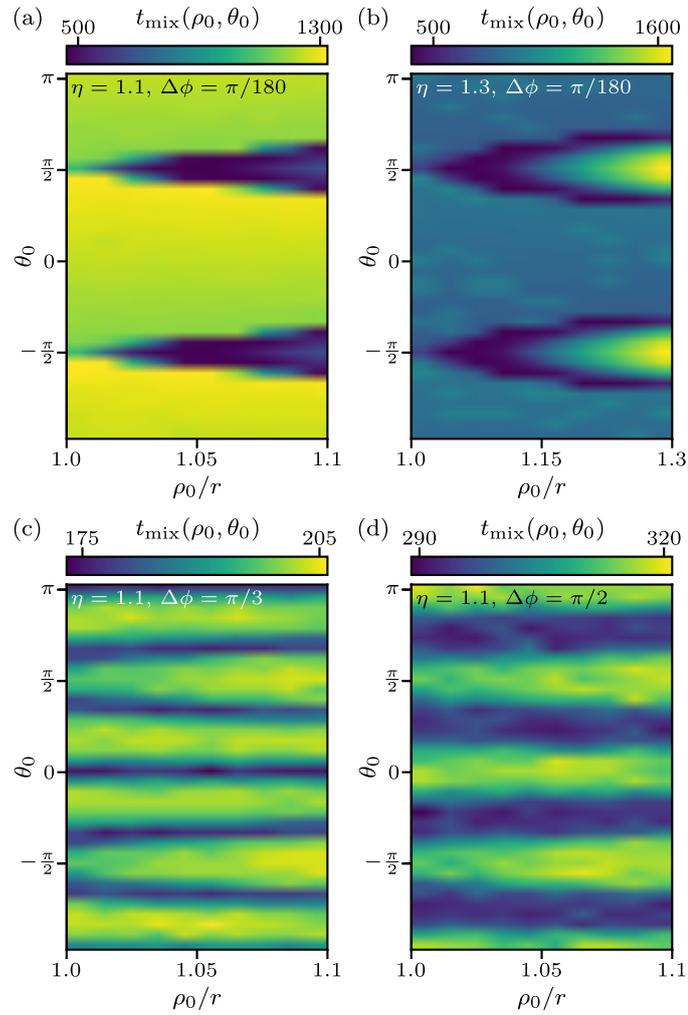
Figure 3.5: Mixing times $t_{\mathrm{mix}}(\overline{\mathbf{x}}_0 = (\rho_0, \theta_0))$ with the initial direction $\phi_0 = 0$ for different values of the tether ratio η and the angle increment $\Delta\phi$. (a) and (b): The most unfavorable initial configuration $\overline{\mathbf{x}}_0$ depends on η for small $\Delta\phi$. (c) and (d): Values of $\Delta\phi$ with small direction sets $\mathcal{D}$ lead to periodic patterns in θ_0.
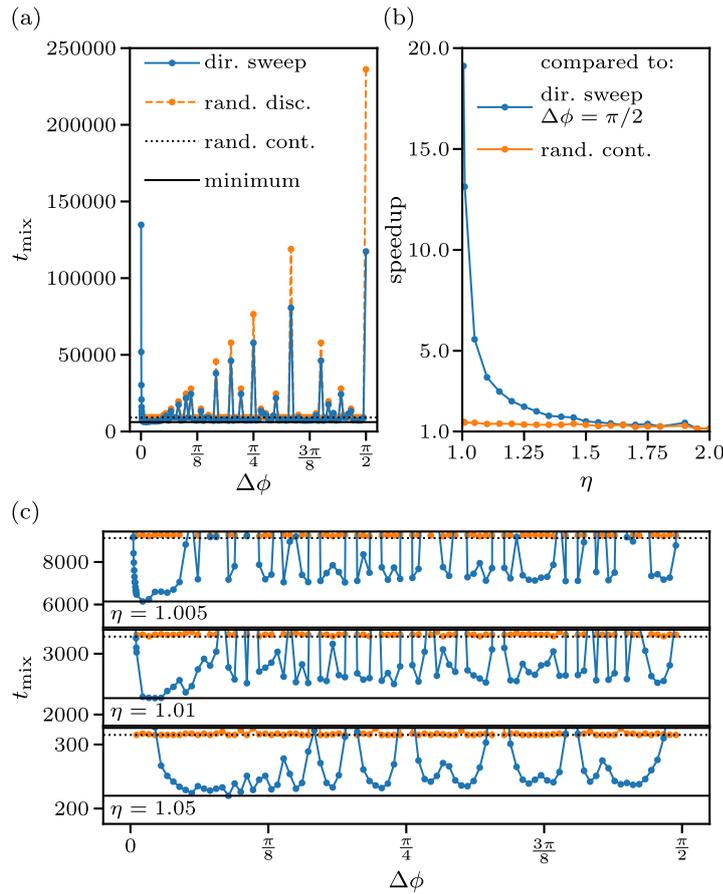
Figure 3.6: Mixing times t_{mix} of direction-sweep MCMC for different angle increments $\Delta\phi$. (a): t_{mix} for the tether ratio $\eta = 1.005$ compared to mixing times of random discrete and continuous MCMC. (b): Optimal speedup with respect to $\Delta\phi = \pi/2$ and to random continuous MCMC. (c): Mixing time t_{mix} of direction-sweep MCMC compared to that of random continuous MCMC for small tether ratios $\eta \gtrsim 1$. The line labeling is the same as in (a).
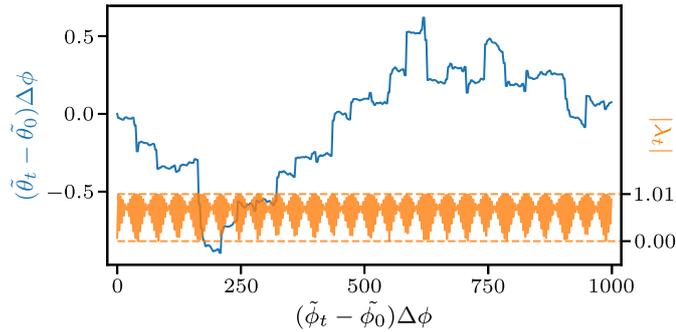
Figure 3.7: Direction-sweep MCMC trajectories of the rescaled rolled-out dipole angle $(\tilde{\theta}_t - \tilde{\theta}_0)\Delta\phi$ (blue, upper) and of the impact parameter $|\lambda_t|$ (yellow, lower) vs. the rescaled change of the rolled-out direction $(\tilde{\phi}_t - \tilde{\phi}_0)\Delta\phi$ for the tether ratio $\eta = 1.01$ and large angle increment $\Delta\phi = 85\,\pi/180$.

that is, for the peculiar trajectories discussed in Section 3.4. This, in our model, can of course only be observed for $\eta \to 1$ because the speedup for small $\Delta\phi$ is cut off by the divergence of t_{mix} for $\Delta\phi \to 0$.

Finally, we find that direction-sweep MCMC is usually faster than the random continuous MCMC even for large values of $\Delta\phi$, except when $|\mathcal{D}|$ is very small (see Fig. 3.6c, large angles $\Delta\phi$ give small direction sets only if $\Delta\phi/(2\pi)$ is a simple fraction). The trajectories remain very regular for generic $\Delta\phi$. They feature intriguing patterns for the impact parameter λ_t and the rolled-out dipole angle $\tilde{\theta}_t$, that require further study (see Fig. 3.7).

3.6 Conclusion

We have discussed a non-reversible MCMC algorithm for particle systems that sweeps through the direction of motion, rather than to sample directions randomly. For a single two-dimensional tethered hard-disk dipole, we proved in a local-equilibrium limit that direction-sweep MCMC induces persistent dipole rotations with a total rolled-out angle that diverges as the direction sweep becomes slower. Persistent rotation takes place in both senses, and the two exactly compensate to zero net rotation.

Direction-lifted MCMC (of which direction-sweep MCMC is a special case) remains valid for general N-body problems. It preserves the independence of the lifted stationary distribution from the lifting variable [80] even if the thermalization condition at fixed lifting variable is dropped. Real-world direction-lifted MCMC may go to much smaller values of angle increments $\Delta\phi$ than the single dipole, simply because mixing times will be much larger in applications. The additional parameter $\Delta\phi$ would have to be optimized in applications, a feature that presents a downside of the method. Nevertheless, we think that the lessons of using small values of the angle increment and of avoiding small direction sets can be carried over. Although the full analytic control of the dynamics can certainly not be preserved in the N-body case, qualitatively similar performance differences in autocorrelation times were observed for direction-lifted ECMC in a model with N tethered hard-disk dipoles [P2]. It will be fascinating to understand the usefulness of direction lifting for applications such as polymer physics, and also in systems of long-range interacting extended molecules at the core of the JeLLyFysh project [P6]. More generally, our model illustrates that non-reversibility profoundly changes the basic

properties of local MCMC algorithms, in the same way as out-of-equilibrium statistical physics is fundamentally different from its equilibrium counterpart.

Acknowledgments

3.A Lifted Markov Chains and Particle-Sweep Monte Carlo

In this appendix, we collect basic notions about lifted Markov chains [111, 112], using the notations of [80]. The lifting framework allows for the systematic construction of non-reversible Markov chains, by integrating some or all of the moves into a lifted sample space. We also discuss the particle-sweep Markov-chain algorithms mentioned in the introduction (see Section 3.1) as particle-liftings of reversible Markov chain, and show their correctness.

3.A.1 Basic Concepts in Lifted Markov Chains

For concreteness, we consider a finite irreducible Markov chain Π with a sample space $\Omega = \{\mathbf{x}\}$, a stationary distribution π, and a transition matrix P (see [74]). The element $P_{\mathbf{x},\mathbf{x}'} \geq 0$ of the transition matrix gives the conditional probability to move to the sample $\mathbf{x}'$ (at time $t + 1$) given that the Markov chain is at sample $\mathbf{x}$ (at time t). The transition matrix is a stochastic matrix, i.e., $\sum_{\mathbf{x}' \in \Omega} P_{\mathbf{x},\mathbf{x}'} = 1$, which encodes that the transition probability from the sample $\mathbf{x}$ to all other samples must be 1. The stationary distribution π and the transition matrix P satisfy the global-balance condition $\pi_{\mathbf{x}} = \sum_{\mathbf{x}' \in \Omega} \pi_{\mathbf{x}'} P_{\mathbf{x}',\mathbf{x}}$. As mentioned in Section 3.1, reversible Markov chains are those that satisfy the more restrictive detailed-balance condition $\pi_{\mathbf{x}} P_{\mathbf{x},\mathbf{x}'} = \pi_{\mathbf{x}'} P_{\mathbf{x}',\mathbf{x}}$.

A lifting (or a "lifted Markov chain") $\widehat{\Pi}$ consists in another lifted sample space $\widehat{\Omega}$, lifted stationary distribution $\hat{\pi}$, and lifted transition matrix $\widehat{P}$. Any sample in the "collapsed" sample space Ω is connected to one or more samples in the lifted sample space $\widehat{\Omega}$. We consider the special case that $\widehat{\Omega}$ can be written as $\widehat{\Omega} = \Omega \times \mathcal{L}$, where $\mathcal{L}$ is a set of lifting variables. Any sample in Ω is then "lifted" into $|\mathcal{L}|$ lifted samples in $\widehat{\Omega}$ that are specified by tuples $(\mathbf{x}, i)$ with $\mathbf{x} \in \Omega$ and $i \in \mathcal{L}$. Examples of $\mathcal{L}$ are the set of particle indices or, as in the main part of this paper, the set of directions. In the lifted sample space, the global-balance condition becomes for every $(\mathbf{x}, i) \in \widehat{\Omega}$:

$$\hat{\pi}_{(\mathbf{x},i)} = \sum_{(\mathbf{x}',j) \in \widehat{\Omega}} \hat{\pi}_{(\mathbf{x}',j)} \widehat{P}_{(\mathbf{x}',j),(\mathbf{x},i)}. \tag{3.A.1}$$

The Markov chain $\widehat{\Pi}$ is a lifting of Π (conversely, Π is referred to as a "collapsing" [112] of $\widehat{\Pi}$), if for every $\mathbf{x} \in \Omega$

$$\pi_{\mathbf{x}} = \sum_{i \in \mathcal{L}} \hat{\pi}_{(\mathbf{x},i)}, \tag{3.A.2}$$

and for every pair $\mathbf{x}, \mathbf{x}' \in \Omega$

$$\pi_{\mathbf{x}} P_{\mathbf{x}, \mathbf{x}'} = \sum_{i,j \in \mathcal{L}} \hat{\pi}_{(\mathbf{x}, i)} \widehat{P}_{(\mathbf{x}, i), (\mathbf{x}', j)}. \tag{3.A.3}$$

Equation (3.A.2) expresses that the combined stationary weights of all lifted copies of a sample $\mathbf{x} \in \Omega$ must be the same as the stationary weight of $\mathbf{x}$, while Eq. (3.A.3) ensures that the probability flow between any two collapsed samples is equal to the total flow between the corresponding lifted samples. These properties make it possible to sample the collapsed probability distribution π using the lifted Markov chain $\widehat{\Omega}$. There are many choices for the lifted transition matrix given a set $\mathcal{L}$ of lifting variables, and these choices greatly influence the performance of the lifted Markov-chain Monte Carlo algorithms [80]. Although a lifting does not modify the basic conductance of the graph, it can allow for a considerable speedup. The mixing time of a lifted Markov chain, that is, the time that describes the approach towards equilibrium from a most unfavorable initial sample (see Section 3.5), is at least the square root of the mixing time of the collapsed Markov chain. This corresponds to replacing the diffusive motion that is usually associated to Markov-chain Monte Carlo with a ballistic motion. It is well understood that in order to realize any significant speed-up, the lifting must be non-reversible [112].

3.A.2 Particle-Sweep Liftings of Reversible Markov Chains

As discussed in Section 3.1, the earliest non-reversible Markov chains—introduced in the original 1953 paper [72]—are "particle-sweep" versions of the reversible Metropolis algorithm for N particles. Sweep algorithms can be interpreted as lifted Markov chains with the set $\mathcal{L} = \{1, \ldots, N\}$ of lifting variables equal to the set of particle indices. The samples $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \Omega$ contain all particle positions.

For concreteness, we consider an arbitrary local reversible Markov-chain algorithm that, at each time step, attempts a move of a randomly chosen particle i. The non-vanishing elements $P_{\mathbf{x}, \mathbf{x}'}$ of the collapsed transition matrix P thus correspond to moves of a single particle i. To simplify notation, we write these elements as $P_{\mathbf{x}_i, \mathbf{x}'_i}$ rather than as $P_{(\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N), (\mathbf{x}_1, \ldots, \mathbf{x}'_i, \ldots, \mathbf{x}_N)}$. Instead of choosing a random particle, the particle-sweep version always attempts a move of particle $i + 1$ after a move of i (with periodic boundary conditions in the particle indices). The lifted Markov chain performs the same moves as the collapsed one, but only for the particle i that is specified through the lifting variable. The possibly non-zero elements of the lifted transition matrix $\widehat{P}$ are

$$\widehat{P}_{(\mathbf{x}_i, i), (\mathbf{x}'_i, i+1)} = N P_{\mathbf{x}_i, \mathbf{x}'_i}, \quad \text{if } \mathbf{x} \neq \mathbf{x}', \tag{3.A.4}$$

and, in addition, because $\widehat{P}$ is a stochastic matrix,

$$\widehat{P}_{(\mathbf{x}'_i, i), (\mathbf{x}'_i, i+1)} = 1 - \sum_{\mathbf{x} \in \Omega, \mathbf{x} \neq \mathbf{x}'} \widehat{P}_{(\mathbf{x}'_i, i), (\mathbf{x}_i, i+1)}. \tag{3.A.5}$$

The factor N in Eq. (3.A.4) stems from the fact that, in the collapsed Markov chain, a move of particle i is attempted with probability $1/N$, so that $P_{\mathbf{x}_i, \mathbf{x}'_i} \sim 1/N$ for $\mathbf{x}' \neq \mathbf{x}$ while in the lifted Markov chain, starting from a lifted sample $(\mathbf{x}, i)$, the same move of particle i is attempted with probability 1. With

the ansatz

$$\hat{\pi}_{(\mathbf{x},i)} = \frac{1}{N}\pi_{\mathbf{x}} \quad \forall i \in \mathcal{L}, \tag{3.A.6}$$

which implies the condition for a lifted Markov chain in Eq. (3.A.2), it follows from Eqs (3.A.4) and (3.A.6) that for $\mathbf{x} \neq \mathbf{x}'$

$$\begin{aligned}
\hat{\pi}_{(\mathbf{x},i)}\widehat{P}_{(\mathbf{x}_i,i),(\mathbf{x}_i',i+1)} &= \pi_{\mathbf{x}}P_{\mathbf{x}_i,\mathbf{x}_i'} \\
&= \pi_{\mathbf{x}'}P_{\mathbf{x}_i',\mathbf{x}_i} \\
&= \hat{\pi}_{(\mathbf{x}',i+1)}\widehat{P}_{(\mathbf{x}_i',i),(\mathbf{x}_i,i+1)},
\end{aligned} \tag{3.A.7}$$

where we used that the reversible collapsed Markov chain satisfies the detailed-balance condition. Equation (3.A.7) can be understood as a skew detailed balance [109, 181]. We now compute, using Eqs (3.A.5) and (3.A.7), the probability flow into the lifted sample $(\mathbf{x}', i + 1)$:

$$\sum_{\mathbf{x}\in\Omega}\hat{\pi}_{(\mathbf{x},i)}\,\widehat{P}_{(\mathbf{x}_i,i),(\mathbf{x}_i',i+1)} = \sum_{\mathbf{x}\in\Omega,\mathbf{x}\neq\mathbf{x}'}\hat{\pi}_{(\mathbf{x},i)}\,\widehat{P}_{(\mathbf{x}_i,i),(\mathbf{x}_i',i+1)} + \hat{\pi}_{(\mathbf{x}',i+1)}\widehat{P}_{(\mathbf{x}_i',i),(\mathbf{x}_i',i+1)} \tag{3.A.8}$$

$$= \hat{\pi}_{(\mathbf{x}',i+1)}\left[\sum_{\mathbf{x}\in\Omega,\mathbf{x}\neq\mathbf{x}'}\widehat{P}_{(\mathbf{x}_i',i),(\mathbf{x}_i,i+1)} + \widehat{P}_{(\mathbf{x}_i',i),(\mathbf{x}_i',i+1)}\right] \tag{3.A.9}$$

$$= \hat{\pi}_{(\mathbf{x}',i+1)}. \tag{3.A.10}$$

This is the global-balance condition for the lifted probability distribution of Eq. (3.A.6) and, because an irreducible Markov chain has a unique stationary distribution, it verifies the above ansatz. The reversibility condition that was used in Eq. (3.A.7) is necessary: there are examples of non-reversible collapsed Markov chains that do not allow for a sweep-style lifting [80, Section 4.2.1].

The direction-lifted Markov chains of the main text perform sweeps in the space of directions, rather than in the space of particle indices. These non-reversible algorithms likewise satisfy the global-balance condition as the collapsed Markov chain is reversible.

Research Paper 2—Hard-Disk Dipoles and Non-Reversible Markov Chains

This chapter considers the publication

Section 4.0 provides a summary of the publication and contextualizes it within this doctoral thesis. Furthermore, the author's contributions are clearly indicated. The remaining part of this chapter reproduces the published version of the publication itself. All co-authors of the publication agree with its use for this doctoral thesis and with the identification of the author's contributions.

4.0 Prologue

The first research paper of this doctoral thesis introduced a two-dimensional tethered hard-disk dipole model that caricatures the SPC/Fw water model (see Chapter 3 and Ref. [P1]). The simplified model retains the polar nature which allows us to study its rotation dynamics and, in the end, to gain insights on how rotations can be enhanced in the framework of event-chain Monte Carlo (ECMC), even for the computationally more complex SPC/Fw water model. Since ECMC does currently not allow for continuous-time rotations of dipoles that appear especially difficult to implement in long-range-interacting molecular systems, rotations of tethered hard-disk dipoles are pieced together from subsequent straight-line motions of the active disks. Combining ECMC with reversible Monte Carlo moves that explicitly rotate dipoles is possible [134]. This would, however, not be adequate in the ultimate application to molecular systems in the overarching objective of this doctoral thesis because this would require an inefficient and inaccurate computation of the global potential-energy change.

Only considering a single tethered hard-disk dipole in the first research paper of this doctoral thesis enabled an analytical analysis and a numerical computation of the mixing time. While mixing times are the primary characteristic of any Markov-chain Monte Carlo (MCMC) algorithm, it is impossible to obtain them in most practical applications and one relies on alternative cues to ensure that the algorithm reached the desired stationary equilibrium distribution (see Section 2.2.1). Assuming that the MCMC simulation reached stationarity, one usually considers the easier-to-evaluate integrated autocorrelation time as a performance characteristic instead (see Section 2.2.2). This approach is followed in the reproduced publication in this chapter. We consider an extensive numerical benchmark of an increasing number of tethered hard-disk dipoles at increasing densities in a periodic cubic box. The length of the tether that connects two hard disks to a dipole is chosen so that it shows a similar amount of flexibility as the O−H bond lengths in the SPC/Fw water model at room temperature.

We consider the reversible Metropolis algorithm with different proposal distributions and compare it to a large number of non-reversible ECMC variants. First, the "periodic" straight ECMC variant periodically aligns the velocities with the coordinate axes after each event chain as in the first version of JᴇLLʏFʏsʜ-Version1.0. Second, the "random" straight ECMC variant chooses the two-dimensional direction of the velocity randomly on a unit circle. Third, "sequential" straight ECMC variants carry out a direction sweep with various values of the angle increment $\Delta\phi$ that, similar to the chain time τ_{chain} of an event chain, acts as another intrinsic parameter that must be fine-tuned. Against the background of the ultimate application to molecular systems, an important point of this paper is that it also considers ECMC variants that differ from straight ECMC in their lifting moves in events. Reflective [115], forward [150], and Newtonian ECMC [124] do not simply transfer the velocity of the previously active disk to its collision partner in an event. Instead, both the active disk and the velocity are changed in an event (see Figs 1.5 and 4.2). Even though the velocity continuously changes during an event chain, one may still consider resampling it in periodic time intervals of the chain time τ_{chain} (and this might even be necessary in certain cases, see below). The chain time τ_{chain} acts as an intrinsic parameter that influences the autocorrelation times in every ECMC variant. All in all, thousands of parameter sets for different ECMC variants are considered and benchmarked against the reversible local Metropolis algorithm (see, e.g., Fig. 4.5).

Once fine-tuned to an optimal acceptance rate of their Monte Carlo moves, the local reversible Metropolis algorithms with the different proposal distributions become equally fast at high densities of the tethered hard-disk dipoles (see Fig. 4.6). This is expected as the precise choice of the proposal distribution typically does not influence the performance of local reversible algorithms very much. Sequential straight ECMC with fine-tuned τ_{chain} and $\Delta\phi$ is indeed the fastest straight ECMC variant, as hoped for after the results of the first research paper of this doctoral thesis, but only by a relatively small margin. Also, similar to the different Metropolis algorithms, all straight ECMC variants, as well as reflective and forward ECMC show a similar performance at high densities (see Fig. 4.7). Here, reflective and forward ECMC are fastest in the $\tau_{\text{chain}} \to \infty$ limit which removes any need for fine-tuning their intrinsic parameter. Straight, reflective, and forward ECMC give an order-of-magnitude speedup compared to the Metropolis algorithm which was also expected from replacing a reversible by a non-reversible algorithm. What comes as a surprise, however, is that Newtonian ECMC, again in the $\tau_{\text{chain}} \to \infty$ limit, is consistently faster than the other ECMC variants by a considerable factor at all densities. We observe an up to an order-of-magnitude spread between the worst periodic straight

ECMC variant and Newtonian ECMC. At high density, optimal Newtonian ECMC demonstrates a 60-fold speedup compared to the reversible Metropolis algorithm (see Fig. 4.7 again). For the overarching objective of this doctoral thesis—the ultimate application to molecular systems such as the SPC/Fw water model—we conclude that non-straight ECMC variants require no fine-tuning and that the performance of non-reversible MCMC can show an extreme dependence on details in its algorithm. Especially Newtonian ECMC appears promising for decorrelating the polarization by rotating dipoles or molecules.

In a more general context, we also discuss the problem of ensuring irreducibility in different ECMC variants. Sequential straight ECMC necessarily requires velocity changes in periodic time intervals of the chain time τ_{chain}. We show for a single tethered hard-disk dipole that this velocity change must be connected to a random choice of the initially active disk in the event chain to ensure irreducibility (see Fig. 4.3). Similarly, although reflective and Newtonian ECMC change the velocity of the active disk in every event, they strictly require random resamplings of the velocity and active disk in periodic time intervals τ_{chain} (see also Ref. [140, Fig. 3]). Only forward ECMC with its event-based randomness is irreducible without any resamplings. These irreducibility problems of reflective and Newtonian ECMC can only be observed for a single dipole (or a set of effectively independent dipoles). At high densities, static observables agree with or without resamplings and it appears safe to omit them in practical simulations. Still, for reflective and Newtonian ECMC that become fastest in the $\tau_{\text{chain}} \to \infty$ limit, it appears safer to choose a large finite value $\tau_{\text{chain}} \to \infty$, *just in case*.

Contributions of the Author

The author contributed in an essential manner to the entire research project underlying this research paper. In particular, he was the sole contributor to the implementation of the Go software for an efficient simulation of tethered hard-disk dipoles with an $O(1)$ complexity per event with different ECMC variants, the data collection for the ECMC algorithms, and the creation of the figures. The author also made essential contributions to the proofs of the different irreducibility problems, and the writing of the research paper itself. Following the Contributor Role Taxonomy (CRediT) [174], the author contributions are as follows:

- Conceptualization (ideas; formulation or evolution of overarching research goals and aims).
- Methodology (development or design of methodology; creation of models).
- Software (programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components).
- Validation (verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs).
- Formal analysis (application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data).
- Investigation (conducting a research and investigation process, specifically performing the experiments, or data/evidence collection).

- Writing—Original Draft [preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)].

- Writing—Review & Editing (preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision—including pre- or post-publication stages).

- Visualization (preparation, creation and/or presentation of the published work, specifically visualization/data presentation).

Hard-Disk Dipoles and Non-Reversible Markov Chains

Abstract   We benchmark event-chain Monte Carlo (ECMC) algorithms for tethered hard-disk dipoles in two dimensions in view of application of ECMC to water models in molecular simulation. We characterize the rotation dynamics of dipoles through the integrated autocorrelation times of the polarization. The non-reversible straight, reflective, forward, and Newtonian ECMC algorithms are all event-driven and only move a single hard disk at any time. They differ only in their update rules at event times. We show that they realize considerable speedups with respect to the local reversible Metropolis algorithm with single-disk moves. We also find significant speed differences among the ECMC variants. Newtonian ECMC appears particularly well-suited for overcoming the dynamical arrest that has plagued straight ECMC for three-dimensional dipolar models with Coulomb interactions.

4.1 Introduction

Markov-chain Monte Carlo [72] (MCMC) is a computational method for sampling high-dimensional probability distributions π including the Boltzmann distribution of statistical physics. In MCMC, an initial sample x, at a time $t = 0$, is drawn from a distribution $\pi^{\{0\}}$ rather than from the "equilibrium" distribution π. The sample x, at time t, moves to the sample y, at time $t + 1$, with a conditional probability P_{xy}, element of a time-independent transition matrix P. The probability distribution $\pi^{\{t\}}$ then evolves to $\pi^{\{t+1\}} = \pi^{\{t\}}P$. The intermediate probability distributions $\pi^{\{t\}}$ are usually not known explicitly for $t > 0$, but the transition matrix is designed for $\pi^{\{t\}}$ to converge to π in the limit of infinite times. To do so, P must satisfy a balance condition that expresses the stationarity of π. In addition, P must be irreducible and aperiodic [74] (that is, eventually reach any sample y from any other sample x and be free of cycles). Reversible transition matrices satisfy the detailed-balance condition $\pi_x P_{xy} = \pi_y P_{yx}$ for all x and y. Detailed balance is equivalent to the statement that the equilibrium flow $\mathcal{F}_{xy} = \pi_x P_{xy}$ from x to y equals the reverse flow from y to x. Non-reversible MCMC algorithms satisfy a weaker global-balance condition $\pi_x = \sum_y \pi_y P_{yx}$ for all x. Under conditions of irreducibility and aperiodicity, they converge to the equilibrium distribution π with non-zero net flows $\mathcal{F}_{xy} - \mathcal{F}_{yx}$. Paradoxically, samples are then drawn from the equilibrium distribution π for $t \to \infty$ by a non-equilibrium random process with non-vanishing net flows.

In past decades, research and applications have focused almost exclusively on reversible Markov chains (and some close relatives, such as sequential schemes [168, 182]). Reversible Markov chains are straightforward to set up for arbitrary probability distributions π and are easy to conceptualize, in particular because of the real-valued eigenvalue spectrum of their transition matrices. Popular

reversible Markov chains include the Metropolis- [72] and the heat-bath [86–88] algorithms. Interest in non-reversible Markov chains has increased in recent years, ever since it was understood [111, 112] that such algorithms often approach equilibrium faster than reversible Markov chains. Systematic non-reversible Markov-chain schemes are becoming available [109, 115].

Event-chain Monte Carlo [80, 115, 121] (ECMC) is a family of non-reversible local Markov chains that have been applied to, e.g., particle [142] and spin systems [125–127], polymers [131, 132], and field-theoretical models [133]. ECMC algorithms are defined in continuous time, and they can be interpreted as piecewise deterministic Markov processes [136]. They have found applications in Bayesian statistics [140, 141]. A number of features set ECMC algorithms apart from reversible Markov chains. First, they may reach the equilibrium distribution π on fast ballistic time scales rather than on the diffusive time scales that are typically associated with local reversibility. Considerable speedups were demonstrated in analytically solvable test cases [111, 112, 171] and recovered in applications of practical relevance [80, 132, 133, 142]. Second, for a given model (probability distribution π), ECMC encompasses a variety of local MCMC algorithms that can behave quite differently. On the one hand, different ECMC algorithms may adopt different factor sets for the breakup of the Boltzmann distribution. As an example, a Hamiltonian describing interacting dipoles might be broken up into factors of pairs of atoms or else into factors of pairs of dipoles [146]. On a finer scale, even for a given set of factors, different variants may have widely different behaviors, as we will discuss in the present paper. Third, ECMC need not evaluate the ratio π_y/π_x to move from a sample x to a sample y (in other words, need not evaluate the change in energy) because the factors are statistically independent [121, 146]. For N-particle systems made up of charge-neutral dipolar molecules with long-range Coulomb interactions implemented as dipole factors, to give an example, all interactions between pairs of dipoles are treated by time-independent bounds. For the next event, a single dipole factor is then identified through advanced sampling methods [148], and the Coulomb interaction is considered to machine precision only for this factor. This eliminates all statistical bias, and the sampling method is without any approximation (although the total energy remains unknown) and of complexity $O(1)$ per event. Since the mean-free path decreases as $O(1/\log N)$, the overall complexity is $O(N \log N)$ to advance every particle by $O(1)$ [146].

The sampling of the Boltzmann distribution without evaluating the energy has been implemented in the general-purpose open-source ECMC application "JᴇLLʏFʏsʜ" [P6] that is aimed at classical molecular simulation in Coulomb systems. In these systems, the evaluation of the long-range energy represents the computational bottleneck for other methods such as the molecular-dynamics-based approach [10], while it renders the use of local reversible Monte Carlo algorithms [98] totally hopeless. In the test case of three-dimensional water with the simple-point-charge-flexible-water (SPC/Fw) potential [28] and the "periodic" variant of straight ECMC that only moves along the coordinate axes (see Section 4.3.2 for a definition), the simulated water molecules, however, resisted rotation and the polarization remained dynamically arrested for long times [149]. The present paper studies many variants of ECMC for a simpler model of polar molecules and suggests that the straight periodic variant is ill-suited to these systems.

In a generic Markov chain, the sample space Ω is distinct from the moves Δ. The latter are then part of a move set and lead from samples $x \in \Omega$ to samples $y \in \Omega$. In the local reversible Metropolis algorithm for particle systems, for example, $\Delta = (\delta, i)$ consists in a small random displacement δ of a

random particle i. ECMC, in contrast, belongs to the class of lifted Markov chains [111, 112], in which parts of the move set are integrated into a lifted sample space $\widehat{\Omega}$. In the above example, the proposed move $\Delta' = (\delta', i')$ at time $t + 1$, after a move Δ at time t, is then no longer independently proposed, but it rather depends, in the lifted transition matrix, on the lifted configuration and, in particular, on Δ. In a sequential MCMC scheme, for example, i' could be forced to equal $i + 1$ (with periodic boundary conditions) by suitable terms in the lifted transition matrix. The probability distribution lives itself in the lifted sample space $\widehat{\Omega}$. Nevertheless, for Markov chains with lifting variables, the stationary probability distribution in $\widehat{\Omega}$ factorizes into the original distribution π and a distribution for the lifting variables, while the lifted transition matrix non-trivially couples the two sectors. In the ECMC algorithms that we consider, the time t is generally continuous.

In this paper, we caricature the SPC/Fw water model as tethered hard-disk dipoles. This replaces a three-dimensional model by a two-dimensional one and lumps the vibrational, bending, Lennard-Jones, and Coulomb interactions into a hard-disk potential. The simplified dipole model retains the polar nature of molecules, and it can be simulated within ECMC orders of magnitude faster than the system of flexible Coulomb dipoles, although the computational complexity of ECMC [$O(1)$ per event both for hard disks and for Coulomb-interacting systems] is comparable. The different algorithms are also much simpler to implement for the hard-disk dipoles than for the Coulomb system. This allows us to scan straight [P1, 115], reflective [115], forward [150], and Newtonian [124] ECMC with thousands of parameter sets. We benchmark the ECMC algorithms, which continuously move a single hard disk at any time, against the local Metropolis algorithms with different displacement sets for single disks. We characterize the speed of algorithms via the autocorrelation of the polarization, the mode that with ECMC equilibrates slowly in the SPC/Fw water model [149].

For hard-disk dipoles with model parameters that roughly correspond to those of SPC/Fw water, we find a 50-fold speedup for the fastest non-reversible ECMC algorithm with respect to the optimized local reversible Metropolis algorithm. We also find a more than order-of-magnitude speed difference among the different ECMC variants. The different versions of straight ECMC, while faster than the local Metropolis algorithm, are clearly the slowest ones at low density, while they are comparable with forward and reflective ECMC at high densities. Newtonian ECMC is by far the fastest. We suggest that this may be rooted in its absence of intrinsic parameters.

The content of this paper is as follows: In Section 4.2, we define the two-dimensional tethered hard-disk dipole model and motivate its parameters with respect to physical systems of the SPC/Fw model in three dimensions. We also discuss the polarization, which tracks the ability of dipoles to rotate. In Section 4.3, we discuss the reversible and non-reversible MCMC algorithms that we have implemented and illustrate their behavior for the case of a single dipole. We also discuss subtle aspects concerning the irreducibility of the Markov chains. In Section 4.4, we benchmark the different variants of ECMC against the local reversible Metropolis algorithm for a range of densities and system sizes, and we evidence the superiority of the non-reversible ECMC algorithms. In Section 4.5, we project how our conclusions can be extended to the more complex water models.
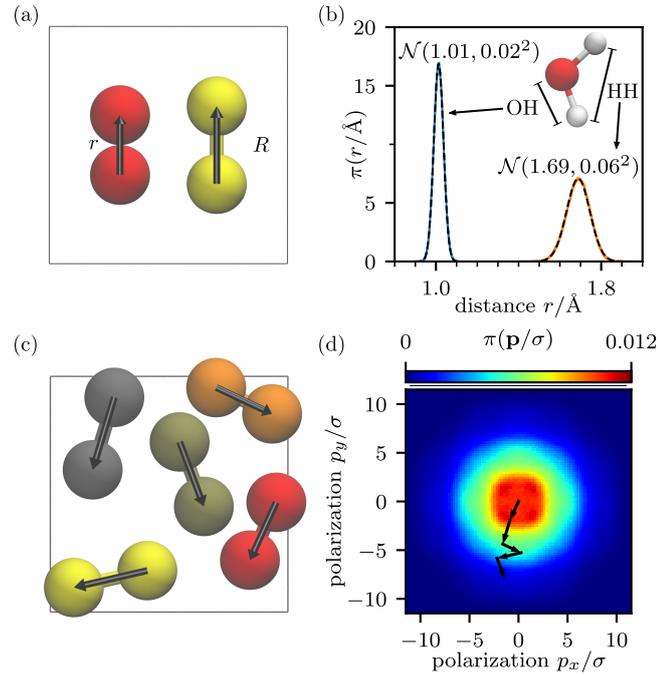
Figure 4.1: Tethered hard-disk dipole model. (a): Dipoles with two hard disks of radius σ, minimum extension $r = 2\sigma$, and tether length R. (b): Probability distribution of the oxygen–hydrogen (OH) and hydrogen–hydrogen (HH) bond lengths for a single SPC/Fw water molecule at room temperature from a straight ECMC simulation using the open-source application JɛLLyFʏsʜ [P6]. (c): Five dipoles with polarization $\mathbf{p}_i$. (d): Total polarization $\mathbf{p}$ of (c), together with its probability distribution.

4.2 Hard-Disk Dipole Model

We consider tethered dipoles in a two-dimensional square box of sides L with periodic boundary conditions. Each dipole consists of two hard disks of radius σ, with a flat inner-dipole interaction that constrains its extension ρ (the separation of the disk centers) between the contact distance $r = 2\sigma$ and the tether length R (see Fig. 4.1a). Dipole configurations without overlapping disks and with all dipole extensions between r and R all have the same Boltzmann weight, whereas all other configurations have zero statistical weight. As the minimum dipole extension at contact equals 2σ (see Fig. 4.1a), each dipole configuration is also a configuration of hard disks. Any dipole system is parameterized by the number of dipoles N, the tether ratio $\eta = R/r$, and the hard-disk density $D = 2N\pi\sigma^2/L^2$.

In a single three-dimensional water molecule at room temperature (as, for example, described in the SPC/Fw model), the oxygen–hydrogen distance fluctuates by $\sim 2.3\,\%$ and the hydrogen–hydrogen distance fluctuates by $\sim 3.4\,\%$, as we observed through a simulation of a single SPC/Fw water molecule using the implementation of straight ECMC in the open-source application JɛLLyFʏsʜ [P6] (see Fig. 4.1b). In the hard-disk dipole model, we thus adopt a tether ratio $\eta = 1.1$, for which the $\sim 2.75\,\%$ fluctuations of the extension are quite similar. The tether length R is thus only 10 % larger than the contact distance between two disks so that two dipoles cannot lock into a crossed state that would be difficult to disentangle in two dimensions. Our results for the dynamics of two-dimensional models

may well extend to three dimensions.

We measure dynamical information through the integrated autocorrelation functions of components p_x, p_y of the total polarization $\mathbf{p}$,

$$\mathbf{p} = (p_x, p_y) = \sum_{i=1}^{N} \mathbf{p}_i, \tag{4.1}$$

where $\mathbf{p}_i$ is the oriented separation vector between the two disks in the ith dipole, possibly corrected for periodic boundary conditions (see Fig. 4.1c and d). For the disordered system studied here, p_x and p_y average to zero, and their fluctuations scale as $\sigma \sim \sqrt{N}$. For a long simulation with n correlated samples, the variance of the time average of p_x or p_y equals $\sigma^2 \tau_{\text{int}}/n$. We thus evaluate the quality of the MCMC algorithms through their value of τ_{int}. This integrated autocorrelation time τ_{int} is related to the characteristic time of fluctuations in the system,

$$\tau_{\text{int}} = \int_{-\infty}^{\infty} \frac{C(\tau)}{C(0)} \, d\tau, \tag{4.2}$$

with the connected autocorrelation function $C(\tau) = \langle p_i(t) p_i(t + \tau) \rangle_c$, where $i \in \{x, y\}$ and $\langle \ldots \rangle_c$ indicates that the product of the estimated mean values are subtracted. In this convention, if $C(\tau)$ has simple exponential decay, $C(\tau) = C(0) \exp(-|\tau|/\tau_0)$, τ_{int} corresponds to $2\tau_0$. For a reversible transition matrix, the exponential autocorrelation time τ_0 corresponds to the inverse gap between the two eigenvalues of the largest absolute value. Non-reversible transition matrices are not necessarily diagonalizable [80], but the correlation decay is still exponential at large times [74]. We estimate the integrated autocorrelation time τ_{int} using the formulation in terms of a time integral [157], and checked that using the variance of the mean signal calculated within blocks [160] yields a similar outcome. We express τ_{int} in units of a single trial (for the Metropolis algorithm) or on the order of a single event (for the ECMC algorithms) with, in each case, a computational effort of $O(1)$ per unit on a serial central processing unit (CPU).

For $N = 1$, analytic results are available for the dynamics of some of the algorithms discussed here [P1]. The center-of-mass motion of the single dipole then factors out, and the sample space Ω simply corresponds to the possible values of the polarization vector, in other words to the two-dimensional ring with inner radius r and outer radius R (see Ref. [P1]). The single-dipole polarization $\mathbf{p}$ is then uniformly distributed in this ring.

4.3 Reversible and Non-Reversible MCMC

In the present section, we introduce MCMC algorithms that will be used in Section 4.4. We also illustrate them for a single dipole ($N = 1$) and review some of their fundamental properties. We only consider algorithms where at each moment a single disk moves, as this can also be realized within ECMC in the presence of long-range Coulomb interactions [P6, 146]. Dipole rotations—that relax the polarization—are thus pieced together from subsequent displacements of single disks. This has already proven efficient in ECMC for polymer models and for elongated hard needles [131, 132]. Explicit continuous-time dipole rotations, in the presence of periodic boundary conditions, have not

been successfully implemented for ECMC and appear difficult to be set up for long-range Coulomb interactions. It is for this reason that we do not take them into consideration for the hard-disk dipole model.

The local MCMC moves discussed here feature small single-disk displacements at each time step (for the Metropolis algorithm) or continuous moves of a single disk with piecewise constant velocities as a function of continuous time (for the ECMC algorithms). Again, this is motivated by the future applications to more complicated systems such as water, for which non-local moves including cluster updates [183, 184] appear out of reach. Although we thus restrict our attention to a subclass of MCMC algorithms with very similar design principles, we will show in Section 4.4 that their properties differ substantially.

In the ECMC algorithms, a single active disk at time t_a and position $\mathbf{x}_a$ moves in a given direction with a constant velocity $\mathbf{v}_a$ as a function of the continuous time t underlying the Markov process. This constant velocity results in a straight-line trajectory of the active disk, and it terminates in an event at time t_t and position $\mathbf{x}_a + (t_t - t_a)\mathbf{v}_a$, which corresponds either to a hard-disk collision or to the dipole extension reaching the tether length. In both cases, a unique target disk is responsible for the violation of a constraint in the hard-disk dipole model that would occur if the active disk continued to move. Instead, after the event, the disk that was moving stops and the target disk becomes active, again with a velocity $\mathbf{v}_t$ that stays constant up to the next event. In addition, resamplings may take place at predefined times t_{res}, which are usually separated by a chain time τ_{chain}, that is, at times $t_{\text{res}} = \tau_{\text{chain}}, 2\tau_{\text{chain}}, \dots$. Resamplings also interrupt the straight-line trajectory of the currently active disk and, for example, draw a new velocity and a new active disk.

4.3.1 Reversible Local Metropolis Algorithms

The local Metropolis algorithm [72], for the dipole model, selects a random disk i at each time step $t = 1, 2, \dots$. For the cross-shaped displacement set, the proposed move for i is sampled uniformly between $-\delta$ and δ, randomly in the x- or y-direction. We refer to δ as the "range." For the square-shaped displacement set, the proposed move is uniformly sampled from a square of sides 2δ centered at zero, with both the x- and the y-components of the displacement sampled independently between $-\delta$ and δ. In both cases, the move is accepted if it violates no constraints. If the move is rejected, the configurations at times t and $t + 1$ are the same. The Metropolis algorithm, with the given displacement sets, satisfies the detailed-balance condition. For $N = 1$, the polarization $\mathbf{p}$ performs a random walk in the sample space Ω, that is, the ring with inner radius r and outer radius R. The algorithm can be considered irreducible for all N, although some blocked configurations [185] may pose problems [P3].

4.3.2 Straight ECMC

In straight ECMC [115], the velocity $\mathbf{v}_a$ of the active disk is maintained until a constraint is violated in an event (where it drops to zero), at which moment the target disk starts to move with the same velocity $\mathbf{v}_t = \mathbf{v}_a$. The direction of the velocity does not change at collision events or when the tether length is reached but only at resamplings at the end of an event chain. The absolute value $|\mathbf{v}_a|$ of the velocity is kept constant in resamplings. We analyze several variants for updating the direction angle
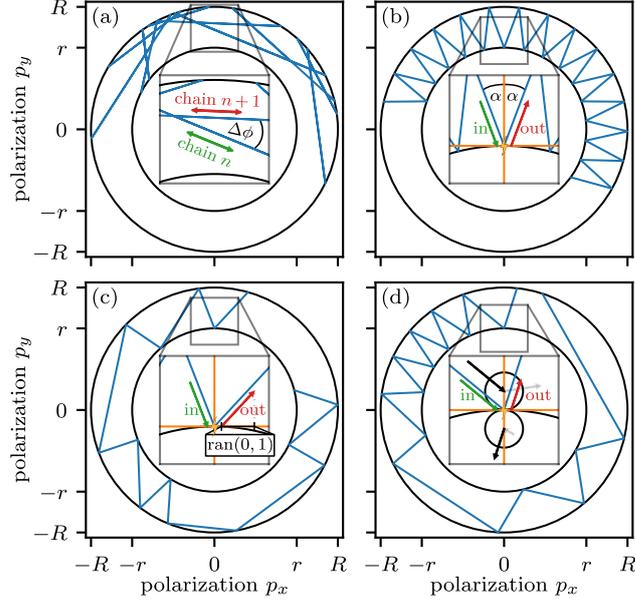
Figure 4.2: Sample trajectories for a single tethered dipole. The polarization **p** is the oriented vector between the two disk centers and evolves along the blue lines. (a): Sequential variant of straight ECMC [P1]. During any event chain between two resamplings, **p** changes on a straight line. At each resampling, the direction angle ϕ of this line with respect to the x-axis is incremented by the angle increment $\Delta\phi$ for the next chain. (b): In reflective ECMC [115], the effective velocity of **p** is reflected off the inner and outer ring in an event. (c): In forward ECMC [150], the outgoing reflection angle is non-deterministic in an event. (d): In Newtonian ECMC [124], every event considers the velocity labels of both hard disks.

ϕ with respect to the x-axis that describes the velocity in two dimensions: $\mathbf{v}_a = |\mathbf{v}_a|(\cos\phi, \sin\phi)$. In the "periodic" variant, the direction angle alternates between $\phi = 0$ and $\phi = \pi/2$ at each resampling. In arbitrary space dimensions, this would be generalized by the velocity cycling through the vectors of a constant absolute value aligned with the positive coordinate axes. The periodic variant of straight ECMC, used in most previous studies, is implemented in the JᴇʟʟʏFʏsʜ application [P6]. In the present paper, we also analyze the "random" variant where ϕ is uniformly sampled in the interval $[0, 2\pi)$ and, finally, the "sequential" variant [P1], where the direction angle ϕ is incremented by a constant angle increment $\Delta\phi$ at each resampling. The polarization trajectory between resamplings of straight ECMC is a straight line because every disk position is only altered with the same velocity vector in the same direction. In the periodic variant, subsequent straight-line trajectories are at right angles. They are at an angle $\Delta\phi$ in the sequential variant (see Fig. 4.2a for an example in $N = 1$). The described variants have very different time evolutions. The polarization trajectories of the sequential variant, for example, persistently rotate for $N = 1$, with a total rotation angle of the dipole that diverges as $\Delta\phi$ goes to zero [P1].

At a resampling of straight ECMC at time t_{res} (usually a multiple of τ_{chain}), the direction angle ϕ is updated, but the active disk must also be chosen randomly in order to satisfy global balance. To motivate this, we show for $N = 1$ that the two disks 1 and 2 must be active with the same probability at time t_{res}. Up until the next resampling, all polarizations $\mathbf{p}(t)$ lie on a line segment of the ring Ω
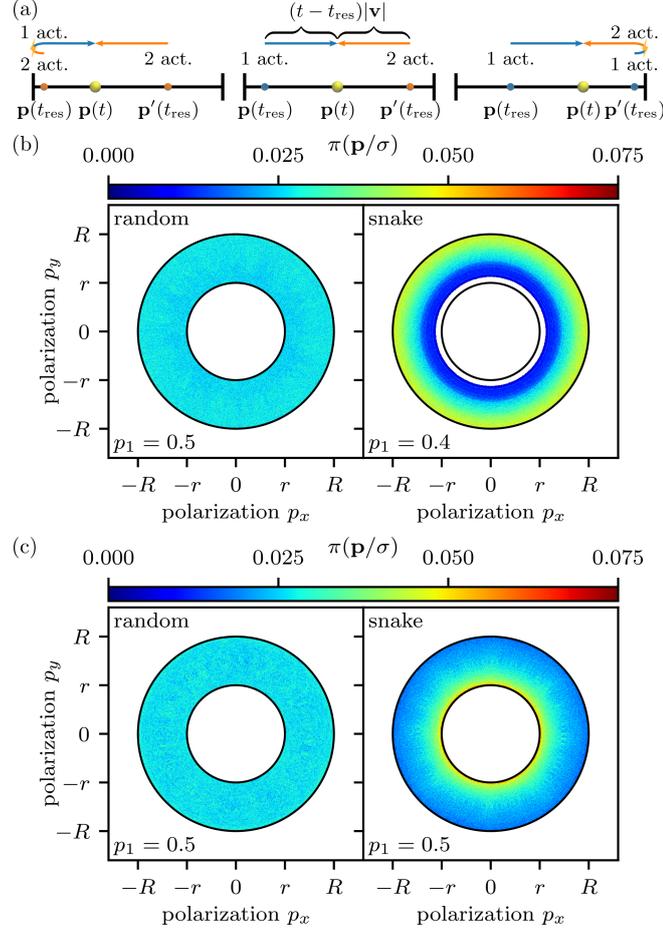
Figure 4.3: Global balance and irreducibility for straight sequential ECMC. (a): Flow into polarization $\mathbf{p}(t)$ from $\mathbf{p}(t_{\text{res}})$ and $\mathbf{p}'(t_{\text{res}})$ on a line segment of Ω as in Fig. 4.2a. The events take place at the boundaries of the segment, and the active disk changes. (b): Single-dipole distribution $\pi(\mathbf{p}/\sigma)$ for straight sequential ECMC with small angle increment $\Delta\phi$ using the random or snake mode for the choice of the active disk after a resampling. The snake-mode distribution is incorrect, and the probability p_1 that disk 1 is active right after a resampling is smaller than p_2. (c): Same as (b) but for a different initial configuration and chain time. The snake-mode distribution is incorrect although both disks are equally likely to be active after a resampling ($p_1 = p_2$).

determined by $\mathbf{p}(t_{\text{res}})$ and the current direction angle ϕ, and each of the two directions of motion on that segment corresponds to one of the two disks being active (see Fig. 4.3a). Global balance requires the flows into $\mathbf{p}(t)$ to be independent of its position. It can be reached from two polarizations $\mathbf{p}(t_{\text{res}})$ and $\mathbf{p}'(t_{\text{res}})$. Depending on the value of $\mathbf{p}(t)$, the corresponding polarizations $\mathbf{p}(t_{\text{res}})$ and $\mathbf{p}'(t_{\text{res}})$ either both correspond to disk 1 being active at t_{res}, or both to disk 2, or one to 1 and one to 2 (see Fig. 4.3a). Disks 1 and 2 must thus be equally likely ($p_1 = p_2$) to be active right after the resampling, a "random"-mode condition that is precisely implemented by the resampling of the active disk at t_{res}.

To illustrate the relevance of the random-mode requirement, we test for $N = 1$ the case when the active disk is the same before and after the resampling while the direction angle ϕ is incremented

by a small $\Delta\phi$ in the sequential variant of straight ECMC ("snake" mode). Clearly, snake mode is incorrect, as it samples Ω non-uniformly ($p_1 \neq p_2$, see Fig. 4.3b). It seems to be incorrect even when both disks happen to be initially active with practically the same probability ($p_1 = p_2$, see Fig. 4.3c). In both cases, the snake-mode trajectory is repetitive, and the lack of irreducibility in the lifted sample space $\widehat{\Omega} = \Omega \times \{\phi\}$ induces a non-uniform distribution in Ω. For $N > 1$ at very low density, sequential ECMC in the snake mode might not be irreducible because a single dipole deterministically rotates on a closed trajectory. However, no difference between the snake mode and random mode is detected for $N > 1$ at the higher densities considered in Section 4.4.

4.3.3 Reflective ECMC

Reflective ECMC [115] differs from the straight ECMC only in its handling of events so that the velocity vector is no longer constant. At an event, the velocity $\mathbf{v}_a$ of the incoming active disk is reflected by the line connecting the active and the target disks, and it becomes the velocity $\mathbf{v}_t$ of the target disk. This target disk is thus active after the event, while the disk that was active before the event stands still. The absolute value of the velocity is preserved in the reflection. For $N = 1$, the polarization follows straight lines that are symmetrically reflected off the inner and outer ring boundaries, with a single reflection angle α (see Fig. 4.2b).

Reflective ECMC satisfies global balance for any number of dipoles. For $N = 1$, and for an initial velocity that is nearly perpendicular to the separation between the two disks, the polarization may follow a "whispering-gallery-mode" trajectory that never visits small separations $\rho \gtrsim r$ (compare with Ref. [140, Fig. 3]), showing that uniform resamplings of the active disk and of the direction angle ϕ of the velocity with respect to the x-axis (keeping the absolute value of the velocity constant) are required for irreducibility. For all initial velocities at $N = 1$, even for those that visit all of sample space Ω (as in Fig. 4.2b), the probability distribution in the lifted sample space $\widehat{\Omega}$ does not separate into independent distributions, and the sampled stationary distribution in Ω is non-uniform, and thus incorrect. With periodic resamplings in intervals of the chain time τ_{chain}, the stationary distribution is uniform in $\widehat{\Omega}$ and therefore also in Ω. For $N > 1$ dipoles, in a large enough box, reflective ECMC fails to be irreducible without resamplings, as the dynamics is then deterministic and a single dipole may effectively rotate in a stationary closed trajectory. Nevertheless, at high enough density, we have not observed any difference in static observables with or without resamplings. It thus appears safe to omit them in practical simulations. As we will see in Section 4.4, resamplings do not improve the decorrelation because they only increase the integrated autocorrelation times. Without resamplings, reflective ECMC has no intrinsic parameters and is somewhat easier to set up. Computational overhead is avoided.

In previous work on hard disks [115] (rather than hard-disk dipoles), reflective ECMC was benchmarked against straight ECMC and found to be considerably slower. In contrast, in the dipole model studied here, reflective ECMC will prove more powerful than straight ECMC, even when the latter runs with optimized intrinsic parameters.

4.3.4 Forward ECMC

Forward ECMC [150] resembles reflective ECMC in that, after an event, the target-disk velocities $\mathbf{v}_t$ of both algorithms are located in the same quadrant of the coordinate system with axes parallel and orthogonal to the line connecting the disks at the event. The velocity is of constant absolute value, $|\mathbf{v}_t| = |\mathbf{v}_a| = 1$, but forward ECMC incorporates a random element into each event. More precisely, the component orthogonal to the line connecting the disks at contact is uniformly sampled between 0 and 1 (while reflecting the orthogonal orientation). Its parallel component is determined so that the velocity is of unit norm. For a single dipole, the polarization describes straight lines that are reflected off the inner and outer ring boundaries. However, the outgoing reflection angle is non-deterministic (see Fig. 4.2c). In contrast to straight and reflective ECMC, forward ECMC is irreducible even for $N = 1$, and it requires no resamplings. Also, as we will show in Section 4.4, uniform resamplings of the direction angle ϕ of the velocity with the x-axis and the active disk every chain time τ_{chain} do not speed up the algorithm.

4.3.5 Newtonian ECMC

Newtonian ECMC [124] mimics event-driven molecular dynamics [186]. All disks have velocity labels in addition to their positions. The label $\mathbf{v}_a$ of the active disk indicates the time derivative of its position, that is, its displacement in time. All other disks are stationary despite their velocity labels. Events (including those where the maximum dipole separation is reached) take place as in molecular dynamics, with all velocity labels treated on an equal footing, and with equal masses for all disks. The identities of the active and the target disks are interchanged in the event like in the other ECMC variants, and the target disk continues to move with its velocity label $\mathbf{v}_t$. During the simulation, the absolute value of the active-disk velocity varies. The equilibrium distribution of all velocity labels is uniform on a $2N$-dimensional sphere, and the root-mean-square velocity v_{rms} is conserved. The event rate per unit distance of the Newtonian ECMC equals the one of the other variants, but the event rate per unit time is smaller by $\sqrt{\pi}/2$ because of the difference between v_{rms} and the mean absolute value for the two-dimensional Gaussian distribution of velocities. In Newtonian ECMC, all the velocity labels are used as lifting variables. At a possible resampling every chain time τ_{chain}, these labels must be sampled from the exact equilibrium distribution (the rescaled Maxwell–Boltzmann distribution for all disks) together with the active disk. Without resampling, the choice of the initial velocity labels is arbitrary.

For $N = 1$, the polarization trajectory of Newtonian ECMC reflects off the inner and outer ring boundaries but considers both velocity labels at each event (see Fig. 4.2d). Without resamplings, Newtonian ECMC breaks irreducibility in the lifted sample space $\widehat{\Omega}$ for this single dipole, and it samples an incorrect stationary probability distribution, even if the full sample space Ω is visited. For $N \geq 2$, we find no influence on observable means of the resampling rate, so that the irreducibility problem is probably again due to the high symmetry of the ring geometry of the polarization for a single dipole. In the larger-N dipole systems in Section 4.4, we furthermore notice that the algorithm becomes faster in the limit of infinite resampling time.
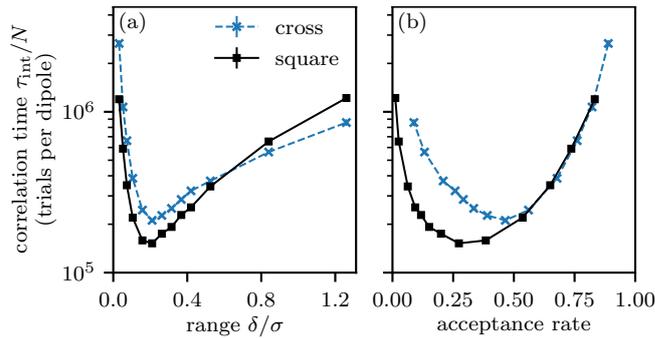
Figure 4.4: Correlation time per dipole τ_{int}/N of the Metropolis algorithm with cross- and square-shaped displacement sets for $N = 81$ dipoles at density $D = 0.70$. (a): τ_{int}/N vs. the rescaled range δ/σ. (b): τ_{int}/N vs. acceptance rate. The minimum of τ_{int}/N provides the reference for the ECMC benchmarks in Fig. 4.5.

4.4 Autocorrelations for *N* Dipoles

In the present section, we characterize the local MCMC algorithms of Section 4.3 via the autocorrelation dynamics of the polarization. The integrated autocorrelation times τ_{int} are separately computed for the components p_x and p_y of the polarization, with a unit of time corresponding to a single trial move for the Metropolis algorithm and to the mean event time for ECMC. As the computational effort per move or per event is $O(1)$, we make no effort to compare the (implementation-dependent) CPU times. Statistical errors are estimated from the difference between the integrated autocorrelation times of p_x and p_y, which, by symmetry, must be the same. For ECMC, we also take into account the uncertainty in the mean event time. Single MCMC-run times, for each parameter set, are at least three orders of magnitude longer than τ_{int}.

4.4.1 Intrinsic Parameters of Metropolis and ECMC

The correlation time of the local Metropolis algorithm depends on its intrinsic parameters, namely, the range δ and the choice of the displacement set (see Fig. 4.4a). Similarly, the acceptance rate of a move in the Metropolis algorithm depends on the displacement set and the range δ. The correlation time can thus be expressed as a function of the acceptance rate. As in many comparable models, the "one-half" rule [78] is roughly respected, and the Metropolis algorithm converges best for a rejection rate on the order of 50 %. For $N = 81$ dipoles at density $D = 0.70$, we observe a broad optimum between 20 % and 40 % for the square-shaped displacement set and an optimum close to 50 % for the cross-shaped displacement set (see Fig. 4.4b).

All ECMC variants allow for resamplings in intervals of the chain time τ_{chain} so that their correlation times depend on the intrinsic parameter τ_{chain}. The performance of the sequential variant of straight ECMC also depends on the angle increment $\Delta\phi$. Without resamplings, straight ECMC is not irreducible and its correlation time is infinite. The optimum τ_{int} is thus at a finite τ_{chain} to be obtained by fine-tuning (see the upper curves in Fig. 4.5). In its sequential variant, the cases $\Delta\phi = \pi/2$ and $\Delta\phi = \pi/3$ make the velocity cycle through a few values only. For these cases, we notice that τ_{int} has two local minima, both of which are rather large. In the absence of a heuristic for the choice of
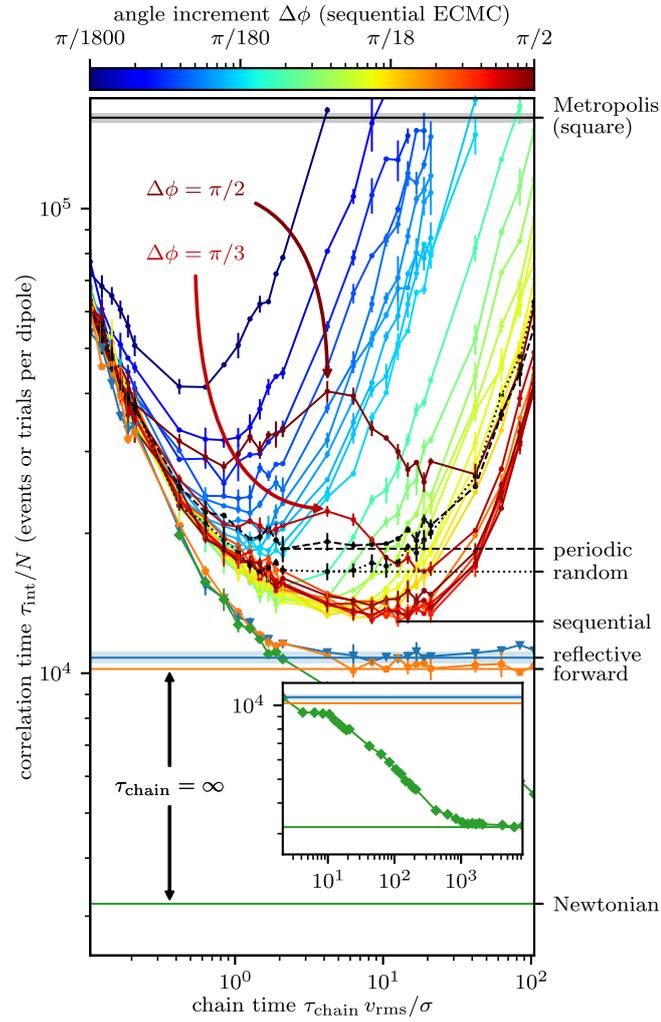
Figure 4.5: Correlation times per dipole τ_{int}/N of ECMC variants for different intrinsic parameters (rescaled chain time $\tau_{chain} v_{rms}/\sigma$, angle increment $\Delta\phi$) for $N = 81$ dipoles at density $D = 0.70$. The optimal values of τ_{int}/N are highlighted on the right y-axis. For reflective, forward, and Newtonian ECMC, they agree with those for $\tau_{chain} \to \infty$. The optimal performance of the Metropolis algorithm is indicated as the benchmark reference (see Fig. 4.4). The inset illustrates the evolution of τ_{int}/N for large $\tau_{chain} v_{rms}/\sigma$ for Newtonian ECMC.

intrinsic parameters, straight ECMC requires explicit fine-tuning, which increases its computational complexity.

As discussed in Section 4.3, reflective, forward and Newtonian ECMC appear irreducible for $N > 1$ even without resamplings. They are also fastest in this limit $\tau_{chain} \to \infty$ (see the lower curves in Fig. 4.5). This was also observed, e.g., for reflective ECMC in two-dimensional hard-disk systems [115], and for Newtonian ECMC in three-dimensional hard spheres [124]. For moderate values of τ_{chain}, Newtonian ECMC is comparable to other variants. However, for much larger τ_{chain}, its correlation times again decrease strongly (see the inset of Fig. 4.5). For $N = 81$ dipoles at density
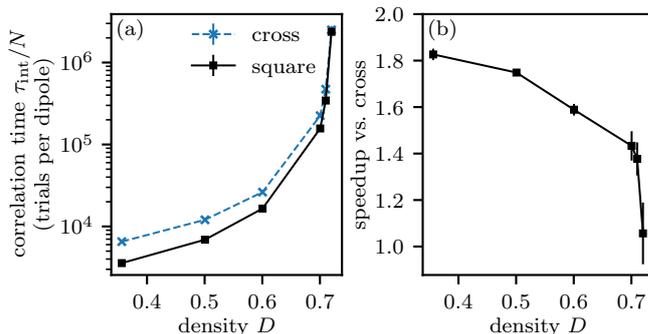
Figure 4.6: Performance of the Metropolis algorithm for $N = 81$ dipoles (optimized intrinsic parameters). (a): Correlation time per dipole τ_{int}/N vs. density D for the cross- and square-shaped displacement sets. (b): Speedup for the square-shaped displacement set compared to the cross-shaped displacement set.

$D = 0.70$, the variants without intrinsic parameters that thus require no fine-tuning perform best. Newtonian ECMC without resampling accelerates with respect to the local Metropolis algorithm by a surprising factor of 50 at density 0.70, which seems to further increase at even higher density (see Section 4.4.2).

4.4.2 Density and Size Dependence of Speedups

For both versions of the Metropolis algorithms, the autocorrelation time τ_{int} naturally increases at higher densities and climbs steeply for densities $D \gtrsim 0.70$, around the liquid–hexatic phase-transition density of the hard-disk system without the tether constraint (see Fig. 4.6a, again for $N = 81$ dipoles). The square-shaped displacement set is somewhat preferable over the cross-shaped displacement set for small densities, but the two become equally fast at large densities (see Fig. 4.6b). This illustrates that, while step-size control is a key feature in reversible MCMC, the details of a local reversible algorithm do not really depend on the displacement sets. For simplicity, we chose the optimal acceptance rates for $D = 0.70$ for these runs at other densities (see Fig. 4.4b).

To identify the optimal correlation times of the straight ECMC variants, we perform systematic scans in the chain time τ_{chain} and the angle increment $\Delta\phi$, as in Section 4.4.1, for each density at $N = 81$. For forward, reflective, and Newtonian ECMC, we use $\tau_{chain} = \infty$ without any fine-tuning. For all ECMC variants, τ_{int} increases with density (see Fig. 4.7a). The periodic variant of straight ECMC, which is the fastest for two-dimensional hard disks, is the slowest ECMC variant for hard-disk dipoles. The random and sequential variants are somewhat faster. Reflective and forward ECMC resemble each other in performance at all densities D. While they are considerably faster than the straight variants at low D, this gap vanishes at $D = 0.72$. Newtonian ECMC is the fastest algorithm for all densities, and as mentioned, it requires no fine-tuning. All ECMC algorithms outperform the Metropolis algorithm with optimized intrinsic parameters by a considerable margin. For the latter, the square-shaped displacement set is somewhat faster than the cross-shaped one (see Fig. 4.7b). With increasing density, the speedup of reflective and forward ECMC drops from roughly 30 to 15, while the speedup of the straight ECMC variants increases with density until they become comparable. These trends were similarly observed near the liquid–hexatic transition of two-dimensional hard disks
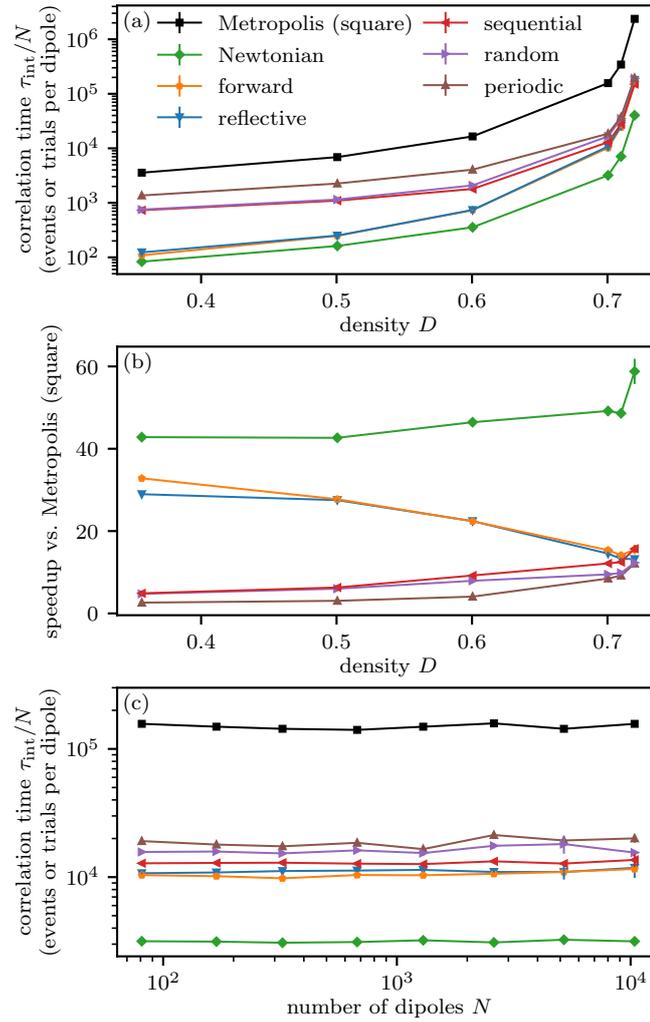
Figure 4.7: Performance of ECMC variants and of the Metropolis algorithm with a square-shaped displacement set (where applicable: with optimized intrinsic parameters). (a): Correlation time per dipole τ_{int}/N for $N = 81$ dipoles vs. density D. (b): Speedup of ECMC with respect to the Metropolis algorithm. (c): Correlation time per dipole τ_{int}/N at density $D = 0.70$ vs. number of dipoles N.

for periodic and reflective ECMC [115] (where the straight periodic variant, however, was considerably faster than reflective ECMC at high densities). The speedup realized by Newtonian ECMC with respect to the Metropolis algorithm with an optimized range and displacement set improves slightly from low to high densities, as was likewise found for three-dimensional hard spheres [124]. We observe the highest speedup of roughly a factor of 60 at the highest density $D = 0.72$ that we studied.

We finally study the dependence of the correlation time τ_{int} of the number of dipoles N at the density $D = 0.70$. For the reflective, forward, and Newtonian ECMC variants, we simply set $\tau_{\text{chain}} = \infty$, given our findings at $N = 81$. For periodic, random, and sequential ECMC, we infer the optimal τ_{chain} and $\Delta\phi$ from the $N = 81$ case (performing occasional sweeps through intrinsic parameters as

cross-checks). For the Metropolis algorithm, we used the optimal square-shaped displacement set and the acceptance rate of the $N = 81$ system. We observe that the autocorrelation time τ_{int} increases linearly with N for all algorithms so that τ_{int}/N is essentially independent of N (see Fig. 4.7c). The considerable speedups realized by ECMC with respect to the Metropolis algorithm thus seem to be preserved for large system sizes.

4.5 Conclusions and Outlook

In this paper, we have systematically studied local MCMC algorithms for complex molecules that we caricatured as hard-disk dipoles with parameters inspired by the SPC/Fw water system. Moving from three to two spatial dimensions and from a liquid of charged molecules to a model of hard-disk dipoles has greatly simplified the algorithm development and its implementation. This allowed us to scan thousands of parameter sets for different ECMC variants and to benchmark them against the reversible local Metropolis algorithm. At this exploratory stage, this would have required prohibitive computing resources for the original three-dimensional model [149]. The broad spread of the performance of different ECMC versions came as a surprise. We found that algorithms with intrinsic parameters perform considerably less well than others and demonstrated a 60-fold speedup of Newtonian ECMC with respect to the Metropolis algorithm. This speedup can be achieved without any fine-tuning. It is unclear why straight ECMC performs worse for hard-disk dipoles than the reflective, forward, and Newtonian variants, while for simple hard disks straight ECMC is clearly the most efficient. Our study has also brought out subtleties of ECMC, which are hidden in simple liquids. For the case of a single dipole, we thus discovered strict requirements for resampling in order to ensure irreducibility of straight, reflective, and Newtonian ECMC. However, these requirements did not seem to play a role for denser systems of dipoles. Event-based randomness as in forward ECMC (or as in all ECMC simulations of soft interactions) strictly ensures irreducibility.

We expect our observations to carry over from the caricature two-dimensional dipoles to the three-dimensional SPC/Fw systems and related models. There are two reasons why we have not yet studied this system directly. One is, as mentioned, the scale of the computing requirements for a full-fledged three-dimensional scan. The other is that further algorithm development is still required. The cell-veto algorithm [148] [which reduces the computational complexity of the Coulomb problem to $O(N \log N)$ per sweep of events, without evaluating the Coulomb energy] is in the present version of our open-source project only implemented for the periodic variant of straight ECMC. However, we expect it to generalize to all ECMC variants for flexible water molecules with an explicit Coulomb interaction. The benchmark against Metropolis MCMC will then be even more favorable, as the change in the Coulomb energy can there only be computed in $O(N^{3/2})$ per sweep of moves [98]. The key question will be whether the dynamic arrest of straight ECMC for the three-dimensional water system can be overcome as in the two-dimensional hard-disk dipole model studied here. The extreme dependence of the performance of non-reversible MCMC on details of the algorithm was also evidenced in the escape times from a tightly confined initial configuration [P3], which might be relevant for overcoming dynamic arrest.

Acknowledgments

Author Declarations

Conflict of Interest

The authors have no conflicts to disclose.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The figures in this paper were prepared with the Matplotlib [187] and Visual Molecular Dynamics [188] (VMD) software packages. Integrated autocorrelation times were computed by the Python `emcee` package [159].

Research Paper 3—Sparse Hard-Disk Packings and Local Markov Chains

This chapter considers the publication

Section 5.0 provides a summary of the publication and contextualizes it within this doctoral thesis. Furthermore, the author's contributions are clearly indicated. The remaining part of this chapter reproduces the published version of the publication itself. All co-authors of the publication agree with its use for this doctoral thesis and with the identification of the author's contributions.

5.0 Prologue

Thematically, the reproduced publication in this chapter is the furthest away from the overarching objective of this doctoral thesis, that is, the introduction of a competitive, rigorous paradigm for molecular simulations that is based on non-reversible Markov-chain Monte Carlo (MCMC) algorithms and that overcomes, in a first step, the dynamical arrest of the polarization of SPC/Fw water molecules of straight event-chain Monte Carlo (ECMC) as implemented in JELLYFYSH-Version1.0. In that context, it shows again that different ECMC variants can show widely different behaviors as already observed in the second research paper of this doctoral thesis (see Chapter 4 and Ref. [P2]).

We consider locally stable packings of N hard disks in a two-dimensional periodic box that were originally introduced by K. Böröczky in 1964 [185]. Böröczky packings are sparse because their density vanishes in the limit $N \to \infty$, yet they are locally stable because no single disk can move infinitesimally as it is blocked in any direction by a neighbored disk (see Fig. 5.1). By slightly reducing all disk radii by a factor $(1 - \varepsilon)$ with a small relaxation parameter $\varepsilon \gtrsim 0$, we obtain tightly confined ε-relaxed Böröczky configurations that can be used to analyze and benchmark local MCMC algorithms.

Specifically, we study how the straight, reflective, forward, and Newtonian ECMC variants escape from ε-relaxed Böröczky configurations. We also consider the Metropolis algorithm with Monte Carlo moves that are chosen small enough so that disks cannot simply jump out of their confinement. (Such jumps may be prevented by the presence of other disks when the ε-relaxed Böröczky configuration is the backbone of a much denser hard-disk configuration.)

A scaling theory for the escape times predicts two entirely different classes of the algorithms. For straight ECMC and the Metropolis algorithm, the escape time diverges algebraically as $\varepsilon \to 0$. For reflective, forward, and Newtonian ECMC, in contrast, it diverges logarithmically (see Fig. 5.6). The logarithmic divergence is tested up to $\varepsilon = 10^{-29}$ in numerical simulations by using an arbitrary-precision implementation of ECMC. Our scaling theory suggests that this power-law-to-logarithm speedup of reflective, forward, and Newtonian ECMC with respect to straight ECMC is a combination of their missing intrinsic scale and their event-driven nature. Straight ECMC moves all hard disks in the same direction in an event chain between changes of the velocity in periodic time intervals of the chain time τ_{chain}. This results in a rigidity problem where all moves in an overly long event chain effectively cancel each other out. One thus has to choose an optimal chain time τ_{chain} that minimizes the escape time (see Fig. 5.4). This choice balances between two limiting cases. First, small τ_{chain} are favored by the requirement for short event chains in the tightly confined ε-relaxed Böröczky configuration that do not cancel each other out. Second, large τ_{chain} are favored by the requirement to move disks by a given distance to actually escape. Similarly, the Metropolis algorithm has to choose an optimal proposal distribution for its Monte Carlo moves. Reflective, forward, and Newtonian ECMC change the direction in every event (see Fig. 1.5) so that they are not subject to the rigidity problem of straight ECMC and may escape from ε-relaxed Böröczky configurations without resamplings of the velocity in the $\tau_{\text{chain}} \to \infty$ limit. The average distance that the active disk travels between events grows as the configuration is gradually escaped (see Fig. 5.5). At the same time, the computational cost for every event remains constant in the event-driven implementation.

The problem of the choice of an intrinsic scale in straight ECMC is somewhat reminiscent of the choice of a particular time step in molecular dynamics (MD). Rare configurations with excessive forces between unusually close interaction sites would require a smaller time step than necessary in the more well-behaved configurations of the greatest part of the simulation. (This motivates variable time step integrators in MD that are, however, not symplectic; see, e.g., Ref. [43] and references therein.) More generally, the time step has to be chosen small enough for an accurate integration of the fastest motion although slower motions could, in principle, use larger time steps (which motivates multiple time-scale integration algorithms [15]). In contrast to MD, straight ECMC does not have any problem with stability. Still, choosing a particular value of the chain time τ_{chain} does not only require a laborious fine-tuning, but the best choice may actually change during a simulation.

In the context of the overarching objective of this doctoral thesis, we conclude that it is essential to find the most appropriate non-reversible MCMC algorithm for molecular simulations because their performance may widely differ. Also, the explicit choice of an intrinsic scale in straight ECMC could be one reason for its dynamically-arrested simulations of the complex SPC/Fw water model.

In a more general context, we conjecture that Böröczky packings exist for any hard-disk density (up to the densest possible packing) for a large enough number N of hard disks. This conjecture relies

on their flexible construction (see Section 5.2.1). Here, they can also appear as backbones in dense hard-disk configurations where the majority of disks is not confined in the packing. This makes it important to consider their implications on general local MCMC simulations of hard-disk systems.

The considered local MCMC algorithms, ECMC and the Metropolis algorithm with small enough Monte Carlo moves, can never escape from a locally stable Böröczky packing. In the definition of the mixing time, locally stable packings could thus serve as the most unfavorable initial configuration which would yield an infinite mixing time (see Section 2.2.1). However, because the packing contains direct contacts between hard disks, they make up only a set of measure zero in the hard-disk configuration space and may thus be deemed irrelevant. Still, even when locally stable packings are excluded from the configuration space, there is no general proof that any local MCMC algorithm is irreducible in the hard-disk system in a (periodic) box. We discuss how the undoubted usefulness of MCMC can be remedied in the canonical (NVT) ensemble and more rigorous in the isothermal–isobaric (NPT) ensemble (see Section 5.4.2).

In contrast to Böröczky packings, ε-relaxed Böröczky configurations (and equivalent "rattled" configurations where every disk is at a random position within its tight confinement) make up a *finite* portion in the configuration space. Because of the divergent escape times as $\varepsilon \rightarrow 0$, there always exists such a finite portion that is practically excluded in simulations of local MCMC algorithms. The simulations cannot escape from this excluded portion in a given upper limit of computation time, and they can probably not even access it. Since Markov chains and their mixing times can be characterized by a single bottleneck that partitions the configuration space into two pieces (see Section 2.3), we argue that escape times ε-relaxed Böröczky configurations may well model the probability flow across such a bottleneck (see Section 5.4.1.2). Therefore, comparing escape times may reflect relative merits of local MCMC algorithm as it already did for the different ECMC variants.

ECMC is a non-reversible lifting of the reversible factorized Metropolis algorithm that is equivalent to the original Metropolis algorithm for the hard-disk system (see Sections 2.4.2 and 2.4.3). As such, it can reduce the mixing time by, at most, a square-root factor (see Section 1.3.2). For the escape times from ε-relaxed Böröczky configurations, we observe, however, a power-law-to-logarithm speedup in the escape time for some of the ECMC variants (see Fig. 5.6 again). Here, it is important to note that the escape time is measured in events for the ECMC algorithms because they have the same $O(1)$ complexity as a Monte Carlo move in the Metropolis algorithm. The average continuous Monte Carlo time between events grows as the ε-relaxed Böröczky configurations is gradually escaped. If the escape times are directly related to mixing times, and if one would measure all escape times in terms of Monte Carlo times, one would most probably observe a reduction of the mixing time that conforms to the mathematical bounds. This emphasizes, however, that the event-driven nature of ECMC can lead to larger speedups than one would expect from the theory of lifted Markov chains when measured in computation time.

Since we propose to use the escape times from ε-relaxed Böröczky configurations to analyze and benchmark general local MCMC algorithms, the reproduced publication of this chapter is accompanied by the open-source, arbitrary-precision software package BIGBORO [P8]. It provides an easy access to the construction of Böröczky packings in a periodic box. Furthermore, it can compute collective escape modes from Böröczky packings that infinitesimally displace all disks at once (which shows that

the packing is only locally but not collectively stable). Finally, it also includes an arbitrary-precision implementation of straight, reflective, forward, and Newtonian ECMC that can start from ε-relaxed Böröczky configurations with arbitrary relaxation parameters $\varepsilon \gtrsim 0$. Besides allowing to reproduce our data, this allows to diagnose and study gridlocks where, even though the configuration was initially relaxed, the ECMC simulation gets trapped in a subset of the configuration in which disks are in direct contact and the event rate diverges (see Section 5.3.2.4). Such gridlocks appear most frequently in Newtonian ECMC. An understanding of this effect may thus give a deeper insight into the nature of its dynamics.

Contributions of the Author

The author contributed in an essential manner to the entire research project underlying this research paper. This includes, in particular, essential contributions to the implementation of the open-source, arbitrary-precision software package BigBoro and the data collection, the creation of the figures, and the writing of the research paper itself. Following the Contributor Role Taxonomy (CRediT) [174], the author contributions are as follows:

- Methodology (development or design of methodology; creation of models).

- Software (programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components).

- Validation (verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs).

- Formal analysis (application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data).

- Investigation (conducting a research and investigation process, specifically performing the experiments, or data/evidence collection).

- Writing—Original Draft [preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)].

- Writing—Review & Editing (preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision—including pre- or post-publication stages).

- Visualization (preparation, creation and/or presentation of the published work, specifically visualization/data presentation).

Sparse Hard-Disk Packings and Local Markov Chains

This reproduces P. Höllmer, N. Noirault, B. Li, A. C. Maggs, and W. Krauth, *Sparse Hard-Disk Packings and Local Markov Chains*, Journal of Statistical Physics **187** (2022) 31, which is licensed under the Creative Commons Attribution 4.0 International License and may be found at `https://doi.org/10.1007/s10955-022-02908-4` (see Ref. [P3]).

Abstract We propose locally stable sparse hard-disk packings, as introduced by Böröczky, as a model for the analysis and benchmarking of Markov-chain Monte Carlo (MCMC) algorithms. We first generate such Böröczky packings in a square box with periodic boundary conditions and analyze their properties. We then study how local MCMC algorithms, namely the Metropolis algorithm and several versions of event-chain Monte Carlo (ECMC), escape from configurations that are obtained from the packings by slightly reducing all disk radii by a relaxation parameter. We obtain two classes of ECMC, one in which the escape time varies algebraically with the relaxation parameter (as for the local Metropolis algorithm) and another in which the escape time scales as the logarithm of the relaxation parameter. A scaling analysis is confirmed by simulation results. We discuss the connectivity of the hard-disk sample space, the ergodicity of local MCMC algorithms, as well as the meaning of packings in the context of the NPT ensemble. Our work is accompanied by open-source, arbitrary-precision software for Böröczky packings (in Python) and for straight, reflective, forward, and Newtonian ECMC (in Go).

5.1 Introduction

The hard-disk system is a fundamental statistical-physics model that has been intensely studied since 1953. Numerical simulations, notably Markov-chain Monte Carlo [72] (MCMC) and event-driven molecular dynamics [186], have played a particular role in its study. The existence of hard-disk phase transitions [144] was asserted as early as 1962. The recent identification of the actual transition scenario [142] required the use of a modern event-chain Monte Carlo (ECMC) algorithm [80, 115].

The hard-disk model has been much studied in mathematics. Even today, the existence of a phase transition has not been proven [189, 190]. A fundamental rigorous result is that the densest packing of N equal hard disks (for $N \rightarrow \infty$) arranges them in a hexagonal lattice [191]. This densest packing is locally stable, which means that no single disk can move infinitesimally in the two-dimensional plane. The densest packing is furthermore collectively stable, which means that no subset of disks can move at once, except if the collective infinitesimal move corresponds to symmetries, as for example uniform translations in the presence of periodic boundary conditions [192–194]. In 1964, Böröczky [185] constructed two-dimensional disk packings that are sparse, that is, have vanishing density in the limit $N \rightarrow \infty$. The properties of these Böröczky packings are very different from those of the densest hexagonal lattice. Infinitesimal motion of just a single disk remains impossible, so that Böröczky packings are locally stable. However, coherent infinitesimal motion of more than one disk does allow escape from Böröczky packings so that they are not collectively stable.

In this work, we construct finite-N Böröczky packings in a fixed periodic box and use them to build initial configurations for local Markov-chain Monte Carlo (MCMC) algorithms, namely the

reversible Metropolis algorithm [72, 78] and several variants [115, 124, 150] of non-reversible ECMC. In the Metropolis algorithm, single disks are moved one by one within a given range δ. A Böröczky packing traps the local Metropolis algorithm if δ is small enough, because all single-disk moves are rejected. ECMC is by definition local. It features individual infinitesimal displacements of single disks, and it also cannot escape from a Böröczky packing. We thus consider ε-relaxed Böröczky configurations that have the same disk positions as the Böröczky packings but with disk radii reduced by a factor $(1 - \varepsilon)$. Here, $\varepsilon \gtrsim 0$ is the relaxation parameter. Our scaling theory for the escape times from ε-relaxed Böröczky configurations predicts the existence of two classes of local Markov-chain algorithms. In one class, escape times grow as a power of the relaxation parameter ε, whereas the other class features only logarithmic growth. Numerical simulations confirm our scaling theory, whose power-law exponents we conjecture to be exact. The ε-relaxed Böröczky configurations are representative of a finite portion of sample space. For a fixed number of disks, the growth of the escape times thus leads to the existence of a small but finite fraction of sample space that cannot be escaped from or even accessed by local MCMC in a given upper limit of CPU time. More generally, we discuss the apparent paradox that the lacking proof for the connectedness of the hard-disk sample space, on the one hand, might render local MCMC non-irreducible (that is, "non-ergodic") but, on the other hand, does not invalidate their practical use. We resolve this paradox by considering the NPT ensemble (where the pressure is conserved instead of the volume). We moreover advocate the usefulness of ε-relaxed Böröczky configurations for modeling bottlenecks in MCMC and consider the comparison of escape times from these configurations as an interesting benchmark. We provide open-source arbitrary-precision software for Böröczky packings and for ECMC. Several of the ECMC algorithms can evolve towards numerical gridlock, that can be diagnosed and studied using our arbitrary-precision software.

This work is organized as follows. In Section 5.2, we construct Böröczky packings following the original proposal [185] and a variant due to Kahle [195], and we analyze their properties. In Section 5.3, we discuss local MCMC algorithms and present analytical and numerical results for the escape times from the ε-relaxed Böröczky configurations. In Section 5.4, we analyze algorithms and their escape times and discuss fundamental aspects, among them irreducibility, statistical ensembles, as well as the question of bottlenecks, and the difference between local and non-local MCMC methods. In the conclusion (Section 5.5), we point to several extensions and place our findings into the wider context of equilibrium statistical mechanics, the physics of glasses and the mechanics of granular materials. In Appendix 5.A, we present further numerical analysis and, in Appendix 5.B, we introduce our open-source arbitrary-precision software package `BigBoro` for Böröczky packings and for ECMC.

5.2 Böröczky Packings

In the present section, we discuss Böröczky packings of N disks of radius $\sigma = 1$ in a periodic square box of sides L. The density η is the ratio of the disk areas to that of the box:

$$\eta = N\pi\sigma^2/L^2. \tag{5.1}$$

For concreteness, the central simulation box ranges from $-L/2$ to $L/2$ in both the x and the y direction. The periodic boundary conditions map the central simulation box onto an infinite hard-disk system with periodically repeated boxes or, equivalently, onto a torus. A Böröczky packing is locally stable, and each of its N disks is blocked—at a distance 2σ—by at least three other disks (taking into account periodic boundary conditions), with the contacts not all in the same half-plane. The opening angle of a disk i, the largest angle formed by the contacts to its neighbors, is then always smaller than π. The maximum opening angle is the largest of the N opening angle of all disks. Clearly, a locally stable packing cannot be escaped from through the infinitesimal single-disk moves of ECMC or, in Metropolis MCMC, through steps of small enough range. Only collective infinitesimal moves of all disks may escape from the packing.

In a nutshell, Böröczky packings (see Section 5.2.1 for their construction) consist in cores and branches (as visible in Fig. 5.1). The original Ref. [185] mainly focused on Böröczky packings in an infinite plane, but also sketched how to generalize the packings to the periodic case. Böröczky packings can exist for different cores, and they depend on a bounding curve (more precisely: a convex polygonal chain) which encloses the branches, and which can be chosen more or less freely (see Section 5.2.2 for the properties of Böröczky packings, including the collective infinitesimal escape modes from them).

5.2.1 Construction of Böröczky Packings

In the central simulation box, a finite-N Böröczky packing is built on a central core placed around $(0, 0)$ (see Section 5.2.1.1 for a discussion of cores). This core connects to four periodic copies of the core centered at $(L, 0)$, $(0, L)$, $(-L, 0)$, and $(0, -L)$ by branches that have k separate layers (see Sections 5.2.1.2 and 5.2.1.3 for a detailed discussion of branches). A Böröczky packing shares the symmetries of the central simulation box. Cores with different shapes, as for example that of a triangle, yield Böröczky packings in other geometries (see Refs [185, 195] and Ref. [196, Section 9.3]).

5.2.1.1 Böröczky Core, Kahle Core

We consider Böröczky packings with two different cores, either the Böröczky core or the Kahle core. Both options are implemented in the `BigBoro` software package (see Appendix 5.B). The Böröczky core [185] consists of 20 disks (see Fig. 5.1a). Using reflection symmetry about coordinate axes and diagonals, this core can be constructed from four disks at coordinates $(\sqrt{2}, 0)$, $(2 + \sqrt{2}, 0)$, $(2 + \sqrt{6}/2 + 1/\sqrt{2}, \sqrt{6}/2 + 1/\sqrt{2})$, and $(2 + \sqrt{6}/2 + 1/\sqrt{2}, 2 + \sqrt{6}/2 + 1/\sqrt{2})$ (see highlighted disks in Fig. 5.1a). The Kahle core [195], with a total of 8 disks, is constructed from two disks at coordinates $(1, 1)$, and $(1 + \sqrt{3}, 0)$, using the same symmetries (see highlighted disks in Fig. 5.1b). The Böröczky core for $k = 0$, that is without the branches included in Fig. 5.1a, is only locally stable if repeated periodically in a central simulation box that fully encloses the core disks, with $L/2 = 3 + \sqrt{6}/2 + 1/\sqrt{2}$. The Kahle core, again without branches, can be embedded in two non-equivalent ways into a periodic structure. When the outer-disk centers are placed on the cell boundaries, with $L/2 = 1 + \sqrt{3}$, it forms a collectively stable packing with no remaining degrees of freedom other than uniform translations. Alternatively, it only forms a locally stable packing, with the possibility of non-trivial collective
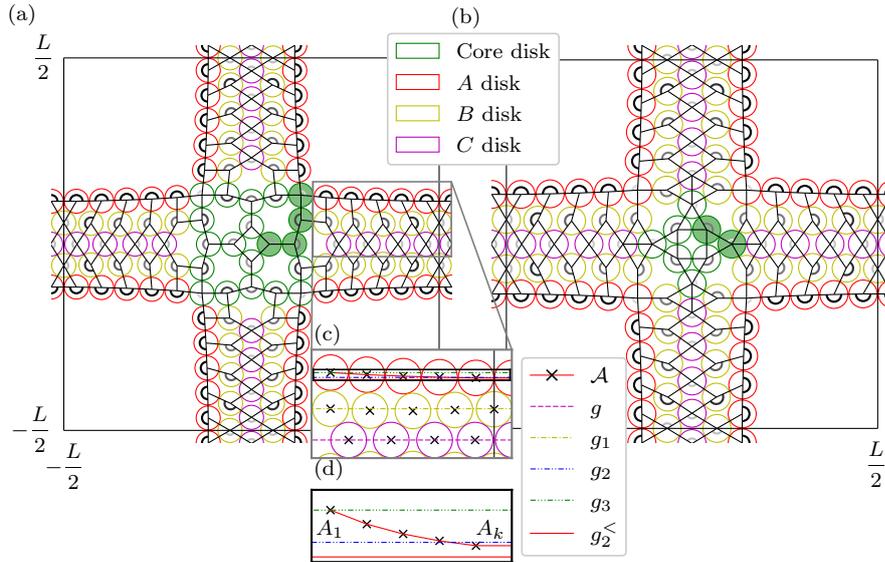
Figure 5.1: Hard-disk Böröczky packings, composed of a core and of four branches with $k = 5$ layers, with contact graphs and highlighted opening angles. (a): Packing with the Böröczky core [185]. (b): Packing with the Kahle core [195]. (c): Detail of a branch. (d): Convex polygonal chain $\mathcal{A}$, and horizontal lines $g_2^<$, g_2, and g_3. Two different classes of polygonal chains, called $\mathcal{A}^{\text{geo}}$ and $\mathcal{A}^{\text{circ}}$, are considered in this work

.

deformations, if the outer disks are enclosed in a larger simulation cell, with $L/2 = 2 + \sqrt{3}$. These two cores are the seeds from which larger and less dense Böröczky packings are now constructed and studied.

5.2.1.2 Branches—Infinite-Layer Case (Infinite N)

Following Ref. [185], we first construct infinite branches ($k = \infty$) that correspond to the $N \to \infty$ and $\eta \to 0$ limits, without periodic boundary conditions. One such branch is attached to each of the four sides of the central core so that all disks are locally stable. The horizontal branch that extends from the central core in the positive x-direction is symmetric about the x-axis. The half branch for $y \geq 0$ uses three sets of disks $\{A_1, A_2, \ldots\}$, $\{B_1, B_2, \ldots\}$, and $\{C_1, C_2, \ldots\}$, where $i = 1, 2, \ldots$ is the layer index.

For the branch that is symmetric about the x-axis, the construction relies on four horizontal lines [185]:

$$
\begin{array}{c|cccc}
\text{horizontal line} & g & g_1 & g_2 & g_3 \\
\hline
y\text{-value} & 0 & \sqrt{3} & 2\sqrt{3} & \sqrt{3} + 2
\end{array}
\qquad (5.2)
$$

The disks A_1 and B_1 are aligned in x at heights g_3 and g_1, respectively. All A disks lie on a given convex polygonal chain $\mathcal{A}$ between g_2 and g_3. The chain segments on $\mathcal{A}$ are of length 2 so that subsequent disks A_i and A_{i+1} block each other, and the position of A_1 fixes all other A disks. All C disks lie on g, and C_i blocks B_i from the right (in particular, C_1 is placed after B_1). The disk B_i, for $i > 1$, lies between g and g_1 and it blocks disks A_i and C_{i-1} from the right. With the position of g_2,

the branch approaches a hexagonal packing for $i \to \infty$. After reflection about the x-axis, all disks except A_1 and B_1 are locally stable in the infinite branch.

The Böröczky packing is completed by attaching the four branches along the four coordinate axes to a core. For the Böröczky core, both A_1 and B_1 are blocked by core disks (see Fig. 5.1a). For the Kahle core, B_1 is blocked by a core disk, and A_1 is locally stable as it also belongs to another branch (see Fig. 5.1b).

5.2.1.3 Branches—Finite-Layer Case (Finite N), Periodic Boundary Conditions

Branches can also be constructed for periodic simulation boxes, with a finite number k of layers and finite N (see Ref. [185]). The branch that connects the central core placed around $(0, 0)$ with its periodic image around $(L, 0)$ is then again symmetric about the x-axis but, in addition, also about the boundary of the central simulation box at $x = L/2$. We describe the construction of the half-branch (for $y \geq 0$) up to this boundary (see Fig. 5.1).

For half-branches with a finite number of layers k and a finite number of disks $\{A_1, \ldots, A_k\}$, $\{B_1, \ldots, B_k\}$, and $\{C_1, \ldots, C_{k-1}\}$ (with their corresponding mirror images), the convex polygonal chain $\mathcal{A}$ lies between $g_2^<$ and g_3 where $g_2^<$ is an auxiliary horizontal line placed slightly below g_2. The horizontal lines g and g_1 and the algorithm for placing the disks are as in Section 5.2.1.2 (see Fig. 5.1c and d). By varying the distance between g_2 and $g_2^<$, one can make disk B_k satisfy the additional requirement $x_{B_k} = x_{A_k} + 1$ that allows for periodic boundary conditions. The position of B_k then fixes the boundary of the square box ($x_{B_k} = L/2$) and B_k blocks A_k as well as the mirror image A_{k+1} of A_k (see Fig. 5.1c again).

5.2.2 Properties of Böröczky Packings

The local stability of Böröczky packings only relies on the fact that all A disks lie on a largely arbitrary convex polygonal chain $\mathcal{A}$ [185]. The choice of $\mathcal{A}$ influences the qualitative properties of the packing. The `BigBoro` software package (see Appendix 5.B) implements two different classes of convex polygonal chains that we discuss in Section 5.2.2.1. Another computer program in the package explicitly determines the space of collective escape modes from a Böröczky packing, which we discuss in Section 5.2.2.2.

5.2.2.1 Convex Polygonal Chains (Geometric, Circular)

In the convex geometric chain $\mathcal{A}^{\text{geo}}$ (which is for instance used in Fig. 5.1), the disks A_i approach the line $g_2^<$ exponentially in i. In contrast, in the convex circular chain $\mathcal{A}^{\text{circ}}$, all A disks lie on a circle (including their mirror images after reflection about $x = L/2$) so that their opening angles are all the same.

For the convex geometric chain $\mathcal{A}^{\text{geo}}$, the distance between A_i and $g_2^<$ follows a geometric progression:

$$\text{dist}\left(A_{i+1}, g_2^<\right) = \phi \, \text{dist}\left(A_i, g_2^<\right), \quad \phi \in (0, 1), \tag{5.3}$$

Table 5.1: Parameters of Böröczky packings for different numbers k of layers with $N \sim 20k$ given by Eq. (5.4). Second and third columns: Density window for the Böröczky and Kahle cores with $\mathcal{A}^{\text{geo}}$, obtained from ϕ between 0.0001 and 0.9. Fourth and fifth columns: Deficit angle with respect to 180° of the maximum opening angle (in degrees, same for both cores) for $\mathcal{A}^{\text{circ}}$ and for $\mathcal{A}^{\text{geo}}$ with attenuation parameter $\phi = 0.8$.

Layers k	Density $\eta_{\text{Bör}}$	Density η_{Kahle}	Def. angle$^{\text{circ}}$	Def. angle$^{\text{geo}}$
5	$0.3957 \pm 3.1 \times 10^{-4}$	$0.4660 \pm 4.3 \times 10^{-4}$	8.3×10^{-1}	3.8×10^{-1}
6	$0.3625 \pm 2.9 \times 10^{-4}$	$0.4204 \pm 3.9 \times 10^{-4}$	5.3×10^{-1}	2.5×10^{-1}
7	$0.3338 \pm 2.6 \times 10^{-4}$	$0.3820 \pm 3.3 \times 10^{-4}$	3.8×10^{-1}	1.8×10^{-1}
8	$0.3089 \pm 2.2 \times 10^{-4}$	$0.3496 \pm 2.8 \times 10^{-4}$	2.8×10^{-1}	1.3×10^{-1}
9	$0.2873 \pm 1.9 \times 10^{-4}$	$0.3219 \pm 2.4 \times 10^{-4}$	2.2×10^{-1}	9.9×10^{-2}
10	$0.2683 \pm 1.7 \times 10^{-4}$	$0.2982 \pm 2.1 \times 10^{-4}$	1.7×10^{-1}	7.6×10^{-2}
15	$0.2010 \pm 9.5 \times 10^{-5}$	$0.2171 \pm 1.1 \times 10^{-4}$	7.3×10^{-2}	2.2×10^{-2}
20	$0.1604 \pm 6.0 \times 10^{-5}$	$0.1704 \pm 6.7 \times 10^{-5}$	4.1×10^{-2}	7.0×10^{-3}
30	$0.1141 \pm 3.0 \times 10^{-5}$	$0.1190 \pm 3.2 \times 10^{-5}$	1.8×10^{-2}	7.4×10^{-4}
50	$0.0722 \pm 1.2 \times 10^{-5}$	$0.0741 \pm 1.2 \times 10^{-5}$	6.3×10^{-3}	8.5×10^{-6}
100	$0.0376 \pm 3.1 \times 10^{-6}$	$0.0381 \pm 3.2 \times 10^{-6}$	1.6×10^{-3}	1.2×10^{-10}
1000	$0.0039 \pm 3.3 \times 10^{-8}$	$0.0039 \pm 3.3 \times 10^{-8}$	1.5×10^{-5}	7.4×10^{-98}

with the attenuation parameter ϕ. (For a horizontal branch, the distances in Eq. (5.3) are simply the difference between y-values.) The densities $\eta_{\text{Bör}}$ and η_{Kahle} of the Böröczky packings that either use the Böröczky or the Kahle core vary with ϕ, and they decrease as $\sim 1/k$ for large k (see Table 5.1). The geometric sequence for A_i induces that the maximum opening angle, usually the one between A_{k-1}, A_k, and A_{k+1}, approaches the angle π as $\theta_k = \phi^{k-2}(1 - \phi)(g_3 - g_2^<)/2 \sim \phi^k$, that is, exponentially in k and in L. This implies that the Böröczky packing with the convex geometric chain $\mathcal{A}^{\text{geo}}$ is for large number of layers k exponentially close to losing its local stability (see fifth column of Table 5.1).

The convex circular chain $\mathcal{A}^{\text{circ}}$ improves the local stability of the Böröczky packing, as the maximum opening angle on $\mathcal{A}$ approaches the critical angle π only algebraically with the number of layers k. Here, all A disks lie on a circle of radius R. This includes A_1, which by construction lies on g_3 (see Section 5.2.1.2). The circle is tangent to $g_2^<$ at $x = L/2$. The center of the circle lies on the vertical line at $x = L/2$. It follows from elementary trigonometry that for large k, the radius of the circle R scales as $\sim k^2$ and that the maximum opening angle approaches the angle π as $\sim k^{-2}$ (see fourth column of Table 5.1).

5.2.2.2 Contact Graphs: Local and Collective Stability

The contact graph of a Böröczky packing connects any two disks whose pair distance equals 2 (including periodic boundary conditions, see Fig. 5.1). In a Böröczky packing with $k \geq 1$ layers, the

number N of disks and the number N_{contact} of contacts are:

	N	N_{contact}
Böröczky core	$20k + 12$	$32k + 20$
Kahle core	$20k - 4$	$32k + 4$

$\qquad$ (5.4)

For all values of $k \geq 1$, the number of contacts is smaller than $2N - 2$. This implies that collective infinitesimal two-dimensional displacements, with $2N - 2$ degrees of freedom (the values of the displacements in x and in y for each disk avoiding trivial translations), can escape from a Böröczky packing, which is thus not collectively stable [195].

When all disks i, at positions $\mathbf{x}_i$, are moved to $\mathbf{x}_i + \boldsymbol{\Delta}_i$ with $\boldsymbol{\Delta}_i = (\Delta_i^x, \Delta_i^y)$, the squared separation between two touching disks from the contact graph i and j changes from $|\mathbf{x}_i - \mathbf{x}_j|^2$ to

$$|\mathbf{x}_i + \boldsymbol{\Delta}_i - (\mathbf{x}_j + \boldsymbol{\Delta}_j)|^2 = |\mathbf{x}_i - \mathbf{x}_j|^2 + \underbrace{2(\mathbf{x}_i - \mathbf{x}_j) \cdot (\boldsymbol{\Delta}_i - \boldsymbol{\Delta}_j)}_{\text{first-order variation}} + |\boldsymbol{\Delta}_i - \boldsymbol{\Delta}_j|^2. \qquad (5.5)$$

If the first-order term in Eq. (5.5) vanishes for all contacts i and j, the separation between touching disks cannot decrease. It then increases to second order in the displacements, if $\boldsymbol{\Delta}_i \neq \boldsymbol{\Delta}_j$, so that contact is lost. Distances between disks that are not in contact need not be considered because the displacements $\boldsymbol{\Delta}_i$ are infinitesimal. The first-order variation in Eq. (5.5) can be written as a product of twice an "escape matrix" $\mathcal{M}^{\text{esc}}$ of dimensions $N_{\text{contacts}} \times 2N$ with a $2N$-dimensional vector $\boldsymbol{\Delta} = (\Delta_1^x, \Delta_1^y, \Delta_2^x, \Delta_2^y, \ldots)$. The row r of $\mathcal{M}^{\text{esc}}$ corresponding to the contact between i and j has four non-zero entries

$$\begin{aligned}
\mathcal{M}^{\text{esc}}_{r,2i-1} &= x_i - x_j, \\
\mathcal{M}^{\text{esc}}_{r,2i} &= y_i - y_j, \\
\mathcal{M}^{\text{esc}}_{r,2j-1} &= -(x_i - x_j), \\
\mathcal{M}^{\text{esc}}_{r,2j} &= -(y_i - y_j).
\end{aligned} \qquad (5.6)$$

The `BigBoro` software package (see Appendix 5.B) solves for

$$\mathcal{M}^{\text{esc}} \boldsymbol{\Delta} = 0 \qquad (5.7)$$

using singular-value decomposition. The solutions of Eq. (5.7) are the directions of the small displacements that break the contacts but do not introduce overlapping disks. For the $k = 5$ Böröczky packing with the Kahle core, we find 28 vanishing singular values. It follows from Eq. (5.4) that, because of $28 = 2N - N_{\text{contact}}$, all contacts are linearly independent. We classify the 28 modes by studying the following cost function on the contact graph:

$$L = \sum_{i,j} (\boldsymbol{\Delta}_i - \boldsymbol{\Delta}_j)^2, \qquad (5.8)$$

where the sum is over all contact pairs i and j. This function, acting on the $2N$ displacements $\boldsymbol{\Delta}$,
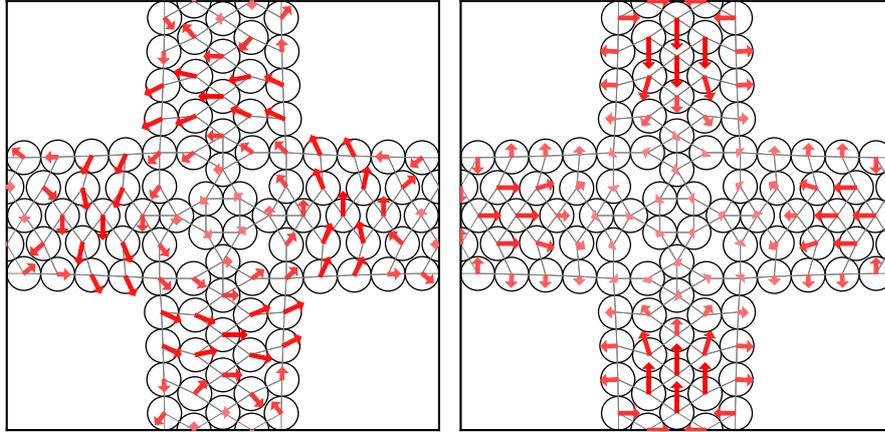
Figure 5.2: Two orthogonal modes (represented as red arrows) out of the 28-dimensional space of all collective escape modes Δ for the $k = 5$ Böröczky packing with the Kahle core and the convex geometric chain $\mathcal{A}^{geo}$ with attenuation parameter $\phi = 0.7$. Lines are drawn between pairs of disks which are in contact.

measures the non-uniformity of a deformation. It acts as a quadratic form within the 28-dimensional space of vanishing singular values, and can be diagonalized within this space. The resulting two lowest eigenmodes (with zero eigenvalue) of Eq. (5.8) describe rigid translation of the packing in the plane. Other low-lying eigenmodes give smooth large-scale deformations which collectively escape the contact constraints (see Fig. 5.2).

For $k \geq 1$, the number of contacts in Eq. (5.4) is larger than $N - 1$. Böröczky packings are thus collectively stable for displacements that are constrained to a single direction, as for example the x or y direction. This strongly constrains the dynamics of MCMC algorithms that for a certain time have only one degree of freedom per disk.

5.2.2.3 Dimension of the Space of Böröczky Packings

As discussed in Section 5.2.2.2, each Böröczky packing has a contact graph. Conversely, a given contact graph describes Böröczky packings for a continuous range of densities η. As an example, changing the attenuation parameter ϕ of the convex polygonal chain $\mathcal{A}^{geo}$ in Eq. (5.3) continuously moves all branch disks, and in particular disk B_k and, therefore, the value of L and the density η (see Table 5.1 for density windows that can be obtained in this way). We conjecture that locally stable packings exist for any density at large enough N. Sparse locally stable packings can also be part of dense hard-disk configurations where the majority of disks are free to move.

Moreover, the space $\mathcal{B}$ of locally stable packings of N disks of radius σ in a given central simulation box is of lower dimension than the sample space Ω: For each contact graph, each independent edge decreases the dimensionality by one. In addition there is only a finite number of contact graphs for a given N. The low dimension of $\mathcal{B}$ also checks with the fact that any packing, and more generally, any configuration with contacts, has effectively infinite pressure (see the detailed discussion in Section 5.4.2.2). As

the ensemble-averaged pressure is finite (except for the densest packing), the packings (and the configurations containing packings) must be of lower dimension. As the dimension of $\mathcal{B}$, for large N, is much lower than that of Ω, we conjecture $\Omega \setminus \mathcal{B}$ to be connected for a given η below the densest packing at large enough N although, in our understanding, this is proven only for $\eta \sim 1/\sqrt{N}$ (see Refs [197, 198]).

5.3 MCMC Algorithms and ε-Relaxed Böröczky Configurations

In this section, we first introduce to a number of local MCMC algorithms (see Section 5.3.1). In Section 5.3.2, we then determine the escape times (in the number of trials or events) after which these algorithms escape from ε-relaxed Böröczky configurations, that is, from Böröczky packings with disk radii multiplied by a factor $(1 - \varepsilon)$ (see Fig. 5.3a and b). A scaling theory establishes the existence of two classes of MCMC algorithms, one in which the escape time from an ε-relaxed Böröczky configuration scales algebraically with ε, with exponents that are predicted exactly, and the other in which the scaling is logarithmic. Numerical simulations confirm the theory.

5.3.1 Local Hard-Disk MCMC Algorithms

We define the reversible Metropolis algorithm with two displacement sets, from which the trial moves are uniformly sampled (see Section 5.3.1.1). We also consider variants of the non-reversible ECMC algorithm that only differ in their treatment of events, that is, of disk collisions (see Section 5.3.1.2). An arbitrary-precision implementation of the discussed ECMC algorithms (in the Go programming language) is contained in the `BigBoro` software package (see Appendix 5.B).

5.3.1.1 Local Metropolis Algorithm: Displacement Sets

The N disks are at positions $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$. In the local Metropolis algorithm [72], at each time $t = 1, 2, \ldots$, a trial move is proposed for a randomly chosen disk i, from its position $\mathbf{x}_i$ to $\mathbf{x}_i + \Delta\mathbf{x}_i$. If the trial produces an overlap, disk i stays put and $\mathbf{x}$ remains unchanged. We study two sets for the trial moves. For the cross-shaped displacement set, the trial moves are uniformly sampled within a range δ along the coordinate axes, that is, either along the x-axis $[\Delta\mathbf{x}_i = (\mathrm{ran}(-\delta, \delta), 0)]$ or along the y-axis $[\Delta\mathbf{x}_i = (0, \mathrm{ran}(-\delta, \delta))]$. Alternatively, for the square-shaped displacement set, the trial moves are uniformly sampled as $\Delta\mathbf{x}_i = (\mathrm{ran}(-\delta, \delta), \mathrm{ran}(-\delta, \delta))$. A Böröczky packing traps the local Metropolis algorithm if the range δ is smaller than a critical range δ_c. This range is closely related to the maximum opening angle (see the discussion in Section 5.2.2.1 and Fig. 5.3c). For these packings, the critical range vanishes for $N \to \infty$ independently of the specific core or of the convex polygonal chain, simply because the maximum opening angle approaches π in that limit. On the other hand, for large range δ, the algorithm can readily escape from the stable configuration. For $\delta = L/2$, the Metropolis algorithm with a square-shaped displacement set proposes a random placement of the disk i inside the central simulation box. This displacement set leads to a very inefficient algorithm at the densities of physical interest, but it mixes very quickly at small finite densities (see Section 5.4.2.1). For the scaling theory of the escape of the Metropolis algorithm from ε-relaxed Böröczky configurations, we consider ranges δ smaller than the critical range δ_c.
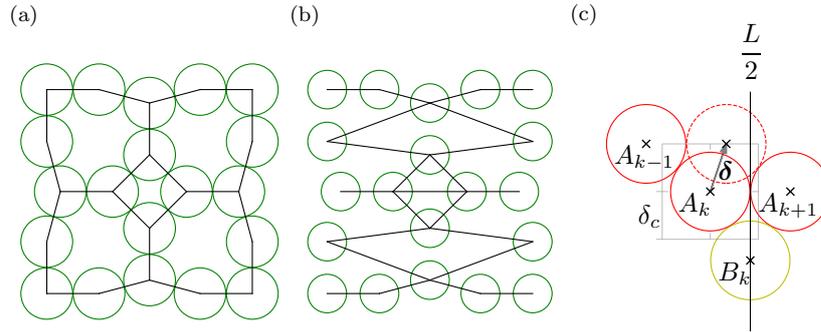
Figure 5.3: Contact graphs, constraint graphs and minimal escape range. (a): Contact graph for a packing consisting solely of the Böröczky core. (b): Constraint graph in x-direction for an ε-relaxed Böröczky configuration derived from the same packing with $\varepsilon = 0.25$. The edges indicate all possible collisions of straight ECMC in x-direction. (c): Escape move $\boldsymbol{\delta}$ and minimal escape range δ_c of the Metropolis algorithm with a square-shaped displacement set.

5.3.1.2 Hard-Disk ECMC: Straight, Reflective, Forward, Newtonian

Straight ECMC [115] is one of the two original variants of event-chain Monte Carlo. This Markov chain evolves in (real-valued) continuous Monte-Carlo time t_{MCMC}, but its implementation is event-driven. The algorithm is organized in a sequence of "chains", each with a chain time τ_{chain}, its intrinsic parameter. In each chain, with Monte-Carlo time between t_{MCMC} and $t_{\mathrm{MCMC}} + \tau_{\mathrm{chain}}$, disks move with unit velocity in one given direction (alternatively in $+x$ or in $+y$). A randomly sampled initial disk thus moves either until the chain time τ_{chain} is used up, or until, at a collision event, it collides with another disk, which then moves in its turn, etc. This algorithm is highly efficient in some applications [115, 142, 143]. During each chain (in between changes of direction), any disk can collide only with three other disks or fewer [199, 200]. A constraint graph with directed edges may encode these relations. This constraint graph (defined for hard-disk configurations) takes on the role of the contact graph (that is defined for packings) (see Fig. 5.3a and b). As the moves in a chain are all in the same direction, straight ECMC has only $N - 1$ degrees of freedom, fewer than there are edges in the constraint graph. It is for this reason that it may encounter the rigidity problems evoked in Section 5.2.2.2.

In reflective ECMC [115], in between events, disks move in straight lines with unit velocity just as in straight ECMC. At a collision event, the target disk does not continue in the same direction as the active disk. Rather, the target-disk direction is the original active-disk direction reflected from the line connecting the two disk centers at contact (see Ref. [115]). As all ECMC variants, reflective ECMC satisfies the global-balance condition. Irreducibility (for connected sample spaces) requires in principle resamplings of the active disk and its velocity in intervals of the chain time τ_{chain} [P2, 140, 150]. However, this seems not always necessary [P2, 150]. Numerical experiments indicate that reflective ECMC requires no resamplings in our case as well. It is also faster without them (see Appendix 5.A.2). A variant of reflective ECMC, obtuse ECMC [124], has shown interesting behavior.

Forward ECMC [150] updates the normalized target-disk direction as follows after an event. The component orthogonal to the line connecting the disks at contact is uniformly sampled between 0 and 1 (reflecting the orthogonal orientation). Its parallel component is determined so that the

direction vector (which is also the velocity vector) is of unit norm. The parallel orientation remains unchanged. In contrast to reflective ECMC, the event-based randomness renders forward ECMC practically irreducible for the considered two-dimensional hard-disk systems even without resamplings. Resamplings in intervals of the chain time τ_{chain} can still be considered but slow the algorithm down (see Appendix 5.A.2). We thus consider forward ECMC without resampling.

Newtonian ECMC [124] mimics molecular dynamics in order to determine the velocity of the target disk in an event. It initially samples disk velocities from the two-dimensional Maxwell distribution with unit root-mean-square velocity. However, at each moment, only a single disk is actually moving with its constant velocity. At a collision event, the velocities of the colliding disks are updated according to Newton's law of elastic collisions for hard disks of equal masses, but only the target disks actually moves after the event. In this algorithm, the velocity (which indexes the Monte-Carlo time) generally differs from unity. Similar to reflective ECMC, we tested that resamplings appear not to be required in our case (and again yield a slower performance, see Appendix 5.A.2), although Newtonian ECMC manifestly violates irreducibility in highly symmetric models [P2]. As in earlier studies for three-dimensional hard-sphere systems [124] and for two-dimensional dipoles [P2], Newtonian ECMC is typically very fast for ε-relaxed Böröczky configurations. However, it suffers from frequent gridlocks (see Section 5.3.2.4).

5.3.2 Escape Times from ε-Relaxed Böröczky Configurations

The principal figure of merit for a Markov chain is its mixing time [74], the number of steps it takes from the worst-case initial condition to approach the stationary probability distribution to some precision level. Böröczky packings trap the local Metropolis dynamics (of sufficiently small range) as well as ECMC dynamics, so that the mixing time is, strictly speaking, infinite. Although they cannot be escaped from, the packings make up only a set of measure zero in sample space, and might thus be judged irrelevant.

However, as we will discuss in the present subsection, the situation is more complex. For every Böröczky packing, an associated ε-relaxed Böröczky configuration keeps the central simulation box and the disk positions, but reduces the disk radii from 1 to $1 - \varepsilon$. An ε-relaxed Böröczky configuration effectively defines a finite portion of the sample space (the spheres of radius ε around each disk position of the packing). All MCMC algorithms considered in this work escape from these configurations in an escape time that diverges as $\varepsilon \to 0$ (see Section 5.3.2.1 for a definition of escape times). Numerical results and a scaling theory for the escape times are discussed in Sections 5.3.2.2 and 5.3.2.3, and a synopsis of our results is contained in Section 5.3.2.4. The divergent escape times as $\varepsilon \to 0$ are specific to the *NVT* ensemble (as we will discuss in Section 5.4.2.2).

5.3.2.1 Nearest-Neighbor Distances and Escape Times

In a Böröczky packing, disks are locally stable, and they all have a nearest-neighbor distance of 2. The packings are sparse, and the nearest-neighbor distance is thus smaller than its $\sim 1/\sqrt{\eta}$ equilibrium value. To track the escape from an ε-relaxed Böröczky configuration, we monitor the maximum

nearest-neighbor distance:

$$d(t) = \max_i \left[\min_{j(\neq i)} |\mathbf{x}_{ij}(t)| \right], \tag{5.9}$$

where $|\mathbf{x}_{ij}(t)| = |\mathbf{x}_j(t) - \mathbf{x}_i(t)|$ is the distance between disks i and j (possibly corrected for periodic boundary conditions). The maximum nearest-neighbor distance signals when a single disk breaks loose from what corresponds to its contacts. In the further time evolution, the configuration then falls apart. For the Metropolis algorithm, we compute $d(t)$ once every N trials, and t denotes the integer-valued number of individual trial moves. For ECMC, we sample $d(t)$ and the number of events in intervals of the sampling Monte-Carlo time. In Eq. (5.9), t then denotes the integer-valued number of events. Both discrete times t increment by one with a computational effort $O(1)$, corresponding to one trial in the Metropolis algorithm and to one event in ECMC. Starting from an ε-relaxed Böröczky configuration, $d(t)$ typically remains at $d(t) \sim 2 + O(\varepsilon)$ for a long time until it approaches the equilibrium value in a way that depends on the algorithm. We define the escape time t_{esc}, an integer, as the time t at which $d(t)$ has increased by ten percent:

$$t_{\text{esc}} = \min \{t : d(t) > 2(1 + \gamma)\}, \tag{5.10}$$

with $\gamma = 0.1$. All our results for the scaling of the escape time with the relaxation parameter ε in the following subsections were reproduced for $\gamma = 0.025$ (see Appendix 5.A.1). The definition of the escape time based on the maximum nearest-neighbor distance $d(t)$ is certainly not the only one to monitor the stability of ε-relaxed Böröczky configurations. It may not be equally well-suited for all considered algorithms. Still, our scaling theory suggests that the algorithms with an intrinsic parameter show a distinctly different behavior than the algorithms without them, which appears to be independent of the precise definition of the escape time.

5.3.2.2 Escape-Time Scaling for Metropolis and Straight ECMC

The local Metropolis algorithm and straight ECMC both have an intrinsic parameter, namely the range δ of the displacement set or the chain time τ_{chain}. These two parameters play a similar role. We numerically measure the escape time t_{esc} of these algorithms for a wide range of their intrinsic parameters and for small relaxation parameters ε (see Fig. 5.4, for the escape times from ε-relaxed Böröczky configurations with $k = 5$ layers and the Kahle core). The escape time diverges for $\delta, \tau_{\text{chain}} \to 0$. For straight ECMC and small ε, t_{esc} also diverges for $\tau_{\text{chain}} \to \infty$ so that the function is "V"-shaped with an optimal chain time $\tau_{\text{chain}}^{\min}$. For the Metropolis algorithm, t_{esc} increases until around the critical range δ_c so that there is an optimal range $\delta^{\min} < \delta_c$.

Two limiting cases can be analyzed in terms of the intrinsic parameter $\delta < \delta_c$ or τ_{chain}, and the internal length scales ε, and σ. For the Metropolis algorithm at small δ, a trajectory spanning a constant distance is required to escape from an ε-relaxed Böröczky configuration. This constant distance can be thought of as the escape distance δ_c in Fig. 5.3, which is on a scale σ and independent of ε for small ε. As the Monte-Carlo dynamics is diffusive, this constant distance satisfies const $= \delta \sqrt{t_{\text{esc}}}$. For straight ECMC with small chain times τ_{chain}, the effective dynamics (after subtraction of the uniform
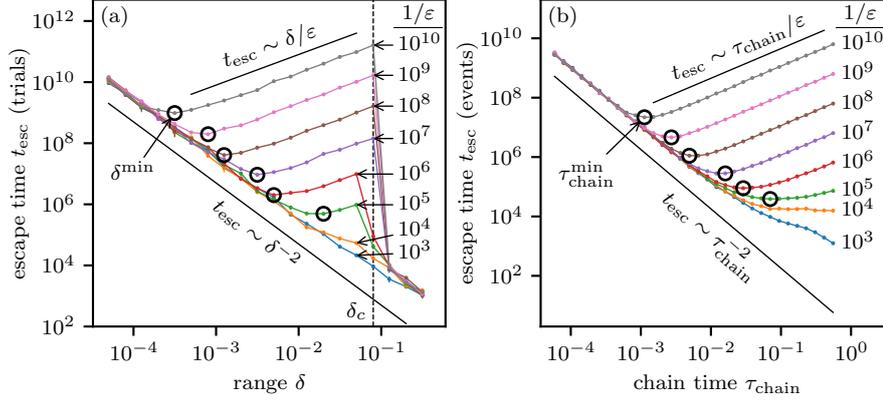
Figure 5.4: Median escape times from the $k = 5$ ε-relaxed Böröczky configurations (Kahle core and convex geometric chain $\mathcal{A}^{\text{geo}}$ with attenuation parameter $\phi = 0.7$, $N = 96$ disks) for different ε. (a): t_{esc} (in trials) vs. range δ for the Metropolis algorithm with the cross-shaped displacement set. (b): t_{esc} (in events) vs. chain time τ_{chain} for straight ECMC. Asymptotes are from Eqs (5.11) and (5.12). Error bars are smaller than the marker sizes.

displacement), is again diffusive. This leads to:

$$t_{\text{esc}} \sim \begin{cases} \text{const}/\delta^2 & \text{(Metropolis),} \\ \text{const}/\tau_{\text{chain}}^2 & \text{(straight ECMC),} \end{cases} \quad \text{(for small } \delta < \delta_c, \tau_{\text{chain}}\text{).} \tag{5.11}$$

The independence of t_{esc} of the relaxation parameter ε for small intrinsic parameters is clearly brought out in the numerical simulations (see Fig. 5.4).

On the other hand, even for large $\delta < \delta_c$ or τ_{chain}, the Markov chain must make a certain number of moves on a length scale ε in order to escape from the ε-relaxed Böröczky configuration. In the Metropolis algorithm, the probability for a trial on this scale is ε/δ for the cross-shaped displacement set and ε^2/δ^2 for the square-shaped displacement set. For the straight ECMC with large τ_{chain}, all displacements beyond a time $\sim \varepsilon$ (or, possibly, $\sim N\varepsilon$) effectively cancel each other, because the constraint graph is rigid. This leads to:

$$t_{\text{esc}} \sim \begin{cases} \delta^2/\varepsilon^2 & \text{(Metropolis—square),} \\ \delta/\varepsilon & \text{(Metropolis—cross),} \\ \tau_{\text{chain}}/\varepsilon & \text{(straight ECMC),} \end{cases} \quad \text{(for large } \delta < \delta_c, \tau_{\text{chain}}\text{).} \tag{5.12}$$

The scaling of t_{esc} as $\sim 1/\varepsilon$ or $\sim 1/\varepsilon^2$ for large intrinsic parameters is confirmed in the numerical simulations for small relaxation parameters ε (see Fig. 5.4). For large ε, the critical range δ_c of the Metropolis algorithm (that slightly decreases with ε) falls below the region of large δ. For large ε, the constraint graph of straight ECMC loses its rigidity, and τ_{chain} no longer appears as a relevant intrinsic parameter. The scaling theory no longer applies.

The two asymptotes of Eqs (5.11) and (5.12) form a "V" with a base $\delta^{\min}$ (or $\tau_{\text{chain}}^{\min}$) that is obtained by equating the two expressions for $t_{\text{esc}}(\delta)$ (or $t_{\text{esc}}(\tau_{\text{chain}})$). This yields $\delta^{\min} \sim \sqrt[3]{\varepsilon}$ for the Metropolis

algorithm with a cross-shaped displacement set, and likewise $\tau_{\text{chain}}^{\min} \sim \sqrt[3]{\varepsilon}$ for straight ECMC. For the Metropolis algorithm with a square-shaped move set, one obtains $\delta^{\min} \sim \sqrt{\varepsilon}$. The resulting optimum, the minimal escape time with respect to ε, is

$$t_{\text{esc}} \sim \begin{cases} \varepsilon^{-1} & \text{(Metropolis—square),} \\ \varepsilon^{-2/3} & \text{(Metropolis—cross),} \\ \varepsilon^{-2/3} & \text{(straight ECMC),} \end{cases} \quad \text{(for optimal } \delta^{\min}, \tau_{\text{chain}}^{\min}). \tag{5.13}$$

These scalings balance two requirements: to move by a constant distance (which favors large δ or τ_{chain}) and to move on the scale ε (which favors small δ or τ_{chain}).

5.3.2.3 Time Dependence of Free Path—Reflective, Forward, and Newtonian ECMC

The forward, reflective, and Newtonian variants of ECMC move in any direction, even in the absence of resamplings, so that their displacement sets are $2N$-dimensional. This avoids the rigidity problem of straight ECMC (the fact that the number of constraints can be larger than the number of degrees of freedom). We consider these algorithms without resamplings, that is, for $\tau_{\text{chain}} = \infty$. Finite chain times yield larger escape times that approach the value at $\tau_{\text{chain}} = \infty$ (see Appendix 5.A.2). Without an intrinsic parameter, the effective free path between events may thus adapt as the configuration gradually escapes from the ε-relaxed Böröczky configuration. The free path is initially on the scale ε, but then grows on average by a constant factor at each event, reaching a scale $\varepsilon' > \varepsilon$ after a time (that is, after a number of events) that scales as $\sim \ln(\varepsilon'/\varepsilon)$. The scale ε' at which the algorithms break free is independent of the initial scale ε, and we expect a logarithmic scaling of the escape time (measured in events):

$$t_{\text{esc}} \sim \ln(1/\varepsilon) \quad \text{(reflective, forward, and Newtonian ECMC).} \tag{5.14}$$

The absence of an imposed scale for displacements manifests itself in the logarithmic growth with time of the average free path, that is, the averaged displacement between events over many simulations starting from the same ε-relaxed Böröczky configuration (see Fig. 5.5 for the example of the escape of forward ECMC from ε-relaxed Böröczky configurations with $k = 5$ layers and the Kahle core). Individual evolutions as a function of time t for small relaxation parameters ε and ε' overlap when shifted by their escape times. Starting from an ε-relaxed Böröczky configuration with $\varepsilon = 10^{-30}$, as an example, the same time is on average required to move from an average free path of $\sim 10^{-30}$ to 10^{-25}, as from an average free path $\sim 10^{-25}$ to 10^{-20}. The time t in this discussion refers to the number of events and not to the Monte-Carlo time t_{MCMC}. As discussed, the velocity in reflective and forward ECMC, and the root-mean-square velocity in Newtonian ECMC, have unit value. The free path between subsequent events—which, as discussed, grows exponentially with t—then equals the difference of Monte-Carlo times $t_{\text{MCMC}}(t + 1) - t_{\text{MCMC}}(t)$. The Monte-Carlo time t_{MCMC} thus grows as a geometric series and depends exponentially on the number of events t. This emphasizes that the escape from an ε-relaxed Böröczky configuration is a non-equilibrium phenomenon.
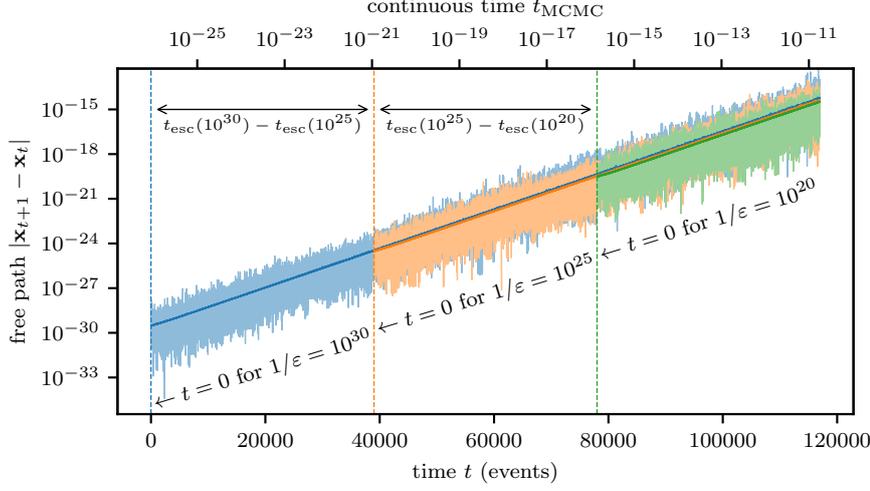
Figure 5.5: Free path (equivalently: Monte-Carlo time between events) for the forward ECMC algorithm started from three $k = 5$ ε-relaxed Böröczky configurations (Kahle core and convex geometric chain $\mathcal{A}^{\mathrm{geo}}$ with attenuation parameter $\phi = 0.7$, $N = 96$ disks) with $\varepsilon = 10^{-30}$, 10^{-25} and 10^{-20}. Integer time t (lower x-axis) counts events, while t_{MCMC} (upper x-axis) is the real-valued continuous Monte-Carlo time. Event times are shifted. Expanded light curves show single simulations for each ε, dark lines average over $10\,000$ simulations.

5.3.2.4 Escape Times: Synopsis of Numerical Results and Scaling Theory

Overall, escape times $t_{\mathrm{esc}}(\varepsilon)$ (with intrinsic parameters optimized through a systematic scan for the Metropolis algorithm and for straight ECMC) validate the algebraic scalings of Eq. (5.13), on the one hand, and the logarithmic scaling of Eq. (5.14), on the other (see Fig. 5.6 for the escape times from ε-relaxed Böröczky configurations with $k = 5$ layers with either the Kahle core or the Böröczky core). Our arbitrary-precision implementation of reflective, forward, and Newtonian ECMC confirms their logarithmic scaling down to $\varepsilon = 10^{-29}$. Newtonian ECMC appears *a priori* as the fastest variant of ECMC. However, it frequently gets gridlocked, i.e., trapped in circles of repeatedly active disks with a diverging event rate. Gridlocks also rarely appear in straight and reflective ECMC. In runs that end up in gridlock, escape times are very large, possibly diverging. (In Figs 5.4 and 5.6, median escape times rather than the means are therefore displayed for all algorithms. Mean and median escape times are similar for the Metropolis algorithm and forward ECMC where gridlocks play no role.) The gridlock rate increases with $1/\varepsilon$. For the Kahle core, this effect is negligible for all ε. For the Böröczky core, the gridlock rate of Newtonian ECMC is $\sim 30\%$ for $\varepsilon = 10^{-29}$ (see Fig. 5.6b, the logarithmic scaling is distorted even for the median). We observe no clear dependence of the gridlock rate on the floating-point precision of our arbitrary-precision ECMC implementation, and it thus appears unlikely that gridlocks are merely numerical artifacts (see Appendix 5.A.3).

Gridlock is the very essence of ECMC dynamics from a locally stable Böröczky packing, but it can also appear as a final state from an ε-relaxed Böröczky configuration. We observe gridlocks in all hard-disk ECMC variants that feature deterministic collision rules. They were previously observed for straight ECMC from tightly packed initial configurations [178, Section 4.2.3]. Only forward ECMC with its event-based randomness is free of them. In a gridlock, the event rate diverges at a given
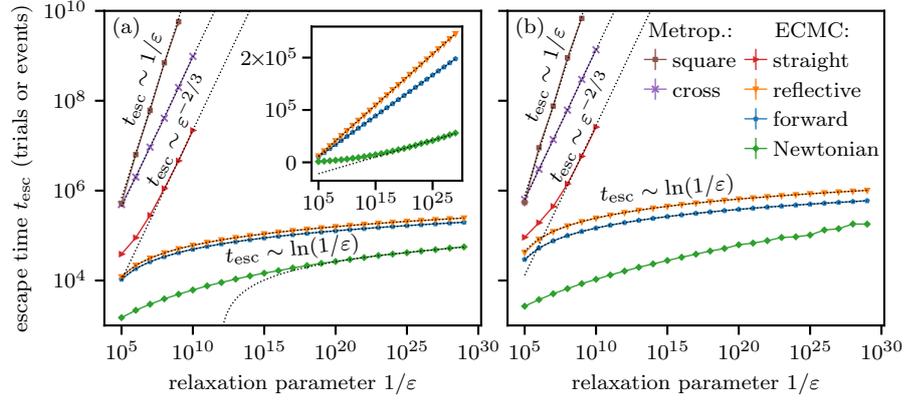
Figure 5.6: Median escape time t_{esc} from $k = 5$ ε-relaxed Böröczky configurations with different cores (with convex geometric chain $\mathcal{A}^{geo}$ and attenuation parameter $\phi = 0.7$) for local MCMC algorithms (where applicable: with optimized intrinsic parameters). (a): t_{esc} for the Kahle core ($N = 96$ disks). The Metropolis algorithm and straight ECMC show an algebraic scaling. Inset: log–lin plots suggesting logarithmic scaling for the forward, reflective, and Newtonian ECMC. (b): t_{esc} for the Böröczky core ($N = 112$ disks). Newtonian ECMC has frequent gridlocks for small ε so that its logarithmic scaling is distorted. Error bars are smaller than the marker sizes.

Monte-Carlo time, which then seems to stand still so that no finite amount of Monte-Carlo time is spent in a configuration with contacts. Because of the divergence of the event rate, gridlocks cannot be cured through resamplings at fixed Monte-Carlo-time intervals. To overcome them in Newtonian ECMC, which appears *a priori* as the fastest of our ECMC variants, one can probably introduce event-based randomness as is done in forward ECMC. Nevertheless, gridlocks play no role in large systems at reasonable densities. Also, ECMC algorithms for soft potentials introduce randomness at each event so that gridlocks should not appear.

5.4 Discussion

In the present section, we discuss our results for the escape times (Section 5.4.1), as well as a number of more fundamental aspects of Böröczky packings in the context of MCMC (Section 5.4.2). We in particular clarify why a packing effectively realizes an infinite-pressure configuration that in a constant-pressure Monte-Carlo simulation is instantly relaxed through a volume increase.

5.4.1 Escape Times: Speedups, Bottlenecks

ECMC is a continuous-time MCMC method, and its continuous Monte-Carlo time t_{MCMC} takes the place of the usual count of discrete-time Monte-Carlo trials. However, ECMC is event-driven. The time t, and especially the escape time t_{esc}, are integers, and they count events. The computational effort in hard-disk ECMC is $O(1)$ per event, using a cell-occupancy system that is also implemented in the `BigBoro` software package. In several of our algorithms, the times t and t_{MCMC} are not proportional to each other, because the free path (roughly equivalent to the Monte-Carlo time between events) evolves during each individual run.

5.4.1.1 Range of Speedups

The speedup realized by lifted Markov chains, of which ECMC is a representative, corresponds to the transition from diffusive to ballistic transport [80, 111, 112]. This speedup refers to what we call the "Monte-Carlo time" t_{MCMC}, that is the underlying time of the Markov process, and not to the time t that is measured in events. For Markov chains in a finite sample space Ω, the Monte-Carlo time for mixing of the lifted Markov chain cannot be smaller than the square root of the mixing time for the original (collapsed) chain. The remarkable power-law-to-logarithm speedup in ε realized by some of the ECMC algorithms concerns escape times which measure the number of events. The Monte-Carlo escape times probably conform to the mathematical bounds, although it is unclear how to approximate hard-disk MCMC for $\varepsilon \to 0$ through a finite Markov chain. Mathematical results for the Monte-Carlo escape times from locally blocked configurations would be extremely interesting, even for models with a restricted number of disks.

5.4.1.2 Space of ε-Relaxed Böröczky Configurations

The definition of an ε-relaxed Böröczky configuration can be generalized. Equivalent legal hard-disk configurations are obtained by reducing the disk radii and choosing random disk positions in a circle of radius ε around the original disk positions in the Böröczky packing. These configurations also feature the escape-time scalings given in Eqs (5.13) and (5.14). Any ε-relaxed Böröczky configuration is thus merely a sample in a space $\mathcal{B}_\varepsilon$ of volume $\sim \varepsilon^{2N}$. For a given upper limit t_{cpu} of CPU time at fixed N, this corresponds to a volume of $\mathcal{B}_\varepsilon$ (that cannot be escaped from in t_{cpu}) scaling with the computer-time budget as $\sim t_{cpu}^{-3N}$ for the straight ECMC and scaling as $\sim \exp(-2Nt_{cpu})$ for the forward ECMC. We expect $\mathcal{B}_\varepsilon$ to have a double role, as a space of configurations that the Monte-Carlo dynamics cannot practically escape from, but maybe also a space that it cannot even access. The volume of $\mathcal{B}_\varepsilon$ (with ε chosen such that it cannot be escaped from in a reasonable CPU time) as well as the corresponding changes in the free energy per disk are probably unmeasurably small except, possibly, at very small N. The existence of a finite fraction of sample space that cannot be escaped from in any reasonable CPU time at finite N is however remarkable. In many MCMC algorithms for physical systems, as for example the Ising model, parts of sample space are practically excluded because of their low Boltzmann weight, but they feature diverging escape times only in the limit $N \to \infty$.

In this context, we note that Markov chains can be interpreted in terms of a single bottleneck partitioning the sample space into two pieces [74, Sect. 7.2]. The algorithmic stationary probability flow across the bottleneck sets the conductance of an algorithm, which again bounds mixing and correlation times. Ideally, MCMC algorithms would be benchmarked through their conductances. In the hard-disk model, the bottleneck has not been identified, so that the benchmarking and the analysis of MCMC algorithms must rely on empirical criteria. However, Böröczky packings and the related ε-relaxed Böröczky configurations may well model a bottleneck, from which the Markov chain has to escape in order to cross from one piece of the sample space into its complement. The benchmarks obtained by comparing escape times from an ε-relaxed Böröczky configuration may thus reflect the relative merits of sampling algorithms.

5.4.2 Böröczky Packings and Local MCMC: Fundamental Aspects

We now discuss fundamental aspects of the present work, namely the question of the irreducibility of local hard-sphere Markov chains and the connection with non-local MCMC algorithms (see Section 5.4.2.1), as well as regularization of Böröczky packings and ε-relaxed Böröczky configurations in the NPT ensemble (see Section 5.4.2.2).

5.4.2.1 Irreducibility of Local and Non-Local Hard-Disk MCMC

Strictly speaking, ECMC can be irreducible only if $\Omega \setminus \mathcal{B}$ is connected, where $\mathcal{B}$ is a suitably defined space of locally stable configurations. Packings in $\mathcal{B}$ (a space of low dimension) are certainly invariant under any version of the ECMC algorithm, so that they cannot evolve towards other samples in Ω. Connectivity in $\Omega \setminus \mathcal{B}$ would at least assure that this space can be sampled. In addition it appears necessary to guarantee that a well-behaved initial configuration cannot evolve towards an ε-environment around $\mathcal{B}$ (e.g., the space $\mathcal{B}_\varepsilon$ of ε-relaxed Böröczky configurations that makes up a finite portion of Ω) or to gridlocks with diverging event rates. These properties appear not clearly established for finite densities η and for large N. In other models, for example the Ising model of statistical physics, irreducibility can be proven for any N.

These unresolved mathematical questions concerning irreducibility do not shed doubt on the practical usefulness of MCMC for particle systems. First, the concept of local stability is restricted to hard disks and hard spheres (that is, to potentials that are either zero or infinite). The phase diagram of soft-disk models can be continuously connected to the hard-disk case [122]. For soft disks, irreducibility is trivial, but the sampling speed of algorithms remains crucial. Second, in applications, one may change the thermodynamic ensemble. In the NPT ensemble, the central simulation box fluctuates in size and can become arbitrarily large. In this ensemble, irreducibility follows from the fact that large enough simulation boxes are free of steric constraints. Again, the question of mixing and correlation time scales is primordial. Third, practical simulations that require some degree of irreducibility are always performed under conditions where the simulation box houses a number of effectively independent copies of the system. This excludes the crystalline or solid phases. Monte Carlo simulations of such phases are more empirical in nature. They require a careful choice of initial states, and are then not expected to visit the entire sample space during their time evolution. Fundamental quantitative results can nevertheless be obtained [201].

In this work, we concentrate on local MCMC algorithms, because global-move algorithms, as the cluster algorithms in spin systems, rely on *a priori* probabilities for many-particle moves that appear too complicated. Also, global single-particle moves are related to the single-particle insertion probabilities, in other words to fugacities (the exponential of the negative chemical potential) that are prohibitively small. At lower (finite) densities, however, placing at each time step a randomly chosen disk at a random position inside the box corresponds to the Metropolis algorithm of Section 5.3.1.1 with a square-shaped displacement set and a range $\delta = L/2$. This non-local algorithm easily escapes from a Böröczky packing. Moreover, it is proven to mix in $O(N \log N)$ steps at densities $\eta < 1/6$ [190, 202] (see also Ref. [203]), a result that implies that the liquid phase in the hard-disk system extends at least to the density $\eta = 1/6$ [190]. The density bound for the algorithm (which yields a bound for the

stability of the liquid phase) is much smaller than the empirical density bound for the liquid phase, at $\eta \simeq 0.70$. At this higher density, the global-move Metropolis algorithm and the more general hard-disk cluster algorithm [183] are almost totally stuck. For applications, we imagine structures resembling ε-relaxed Böröczky configurations to be backbones of configurations at high density, where global moves cannot be used.

5.4.2.2 Böröczky Packings and the NPT Ensemble

The concepts of packings and of local and collective stability make sense only in the NVT ensemble, that is, for a constant number of particles and for a simulation box with fixed shape and volume (the temperature $T = 1/\beta$ that appears in NVT plays no role in hard-disk systems [78]). In the NPT ensemble, the pressure P is constant, and the size of the simulation box may vary. The equivalence of the two ensembles is proven [204] for large N, so that the choice of ensemble is more a question of convenience than of necessity. As we will see, in the NPT ensemble, tiny relaxation parameters (as $\varepsilon = 10^{-29}$ in Fig. 5.6) are instantly relaxed to $\varepsilon \sim 10^{-3}$ for normal pressures and system sizes.

To change the volume at constant pressure, one may, among others, proceed to "rift volume changes" (see Ref. [121, Sect. VI]) or else to homothetic transformations of the central simulation box. We discuss this second approach (see Ref. [78, Sect. 2.3.4]), where the disk positions (but not the radii) are rescaled by the box size L as:

$$\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \rightarrow \boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N) \quad \text{with } \alpha_i = \mathbf{x}_i/L. \tag{5.15}$$

Each configuration is then specified by an $\boldsymbol{\alpha}$ vector in the $2N$-dimensional periodic unit square and an associated volume $V = L^2$, which must satisfy $V \geq V_{\text{cut}}(\boldsymbol{\alpha})$. A classic MCMC algorithm [205] directly samples the volume at fixed $\boldsymbol{\alpha}$ from a gamma distribution above $V_{\text{cut}}(\boldsymbol{\alpha})$, below which $(\boldsymbol{\alpha}, V)$ ceases to represent a valid hard-disk configuration [78, Eq. (2.19)]. Typical sample volumes are characterized by $\beta P(V - V_{\text{cut}}) \sim 1$, and with $V = (L_{\text{cut}} + \Delta L)^2$, it follows that

$$\frac{\Delta L}{L} \sim \varepsilon \sim \frac{1}{\beta P V_{\text{cut}}} \quad \text{(at fixed } \boldsymbol{\alpha}). \tag{5.16}$$

This equation illustrates that a packing, with $\varepsilon \rightarrow 0$, is realized as a typical configuration only in the limit $\beta P \rightarrow \infty$. For the Böröczky packings of Fig. 5.1, we have $L \simeq 20$, and a typical value for the pressure for hard-disk systems is $\beta P \sim 1$, which results in $\varepsilon \sim 10^{-3}$. In the NPT ensemble, as a consequence, escape times from a packing naturally correspond to a relaxation parameter $\varepsilon \sim 1/(\beta PV)$, in our example to $t_{\text{esc}}(\varepsilon \sim 10^{-3})$, which is $O(1)$.

The above NPT algorithm combines constant-volume NVT-type moves of $\boldsymbol{\alpha}$ with the mentioned direct-sampling moves of V at fixed $\boldsymbol{\alpha}$. In practice, however, NPT calculations are rarely performed in hard-disk systems [206, 207]. This is because, as discussed in Eq. (5.16), the expected single-move displacement in volume at fixed $\boldsymbol{\alpha}$ is $\Delta V \sim 1/(\beta P)$, so that $\Delta V/V \sim 1/N$ (because $N \sim V$ and $\beta P \sim 1$). The fluctuations of the equilibrium volume V^{eq} (averaged over $\boldsymbol{\alpha}$) scale as $\sqrt{V^{\text{eq}}}$, which implies $\Delta V^{\text{eq}}/V^{\text{eq}} \sim 1/\sqrt{N}$. The volume-sampling algorithm requires $\sim N$ single updates of the volume to go from the $1/N$ scale of volume fluctuations at fixed $\boldsymbol{\alpha}$ to the $1/\sqrt{N}$ scale of the fluctuations of V^{eq} at

equilibrium. This multiplies with the number of steps to decorrelate at a given volume. In practice, it has proven more successful to perform single *NVT* simulations, but to restrict them to physical parameters where the central simulation box houses a finite number of effectively independent systems mimicking constant-pressure configurations.

5.5 Conclusion

Building on an early breakthrough by Böröczky, we have studied in this work locally stable hard-disk packings. Böröczky packings are sparse, with arbitrarily small densities for large numbers N of disks. We constructed different types of these packings to arbitrary precision for finite N, namely Böröczky packings with the original Böröczky core [185] and those with the Kahle core [195]. In addition to the core and the number k of layers, Böröczky packings are defined by the convex polygonal chain which bounds their branches. We constructed Böröczky packings in a continuous range of densities, and made our software implementation of the construction openly accessible. Böröczky packings are locally, but not collectively stable. Using singular-value decomposition (in an implementation that is included in our open-source software) we explicitly exposed the unstable collective modes. We furthermore reduced the radius of Böröczky packings slightly, and determined the escape times from ε-relaxed Böröczky configurations as a function of the parameter ε for a number of local MCMC algorithms, including several variants of ECMC, arbitrary-precision implementations of which are also made openly available. Although the algorithms depart from each other in seemingly insignificant details only, we witnessed widely different escape times, ranging from $1/\varepsilon$ to $\log(1/\varepsilon)$. Our theory suggested that the significant speedup of some of the algorithms is rooted in their event-driven nature coupled to their lack of an intrinsic scale. We noted that the space of ε-relaxed Böröczky configurations is a finite portion of the sample space, and that a given computer-time budget implies such a finite fraction of sample space that is practically excluded in local MCMC at finite N. Here, the excluded volume only vanishes in the limit of infinite CPU time. More generally, connectedness of the hard-disk sample space is not proven. We pointed to the importance of statistical ensembles to reconcile the possible loss of irreducibility with the proven practical usefulness of local hard-disk MCMC algorithms. Although Böröczky packings or ε-relaxed Böröczky configurations are sparse, they could form the locally stable (or almost locally stable) backbones of hard-disk configurations at the much higher density which are of practical interest.

We expect the observed differences in escape times to carry over to real-world ECMC implementations. Qualitatively similar performance differences were already observed in autocorrelation times of hard-disk dipoles [P2]. In statistical mechanics, bottlenecks and escape times possibly play an important role in polymer physics and complex molecular systems and some of the algorithms studied here may find useful applications. Escape times may also play an important role in the study of glasses and in granular matter, where the high or even infinite pressures favor local configurations that resemble the mutually blocked disks in the ε-relaxed Böröczky configurations. We finally point out that the very concept of locally stable packings naturally extends to higher dimensions.

Data Availability Statement

This work is based on computer programs that are all publicly available (see Appendix 5.B). Data will also be made available on reasonable request.

Funding

Conflict of Interest

The authors declare that they have no conflict of interest.

Additional Information

Communicated by Ludovic Berthier.

Acknowledgments

5.A Escape Times, Resamplings and Gridlocks

In this appendix we collect a number of numerical results that support statements made in the main text.

5.A.1 Critical Maximum Nearest-Neighbor Distance

In the escape time t_{esc} of Eq. (5.10), the parameter γ sets the critical maximum nearest-neighbor distance $d(t)$ for the escape from an ε-relaxed Böröczky configuration. In Section 5.3.2, we use $\gamma = 0.1$ which corresponds to a 10 %-increase of the initial value $d(t = 0) = 2$. Using the alternative value $\gamma = 0.025$, we find that the escape time of straight ECMC again varies algebraically as $t_{esc} \sim \varepsilon^{-2/3}$ and, for forward, reflective, and Newtonian ECMC we again find $t_{esc} \sim \ln(1/\varepsilon)$ (see Fig. 5.7). Our conclusions thus appear robust with respect to the value of γ.

5.A.2 Escape Times with Resamplings

The reflective, forward and Newtonian variants of ECMC, at a difference of straight ECMC, appear to not always require resampling. In the main text, we therefore use $\tau_{chain} = \infty$, which, given our discussion in Section 5.3.1, is appropriate. Moreover, resamplings after chain times τ_{chain} considerably
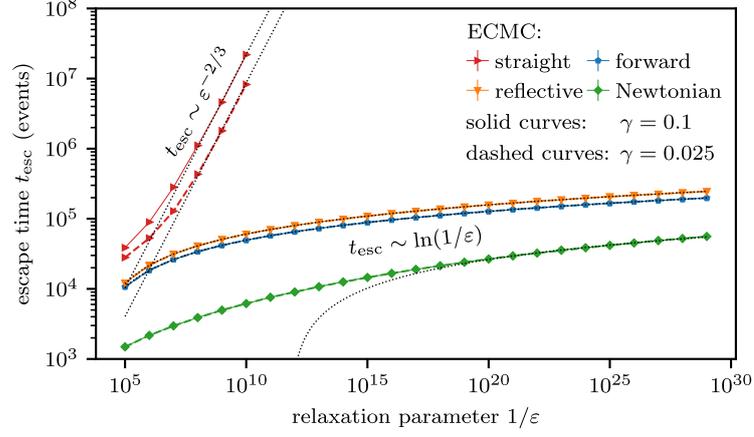
Figure 5.7: Median escape times t_{esc} from $k = 5$ ε-relaxed Böröczky configurations (Kahle core and convex geometric chain $\mathcal{A}^{geo}$ with attenuation parameter $\phi = 0.7$, $N = 96$ disks) for ECMC algorithms. Solid curves use $\gamma = 0.1$ for the definition of t_{esc} (as in Section 5.3.2), dashed curves use $\gamma = 0.025$ [see Eq. (5.10)]. For both values of γ, straight ECMC with optimized chain time τ_{chain} shows algebraic scaling with identical exponents, whereas forward, reflective, and Newtonian ECMC scale logarithmically. Error bars are smaller than the marker sizes.
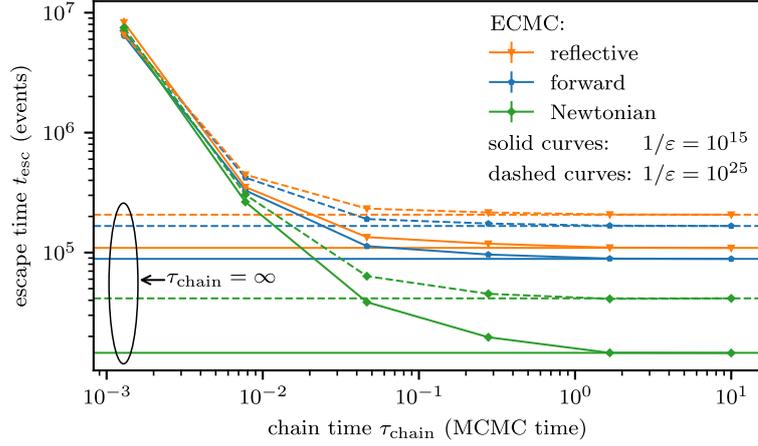
Figure 5.8: Median escape times t_{esc} from $k = 5$ ε-relaxed Böröczky configurations (Kahle core and convex geometric chain $\mathcal{A}^{geo}$ with attenuation parameter $\phi = 0.7$, $N = 96$ disks) for forward, reflective, and Newtonian ECMC vs. chain time τ_{chain} for two different relaxation parameters ε. Horizontal lines indicate the escape times without any resamplings. Error bars are smaller than the marker sizes.

deteriorate the escape time for all three variants (see Fig. 5.8). This again illustrates the power of lifted Markov chains, in which the proposed moves are correlated over long Monte-Carlo times.

5.A.3 Gridlock Rates with Different Numerical Precisions

The straight, reflective, and Newtonian variants of ECMC feature deterministic collision rules, and they may run into gridlocks if started from ε-relaxed Böröczky configurations for very small ε (see
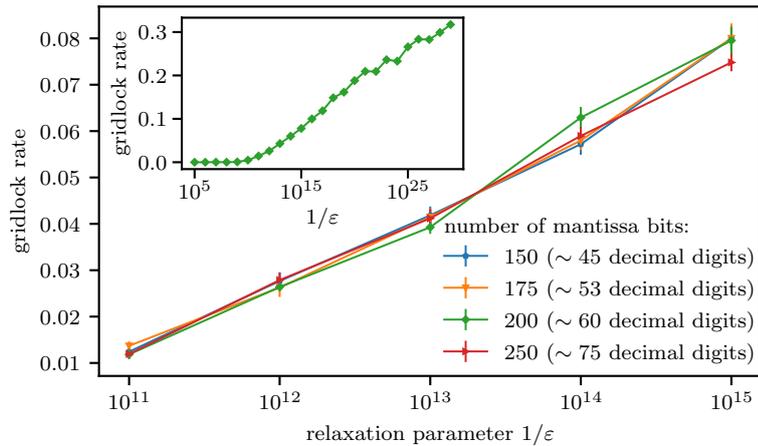
Figure 5.9: Gridlock rate of Newtonian ECMC simulations with different numerical precisions starting from a $k = 5$ ε-relaxed Böröczky configuration (Böröczky core and convex geometric chain $\mathcal{A}^{\text{geo}}$ with attenuation parameter $\phi = 0.7$, $N = 112$ disks). The inset shows the gridlock rate as a function of $1/\varepsilon$ for the Newtonian ECMC simulations with 200 mantissa bits that were used to measure the escape times in Fig. 5.6b.

Section 5.3.2.4). In a gridlock, the active-disk label loops through a subset of the N disks which are in contact. The event rate diverges, and so does the CPU time spent in the gridlock. The Monte-Carlo time, however, stands still. Newtonian ECMC starting from $k = 5$ ε-relaxed Böröczky configurations with the Böröczky core appears particularly prone to gridlocks.

It remains an open question whether gridlocks are a numerical artifact related to the finite-precision computer arithmetic. In our arbitrary-precision `BigBoro` software, the number of mantissa bits (in base 2) can be set freely. We have studied the gridlock rate of Newtonian ECMC (the fraction of simulations that run into gridlock) for the problematic $k = 5$ ε-relaxed Böröczky configurations (using the convex geometric chain $\mathcal{A}^{\text{geo}}$ with attenuation parameter $\phi = 0.7$) with the Böröczky core, and observed no clear influence of the numerical precision. It thus appears unlikely that gridlocks are a precision issue (see Fig. 5.9).

5.B `BigBoro` Software Package: Outline, License, Access

The `BigBoro` software package is published as an open-source project under the GNU GPLv3 license. It is available on GitHub as part of the `JeLLyFysh` organization.[1] The software package consists of three parts: First, the arbitrary-precision Python script `construct_packing.py` constructs finite-N Böröczky packings of hard disks in a periodic square box. Second, the Python script `collective_escape_modes.py` computes collective infinitesimal displacements of hard disks in a packing that result in an escape. Third, the arbitrary-precision Go application `go-hard-disks` performs hard-disk ECMC simulations that may start from ε-relaxed Böröczky configurations derived from Böröczky packings.

---

[1] The url of repository is `https://github.com/jellyfysh/BigBoro`.

5.B.1 Python Script `construct_packing.py`

The arbitrary-precision Python script `construct_packing.py` implements the construction of Böröczky packings. It allows for the Böröczky or Kahle cores (see Section 5.2.1.1), and connects them to branches with a finite number of layers (see Section 5.2.1.3). The convex geometric chain $\mathcal{A}^{\mathrm{geo}}$ with different attenuation parameters ϕ, and the convex circular chain $\mathcal{A}^{\mathrm{circ}}$ are implemented (see Section 5.2.2.1). The core, the number of layers, and the convex polygonal chain are specified using command-line arguments. The construction of the Böröczky packings uses arbitrary-precision decimal floating-point arithmetic. Two additional command-line options specify the number of decimal digits, and the precision of the bisection search for the value $g_2^<$ that renders the Böröczky packing compatible with periodic boundary conditions (see Section 5.2.1.3). The final configuration and its parameters (as for example the system length) are stored in a human-readable format in a specified output file.

The `example_packings` directory of `BigBoro` contains several Böröczky packings in corresponding subdirectories (as for example `kahle_geometric_5`). The headers of these files contain the values of the command-line arguments for `construct_packing.py`. A plot of each example configuration is provided. The different packings in `kahle_geometric_5` and `boro_geometric_5` (see Fig. 5.1) were used in this work. Although the bisection search for the construction of the Böröczky packing usually requires an increased precision, the high-precision packings with small enough number of layers may be used as input for standard double-precision applications. For simplicity and improved readability, we provide `packing_double.txt` files that store the configurations in double precision, where applicable.

5.B.2 Python Script `collective_escape_modes.py`

The double-precision Python script `collective_escape_modes.py` identifies the orthonormal basis vectors of the escape matrix $\mathcal{M}^{\mathrm{esc}}$ from a packing $\mathbf{x}$ [see Eq. (5.6)] that have zero singular values. Afterwards, these modes are classified using the cost function in Eq. (5.8). The resulting basis vectors $\mathbf{\Delta}_a$ form the solution space for $2N$-dimensional displacements $\mathbf{\Delta} = (\Delta_1^x, \Delta_1^y, \Delta_2^x, \Delta_2^y, \ldots)$ that have a vanishing first-order term in Eq. (5.5) and thus for collective infinitesimal displacements $\mathbf{\Delta}$ of all disks that escape from the packing. The basis vectors $\mathbf{\Delta}_a$ are stored in a human-readable output file, and optionally represented as in Fig. 5.2. The input filename of the packing, and the output filename for the collective escape modes are specified in command-line arguments. Further optional arguments specify the filename for the plots of the escape modes, and the system length of the central simulation box (that is unnecessary for packings generated by the Python script `construct_packing.py` in which case the system length is parsed from the packing file).

5.B.3 Go Application `go-hard-disks`

The Go application `go-hard-disks` relies on a cell-occupancy system for the efficient simulations of large-N hard-disk systems using several variants of the ECMC algorithm. Straight, reflective, forward, and Newtonian ECMC are implemented. After each sampling interval, it samples the maximum nearest-neighbor distance $d(t)$ [see Eq. (5.9)]. All computations use a fixed number of mantissa bits (in base 2) that may exceed the usual 24 or 53 bits for single- or double-precision floating-point

values. The ECMC variant, its parameters (as for example the sampling time or chain time), and further specifications (the number of mantissa bits, the cell specifications, the filename for the initial configuration, etc.) are again set using command-line arguments.

Research Paper 4—Molecular Simulation from Modern Statistics: Continuous-Time, Continuous-Space, Exact

This chapter considers the preprint

[P4] **Philipp Höllmer**, Anthony C. Maggs, and Werner Krauth,
Molecular simulation from modern statistics: Continuous-time, continuous-space, exact,
arXiv:2305.02979v2 [physics.chem-ph] (2023),
URL: https://doi.org/10.48550/arXiv.2305.02979.

For convenience, this manuscript, which is currently submitted for publication, is reproduced in the present chapter. However, since the manuscript is at this point in time only a preprint and not a peer-reviewed publication (as in Chapters 3–5), Section 6.0 rephrases its entire main part for the context of this doctoral thesis. Sections 6.1–6.5 then reproduce the preprint version. Because the author of this doctoral thesis was the sole creator of the figures in the preprint, they are still explicitly referenced in the independent contextualization in Section 6.0 to avoid unnecessary duplication. Likewise, the supplementary materials, which are reproduced in Sections 6.A and 6.B, were single-handedly written by the author of this doctoral thesis (with review and editing by the co-authors). They are therefore not explicitly rephrased but referenced where necessary. At the end of Section 6.0, the author's contributions to the preprint are clearly indicated and confirm the responsibility statements on the figures and supplementary materials. All co-authors of the preprint agree with its inclusion in this doctoral thesis and with the identification of the author's contributions.

6.0 Prologue

The final research paper of this doctoral thesis, which is included in this chapter and contextualized within this thesis in the following, achieves the overarching objective by introducing, *at long last*, the competitive, rigorous paradigm for molecular simulations in the thermodynamic equilibrium. It is

based on non-reversible Markov-chain Monte Carlo (MCMC) algorithms and rigorously exact from
the beginning. By construction, the Boltzmann distribution of the canonical (*NVT*) is strictly sampled
in a non-reversible Markov chain. Long-range-interactions in molecular systems like the Coulomb or
Lennard-Jones interactions are treated without any approximations, yet with competitive efficiency as
shown in this chapter.

As a specific example of the introduced paradigm, we consider the non-reversible event-chain
Monte Carlo (ECMC) algorithm. Its general event-driven approach is implemented in the open-source
JELLYFYSH application whose second version was developed and released together with this research
paper (see Fig. 6.1) [P6, P9]. The potential energy $U = \sum_M U_M$ in molecular systems is a sum
over factor potentials U_M. The Boltzmann distribution factorizes into a product of distributions of
factors M that are treated statistically independent in ECMC (see Section 2.4.1). At every point in
time, a single active interaction site i moves in a straight-line trajectory with constant velocity $\mathbf{v}_a$.
The positions and velocities of the interaction sites define the *global state* of the physical system.
In JELLYFYSH-Version2.0, the global state is only accessed by the central *mediator* which imposes
coherency of the physical system [208]. The piecewise-deterministic motion of the active interaction
site is interrupted by an event that results from a competition between stochastic inhomogeneous
Poisson processes for every factor M (see Section 2.4.3). The corresponding *candidate event times* of
every factor M are computed by *event handlers*. A *scheduler* determines the winning, event-triggering
factor M_{ev} with the minimum candidate event time. The next active interaction site $j \in M_{\text{ev}}$ and its
new velocity $\mathbf{v}'_a$ follow from a lifting move of M_{ev} (see Section 2.4.7). Every factor potential only
depends on a small number of interaction sites and can be accurately treated up to machine precision
(see Section 2.4.2). At the same time, the cell-veto algorithm heavily reduces the required number of
these small computations (see Section 2.4.6).

In order to prove the competitiveness of the introduced paradigm, we consider large-scale systems
of SPC/Fw water molecules in JELLYFYSH-Version2.0. The SPC/Fw water model already contains
the computationally intensive Coulomb interactions and appears in many molecular simulations as
an explicit aqueous solution (see Fig. 1.1). The benchmark is the integrated autocorrelation time
(see Section 2.2.2) of the electric polarization (or total electric dipole moment) that characterizes
the rotation dynamics of the water molecules. In ECMC, the polarization is expected to decorrelate
especially slowly because the rotation of individual water molecules has to be pieced together from
straight-line motions of single interaction sites in between events (see Section 1.3.2).

In the three previous research papers of this doctoral thesis, Newtonian ECMC emerged as the most
promising ECMC variant for the required rotations of SPC/Fw water molecules (see Chapters 3–5 and
Refs [P1, P2, P3]). In these cases, we only considered hard-disk systems or close equivalents for which
Newtonian ECMC was originally proposed [124]. The peculiar factor potentials U_M contain infinite
steps and consider two interaction sites. In contrast, the factor potentials U_M in molecular systems are
smooth and may depend on an arbitrary number of interaction sites (as the bending potential between
three interaction sites or the molecular Coulomb potential between six interaction sites in Fig. 1.1; see
also Section 2.4.1).

We introduce generalized Newtonian ECMC in the supplementary materials of this research paper
(see Section 6.B.1). As in the original Newtonian ECMC algorithm, every interaction site has its

own hypothetical velocity where the velocities are initially drawn from the Maxwell–Boltzmann distribution. At any time of the simulation, only one of the interaction sites actually moves with its velocity. An event in the hard-disk system changes the velocities of both colliding disks according to an elastic Newtonian collision. The previously active disk stops moving and its collision partner starts to move instead with its velocity (see Fig. 1.5). An event of generalized Newtonian ECMC in the factor M_{ev} generalizes the Newtonian collision to a "force kick." Here, the hypothetical velocity of every interaction site $k \in M_{\text{ev}}$ is updated in the direction of the gradient of the factor potential $\boldsymbol{\nabla}_{\mathbf{x}_k} U_{M_{\text{ev}}}$ at the position $\mathbf{x}_k$ of k. Such a force kick is, however, not applied in every event. Instead, generalized Newtonian ECMC decides stochastically whether the force kick changes all hypothetical velocities of the interaction sites $k \in M_{\text{ev}}$ and which interaction site $j \in M_{\text{ev}}$ actually becomes active with its (possibly unmodified) velocity. As in molecular dynamics, one can introduce a general mass matrix for the interaction sites in this generalized scheme. In contrast to straight, reflective, and forward ECMC, one does not require translationally invariant factor potentials anymore to prove that the necessary global-balance condition is satisfied (see Section 2.4.7). The original Newtonian ECMC algorithm for the hard-disk system can be recovered in the limit of equal masses and translationally invariant factor potentials.

In the supplementary materials, this research paper also generalizes the cell-veto algorithm from a finite number of possible velocities $\mathbf{v}_a$ of the active interaction site i, as it appears in the straight ECMC variant (see Section 2.4.6), to the continuous space of possible velocities, as it appears, e.g., in generalized Newtonian ECMC (see Section 6.B.3). This is achieved by discretizing the continuous velocity space into Voronoi cells of a finite number of Fibonacci vectors (see Fig. 6.2). The cell bounds in the original cell-veto algorithm are then computed for the finite set of Fibonacci vectors. During the simulation, the deviation between the actual velocities of the active interaction sites and the Fibonacci vectors were already considered in the estimation of the cell bounds. This ensures that the cell-veto algorithm still treats long-range interactions rigorously exact. Generalized Newtonian ECMC and the generalized cell-veto algorithm are implemented in JᴇLLʏFʏsʜ-Version2.0 that accompanies the research paper in this chapter [P9].

For the factorization of the SPC/Fw water model in Section 2.4.1 that, in particular, combines the set of Coulomb interactions between two water molecules into a single factor (as depicted in Fig. 1.1), theory predicts that the number of events per unit distance grows as $O(\log N)$ with an increasing number N of water molecules [146]. The cell-veto algorithm should further yield a constant computer time per event. Both of these results are confirmed in JᴇLLʏFʏsʜ-Version2.0 for large N (see Fig. 6.3). This numerically confirms a computational complexity of $O(N \log N)$ to move every interaction site by a constant distance, which matches the computational complexity of fast mesh-based Ewald methods in molecular dynamics [50].

As a first measure of the efficiency in the dynamics of different simulation methods, we considered the characteristic distance which the interaction sites of any water molecule have to travel to decorrelate the polarization (see Section 6.A.6 in the supplementary material for details). Straight ECMC, for which we could only consider small numbers of water molecules, is only slightly faster than the reversible Metropolis algorithm with local Monte Carlo moves of single interaction sites. Here, we use the DL_MONTE software package for the simulations of the Metropolis algorithm [108]. (Reflective

ECMC yields very comparable results to straight ECMC which is, however, not shown in the research paper.) Generalized Newtonian ECMC, in contrast, is an order-of-magnitude faster which resolves the dynamical arrest of straight ECMC. Its local dynamics even comes close to the fastest collective dynamics of molecular dynamics (MD; see Fig. 6.3). This may be related to the fact that generalized Newtonian ECMC injects ideas of MD into its events. Because of the factorization of the Boltzmann distribution, event rates between different factors cannot compensate each other. However, as the favorable results for generalized Newtonian ECMC show, this does not penalize the dynamics.

Generalized Newtonian ECMC, as implemented in JeLLyFysh-Version2.0, treats the long-range Coulomb interaction exact up to machine precision. MD, as implemented in Lammps [54], sets a target accuracy instead and its computational cost per discrete time step increases with increasing accuracy. We show that JeLLyFysh-Version2.0 reaches a break-even point with Lammps at a high target accuracy which is well below machine precision (see Fig. 6.4). This is the first proof of concept that the proposed rigorous paradigm for molecular simulations based on non-reversible Markov chains is competitive. This is particularly highlighted by the fact that JeLLyFysh-Version2.0 is a demonstration software that is developed by a very small number of people and runs with a just-in-time compiler of Python called PyPy. Lammps, in contrast, is a long-developed software package with hundreds of contributors that is implemented in C++. Hence, already re-engineering JeLLyFysh-Version2.0 and, in particular, optimizing the parameters of the cell-veto algorithm may push down the break-even point with Lammps considerably. The general, far-reaching implications of this first proof for the competitive efficiency of the rigorous paradigm for molecular simulations based on non-reversible Markov chains are discussed in the conclusion of this doctoral thesis (see Section 7.1).

Contributions of the Author

The author contributed in an essential manner to the entire research project underlying this research paper. In particular, he was the sole contributor to the entire update of the open-source JeLLyFysh software to JeLLyFysh-Version2.0, the generalization of Newtonian ECMC, the data collection in JeLLyFysh-Version2.0 and DL_MONTE, and the creation of the figures. He single-handedly wrote the supplementary materials (with review and editing by the co-authors). The author also made essential contributions to the generalization of the cell-veto algorithm and the writing of the research paper itself. Following the Contributor Role Taxonomy (CRediT) [174], the author contributions are as follows:

- Conceptualization (ideas; formulation or evolution of overarching research goals and aims).
- Methodology (development or design of methodology; creation of models).
- Software (programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components).
- Validation (verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs).
- Formal analysis (application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data).

- Investigation (conducting a research and investigation process, specifically performing the experiments, or data/evidence collection).

- Writing—Original Draft [preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)].

- Writing—Review & Editing (preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision—including pre- or post-publication stages).

- Visualization (preparation, creation and/or presentation of the published work, specifically visualization/data presentation).

Molecular Simulation from Modern Statistics: Continuous-Time, Continuous-Space, Exact

This reproduces P. Höllmer, A. C. Maggs, and W. Krauth, *Molecular simulation from modern statistics: Continuous-time, continuous-space, exact*, arXiv:2305.02979 [physics.chem-ph] (2023), which is licensed under the Creative Commons Attribution 4.0 International License and may be found at `https://doi.org/10.48550/arXiv.2305.02979` (see Ref. [P4]). Note that the entire main part of this research paper in Sections 6.1–6.5 was rephrased for the context of this doctoral thesis in Section 6.0. The supplementary materials in Sections 6.A and 6.B were single-handedly written by the author of this doctoral thesis (with review and editing by the co-authors). The author was also the sole creator of all figures.

Abstract   In a world made of atoms, the computer simulation of molecular systems, such as proteins in water, plays an enormous role in science. Software packages that perform these computations have been developed for decades. In molecular simulation, Newton's equations of motion are discretized and long-range potentials are treated through cutoffs or spacial discretization, which all introduce approximations and artifacts that must be controlled algorithmically. Here, we introduce a paradigm for molecular simulation that is based on modern concepts in statistics and is rigorously free of discretizations, approximations, and cutoffs. Our demonstration software reaches a break-even point with traditional molecular simulation at high precision. We stress the promise of our paradigm as a gold standard for critical applications and as a future competitive approach to molecular simulation.

One-sentence summary   A rigorous paradigm for exact non-reversible Markov processes is benchmarked for classical long-range-interacting water.

6.1 Introduction

The fact that all matter consists of atoms has been described as the greatest insight of science [1]. The consequence that matter can be modeled on a computer by following the motion of its atoms leads to the founding paradigm of molecular simulation. It tracks the dynamics and explores the thermodynamic equilibrium of complex molecular systems, for example, a peptide in an explicit water solution with tens of thousands of atoms, all interacting through classical empirical potentials [17]. Molecular simulation is of enormous importance to numerous fields ranging from biology and physics to engineering [11, 13]. Powerful computer packages have been developed over decades [54, 103–106]. They compute the forces on all atoms at discretized time steps and then update the atomic positions and velocities to integrate the classical equations of motion of molecular dynamics. A voluminous literature is dedicated to the analysis and control of time-discretization errors (see, e.g., Ref. [44]). Thermostats, understood as "necessary evils" [46], mimic the effect of a coupled thermal reservoir and, in a symptomatic but non-curative treatment, hide the accumulated errors. The limiting factor in molecular dynamics is the computation of forces. The Lennard-Jones interaction is typically cut off

beyond a certain distance so that only a few neighbors exert a force on any given atom. The long-range nature of the Coulomb potential, which must be preserved, is usually treated through fast mesh-based Ewald methods [49–51, 96, 97] that solve the Poisson equation in discretized space. Thermostats and cutoffs as well as the discretizations inherent in long-range interactions introduce approximations and artifacts [46, 209] that complex algorithms aim to keep under control.

We present here an alternative paradigm for molecular simulation that is based on modern concepts in statistics. It is rigorously exact from the start and, by construction, strictly simulates the canonical ensemble without thermostats. Straight-line trajectories of atoms in continuous Monte-Carlo time are interrupted by *events*. This non-reversible piecewise-deterministic Markov process [136] violates the detailed-balance condition normally associated with thermal equilibrium, but still samples the Boltzmann distribution. Its use has led to spectacular speedups of local Markov chains in statistical physics [142]. Short- and long-range potentials are handled without any cutoffs or discretizations and, as we show in this paper, with competitive efficiency. The Boltzmann weight $\pi = \exp(-\beta U)$ (with β the inverse temperature and U the potential) is expressed as a factorized product $\pi = \prod_M \exp(-\beta U_M)$ of statistically independent factors M with factor potentials U_M with $\sum_M U_M = U$ that each depend only on a small subensemble of atoms [121]. Every factor stochastically generates a time when the piecewise-deterministic motion must be interrupted. The minimum of these times triggers an event, and determines the initial conditions for the next piece. The total potential U and the corresponding forces never need be evaluated, yet the stationary state is rigorously the Boltzmann distribution.

We implement the event-driven paradigm in demonstration software for N flexible SPC/Fw water molecules [28] that interact with the long-range Coulomb potential. As a benchmark, we concentrate on the electric polarization [48] (in other words, the total electric dipole moment) that measures the capacity of individual water molecules to rotate and determines the dielectric properties of water. We find that a particular non-reversible Markov process greatly reduces autocorrelation times and overcomes the slow diffusive dynamics of reversible Monte Carlo algorithms, while the factorization does not penalize the dynamics in comparison to molecular dynamics. The polarization decorrelates in a computer time that scales as $N \log N$, similar to mesh-based Ewald methods in molecular dynamics [50] but without their diverging precision-dependent prefactor. Our code reaches a break-even point with respect to a standard molecular-dynamics code much below machine precision, and we point out its great potential for improvement.

6.2 Modern-Statistics Paradigm for SPC/Fw Water

In a molecular system with long-range interactions, the force on an atom depends on the position of all other atoms, rendering its evaluation tedious unless one introduces cutoffs or discretizes space. In contrast, we implement a piecewise-deterministic Markov process through the event-chain Monte Carlo algorithm, where a single atom moves at any given moment [80, 115, 121]. The deterministic motion of this atom is interrupted by an event that stops it and sets off a similar motion of a new atom. Factors M in the SPC/Fw water model describe O−H bonds, the bending of H−O−H opening angles, O−O intermolecular Lennard-Jones interactions, and the Coulomb interaction between two water molecules. Each factor proposes an independent *candidate event time*. The minimum over all the candidate event

times then realizes the next event, and motion is transferred to another atom contributing to U_M. This succession of events, in which rejections are replaced by transfers, distinguishes our formulation from the usual Metropolis Monte Carlo algorithm [72].

Positions and velocities of atoms define the *global state* of the physical system. To impose coherency of the physical system, the global state is accessed only through a central *mediator* [208] that dispatches physically independent computations of candidate events to *event handlers*. A *scheduler* weeds through candidate events and identifies the unique event that provides the subsequent transfer of motion, leading to an update of the global state (see Fig. 6.1). The event handlers within the mediator architecture mirror the statistical independence of the factors composing the physical system. This allows us to compose complex interactions in a transparent and independent manner. A number of inequivalent options have been constructed for the update of active particles within event-chain Monte Carlo [145, 146]. Similar flexibility is possible in the updating of velocities [P2, 115, 124, 150], as well as in parallelizing [P6, 200] it. In this paper, we replace the original straight variant by the substantially more efficient Newtonian event-chain Monte Carlo [P2, 124]. It requires no fine-tuning and again exactly samples the canonical Boltzmann distribution. Modern statistics offers an even wider choice of options for the management of events, the choice of factors, and the piecewise-deterministic trajectories, that may well apply to molecular simulation in the approximation-free non-reversible Markov-chain framework.

6.3 Implementation for SPC/Fw Water: Cell-Veto—Fibonacci Sphere

In the event-driven implementation of our method (see Fig. 6.1), a single atom moves among the other atoms and molecules, so that $O(N)$ factors are changing with time and in principle yield independent events that would require sorting and managing. Modern statistics allows one to bundle most of these factors and, in the SPC/Fw water model, the mediator interacts with only ~ 50 event handlers that propose candidate event times to the scheduler (see Fig. 6.2a). The bundling allows the processing of each event in constant computer time (for large N) while treating the long-range interactions without approximations.

In the cell-veto algorithm [148], which we use for the bundling, event rates for pairs of molecules interacting with the Coulomb potential are upper-bounded by precomputed, time-independent bounds for these molecules somewhere within a pair of cells (see Fig. 6.2b). The full set of these *cell bounds* corresponds to the set of bundled factors of the long-range interaction. Walker's method of aliases [166] conserves the cell bounds in a *Walker table*. We build separate such tables for multiple directions of the velocity of the moving atom corresponding to *Fibonacci vectors* on the unit sphere (see Fig. 6.2c). During the simulation, the actual velocity of the moving atom is mapped to the closest Fibonacci vector. The set of cell bounds in the corresponding table provides a single candidate event time for the entire set of bundled factors, and Walker's method samples an associated single cell bound—and thus a single associated factor—with constant algorithmic complexity. The overestimation of the event rate by the cell bound is corrected in a procedure akin to the thinning of non-homogeneous Poisson processes [165] by confirming the transfer of motion in the event (see Fig. 6.2d). This thinning is performed with the actual positions of the atoms, leading to an exact treatment of the long-range interaction that is independent of the set of cell bounds and the discretization of space.
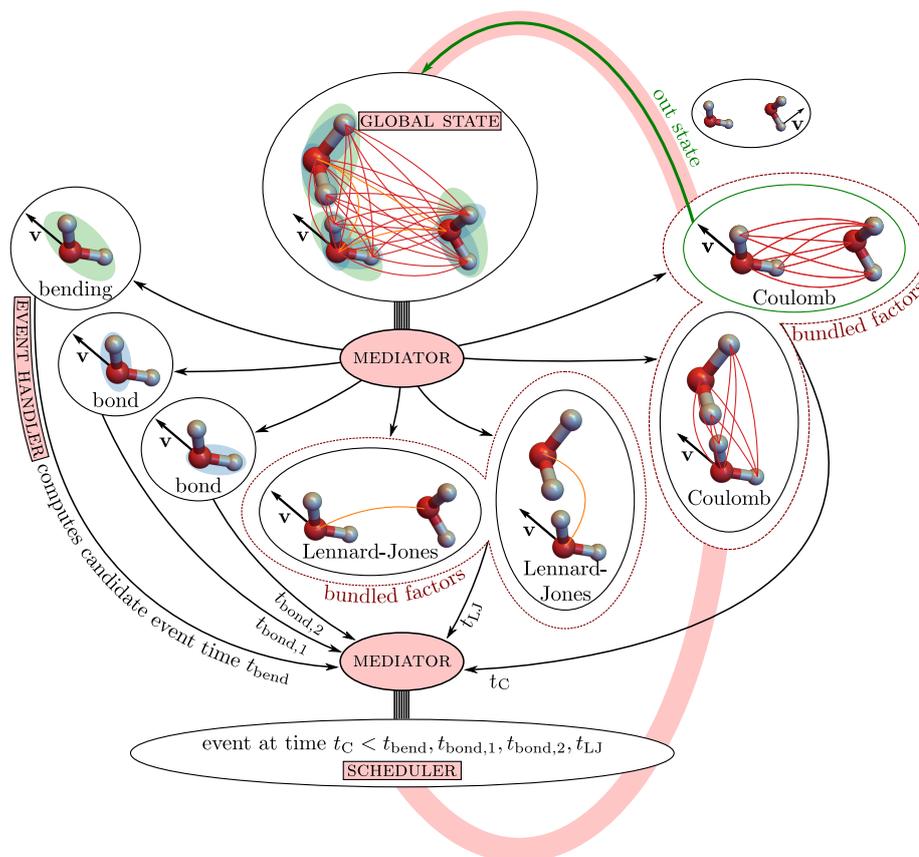
Figure 6.1: JᴇʟʟʏFʏsʜ implementation of our Markov-chain paradigm. The *mediator* splits the *global state* into statistically independent factors. Factors communicate independent candidate event times, the earliest of which defines the next event. Factors for long-range interactions are bundled, so that the number of *event handlers* remains limited. Candidate event times are collected by the mediator and then treated in the *scheduler*. The factor triggering the event then updates the global state, again *via* the mediator.

6.4 Benchmarking for SPC/Fw Water

For our benchmark, we use JᴇʟʟʏFʏsʜ [P6] to sample configurations of SPC/Fw water molecules in a periodic box at standard density and temperature. We implement long-range molecular Coulomb factors with Walker tables that we also adopt for the Lennard-Jones interaction. We find that Newtonian event-chain Monte Carlo, for large N, requires a computer time per event that remains constant (see Fig. 6.3a). A large number of unconfirmed events stems from the overestimated cell bounds which, e.g., do not account for the relative orientation of molecules. While the computer time per event is constant, the number of events per Ångström (that is, per unit Monte-Carlo time) increases logarithmically for the Coulomb factors (see Fig. 6.3b), as predicted by theory [146]. In summary, our approach requires a computer time scaling as $N \log N$ to advance N water molecules by a constant distance. This matches the complexity of mesh-based Ewald methods [50], but without their slowdown as the precision is increased.
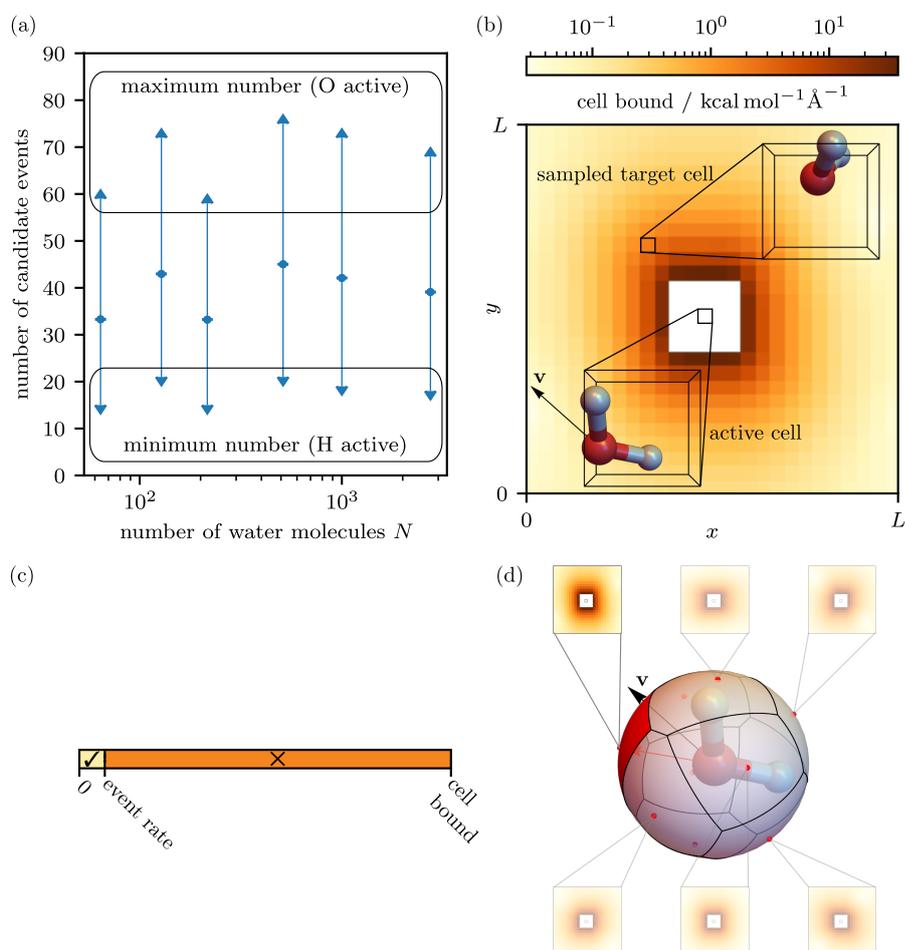
Figure 6.2: Long-range interactions with constant computer time per event. (a): The number of candidate events (event handlers) is constant for increasing system sizes. (b): Walker table from which a target cell is sampled according to its cell bound with respect to the active cell containing the moving atom. (c): Different Walker tables for Fibonacci vectors on the unit sphere. The active atom obtains cell bounds from the nearest vector. (d): The thinning procedure confirms (✓) or rejects (✗) the event using the actual event rate of the molecules in the active and target cell. Figure adapted to the layout of this thesis from Fig. 2 in Ref. [P4] that is licensed under the Creative Commons Attribution 4.0 International License.
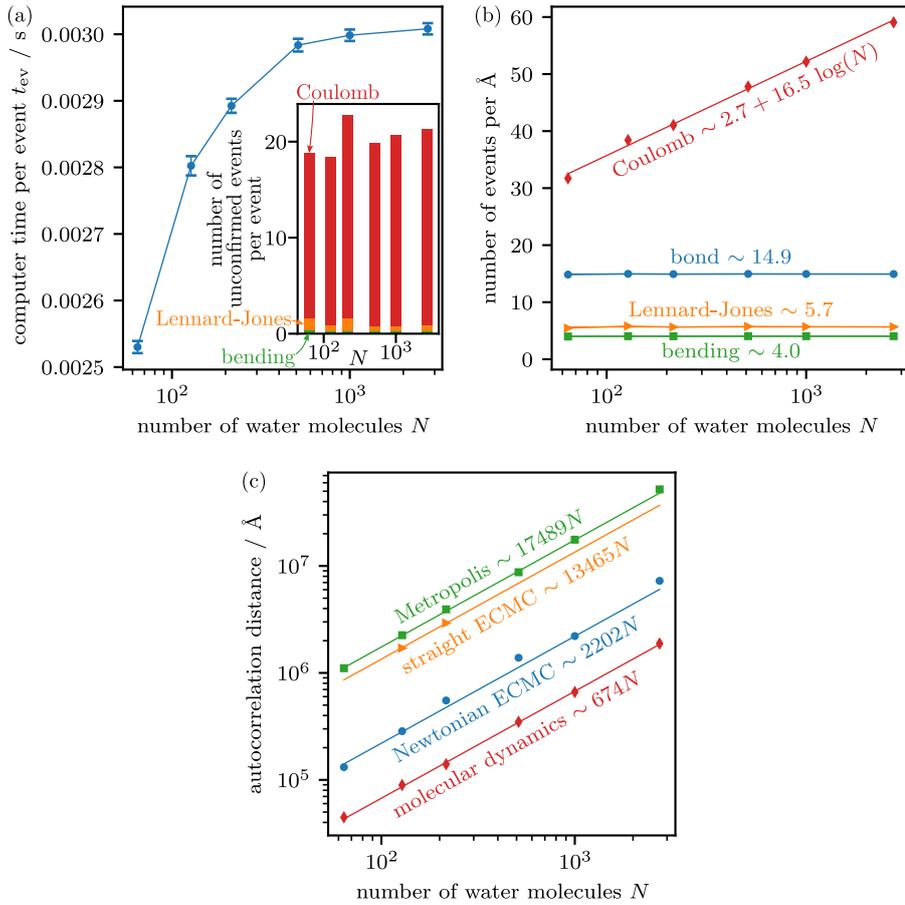
Figure 6.3: Event rates and decorrelation in the SPC/Fw water model. (a): Computer time per event in JᴇLLʏFʏsʜ. Inset: Number of unconfirmed events per event for different factor types. (b): Event rate in JᴇLLʏFʏsʜ for different factor types. (c): Distance to decorrelate the polarization for different sampling algorithms (for molecular dynamics: sum over the average displacements of all atoms per time step). Figure adapted to the layout of this thesis from Fig. 3 in Ref. [P4] that is licensed under the Creative Commons Attribution 4.0 International License.

For concreteness, we compare the decorrelation of the polarization within JᴇLLʏFʏsʜ to the LᴀᴍᴍᴘS software on a single processor with default parameters and a 1 fs time step. To decorrelate this local quantity, both LᴀᴍᴍᴘS and JᴇLLʏFʏsʜ must move the atoms of any water molecule by a characteristic distance, providing a first measure of efficiency (see Fig. 6.3c). Different variants of our method vary in their efficiency, and the recently developed Newtonian event-chain Monte Carlo is an order of magnitude faster than the straight variant. The reversible Metropolis algorithm with single-atom moves (that also reaches an $N \log N$ scaling [99] by using a recent variant of the fast multipole method [95]), as implemented in the DL_MONTE software package [108], is clearly inferior to our non-reversible methods.

The mesh-based Ewald method implemented in LᴀᴍᴍᴘS comes with a target accuracy that is based on analytic error estimates obtained from a specific charge distribution [96, 147, 210]. The charges
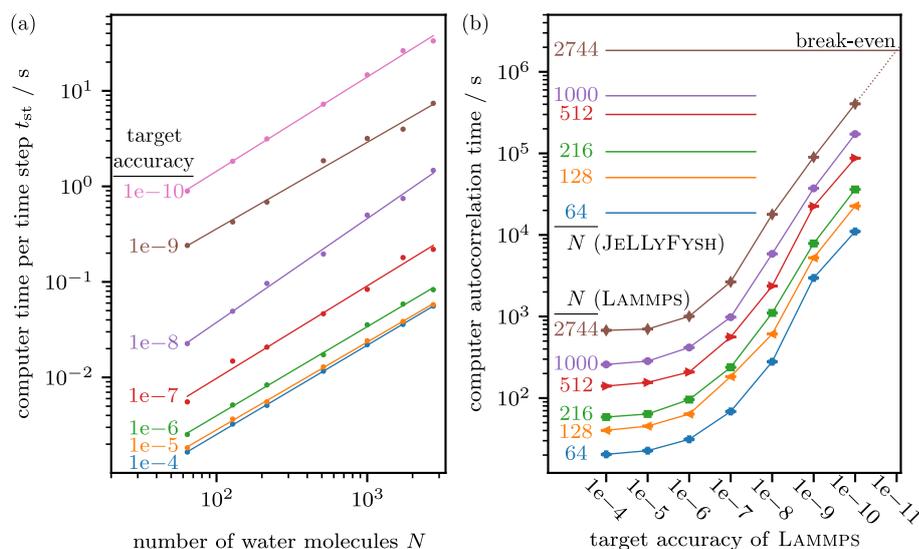
Figure 6.4: LAMMPS–JELLYFYSH benchmark for the SPC/Fw water model. (a): Computer time per step of LAMMPS for different target accuracies of its particle–particle particle–mesh solver. (b): Computer time used by LAMMPS to decorrelate the polarization depends on the target accuracy and the number of water molecules N. JELLYFYSH is exact up to machine precision. The break-even precision is indicated. Figure adapted to the layout of this thesis from Fig. 4 in Ref. [P4] that is licensed under the Creative Commons Attribution 4.0 International License.

of atoms (which live in continuous space) are mapped onto the grid using an interpolation scheme. Finer grids and higher interpolation orders yield higher target accuracies. We estimate the necessary computer times by changing the grid spacing, using the LAMMPS defaults (see Fig. 6.4a). JELLYFYSH uses the two-atom Coulomb interaction calculated with the historic Ewald summation in continuous space. We tune it to machine precision without any assumptions on the global charge distribution. JELLYFYSH reaches the break-even point to LAMMPS well below machine precision (see Fig. 6.4b), although it has only the status of demonstration software.

6.5 Discussion

In this paper, we benchmarked an implementation of a modern-statistics paradigm for molecular simulations in the standard SPC/Fw water model. The time dependence of the corresponding Markov process differs from the physical dynamics yet it exactly approaches thermal equilibrium on time scales that are potentially faster than in nature. Its remarkable efficiency (that we expressed as a $N \log N$ computer time to decorrelate a local observable, the polarization) is rooted in three paradoxes. First, the Markov process is non-reversible (that is, effectively out-of-equilibrium), yet its steady state coincides with the equilibrium Boltzmann distribution. In contrast to standard Monte Carlo algorithms that satisfy the detailed-balance condition and only move diffusively, it features finite probability flows, making it capable of moving ballistically. This has already led to considerable speedups in a variety of fields ranging from physics to statistics and machine learning (see, e.g., Refs [80, 138, 139, 211]).

It remains to be seen whether non-local observables, for example large-scale hydrodynamic modes, macroscopic conformations, and order parameters, can similarly benefit from non-reversibility in chemical physics. The second paradox is that in our approach, the Boltzmann distribution $\exp(-\beta U)$ is sampled without any approximation yet with great efficiency because the potential U (or its derivatives, the forces) are never evaluated. This sidesteps all the problems with limited-precision calculations of energies and forces. The third paradox is the bundling of $O(N)$ independent decisions to interrupt the straight-line trajectory into an expression that can be evaluated in constant time. The Walker tables channel long-range Coulomb factors into a single candidate event, which allows us to handle a complex decision (a conjunction of $O(N)$ factor-wise decisions of independent factor potentials U_M) in a few operations, even in the $N \rightarrow \infty$ limit.

Our demonstration software is openly available and fully functional. As discussed, it becomes competitive with a traditional molecular-dynamics code at high intrinsic precision. Our method is exact from the very beginning, and future research will be able to concentrate on the most efficient ones among a large choice of cell bounds, factorizations, Fibonacci vectors and variants of the piecewise-deterministic Markov processes. Clearly, more interdisciplinary research from statistics to computational chemistry will clarify whether this provides a sufficiently strong basis for an alternative approach for practical molecular simulation.

With its guarantee for the unbiased sampling of the Boltzmann distributions, our paradigm may serve as a gold standard for molecular simulation in general, capable of identifying artifacts and approximations that may not have been totally eliminated through the symptomatic algorithmic approach to molecular dynamics. Furthermore, given the greater algorithmic freedom for the Markov-chain approach than for molecular dynamics, it may actually become faster than molecular dynamics. Several orders of magnitude in algorithmic speed can certainly be gained by re-engineering our software, which would then be able to tackle the peptide-in-water benchmark problem [17] that has had a major influence over the last decade. The great simplicity of our approach and its present implementation in the JeLLyFysh software may well facilitate further developments.

Acknowledgments

Visualization: PH
Writing—original draft: PH, WK
Writing—review & editing: PH, ACM, WK
Competing interests: The authors declare that they have no competing interests.
Data and materials availability: The JeLLyFysh software is made available under the GNU GPLv3 license at https://github.com/jellyfysh. All configuration files for the simulations of this paper are part of JeLLyFysh.

6.A Supplementary Materials—Materials and Methods

In this section, we describe the SPC/Fw water model that we studied in this paper. We further specify the simulation protocols for the different sampling algorithms. Finally, we report on our analysis procedure of the time series of the electric polarization.

6.A.1 SPC/Fw Water Model

The atomistic flexible simple point-charge water model SPC/Fw defines a water molecule by three charged interaction sites that represent the oxygen and hydrogen atoms [28]. We treat the canonical ensemble in a cubic box with periodic boundary conditions, i.e., N water molecules with $N_a = 3N$ atoms in a periodically repeated cubic box of side length L at a given temperature $T \sim 300\,\text{K}$. Within any water molecule $i \in \{1, \ldots, N\}$, the intramolecular interactions of the SPC/Fw model consist of two harmonic bond potentials $U_{\text{bond}}^{i,1}$ and $U_{\text{bond}}^{i,2}$ that lead to fluctuations of the O$-$H bond lengths around their equilibrium length. Likewise, a harmonic bending potential U_{bend}^i yields a fluctuation of the H$-$O$-$H opening angle around an equilibrium value. The intermolecular interactions between two different water molecules i and j consist of a Lennard-Jones potential U_{LJ}^{ij} between the two oxygen atoms, and a Coulomb potential U_{C}^{ij} between all nine pairs of charged atoms. The intermolecular potentials explicitly include the interactions between all periodic images of the two involved water molecules. The total potential U of an all-atom configuration $\mathbf{x}$ is a sum of these factor potentials:

$$U(\mathbf{x}) = \sum_{i=1}^{N} \left[U_{\text{bond}}^{i,1}(\mathbf{x}) + U_{\text{bond}}^{i,2}(\mathbf{x}) + U_{\text{bend}}^{i}(\mathbf{x}) \right] + \sum_{i=1}^{N} \sum_{j=1}^{i-1} \left[U_{\text{LJ}}^{ij}(\mathbf{x}) + U_{\text{C}}^{ij}(\mathbf{x}) \right]. \tag{6.A.1}$$

The bond and Lennard-Jones factor potentials depend on two atomic positions; the bending and Coulomb factor potentials depend on three and six atomic positions, respectively (see also Ref. [146, Section V A]). The empirical parameters of the different potentials of the SPC/Fw water model are as in Ref. [28].

6.A.2 Simulation Protocols—Molecular Dynamics

We use the feature release from February 8, 2023, of Lammps [54, 55] for the molecular-dynamics simulations of the SPC/Fw water model in this paper. We use a spherical cut-off for the Lennard-Jones potential. The Coulomb potential is treated by a particle–particle particle–mesh solver [49]. Unless

explicitly specified otherwise (as, e.g., in Fig. 6.4b), the solver uses a target accuracy of 10^{-6}. Coulomb interactions are partly treated in discretized reciprocal space, whereby the grid size is chosen to meet the target accuracy based on analytic error estimates obtained from a specific global charge distribution [96, 147, 210]. Thermostatting is achieved by integrating Nosé–Hoover-chain equations of motion for the canonical ensemble [212] with a time-reversible measure-preserving Verlet integrator [213]. The time step is 1 fs, and the temperature is relaxed in a timespan of roughly 300 time steps. This simulation protocol for molecular dynamics in LAMMPS is validated by comparing numerical results for $N = 2$ SPC/Fw water molecules of all sampling methods of this paper (see Fig. 6.5).

6.A.3 Simulation Protocols—Metropolis Algorithm

We use version 2.07 of the DL_MONTE software package [108, 214, 215] to sample the canonical ensemble of the SPC/Fw water model with the reversible Metropolis algorithm [72], and minimally amended the software to output the electric polarization. We use a spherical cut-off for the Lennard-Jones potential. The Coulomb interaction is treated by an Ewald summation [47] with a target-accuracy tolerance of 10^{-6}. The simulations with the event-chain Monte Carlo algorithm in this paper move a single atom at a time. In order to directly compare the dynamics of the reversible Metropolis algorithm with the non-reversible event-chain Monte Carlo algorithm (see Fig. 6.3c), each trial proposes a new position of a single random atom in a proposal cube around its original position. The new position is accepted with the usual Metropolis criterion based on the change of energy. The size of the proposal cube is separately adapted for the hydrogen and oxygen atoms during the simulation to obtain a target acceptance rate of 37 %. This simulation protocol for the Metropolis algorithm in DL_MONTE is validated by comparing numerical results for $N = 2$ SPC/Fw water molecules of all sampling methods of this paper (see Fig. 6.5).

6.A.4 Simulation Protocols—Event-Chain Monte Carlo

We develop version 2.0 of JELLYFYSH [P6, P9] for the event-chain Monte Carlo simulations of the SPC/Fw water model in this paper. In addition to the straight variant implemented in versions < 2.0, version 2.0 implements the generalized Newtonian event-chain Monte Carlo variant (see Section 6.B.1, for details). We use the factorization in Eq. (6.A.1). In particular, we group all the charge–charge Coulomb interactions between two water molecules into a single factor. Theory then predicts an optimal logarithmic increase of the number of events from these molecular Coulomb factors [146] (that is numerically confirmed in Fig. 6.3b).

Straight event-chain Monte Carlo (used in Fig. 6.3c) periodically aligns the unit velocity $|\mathbf{v}_a| = 1$ of the active atom with the coordinate axes. We change the velocity after a chain time of $\tau_{\text{chain}} = 0.2N$ [149]. In an event triggered by a factor M, the velocity is transferred to another atom which belongs to M. For factors with more than two atoms, the next active atom is chosen according to the ratio lifting scheme [145, 146].

Newtonian event-chain Monte Carlo assigns a velocity label to every atom, but only moves a single one at any time. The velocities are initialized so that the average speed of the active atom during the simulation is approximately one: $\langle |\mathbf{v}_a| \rangle \approx 1$ (see Section 6.B.1 for details). This is possible because the

velocities have no kinematic meaning as they have in molecular dynamics. The physical temperature only rescales the event rates [see Eq. (6.B.5) in Section 6.B.1]. The velocities are resampled in large time intervals of $\tau_{\text{chain}} = 10\,000\,N$. This is because Newtonian event-chain Monte Carlo rotates dipoles most efficiently in the limit $\tau_{\text{chain}} \to \infty$, while a finite value of τ_{chain} ensures irreducibility [P2, P3]. Although our generalized Newtonian event-chain Monte Carlo can consider a general mass matrix of the atoms in its events, we set the masses of all atoms to be equal. Then, the events of factors M of two atoms can be treated in a deterministic way in which the two involved velocities and the active atom always change. For factors M of more than two atoms, we apply a stochastic scheme that corresponds to a generalized ratio lifting scheme (see the Newtonian-pair and Newtonian-general schemes in Section 6.B.1).

The events of the bond factors are computed exactly [see Eq. (6.B.6) in Section 6.B.1], while the bending factors are treated with a piecewise-linear bounding potential followed by a Poisson thinning procedure (see Ref. [P6, Section 4.4.5]). The cell-veto algorithm [148] bundles most molecular Coulomb factors and relies on a cell-occupancy system that tracks the molecular barycenters (we generally adapt the setup described in Ref. [P6, Section 5.3.4]). Very close molecule pairs or surplus molecules in the cell-occupancy system are separately treated with a piecewise-linear bounding potential. The cell bounds for 10 different Fibonacci vectors (see Section 6.B.3 for details) are estimated by varying the position of an atom in one cell, and the position of a dipole in the other cell. Here, the dipole is aligned with the direction of the gradient of the charge–dipole Coulomb interaction. The maximum event rate multiplied by an empirical prefactor yields the cell bound. During the simulation, the cell bounds in the Walker table are rescaled to the actual speed and charge of the active atom. Events from the cell-veto algorithm are confirmed with a Poisson thinning procedure that compares the actual event rate of the Coulomb interaction between two molecules with the sampled cell bound. To compute the real event rate, we compute the gradients of the Coulomb potential between the nine pairs of charged atoms [see also Eq. (6.B.5) in Section 6.B.1] with an Ewald summation [47] that is tuned to machine precision without any assumption on the global charge distribution. We also use the cell-veto algorithm to treat most Lennard-Jones factors, although we cut off the interaction beyond the closest images for consistency with the other sampling algorithms. We carefully checked that an alternative spherical cut-off does not change the results of this paper. The cell-veto algorithm for the Lennard-Jones interaction relies on a cell-occupancy system that tracks the oxygens. Very close or surplus oxygen pairs are treated directly [see Eq. (6.B.6) in Section 6.B.1]. The cell bounds for 10 different Fibonacci vectors are estimated by varying the oxygen positions evenly in the two cells, and again including an empirical prefactor. The parameters of the cell-occupancy systems, as the cell sizes and the number of nearby cells that are excluded from the cell-veto algorithm, are tuned to minimize computer time. As a rule of thumb, the largest possible cells that do not yield any surplus water molecules, and two excluded layers are found to be a good choice.

These simulation protocols for straight and Newtonian event-chain Monte Carlo in JᴇLLʏFʏsʜ are validated by comparing numerical results for $N = 2$ SPC/Fw water molecules of all sampling methods of this paper (see Fig. 6.5).
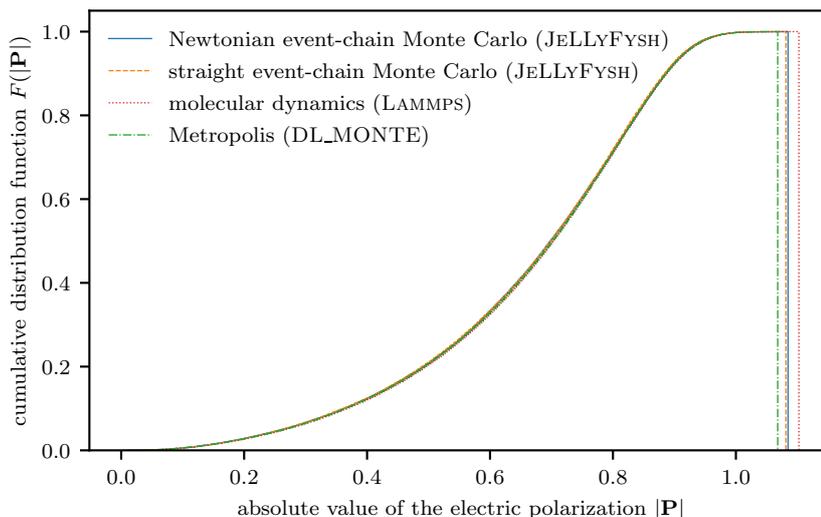
Figure 6.5: Validation for the SPC/Fw water model. Cumulative distribution function of the absolute value of the electric polarization of $N = 2$ SPC/Fw water molecules in a cubic box of side length $L = 20$ Å. The simulation protocols of the different sampling methods are described in Section 6.A.

6.A.5 Simulation Protocols—Creation of Initial Configurations

We create initial configurations in the range from $N = 64$ to $N = 2744$ SPC/Fw water molecules with hydrogen mass $m_H = 1.0079$ Da and oxygen mass $m_O = 15.9994001$ Da at a density of $\rho = 0.97$ g cm^{-3} using the software package PLAYMOL [56] (commit `67eb56c` from 26 November 2019). Initial configurations are equilibrated using a molecular-dynamics simulation of LAMMPS in the isothermal–isobaric ensemble at pressure $P = 1$ atm. Similar to thermostatting, barostatting is achieved by integrating the appropriate Nosé–Hoover-chain equations of motion with a Verlet integrator [212, 213] with a time step of 1 fs. This results in initial configurations for the simulations in the canonical ensemble with the different sampling algorithms at slightly different densities for different N. We confirmed that these small density variations do not influence the results of this paper.

6.A.6 Analysis of the Electric Polarization

We sample the electric polarization $\mathbf{P}$ (or the total electric dipole moment $\mathbf{P} = \sum_{i=1}^{N} \mathbf{p}_i$, where $\mathbf{p}_i$ is the molecular dipole moment of the water molecule i) of the SPC/Fw water system with the different sampling algorithms. The generated "time"-series $\mathbf{P}(t)$ (where the "time" t only has a a physical meaning in molecular dynamics) yields the normalized autocorrelation function

$$\Phi_{\mathbf{P}}(\tau) = \frac{\langle \mathbf{P}(t) \cdot \mathbf{P}(t + \tau) \rangle_t}{\langle \mathbf{P}(t) \cdot \mathbf{P}(t) \rangle_t}, \tag{6.A.2}$$

where we explicitly assume that $\langle \mathbf{P} \rangle = 0$ for $t \to \infty$. The first part of each trajectory in the canonical ensemble is not considered in the equilibrium average. We observe an exponential decay in all sampling algorithms and fit an exponential $\sim \exp(-\tau/\tau_\Phi)$ to extract the time constant τ_Φ and its

standard error (see also Ref. [28]). We naïvely parallelize the Metropolis and event-chain Monte Carlo algorithms by running 20 to 50 independent simulations (where the number of runs grows with N). For every τ, we then compute the median autocorrelation function $\Phi_{\mathbf{P}}(\tau)$ and estimate its error with the bootstrap method. These errors are then considered in the least-squares fit.

The units of the autocorrelation time τ_{Φ} for the different sampling algorithms are physical time for molecular dynamics, Monte-Carlo trials for the Metropolis algorithm, and continuous Monte-Carlo time for the event-chain Monte Carlo algorithms. As a first measure of efficiency of the different dynamics, we compare the autocorrelation distances d_{Φ} (see Fig. 6.3c). It gives the average cumulative distance moved by the atoms in each of the sampling schemes. Since only a single atom moves in our implementation of the event-chain Monte Carlo algorithm, it follows that $d_{\Phi}^{\mathrm{ECMC}} = \langle|\mathbf{v}_a|\rangle^{\mathrm{ECMC}} \tau_{\Phi}^{\mathrm{ECMC}}$, where $\langle|\mathbf{v}_a|\rangle^{\mathrm{ECMC}}$ is the average speed of the active atom during the simulation. For molecular dynamics, we can compute the average speed of the hydrogen and oxygen atoms $t \in \{\mathrm{H}, \mathrm{O}\}$ with the Maxwell–Boltzmann distribution:

$$\langle|\mathbf{v}_t|\rangle^{\mathrm{MD}} = \sqrt{\frac{8 k_{\mathrm{B}} T}{\pi m_t}}, \tag{6.A.3}$$

which yields $d_{\Phi}^{\mathrm{MD}} = N(2\langle|\mathbf{v}_{\mathrm{H}}|\rangle^{\mathrm{MD}} + \langle|\mathbf{v}_{\mathrm{O}}|\rangle^{\mathrm{MD}}) \tau_{\Phi}^{\mathrm{MD}}$. In the Metropolis algorithm, each trial samples a random displacement vector $\mathbf{d}_t = [\mathrm{ran}(-\delta_t, \delta_t), \mathrm{ran}(-\delta_t, \delta_t), \mathrm{ran}(-\delta_t, \delta_t)]^T$ for a random atom, where δ_t is tuned independently for the oxygen and hydrogen atoms $t \in \{\mathrm{H}, \mathrm{O}\}$ to accept the displacement with probability $p \approx 0.37$. We can compute the average length $\langle|\mathbf{d}_t|\rangle^{\mathrm{Met}}$ of the sampled displacement vector as

$$\langle|\mathbf{d}_t|\rangle^{\mathrm{Met}} = \frac{1}{8\delta_t^3} \int_{-\delta_t}^{\delta_t} \mathrm{d}x \int_{-\delta_t}^{\delta_t} \mathrm{d}y \int_{-\delta_t}^{\delta_t} \mathrm{d}z \sqrt{x^2 + y^2 + z^2} \approx 0.961\, \delta_t. \tag{6.A.4}$$

Neglecting correlations between rejection probability and the displacement for this first measure of efficiency, we get $d_{\Phi}^{\mathrm{Met}} = p\, (2\langle|\mathbf{d}_{\mathrm{H}}|\rangle^{\mathrm{Met}}/3 + \langle|\mathbf{d}_{\mathrm{O}}|\rangle^{\mathrm{Met}}/3)\, \tau_{\Phi}^{\mathrm{Met}}$.

Finally, the computer autocorrelation time of our generalized Newtonian event-chain Monte Carlo algorithm (see Fig. 6.4b) is obtained by combining the autocorrelation distance (see Fig. 6.3c), the event rates (see Fig. 6.3b), and the computer time per event in JeLLyFysh (see Fig. 6.3a). Similarly, we measure the computer time per time step of molecular dynamics in Lammps for different target accuracies of its particle–particle particle–mesh solver (see Fig. 6.4a). This is then combined with the autocorrelation time in the number of time steps.

6.B Supplementary Materials—Supplementary Text

In this supplementary text, we generalize Newtonian event-chain Monte Carlo to smooth factor potentials depending on an arbitrary number of atoms. We further prove that this algorithm satisfies the necessary global-balance condition. Finally, we report on our generation of Fibonacci vectors on the unit sphere.

6.B.1 Generalized Newtonian Event-Chain Monte Carlo

Newtonian event-chain Monte Carlo was originally formulated for the hard-sphere model that contains peculiar stepwise-changing two-body factor potentials [124]. Among various event-chain Monte Carlo variants, it was shown to escape faster from sparse hard-disk packings [P3] and to produce favorable rotation dynamics in tethered hard-disk dipoles [P2]. Its superiority is confirmed for the rotation dynamics of SPC/Fw water molecules in this paper, but requires its generalization to smooth factor potentials that depend on an arbitrary number of atom positions.

We follow the initial introduction of non-reversible event-chain Monte Carlo as a lifted continuous-time Markov process [121] (see also the formulation as a piecewise-deterministic Markov process in Ref. [137]). For Newtonian event-chain Monte Carlo for N_a three-dimensional atoms, the lifting framework [111, 112] (see also Refs [80] and [P1, Appendix A]) extends the physical all-atom configuration $\mathbf{x}$ to $(\mathbf{x}, \mathbf{v}, i)$, with auxiliary lifting variables that represent an all-atom velocity $\mathbf{v}$ and an activity label i, respectively. The physical sample space $\mathbf{x} \in \Omega$ is thus extended to the lifted sample space $(\mathbf{x}, \mathbf{v}, i) \in \widehat{\Omega} = \Omega \times \mathcal{V}^{3N_a}(E_{\text{kin}}) \times \mathcal{N}$, where $\mathcal{N} = \{1 \ldots N_a\}$ the set of atom indices, and $\mathcal{V}^{3N_a}(E_{\text{kin}}) = \{\mathbf{v} \in \mathbb{R}^{3N_a} : \mathbf{v}^T \cdot \mathbf{M} \cdot \mathbf{v} = 2E_{\text{kin}}\}$ with a positive-definite symmetric mass matrix $\mathbf{M}$ and total kinetic energy E_{kin}. The Markov process targets the lifted stationary distribution $\hat{\pi}(\mathbf{x}, \mathbf{v}, i) = \pi(\mathbf{x}) \times \mu_{\mathcal{V}}(\mathbf{v}) \times \mu_{\mathcal{N}}(i)$ that separates into the factorized Boltzmann distribution $\pi(\mathbf{x}) \propto \exp[-\beta U(\mathbf{x})] = \prod_M \exp[-\beta U_M(\mathbf{x})]$, and the uniform distributions $\mu_{\mathcal{V}}(\mathbf{v})$ on $\mathcal{V}^{3N_a}$, and $\mu_{\mathcal{N}}(i)$ on $\mathcal{N}$.

Given a lifted configuration $(\mathbf{x}(t_0), \mathbf{v}(t_0), i)$ at time t_0, Newtonian event-chain Monte Carlo continuously moves the single active atom i starting from its position $\mathbf{x}_i(t_0)$ with its constant velocity $\mathbf{v}_i = \mathbf{v}_i(t_0)$ up to an event at time $t_{\text{ev}} > t_0$, which interrupts the motion: $\mathbf{x}_i(t) = \mathbf{x}_i(t_0) + \mathbf{v}_i(t - t_0)$ for $t_0 <= t < t_{\text{ev}}$. The velocities $\mathbf{v}_j$ of all other atoms $j \neq i$ in the all-atom velocity $\mathbf{v}$ are, at this point, hypothetical, that is, mere labels. The time-dependent event rate $\lambda(t)$, that is, the probability density to interrupt the piecewise-deterministic motion of the active atom i, is given by a sum of factor event rates $\lambda_M(t)$:

$$\lambda(t) = \sum_M \lambda_M(t) = \sum_M \beta \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_i} U_M\left(\mathbf{x}(t)\right) \cdot \mathbf{v}_i\right]. \tag{6.B.5}$$

Every factor M can be considered as statistically independent. Each of them stochastically generates a candidate event time $t_{\text{ev},M}$ in an inhomogeneous Poisson process based on its factor event rate $\lambda_M(t) \geq 0$ (that is only nonzero for factor potentials U_M that are actually changed by the motion of the active atom):

$$\text{ran}_M(0, 1) = \exp\left[-\int_{t_0}^{t_{\text{ev},M}} \lambda_M(t)\, dt\right]. \tag{6.B.6}$$

Here, $\text{ran}_M(0, 1)$ is a uniformly distributed random number between 0 and 1 that is drawn separately for each factor M. In Eq. (6.B.6), the cumulative increments of the factor potential U_M under the motion of the active atom with its velocity $\mathbf{v}_i$ since t_0 are equal to an exponentially distributed random number with mean $1/\beta$ at the time $t_{\text{ev},M}$. If the computation of $t_{\text{ev},M}$ from Eq. (6.B.6) is tedious or even impossible, a Poisson thinning procedure may yield candidate event times [165].

At the event time $t_{\text{ev}} = \min_M t_{\text{ev},M}$, the motion of the active atom i is interrupted at the lifted configuration $(\mathbf{x} = \mathbf{x}(t_{\text{ev}}), \mathbf{v}, i)$. An event changes the lifting variables and sets the initial lifted configuration $(\mathbf{x}, \mathbf{v}', j)$ for the next piece in the piecewise-deterministic Markov process. The event is

realized by a unique event-factor $M_{ev} = \arg\min_M t_{ev,M}$. Generalized Newtonian event-chain Monte Carlo proposes an update of the all-atom velocity $\mathbf{v} \to \mathbf{v}'$ at time t_{ev} with a force kick in the direction of the gradient of the event-factor potential:

$$\mathbf{v}' = \mathbf{v} - 2 \frac{\mathbf{v}^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_{M_{ev}}}{(\boldsymbol{\nabla}_{\mathbf{x}} U_{M_{ev}})^T \cdot \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_{M_{ev}}} \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_{M_{ev}}. \tag{6.B.7}$$

This force kick leaves the kinetic energy $2E_{kin} = \mathbf{v}^T \cdot \mathbf{M} \cdot \mathbf{v}$ invariant and applying it twice yields the original $\mathbf{v}$ [116]. The event-factor potential only acts on a small subensemble $k \in M_{ev}$ of atoms, and thus only the velocities $\mathbf{v}_k$ of these contributing atoms are possibly modified. For the simple choice of an identity mass matrix $\mathbf{M} = \mathbf{I}$, Eq. (6.B.7) may be written as

$$\mathbf{v}'_k = \mathbf{v}_k - 2 \frac{\sum_{j \in M_{ev}} \mathbf{v}_j^T \cdot \boldsymbol{\nabla}_{\mathbf{x}_j} U_{M_{ev}}}{\sum_{j \in M_{ev}} \left| \boldsymbol{\nabla}_{\mathbf{x}_j} U_{M_{ev}} \right|^2} \boldsymbol{\nabla}_{\mathbf{x}_k} U_{M_{ev}}. \tag{6.B.8}$$

In Newtonian event-chain Monte Carlo, only a single active atom moves with its velocity at any time. At an event, it must choose the next active atom $k \in M_{ev}$ from the event factor. It also chooses to either change the velocity $\mathbf{v} \to \mathbf{v}'$ of all contributing atoms according to Eq. (6.B.7), or to keep the all-atom velocity constant $\mathbf{v} \to \mathbf{v}$. Generalized Newtonian event-chain Monte Carlo offers two schemes to treat events. The *Newtonian-general* scheme applies to general mass matrices $\mathbf{M}$ and factors that depend on an arbitrary number of atoms. The *Newtonian-pair* scheme only applies to distance-dependent pair potentials and identity mass matrix $\mathbf{M} = \mathbf{I}$.

In the Newtonian-general scheme, the probability for each possible choice $(k, \mathbf{v}')$ or $(k, \mathbf{v})$ of the lifting variables is given by $\max[0, -\boldsymbol{\nabla}_{\mathbf{x}_k} U_{M_{ev}} \cdot \mathbf{v}'_k]/C$ or $\max[0, -\boldsymbol{\nabla}_{\mathbf{x}_k} U_{M_{ev}} \cdot \mathbf{v}_k]/C$, respectively. Here, C is a common normalization factor. If the force kick is not applied and $\mathbf{v}$ stays constant, the active atom always changes because the factor event rate was positive at the time of the event [see Eqs (6.B.5) and (6.B.6)]. Likewise, if the event-factor potential only depends on a single atom (as for external potentials), the force kick is always applied.

In the Newtonian-pair scheme for distance-dependent pair potentials between two atoms a and b, $U_{M_{ev}} = U_{M_{ev}}(|\mathbf{x}_a - \mathbf{x}_b|)$ and an identity mass matrix $\mathbf{M} = \mathbf{I}$, we alternatively exploit the inherent translational invariance to treat the event deterministically. We always apply the force kick in Eq. (6.B.8) to modify both $\mathbf{v}_a$ and $\mathbf{v}_b$, and always change the active atom within the event factor. In this case, Eq. (6.B.8) is equivalent to an elastic Newtonian collision between two particles of equal mass, and this alternative scheme recovers the original formulation of Newtonian event-chain Monte Carlo for the hard-sphere model [124].

Samples of all-atom configurations $\mathbf{x}$ that can be used to compute observables are taken at periodic time intervals [124]. Furthermore, Newtonian event-chain Monte Carlo requires resamplings of the all-atom velocity $\mathbf{v}$ and the active atom i in periodic time intervals τ_{chain} to become irreducible [P2]. We sample $\mathbf{v}$ and i from their respective stationary distributions $\mu_{\mathcal{V}}(\mathbf{v})$ and $\mu_{\mathcal{N}}(i)$. For $\mathbf{v}$, this is equivalent to sampling a multivariate Gaussian distribution with mean $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} = \mathbf{M}^{-1}$, followed by corrections of the total velocity $\mathbf{1}^T \cdot \mathbf{v} \to 0$ and $\mathbf{v}^T \cdot \mathbf{M} \cdot \mathbf{v} \to 2E_{kin}$. The activity label i is sampled uniformly from the the set $\mathcal{N} = \{1 \dots N_a\}$.

6.B.2 Proof of Correctness of Generalized Newtonian Event-Chain Monte Carlo

In order to prove the correctness of the Newtonian event-chain Monte Carlo algorithm for general smooth interactions, we must show that the global balance condition is satisfied. The total probability flow into any lifted configuration $(\mathbf{x}, \mathbf{v}, i) \in \widehat{\Omega}$ consists of a physical flow $\mathcal{F}^{\text{phys}}(\mathbf{x}, \mathbf{v}, i)$ and a lifting flow $\mathcal{F}^{\text{lift}}(\mathbf{x}, \mathbf{v}, i)$ and must equal its statistical weight $\hat{\pi}(\mathbf{x}, \mathbf{v}, i)$:

$$\hat{\pi}(\mathbf{x}, \mathbf{v}, i) = \mathcal{F}^{\text{phys}}(\mathbf{x}, \mathbf{v}, i) + \mathcal{F}^{\text{lift}}(\mathbf{x}, \mathbf{v}, i). \tag{6.B.9}$$

The physical flow into $(\mathbf{x}, \mathbf{v}, i)$ stems from the continuous movement of the active atom i that was not interrupted by an event with the event rate given in Eq. (6.B.5). With $\mathbf{x}' = (\mathbf{x}_1, \dots, \mathbf{x}_i - \mathbf{v}_i \mathrm{d}t, \dots, \mathbf{x}_{N_a})$, we get

$$
\begin{aligned}
\mathcal{F}^{\text{phys}}(\mathbf{x}, \mathbf{v}, i) &= \hat{\pi}(\mathbf{x}', \mathbf{v}, i) \left\{ 1 - \beta \sum_M \max \left[0, \boldsymbol{\nabla}_{\mathbf{x}_i'} U_M \cdot \mathbf{v}_i \right] \right\} \\
&= \hat{\pi}(\mathbf{x}, \mathbf{v}, i) \left\{ 1 - \beta \sum_M \max \left[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M \cdot \mathbf{v}_i \right] \right\} \\
&=: \hat{\pi}(\mathbf{x}, \mathbf{v}, i) + \sum_M \mathcal{F}_M^{\text{phys}}(\mathbf{x}, \mathbf{v}, i),
\end{aligned}
\tag{6.B.10}
$$

where the second line uses the detailed-balance property of the factorized Metropolis filter that yields the event rates in Eq. (6.B.5) and allows to treat all factors as statistically independent [121]. The lifting flow into $(\mathbf{x}, \mathbf{v}, i)$ stems from interrupted motions of lifted configurations $(\mathbf{x}, \mathbf{v}', j)$ with different lifting variables $\mathbf{v}'$ and j but the same configuration $\mathbf{x}$:

$$
\begin{aligned}
\mathcal{F}^{\text{lift}}(\mathbf{x}, \mathbf{v}, i) &= \beta \sum_M \sum_{j \in \mathcal{N}} \int_{\mathcal{V}^{3N_a}(E_{\text{kin}})} \mathrm{d}^{3N_a} \mathbf{v}' \, \hat{\pi}(\mathbf{x}, \mathbf{v}', j) \max \left[0, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M \cdot \mathbf{v}_j' \right] p_{(\mathbf{v}', j),(\mathbf{v}, i)}^M \\
&=: \sum_M \mathcal{F}_M^{\text{lift}}(\mathbf{x}, \mathbf{v}, i).
\end{aligned}
\tag{6.B.11}
$$

Here, $p_{(\mathbf{v}', j),(\mathbf{v}, i)}$ is the probability to change the lifting variables $(\mathbf{v}', j)$ to $(\mathbf{v}, i)$. For every factor M, the lifting flow $\mathcal{F}_M^{\text{lift}}(\mathbf{x}, \mathbf{v}, i)$ in Eq. (6.B.11) exactly cancels the physical flow $\mathcal{F}_M^{\text{phys}}(\mathbf{x}, \mathbf{v}, i)$ in Eq. (6.B.10) so that the global-balance condition in Eq. (6.B.9) is satisfied. A trivial (inefficient) solution would be to simply invert the velocity of the active atom i in an event while keeping it active: $p_{(\mathbf{v}', j),(\mathbf{v}, i)}^M = \delta_{ij} \, \delta^{(3)}(\mathbf{v}_i - \mathbf{v}_i')$.

We presented two schemes for the update of the lifting variables in an event of generalized Newtonian event-chain Monte Carlo. The first stochastic Newtonian-general scheme applies for a general number of atoms on the event factor. The second deterministic Newtonian-pair scheme exploits the translational

invariance of distance-dependent pair potentials. For the Newtonian-general scheme, we get

$$
\begin{aligned}
p^M_{(\mathbf{v}',j),(\mathbf{v},i)} &= \frac{\max\left[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M \cdot \mathbf{v}_i\right]}{C} \left[\delta^{(3N_a)}\left(\mathbf{v} - \mathbf{v}' + 2\frac{(\mathbf{v}')^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}{(\boldsymbol{\nabla}_{\mathbf{x}} U_M)^T \cdot \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M} \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M\right)\right. \\
&\qquad\qquad \left. + \delta^{(3N_a)}\left(\mathbf{v} - \mathbf{v}'\right)\right] \\
&= \frac{\max\left[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M \cdot \mathbf{v}_i\right]}{C} \left[\delta^{(3N_a)}\left(\mathbf{v}' - \mathbf{v} + 2\frac{\mathbf{v}^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}{(\boldsymbol{\nabla}_{\mathbf{x}} U_M)^T \cdot \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M} \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M\right)\right. \\
&\qquad\qquad \left. + \delta^{(3N_a)}\left(\mathbf{v}' - \mathbf{v}\right)\right],
\end{aligned}
$$

(6.B.12)

where the first term applies the force kick, while the second does not. From the necessary condition

$$
1 = \sum_{k \in \mathcal{N}} \int_{\mathcal{V}^{3N_a}(E_{\text{kin}})} \mathrm{d}^{3N_a} \mathbf{v} \, p^M_{(\mathbf{v}',j),(\mathbf{v},k)},
$$

(6.B.13)

it follows that

$$
\begin{aligned}
C &= \sum_{k \in \mathcal{N}} \left\{ \max\left[0, -\boldsymbol{\nabla}_{\mathbf{x}_k} U_M \cdot \left(\mathbf{v}'_k - 2\frac{(\mathbf{v}')^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}{(\boldsymbol{\nabla}_{\mathbf{x}} U_M)^T \cdot \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M} \left(\mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M\right)_k\right)\right]\right. \\
&\qquad\qquad \left. + \max\left[0, -\boldsymbol{\nabla}_{\mathbf{x}_k} U_M \cdot \mathbf{v}'_k\right]\right\} \\
&= \sum_{k \in \mathcal{N}} \left\{ \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_k} U_M \cdot \left(\mathbf{v}'_k - 2\frac{(\mathbf{v}')^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}{(\boldsymbol{\nabla}_{\mathbf{x}} U_M)^T \cdot \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M} \left(\mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M\right)_k\right)\right]\right. \\
&\qquad\qquad \left. + \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_k} U_M \cdot \mathbf{v}'_k\right]\right\}.
\end{aligned}
$$

(6.B.14)

Here, the second rewriting follows from expressing the trivial equality

$$
\mathbf{v}^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M + \left(\mathbf{v} - 2\frac{\mathbf{v}^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}{(\boldsymbol{\nabla}_{\mathbf{x}} U_M)^T \cdot \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M} \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M\right)^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M = 0
$$

(6.B.15)

as a sum over $k \in \mathcal{N}$ and from grouping the positive and negative terms separately. The normalization factor C in $p^M_{(\mathbf{v}',j),(\mathbf{v},i)}$ is the same for every possible next active atom i. Because of the $\delta^{(3N_a)}$-functions in Eq. (6.B.12), the normalization factor C can also be written in terms of the velocity $\mathbf{v}$ after the force

kick by replacing $\mathbf{v}' \to \mathbf{v}$ in Eq. (6.B.14). This yields

$$
\begin{aligned}
\mathcal{F}_M^{\text{lift}}(\mathbf{x}, \mathbf{v}, i) = {}& \beta\, \hat{\pi}(\mathbf{x}, \mathbf{v}, i)\, \max\left[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M \cdot \mathbf{v}_i\right] \frac{1}{C} \sum_{j \in \mathcal{N}} \Bigg\{ \\
& \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M \cdot \left(\mathbf{v}_j - 2\frac{\mathbf{v}^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}{(\boldsymbol{\nabla}_{\mathbf{x}} U_M)^T \cdot \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}\left(\mathbf{M}^{-1} \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M\right)_j\right)\right] \\
& + \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M \cdot \mathbf{v}_j\right] \Bigg\} \\
= {}& \beta\, \hat{\pi}(\mathbf{x}, \mathbf{v}, i)\, \max\left[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M \cdot \mathbf{v}_i\right] \\
= {}& -\mathcal{F}_M^{\text{phys}}(\mathbf{x}, \mathbf{v}, i),
\end{aligned}
\tag{6.B.16}
$$

where we used that the velocity $\mathbf{v}$ and activity label i are uniformly distributed over their sample spaces $\mathcal{V}^{3N_a}(E_{\text{kin}})$ and $\mathcal{N}$, respectively. Equation (6.B.16) concludes the proof that the presented Newtonian-general scheme for the update of the lifting variables in an event by a general factor potential satisfies the global-balance condition of Eq. (6.B.9). The reflective [115] and forward [150] event-chain Monte Carlo variants can likewise be generalized to factor potentials depending on an arbitrary number of atoms.

For the Newtonian-pair scheme for distance-dependent pair potentials and identity mass matrix $\mathbf{M} = \mathbf{I}$, we get

$$
\begin{aligned}
p_{(\mathbf{v}', j),(\mathbf{v}, i)}^M &=
\begin{cases}
\delta^{(3N_a)}\left(\mathbf{v} - \mathbf{v}' + 2\frac{(\mathbf{v}')^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}{(\boldsymbol{\nabla}_{\mathbf{x}} U_M)^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}\, \boldsymbol{\nabla}_{\mathbf{x}} U_M\right) & \text{if } i, j \in M \text{ and } i \neq j, \\
0 & \text{otherwise},
\end{cases} \\
&=
\begin{cases}
\delta^{(3N_a)}\left(\mathbf{v}' - \mathbf{v} + 2\frac{\mathbf{v}^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}{(\boldsymbol{\nabla}_{\mathbf{x}} U_M)^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}\, \boldsymbol{\nabla}_{\mathbf{x}} U_M\right) & \text{if } i, j \in M \text{ and } i \neq j, \\
0 & \text{otherwise}.
\end{cases}
\end{aligned}
\tag{6.B.17}
$$

With $i, j \in M$ and $i \neq j$, this yields

$$
\begin{aligned}
\mathcal{F}_M^{\text{lift}}(\mathbf{x}, \mathbf{v}, i) &= \beta\, \hat{\pi}(\mathbf{x}, \mathbf{v}, i)\, \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M \cdot \left(\mathbf{v}_j - 2\frac{\mathbf{v}^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}{(\boldsymbol{\nabla}_{\mathbf{x}} U_M)^T \cdot \boldsymbol{\nabla}_{\mathbf{x}} U_M}\, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M\right)\right] \\
&= \beta\, \hat{\pi}(\mathbf{x}, \mathbf{v}, i)\, \max\left[0, \boldsymbol{\nabla}_{\mathbf{x}_j} U_M \cdot \mathbf{v}_j - 2\frac{\mathbf{v}_i^T \cdot \boldsymbol{\nabla}_{\mathbf{x}_i} U_M + \mathbf{v}_j^T \cdot \boldsymbol{\nabla}_{\mathbf{x}_j} U_M}{\left|\boldsymbol{\nabla}_{\mathbf{x}_i} U_M\right|^2 + \left|\boldsymbol{\nabla}_{\mathbf{x}_j} U_M\right|^2}\left|\boldsymbol{\nabla}_{\mathbf{x}_j} U_M\right|^2\right] \\
&= \beta\, \hat{\pi}(\mathbf{x}, \mathbf{v}, i)\, \max\left[0, -\boldsymbol{\nabla}_{\mathbf{x}_i} U_M \cdot \mathbf{v}_i\right] \\
&= -\mathcal{F}_M^{\text{phys}}(\mathbf{x}, \mathbf{v}, i).
\end{aligned}
\tag{6.B.18}
$$

Here, we again used the uniform distributions of $\mathbf{v}$ and i and the translational invariance $\boldsymbol{\nabla}_{\mathbf{x}_i} U_M = -\boldsymbol{\nabla}_{\mathbf{x}_j} U_M$. Thus, also the Newtonian-pair scheme satisfies the global-balance condition of Eq. (6.B.9). We note that it always applies the force kick.

6.B.3 Fibonacci Vectors

The cell-veto algorithm was previously only used for the straight variant of event-chain Monte Carlo [P6, 146, 148]. Its finite set of possible velocities $\mathcal{D}$ of the single active atom a allows to set up a Walker table for every velocity $\mathbf{v}_a \in \mathcal{D}$. For instance, for factor pair potentials between the active atom a and another atom b at positions $\mathbf{x}_a$ and $\mathbf{x}_b$, $U_M(\mathbf{x}) = U_M(\mathbf{x}_a, \mathbf{x}_b)$, every cell bound $q_M^{\mathrm{cell}}(C_a, C_b, \mathbf{v}_a)$ for the pair of cells C_a and C_b in the Walker table may be written as

$$q_M^{\mathrm{cell}}(C_a, C_b, \mathbf{v}_a) = \max_{\mathbf{x}_a \in C_a, \mathbf{x}_b \in C_b} \beta \max \left[0, \boldsymbol{\nabla}_{\mathbf{x}_a} U_M(\mathbf{x}_a, \mathbf{x}_b) \cdot \mathbf{v}_a \right]. \tag{6.B.19}$$

In principle, a Walker table must be precomputed for every velocity $\mathbf{v}_a \in \mathcal{D}$ and for every possible cell C_a of the active atom. Symmetries, such as a translational invariance of the factor potential, may heavily reduce the necessary number of Walker tables [146]. During the straight event-chain Monte Carlo simulation, the relevant Walker table is determined by the cell of the currently active atom and its velocity. The generalization of Eq. (6.B.19) to more complex factors is straightforward. For example, in order to treat the Coulomb interaction between two water molecules in the SPC/Fw water model, the position and orientation of the water molecule containing the active atom may be varied in C_a, while the position and orientation of the other molecule is varied in C_b. From a practical point of view, the cell bounds $q_M^{\mathrm{cell}}(C_a, C_b)$ are usually not computed exactly but rather approximated, e.g., by considering a finite set of positions $\mathbf{x}_a \in C_a$ and $\mathbf{x}_b \in C_b$, and including a corrective multiplicative prefactor [P6]. As long as the approximated cell bound satisfies $\tilde{q}_M^{\mathrm{cell}}(C_a, C_b, \mathbf{v}_a) \geq q_M^{\mathrm{cell}}(C_a, C_b, \mathbf{v}_a)$, a Poisson thinning procedure [165] corrects *any* overestimate. The quality of $\tilde{q}_M^{\mathrm{cell}}(C_a, C_b, \mathbf{v}_a)$, however, does influence the performance because higher cell bounds in the cell-veto algorithm yield more events per unit distance that have to be confirmed by computing the actual event rate.

The inherent discretization of continuous space in Eq. (6.B.19) can be translated to a continuous velocity space, as it appears, e.g., in Newtonian event-chain Monte Carlo. Consider a finite number of unit vectors $\hat{\mathbf{v}}_d$ on the two-dimensional unit sphere, and let $\mathcal{V}_d$ be the associated Voronoi cells under some distance function. We can then formally compute the cell bounds for every Voronoi cell as

$$q_M^{\mathrm{cell}}(C_a, C_b, \mathcal{V}_a) = \max_{\mathbf{x}_a \in C_a, \mathbf{x}_b \in C_b, \mathbf{v}_a \in \mathcal{V}_a} \beta \max \left[0, \boldsymbol{\nabla}_{\mathbf{x}_a} U_M(\mathbf{x}_a, \mathbf{x}_b) \cdot \mathbf{v}_a \right]. \tag{6.B.20}$$

The results are used to precompute Walker tables for every Voronoi cell $\mathcal{V}_a$ and for every cell C_a, where symmetries may again heavily reduce the number of actually necessary tables. During the Newtonian event-chain Monte Carlo simulation, the relevant Walker table is determined by finding the cell of the currently active atom and the Voronoi cell of its normalized velocity $\mathbf{v}_a/|\mathbf{v}_a|$. We can then correct all cell bounds in the relevant Walker table to the actual speed of the active atom by multiplying them with $|\mathbf{v}_a|$.

In this paper, as proof of concept, we map a generalized Fibonacci lattice onto a two-dimensional unit sphere by the Lambert cylindrical equal-area projection to generate D unit vectors $\hat{\mathbf{v}}_d^{\mathrm{fib}}$:

$$\hat{\mathbf{v}}_d^{\mathrm{fib}} = (\cos \phi_d \sin \theta_d, \sin \phi_d \sin \theta_d, \cos \theta_d)^T, \tag{6.B.21}$$

where $0 \leq d < D$ and

$$
\begin{aligned}
\phi_d &= \frac{2\pi d}{\varphi}, \\
\theta_d &= \arccos\left(1 - \frac{2(d + \varepsilon)}{D - 1 + 2\varepsilon}\right),
\end{aligned}
\tag{6.B.22}
$$

with the golden ratio $\varphi = (1 + \sqrt{5})/2$ [216–218]. The empirical choice of the parameter $\varepsilon = 0.36$ optimizes the average nearest-neighbor distance of the Fibonacci vectors $\hat{\mathbf{v}}_d$. As a distance function to construct the Voronoi cells $\mathcal{V}_d$, we use the quick-to-evaluate cosine distance. Equations (6.B.21) and (6.B.22) efficiently generate an arbitrary number of Fibonacci vectors with a nearly uniform distribution. However, there is no efficient inverse mapping from a general velocity $\mathbf{v}_a$ to the closest Fibonacci vector $\hat{\mathbf{v}}_d$. This is not a problem in this paper because we choose $D = 10$ small. Then, $\hat{\mathbf{v}}_d$ can be found by brute force. If considerably larger values of D become necessary, other point configurations on the two-dimensional unit sphere may be considered [218, 219]. For the uniform SPC/Fw water systems in this paper, the Walker tables do not strongly depend on the respective Fibonacci vector. The proper discretization of velocity space will, however, impact the performance of Newtonian event-chain Monte Carlo in nonuniform systems.

Conclusion

The four research papers of this doctoral thesis are clearly connected in one way or another. On the one hand, they all considered theoretical foundations of non-reversible Markov-chain Monte Carlo (MCMC) algorithms. On the other hand, they all pursued the overarching objective of this doctoral thesis—the introduction of a competitive, rigorous paradigm for molecular simulation that is based on non-reversible Markov chains.

For the moment, let us recapitulate the novel scientific insights resulting from this doctoral thesis focusing on the theoretical point of view. Chapter 3 (see Ref. [P1]) introduced non-reversible direction sweeps. These are reminiscent of the widely-used sequential sweeps in the reversible Metropolis algorithm that propose a local Monte Carlo move for every interaction site in succession. In a direction sweep, the local move is proposed along a direction that is drawn from a given direction set in succession. For a single tethered hard-disk dipole, it was shown analytically that the introduction of a non-reversible direction sweep into a local reversible MCMC algorithm fundamentally changes its dynamics. This is equivalent to the elementary difference between equilibrium and non-equilibrium statistical physics. Numerical computations of mixing times showed that small direction sets should be avoided. In comparison to a completely random choice of the direction, a direction sweep over large direction sets somewhat reduced mixing times, which is similar to the slight speedup that can be typically obtained with sequential proposal sweeps. Chapter 4 (see Ref. [P2]) considered large systems of the hard-disk dipoles. For the non-reversible straight event-chain Monte Carlo (ECMC) algorithm, we found that a direction sweep is indeed slightly faster for decorrelating the polarization (i.e., for rotating the individual dipoles) than choosing the direction at random. It would be interesting to see in the future whether direction sweeps over large direction sets could also maintain that advantage in practical applications of the Metropolis algorithm.

From a theoretical point of view, Chapter 4 (see Ref. [P2]) also studied subtleties concerning the irreducibility of different ECMC variants for a single tethered hard-disk dipole. Reflective and Newtonian ECMC change the velocity of the active disk in every event. Still, it was shown that they strictly require random resamplings of the active disk and its velocity after the end of an event chain in periodic time intervals. Straight ECMC transfers the velocity between disks in an event. It may

consider deterministic changes of the velocity after each event chain, for instance, by using a direction sweep, but it strictly requires a random resampling of the active disk. Although no differences in static observables were observed with or without resamplings in large systems of the hard-disk dipoles, simply including the appropriate resamplings appears more rigorous. Event-based randomness as in forward ECMC can resolve these issues more generally and does not require any resamplings.

Chapter 5 (see Ref. [P3]) examined tightly confined ε-relaxed Böröczky configurations of hard disks in a periodic box that are constructed from locally stable sparse Böröczky packings. Escape times from ε-relaxed Böröczky configurations were proposed for the analysis and benchmarking of local MCMC algorithms. They may well model the probability flow across a bottleneck in the Markov chains. Within the family of ECMC algorithms, a scaling theory resulted in widely different scalings of the escape times in the relaxation parameter ε. The significant power-law-to-logarithm speedup of reflective, forward, and Newtonian ECMC with respect to straight ECMC is rooted in their lack of an intrinsic scale and their event-driven implementation. Locally stable sparse Böröczky packings were constructed for continuous ranges of densities and may appear as backbones in hard-disk systems at higher densities. The space of the derived ε-relaxed Böröczky configurations is thus a finite portion of the configuration space that is practically excluded from any local MCMC simulation of the hard-disk system. Since, more generally, one cannot prove the irreducibility of local MCMC algorithms in the hard-disk system in a (periodic) box, the question of how to remedy the undoubted usefulness of MCMC simulations was discussed. The open-source software package BigBoro provides an easy access to the construction of Böröczky packings that, as advocated, can be subsequently used to benchmark local MCMC algorithms. For example, they can be directly used to analyze ECMC with the arbitrary-precision implementations of various ECMC variants that are also included in BigBoro.

Again only considering theoretical aspects of non-reversible MCMC algorithms, Chapter 6 (see Ref. [P4]) generalized Newtonian ECMC from the stepwise-changing translationally-invariant pair interactions of the hard-disk system to the smooth interactions between an arbitrary number of interaction sites of molecular systems. In every event, generalized Newtonian ECMC decides stochastically about the application of a force kick that may further consider general mass matrices of the interaction sites. Naturally, it was proven that generalized Newtonian ECMC satisfies the necessary global-balance condition. In addition, it was shown that the cell-veto algorithm can be used for a continuous space of possible velocities of the active interaction site in ECMC. This was achieved by discretizing the continuous velocity space into Voronoi cells of a finite number of Fibonacci vectors for the estimation of the cell bounds in the cell-veto algorithm. This generalization still treats long-range interactions rigorously exact with an $O(1)$ computational complexity per event.

This doctoral thesis ultimately fulfills its overarching objective of the introduction of a competitive, rigorous paradigm for molecular simulation. The framework of non-reversible ECMC or, more generally, the framework of piecewise-deterministic Markov processes, rigorously samples the canonical ensemble of molecular systems. The Boltzmann distribution is expressed as a product of statistically independent factor distributions that each depend on their own factor potential. Independently of the system size, a factor potential only involves the interactions between a small number of interaction sites. Every long-range-interacting factor potential between such a small number of interaction sites can be efficiently treated with constant computational cost and without

approximations. The cell-veto algorithm further bundles the long-range-interacting factors together. Instead of treating every factor individually, a discrete probability distribution is sampled with constant computational complexity. This provides the only factor from the bundled set whose potential is treated with the actual positions of its interaction sites. In order to advance N long-range-interacting sites by a constant distance, the non-reversible MCMC approach to molecular simulation only requires an $O(N \log N)$ computational complexity. The open-source JᴇLLʏFʏsʜ application implements ECMC as a specific example of the introduced paradigm for molecular simulation based on non-reversible Markov chains. The simulation of the thermodynamic equilibrium of long-range-interacting SPC/Fw water molecules in a periodic cubic box with straight ECMC in JᴇLLʏFʏsʜ-Version1.0, however, showed an unfavorable dynamical arrest of the polarization. Hence, the competitiveness of the novel paradigm for molecular simulation could not be proven. Gaining an understanding and finding a solution for this problem was an essential motive underlying a great part of this doctoral thesis.

Now, let us recapitulate the novel scientific insights resulting from this doctoral thesis in terms of the search for a proof that the rigorous paradigm for molecular simulation based on non-reversible Markov chain is competitive. Chapter 3 (see Ref. [P1]) considered the mixing time of a single tethered hard-disk dipole and found that the small number of possible velocities in the straight ECMC variant of JᴇLLʏFʏsʜ-Version1.0 was particularly ill-suited for inducing rotations of dipoles or molecules. In large systems of tethered hard-disk dipoles, Chapter 4 (see Ref. [P2]) found that straight, reflective, and forward ECMC show similar performances for the decorrelation of the polarization at high densities. Newtonian ECMC, however, stood out and gave a considerable speedup with respect to straight ECMC. Chapter 5 (see Ref. [P3]) witnessed widely different behaviors of different ECMC variants in their escape times from tightly confined ε-relaxed Böröczky configurations. Here, straight ECMC suffers from the explicit choice of an intrinsic scale by choosing the length of its event chains. Reflective, forward, and Newtonian ECMC become fastest in the limit of infinitely long event chains. This was also found for the decorrelation of the polarization in the tethered hard-disk dipole system in Chapter 4 which, from a practical point of view, conveniently removes the need for fine-tuning the chain time. After this systematic evaluation of the manifold of ECMC variants, it was clear that the straight ECMC variant in JᴇLLʏFʏsʜ-Version1.0 was the wrong choice for competitive simulations of the SPC/Fw water model.

Chapter 6 reconsidered the problem of sampling the canonical ensemble of long-range-interacting SPC/Fw water molecules in a periodic cubic box. Now, JᴇLLʏFʏsʜ-Version2.0 that implements generalized Newtonian ECMC was used. The culmination of this doctoral thesis showed that generalized Newtonian ECMC yields an order-of-magnitude speedup over straight ECMC in the characteristic distance that every interaction site has to travel to decorrelate the polarization. Hence, JᴇLLʏFʏsʜ-Version2.0 solves the problem of the dynamical arrest of the polarization in JᴇLLʏFʏsʜ-Version1.0 and allows to consider much larger system sizes. Furthermore, a direct comparison to molecular dynamics (MD) was carried out. MD is nowadays the prevalent method of choice in long-range-interacting molecular systems. The computational complexity of MD to advance N long-range-interacting sites by a constant distance when using widespread fast mesh-based Ewald methods is $O(N \log N)$, as in ECMC. In contrast to ECMC, however, MD only computes the long-range interactions up to a given target accuracy. Increasing the desired accuracy also increases the prefactor of the $O(N \log N)$ scaling. Therefore, it was numerically shown that generalized Newtonian ECMC

in JELLYFYSH-Version2.0, whose treatment of long-range interactions is tuned to machine precision, reaches a break-even point with MD in LAMMPS at a high precision that is below machine precision. This concluded the first practical proof of the competitiveness of the novel rigorous paradigm for molecular simulation based on non-reversible Markov chains in this doctoral thesis.

7.1 Discussion

The two corner stones of molecular simulation have always been MD and MCMC. MCMC algorithms are designed from the beginning to sample a given probability distribution such as the Boltzmann distribution of the canonical ensemble (see Section 1.3). By introducing thermostats that mimic the effects of a thermal reservoir, it was also understood that MD can achieve the same goal (see Section 1.2). Historically, MCMC algorithms for molecular simulations were for a long time restricted to local reversible MCMC algorithms with slow diffusive dynamics and an inefficient treatment of long-range interactions (see Section 1.3.1). This resulted in the dominance of MD in the field of molecular simulations that one observes today, especially when long-range interactions are present.

Non-reversible MCMC algorithms such as ECMC are fundamentally different from their reversible counterparts (see Section 1.3.2). First, their non-reversibility makes them capable of moving ballistically instead of diffusively. Even when they reached their stationary distribution, finite probability flows effectively induce non-equilibrium dynamics. Second, the factorization of the Boltzmann distribution removes global computations of the potential-energy change (or, analogously, the force in MD) that depend on the entire configuration. Instead, factor potentials that only depend on a limited number of interaction sites are evaluated. This enables a simple and exact treatment of long-range interactions in their respective factors and sidesteps the necessity for limited-precision computations of potential energies (or forces). Third, the cell-veto algorithm allows to achieve a competitive computationally complexity by treating the large set of long-range interacting factor potentials at once.

The rigorous paradigm for molecular simulations of the thermodynamic equilibrium based on non-reversible MCMC algorithms has to compete with the thermostatted MD approach. MD controls its approximations and unphysical artifacts algorithmically (see Section 1.2). Here, the choice of the accuracy of an MD simulation, which, for instance, follows from the size of the discrete time steps or an inaccurate treatment of long-range interactions, attempts a balancing act between reducing computational cost and avoiding unphysical results. In critical applications that have to strictly distinguish between unphysical results arising from artifacts and approximations of MD, on the one hand, and between predictions of real physical phenomena from the underlying model, on the other hand, the rigorous paradigm of this doctoral thesis may act as a gold standard. Because it is exact from the beginning, it is capable of filtering out the unwanted unphysical results. Here, one should note that, naturally, every molecular-mechanics model is an approximation of the physical reality in itself. The choice of a particular model depends on the specific research question. Therefore, any numerical result from a molecular simulation of the model will not perfectly agree with experiments. Still, the predictions of any model should be fixed once its parameterization is established. They should, in particular, not change when different computational methods are used. Rigorous molecular simulation based on non-reversible MCMC algorithms may yield definite results for any molecular-mechanics model.

The rigorous approach as implemented in JᴇLLʏFʏsʜ-Version2.0 reaches a break-even point with a long-developed MD software at high precision. Still, simulations of JᴇLLʏFʏsʜ require considerably more time than practical MD simulations that are never carried out at such high precisions. In contrast to JᴇLLʏFʏsʜ, however, MD is not designed for an exact treatment of long-range interactions from the very beginning. JᴇLLʏFʏsʜ should be viewed as a first demonstration software. If the same amount of sheer work force that optimized MD software packages over decades would be used for JᴇLLʏFʏsʜ, a large reduction of the break-even point is certainly possible. Furthermore, the break point was defined based on the decorrelation of a local observable, the polarization. In statistical physics, the largest speedups by introducing non-reversibility were obtained for non-local observables (as, e.g., in the hard-disk system [142]). Whether the dynamics of large-scale hydrodynamic modes, macroscopic conformational changes, and order parameters in molecular simulations may similarly benefit, remains to be seen.

More generally, generalized Newtonian ECMC as implemented in JᴇLLʏFʏsʜ-Version2.0 is only a first step from an algorithmic point of view. The introduced paradigm includes the entire framework of ECMC or, more generally, piecewise-deterministic Markov processes, which is far from fully explored. As this doctoral thesis showed, different ECMC variants can show widely different behaviors in different scenarios. This underlines one of the greatest strengths of MCMC algorithms in contrast to MD: They are not restricted to real physical dynamics but may instead choose from a wide range of different dynamics to sample the Boltzmann distribution of the canonical ensemble in the most efficient way. The degrees of freedom in choosing the most appropriate dynamics are even further increased by a transition from traditional reversible to non-reversible Markov chains that removes the reliance on the detailed-balance condition.

As shown by the dynamical arrest of the polarization in the SPC/Fw water model in straight ECMC, the huge amount of possibilities in non-reversible MCMC algorithms also results in the risk of finding a particularly bad choice for the molecular system at hand. Generalized Newtonian ECMC moves closer to MD in its treatment of events and leads to comparable dynamics as MD. It therefore appears as a first default choice for rigorous molecular simulation with well-working dynamics. In the next step, however, it is essential to find non-reversible rejection-free MCMC algorithms that grow beyond the real physical dynamics of MD. This is a fundamental difference to Hybrid (or Hamiltonian) Monte Carlo that uses short MD simulations to propose global moves in the Metropolis–Hastings algorithm [220, 221]. Hybrid Monte Carlo is successfully used for molecular simulations [11, 12, 222]. However, while this procedure removes the need for thermostats, it is subject to the dynamics of MD for the exploration of configuration space. In the context of Bayesian statistics, it was already shown that non-reversible rejection-free MCMC algorithms outperform Hybrid Monte Carlo in different scenarios (see, e.g., Refs [138, 140, 150]). From a practical point of view, these favorable results also require a much smaller amount of fine tuning than necessary in Hybrid Monte Carlo which has to optimize its acceptance rate. Finally, as for the local reversible MCMC algorithms, the computation of the acceptance probability of the proposed global move in Hybrid Monte Carlo involves inaccurate global computations of long-range interactions. This conflicts with the fundamental motivation of the paradigm of this doctoral thesis that treats molecular systems rigorously exact.

In order to grow beyond the real physical dynamics of MD, one may, first, choose different factorizations of the Boltzmann distribution. Second, one may find the optimal non-reversible MCMC

algorithm for the specific problem at hand. Here, the possible range of non-reversible schemes remains largely unexplored. For instance, although explicit continuous-time rotations of molecules are currently not possible and appear hard to implement under the presence of long-range interactions in periodic boundary conditions, it was recently understood that it is possible to move the interaction sites on curved paths instead of straight-line trajectories between events [211, 223]. This may well enhance rotation dynamics in molecular systems. Third, it is possible to change the algorithm within a molecular simulation. For example, one could periodically switch between generalized Newtonian ECMC and straight ECMC that may each speedup the decorrelation of different observables. Fourth, one may switch between motions of single interaction sites to motions of whole molecules (or any other subset of interaction sites) during the simulation [P6]. Fifth, as this doctoral thesis showed (see Chapter 6), the factorization of the Boltzmann distribution does not penalize the dynamics of generalized Newtonian ECMC for the case of the SPC/Fw water model. If the event rates of ECMC become too high in other systems because factors cannot compensate each other, factor fields may be a unique solution [80]. They add an invariant potential-energy term whose subsequent breakup into factors allows to decrease event rates. Also, they can profoundly modify the non-reversible dynamics. In the fluid phase of the hard-disk system, they reduced the dynamical scaling exponent to the optimal value for local dynamics which is, in particular, superior to MD [92].

Eventually, the rigorous paradigm for molecular simulation based on non-reversible MCMC algorithms also needs to be systematically parallelized. In MD, the most time-consuming step, the calculation of the forces, is inherently parallel by splitting both loops over all interaction sites and over their interaction partners. This is even possible for the long-range Coulomb potential. If it is treated by a mesh-based Ewald method, the short-range contribution in real space can be treated in a domain decomposition and the discretized grid in reciprocal space can also be distributed over several processors. Such a highly parallel computation then advances all interaction sites at once [224]. First of all, the non-reversible MCMC algorithms are, as every MCMC algorithm, "embarrassingly parallel" because one can always consider a set of entirely different simulations on separate processors. Moreover, one can treat the statistically independent factors in parallel [P6]. However, when only a single interaction site moves as, for instance, in generalized Newtonian ECMC in JɛLLʏFʏsʜ-Version2.0, the number of factors stays constant as the system size is increased. That degree of parallelization does not allow to reach larger system sizes. This problem could be resolved by moving several interaction sites at once and, at best, by treating every moving interaction site and the corresponding factors in parallel. For the hard-disk system with its short-range interactions, this approach has been successfully implemented [200]. Other approaches considered the natural domain decomposition in short-range interacting systems [130, 199]. A systemic formulation for the parallelization of the non-reversible MCMC algorithms in the rigorous paradigm for long-range-interacting molecular systems is, however, a long-term project in its own and may well be part of another doctoral thesis.

This doctoral thesis introduced an approach to molecular simulation that rigorously samples the canonical ensemble without thermostats, without approximating long-range interactions, without discretizing time nor (reciprocal) space, and without unphysical artifacts. It showed that such a rigorous paradigm can become competitive with MD. The great potential for improvements of the novel approach in the future is exciting. The realization of these improvements would, however,

require a massive amount of interdisciplinary research ranging from statistics to physics to chemistry to computational science. MD has the huge advantage that such an astonishing amount of research has already been invested over the last decades. Despite its approximations and artifacts, this effort shaped it into a very efficient, powerful, and widely-used computational method. In order to convince the research community to switch away from the long-developed MD software packages and start anew, the most persuasive argument would be a demanding molecular simulation where state-of-the-art MD fails and only non-reversible MCMC perseveres. In that possible future, a wide range of exact non-reversible MCMC methods computes thermodynamic averages and generates independent configurations, whereas the undoubted power of MD yields dynamical properties. The future will show, however, whether the results of this doctoral thesis provide a sufficiently strong basis that this path will be followed by an increasing number of researchers in a self-amplifying effect. Independent of the outcome, on a final personal note, the author will be glad that he, together with his collaborators, tried to establish a new perspective on molecular simulation.

Bibliography

[P1]   L. Qin, P. Höllmer, and W. Krauth, *Direction-sweep Markov chains*,
       Journal of Physics A: Mathematical and Theoretical **55** (2022) 105003,
       URL: https://doi.org/10.1088/1751-8121/ac508a
       (cit. on pp. v, 2, 20, 49, 52, 71, 77, 79, 81, 122, 139, 147, 149).

[P2]   P. Höllmer, A. C. Maggs, and W. Krauth,
       Hard-disk dipoles and non-reversible Markov chains,
       The Journal of Chemical Physics **156** (2022) 084108,
       URL: https://doi.org/10.1063/5.0080101 (cit. on pp. v, 2, 20, 21, 54, 57, 66, 71, 75,
       91, 104, 105, 114, 122, 128, 136, 139, 140, 147, 149).

[P3]   P. Höllmer, N. Noirault, B. Li, A. C. Maggs, and W. Krauth,
       Sparse Hard-Disk Packings and Local Markov Chains,
       Journal of Statistical Physics **187** (2022) 31,
       URL: https://doi.org/10.1007/s10955-022-02908-4
       (cit. on pp. v, vi, 2, 20, 21, 80, 89, 91, 95, 122, 136, 139, 148, 149).

[P4]   P. Höllmer, A. C. Maggs, and W. Krauth,
       Molecular simulation from modern statistics: Continuous-time, continuous-space, exact,
       arXiv: 2305.02979v2 [physics.chem-ph] (2023),
       URL: https://doi.org/10.48550/arXiv.2305.02979
       (cit. on pp. v, vi, 2, 19, 20, 22, 44, 121, 126, 130–132, 148).

[P5]   B. Li, Y. Nishikawa, P. Höllmer, L. Carillo, A. C. Maggs, and W. Krauth,
       Hard-disk pressure computations—a historic perspective,
       The Journal of Chemical Physics **157** (2022) 234111,
       URL: https://doi.org/10.1063/5.0126437 (cit. on pp. vi, 16, 32, 42).

[P6]   P. Höllmer, L. Qin, M. F. Faulkner, A. C. Maggs, and W. Krauth,
       JeLLyFysh-Version1.0 — a Python application for all-atom event-chain Monte Carlo,
       Computer Physics Communications **253** (2020) 107168,
       URL: https://doi.org/10.1016/j.cpc.2020.107168 (cit. on pp. vi, 19, 36, 37, 42–44,
       49, 53, 54, 66, 76, 78, 79, 81, 122, 128, 129, 135, 136, 144, 152).

[P7]   P. Höllmer, J.-S. Bernier, C. Kollath, C. Baals, B. Santra, and H. Ott,
       Talbot effect in the presence of interactions, Physical Review A **100** (2019) 063613,
       URL: https://doi.org/10.1103/PhysRevA.100.063613 (cit. on p. vi).

[P8]   P. Höllmer, N. Noirault, B. Li, A. C. Maggs, and W. Krauth, *BigBoro, version 1.0*, 2021,
       URL: https://github.com/jellyfysh/BigBoro, last accessed: 17 August 2023
       (cit. on pp. vi, 93).

[P9]   P. Höllmer, A. C. Maggs, and W. Krauth, *JeLLyFysh, version 2.0-alpha*, 2023,
       URL: https://github.com/jellyfysh/JeLLyFysh/tree/Version2.0-alpha,
       last accessed: 17 August 2023 (cit. on pp. vi, 22, 37, 43, 44, 122, 123, 135).

[P10] B. Li, Y. Nishikawa, P. Höllmer, L. Carillo, A. C. Maggs, and W. Krauth,
       HistoricDisks, version 1.0, 2023,
       URL: https://github.com/jellyfysh/HistoricDisks, last accessed: 17 August 2023
       (cit. on p. vi).

[1]   R. P. Feynman, R. B. Leighton, and M. L. Sands,
       The Feynman Lectures on Physics, Vol. I: Mainly Mechanics, Radiation, and Heat,
       Addison–Wesley, 1964 (cit. on pp. 3, 4, 126).

[2]   R. P. Feynman, R. B. Leighton, and M. L. Sands,
       The Feynman Lectures on Physics, Vol. II: Mainly Electromagnetism and Matter,
       Addison–Wesley, 1964 (cit. on p. 3).

[3]   R. P. Feynman, R. B. Leighton, and M. L. Sands,
       The Feynman Lectures on Physics, Vol. III: Quantum Mechanics, Addison–Wesley, 1964
       (cit. on p. 3).

[4]   R. Phillips, *In retrospect: The Feynman Lectures on Physics*, Nature **504** (2013) 30,
       URL: https://doi.org/10.1038/504030a (cit. on p. 3).

[5]   S. Berryman, *Ancient Atomism*,
       The Stanford Encyclopedia of Philosophy (2022), ed. by E. N. Zalta and U. Nodelman,
       URL: https://plato.stanford.edu/entries/atomism-ancient/ (cit. on p. 3).

[6]   J. P. Millington, *John Dalton*, J. M. Dent & Co, 1906 (cit. on p. 3).

[7]   R. Oerter, *The Theory of Almost Everything: The Standard Model, the Unsung Triumph of
       Modern Physics*, Penguin, 2006 (cit. on p. 3).

[8]   P. W. Anderson, *More Is Different*, Science **177** (1972) 393,
       URL: https://doi.org/10.1126/science.177.4047.393 (cit. on p. 3).

[9]   M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, 2nd ed.,
       Oxford University Press, 2017 (cit. on pp. 4, 6, 7, 10).

[10]   D. Frenkel and B. Smit,
       Understanding Molecular Simulation: From Algorithms to Applications, 2nd ed.,
       Academic Press, 2002 (cit. on pp. 4, 6, 7, 10, 76).

[11]   T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, Springer, 2002
        (cit. on pp. 4, 6, 7, 10, 126, 151).

[12]   M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*,
        Oxford University Press, 2010 (cit. on pp. 4, 6, 7, 9, 10, 13, 38, 151).

[13]   M. Karplus, *Development of Multiscale Models for Complex Chemical Systems: From
        $H + H_2$ to Biomolecules (Nobel Lecture)*,
        Angewandte Chemie International Edition **53** (2014) 9992,
        URL: https://doi.org/10.1002/anie.201403924 (cit. on pp. 4, 7, 126).

[14]   K. Vanommeslaeghe, O. Guvench, and A. D. MacKerell Jr., *Molecular Mechanics*,
        Current Pharmaceutical Design **20** (2014) 3281,
        URL: https://doi.org/10.2174/13816128113199990600 (cit. on pp. 4, 7).

[15]   G. Ciccotti, C. Dellago, M. Ferrario, E. R. Hernández, and M. E. Tuckerman,
        Molecular simulations: past, present, and future (a Topical Issue in EPJB),
        The European Physical Journal B **95** (2022) 3,
        URL: https://doi.org/10.1140/epjb/s10051-021-00249-x (cit. on pp. 4, 7, 9, 92).

[16]   C. D. Snow, H. Nguyen, V. S. Pande, and M. Gruebele,
        Absolute comparison of simulated and experimental protein-folding dynamics,
        Nature **420** (2002) 102, URL: https://doi.org/10.1038/nature01160 (cit. on pp. 4, 7).

[17]   D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood,
        J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers,
        Atomic-Level Characterization of the Structural Dynamics of Proteins,
        Science **330** (2010) 341, URL: https://doi.org/10.1126/science.1187409
        (cit. on pp. 4, 7, 126, 133).

[18]   W. L. Jorgensen, *The Many Roles of Computation in Drug Discovery*,
        Science **303** (2004) 1813, URL: https://doi.org/10.1126/science.1096361
        (cit. on pp. 4, 7).

[19]   M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli,
        Role of Molecular Dynamics and Related Methods in Drug Discovery,
        Journal of Medicinal Chemistry **59** (2016) 4035,
        URL: https://doi.org/10.1021/acs.jmedchem.5b01684 (cit. on pp. 4, 7).

[20]   D. B. Boyd and K. B. Lipkowitz,
        Molecular mechanics: The method and its underlying philosophy,
        Journal of Chemical Education **59** (1982) 269,
        URL: https://doi.org/10.1021/ed059p269 (cit. on pp. 4, 6).

[21]   N. L. Allinger,
        Molecular Structure: Understanding Steric and Electronic Effects from Molecular Mechanics,
        Wiley, 2010 (cit. on pp. 4, 6).

[22]   C. Kittel, *Introduction to Solid State Physics*, 8th ed., Wiley, 2005 (cit. on p. 4).

[23] J. D. Bernal and R. H. Fowler, *A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions*, The Journal of Chemical Physics **1** (1933) 515, URL: https://doi.org/10.1063/1.1749327 (cit. on p. 4).

[24] D. van der Spoel, P. J. van Maaren, and H. J. C. Berendsen, *A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field*, The Journal of Chemical Physics **108** (1998) 10220, URL: https://doi.org/10.1063/1.476482 (cit. on p. 4).

[25] A. Wallqvist and R. D. Mountain, *Molecular Models of Water: Derivation and Description*, Reviews in Computational Chemistry **13** (1999) 183, URL: https://doi.org/10.1002/9780470125908.ch4 (cit. on p. 4).

[26] B. Guillot, *A reappraisal of what we have learnt during three decades of computer simulations on water*, Journal of Molecular Liquids **101** (2002) 219, URL: https://doi.org/10.1016/S0167-7322(02)00094-6 (cit. on p. 4).

[27] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, *Building Water Models: A Different Approach*, The Journal of Physical Chemistry Letters **5** (2014) 3863, URL: https://doi.org/10.1021/jz501780a (cit. on p. 4).

[28] Y. Wu, H. L. Tepper, and G. A. Voth, *Flexible simple point-charge water model with improved liquid-state properties*, The Journal of Chemical Physics **124** (2006) 024503, URL: https://doi.org/10.1063/1.2136877 (cit. on pp. 5, 37, 38, 54, 76, 127, 134, 138).

[29] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *Comparison of simple potential functions for simulating liquid water*, The Journal of Chemical Physics **79** (1983) 926, URL: https://doi.org/10.1063/1.445869 (cit. on p. 5).

[30] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *The missing term in effective pair potentials*, The Journal of Physical Chemistry **91** (1987) 6269, URL: https://doi.org/10.1021/j100308a038 (cit. on p. 5).

[31] F. H. Westheimer and J. E. Mayer, *The Theory of the Racemization of Optically Active Derivatives of Diphenyl*, The Journal of Chemical Physics **14** (1946) 733, URL: https://doi.org/10.1063/1.1724095 (cit. on p. 5).

[32] F. H. Westheimer, *A Calculation of the Energy of Activation for the Racemization of 2,2′-Dibromo–4,4′-Dicarboxydiphenyl*, The Journal of Chemical Physics **15** (1947) 252, URL: https://doi.org/10.1063/1.1746491 (cit. on p. 5).

[33] P. H. Hünenberger, *Thermostat Algorithms for Molecular Dynamics Simulations*, in: *Advanced Computer Simulation Approaches for Soft Matter Sciences I*, ed. by C. Holm and K. Kremer, Springer, 2005 (cit. on p. 6).

[34] E. Braun, J. Gilmer, H. B. Mayes, D. L. Mobley, J. I. Monroe, S. Prasad, and D. M. Zuckerman, *Best Practices for Foundations in Molecular Simulations [Article v1.0]*, Living Journal of Computational Molecular Science **1** (2018) 5957, URL: https://doi.org/10.33011/livecoms.1.1.5957 (cit. on pp. 6, 8, 9).

[35] G. Benettin and A. Giorgilli, *On the Hamiltonian interpolation of near-to-the identity symplectic mappings with application to symplectic integration algorithms*, Journal of Statistical Physics **74** (1994) 1117, URL: https://doi.org/10.1007/BF02188219 (cit. on p. 6).

[36] E. Hairer and C. Lubich, *The life-span of backward error analysis for numerical integrators*, Numerische Mathematik **76** (1997) 441, URL: https://doi.org/10.1007/s002110050271 (cit. on p. 6).

[37] S. Reich, *Backward Error Analysis for Numerical Integrators*, SIAM Journal on Numerical Analysis **36** (1999) 1549, URL: https://doi.org/10.1137/S0036142997329797 (cit. on p. 6).

[38] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics*, Cambridge University Press, 2005 (cit. on pp. 6, 13).

[39] A. Drăgănescu, R. Lehoucq, and P. Tupper, *Hamiltonian Molecular Dynamics for Computational Mechanicians and Numerical Analysts*, Sandia National Laboratories, 2009 (cit. on p. 6).

[40] R. D. Engle, R. D. Skeel, and M. Drees, *Monitoring Energy Drift with Shadow Hamiltonians*, Journal of Computational Physics **206** (2005) 432, URL: https://doi.org/10.1016/j.jcp.2004.12.009 (cit. on pp. 6, 7).

[41] D. Cottrell and P. F. Tupper, *Energy drift in molecular dynamics simulations*, BIT Numerical Mathematics **47** (2007) 507, URL: https://doi.org/10.1007/s10543-007-0134-z (cit. on p. 6).

[42] S. Toxvaerd and J. C. Dyre, *Communication: Shifted forces in molecular dynamics*, The Journal of Chemical Physics **134** (2011) 081102, URL: https://doi.org/10.1063/1.3558787 (cit. on p. 6).

[43] P. Eastman and V. S. Pande, *Energy Conservation as a Measure of Simulation Accuracy*, bioRxiv: 083055 (2016), URL: https://doi.org/10.1101/083055 (cit. on pp. 6, 92).

[44] K. D. Hammonds and D. M. Heyes, *Shadow Hamiltonian in classical NVE molecular dynamics simulations: A path to long time stability*, The Journal of Chemical Physics **152** (2020) 024114, URL: https://doi.org/10.1063/1.5139708 (cit. on pp. 6, 7, 126).

[45] M. Patra, M. Karttunen, M. T. Hyvönen, E. Falck, P. Lindqvist, and I. Vattulainen,
*Molecular Dynamics Simulations of Lipid Bilayers: Major Artifacts Due to Truncating
Electrostatic Interactions*, Biophysical Journal **84** (2003) 3636,
URL: https://doi.org/10.1016/S0006-3495(03)75094-2 (cit. on pp. 6, 7).

[46] J. Wong-ekkabut and M. Karttunen, *The good, the bad and the user in soft matter simulations*,
Biochimica et Biophysica Acta - Biomembranes **1858** (2016) 2529,
URL: https://doi.org/10.1016/j.bbamem.2016.02.004 (cit. on pp. 6, 7, 126, 127).

[47] P. P. Ewald, *Die Berechnung optischer und elektrostatischer Gitterpotentiale*,
Annalen der Physik **369** (1921) 253,
URL: https://doi.org/10.1002/andp.19213690304 (cit. on pp. 6, 13, 135, 136).

[48] S. W. de Leeuw, J. W. Perram, and E. R. Smith, *Simulation of electrostatic systems in
periodic boundary conditions. I. Lattice sums and dielectric constants*, Proceedings of the
Royal Society A: Mathematical, Physical and Engineering Sciences **373** (1980) 27,
URL: https://doi.org/10.1098/rspa.1980.0135 (cit. on pp. 6, 13, 19, 127).

[49] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles*, CRC Press, 1988
(cit. on pp. 6, 7, 13, 127, 134).

[50] T. Darden, D. York, and L. Pedersen,
Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems,
The Journal of Chemical Physics **98** (1993) 10089,
URL: https://doi.org/10.1063/1.464397 (cit. on pp. 6, 13, 123, 127, 129).

[51] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen,
A smooth particle mesh Ewald method, The Journal of Chemical Physics **103** (1995) 8577,
URL: https://doi.org/10.1063/1.470117 (cit. on pp. 6, 13, 127).

[52] R. D. Skeel, D. J. Hardy, and J. C. Phillips, *Correcting mesh-based force calculations to
conserve both energy and momentum in molecular dynamics simulations*,
Journal of Computational Physics **225** (2007) 1,
URL: https://doi.org/10.1016/j.jcp.2007.03.010 (cit. on p. 6).

[53] K. D. Hammonds and D. M. Heyes, *Shadow Hamiltonian in classical NVE molecular
dynamics simulations involving Coulomb interactions*,
The Journal of Chemical Physics **154** (2021) 174102,
URL: https://doi.org/10.1063/5.0048194 (cit. on pp. 6, 7).

[54] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier,
P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens,
J. Tranchida, C. Trott, and S. J. Plimpton, *LAMMPS - a flexible simulation tool for
particle-based materials modeling at the atomic, meso, and continuum scales*,
Computer Physics Communications **271** (2022) 108171,
URL: https://doi.org/10.1016/j.cpc.2021.108171
(cit. on pp. 7, 8, 14, 124, 126, 134).

[55] S. Plimpton, A. Thompson, S. Moore, A. Kohlmeyer, and R. Berger,
LAMMPS, Feature release 8 February 2023, 2023, URL: https://www.lammps.org,
last accessed: 17 August 2023 (cit. on pp. 7, 8, 14, 134).

[56] C. R. A. Abreu, *Playmol, commit 67eb56c from 26 November 2019*, 2019,
URL: http://atoms.peq.coppe.ufrj.br/playmol/index.html,
last accessed: 17 August 2023 (cit. on pp. 8, 137).

[57] M. R. Shirts,
Simple Quantitative Tests to Validate Sampling from Thermodynamic Ensembles,
Journal of Chemical Theory and Computation **9** (2013) 909,
URL: https://doi.org/10.1021/ct300688p (cit. on p. 7).

[58] M. Hopkin, *Mmm. . . pi*, Nature **448** (2007) 404,
URL: https://doi.org/10.1038/448404a (cit. on p. 7).

[59] S. Piana, K. Lindorff-Larsen, R. M. Dirks, J. K. Salmon, R. O. Dror, and D. E. Shaw,
*Evaluating the Effects of Cutoffs and Treatment of Long-range Electrostatics in Protein
Folding Simulations*, PLoS ONE **7** (2012) e39918,
URL: https://doi.org/10.1371/journal.pone.0039918 (cit. on p. 7).

[60] H. Kim, B. Fábián, and G. Hummer,
Neighbor List Artifacts in Molecular Dynamics Simulations, ChemRxiv: zbj6j (2023),
URL: https://doi.org/10.26434/chemrxiv-2023-zbj6j (cit. on p. 7).

[61] T. Morishita,
Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath,
The Journal of Chemical Physics **113** (2000) 2976,
URL: https://doi.org/10.1063/1.1287333 (cit. on p. 7).

[62] M. D'Alessandro, A. Tenenbaum, and A. Amadei,
Dynamical and Statistical Mechanical Characterization of Temperature Coupling Algorithms,
The Journal of Physical Chemistry B **106** (2002) 5050,
URL: https://doi.org/10.1021/jp013689i (cit. on p. 7).

[63] S. C. Harvey, R. K.-Z. Tan, and T. E. Cheatham III, *The flying ice cube: Velocity rescaling in
molecular dynamics leads to violation of energy equipartition*,
Journal of Computational Chemistry **19** (1998) 726,
URL: https://doi.org/10.1002/(SICI)1096-987X(199805)19:7%3C726::AID-
JCC4%3E3.0.CO;2-S (cit. on p. 7).

[64] K. Oda, H. Miyagawa, and K. Kitamura, *How does the Electrostatic Force Cut-Off Generate
Non-uniform Temperature Distributions in Proteins?* Molecular Simulation **16** (1996) 167,
URL: https://doi.org/10.1080/08927029608024070 (cit. on p. 7).

[65] M. Lingenheil, R. Denschlag, R. Reichold, and P. Tavan,
The "Hot-Solvent/Cold-Solute" Problem Revisited,
Journal of Chemical Theory and Computation **4** (2008) 1293,
URL: https://doi.org/10.1021/ct8000365 (cit. on p. 7).

[66] M. E. Tuckerman and G. J. Martyna,
Understanding Modern Molecular Dynamics: Techniques and Applications,
The Journal of Physical Chemistry B **104** (2000) 159,
URL: https://doi.org/10.1021/jp992433y (cit. on p. 7).

[67] M. C. Zwier and L. T. Chong,
Reaching biological timescales with all-atom molecular dynamics simulations,
Current Opinion in Pharmacology **10** (2010) 745,
URL: https://doi.org/10.1016/j.coph.2010.09.008 (cit. on p. 9).

[68] J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsson, and L. Delemotte,
Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0],
Living Journal of Computational Molecular Science **4** (2022) 1583,
URL: https://doi.org/10.33011/livecoms.4.1.1583 (cit. on p. 9).

[69] C. R. A. Abreu and M. E. Tuckerman,
Multiple timescale molecular dynamics with very large time steps: avoidance of resonances,
The European Physical Journal B **94** (2021) 231,
URL: https://doi.org/10.1140/epjb/s10051-021-00226-4 (cit. on p. 9).

[70] L. López-Villellas, C. C. Kjelgaard Mikkelsen, J. J. Galano-Frutos, S. Marco-Sola,
J. Alastruey-Benedé, P. Ibáñez, M. Moretó, J. Sancho, and P. García-Risueño,
*Accurate and efficient constrained molecular dynamics of polymers using Newton's method
and special purpose code*, Computer Physics Communications **288** (2023) 108742,
URL: https://doi.org/10.1016/j.cpc.2023.108742 (cit. on p. 9).

[71] J. E. Basconi and M. R. Shirts, *Effects of Temperature Control Algorithms on Transport
Properties and Kinetics in Molecular Dynamics Simulations*,
Journal of Chemical Theory and Computation **9** (2013) 2887,
URL: https://doi.org/10.1021/ct400109a (cit. on p. 9).

[72] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller,
Equation of State Calculations by Fast Computing Machines,
The Journal of Chemical Physics **21** (1953) 1087,
URL: https://doi.org/10.1063/1.1699114
(cit. on pp. 9, 11, 16, 39, 50, 52, 53, 68, 75, 76, 80, 95, 96, 103, 128, 135).

[73] S. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd ed.,
Cambridge University Press, 2009 (cit. on pp. 9, 10, 30).

[74] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*, 2nd ed.,
American Mathematical Society, 2017
(cit. on pp. 9, 10, 24–27, 29, 31, 34, 35, 52, 53, 62, 67, 75, 79, 105, 111).

[75] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*,
Springer, 1996 (cit. on pp. 9, 29, 30, 33).

[76] S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng,
Handbook of Markov Chain Monte Carlo, CRC Press, 2011 (cit. on pp. 9, 31).

[77] D. B. Dunson and J. E. Johndrow, *The Hastings algorithm at fifty*, Biometrika **107** (2019) 1,
URL: https://doi.org/10.1093/biomet/asz066 (cit. on p. 9).

[78] W. Krauth, *Statistical Mechanics: Algorithms and Computations*,
Oxford University Press, 2006 (cit. on pp. 10, 33, 56, 85, 96, 113).

[79] A. Robison, *Did Einstein really say that?* Nature **557** (2018) 30,
URL: https://doi.org/10.1038/d41586-018-05004-4 (cit. on p. 10).

[80] W. Krauth, *Event-Chain Monte Carlo: Foundations, Applications, and Prospects*,
Frontiers in Physics **9** (2021) 663457,
URL: https://doi.org/10.3389/fphy.2021.663457
(cit. on pp. 10, 12, 15, 29, 31, 34, 35, 53, 57, 66–69, 76, 79, 95, 111, 127, 132, 139, 152).

[81] G. O. Roberts and J. S. Rosenthal, *General state space Markov chains and MCMC algorithms*,
Probability Surveys **1** (2004) 20,
URL: https://doi.org/10.1214/154957804100000024 (cit. on pp. 10, 28–30, 32).

[82] J. Dongarra and F. Sullivan, *Guest Editors Introduction to the top 10 algorithms*,
Computing in Science & Engineering **2** (2000) 22,
URL: https://doi.org/10.1109/MCISE.2000.814652 (cit. on p. 11).

[83] F. Sullivan and I. Beichl, *The Metropolis Algorithm*,
Computing in Science & Engineering **2** (2000) 65,
URL: https://doi.org/10.1109/5992.814660 (cit. on p. 11).

[84] R. C. Tolman, *The Principles of Statistical Mechanics*, Oxford University Press, 1938
(cit. on p. 12).

[85] W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*,
Biometrika **57** (1970) 97, URL: https://doi.org/10.1093/biomet/57.1.97
(cit. on p. 12).

[86] R. J. Glauber, *Time-Dependent Statistics of the Ising Model*,
Journal of Mathematical Physics **4** (1963) 294,
URL: https://doi.org/10.1063/1.1703954 (cit. on pp. 12, 76).

[87] M. Creutz, *Monte Carlo study of quantized SU(2) gauge theory*,
Physical Review D **21** (1980) 2308,
URL: https://doi.org/10.1103/PhysRevD.21.2308 (cit. on pp. 12, 76).

[88] S. Geman and D. Geman,
Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images,
IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-6** (1984) 721,
URL: https://doi.org/10.1109/tpami.1984.4767596 (cit. on pp. 12, 76).

[89] R. H. Swendsen and J.-S. Wang, *Nonuniversal critical dynamics in Monte Carlo simulations*,
Physical Review Letters **58** (1987) 86,
URL: https://doi.org/10.1103/PhysRevLett.58.86 (cit. on p. 12).

[90]   U. Wolff, *Collective Monte Carlo Updating for Spin Systems*,
       Physical Review Letters **62** (1989) 361,
       URL: https://doi.org/10.1103/PhysRevLett.62.361 (cit. on p. 12).

[91]   Z. Lei, W. Krauth, and A. C. Maggs, *Event-chain Monte Carlo with factor fields*,
       Physical Review E **99** (2019) 043301,
       URL: https://doi.org/10.1103/PhysRevE.99.043301 (cit. on p. 12).

[92]   A. C. Maggs and W. Krauth, *Large-scale dynamics of event-chain Monte Carlo*,
       Physical Review E **105** (2022) 015309,
       URL: https://doi.org/10.1103/PhysRevE.105.015309 (cit. on pp. 12, 152).

[93]   W. L. Jorgensen and J. Tirado-Rives,
       Monte Carlo vs Molecular Dynamics for Conformational Sampling,
       The Journal of Physical Chemistry **100** (1996) 14508,
       URL: https://doi.org/10.1021/jp960880x (cit. on p. 13).

[94]   A. Vitalis and R. V. Pappu,
       Chapter 3 Methods for Monte Carlo Simulations of Biomacromolecules,
       Annual Reports in Computational Chemistry **5** (2009) 49, ed. by R. A. Wheeler,
       URL: https://doi.org/10.1016/S1574-1400(09)00503-9 (cit. on pp. 13, 14).

[95]   L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*,
       Journal of Computational Physics **73** (1987) 325,
       URL: https://doi.org/10.1016/0021-9991(87)90140-9 (cit. on pp. 13, 131).

[96]   H. G. Petersen, *Accuracy and efficiency of the particle mesh Ewald method*,
       The Journal of Chemical Physics **103** (1995) 3668,
       URL: https://doi.org/10.1063/1.470043 (cit. on pp. 13, 18, 127, 131, 135).

[97]   B. Kohnke, C. Kutzner, and H. Grubmüller,
       A GPU-Accelerated Fast Multipole Method for GROMACS: Performance and Accuracy,
       Journal of Chemical Theory and Computation **16** (2020) 6938,
       URL: https://doi.org/10.1021/acs.jctc.0c00744 (cit. on pp. 13, 127).

[98]   H. Berthoumieux and A. C. Maggs, *Collective dispersion forces in the fluid state*,
       Europhysics Letters **91** (2010) 56006,
       URL: https://doi.org/10.1209/0295-5075/91/56006 (cit. on pp. 13, 76, 89).

[99]   W. R. Saunders, J. Grant, and E. H. Müller,
       A new algorithm for electrostatic interactions in Monte Carlo simulations of charged particles,
       Journal of Computational Physics **430** (2021) 110099,
       URL: https://doi.org/10.1016/j.jcp.2020.110099 (cit. on pp. 13, 131).

[100] M. Jorge and N. A. Seaton,
       Long-range interactions in Monte Carlo simulation of confined water,
       Molecular Physics **100** (2002) 2017,
       URL: https://doi.org/10.1080/00268970110099585 (cit. on p. 13).

[101] J. M. Martínez and L. Martínez, *Packing Optimization for Automated Generation of Complex System's Initial Configurations for Molecular Dynamics and Docking*,
Journal of Computational Chemistry **24** (2003) 819,
URL: https://doi.org/10.1002/jcc.10216 (cit. on p. 13).

[102] M. Salanne, J. M. Buriak, X. Chen, W. Chueh, M. C. Hersam, and R. E. Schaak,
Best Practices for Simulations and Calculations of Nanomaterials for Energy Applications: Avoiding "Garbage In, Garbage Out", ACS Nano **17** (2023) 6147,
URL: https://doi.org/10.1021/acsnano.3c02368 (cit. on p. 14).

[103] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev,
C. Simmerling, B. Wang, and R. J. Woods, *The Amber biomolecular simulation programs*,
Journal of Computational Chemistry **26** (2005) 1668,
URL: https://doi.org/10.1002/jcc.20290 (cit. on pp. 14, 126).

[104] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux,
Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner,
M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma,
V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor,
R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus,
CHARMM: The biomolecular simulation program,
Journal of Computational Chemistry **30** (2009) 1545,
URL: https://doi.org/10.1002/jcc.21287 (cit. on pp. 14, 126).

[105] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen,
GROMACS: A message-passing parallel molecular dynamics implementation,
Computer Physics Communications **91** (1995) 43,
URL: https://doi.org/10.1016/0010-4655(95)00042-E (cit. on pp. 14, 126).

[106] J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch,
G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel,
A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé,
K. Schulten, C. Chipot, and E. Tajkhorshid,
Scalable molecular dynamics on CPU and GPU architectures with NAMD,
The Journal of Chemical Physics **153** (2020) 044130,
URL: https://doi.org/10.1063/5.0014475 (cit. on pp. 14, 126).

[107] W. L. Jorgensen and J. Tirado–Rives,
Molecular modeling of organic and biomolecular systems using BOSS and MCPRO,
Journal of Computational Chemistry **26** (2005) 1689,
URL: https://doi.org/10.1002/jcc.20297 (cit. on p. 14).

[108] J. A. Purton, J. C. Crabtree, and S. C. Parker,
DL_MONTE: a general purpose program for parallel Monte Carlo simulation,
Molecular Simulation **39** (2013) 1240,
URL: https://doi.org/10.1080/08927022.2013.839871
(cit. on pp. 14, 123, 131, 135).

[109] K. S. Turitsyn, M. Chertkov, and M. Vucelja,
Irreversible Monte Carlo algorithms for efficient sampling,
Physica D: Nonlinear Phenomena **240** (2011) 410,
URL: https://doi.org/10.1016/j.physd.2010.10.003 (cit. on pp. 14, 15, 53, 69, 76).

[110] M. F. Eggl and P. J. Schmid, *Mixing by stirring: Optimizing shapes and strategies*,
Physical Review Fluids **7** (2022) 073904,
URL: https://doi.org/10.1103/PhysRevFluids.7.073904 (cit. on p. 14).

[111] P. Diaconis, S. Holmes, and R. M. Neal, *Analysis of a nonreversible Markov chain sampler*,
The Annals of Applied Probability **10** (2000) 726,
URL: https://doi.org/10.1214/aoap/1019487508
(cit. on pp. 15, 16, 34, 35, 53, 67, 76, 77, 111, 139).

[112] F. Chen, L. Lovász, and I. Pak, *Lifting Markov Chains to Speed up Mixing*, Proceedings of the
Thirty-First Annual ACM Symposium on Theory of Computing (1999) 275,
URL: https://doi.org/10.1145/301250.301315
(cit. on pp. 15, 34, 35, 67, 68, 76, 77, 111, 139).

[113] M. Vucelja, *Lifting—A nonreversible Markov chain Monte Carlo algorithm*,
American Journal of Physics **84** (2016) 958,
URL: https://doi.org/10.1119/1.4961596 (cit. on p. 15).

[114] F. H. L. Essler and W. Krauth,
Lifted TASEP: a Bethe ansatz integrable paradigm for non-reversible Markov chains,
arXiv: 2306.13059 [cond-mat.stat-mech] (2023),
URL: https://doi.org/10.48550/arXiv.2306.13059 (cit. on p. 15).

[115] E. P. Bernard, W. Krauth, and D. B. Wilson,
Event-chain Monte Carlo algorithms for hard-sphere systems,
Physical Review E **80** (2009) 056704,
URL: https://doi.org/10.1103/PhysRevE.80.056704 (cit. on pp. 15, 20, 21, 36, 40,
46, 47, 53, 72, 76, 77, 80, 81, 83, 86, 88, 95, 96, 104, 127, 128, 143).

[116] E. A. J. F. Peters and G. de With, *Rejection-free Monte Carlo sampling for general potentials*,
Physical Review E **85** (2012) 026703,
URL: https://doi.org/10.1103/PhysRevE.85.026703 (cit. on pp. 15, 41, 140).

[117] F. Faizi, G. Deligiannidis, and E. Rosta, *Efficient Irreversible Monte Carlo Samplers*,
Journal of Chemical Theory and Computation **16** (2020) 2124,
URL: https://doi.org/10.1021/acs.jctc.9b01135 (cit. on p. 15).

[118] H. Suwa and S. Todo, *Markov Chain Monte Carlo Method without Detailed Balance*,
Physical Review Letters **105** (2010) 120603,
URL: https://doi.org/10.1103/PhysRevLett.105.120603 (cit. on pp. 15, 53).

[119] H. Suwa, *Lifted directed-worm algorithm*, Physical Review E **106** (2022) 055306,
URL: https://doi.org/10.1103/PhysRevE.106.055306 (cit. on p. 15).

[120] H. Zhao and M. Vucelja, *Nonreversible Markov Chain Monte Carlo Algorithm for Efficient Generation of Self-Avoiding Walks*, Frontiers in Physics **9** (2022) 782156, URL: https://doi.org/10.3389/fphy.2021.782156 (cit. on p. 15).

[121] M. Michel, S. C. Kapfer, and W. Krauth, *Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps*, The Journal of Chemical Physics **140** (2014) 054116, URL: https://doi.org/10.1063/1.4863991 (cit. on pp. 15, 17, 36, 39–41, 46, 53, 76, 113, 127, 139, 141).

[122] S. C. Kapfer and W. Krauth, *Two-Dimensional Melting: From Liquid-Hexatic Coexistence to Continuous Transitions*, Physical Review Letters **114** (2015) 035702, URL: https://doi.org/10.1103/PhysRevLett.114.035702 (cit. on pp. 15, 112).

[123] M. Isobe and W. Krauth, *Hard-sphere melting and crystallization with event-chain Monte Carlo*, The Journal of Chemical Physics **143** (2015) 084509, URL: https://doi.org/10.1063/1.4929529 (cit. on p. 15).

[124] M. Klement and M. Engel, *Efficient equilibration of hard spheres with Newtonian event chains*, The Journal of Chemical Physics **150** (2019) 174108, URL: https://doi.org/10.1063/1.5090882 (cit. on pp. 15, 20–22, 36, 37, 53, 72, 77, 81, 84, 86, 88, 96, 104, 105, 122, 128, 139, 140).

[125] M. Michel, J. Mayer, and W. Krauth, *Event-chain Monte Carlo for classical continuous spin models*, Europhysics Letters **112** (2015) 20003, URL: https://doi.org/10.1209/0295-5075/112/20003 (cit. on pp. 16, 76).

[126] Y. Nishikawa, M. Michel, W. Krauth, and K. Hukushima, *Event-chain algorithm for the Heisenberg model: Evidence for $z \simeq 1$ dynamic scaling*, Physical Review E **92** (2015) 063306, URL: https://doi.org/10.1103/PhysRevE.92.063306 (cit. on pp. 16, 76).

[127] K. Kimura and S. Higuchi, *Anomalous diffusion analysis of the lifting events in the event-chain Monte Carlo for the classical XY models*, Europhysics Letters **120** (2017) 30003, URL: https://doi.org/10.1209/0295-5075/120/30003 (cit. on pp. 16, 76).

[128] Z. Lei and W. Krauth, *Irreversible Markov chains in spin models: Topological excitations*, Europhysics Letters **121** (2018) 10008, URL: https://doi.org/10.1209/0295-5075/121/10008 (cit. on p. 16).

[129] T. A. Kampmann, H.-H. Boltz, and J. Kierfeld, *Monte Carlo simulation of dense polymer melts using event chain algorithms*, The Journal of Chemical Physics **143** (2015) 044105, URL: https://doi.org/10.1063/1.4927084 (cit. on p. 16).

[130] T. A. Kampmann, H.-H. Boltz, and J. Kierfeld,
Parallelized event chain algorithm for dense hard sphere and polymer systems,
Journal of Computational Physics **281** (2015) 864,
URL: https://doi.org/10.1016/j.jcp.2014.10.059 (cit. on pp. 16, 152).

[131] D. Müller, T. A. Kampmann, and J. Kierfeld, *Chaining of hard disks in nematic needles:
particle-based simulation of colloidal interactions in liquid crystals*,
Scientific Reports **10** (2020) 12718,
URL: https://doi.org/10.1038/s41598-020-69544-4 (cit. on pp. 16, 54, 76, 79).

[132] T. A. Kampmann, D. Müller, L. P. Weise, C. F. Vorsmann, and J. Kierfeld,
Event-Chain Monte-Carlo Simulations of Dense Soft Matter Systems,
Frontiers in Physics **9** (2021) 635886,
URL: https://doi.org/10.3389/fphy.2021.635886 (cit. on pp. 16, 46, 54, 76, 79).

[133] M. Hasenbusch and S. Schaefer,
Testing the event-chain algorithm in asymptotically free models,
Physical Review D **98** (2018) 054502,
URL: https://doi.org/10.1103/PhysRevD.98.054502 (cit. on pp. 16, 76).

[134] M. Klement, S. Lee, J. A. Anderson, and M. Engel,
Newtonian Event-Chain Monte Carlo and Collision Prediction with Polyhedral Particles,
Journal of Chemical Theory and Computation **17** (2021) 4686,
URL: https://doi.org/10.1021/acs.jctc.1c00311 (cit. on pp. 16, 21, 71).

[135] T. A. Kampmann, T. Sathiyanesan, and J. Kierfeld,
Kinetic Event-Chain Algorithm for Active Matter, arXiv: 2111.06760 [cond-mat.soft] (2021),
URL: https://doi.org/10.48550/arXiv.2111.06760 (cit. on p. 16).

[136] M. H. A. Davis, *Piecewise-Deterministic Markov Processes: A General Class of
Non-Diffusion Stochastic Models*,
Journal of the Royal Statistical Society: Series B (Methodological) **46** (1984) 353,
URL: https://doi.org/10.1111/j.2517-6161.1984.tb01308.x
(cit. on pp. 16, 40, 76, 127).

[137] A. Monemvassitis, A. Guillin, and M. Michel,
PDMP Characterisation of Event-Chain Monte Carlo Algorithms for Particle Systems,
Journal of Statistical Physics **190** (2023) 66,
URL: https://doi.org/10.1007/s10955-023-03069-8 (cit. on pp. 16, 40, 139).

[138] P. Vanetti, A. Bouchard-Côté, G. Deligiannidis, and A. Doucet,
Piecewise-Deterministic Markov Chain Monte Carlo, arXiv: 1707.05296 [stat.ME] (2018),
URL: https://doi.org/10.48550/arXiv.1707.05296 (cit. on pp. 16, 132, 151).

[139] P. Fearnhead, J. Bierkens, M. Pollock, and G. O. Roberts,
Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo,
Statistical Science **33** (2018) 386, URL: https://doi.org/10.1214/18-STS648
(cit. on pp. 16, 42, 132).

[140] A. Bouchard-Côté, S. J. Vollmer, and A. Doucet, *The Bouncy Particle Sampler: A Nonreversible Rejection-Free Markov Chain Monte Carlo Method*, Journal of the American Statistical Association **113** (2018) 855, URL: https://doi.org/10.1080/01621459.2017.1294075 (cit. on pp. 16, 42, 43, 73, 76, 83, 104, 151).

[141] J. Bierkens, P. Fearnhead, and G. Roberts, *The Zig-Zag process and super-efficient sampling for Bayesian analysis of big data*, The Annals of Statistics **47** (2019) 1288, URL: https://doi.org/10.1214/18-AOS1715 (cit. on pp. 16, 76).

[142] E. P. Bernard and W. Krauth, *Two-Step Melting in Two Dimensions: First-Order Liquid-Hexatic Transition*, Physical Review Letters **107** (2011) 155704, URL: https://doi.org/10.1103/PhysRevLett.107.155704 (cit. on pp. 16, 20, 40, 76, 95, 104, 127, 151).

[143] M. Engel, J. A. Anderson, S. C. Glotzer, M. Isobe, E. P. Bernard, and W. Krauth, *Hard-disk equation of state: First-order liquid-hexatic transition in two dimensions with three simulation methods*, Physical Review E **87** (2013) 042134, URL: https://doi.org/10.1103/PhysRevE.87.042134 (cit. on pp. 16, 17, 104).

[144] B. J. Alder and T. E. Wainwright, *Phase Transition in Elastic Disks*, Physical Review **127** (1962) 359, URL: https://doi.org/10.1103/PhysRev.127.359 (cit. on pp. 16, 95).

[145] J. Harland, M. Michel, T. A. Kampmann, and J. Kierfeld, *Event-chain Monte Carlo algorithms for three- and many-particle interactions*, Europhysics Letters **117** (2017) 30001, URL: https://doi.org/10.1209/0295-5075/117/30001 (cit. on pp. 17, 36, 46, 128, 135).

[146] M. F. Faulkner, L. Qin, A. C. Maggs, and W. Krauth, *All-atom computations with irreversible Markov chains*, The Journal of Chemical Physics **149** (2018) 064113, URL: https://doi.org/10.1063/1.5036638 (cit. on pp. 17, 19, 36, 38–40, 43, 44, 47, 49, 50, 53, 54, 76, 79, 123, 128, 129, 134, 135, 144).

[147] M. Deserno and C. Holm, *How to mesh up Ewald sums. II. An accurate error estimate for the particle–particle–particle-mesh algorithm*, The Journal of Chemical Physics **109** (1998) 7694, URL: https://doi.org/10.1063/1.477415 (cit. on pp. 18, 131, 135).

[148] S. C. Kapfer and W. Krauth, *Cell-veto Monte Carlo algorithm for long-range systems*, Physical Review E **94** (2016) 031302, URL: https://doi.org/10.1103/PhysRevE.94.031302 (cit. on pp. 19, 36, 43, 76, 89, 128, 136, 144).

[149] L. Qin, *Application of irreversible Monte Carlo in realistic long-range systems*,
       PhD Thesis: Université Paris Sciences et Lettres, 2020,
       URL: https://tel.archives-ouvertes.fr/tel-02998657,
       last accessed: 17 August 2023 (cit. on pp. 20, 76, 77, 89, 135).

[150] M. Michel, A. Durmus, and S. Sénécal, *Forward Event-Chain Monte Carlo: Fast Sampling by
       Randomness Control in Irreversible Markov Chains*,
       Journal of Computational and Graphical Statistics **29** (2020) 689,
       URL: https://doi.org/10.1080/10618600.2020.1750417
       (cit. on pp. 20, 21, 36, 37, 42, 53, 72, 77, 81, 84, 96, 104, 128, 143, 151).

[151] C. J. Geyer, *Markov Chain Monte Carlo Lecture Notes*, 2005,
       URL: https://www.stat.umn.edu/geyer/f05/8931/n1998.pdf,
       last accessed: 17 August 2023 (cit. on pp. 28, 32, 33).

[152] D. F. Anderson, *Lecture Notes on Stochastic Processes with Applications in Biology*, 2017,
       URL: https://u.math.biu.ac.il/~amirgi/SBA.pdf, last accessed: 17 August 2023
       (cit. on p. 29).

[153] J. Bierkens, G. O. Roberts, and P.-A. Zitt, *Ergodicity of the zigzag process*,
       The Annals of Applied Probability **29** (2019) 2266,
       URL: https://doi.org/10.1214/18-AAP1453 (cit. on p. 30).

[154] G. Deligiannidis, A. Bouchard-Côté, and A. Doucet,
       Exponential ergodicity of the bouncy particle sampler,
       The Annals of Statistics **47** (2019) 1268, URL: https://doi.org/10.1214/18-AOS1714
       (cit. on p. 30).

[155] A. Durmus, A. Guillin, and P. Monmarché,
       Geometric ergodicity of the Bouncy Particle Sampler,
       The Annals of Applied Probability **30** (2020) 2069,
       URL: https://doi.org/10.1214/19-AAP1552 (cit. on p. 30).

[156] A. Monemvassitis, A. Guillin, and M. Michel,
       PDMP Characterisation of Event-Chain Monte Carlo Algorithms for Particle Systems,
       Journal of Statistical Physics **190** (2023) 66,
       URL: https://doi.org/10.1007/s10955-023-03069-8 (cit. on p. 30).

[157] A. Sokal, *Functional Integration: Basics and Applications, Monte Carlo Methods in
       Statistical Mechanics: Foundations and New Algorithms*,
       ed. by C. DeWitt-Morette, P. Cartier, and A. Folacci, Springer, 1997 (cit. on pp. 31–33, 79).

[158] M. B. Priestley, *Spectral Analysis and Time Series*, Academic Press, 1982 (cit. on pp. 32, 33).

[159] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *emcee: The MCMC Hammer*,
       Publications of the Astronomical Society of the Pacific **125** (2013) 306,
       URL: https://doi.org/10.1086/670067 (cit. on pp. 33, 90).

[160] H. Flyvbjerg and H. G. Petersen, *Error estimates on averages of correlated data*,
The Journal of Chemical Physics **91** (1989) 461,
URL: https://doi.org/10.1063/1.457480 (cit. on pp. 33, 79).

[161] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, CRC Press, 1993
(cit. on p. 34).

[162] Y. Nishikawa, J. Takahashi, and T. Takahashi,
Stationary Bootstrap: A Refined Error Estimation for Equilibrium Time Series,
arXiv: 2112.11837 [cond-mat.stat-mech] (2021),
URL: https://doi.org/10.48550/arXiv.2112.11837 (cit. on p. 34).

[163] M. Michel, X. Tan, and Y. Deng, *Clock Monte Carlo methods*,
Physical Review E **99** (2019) 010105,
URL: https://doi.org/10.1103/PhysRevE.99.010105 (cit. on p. 39).

[164] L. Devroye, *Non-Uniform Random Variate Generation*, Springer, 1986 (cit. on p. 41).

[165] P. A. W. Lewis and G. S. Shedler,
Simulation of nonhomogeneous Poisson processes by thinning,
Naval Research Logistics Quarterly **26** (1979) 403,
URL: https://doi.org/10.1002/nav.3800260304 (cit. on pp. 42, 128, 139, 144).

[166] A. J. Walker,
An Efficient Method for Generating Discrete Random Variables with General Distributions,
ACM Transactions on Mathematical Software **3** (1977) 253,
URL: https://doi.org/10.1145/355744.355749 (cit. on pp. 44, 128).

[167] H. Risken, *The Fokker–Planck Equation*, Springer, 1989 (cit. on pp. 50, 59).

[168] B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis: With
Web-Based Fortran Code*, World Scientific, 2004 (cit. on pp. 50, 53, 75).

[169] C. J. O'Keeffe and G. Orkoulas,
Parallel canonical Monte Carlo simulations through sequential updating of particles,
The Journal of Chemical Physics **130** (2009) 134109,
URL: https://doi.org/10.1063/1.3097528 (cit. on pp. 50, 53).

[170] S. C. Kapfer and W. Krauth,
Irreversible Local Markov Chains with Rapid Convergence towards Equilibrium,
Physical Review Letters **119** (2017) 240603,
URL: https://doi.org/10.1103/PhysRevLett.119.240603 (cit. on pp. 50, 53).

[171] Z. Lei and W. Krauth, *Mixing and perfect sampling in one-dimensional particle systems*,
Europhysics Letters **124** (2018) 20003,
URL: https://doi.org/10.1209/0295-5075/124/20003 (cit. on pp. 50, 53, 76).

[172] R. Ren, C. J. O'Keeffe, and G. Orkoulas,
Sequential updating algorithms for grand canonical Monte Carlo simulations,
Molecular Physics **105** (2007) 231,
URL: https://doi.org/10.1080/00268970601143341 (cit. on pp. 50, 53).

[173] V. I. Manousiouthakis and M. W. Deem,
Strict detailed balance is unnecessary in Monte Carlo simulation,
The Journal of Chemical Physics **110** (1999) 2753,
URL: https://doi.org/10.1063/1.477973 (cit. on p. 50).

[174] L. Allen, A. O'Connell, and V. Kiermer,
How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship,
Learned Publishing **32** (2019) 71, URL: https://doi.org/10.1002/leap.1210
(cit. on pp. 51, 73, 94, 124).

[175] H. C. Fernandes and M. Weigel, *Non-reversible Monte Carlo simulations of spin models*,
Computer Physics Communication **182** (2011) 1856,
URL: https://doi.org/10.1016/j.cpc.2010.11.017 (cit. on p. 53).

[176] J. Bierkens, A. Bouchard-Côté, A. Doucet, A. B. Duncan, P. Fearnhead, T. Lienart,
G. Roberts, and S. J. Vollmer,
Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains,
Statistics & Probability Letters **136** (2018) 148,
URL: https://doi.org/10.1016/j.spl.2018.02.021 (cit. on p. 53).

[177] J. Bierkens and G. Roberts, *A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model*,
The Annals of Applied Probability **27** (2017) 846,
URL: https://doi.org/10.1214/16-aap1217 (cit. on p. 53).

[178] R. F. B. Weigel,
Equilibration of Orientational Order in Hard Disks via Arcuate Event-Chain Monte Carlo,
Master Thesis: Friedrich-Alexander-Universität Erlangen-Nürnberg, 2018,
URL: https://theorie1.physik.uni-erlangen.de/research/theses/2018-ma-roweigel.html, last accessed: 17 August 2023 (cit. on pp. 53, 54, 109).

[179] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003
(cit. on p. 56).

[180] P. Diaconis, *The Mathematics of Mixing Things Up*,
Journal of Statistical Physics **144** (2011) 445,
URL: https://doi.org/10.1007/s10955-011-0284-x (cit. on p. 62).

[181] Y. Sakai and K. Hukushima,
Eigenvalue analysis of an irreversible random walk with skew detailed balance conditions,
Physical Review E **93** (2016) 043318,
URL: https://doi.org/10.1103/PhysRevE.93.043318 (cit. on p. 69).

[182] R. Ren, C. J. O'Keeffe, and G. Orkoulas,
Sequential Metropolis Algorithms for Fluid Simulations,
International Journal of Thermophysics **28** (2007) 520,
URL: https://doi.org/10.1007/s10765-007-0193-z (cit. on p. 75).

[183] C. Dress and W. Krauth, *Cluster algorithm for hard spheres and related systems*,
Journal of Physics A: Mathematical and General **28** (1995) L597,
URL: https://doi.org/10.1088/0305-4470/28/23/001 (cit. on pp. 80, 113).

[184] J. Liu and E. Luijten, *Rejection-Free Geometric Cluster Algorithm for Complex Fluids*,
Physical Review Letters **92** (2004) 035504,
URL: https://doi.org/10.1103/physrevlett.92.035504 (cit. on p. 80).

[185] K. Böröczky, *Über stabile Kreis- und Kugelsysteme*, Annales Universitatis Scientarium
Budapestinensis de Rolando Eötvös Nominatae Sectio Mathematica **7** (1964) 79,
URL: https://annalesm.elte.hu/archive.html, last accessed: 17 August 2023
(cit. on pp. 80, 91, 95–99, 114).

[186] B. J. Alder and T. E. Wainwright, *Phase Transition for a Hard Sphere System*,
The Journal of Chemical Physics **27** (1957) 1208,
URL: https://doi.org/10.1063/1.1743957 (cit. on pp. 84, 95).

[187] J. D. Hunter, *Matplotlib: A 2D graphics environment*,
Computing in Science & Engineering **9** (2007) 90,
URL: https://doi.org/10.1109/MCSE.2007.55 (cit. on p. 90).

[188] W. Humphrey, A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*,
Journal of Molecular Graphics **14** (1996) 33,
URL: https://doi.org/10.1016/0263-7855(96)00018-5 (cit. on p. 90).

[189] J. L. Lebowitz and O. Penrose, *Convergence of Virial Expansions*,
Journal of Mathematical Physics **5** (1964) 841,
URL: https://doi.org/10.1063/1.1704186 (cit. on p. 95).

[190] T. Helmuth, W. Perkins, and S. Petti, *Correlation decay for hard spheres via Markov chains*,
The Annals of Applied Probability **32** (2022) 2063,
URL: https://doi.org/10.1214/21-AAP1728 (cit. on pp. 95, 112).

[191] L. Fejes, *Über einen geometrischen Satz*, Mathematische Zeitschrift **46** (1940) 83,
URL: https://doi.org/10.1007/BF01181430 (cit. on p. 95).

[192] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, Springer, 1999
(cit. on p. 95).

[193] S. Torquato and F. H. Stillinger,
Jammed hard-particle packings: From Kepler to Bernal and beyond,
Reviews of Modern Physics **82** (2010) 2633,
URL: https://doi.org/10.1103/RevModPhys.82.2633 (cit. on p. 95).

[194] A. Donev, S. Torquato, F. H. Stillinger, and R. Connelly,
Jamming in hard sphere and disk packings, Journal of Applied Physics **95** (2004) 989,
URL: https://doi.org/10.1063/1.1633647 (cit. on p. 95).

[195] M. Kahle, *Sparse Locally-Jammed Disk Packings*, Annals of Combinatorics **16** (2012) 773,
URL: https://doi.org/10.1007/s00026-012-0159-0 (cit. on pp. 96–98, 101, 114).

[196] J. Pach and M. Sharir,
Combinatorial Geometry and Its Algorithmic Applications: The Alcalá Lectures,
American Mathematical Society, 2009 (cit. on p. 97).

[197] P. Diaconis, G. Lebeau, and L. Michel,
Geometric analysis for the Metropolis algorithm on Lipschitz domains,
Inventiones mathematicae **185** (2011) 239,
URL: https://doi.org/10.1007/s00222-010-0303-6 (cit. on p. 103).

[198] Y. Baryshnikov, P. Bubenik, and M. Kahle,
Min-Type Morse Theory for Configuration Spaces of Hard Spheres,
International Mathematics Research Notes **2014** (2013) 2577,
URL: https://doi.org/10.1093/imrn/rnt012 (cit. on p. 103).

[199] S. C. Kapfer and W. Krauth, *Sampling from a polytope and hard-disk Monte Carlo*,
Journal of Physics: Conference Series **454** (2013) 012031,
URL: https://doi.org/10.1088/1742-6596/454/1/012031 (cit. on pp. 104, 152).

[200] B. Li, S. Todo, A. C. Maggs, and W. Krauth,
Multithreaded event-chain Monte Carlo with local times,
Computer Physics Communications **261** (2021) 107702,
URL: https://doi.org/10.1016/j.cpc.2020.107702 (cit. on pp. 104, 128, 152).

[201] P. G. Bolhuis, D. Frenkel, S.-C. Mau, and D. A. Huse,
Entropy difference between crystal phases, Nature **388** (1997) 235,
URL: https://doi.org/10.1038/40779 (cit. on p. 112).

[202] R. Kannan, M. W. Mahoney, and R. Montenegro,
Rapid Mixing of Several Markov Chains for a Hard-Core Model,
International Symposium on Algorithms and Computation (2003) 663,
URL: https://doi.org/10.1007/978-3-540-24587-2_68 (cit. on p. 112).

[203] E. P. Bernard, C. Chanal, and W. Krauth, *Damage spreading and coupling in Markov chains*,
Europhysics Letters **92** (2010) 60004,
URL: https://doi.org/10.1209/0295-5075/92/60004 (cit. on p. 112).

[204] D. Ruelle, *Statistical Mechanics: Rigorous Results*, World Scientific, 1999 (cit. on p. 113).

[205] W. W. Wood, *Monte Carlo Calculations for Hard Disks in the Isothermal-Isobaric Ensemble*,
The Journal of Chemical Physics **48** (1968) 415,
URL: https://doi.org/10.1063/1.1667938 (cit. on p. 113).

[206] W. W. Wood, *NpT-Ensemble Monte Carlo Calculations for the Hard-Disk Fluid*,
The Journal of Chemical Physics **52** (1970) 729,
URL: https://doi.org/10.1063/1.1673047 (cit. on p. 113).

[207] J. Lee and K. J. Strandburg, *First-order melting transition of the hard-disk system*,
Physical Review B **46** (1992) 11190,
URL: https://doi.org/10.1103/physrevb.46.11190 (cit. on p. 113).

[208] E. Gamma, R. Helm, R. Johnson, and J. M. Vlissides,
Design Patterns: Elements of Reusable Object-Oriented Software, Addison–Wesley, 1994
(cit. on pp. 122, 128).

[209] C. L. Wennberg, T. Murtola, B. Hess, and E. Lindahl, *Lennard-Jones Lattice Summation in Bilayer Simulations Has Critical Effects on Surface Tension and Lipid Properties*,
Journal of Chemical Theory and Computation **9** (2013) 3527,
URL: https://doi.org/10.1021/ct400140n (cit. on p. 127).

[210] J. Kolafa and J. W. Perram,
Cutoff Errors in the Ewald Summation Formulae for Point Charge Systems,
Molecular Simulation **9** (1992) 351,
URL: https://doi.org/10.1080/08927029208049126 (cit. on pp. 131, 135).

[211] J. Bierkens, S. Grazzi, K. Kamatani, and G. Roberts, *The Boomerang Sampler*,
Proceedings of the 37th International Conference on Machine Learning, Proceedings of
Machine Learning Research **119** (2020) 908,
URL: https://proceedings.mlr.press/v119/bierkens20a.html,
last accessed: 17 August 2023 (cit. on pp. 132, 152).

[212] W. Shinoda, M. Shiga, and M. Mikami,
Rapid estimation of elastic constants by molecular dynamics simulation under constant stress,
Physical Review B **69** (2004) 134103,
URL: https://doi.org/10.1103/PhysRevB.69.134103 (cit. on pp. 135, 137).

[213] M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, and G. J. Martyna,
*A Liouville-operator derived measure-preserving integrator for molecular dynamics
simulations in the isothermal–isobaric ensemble*,
Journal of Physics A: Mathematical and General **39** (2006) 5629,
URL: https://doi.org/10.1088/0305-4470/39/19/S18 (cit. on pp. 135, 137).

[214] A. Brukhno, J. Grant, T. Underwood, K. Stratford, S. Parker, J. Purton, and N. Wilding,
DL_MONTE, version 2.07, 2020,
URL: https://gitlab.com/dl_monte/user-hub/-/wikis/home,
last accessed: 17 August 2023 (cit. on p. 135).

[215] A. V. Brukhno, J. Grant, T. L. Underwood, K. Stratford, S. C. Parker, J. A. Purton, and
N. B. Wilding, *DL_MONTE: a multipurpose code for Monte Carlo simulation*,
Molecular Simulation **47** (2021) 131,
URL: https://doi.org/10.1080/08927022.2019.1569760 (cit. on p. 135).

[216] M. Roberts, *How to evenly distribute points on a sphere more effectively than the canonical
Fibonacci Lattice*, 2020,
URL: https://extremelearning.com.au/how-to-evenly-distribute-points-on-
a-sphere-more-effectively-than-the-canonical-fibonacci-lattice/,
last accessed: 17 August 2023 (cit. on p. 145).

[217]   Y. Feng, *An effective energy-conserving contact modelling strategy for spherical harmonic particles represented by surface triangular meshes with automatic simplification*, Computer Methods in Applied Mechanics and Engineering **379** (2021) 113750, URL: https://doi.org/10.1016/j.cma.2021.113750 (cit. on p. 145).

[218]   D. P. Hardin, T. J. Michaels, and E. B. Saff, *A Comparison of Popular Point Configurations on $\mathbb{S}^2$*, arXiv: 1607.04590 [math.NA] (2016), URL: https://doi.org/10.48550/arXiv.1607.04590 (cit. on p. 145).

[219]   B. Keinert, M. Innmann, M. Sänger, and M. Stamminger, *Spherical Fibonacci Mapping*, ACM Transactions on Graphics **34** (2015) 193, URL: https://doi.org/10.1145/2816795.2818131 (cit. on p. 145).

[220]   S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Physics Letters B **195** (1987) 216, URL: https://doi.org/10.1016/0370-2693(87)91197-X (cit. on p. 151).

[221]   M. Betancourt, *A Conceptual Introduction to Hamiltonian Monte Carlo*, arXiv: 1701.02434 [stat.ME] (2018), URL: https://doi.org/10.48550/arXiv.1701.02434 (cit. on p. 151).

[222]   S. Prokhorenko, K. Kalke, Y. Nahas, and L. Bellaiche, *Large scale hybrid Monte Carlo simulations for structure and property prediction*, npj Computational Materials **4** (2018) 80, URL: https://doi.org/10.1038/s41524-018-0137-0 (cit. on p. 151).

[223]   T. Guyon, A. Guillin, and M. Michel, *Necessary and sufficient symmetries in Event-Chain Monte Carlo with generalized flows and Application to hard dimers*, arXiv: 2307.02341 [cond-mat.stat-mech] (2023), URL: https://doi.org/10.48550/arXiv.2307.02341 (cit. on p. 152).

[224]   G. S. Heffelfinger, *Parallel atomistic simulations*, Computer Physics Communications **128** (2000) 219, URL: https://doi.org/10.1016/S0010-4655(00)00050-3 (cit. on p. 152).

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Hartmut and Werner. I greatly appreciate that both of you always supported, even encouraged the long-lasting collaboration between Bonn and Paris for the research projects underlying this doctoral thesis. Moreover, both of you always believed in my capabilities in academic research (maybe even more than I did) what finally gave me the courage to try to stay in academia even after this doctoral thesis. I hope to see you both in New York!

Hartmut, ever since you initiated the contact with Werner, you understood that I found a research topic that captivated me completely. If I would have ever gotten enough of that subject at some point in time, I am entirely sure that you would have offered me all of your support and insight until I would have found another topic to be enthusiastic about. At the same time, you saw my desire to continue the long-term project of this doctoral thesis and never even tried to change my mind. This fundamental view of research, which encourages independence and individualism, inspires me. On top of that, I am glad that I had the opportunity to be the assistant for your lectures already during my doctoral studies. Although I certainly complained about the amount of work back then, in hindsight, I greatly appreciate the experience and realize that I learned a lot about your view of physics. Besides that, the weekly lunch meetings for burgers (until they unfortunately remembered to remove their unbeatable lunch offer) and our frequent chats were always a welcome variety in my days and additionally taught me a lot of facts outside of physics. Now, with this doctoral thesis finally finished, we should have that dinner. I look forward to it!

Werner, your enthusiasm about non-reversible Markov chains truly was contagious. Ever since you directly replied to my mail asking for a small research project in 2018, I never stopped thinking about them. You taught me everything I know about this topic and always when I thought I finally understood it, you surprised me with a new insight. I am deeply grateful that you never questioned my decision to stay in Bonn for my doctoral studies and even then, never doubted that we can successfully continue the research presented in this doctoral thesis. During our research, I always appreciated that you really cared about my opinions and insights. Your unshakable faith in the power of non-reversible Markov chains, especially during the initial part of my doctoral studies when we only saw "piggies," inspires me. It also taught me how to regard setbacks in research as an opportunity. On top of that, I am very glad that you facilitated repeated visits in Paris during which you introduced me to great food, great coffee, and great wine. By the way, when is the next salon des vins des vignerons indépendants?

I would also like to thank my collaborators Liang, Botao, Nicolas, and Tony with whom I had the pleasure of working together in the research projects underlying this doctoral thesis. A special thank you goes to Tony, who also supported me in my final decision to stay in academia.

A big thank you to my office mate David for the fun times during the years of our doctoral studies and for your comments on my thesis.

I acknowledge support from the Studienstiftung des deutschen Volkes and from Institut Philippe Meyer.

Last but not least, I reserve a special place of gratitude to my mother Nana, my father Helge, my second father Karsten, and my brothers Niklas and Jan. You supported me in every step on my way. Although it appears like I keep on moving further and further away from you, be assured that it has nothing to do with you! I love all of you.

Joana, thank you for proofreading this entire doctoral thesis. We went through our entire studies together, and now we finish it together (although you were indeed faster). Handing in this doctoral thesis is certainly an important milestone. It is, however, not even close to the most important event of this year: my marriage to you. My words from that day stay true: *Nur Du*.

(Finally, in the name of Joana and myself, we thank our cats.)