Irreversible Markov-chains for particle systems and spin models: Mixing and dynamical scaling

Ze Lei

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École Normale Supérieure de Paris
au Laboratoire de Physique Statistique

Irreversible Markov Chains for Particle Systems and Spin Models: Mixing and Dynamical Scaling

École doctorale n° 564

Physique en Île de France EDPIF

Spécialité : Physique

COMPOSITION DU JURY :

M^me^ Martine BENAMAR
ENS Paris, Examinatrice

M. Werner KRAUTH
ENS Paris, Directeur de thèse

M. Tony LELIÈVRE
École des Ponts ParisTech, Rapporteur

M. Kirone MALLICK
CEA Saclay, Examinateur

M. Clément SIRE
Université Toulouse III, Rapporteur

M^me^ Cristina TONINELLI
Université Paris-Dauphine, Examinatrice,
Présidente du jury

Soutenue par Ze LEI
le 19 décembre 2018

Dirigée par Werner KRAUTH

ENS ÉCOLE NORMALE SUPÉRIEURE

PSL RESEARCH UNIVERSITY PARIS

Département de Physique
École normale supérieure

THÈSE DE DOCTORAT
DE L'ÉCOLE NORMALE SUPÉRIEURE

Spécialité : Physique

École doctorale n°564: Physique en Île-de-France

réalisée

au Laboratoire de Physique Statistique

sous la direction de Werner KRAUTH

présentée par

Ze Lei

pour obtenir le grade de :

DOCTEUR DE L'ÉCOLE NORMALE SUPÉRIEURE

Sujet de la thèse :

Irreversible Markov Chains for Particle Systems and Spin Models: Mixing and Dynamical Scaling

soutenue le 19 décembre 2018

devant le jury composé de :

M^me	Martine BENAMAR,	Examinatrice
M.	Werner KRAUTH,	Directeur de thèse
M.	Tony LELIÈVRE,	Rapporteur
M.	Kirone MALLICK,	Examinateur
M.	Clément SIRE,	Rapporteur
M^me	Cristina TONINELLI,	Présidente du jury

Acknowledgement

Prima facie, I would like to express my sincere gratitude to my supervisor Prof. Werner Krauth, for his patience, motivation, immense knowledge, and for his guidance and support since I entered ENS, more specifically in the past three years of my Ph.D study. I also would like to thank Prof. Florent Krzakala for his "parrainage", Prof. Henk Hilhorst for being my "tuteur" of this thesis. I thank the other committee members of my thesis: Prof. Tony Leliévre, Prof. Clément Sire, Prof. Martine Benamar, Prof. Cristina Toninelli and Prof. Kirone Mallick, for their insightful comments and questions prompting me to widen my research and to incorporate various new perspectives.

For fruitful collaborations and stimulating discussions, I thank Prof. Anthony C. Maggs. I also thank Prof. Youjin Deng for the discussions of the XY and Harmonic models, Prof. Masaharu Isobe in the exploration of the two-dimensional relaxation dynamics, Prof. Florent Krzakala (again) and Dr. Sebastian Kapfer for the discussions of mixing in one-dimensional hard-sphere model. It has been my pleasure to work with them.

My sincere thanks also go to the professors having taught and instructed me in the past years. During the years of undergraduate at Peking University, I was fortunate to be guided by Prof. Shulin Zhang and Prof. Tieshuan Fan in research of physics. After my arriving in ENS, it was my pleasure to listen to the courses exhibiting various mechanisms and fascinating phenomena of physics by Prof. Alain Comet, Frédéric van Wijland, Rémi Monasson, Jean-Michel Raimond, Frédéric Chevy, Emmanuel Trizac, Philippe Lecheminant, Vincent Croquette and many other professors. I would like to thank them, also for taking special care of me as the only student who didn't understand French in the early days.

I would like to thank Prof. Matthias Troyer, Dr. Troels Frimodt Rønnow and Dr. Lei Wang for providing me the opportunity to join their team as intern and giving me the access to "Brutus" in Eidgenössische Technische Hochschule Zürich. I also thank Prof. Marcello Civelli and Marcelo Rozenberg for the internship in l'Université Paris-Sud, Dr. Óscar Nájera Ocampo and Dr. Yunlong Lian for the discussions during the internship. These were precious experiences that led me to my thesis today.

For the days of PhD studies in the laboratoire de physique statistique of ENS, I thank Prof. Jorge Kurchan as the director of the laboratory, Annie Ribaudeau, Nora Sadoui, Fabienne Renia and Benoît Paulet in the secretary office, Frédéric Desprez in the technical office, Prof. Jean-François Allemand and Laura Ledez in École doctorale PIF, and M. Auguste Filippi in ENS. With their help I (and also other members in

the laboratory) am able to concentrate on scientific research without being disturbed by the administrative affairs and technical problems. I also thank Prof. Roderich Moessner in Max-Planck-Institut für Physik komplexer Systeme for the kind invitation; though the stay was short, it was a nice and eye-opening experience in Dresden and Berlin.

I would like to thank my fellow group members: Manon Michel and Tommaso Comparin, for being the charismatic "leaders" of the group members; Juliane Uta Klamser, for taking over their role and "making the group united again"; Liang Qin, for providing valuable suggestions (of restaurants); Yoshihiko Nishkawa, for the useful discussions on physics, culture, music. I also thank my fellow colleagues in the laboratoire de physique statistique for creating the good atmosphere: Arthur Prat-Carrabin, Quentin Feltgen, Anirudh Kulkarni, Samuel Poincloux, Lorenzo Posani, Louis Brézin, Diego Contreras, Marco Molari, Tridib Sadhu, Volker Pernice and many others. More specially I would like to thank Félix Werner and Kris van Houcke, for their valuable advice since my master studies in ENS.

I would also like to thank my friends since undergraduate (who were once and who are now) in Paris. In the past seven years I have benefited a lot from the discussions with my friends in mathematics, and my special thanks go to Shen Lin, Daxin Xu, Zhiyuan Zhang, Shengquan Xiang, Shuoqing Deng, Jianyu Xu. For my friends in physics, I thank Bin Xu, the "big brother" guiding me to ENS, and Yongfeng Zhao, for the precious discussions since undergraduate. I also thank Shanshan Lü and Rui Jiao for helping me with my French. Thanks to my friends in "physique-ENS promo 2011": Pierre Fromholz, Bizi Nadir, Virgile Andreani, Federico Mogavero, and many others, for the help during my licence and master studies in ENS.
Again, I would like to thank Virgile and Rui for the correcting the abstract in French, Yoshihiko and Yongfeng for reading my thesis, providing valuable opinions and pointing out mistakes, and Werner for revising and finalizing this manuscript.

Finally, thanks to my parents for their support throughout my studies.

General Introduction

Collective behavior is universal in nature: from molecules to cells, from tissues to organs, from one bird to a flock of birds, from a single node to a network, simple interactions can give rise to complex system behaviors; likewise, slight changes of conditions can lead to different phenomena. These phenomena can often be well modeled by statistical mechanics and they are open to mathematical analysis. In some cases where the analytical approaches fail, computational methods and numerics can provide the missing pieces of knowledge. Computation has been applied to many of the aforementioned, such as non-linear dynamics, high-dimensional integration, etc.

The Monte Carlo method uses random samplings to approach the "actual" probability distribution or an integration of the observable function in high dimensions, via simulations and experiments. The throwing pins experiment by Comte de Buffon in the evaluation of π [1] is the earliest Monte Carlo calculation on record. The idea of modern Monte Carlo method using random numbers generators and computers comes from J. von Neumann and S. Ulam [2], and proceeded by S. Ulam, N. Metropolis, et al. [3, 4] This method is widely applied in modern research, including physics [5], chemistry [6], biology [7], computer science [8] and many more. Theories and applications of the Monte Carlo methods have been progressing since some 70 years, and they still provide us with surprises, such as the irreversible Markov chains presented in this thesis.

Since its invention in 2009 [9], the "event-chain" Monte Carlo (ECMC) has proven to be a general method based on irreversible Markov chains. It was successfully extended from hard spheres [9] to general interaction potentials for particle systems [10] and for spin models [11, 12]. It has been instrumental in research on melting in two dimensions (both for the hard-sphere models [9, 13, 14] and for soft potentials [15]). It is now used for the extremely efficient simulation of particle systems with long-range interactions [16]. For the SPC/Fw (simple point charge with flexible variant) water model, a staple in molecular simulation, it is currently being benchmarked against the molecular-dynamics code (LAMMPS), that has been developed over decades.

This PhD thesis concerns the different dynamical properties of the event-chain algorithm in different physical systems, and attempts to optimize and unify the dynamics. It reviews the work during my PhD studies at Laboratoire de Physique Statistique since September 2015.

Chapter 1 reviews some basics of probability theory, Markov chains, and of statistical physics. In chapter 2, based on the general "global-balance" condition of Markov chains, we discuss a particular irreversible Markov chain using the concepts of "lifting" and replicas, together with the "factorized Metropolis filter" for continu-

ous Boltzmann-distributed systems. These concepts contribute to a full construction of the "event-chain" algorithm.

Chapter 3 summarizes the Publication II [17]. It studies irreversible Markov chains for the one-dimensional hard-sphere model, based on the previous framework of classification of algorithms, including the TASEP and the lifted TASEP [18]. We obtain an exact result for the mixing time in a continuous case, and relate this dynamics to the coupon-collector problem [19]. It explains the mixing property of ECMC of previous findings [18]. This chapter furthermore proposes a "particle-swap" approach, which renders possible the sequential lifted Metropolis algorithm and sequential "event-chain" algorithm. These two algorithms accelerate mixing further, and this approach remains valid in higher dimensions.

Chapter 4 summarizes the Publication I [12]. It presents the dynamics of the "event-chain" algorithm with the factorized Metropolis filter for continuous spin models in the presence of topological excitations, based on the previous work [11]. The local nature of the Markov-chain dynamics leads to a slow "vortex" mode and a fast "spin-wave" mode in the two-dimensional XY model. We identify the probability distribution of the maximum vortex-pair distance as a Fréchet distribution, with a shape parameter α determined by the temperature. Other topological defects such as the monopoles in the three-dimensional Heisenberg model are also described. With a good understanding that the ECMC reaches equilibrium with an exponent $z = 0$ for the spin-wave mode (that has $z = 2$ in local reversible MC simulations), we expect this result to carry over to the relaxation of phonon modes in higher-dimensional particle systems.

Chapter 5 presents the submitted Publication III [20]. It explains the reason that ECMC has a dynamics scaling than other local reversible Markov chains (and even faster than molecular dynamics), and proposes an optimization of the event-chain algorithm for general particle models, by introducing a "factor field". Simulations on the one-dimensional Lennard-Jones model indicate a much improved value of the dynamical scaling exponent for the autocorrelation time, and a super-diffusive motion of "lifting index" in the "event-chain" algorithm for general one-dimensional models. It may help explain the high speed of dynamics and lead to the design of new algorithms.

Contents

Chapter 1

Probability Theory and Markov Chains

Markov chains, a special type of stochastic processes, were first studied by A. Markov in 1906, and then named after him [21]. Einstein's 1905 work on diffusion, which applied a similar idea to the Brownian motion, attracted the attention of physicists and statisticians almost at the same time [22]. Tens of years later, A. Kolmogorov established the axiomatic system of probability theory (including the continuous-time theory of Markov process) as a solid foundation of statistics. [23]. Since then, many other mathematicians, including M. Fréchet and P. Lévy, contributed to this field and extended the theory of stochastic processes.

In this chapter, we will start from some aspects of probability theory and the theory of Markov chains, focus on their convergence and relaxation, discuss a few mathematical problems, and relate them to the dynamics and numerical algorithms in statistical physics.

1.1 Some Aspects of Probabilities

The idea of probability originates from the uncertainty in the world with unknown description (though quantum physics tells us the probability is an essential property of any matter, which we will not investigate in detail). The mathematical theory of probability was initiated by P. Laplace in the 19th century [24] and built into an axiom system by A. N. Kolmogorov in 1933 [23].

Thermodynamics and statistical mechanics were developed in the 19th century, before the foundation of the axiom system of probability theory. Nowadays, statistical physics is largely based on probability theory, nevertheless the terminology of mathematics is sometimes ambiguous to physicists due to historical reasons, and the arbitrary descriptions in physics confuse mathematicians on the other hand.

In this section, we will provide simple definitions of some mathematical terms[1], followed by examples of physical models related to the work in the thesis as illustra-

[1] We refer to *Probability I* by A. N. Shiryaev [25].

tions. We only collect and present the mathematical statements needed in the remainder of the thesis, without the goal of completeness or full rigor.

1.1.1 Probability Space and Random Variables

The *sample space* (or *event space*) U of a finite experiment or a random trial is the set of all possible outcomes or results of that experiment or trial. The elements inside the sample space have to be mutually exclusive and collectively exhaustive, as no two results will occur in one single trial, and after a trial there must be one result found in U. (It is usually denoted by S, U or Ω. Most physicists use Ω for the "state space" which will be explained later, so we pick U, indicating "universe".)

For example, there are two results from throwing a coin once: the head and the tail. The sample space is:

$$U = \{H, T\}.$$

In throwing a coin n times, or throwing n distinguishable coins all at once, the sample space is:

$$U = \{H, T\}^n.$$

If we assume there are n magnetic coins with interaction, and they are thrown t times (for physicists, this corresponds to the simulation of Ising model, which is introduced in section 4.1.1), the sample space is

$$U = \{H, T\}^{nt}.$$

As we see, the sample space depends on the definition of the "experiment": it can describe an experiment that happens once, as well as the one that repeats multiple times, regardless of the connection between the results of different times. In the following, unless otherwise stated, the sample space is always considered as the results from all the experiments.

An *event* is a subset of the sample space. It can be restricted by rules, such as "heads appear three times more than the tails do", "only one head in the sixth test", or can be even more arbitrary by selecting none or one or more elements from the sample space. For a more precise definition of events, *σ-algebra*, an algebraic structure on the sample space U, is needed. It is closed under *complement*, *countable unions* and *countable intersections*, in other words, if $\{A_i\}$ is a set of countable events, then $U \backslash A_i, (\cup A_i)$ and $(\cap A_i)$ are all events as well. (For example, the power set of a finite U, namely 2^U, is a σ-algebra, and also the largest one, having all the other σ-algebra as its subsets.)

Now we can define "probability". Mathematically it is a *measure* (using the measure theory), which indicates a function $\mu : \mathcal{E}(U) \to \mathbb{R}$ (here $\mathcal{E}(U)$ is the set of all possible events, or the corresponding σ-algebra of U) with the following properties:

- Non-negativity: $\forall S \in \mathcal{E}(U), \mu(S) \geq 0$;
- Null empty set: $\mu(\emptyset) = 0$;
- Countable additivity: For countable collections of $\{S_i\}$ that $\forall i, j : S_i \cap S_j = \emptyset$, we have $\mu(\cup S_i) = \sum \mu(S_i)$.

The probability function needs a *normalization* other than these three points, that $\mu(U) = 1$. Hence, one defines the *probability* as the function $\mathbf{P} : \mathcal{E}(U) \to [0,1]$. The sample space U, the σ-algebra $\mathcal{E}(U)$ splitting events and the probability measure $\mathbf{P}$, they construct the well-defined *probability space* or *probability triple*.

A *random variable* (or *stochastic variable*, in the context of a stochastic process) is a mapping:

$$\xi : U \to \xi(U),$$

where the domain is the sample space U and the range $\xi(U)$ is a subset of $\mathbb{R}$. If $\xi(U)$ is discrete, the *(discrete) probability distribution* of a random variable ξ is a mapping $P_\xi : \xi(U) \to \mathbb{R}$:

$$P_\xi(x) = \mathbf{P}\{\omega : \xi(\omega) = x\}. \tag{1.1.1}$$

And this mapping can also be extended to the subset of $\xi(U)$:

$$P_\xi(A) = \mathbf{P}\{\omega : \xi(\omega) \in A\}. \tag{1.1.2}$$

The *cumulative distribution function* (or *distribution function*) $F_\xi : \xi(U) \to [0,1]$ is defined as:[2]

$$F_\xi(x) = \mathbf{P}\{\omega : \xi(\omega) \leq x\}, \tag{1.1.3}$$

and the *probability density function* is defined as the first derivative of the cumulative distribution function, if $F_\xi(x)$ is "absolutely continuous" [25]

$$f_\xi(x) = \frac{\mathrm{d}F_\xi(x)}{\mathrm{d}x}. \tag{1.1.4}$$

In some occasions concerning a continuous distribution, $f_\xi(x)$ is written as $p_\xi(x)$ (or even as $p(x)$). The probability distribution $P_\xi(x)$ is used for discrete ξ and the probability density function $f_\xi(x)$ is used for continuous ξ. In some cases they can be connected with Dirac's δ function. In the limit of continuous U,

$$\sum_{x \in \xi(U)} P_\xi(x) O(x) \sim \int_{\xi(U)} f_\xi(x) O(x) \mathrm{d}x \sim \mathbb{E}[O]$$

for a continuous function $O(x)$ well-defined on $\xi(U)$ (and this term $\mathbb{E}[O]$ is defined as the *expectation* of the function/random variable O). We will base most mathematical discussions on the discrete U in the following sections, and one may derive the corresponding equations in the continuous case.

The σ-algebra and the measure construct important parts in the axiomatic system of probability theory and guarantee its consistence. Nevertheless, we will not discuss anymore about them in the remainder of the thesis, as the sample space of most physical models is either a finite discrete one or a continuous Lebesgue measurable subspace of $\mathbb{R}^n$, with well-defined measure.

[2] Following the formalism of probability theory, the distribution function is introduced before defining the measure (Lebesgue), and the discrete probability distribution is a special case [25]. Here we state the facts in a more comprehensible way (which may not be rigorous in mathematics).

In physics, the term "*observable*" has a similar property as "random variables" in mathematics, in the context of "one experiment". For example in one experiment of throwing multiple magnetic coins, the difference between the numbers of heads and tails of one test is an observable (random variable). (The physical analogy is the magnetization of the Ising model, as mentioned in section 4.1.3.) More generally, some physical observables are n-vectors, which be regarded as n (dependent) random variables. Most observables are functions of one or multiple *configurations*, which corresponds to an element in the sample space of a single or multiple tests. We will continue the discussion in section 1.2.1.

1.1.2 Stability and Max-stability

Stability in mathematics is usually a property of being invariant under some non-trivial transformations. In probability theory, a *stable distribution* is invariant under linear combinations and a *max-stable distribution* is invariant under maximum selections.

The stability of distributions leads to the central limit theorem in statistics. It also characterizes the behavior of the random walk, which is discussed in section 1.5.3 and then applied in section 5.3. Max-stability, describing the properties of extreme events, is applied in our discussions in sections 1.5.1 and 4.3.2.

We assemble them in this section, for these two theories share many similarities in the definitions, the statements of theorems, and their significance within probability theory. [3]

Partial Sums and Stability

Definition of *stable distributions:* For a sequence of independent and identically distributed random variables $\{X_i\}, i = 1, 2, \ldots$, we assume they have the same distribution as X. X follows a *stable* distribution if there exist two sequences of real numbers $\{c_i\}, (c_i > 0)$ and $\{d_i\}$ that:

$$\mathbf{P}(\frac{S_n - d_n}{c_n} = x) = \mathbf{P}(X = x), \tag{1.1.5}$$

where

$$S_n = \sum_{i=1}^{n} X_i,$$

is known as the *partial sum* of the sequence $\{X_1, ..., X_n, ...\}$.

The *characteristic function* of a random variable ξ is defined as:

$$\varphi_\xi(t) = \int_{\mathbb{R}} \mathrm{e}^{itx} f_\xi(x) \mathrm{d}x = \mathbb{E}(\mathrm{e}^{it\xi}). \tag{1.1.6}$$

[3] In this section, we refer to *Limit Distributions for Sums of Independent Random Variables* by B. V. Gnedenko and A. N. Kolmogorov [26, chapter 6] for the discussion of the partial sums and the stability (with some modification of the notation), and refer to *Extreme Value Theory: An Introduction* by L. de Haan [27, chapter 1] for the discussion of the partial maxima and the max-stability.

For the sum of independent variables $\Xi = \sum_{i=1}^{n} \xi_i$:

$$\varphi_{\Xi}(t) = \mathbb{E}(\mathrm{e}^{it\Xi}) = \prod_{i=1}^{n} \mathbb{E}(\mathrm{e}^{it\xi_i}) = \prod_{i=1}^{n} \varphi_{\xi_i}(t). \tag{1.1.7}$$

The distributions of the ξ_i need not be identical as long as they are independent. This result is very useful in the later discussion of single-particle sampling in section 3.2.1.

For a stable distribution, its characteristic function should be somehow "invariant" under multiplications with itself. The Gaussian distribution and the Cauchy distribution are well-known examples of stable distributions. The class of all stable distributions will be provided in the following.

Generalized Central Limit Theorem

The *Lévy–Khinchin Representation* [28, 29] tells that all stable distributions have their characteristic functions as

$$\phi(t;\alpha,\beta,c,\mu) = \mathrm{e}^{it\mu - |ct|^{\alpha}(1 - i\beta \mathrm{sign}(t)\Phi(\alpha,t))}, \tag{1.1.8}$$

where $\beta \in [-1,1]$ (as *skewness parameter*), $\alpha \in (0,2]$, μ is the shift, c provides the scale, and

$$\Phi(\alpha,t) = \begin{cases} \tan(\frac{\pi\alpha}{2}) & \text{if } \alpha \neq 1, \\ -\frac{2}{\pi}\log|t| & \text{if } \alpha = 1. \end{cases} \tag{1.1.9}$$

The family of distributions is called *Lévy alpha-stable distribution family*.

In eq. (1.1.5), the scaling factors c_n is $n^{1/\alpha}$ in Lévy alpha-stable distributions. This leads us to the *generalized central limit theorem*, which describes the partial average of a sequence of N copies of a variable X:

- If the distribution of X has a finite variance, the partial averages converge to a Gaussian distribution ($\mathcal{N}(\mu,\sigma)$) as $\alpha = 2$, with the standard deviation $\sigma \sim N^{-\frac{1}{2}}$, which is the classical central limit theorem;
- If the distribution of X has a heavy tail, that $\lim_{x\to\infty} p(x) \sim \frac{1}{|x|^{1+\alpha}}, 1 < \alpha \leq 2$), it scales as the corresponding Lévy α distribution, with the standard deviation $\sigma \sim N^{\frac{1}{\alpha}-1}$, and converges to the true mean value;
- If the distribution of X has a heavy tail, that $\lim_{|x|\to\infty} p(x) \sim \frac{1}{|x|^{1+\alpha}}$, $\alpha \leq 1$, it scales as the corresponding α distribution, but it does not converge to the "mean" (in the sense of Cauchy principal value), with a diverging standard deviation.

These stable distributions provide the basis for stable processes, which we will present in section 1.5.3 and use in section 5.3.

Partial Maxima and Max-stability

Definition of *max-stable distributions:* For a sequence of independent and identically distributed random variables $\{X_i\}, i = 1,2,\ldots$, we assume they have the same distribution as X. X follows a *max-stable* distribution if there exist two sequences of real

numbers $\{a_i\}, (a_i > 0)$ and $\{b_i\}$ that:

$$\mathbf{P}(\frac{D_n - b_n}{a_n} \leq x) = \mathbf{P}(X \leq x), \tag{1.1.10}$$

where

$$D_n = \max_{i=1,...,n}(\{X_i\}),$$

is the *partial maximum* of a sequence $\{X_1, ..., X_n, ...\}$.

Similar to the treatment with partial sums using the characteristic functions, a cumulative distribution function itself shows the stable behavior:

$$\mathbf{P}(\max_n\{X_i\} \leq x) = \prod_i^n \mathbf{P}(X_i \leq x). \tag{1.1.11}$$

All possible non-trivial max-stable distributions form a class, which will be discussed in the following.

Fisher–Tippett–Gnedenko Theorem

For a sequence of independent and identically distributed random variables $\{X_i\}, i = 1, 2, \ldots$, we assume they have the same distribution as X, and the cumulative distribution function of X is $F(x)$. If there exist two sequences of real numbers $\{a_i\}$ $(a_i > 0)$ and $\{b_i\}$, such that the regulated variable $\frac{\max_n(\{X_i\})-b_n}{a_n}$ has a non-degenerate limit distribution as $n \to \infty$:

$$\lim_{n\to\infty}[F(a_n x + b_n)]^n = G(x). \tag{1.1.12}$$

Then all the possible distributions with the cumulative distribution function as $G(x)$ are called the *extreme value distribution*[4], which plays the same role as the Lévy alpha-stable distribution. The following theorem has the equivalent status in extreme-value theory as the central limit theorem in statistics.

Fisher–Tippett–Gnedenko Theorem (M. Fréchet (1927) [30], R. Fisher and L. Tippett (1928) [31], B. V. Gnedenko (1943) [32]): The family of all extreme value distributions is in the form of $G_\gamma(ax + b)$, where $a > 0, b$ are real parameters, and

$$G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right). \tag{1.1.13}$$

When $\gamma = 0$, the right-hand side can take the limit as $G_0(x) = \exp(-\mathrm{e}^{-x})$. The class of distributions is divided into three types according to the value of γ (in the following three cases, μ is the shift factor, s is the scale factor and $\alpha = 1/\gamma$):

1. $\gamma > 0$, known as the *Fréchet distribution*, is usually written as

$$P(\mu < X < x) = \exp\left[-\left(\frac{x-\mu}{s}\right)^{-\alpha}\right]. \tag{1.1.14}$$

   It is bounded on the left side ($x = \mu$), and it has a heavy tail ($p(x) = \mathcal{O}(x^{-1-\alpha})$) on the right side. We will see a physical example in section 4.3.2.

[4] This name is used in *Extreme Value Theory: An Introduction* by L. de Haan [27].

2. $\gamma = 0$, known as the *Gumbel distribution*, is usually written as

$$P(X < x) = \mathrm{e}^{-\mathrm{e}^{-(x-\mu)/s}}. \tag{1.1.15}$$

It has a double-exponential shape, with light tails ($p(x) = \mathcal{O}(\mathrm{e}^{-|x|/s})$) on both sides. For example, the largest element of a sequence of exponentially distributed random variables follows a Gumbel distribution with $\gamma = 0$, in the limit of an infinite sequence (which we use in section 1.5.1).

3. $\gamma < 0$, known as the *reverse-Weibull distribution*, is usually written as

$$P(X < x) = \mathrm{e}^{-(-\frac{x-\mu}{s})^{\alpha}}. \tag{1.1.16}$$

It is bounded on the right side ($x = \mu$), and it has a light tail on the left side ($p(x) = \mathcal{O}(\mathrm{e}^{-(-x)/s})$). For example, the largest element of a sequence of uniformly distributed random variables follows a reverse-Weibull distribution with $\gamma = -1$, in the limit of an infinite sequence.

Different values of γ imply different scaling behaviors, as the scaling factor $a_n = n^{\gamma}$ in eq. (1.1.10). ($\mathbb{E}(X) \propto b_n$ and $\mathrm{var}(X) \propto a_n^2 = n^{2\gamma}$, if they are well defined.)

1.2 Markov Chains and Master Equations

1.2.1 Stochastic Processes and Markov Chains

For two events A and B, the *conditional probability* of B given A is

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)},$$

where $(B|A)$ is interpreted as the event B on condition of A. If $\mathbf{P}(A) = 0$, then $\mathbf{P}(B|A) = 0$ as well. Considering two random variables X and Y instead of events A and B, the *conditional probability distribution* of Y given the occurrence of the value x of X is

$$\mathbf{P}(Y = y|X = x) = P_Y(y|X = x) = \frac{\mathbf{P}(Y = y \cap X = x)}{\mathbf{P}(X = x)}, \tag{1.2.1}$$

and the *marginal probability distribution* of Y is

$$\mathbf{P}(Y = y) = P_Y(y) = \sum_{x \in \Omega_x} \mathbf{P}(Y = y \cap X = x) = \sum_{x \in \Omega_x} \mathbf{P}(Y = y|X = x)\mathbf{P}(X = x). \tag{1.2.2}$$

These are useful for a time-ordered stochastic process.

A *stochastic process* is defined as an *ordered* sequence of random variables, e.g. a random variables sequence $(\xi_1, \xi_2, .., \xi_n)$. There are some special processes that are usually discussed in mathematics:

- If

$$\mathbb{E}(\xi_{k+1}|\xi_1, \xi_2, ...\xi_k) = \xi_k, \tag{1.2.3}$$

then the sequence $\{\xi_i\}$ is a *martingale*;

- If
$$\mathbf{P}(\xi_{k+1} = a_{k+1}|\xi_1, \xi_2, ..., \xi_k) = \mathbf{P}(\xi_{k+1} = a_{k+1}|\xi_k), \tag{1.2.4}$$
then the sequence $\{\xi_i\}$ is a *Markov chain*.

The concept of a martingale leads to the *optional stopping theorem* [33], which determines the mean value of a martingale at an arbitrary bounded stopping time. This theorem is useful in calculations, and we will use it in section 1.5.1. (We present the "martingale" in the thesis only to facilitate some calculations, and it is irrelevant to the other chapters.)

In a stochastic process, its future states are usually conditional on both past and present states. A Markov chain holds the Markov property that the conditional probability distribution of future states depends only upon the present state, meaning that the process is memoryless of all its previous history.

For a Markov chain $(\xi_1, \xi_2, \ldots, \xi_n)$, we may associate it to a time series $(t_1, t_2, \ldots, t_n)$ $(t_1 < t_2 < ... < t_n)$. Characterizing the random variable by time $(\xi(t_1) = \xi_1)$ facilitates the descriptions of a continuous process. The domain of each random variable ξ_i or $\xi(t_i)$ (which is denoted by the set Ω), is named *state space* (if Ω is discrete) or *phase space* (if Ω is continuous). If there are no other events, the sample space can be written as Ω^n. It can be extended for an infinite sequence of $\{t_i\}$, or generalized for $t \in [0, +\infty)$. In physical processes, such random variables ξ_i are also called *observables*.

In the following, we will write the probability distribution as "$P(x_i)$" or "$P(x(t_i))$" instead of $\mathbf{P}(\xi(t_i) = x_i)$ or $P_{\xi(t_i)}(x_i)$, if it cannot be confused with the probability density functions. Hence, eq. (1.2.4) is rewritten as

$$P(x_n|x_1, x_2, \ldots, x_{n-1}) = P(x_n|x_{n-1}). \tag{1.2.5}$$

Following the definition of the conditional probability distribution and that of the marginal distribution, we have

$$\begin{aligned} P(x_3, x_1) &= \sum_{x_2 \in \Omega} P(x_3, x_2, x_1) \\ &= \sum_{x_2 \in \Omega} P(x_3|x_2, x_1) P(x_2|x_1) P(x_1) \\ &= \sum_{x_2 \in \Omega} P(x_3|x_2) P(x_2|x_1) P(x_1). \end{aligned} \tag{1.2.6}$$

By dividing $P(x_1)$ on both sides:

$$P(x_3|x_1) = \sum_{x_2 \in \Omega} P(x_3|x_2) P(x_2|x_1), \tag{1.2.7}$$

which is known as the *Chapman–Kolmogorov Equation*, characterizing the master equation of a Markov chain.

In a model of an experiment that forms a Markov chain, we start from the initial distribution $p_0(x) = \mathbf{P}(\xi_1 = x)$ and evolve with $P(x_k, x_{k-1}) = \mathbf{P}(\xi_k = x_k|\xi_{k-1} = x_{k-1})$, known as the matrix of transition probabilities at time k (or t_k). If it is k-independent

(or it is only dependent of the time interval $t_k - t_{k-1}$), it is called the *transition matrix*[5], and this Markov chain is called a *homogeneous* Markov chain.

1.2.2 Transition Matrix and Master Equation

The transition matrix in a homogeneous Markov chain is denoted by:

$$P(x, t_3|y, t_2) = T_{x,y}(t_3 - t_2). \tag{1.2.8}$$

Homogeneity is a natural prerequisite for the physical systems with *time translation symmetry*. In the form of T, the Chapman–Kolmogorov Equation eq. (1.2.7) is:

$$T_{x_1,x_3}(t_3 - t_1) = \sum_{x_2 \in \Omega} T_{x_1,x_2}(t_2 - t_1) T_{x_2,x_3}(t_3 - t_2). \tag{1.2.9}$$

By taking $\{x_i\}$ as their indices, eq. (1.2.9) matches the law of matrix multiplication; in matrix form, it is equivalent to:

$$\mathbf{T}(\tau_1 + \tau_2) = \mathbf{T}(\tau_1) \cdot \mathbf{T}(\tau_2). \tag{1.2.10}$$

In this formalism, a probability distribution of all possible states forms a vector $\boldsymbol{\pi}$

$$\boldsymbol{\pi} = (\pi(x_1), \pi(x_2), \cdots),$$

and the transition matrix acts as a right operator on it:

$$\boldsymbol{\pi}(t + \tau) = \boldsymbol{\pi}(t)\mathbf{T}(\tau). \tag{1.2.11}$$

For Markov chains in discrete time (we may pick the unit time interval as "1", and use the notation $\mathbf{T}(1) = \mathbf{T}$ for short):

$$\mathbf{T}(\tau) = \mathbf{T}^\tau(1) = \mathbf{T}^\tau. \tag{1.2.12}$$

It is proven that all homogeneous transition matrices have the exponential form. In the limit of continuous time, the transition matrix $\mathbf{T}$ of an infinitesimal time difference is expanded:

$$\mathbf{T}(t) = \mathbf{1} + t\mathbf{W} + o(t), \tag{1.2.13}$$

where the matrix $\mathbf{W}$ corresponds to the first-order term of the time in the expansion. The off-diagonal terms in $\mathbf{W}$ stand for the non-negative transition rate from one state to another, whose sum compensates the diagonal terms:

$$W_{x,x} = \sum_{y \in \Omega, y \neq x} W_{x,y}. \tag{1.2.14}$$

According to eq. (1.2.10), $\mathbf{T}(t)$ is written as:

$$\mathbf{T}(t) = \lim_{\epsilon \to 0} (\mathbf{1} + \epsilon \mathbf{W})^{\frac{t}{\epsilon}} = \mathrm{e}^{t\mathbf{W}}. \tag{1.2.15}$$

[5]This term is sometimes confused with another term *transfer matrix*. Transfer matrix is a technique in treating some special partition functions, e.g. the partition function of the one-dimension Ising model [34].

The term $o(t)$ in eq. (1.2.13) plays no role in eq. (1.2.15). We will not go too far in the continuous form of matrix, but it will be very extensively used in section 2.3.2.

By replacing T in Chapman–Kolmogorov Equation 1.2.9 with eq. (1.2.13), we have

$$T^{\tau+t}_{x1,x3} = \sum_{x_2\in\Omega} T^{\tau}_{x_1,x_2}\left(\delta_{x_2,x_3} + tW_{x_2,x_3}\right), \tag{1.2.16}$$

where δ_{x_1,x_2} is the Kronecker delta. After taking the derivative of t, eq. (1.2.16) is transformed into:

$$\left.\frac{\partial T^{t}_{x_1,x_3}}{\partial t}\right|_{t=\tau} = \sum_{x_2\in\Omega, x_2\neq x_3} T^{\tau}_{x_1,x_2}W_{x_2,x_3} - T^{\tau}_{x_1,x_3}\sum_{x\in\Omega, x\neq x_3} W_{x_3,x} \tag{1.2.17}$$

$$= \sum_{x_2\in\Omega, x_2\neq x_3}\left[T^{\tau}_{x_1,x_2}W_{x_2,x_3} - T^{\tau}_{x_1,x_3}W_{x_3,x_2}\right]. \tag{1.2.18}$$

By taking x_1 as the initial state, we have a partial differential equation of the distribution, based on the transition rate:

$$\frac{\partial P(x,t)}{\partial t} = \sum_{x'\in\Omega, x'\neq x}\left[P(x',t)W_{x',x} - P(x,t)W_{x,x'}\right]. \tag{1.2.19}$$

This partial differential equation is the *master equation* of the Markov process that describes the dynamics throughout the evolution of distribution, even in the discrete time case. By summing up different states, the conservation of probability in transitions is verified. It is the fundamental formula providing the cornerstone of Monte Carlo methods, which we will discuss later.

1.3 Convergence Theory

In statistical physics, most systems end up in a "macroscopic steady state", which corresponds to a certain probability distribution. We will then discuss the convergence in the dynamics determined by Markov chains.

1.3.1 Steady State and Irreducibility

With the notation of transition matrix, the time evolution of the distribution $\boldsymbol{\pi}^{(t)}$ is given by

$$\boldsymbol{\pi}^{(t)} = \boldsymbol{\pi}^{(t-1)}\mathbf{T}, \tag{1.3.1}$$

in its explicit form:

$$\pi^{(t)}(x) = \sum_{x'}\pi^{(t-1)}(x')T(x'\to x). \tag{1.3.2}$$

If

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{T}, \tag{1.3.3}$$

then $\boldsymbol{\pi}$ is the *steady state*. As eq. (1.3.3) iterates, it is expected that the distribution would converge to a single one starting from any initial distribution of states $\boldsymbol{\pi}^{(0)}$, for most physical processes:

$$\lim_{t\to\infty} \boldsymbol{\pi}^{(t)} = \boldsymbol{\pi}. \tag{1.3.4}$$

The uniqueness of the steady state is guaranteed by *Perron–Frobenius theorem*[6]. For an $n \times n$ strictly positive matrix $\mathbf{A} = (a_{ij})$ that $\forall 1 \leq i, j \leq n$, $a_{ij} > 0$:

- There exists a positive real number r as the unique eigenvalue with the largest (strictly) absolute value;
- All the elements of the left eigenvector $\vec{v}$ and the right one $\vec{w}$ of r are strictly positive, and

$$\lim_{t\to\infty} \vec{v}\mathbf{A}^t = r^t\vec{v}, \tag{1.3.5}$$

$$\lim_{t\to\infty} \mathbf{A}^t\vec{w} = r^t\vec{w}; \tag{1.3.6}$$

- The largest eigenvalue r satisfies $\min_i \sum_j a_{ij} \leq r \leq \min_j \sum_i a_{ij}$.

The first two statements indicate the existence of the steady-state distribution, on the condition that the transition matrix $\mathbf{T}$ has the largest eigenvalue as 1, which is then verified by the third statement. A positive transition matrix can always lead any initial distribution to the steady one. For a non-negative matrix $\mathbf{A} = (a_{ij}), a_{ij} \geq 0 : \forall 1 \leq i, j \leq n$, the theorem is not valid for general cases. But if $\mathbf{A}$ is *irreducible* and *aperiodic*, the theorem is applicable.

Irreducibility in Markov chain presents a non-zero probability of transition from any subset of the phase space to any other subset, in one or more steps. It is closely related to the connectivity of the space. An irreducible Markov chain has an irreducible transition matrix $\mathbf{A}$, which is not conjugated into a block upper triangular form by permutation. (A more illustrative interpretation is that the graph G_A associated with the matrix $\mathbf{A}$, whose number of vertices is the size of $\mathbf{A}$, and each oriented weighted line represents the corresponding element in $\mathbf{A}$, is *strongly connected*, as illustrated in Fig. 1.1.)

In an irreducible non-negative matrix, the group of the eigenvalues $\{\lambda_l\}$ with the absolute maximum r are in the form of

$$\lambda_l = r\mathrm{e}^{2\pi i l/h}, l = 0, 1, \ldots, h-1.$$

Aperiodicity restricts the *period* h to 1, and guarantees the existence of a single absolute maximum eigenvalue.

As a summary, the *irreducible aperiodic non-negative matrices* (or primitive matrices [35]) follows the same convergence theorem as the positive matrices, and an irreducible aperiodic non-negative transition matrix uniquely determines a stationary distribution, starting from any initial ones.

[6] We refer the theorem and the terminology to *Matrix analysis and applied linear algebra* by C. D. Meyer [35, chapter 8]

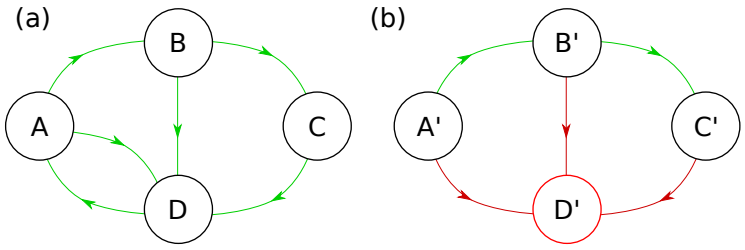


Figure 1.1: Graph representation of Markov chains. Each node represents a state, and each oriented edge (green and red) connecting two nodes represents a transition between states. (a) In a strongly connected graph whose corresponding transition matrix is irreducible, one can explore all the states starting from any of them. (b) In a weakly connected graph, one will be trapped in a subset (the node D') of all the nodes. The transition matrices of weakly connected graphs as in (b) and those of unconnected graphs are reducible.

1.3.2 Global Balance and Detailed Balance

Towards a single stationary distribution there exist multiple transition matrices. As the master equation already provides us with a general dynamical framework, we will then explore the conditions these matrices preserve. In analogy to Liouville's theorem, the master equation eq. (1.2.19) presents the conservation of the probability of one state. On its right-hand side is the probability flow:

$$F(x' \to x, t) = P(x', t) W_{x',x}. \tag{1.3.7}$$

We name the positive components "incoming flow" and the negative ones "outgoing flow":

$$\begin{aligned} F_{\text{in}}(x,t) &= \sum_{x' \in \Omega, x' \neq x} F(x' \to x, t), \\ F_{\text{out}}(x,t) &= \sum_{x' \in \Omega, x' \neq x} F(x \to x', t). \end{aligned} \tag{1.3.8}$$

The net probability flow, defined as the difference between $F_{\text{in}}(x,t)$ and $F_{\text{out}}(x,t)$ in eq. (1.3.8), is 0 in the steady state:

$$\begin{aligned} 0 &= F_{\text{in}}(x) - F_{\text{out}}(x) \\ &= \sum_{x' \in \Omega, x' \neq x} \left[\pi(x') W_{x',x} - \pi(x) W_{x,x'} \right], \end{aligned} \tag{1.3.9}$$

where $\pi(x)$ is probability distribution of the state x in the steady state. This is the *global-balance* condition, a necessary condition of convergence to the steady state (so are irreducibility and aperiodicity, but all of them three are sufficient), where the incoming and outgoing flows compensate, resulting in no macroscopic change of distribution.

A further *detailed-balance* condition requires that the net probability flow between any

two states in the stationary distribution vanishes:

$$\pi(x')W_{x',x} = \pi(x)W_{x,x'}, \tag{1.3.10}$$
$$F(x' \to x) = F(x \to x'), \tag{1.3.11}$$

which is more restrictive than the global balance. The transition matrix preserving the detailed-balance condition is a *reversible transition matrix*, and such a Markov chain is a *reversible Markov chain*. Many algorithms, such as the famous Metropolis–Hastings algorithm, which we will introduce in section 1.6.2, are based on the detailed-balance condition.

On the contrary, the relative freedom provided by the global-balance condition gives enormous possibilities for the construction of Monte Carlo algorithms. These delicate schemes will be presented in detail in chapter 2, and used in all later chapters.

1.4 Characterizing the Speed of Markov Chains

In the study of the dynamics of the Markov chains, an important property is its speed of convergence. In this section, without losing any universality, we consider discrete-time Markov chains to quantify this property.

1.4.1 Relaxation Time

As the transition matrix contains all the information of the Markov chain[7], we will evaluate its time scale from its eigenvalues. Given a reversible transition matrix $\mathbf{T}$, a symmetric matrix $\mathbf{A}$ is constructed

$$A(x', x) = [\pi(x')]^{\frac{1}{2}} T_{x',x} [\pi(x)]^{-\frac{1}{2}}, \tag{1.4.1}$$

according to eq. (1.3.11). Hence, all the eigenvalues of a reversible transition matrix $\mathbf{T}$ are real. The eigenvalues can be listed in a decreasing order:

$$1 = \lambda_1 > \lambda_2 \geq ... \geq \lambda_{|\Omega|} \geq -1, \tag{1.4.2}$$

and the *spectral gap* is defined by:

$$\gamma = 1 - \lambda^*, \tag{1.4.3}$$

where $\lambda^* = \max\{\lambda_2, |\lambda_{|\Omega|}|\}$, and γ is a positive number.

If the Markov chain is *irreversible* but *irreducible* and *aperiodic*, the eigenvalues may not be real, but the one with the largest absolute value λ^* is real according to the discussion in section 1.3.1. The spectral gap is defined the same as in eq. (1.4.3).
The relaxation time of Markov chains is defined as:

$$t_{\text{rel}} = \frac{1}{\gamma}, \tag{1.4.4}$$

[7] We refer to *Markov Chains and Mixing Times* by D. A. Levin et al. [36, chapter 4] for the discussion of the eigenvalues of a reversible transition matrix and the relaxation time.

which characterizes the slowest time scale of the dynamics in the limit of convergence.

Any probability distribution of the states in Ω (denoted by $\mathbf{a}$) can be written as a linear combination of the left eigenfunctions of the transition matrix (denoted by $\mathbf{a}_i$):

$$\mathbf{a} = \sum_{i=1}^{|\Omega|} c_i \mathbf{a}_i, \tag{1.4.5}$$

then

$$\begin{aligned} \mathbf{a}\mathbf{T}^n &= \sum_{i=1}^{|\Omega|} c_i \lambda_i^n \mathbf{a}_i \\ &= \sum_{i=1}^{|\Omega|} c_i \mathrm{e}^{in \operatorname{Arg} \lambda_i} \mathrm{e}^{n \log |\lambda_i|} \mathbf{a}_i. \end{aligned} \tag{1.4.6}$$

Except for $\lambda_1 = 1$ which corresponds to the steady state $\boldsymbol{\pi}$, all the other eigenvalues have exponential decreasing coefficients. The slowest one of them corresponds to λ^*, in the limit $\lambda^* \to 1^-$, $\log \lambda^* \to \gamma$.

1.4.2 Autocorrelation Time

Relaxation times describe the converging speed of Markov chains from the aspect of the tail in long-time evolution. Nevertheless, building the full transition matrix or fully characterizing the probability distribution in the phase space is an enormous job. In the alternative, measuring the "autocorrelation times" of the observables is a more pragmatic option.

From each configuration, we can obtain an "observable", which is a scalar (or a vector) characterizing some features of the system, as we indicated in section 1.2.1. The observable also forms a sequence in the Markov chain, and it has a steady-state distribution as well.

In order to see the correlation between two sequences of observables $\{X\}, \{Y\}$, we use the correlation function:

$$\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \left\langle \frac{(X - \bar{X})(Y - \bar{Y})}{\sigma_X \sigma_Y} \right\rangle. \tag{1.4.7}$$

whose range is $[-1, 1]$. $\operatorname{corr}(X,Y) = \pm 1$ indicates a linear relationship between these two variables, while $\operatorname{corr}(X,Y) = 0$ is a necessary but not sufficient condition to determine that $\{X\}, \{Y\}$ are independent.

By choosing the sequence $\{Y\}$ as a "shifted" $\{X\}$, the autocorrelation function of $\{X\}$ is defined as:

$$C_x(\tau) = \left\langle \frac{(X_t - \mu_X)(X_{t+\tau} - \mu_X)}{\sigma_X^2} \right\rangle_t. \tag{1.4.8}$$

In analogy to the correlation of two different sequences $\{X\}, \{Y\}$, if $C_x(\tau)$ is close enough to 0, it means that the configuration is almost independent of its history τ

steps ago. Here we define the *exponential autocorrelation time* of an observable f[8]:

$$\tau_{\text{exp}}(f) = \lim_{t\to\infty} \sup \frac{1}{-\log|C_f(t)|}, \tag{1.4.9}$$

and the *integrated autocorrelation time*:

$$\tau_{\text{int}}(f) = \frac{1}{2}\sum_{t=-\infty}^{t=+\infty} C_f(t). \tag{1.4.10}$$

The difference of these two autocorrelation times is up to a constant coefficient, only in the case that $C_f(t)$ is purely exponential.

As a surjective function of the state, an observable can not decorrelate faster than the state. For all possible observables, the exponential autocorrelation time of the Markov chain is

$$\tau_{\text{exp}} = \sup_f \tau_{\text{exp}}(f), \tag{1.4.11}$$

which is almost in the same order as the relaxation time t_{rel} ($\tau_{\text{exp}} = \frac{1}{\log(1-\gamma)}$, γ is the spectral gap) [38].

1.4.3 Total Variation Distance and Mixing Times

The total variation distance (TVD)[9] is used to characterize the difference between two probability distributions in the same state space. The TVD of two probability distributions $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ (whose probability density functions are f_μ and f_ν respectively) on the state space Ω is defined as:

$$\|\boldsymbol{\mu}-\boldsymbol{\nu}\|_{\text{TV}} = \frac{1}{2}\sum_{x\in\Omega}|\mu(x)-\nu(x)| \quad \text{(for discrete } \Omega) \tag{1.4.12}$$

$$= \frac{1}{2}\int_\Omega |f_\mu(x)-f_\nu(x)|\mathrm{d}x \quad \text{(for continuous } \Omega) \tag{1.4.13}$$

$$= \max_{\mathcal{A}\subseteq\Omega}|\mu(\mathcal{A})-\nu(\mathcal{A})|. \tag{1.4.14}$$

Eq. (1.4.14) is obtained by partitioning Ω into two subsets $\mathcal{S}^+$ and $\mathcal{S}^-$, that $\mu(x) \geq \nu(x), \forall x \in \mathcal{S}^+$ and analogously in $\mathcal{S}^-$. The difference of the two probability distributions over $\mathcal{S}^+$ in eqs (1.4.12) and (1.4.13) equals the corresponding one over $\mathcal{S}^-$, because of the normalization of probability, and it accounts for the prefactor of $\frac{1}{2}$ in these two equations. Eq. (1.4.14) reaches its maximum when $\mathcal{A} = S^+$ or S^-.
As a distance defined on the space for all the probability distributions on Ω, it is non-negative, it satisfies the triangular relation, and $\|\boldsymbol{\mu}-\boldsymbol{\nu}\|_{\text{TV}} = 0$ if and only if $\mu(x)-\nu(x) = 0$ *almost everywhere* on Ω.

In a Markov chain, we pick $\boldsymbol{\mu} = \boldsymbol{\pi}^{(t)}$ as the probability distribution of states after a certain number of steps and $\boldsymbol{\nu} = \boldsymbol{\pi}$, the distribution of steady state. As the Markov

[8]The following discussions, including exponential and integral autocorrelation time, are largely based on *Monte Carlo methods in statistical mechanics: foundations and new algorithms* by A. D. Sokal [37].

[9]We refer to *Markov Chains and Mixing Times* by D. A. Levin et al. [36, chapter 12] for the discussions of total variance distance, convergence theorem and mixing times.

chain converges, the TVD goes to 0 as time goes to infinity. Similar to the exponential decay of the fluctuations in the discussion of the relaxation time, there is a *convergence theorem*, in the language of TVD: for an irreducible and aperiodic transition matrix $\mathbf{T}$ with the stationary distribution $\boldsymbol{\pi}$, there exist constants $\alpha \in (0,1)$ and $C > 0$, that

$$\max_{x\in\Omega} \|\boldsymbol{\delta}(x)\mathbf{T}^t - \boldsymbol{\pi}\|_{\text{TV}} \leq C\alpha^t, \tag{1.4.15}$$

where $\boldsymbol{\delta}(x)$ is the distribution of one single initial state $x \in \Omega$.

For a given positive number ϵ, the *mixing time* $t_{\text{mix}}(\epsilon)$ is defined as the minimum of the time needed to have TVD $\leq \epsilon$, from any initial distribution:

$$t_{\text{mix}}(\epsilon) = \min\{t : \text{TVD}(t) \leq \epsilon\}. \tag{1.4.16}$$

According to the above convergence theorem, it is deduced that

$$t_{\text{mix}}(\epsilon) \propto -\log\epsilon, \text{ for small } \epsilon.$$

The mixing time tells the time needed for convergence from the worst possible initial state, while the relaxation time tells the slowest decreasing speed. Though their definitions are different, there is a connection between the two in reversible Markov chains in discrete state space, concerning the limit of small ϵ and large t [36]:

$$(t_{\text{rel}} - 1)\log\left(\frac{1}{2\epsilon}\right) \leq t_{\text{mix}}(\epsilon) \leq \log\left(\frac{1}{e\pi_{\min}}\right) t_{\text{rel}}, \text{ where } \pi_{\min} = \min_{x\in\Omega}\{\pi(x)\} \tag{1.4.17}$$

In more general cases (e.g. irreversible Markov chain and continuous phase space), their connection is not very clear, and in chapters 3, 4, and 5 we will see that the mixing times and autocorrelation times can be very different.

For a relatively large ϵ (e.g., $\epsilon = 1/4$), the TVD sometimes exhibits a cut-off in the t_{mix}; it drops to zero very fast, as if t_{mix} is independent of ϵ [39]. This phenomenon is not implied in the convergence theorem. Hence, the mixing process is more complex than an exponential convergence. In section 1.5 we will illustrate it with the example of the "coupon collector".

1.5 Simple Problems of Markov Chains

In this section some models of Markov chains are presented. These "toy models" carry non-trivial ideas which will shed light on the physical models that we will study in chapters 3, 4 and 5.

1.5.1 Coupon-collector Problem

Discrete-time Representation

In a set of n coupons, at the time $t = 1, 2, \ldots$, one of them is marked randomly. We now determine the probability that all the coupons are marked after a period T, or the probability distribution of the first-complete-collection time T_1.

In probability theory, the *stopping time* is the time (also a random variable) that some (well-defined) events are met in the stochastic process. Clearly, T_1 in this question is a stopping time.

The problem is simple but also related to many processes in statistical models, and it plays an important role in chapter 3. We first calculate the mean time of a complete collection. Having collected k different coupons, the probability of collecting a new one next time is

$$p_i = \frac{n-k}{n},$$

which means on average after $1/p_i$ pickings, a new coupon is collected. As the time interval between the first collection of the next new coupon is independent, we have:

$$\begin{aligned}\mathbb{E}(T_1) &= \mathbb{E}(\tau_1) + \mathbb{E}(\tau_2) + \cdots + \mathbb{E}(\tau_n) \\ &= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1} \\ &= n(1 + \frac{1}{2} + \cdots + \frac{1}{n}) = nH_n \\ &= n(\log n + \gamma) + \frac{1}{2} + \mathcal{O}(1/n)\end{aligned} \tag{1.5.1}$$

where H_n is the n-th harmonic number, and $\gamma \approx 0.5772156649\ldots$ is known as the Euler–Mascheroni constant. Hence, the expected first-complete-collection time is $\mathcal{O}(n \log n)$.

The event that the complete collection is not achieved at time t (or $T_1 > t$) is denoted by A, and the event that the coupon i is not marked during all the t times is A_i, then:

$$\mathbf{P}(A) = \mathbf{P}(\cup_i A_i) < \sum_i \mathbf{P}(A_i). \tag{1.5.2}$$

By taking $t = n \log n + cn$,

$$\mathbf{P}(T_1 > n \log n + cn) < n(1 - \frac{1}{n})^{n \log n + cn} < n\mathrm{e}^{-(\log n + c)} = \mathrm{e}^{-c}. \tag{1.5.3}$$

By subtracting the intersections between A_i, a more precise evaluation is given by P. Erdös and A. Rényi [19], that

$$\mathbf{P}(T_1 < n \log n + cn) \sim \mathrm{e}^{-\mathrm{e}^{-c}}. \tag{1.5.4}$$

The problem can be generalized for multiple complete collections [19], that

$$\mathbf{P}(T_m < n \log n + (m-1)n \log \log n + cn) \sim \mathrm{e}^{-\mathrm{e}^{-c}/(m-1)!}. \tag{1.5.5}$$

where T_m is the first time that each coupon is marked m times. The estimations of T_1 and T_m in eqs (1.5.4) and (1.5.5) are extensively used in section 3.2.2. There are other generalizations such as the coupons with non-uniform probabilities of being collected [40], which we will not discuss in further detail.

Continuous-time Representation

Another representation of the coupon-collector problem is described with a *continuous-time Poisson process*. Suppose there are n coins, each of which flips at the rate of 1, in other words, the time to the next flip τ is given by

$$\tau \sim \text{Exp}(1);\ p_\tau(x) = \mathrm{e}^{-x},$$

so that the flipping of each coin is independent of the condition of others and its previous history. We now study, firstly, the expected time t_1 that all coins are flipped at least once; and secondly, the number of flips (of all the coins) during this period.

For the first question, t_1 is equivalent to the maximum of n independent exponentially distributed random variables, whose cumulative distribution function is given in section 1.1.2

$$F_{t_1}(x;n) = \left(1-\mathrm{e}^{-x}\right)^n, \tag{1.5.6}$$

and the expectation is $\mathbb{E}(t_1) = H_n$ (the n-th harmonic number).

For the second question, the total number of flips from the beginning to time t is denoted by $N(t)$, which is a random variable with $\langle N(t)\rangle = nt$. According to the *optional stopping theorem*, the expectation of a martingale (as $[N(t)-nt]$) does not depend on the stopping time, as long as the stopping time is bounded or has a finite expectation [41]. Hence,

$$\mathbb{E}[N(t_1)-nt_1] = N(0)-0 = 0, \mathbb{E}[N(t_1)] = n\mathbb{E}(t_1), \tag{1.5.7}$$

which shows $\langle N(t_1)\rangle = nH_n$, the same as the calculation in the discrete time version.

Furthermore, as a max-stable statistic, in the limit $n \to \infty$, the distribution of t_1 converges to the Gumbel distribution with $\mu = H_n$ in eq. (1.1.15), as described in section 1.1.2. This explains the Gumbel-tail in the estimation of P. Erdös (eq. (1.5.4)).

If no two coins flip at the same time, the times when a flip occurs $\{t_1^{\text{flip}}, t_2^{\text{flip}}, \dots\}$ form a sequence, which maps exactly to the sequence $\{1, 2, \dots\}$. The probability of more than two flips occurring at the same time is zero (in measure), hence the continuous-time and the discrete-time models are essentially the same (almost everywhere).

The continuous-time representation is commonly used in the calculation of the mixing times, such as the single-particle random walk in section 1.5.2 and the multi-particle random walk in section 1.5.4. The expected number of events in one time unit is n in the continuous-time representation, while in the discrete-time representation, one event occurs exactly after one time step: their difference brings a factor of n in the calculations. In our later discussion of irreversible Markov chains (as in chapter 3), we mainly use the notation of the discrete-time representation.

1.5.2 Random Walk on the Vertices of a Hypercube

The coupon-collector model is a good example of Markov chains, however, at first sight it is not directly linked to a steady distribution of random variables (or there is only one steady state: all coupons are collected). We propose a modification to the

process as following:
Suppose there are n coins on the ground, either showing their heads or tails. Each time a random coin is picked up and tossed. (The model is equivalent to the simulation of an Ising model at $T=\infty$, whose state is pace is $\Omega=\{-1,1\}^n$.) A more mathematical statement is a random walk on the vertices of a n-dimensional hypercube, that at each time the walker randomly picks a dimension and randomly chooses to move or stay. We now study the time that it reaches the steady state (equal probability on each state), and how the TVD evolves.

In this set of problems, in analogy to the coupon-collector problem, the steady state is achieved after the first-complete-collection time T_1 that all the coins are picked (the moves on all the dimensions are proposed at least once). In the theory of Markov chains, there is a similar argument of a *strong stationary time*, after which the probability distribution of states no longer depends on time, such as T_1 in this process. The strong stationary time itself is a random variable and has its own distribution, which does not contradict with the exponential convergence of Markov chains.

If the walking occurs as a Poisson process with the rate $\frac{1}{n}$ on each dimension, an exact expression of TVD is given by [42]:

$$\mathrm{TVD}(t)=2^{-n-1}\sum_{k=0}^{n}\binom{n}{k}\left|(1-\mathrm{e}^{-2t/n})^{n-k}(1+\mathrm{e}^{-2t/n})^{k}-1\right|. \tag{1.5.8}$$

The cut-off occurs precisely at time $t_{\mathrm{mix}}=\frac{1}{4}n\log n$.

This model provides some ideas between the cut-off phenomenon in the evolution of TVD and the coupon-collector problem, as the mixing time is $\mathcal{O}(n\log n)$, while the distribution of the strong stationary time has a tail of $\mathrm{e}^{-t/n}$. The two different time scalings indicate unusual behavior, which we will discuss in section 1.5.5.

1.5.3 Single Random Walker

More generally, a random walk describing a path that consists of a succession of random steps can be defined on a more general discrete or continuous space other than the vertices of a hypercube in section 1.5.2. The steady state is a uniform distribution of the walker on each site, if the space is finite.

Random Walk on $\mathbb{Z}^d$

For a random walker starting from the origin, each step it has an equal probability of choosing one direction and moving to its neighbor site. The sequence of displacement $\{\vec{S}_i\}$ is both a martingale and a Markov chain. We may take the displacement of each step $\{\vec{X}_i\}$ as random variables, the total displacement is then

$$\vec{S}_n=\sum_{i=1}^{n}\vec{X}_i.$$

As the direction of each step is random, the correlation between any two different steps is 0:

$$\left\langle \vec{X}_i\cdot\vec{X}_j\right\rangle=\delta_{i,j},$$

where $\delta_{i,j}$ is the Kronecker delta. The mean-square distance is:

$$\langle \vec{S}_n^2 \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \vec{X}_i \cdot \vec{X}_j \rangle = n. \tag{1.5.9}$$

For more general random walks with randomized step size, the mean-square distance behaves as:

$$\langle \vec{S}(t)^2 \rangle \propto t^a. \tag{1.5.10}$$

If $a = 1$, this process is called a *diffusion*, which is the random walk in a discrete space we just discussed. In the similar way, it is named a *super-diffusion* if $a > 1$ and a *sub-diffusion* if $a < 1$. We will continue this discussion in the following, and it will be used in section 5.3.

Random Walk on $\mathbb{R}$

We generalize the random walk in section 1.5.3 to continuous time ($t \in \mathbb{R}$) and space ($x \in \mathbb{R}^n$), with continuous probability distribution of the step $\vec{X}$ and the total displacement $\vec{S}$. According to the discussion in section 1.1.2, every step must follow a stable distribution, if the process is homogeneous and infinitely divisible in time and space. Such a stochastic process with continuous path is called stable process, with zero-shift it is called *Lévy process*. The special case of $\alpha = 2$ is a *Wiener process*, also known in physics as a *Brownian-motion process*. [5]

From the discussion of partial sums in section 1.1.2, we see that for a Lévy process

$$\langle S(t)^2 \rangle \propto t^{2/\alpha}. \tag{1.5.11}$$

Hence, $\alpha = 2$ corresponds to a diffusion process, $\alpha < 2$ corresponds to a super-diffusion process, and a sub-diffusion is not compatible with Lévy processes.

Pólya's recurrence theorem states that the expected total times of returning to the origin is infinite for a random walk in the discrete lattice, if the dimension of the space is ≤ 2, and it is only finite otherwise [43]. For one-dimensional Lévy process, similarly, the walker will return to the origin an infinite number of times if $1 < \alpha \leq 2$, but if $\alpha \leq 1$ it will only return a finite number of times. One may find its relation to the law of large numbers in section 1.1.2. In chapter 5, we will study the return times of random processes in fairly complicated models, for which we have to rely on numerical simulations.

Random Walk on a loop

If the walker is restricted on a loop, such as $\mathbb{Z}_L$ or a circle (S^1) of length L, the mixing time is proportional to L^α. (For example, $\alpha = 2$ in the discrete case where the walker only move to its neighbor site.)

We will use the special case of a continuous loop and $\alpha = 2$ (Wiener's process) for example. The probability distribution after one step (which follows the distribution $\mathcal{N}(0, \sigma^2)$) is

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} \mathrm{e}^{-\frac{(x+kL)^2}{2\sigma^2}} = \vartheta_3(\pi x/L, \mathrm{e}^{-2\pi^2\sigma^2/L^2}), \tag{1.5.12}$$

where

$$\vartheta_3(x,q) = 1 + 2\sum_{n=1}^{\infty} q^{n^2}\cos(2nx) \tag{1.5.13}$$

is the Jacobi ϑ function. Agreeing well with its first-order approximation (see Fig. 3.6 (b)), the TVD decays exponentially with $t = \sigma^2$. In this case, there is no the cut-off in the mixing time.

The case of $\alpha < 2$ (super-diffusive) would lead to a faster convergence in the similar calculation. However a random variable with an infinite variance may get uncontrollable in the evaluation of errors [5], which we try to avoid in simulations.

1.5.4 Multiple Random Walkers

The dynamics and analysis become much more complex with an increasing number of random walkers, if a walker can block the others from passing through. In this section we will start with the random walks on a discrete space (e.g. lattice), that an occupied site does not allow a second walker (later we refer it as a "particle") to get in or get through, and then discuss the continuous case.

In the *symmetric simple exclusion process* (SEP), every particle attempts to move to all its neighbor sites with equal probabilities at a given rate, and the movement is accepted only if the neighbor site is empty, as illustrated in Fig. 1.2 (a). (One may interpret as a graph with vertices occupied by two kinds of indistinguishable "balls", the black for the particles and the white for the "holes". Each time the two balls on the endpoints of an edge switch their positions.) The mixing time of SEP in a one-dimensional lattice with L sites and n particles under periodic boundary condition is $L^3 \log n/(4\pi^2)$ single steps, obtained by H. Lacoin in 2016 [44]; for a d-dimensional lattice, the relaxation time is $\mathcal{O}(L^3 \log d)$ [45]. The difference between the mixing time and the relaxation time is seen once again, which we will discuss in section 1.5.5.

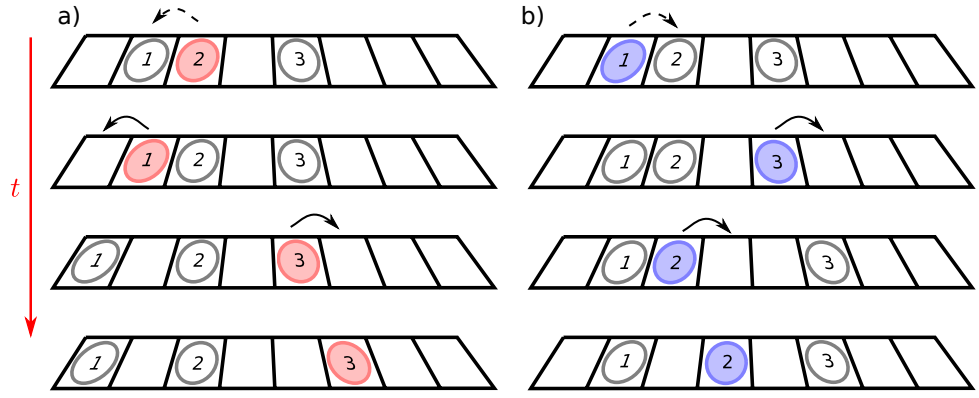


Figure 1.2: The dynamics of SEP and TASEP, the red arrow indicates time. (a) In SEP dynamics, each time a random particle attempts to move to its neighbor site, and it is accepted if it is empty, or rejected if it is occupied. (b) In TASEP dynamics, each time a random particle attempts to move to its right-hand-side neighbor, and the acceptance criterion is the same as SEP.

If each particle has larger probabilities for moving in some directions than the others, the process is called asymmetric simple exclusion process (ASEP) [46]; in the extreme case that it only attempts to move in one direction (on each dimension), the process is known as *totally asymmetric simple exclusion process* (TASEP) [47, 48], as shown in Fig. 1.2 (b).

TASEP is usually studied without periodic boundary conditions, and with imposed flows through the lattice. It has very rich phase behavior, such as a transition between an empty phase and a jammed phase [49]. Under periodic boundary condition, TASEP preserves the global-balance condition. It was proved by J. Baik and Z. Liu in 2016 that the mixing time is $\mathcal{O}(L^{5/2})$ on one-dimensional periodic lattice [50]. These recent exact results of the mixing times of SEP and TASEP demonstrate that irreversible Markov chains can gain an acceleration over reversible ones.

A variant of TASEP, namely the "lifted TASEP" [18] (the term "lift" will be explained in details in chapter 2), that the next active particle is determined by the last move. This process is deterministic, but if the active particle is reset after a randomly distributed period, the mixing time is $\mathcal{O}(n^2 \log n)$.

In a continuous space, a point-like walker can hardly block the others if the dimension is higher than one. For a Glauber dynamics of multiple walkers on a one-dimensional interval (or circle), D. Randall and P. Winkler proved in 2005 that the mixing time is $\mathcal{O}(n^3 \log n)$ on an interval [51], and between $\mathcal{O}(n^3)$ and $\mathcal{O}(n^3 \log n)$ on a circle [52] (simulations [18] considering the relaxation from a specific initial configuration agree with the latter bound as also stated in chapter 3, though the TVD cannot be directly calculated). The mixing times of these processes are similar to that of SEP on a one-dimensional lattice.

These processes set up different classes of dynamics (see Table 3.1), which help us to categorize the algorithms in chapter 3.

1.5.5 The Cut-off of Mixing Time

Based on the discussion in sections 1.5.2 and 1.5.4, we see that the scale of the mixing time is not necessarily the same as that of the relaxation time (and the scale of mixing time is always larger or equal).

The cut-off does not universally exist in the evolution of the TVD: in the random walk of a single particle on a one-dimensional interval, the TVD exhibits an exponential decay with time, and the mixing time has the same scale as the relaxation time.

In some cases, we relate the cut-off behavior of the mixing time to the effect of coupon-collector, as in the example of random walk on the vertices of a hypercube section 1.5.2:

- Before n steps, a complete collection is impossible. The expectation value claims a mixing time of $n \log n$ steps;
- After $n \log n$ steps, the stationary time has an exponentially decaying tail ($P(t < T_1) \propto \mathrm{e}^{-t/n}$, from Erdös and Rényi's estimation [19]), indicating a relaxation time of n.

In more general cases, the cut-off in the mixing time is a result of the different scalings of mixing time (starting from the least favorable case) and relaxation time (the long-term exponential decay). For example in the dynamics of SEP, it is shown that the cut-off does not exist in the evolution of TVD, if the initial configuration is well-chosen [44].

Another interpretation of the reason of the cut-off is the high multiplicity of the second largest eigenvalue (often in the order of the system size n) [39]. It is a general property for the process with permutation symmetry, such as in the coupon-collecting process, each coupon could be the last collected, leading to a multiplicity of n. We will continue the discussion of coupon collection and mixing processes in section 3.2.2.

1.6 Statistical Physics and Monte Carlo Methods

So far in the current chapter we discussed the properties of Markov chains. As a direct application, Markov-chain Monte Carlo is aimed at providing a good sampling of the model in high dimensions.

Monte Carlo is an application of the weak law of large numbers, that the partial average of a sample of independent and identically distributed random variables will approach its expectation, as the sample size goes to infinity:

$$\mathbb{E}(X) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{1.6.1}$$

In simulations, a random number generator is capable of generating a random (float) number ξ uniformly distributed in $(0, 1]$. For an arbitrary distribution, once the cumulative distribution function $F(x)$ and the probability density function $f(x)$ are given, one may establish a connection between ξ and x:

$$\int_0^1 \mathrm{d}\xi = \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_0^1 \mathrm{d}F(x), \tag{1.6.2}$$

that a given $\xi \in (0, 1]$ can be interpreted as $F(x)$ and the corresponding x can be found. For the expectation of a random variable $g(x)$:

$$\mathbb{E}[g(x)] = \int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x = \int_0^1 g(x)\mathrm{d}F(x) = \int_0^1 g(F^{-1}(\xi))\mathrm{d}\xi = \left\langle g(F^{-1}(\xi)) \right\rangle_\xi. \tag{1.6.3}$$

According to the law of large numbers, multiple samples of $g(F^{-1}(\xi))$ provide a good evaluation of $\mathbb{E}[g(x)]$, and the error is determined by the variance of the probability distribution and the number of samples.

Markov-chain Monte Carlo uses Markov chains in the sampling process to avoid high rejection rate in a general phase space. In the following we will begin with statistical physics (whose phase space may be complicated) and find its connection with Markov process.

1.6.1 Ensembles and Boltzmann Distribution

As mentioned in the beginning of this chapter, we introduced the probability theory using the mathematical axiomatic system. In practice, there are two common interpretations of probabilities: frequencies and degrees of beliefs. In the frequency interpretation (the "frequentist"), the probability is the long run proportion of times that an event happens in repetitions; in the degree-of-belief interpretation (the "Bayesian schools"), the probability measures the observer's strength of belief that an event happens [53]. For the evaluation of integrals, as partition functions and physical observables, the frequentist interpretation is appropriate, and the difference between these two interpretations does not matter too much in the philosophy of statistical physics (unless we talk about statistical inference).

In physics when we consider a system, an *ensemble* is an idealization consisting of numerous virtual copies of such a system, considered all at once, and each of them represents a possible state that the real system might be in. [54] In other words, an ensemble is a probability distribution of the state of the system given a set of parameters.

Physical ensembles are categorized to several types with respect to the set of parameters. Inside a system of a fixed number of particles N, a fixed volume V, and a fixed total energy E, every possible state has an equal probability to show up. The total number of states inside the system (as the cardinality of the state space) is denoted by $|\Omega(N,V,E)|$, and this type of distribution of states (or ensemble) is the *micro-canonical ensemble*. The entropy of this ensemble S, as an extensive observable, is defined as:

$$S = k_B \log |\Omega|, \tag{1.6.4}$$

where k_B is the Boltzmann constant. The state/phase space in physical models is often given by the generalized coordinates and momenta with their metric, that $\Omega = \{\omega : \omega = (\vec{x}_1, \ldots, \vec{x}_N, \vec{p}_1, \ldots, \vec{p}_N)\}$. In the continuous limit, each state occupies a volume of h^{dN} in the d-dimensional space, where h is the Planck constant.

We now think of a small system without degeneracy of energy (that each value of energy only matches one state in the state space), connected to a huge heat reservoir. The energy of the system can be exchanged until the equilibrium is achieved, then:

$$P_s(\epsilon | E_{\text{total}}) = P_{res}(E_{\text{total}} - \epsilon | E_{\text{total}}).$$

On the right hand, the probability is proportional to the number of states, as the reservoir is approximated by a micro-canonical ensemble. For the probability distribution of states of the small system, we have

$$\frac{P_s(E_1|E_{\text{total}})}{P_s(E_2|E_{\text{total}})} = \frac{|\Omega_{res}(E_{\text{total}} - E_1)|}{|\Omega_{res}(E_{\text{total}} - E_2)|} \approx \exp\left(\frac{1}{k_B}\left[\frac{1}{T}(E_2 - E_1)\right]\right), \tag{1.6.5}$$

where $T = \frac{\partial E}{\partial S}$ is the temperature of the whole system in equilibrium. Hence:

$$\frac{P_s(E_1)}{P_s(E_2)} = \mathrm{e}^{-\beta(E_1 - E_2)},\ \beta = \frac{1}{K_B T}. \tag{1.6.6}$$

The *canonical ensemble* describes the distribution of the states of the system with a group of parameters (N,V,T), where N is the number of particles, V is the volume

of the system and T is the temperature. In such ensembles, each state of energy E has a weight of $\mathrm{e}^{-\beta E}$ in its distribution. The partition function (generating function in mathematics) of this ensemble is:

$$Z = \sum_{\Omega} \mathrm{e}^{-\beta E} = \int \rho(E) \mathrm{e}^{-\beta E} \mathrm{d}E, \tag{1.6.7}$$

where $\rho(E)$ is the degeneracy/density of states of energy E. This is the Boltzmann distribution, a description of the physical ensembles.

There are more ensembles obtained by changing the state parameter, such as the *grand-canonical ensemble* with (μ, V, T), and the ones with constant pressure instead of volume. They all have similar forms of distribution. Using ensembles and Boltzmann distribution, it becomes feasible to design a Markov chain in physical models.

1.6.2 Monte Carlo of Detailed Balance

In Markov-chain Monte Carlo, the procedure of algorithm usually goes as following: a move from the old configuration x to a new one x' is proposed, the acceptance probability of this move $P^{\text{acc}}(x' \to x)$ is calculated, a random number is generated to decide whether this move is accepted, and so on and so forth. The detailed-balance condition provides detached equations between any two states, generally simplifying the implementation.

As suggested in eq. (1.3.10), in all reversible dynamics:

$$\pi(x') P^{\text{acc}}(x' \to x) = \pi(x) P^{\text{acc}}(x \to x'), \tag{1.6.8}$$

where P^{acc} is the acceptance probability. Based on this equation, we have the Metropolis dynamics and the Glauber dynamics.

Metropolis Dynamics

As the earliest proposed computational Monte Carlo algorithm, the Metropolis algorithm inaugurates the age of numerical analysis in physics. It is the first construction of general Markov chains with irreducibility for physical system under the detailed-balance condition. [4]

In eq. (1.3.10), both sides of the equation stand for the probability flows in the system, which drive the whole system towards its steady-state distribution. As P^{acc} can not be larger than 1, we have

$$P^{\text{acc}}_{\text{Metro}}(x \to x') = \min\left(1, \frac{\pi(x')}{\pi(x)}\right), \tag{1.6.9}$$

in order to maximize the probability flow between any two states. It is verified with eq. (1.6.8).

A generalization of Metropolis algorithm was given by Hastings [55], by introducing a priori conditional probability of proposal. The equation is written as:

$$P^{\text{acc}}_{\text{Metro}}(x \to x') = \min\left(1, \frac{\pi(x') g(x|x')}{\pi(x) g(x'|x)}\right), \tag{1.6.10}$$

where $g(x'|x)$ is the probability of proposing the new configuration x' starting from the configuration x, which follows the normalizing condition:

$$1 = \int g(x'|x)\mathrm{d}x'. \tag{1.6.11}$$

By choosing the proposing probability $g(x'|x)$ and sampling the next configuration x' wisely, high rejection rate can be avoided in the simulations. This is known as the *Metropolis–Hastings algorithm*. In the case of a symmetric proposal that $g(x'|x) = g(x|x')$, we recover the normal Metropolis algorithm.

Glauber Dynamics

As an alternative of Metropolis Dynamics, Glauber Dynamics was first developed by A. Barker in the study of a proton-electron plasma [56], and then widely used in the mathematical study of Markov chains. It is also adapted by Hastings in the improvement of Metropolis algorithm [55].

The idea of this dynamics is to sample directly in a subspace of Ω each time. The subspace $\Omega_i(x)$ generated from the old configuration x only has one degree of freedom on the dimension i:

$$\Omega_i(x) = \{x'|x'_j = x_j, \text{ if } j \neq i\}. \tag{1.6.12}$$

An integration in the subspace $\Omega_i(x)$ will give the distribution of x'_i, and a new configuration x' can be obtained from direct sampling in $\Omega_i(x)$.

It is the extreme case of Hastings' proposal: $P^{\text{acc}} = 1$ if $g(x'|x)$ is perfectly given by direct sampling. The dynamics is reversible, as the probabilities of $(x^{(1)} = x, x^{(2)} = x')$ and of $(x^{(1)} = x', x^{(2)} = x)$ are the same.

As the dynamics equilibrates the subspace $\omega(x, v)$, the algorithm is also known as the "heat-bath algorithm" or the "Gibbs sampler". This algorithm does not maximize the probability flows, but instead proposes the most reasonable choice of the next step. In some special cases, this method has a simpler form than the Metropolis algorithm. In the following we show an example of multiple balls attached to an elastic string.

As illustrated in Fig. 1.3, the energy of a harmonic string (we take $k = 1$ for simplicity) is:

$$E(\{X_i\}) = \sum_{i=0}^{N} \frac{1}{2}(X_{i+1} - X_i)^2, \tag{1.6.13}$$

with $X_0 = 0, X_{N+1} = L$ fixed. In order to sample this model from a given configuration $\vec{x} = (x_1, ..., x_n)$, both algorithms propose a move on site i:

- In the Metropolis algorithm, a move $x'_i = x_i + \delta x$ is proposed (where δx follows some symmetric distribution), and the new configuration is accepted with the probability:

$$\begin{aligned} P^{\text{acc}}_{\text{Metro}}(x \to x') &= \min(1, \mathrm{e}^{-\frac{\beta}{2}\left[(x_{i+1}-x'_i)^2-(x_{i+1}-x_i)^2+(x'_i-x_{i-1})^2-(x_i-x_{i+1})^2\right]}) \\ &= \min(1, \mathrm{e}^{\beta\delta x(x_{i+1}+x_{i+1}-2x_i-\delta x)}). \end{aligned} \tag{1.6.14}$$

- In the heat-bath algorithm, we know:

$$\begin{aligned} E(x_i; x_{i-1}, x_{i+1}) &= \frac{1}{2}[(x_{i+1} - x_i)^2 + (x_i - x_{i-1})^2] \\ &= x_i^2 - x_i(x_{i-1} + x_{i+1}) + \frac{1}{2}(x_{i-1}^2 + x_{i+1}^2), \\ p(x_i | x_{i-1}, x_{i+1}) &= \frac{\mathrm{e}^{-\beta E(x_i; x_{i-1}, x_{i+1})}}{Z_i} \\ &= \sqrt{\frac{\beta}{\pi}} \mathrm{e}^{-\beta(x_i - \frac{x_{i-1}+x_{i+1}}{2})^2}. \end{aligned} \tag{1.6.15}$$

  Hence $x_i' \sim \mathcal{N}(\frac{x_{i-1}+x_{i+1}}{2}, \frac{1}{2\beta})$ in the subspace.

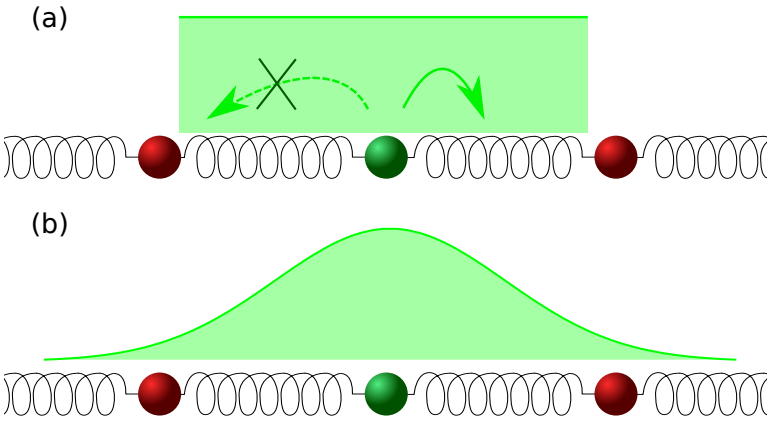


Figure 1.3: Multiple balls attached to a harmonic chain. (a) In the Metropolis algorithm, a movement is proposed with a uniform distribution (other distributions are possible). It will be either accepted or rejected by comparing a random number and the corresponding element in the transition matrix. (b) In the heat bath algorithm, given the position of its neighbors, the next position of the ball follows a Gaussian distribution, which enable directing sampling with no rejection.

This process is then totally rejection-free in sampling. However, it is not as efficient as Hastings' proposal, if the phase space is discrete, as there is a finite chance of staying. Another difficulty comes from the one dimensional integration: it does not always have such a simple form. As a result, Glauber dynamics is the most considered in mathematical discussions, while the Metropolis–Hastings algorithm is more widely used in practice.

1.6.3 Dynamic Scaling law of Monte Carlo Methods

In section 1.5, we presented some temporal properties of Markov chains in simple models. Physical models are usually more complex, especially when a phase transition is involved. For example, in the *first-order phase transition* (also known as discontinuous phase transition) concerning two coexisting phases and a latent heat (e.g.

evaporation of water), the difficulty in the simulation is due to the interface tension created by two separated phases, which leads to an energy barrier of $E = \sigma L^{D-1}$ (σ is the interface tension). It will take $\mathcal{O}(\mathrm{e}^{\beta\sigma L^{D-1}})$ time to get over the barrier on average. New approaches such as multi-canonical simulations have been developed to overcome this problem. [57]

In some physical models, the spatial correlation decreases exponentially, characterized by the correlation length ξ. [58] It concerns many other observables, such as the magnetization, the susceptibility and the specific heat in spin systems. A significant feature in the *second-order phase transition* is a divergent correlation length:

$$\xi \propto |1 - \frac{T}{T_c}|^{-\nu}, \tag{1.6.16}$$

where T_c is the transition temperature.

In the dynamical evolution, the system can be regarded as "almost independent" blocks of the size ξ^d. The time needed to sample the whole system (counted in sweeps) is almost the time needed to sample each small block. Then the relaxation time is only dependent of ξ. However near the critical point:

$$t_{\mathrm{rel}} \propto \xi^z \propto \left|1 - \frac{T}{T_c}\right|^{-\nu z}, \tag{1.6.17}$$

where z is the dynamical critical exponent. At the critical point, the divergent spatial correlation will lead to a divergent temporal correlation, which is known as the critical slowing down, while in numerical simulation, as the correlation length is restricted by system size, we have

$$t_{\mathrm{rel}} \propto L^z. \tag{1.6.18}$$

z is independent of the static exponents such as ν, but it is dependent of the dynamics, as Metropolis and Glauber dynamics usually have $z \geq 2$, and optimized algorithms (such as the cluster algorithm introduced below) can have $z \approx 0$.

Reversely and notably, many pure mathematical non-deterministic polynomial time (known as "NP-complete") problems (such as Boolean satisfiability problem) also exhibit transition-like "phase boundary" induced by the input parameters, above which the computing time increases from polynomial to exponential [59]. The connection between phase transition and computational complexity is even more strengthened. Nevertheless, for some models, faster and more optimal methods are designed based on their physical properties, as introduced in the following.

1.6.4 Non-local Algorithms

Critical slowing down in the second-order phase transition is often due to the diverging correlation length, which makes the whole system an unbreakable cluster. One idea is to evolve the whole cluster instead of single particles or sites, which leads to the rise of non-local Monte Carlo methods. In the following we will introduce some non-local methods applied in various systems.

Spin Systems

In the two-dimensional Ising model, below the critical temperature $T_c = \frac{2}{\log(1+\sqrt{2})}$ [60], the system is more likely to have non-zero magnetization. We consider a configuration in a square lattice, with the spins in the left half all pointing upwards, and the ones in the right half pointing downwards. This configuration should have a small weight in the partition function at low temperature, however the spin near the interface can hardly flip, since there are always more neighbors with the same orientation. Local Monte Carlo methods will be trapped in this local minimum of energy.

A generalization of the Ising model is the Potts model, which has more than 2 states on each site. R. Swendsen and J. Wang proposed a non-local algorithm based on the Fortuin-Kasteleyn scheme, which maps the Potts model to a percolation model via a duality transformation of vertices and edges [61, 62]. The Swendsen-Wang algorithm builds a cluster as percolated blocks, and flip the whole cluster each time. As the size of the cluster grows as large as the correlation length, the time needed for equilibrating is much reduced.

A more advanced algorithm for spin systems (even with continuous phase space) was proposed by U. Wolff. [63] The idea begins with the selection of an axis of rotation (indicated by a unit vector $\mathbf{r}$) and the construction of a cluster C from a random site, by accepting the adjacent sites with a certain probability ($P_{\mathbf{r}}(\sigma_x, \sigma_y)$, depending on the axis and the spins). Once the construction is finished, a new configuration is obtained by flipping the whole cluster along the axis. The transition matrix fulfills the detailed-balance condition

$$\frac{W(\{\sigma_x\} \to \{\sigma'_x\}|C, \mathbf{r})}{W(\{\sigma'_x\} \to \{\sigma_x\}|C, \mathbf{r})} = \frac{\pi(\{\sigma'_x\})}{\pi(\{\sigma_x\})}, \tag{1.6.19}$$

where $\{\sigma_x\}$ and $\{\sigma'_x\}$ indicate the two configurations connected by the rotation of the cluster c along $\mathbf{r}$. The expansion of the left gives

$$\frac{W(\{\sigma_x\} \to \{\sigma'_x\}|C, \mathbf{r})}{W(\{\sigma'_x\} \to \{\sigma_x\}|C, \mathbf{r})} = \prod_{\{x,y\}\in\partial C} \frac{1 - P_r(\sigma_x, \sigma_y)}{1 - P_r(\sigma'_x, \sigma_y)}, \tag{1.6.20}$$

while the right is

$$\frac{\pi(\{\sigma'_x\})}{\pi(\{\sigma_x\})} = \frac{\mathrm{e}^{-\beta E(\{\sigma'_x\})}}{\mathrm{e}^{-\beta E(\{\sigma_x\})}} = \prod_{\{x,y\}\in\partial C} \mathrm{e}^{-\beta[V(\sigma'_x,\sigma_y) - V(\sigma_x,\sigma_y)]}, \tag{1.6.21}$$

where $\{x, y\} \in \partial C$ indicates a two neighbor sites x, y with x inside the cluster C and y outside. The detailed balance is perfectly matched by choosing

$$P_{\mathbf{r}}(\sigma_x, \sigma_y) = \max(0, 1 - \mathrm{e}^{\beta(V(\sigma_x,\sigma_y) - V(\mathbf{R}_r\sigma_x,\sigma_y))}), \tag{1.6.22}$$

where $\mathbf{R}_r$ is the rotation operator.

In the case of $O(n)$, where σ_x is an n-dimensional unit vector, the probability has a simple form:

$$P_{\mathbf{r}}(\sigma_x, \sigma_y) = \max(0, 1 - \mathrm{e}^{\beta\sigma_x\cdot(1-\mathbf{R}_r)\sigma_y}). \tag{1.6.23}$$

This algorithm also works beyond the $O(n)$ model, as long as only the closest two-body interaction is considered on the lattice. For example in Villain's model [64] (a modified version of the two-dimensional XY model, which is mentioned and studied in chapter 4), whose partition function is

$$\begin{aligned} Z &= [\prod_i \int_{-\pi}^{\pi} d\phi_i] \prod_{\langle i,j\rangle} \sum_{n_{i,j}=\infty}^{\infty} \mathrm{e}^{-\beta(\phi_i-\phi_j-2\pi n_{ij})^2} \\ &= \left(\frac{1}{2\sqrt{\pi\beta}}\right)^{nd} [\prod_i \int_{-\pi}^{\pi} d\phi_i] \prod_{\langle i,j\rangle} \vartheta_3(\frac{\phi_i-\phi_j}{2}, e^{-\frac{1}{4\beta}}), \end{aligned} \tag{1.6.24}$$

($\phi_i \in \mathbb{R}$, ϑ_3 is the Jacobi's ϑ function in eq. (1.5.13)) the corresponding acceptance rate is

$$P_{\mathbf{r}}(\phi_x, \phi_y) = \max(0, 1-\mathrm{e}^{\beta\sigma_x\cdot(1-\mathbf{R}_r)\sigma_y}) = \max(0, 1-\frac{\vartheta_3((\mathbf{R}_r\phi_x-\phi_y)/2, e^{-\frac{1}{4\beta}})}{\vartheta_3((\phi_x-\phi_y)/2, e^{-\frac{1}{4\beta}})}). \tag{1.6.25}$$

Wolff's algorithm has a very small relaxation time ($z \leq 0.1$ in XY model) [63], which facilitates the evaluation of the critical behaviors of Kosterlitz-Thouless transition in chapter 4. The algorithm also works very well in Villain's model [65], however it does not show significant advantage over the other algorithms in spin glass system [66].

Geometric Cluster Algorithms

The idea of rotation around a pivot, or reflection along an axis, is also applied in many other systems, e.g, in the dimer model and in the hard-sphere model. These methods are known as geometrical cluster algorithms, which also concern global movements.

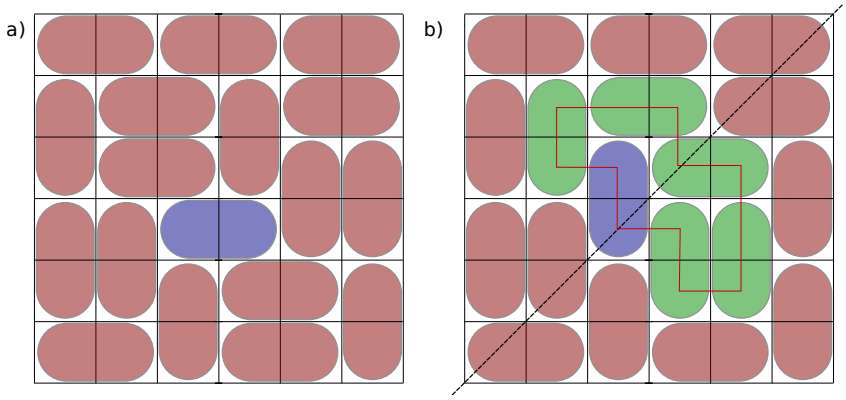


Figure 1.4: Dimer model as a Domino tiling problem. (a) The blue dimer initiates the flipping in the original fully covered configuration. (b) After flipping the other dimers (green) and forming a loop, the new configuration is obtained.

The dimer model has attracted people's attention for a long time. In this model, every site is paired with only one of its neighbor site, forming a dimer. Via a simple duality

of vertices and blocks, the dimer model is equivalent to the Domino tiling problem in combinatorics, which uses 1×2 and 2×1 Domino cards to cover an area without overlapping (see Fig. 1.4). The partition function is analytically calculated when the configuration is fully covered by the dimers [67]. In limit of few dopings, the sampling using local Monte Carlo meets a high rejection rate, while the geometrical cluster algorithm [68] works well. It performs as following:

- One of the diagonals (it will be generalized in periodic boundary condition) of the lattice is chosen as the axis, and the initial dimer is flipped with respect to the diagonal;
- If it overlaps with other dimers, flip the dimer overlapped with respect to the diagonal again, the process proceeds until no overlapping.

A similar strategy is implemented in hard sphere systems, with the rotation around a pivot [69]. For a given pivot and an initially selected sphere, the sphere is displaced to the symmetric position of the pivot. If it overlaps with other spheres, the overlapped spheres are displaced in the similar way, until there is no more overlapping. This method is useful in the simulation of polydisperse models [70].

Nevertheless, the methods based on the characteristics and symmetries of the systems may not apply to the all the cases, such as the glass systems. With the development of technology and hardware, new algorithms such as parallel tempering [71], and Monte Carlo with unsupervised machine learning [72] are invented to speed up the simulations and solve the physical problems.

1.7 Conclusion

In this chapter, we reviewed some aspects of the probability theory and the theory of Markov chains, analyzed the condition of convergence and time scales of a Markov chain. We discussed a few examples of Markov chains related to the coupon-collector problem, which helps explain the difference between the term "relaxation time" and "mixing time", and the cut-off phenomenon in the mixing time. Based on the detailed-balance condition, we presented several reversible Markov-chain Monte Carlo methods such as the Metropolis-Hastings algorithm and the heat-bath algorithm, that are applied in many physical systems, and some optimized non-local methods such as Wolff's cluster algorithm to overcome the limit of phase transitions.

Chapter 2

Irreversible Markov Chains and the Factorized Metropolis Filter

In chapter 1 we discussed algorithms based on the detailed-balance condition: from local algorithms to cluster algorithms, the modern Markov chains have much accelerated the speed of sampling and to overcome the critical limit of the physical system.

Irreversible Markov chains, based on a weaker but more general global-balance condition, are able to mix the system faster than reversible ones (e.g. the circular flow construction method in section 2.1). Although global-balance algorithms were used in the physics literature (such as sequential-sweep updates for particle systems and spin models), a general irreversible algorithm with fast-mixing properties hasn't shown up for long, until the pioneering work using "lifting schemes" proposed by P. Diaconis et al [73]. In this work, a standard method of irreversible Markov Chain construction came to be known.

In the present chapter, we introduce the concept of"lifting schemes", generalize it to multi-dimensional and multi-particle systems, and eventually, introduce the "factorized Metropolis filter" for Boltzmann sampling.

2.1 General Irreversible Markov Chains

According to eq. (1.3.3), an aperiodic irreducible transition matrix $\mathbf{T}$ of a discrete-time Markov chain satisfies the global-balance condition:

$$\begin{aligned} \boldsymbol{\pi} &= \boldsymbol{\pi}\mathbf{T}, \text{ or equivalently,} \\ \pi(x) &= \sum_{x'} \pi(x')T(x' \to x) \equiv \mathcal{F}_x, \end{aligned} \tag{2.1.1}$$

where $\mathcal{F}_x$ is the total probability flow into the state x (including the portion from itself). It can be split into two contributions, namely the "accepted" flow from other states $x' \neq x$

$$\mathcal{A}(x' \to x) = \pi(x')T(x' \to x), \tag{2.1.2}$$

and the "rejected" flow from x itself

$$\mathcal{R}(x) = \pi(x)T(x \to x). \tag{2.1.3}$$

The rejected flow comes either from the rejection of an attempted move (to other states), or from the probability it proposes to stay. As staying will not help the evolution of the system, $\mathcal{R}(x)$ is usually minimized in Monte Carlo, such as the Metropolis algorithm in section 1.6.2 (sometimes we keep it in a considerably small value to break the periodicity). On the other hand, the outgoing probability flows should be as large as possible.

The detailed-balance condition is given by

$$\mathcal{A}(x' \to x) = \mathcal{A}(x \to x'). \tag{2.1.4}$$

In much of the work of Markov-chain Monte Carlo we have discussed up to now, the design of algorithms consists in checking for any pair of states x and x' that eq. (2.1.4) is satisfied. In contrast, to satisfy the global-balance condition, one has to check that the sum of the rejected and accepted flow into x equals $\pi(x)$ (while the sum of the outgoing flow is guaranteed by the normalization of marginal distribution), as implied in eq. (2.1.1). As the most basic equation of global balance, eq. (2.1.1) will be used throughout the thesis, in the examination of the validity of an irreversible Markov chain.

Sequential-sweep Updates

As we mention in 1.5.1, the effect of "coupon-collector" can lead to a $(\log n)$ factor in the mixing time. One may naturally think of applying the Metropolis algorithm sequentially on each particle in each sweep, instead of randomly picking. This leads to the simplest irreversible Markov chain in Monte Carlo sampling.

This algorithm has a long history: it was discussed by N. Metropolis et al. in 1953, as they proposed the reversible Metropolis algorithm for the first time [4]. Compared to the reversible one it has a different autocorrelation time, nevertheless, this method does not reduce the mixing time largely, as the difference is only up to a constant factor [74].

Though not very effective, it is an example showing that steady-state distribution is reachable without the detailed-balance condition. The algorithms with "swap", which we will introduce in section 3.2.3 and 3.3.2, also use a similar idea in the optimization.

Probability Flow Construction

Reversible Markov chains usually lead to a "diffusion" in the graph of states, while a "propagation" possibly transmits the information about the stationary probability distribution much faster. It gives rise to the idea of creating a non-zero macroscopic net flow inside the graph associated to the state space.

One basic idea of acceleration is to minimize the probability of staying and rejecting. If we construct several loops with optimal flow in the exploration of phase space, the algorithm may have a better convergence. The problem is then transformed into a decomposition of the overall probability flow into multiple circular flows. For the

circular flows with the weight $\{f_i\}$, we have:

$$\pi(x) = \mathcal{F}_x = \sum_i n(i,x) f_i, \tag{2.1.5}$$

where $n(i,x)$ is the number of times that the probability flow i crosses the state x. An example is showed in Fig. 2.1, which satisfies the global-balance condition.

Furthermore, to satisfy irreducibility, there exists at least one loop connecting one state from a side to the other in any bipartitions of the state space; as for aperiodicity, one may introduce the probability of a "staying" proposal.

However, this algorithm is not universal: (1) Before the construction we need to know every detailed state and manipulate its connections, which is no less work than a direct integral; (2) The probability of staying is tricky to choose. We will try to solve the first question by the construction of a general scheme in the present chapter. The second one is equivalent to the "optimal stopping times" problem and will be discussed in chapter 3.

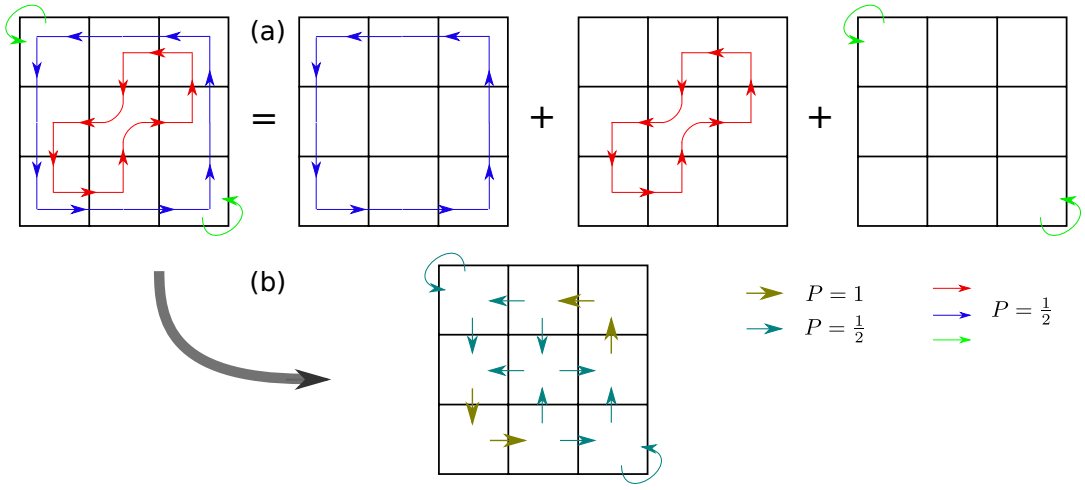


Figure 2.1: One construction of probability flows in the sampling of a particle on a 3×3 square lattice. (a) The total probability flow is decomposed into four loops, and each of them (green, red and blue) carries a weight $f_i = 1/2$ (and they do not compensate each other). (b) The actual total probability flow is obtained by summing all the loops (which indicates the element in the transition matrix). The Markov chain satisfies the global balance, with the uniform steady-state distribution. The irreducibility is guaranteed by the blue and red flows, and the periodicity is broken by the green "rejection" (or "staying") flow.

2.2 Lifting Schemes

The lifting scheme based on global balance is proposed for general systems to overcome the limitations of the probability-flow construction in section 2.1. The key of the scheme is building up multiple replicas, and constructing non-zero net circular

probability flows inside each of them. We will start with a simple example of a single random walker, following the work of P. Diaconis et al.[73].

2.2.1 Single One-dimensional Random Walker

For a single random walker on a bounded one-dimensional lattice $\mathbf{R}$, we may use the trick as shown in the Fig. 2.2, by "gluing" the original lattice with its replica to construct a circle. The new periodic lattice $\mathbf{R}'$ consists of $\mathbf{R}'_1$ and $\mathbf{R}'_2$.

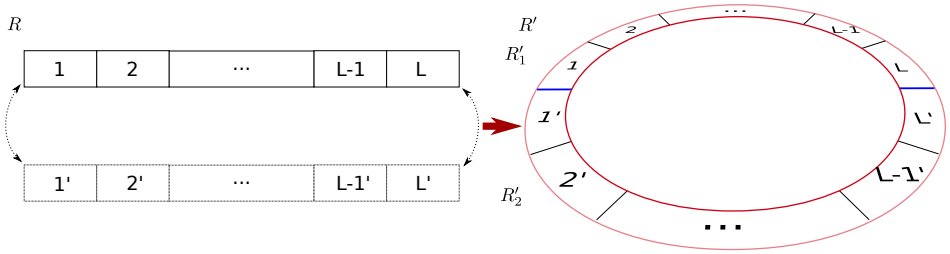


Figure 2.2: The "gluing" trick that convert a bounded one-dimensional lattice $\mathbf{R}$ to a periodic one $\mathbf{R}'$. A random walk of a single walker in $\mathbf{R}'$ is surjected onto a single-particle random walk in $\mathbf{R}$.

For the steady-state distribution on each site of $\mathbf{R}$, we have

$$\pi(x) = \pi'_1(x) + \pi'_2(x), \tag{2.2.1}$$

where $\pi(x)$ is the steady-state probability of state x in $\mathbf{R}$, $\pi'_i(x)$ is the steady-state probability in $\mathbf{R}'_i \subset \mathbf{R}'$. One may check that the "projected" random walk in $\mathbf{R}'$ is also a random walk in $\mathbf{R}$. For bounded systems, the "gluing trick" can be applied to create a "loop" with a non-zero net probability flow.

Here, without loss of universality, we will use a periodic one-dimensional lattice of size L ($\Omega = \mathbb{Z}_L$) to simplify the discussion. (However, in the case of multiple random walkers, the mixing on a bounded interval is very different from that on a circle [51, 52].) As shown in Fig. 2.3 (a), we consider a uniform steady distribution on $\mathbb{Z}_L$:

$$\pi(x) = \frac{1}{L}, x = 1, 2, \ldots, L.$$

For a random walk, at each time step the walker has a probability of one half to move to the left or to the right. The transition matrix is:

$$T(x \to y) = \begin{cases} 0.5 & \text{if } y = x - 1, \\ 0.5 & \text{if } y = x + 1. \end{cases} \tag{2.2.2}$$

When L is odd, the relaxation time of the dynamics is $\mathcal{O}(L^2)$, while the transition matrix is periodic if L is even (the periodicity can be broken with a "staying" probability).

In the lifting scheme (see Fig. 2.3 (b)), the state space is extended to two replicas. In one of them (denoted by the replica $\sigma = -1$), the walker preferentially moves to its

left, while in the other (denoted by $\sigma = +1$) it moves to the right symmetrically. The transition matrix is:

$$T(\{x,\sigma_1\} \to \{y,\sigma_2\}) = \begin{cases} 1-\alpha & \text{if } y = x+\sigma_1, \sigma_1 = \sigma_2, \\ \alpha & \text{if } y = x, \sigma_1 \neq \sigma_2 \\ 0 & \text{other cases,} \end{cases} \tag{2.2.3}$$

where $\alpha \in [0,1]$. In this case, the detailed-balance condition no longer holds true: inside the lifting scheme, the movement $\{x,\sigma\} \to \{x+\sigma,\sigma\}$ is not reversible. The irreversibility acts as a persistence of the motion in the original physical model.

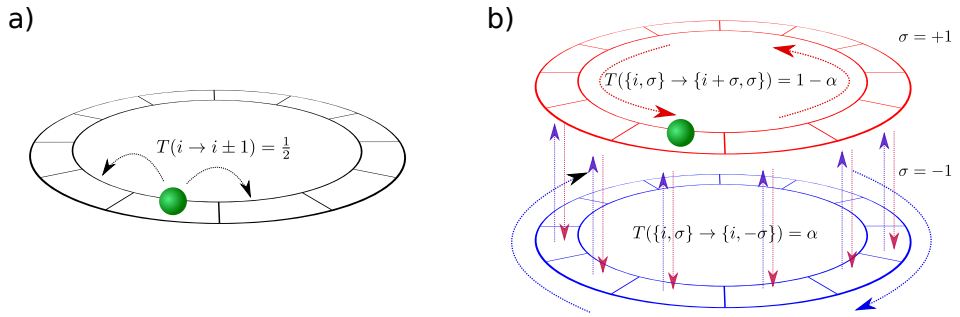


Figure 2.3: Random walk of a single particle on the one-dimensional lattice with periodic boundary. (a) In the normal scheme, the particle (green) moves either to the left or to the right, with equal probability. This process mixes in $\mathcal{O}(L^2)$ moves. (b) In the lifted scheme, there are two replicas of the original lattice. The walker moves anti-clockwisely in the red replica, or clockwisely in the blue replica (indicated by the round solid arrows), with the same probability of $(1-\alpha)$, and hops to the corresponding site in the other replica with the probability of α (indicated by the dashed arrows between the replicas).

By choosing $\alpha = 0$, although the internal flow is maximized, the dynamics is purely deterministic and periodic; if $\alpha = 1/2$, it leads to a random walk as discussed before. In the case where $\alpha = 1/L$, the relaxation time and mixing time of this dynamics both are $\mathcal{O}(L)$ [73], which exhibits a propagative behavior in mixing rather than diffusive.

Continuous Limit

In the limit $L \to \infty$, we think of a continuous rescaled interval of $[0,1)$ instead of $\mathbb{Z}_L$. When $\alpha = 1/L$, the distance ℓ between 2 turnings (which is also called a "chain length") follows an exponential distribution:

$$p(\ell) = \mathrm{e}^{-\ell}, \tag{2.2.4}$$

and the number of turning points during a period follows a Poisson distribution with $\lambda = 1$. In this process the mixing time (proportional to the total distance) of $\mathcal{O}(1)$ is somehow self-evident, in agreement with the mixing time of $\mathcal{O}(L)$ in the discrete case.

Nevertheless, an exponentially distributed "chain length" l is not the whole story. Rather than moving back and forth, a walker moving only forward at a randomized stopping time will bring similar dynamics, which we will explain in section 3.2.

Non-uniform Random Walk

This algorithm also applies to non-uniform steady-state distribution π, by introducing *a priori* acceptance probabilities.

Given the stationary distribution $\pi(x), x \in \mathbb{Z}_L$, the transition matrix of the Metropolis algorithm is:

$$T(x \to y) = \begin{cases} \frac{1}{2}\min(1, \frac{\pi(x-1)}{\pi(x)}) & \text{if } y = x-1, \\ \frac{1}{2}\min(1, \frac{\pi(x+1)}{\pi(x)}) & \text{if } y = x+1, \\ 1 - T(x \to x-1) - T(x \to x+1) & \text{if } y = x. \end{cases} \tag{2.2.5}$$

In the lifting scheme, any rejection will trigger a hopping between replicas:

$$T(\{x,\sigma_1\} \to \{y,\sigma_2\}) = \begin{cases} (1-\alpha)\min(1, \frac{\pi(x+\sigma_1)}{\pi(x)}) & \text{if } y = x+\sigma_1, \sigma_1 = \sigma_2, \\ 1 - (1-\alpha)\min(1, \frac{\pi(x+\sigma_1)}{\pi(x)}) & \text{if } y = x, \sigma_1 \neq \sigma_2, \\ 0 & \text{other cases,} \end{cases} \tag{2.2.6}$$

and the global-balance condition is once again preserved.

The optimal $\alpha \propto \frac{1}{L}$ works in the uniform case, while in this non-uniform case, its value depends on the explicit form of the steady-state distribution. We will discuss the limit of the efficiency in non-uniform case in section 2.4.2.

So far we have discussed the "lifting scheme" on a one-dimensional lattice, where the original model is expanded to two replicas with a "lifting variable" (σ). The probability-flow circuits are constructed on the replicas, with a certain probability to "lift" from one to the other. In the later models, the concept of lifting variable is much generalized, with more replicas built for fast mixing.

2.2.2 Lifting Schemes with Multiple Variables

For a single-particle random walk on a more general strongly connected graph than the one-dimensional lattice, multiple lifting variables can be introduced depending on the dimension of the graph.

The lifting schemes and the probability-flow construction in section 2.1 share some similarities in dividing the flows. More specifically, in lifting schemes, the states are weakly connected following eq. (2.1.1) inside each replica (with the set of lifting variables fixed), and each replica has a similar structure. The sum of probability flows in all the replicas (some of them will compensate) in the one-dimensional single-particle random walk in section 2.2.1 is

$$\tilde{\mathcal{A}}(x' \to x) = \sum_{\sigma} \pi_\sigma(x') T_\sigma(x' \to x), \tag{2.2.7}$$

which preserves the detailed-balance condition in eq. (2.1.4), ensuring the conformity with the conventional Monte Carlo method. (In more general cases, eq. (2.2.7) is not necessary. In section 2.4.1, as well as in chapter 3, 4 and 5, the lifting schemes we propose only concern the global balance.)

Lifting schemes for random walks can be generalized to multiple dimensions with multiple lifting variables. For example, we may use $2 \times d$ lifting variables/replicas for a d-dimensional lattice of size L^d, with the lifting scheme [75]. The mixing in each dimension takes a period of $\mathcal{O}(L)$, and the total mixing time is expected to be $\mathcal{O}(Ld \log d)$ (with random-hopping between d pairs of replicas). By hopping sequentially in all the dimensions, the mixing time will be reduced to $\mathcal{O}(Ld)$. In chapter 3 we will discuss the constructions in more detail.

2.3 Lifting Scheme in Multi-particle Systems

The lifting scheme discussed in section 2.2 implies a faster dynamics of exploration of phase space. For more general physical systems concerning interactions and external fields, with the Boltzmann distribution as the stationary distribution, algorithms with lifting schemes are proposed [76], but an effective acceleration requires optimal connections of different schemes.

In the following, we will introduce the event-chain algorithm [9, 77] and a general factorized Metropolis filter [10, 11, 75] for the systems of multiple interacting particles.

2.3.1 Lifting Variable: Active Particle

We will continue with sampling of N particles on a one-dimensional lattice of L sites under periodic boundary condition. As in each replica a single particle moves in a single direction, we need at least $2N$ replicas to include all possible active particles and directions. In the following discussion, two lifting variables are denoted by $\{a, \sigma\}$, where $a = 1, 2, \ldots, N$ identifies the active particle, and $\sigma = \pm 1$ indicates the direction.

In section 2.2, the hopping between replicas is triggered either spontaneously (as in the uniform case) or by rejection (as in the non-uniform case). In the present model, the blocking by the particle in front will also trigger a hopping. However, if such a hopping only results in inverting the moving direction (σ) of the same active particle a as we discussed before, the mixing process will be confined to the interval created by two neighboring particles, and finally reduces to a local Glauber dynamics.

In expectation of a fast-mixing dynamics, we propose a scheme where the rejection triggers a "lifting" of the parameter a instead of σ (in other words, the one who rejects is activated in the next step). The Monte Carlo method based on this scheme, also known as "event-chain" Monte Carlo, was first used in the two-dimensional hard-disk model [9, 77], which shows advantage over the Metropolis method, and helps identify the liquid-hexatic phase transition [13] (as we will discuss in section 3.3.1).

The transition matrix consists of a shuffling process of the lifting variables $\{a, \sigma\}$,

and a propagative sampling as a generalization of the single-particle model:

$$T_{\text{shuffle}}(\{\mathbf{x};a_1,\sigma_1\}\to\{\mathbf{x};a_2,\sigma_2\}) = \frac{1}{2n} \text{ for } \sigma_1,\sigma_2=\pm1, \text{and } a_1,a_2=1,2,\ldots,N;$$
$$T_{\text{propagate}}(\{\mathbf{x};a_1,\sigma\}\to\{\mathbf{y};a_2,\sigma\}) =$$
$$\begin{cases} 1 & \text{if } x_{a_1+\sigma}\neq x_{a_1}+\sigma, a_1=a_2, y_{a_1}=x_{a_1}+\sigma, y_i=x_i|_{i\neq a_1}, \\ 1 & \text{if } x_{a_1+\sigma}=x_{a_1}+\sigma, a_1+\sigma=a_2, \mathbf{x}=\mathbf{y}, \\ 0 & \text{if others.} \end{cases} \tag{2.3.1}$$
$$\mathbf{T} = \alpha\mathbf{T}_{\text{shuffle}} + (1-\alpha)\mathbf{T}_{\text{propagate}}.$$

In eq. (2.3.1), the first case of $\mathbf{T}_{\text{propagate}}$ indicates no rejection, in which the active particle continues moving in the same direction; the second case implies that the next particle is activated if this move is rejected, as illustrated in Fig. 2.4. $\mathbf{T}_{\text{shuffle}}$ can be modified in many ways as long as it preserves the global-balance condition (e.g. σ and a can be treated separately, or the lifting variables (σ,a) must change their values after a shuffle). This method also applies to the hard-sphere system in continuous space (see section 3.2.2), and it can be proved that this dynamics has a mixing time of $\mathcal{O}(NL\log N)$ by choosing $\alpha=\frac{1}{L}$.

The scheme can be generalized for the non-uniform probability distribution as in section 2.2.1, and eventually for a continuous phase space as we will discuss below.

2.3.2 Factorized Metropolis Filter

For a system with other than the hard-core interactions, the idea of lifting also helps in the mixing process. As an important distinction from the hard-sphere problem, in general physical systems each valid configuration does not share the same weight in the steady-state distribution. K. S. Turitsyn et al. made an early effort in the irreversible sampling of an Ising model (we will introduce this model in section 4.1.1) based on the Metropolis-Hasting-Glauber method, by a lifting on the parameter σ [76]. However, as we discussed in section 2.3.1, changing directions σ alone fails to improve mixing times; the active particle a is also a key to the faster Monte Carlo dynamics.

In order to achieve the Boltzmann distribution, a similar method named "factorized Metropolis filter" [10, 11, 75], sharing similarities with the Metropolis algorithm in detailed balance is proposed. For a system of interacting particles, the energy related to one particle i is denoted by

$$E_i = \sum_{j\neq i} E_{ij}.$$

According to eq. (1.6.6) and eq. (1.6.9), for a proposed move, the Metropolis filter accepts the move with the probability

$$P^{\text{acc}}_{\text{Metro}}(\delta E_i) = \min(1, \mathrm{e}^{-\beta(\delta E_i)}) = \min\left(1, \prod_{j\neq i}\mathrm{e}^{-\beta(\delta E_{ij})}\right), \tag{2.3.2}$$

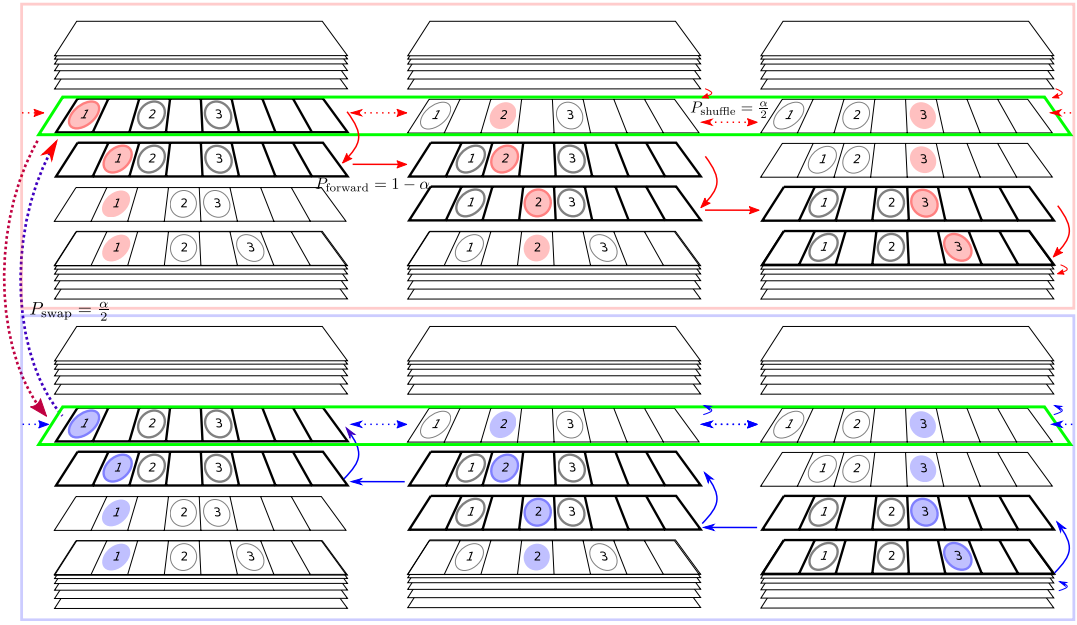


Figure 2.4: Random walk of N particles on the one-dimensional lattice under periodic boundary condition $\mathbb{Z}_L$ (not drawn explicitly). The scheme has $2N$ groups of replicas, corresponding to the directions (the red for moving forward and the blue for moving backward) and the active particles (the replicas with the same active particle are grouped by columns). In each step, the state will either (1) reshuffle the active particle (as the states inside the green boxes) and the direction (indicated by the dashed arrows), with a probability of α, or (2) move the active particle forward or trigger the one in front to be the new active one (indicated by the solid arrows, which is the time-line), with a probability of $1-\alpha$. Global balance is ensured in this scheme, and this system mixes in $\mathcal{O}(NL\log N)$ moves if $\alpha = 1/L$.

where δE indicates the change of energy.
The "factorized Metropolis filter" provides another rate of acceptance:

$$P^{\text{acc}}_{\text{FacMetro}}(\{\delta E_{ij}\}) = \prod_{j\neq i} \min(1, \mathrm{e}^{-\beta(\delta E_{ij})}) = \prod_{j\neq i} P^{\text{acc}}_{\text{Metro}}(\delta E_{ij}) < P^{\text{acc}}_{\text{Metro}}(\delta E_i), \qquad (2.3.3)$$

where the detailed-balance condition eq. (1.6.8) is satisfied for a symmetrically proposed move. Since the Metropolis algorithm maximizes the probability flow between different configurations, this "factorized version" performs no better in reversible Markov chains; the importance of factorization comes along with the option of "lifting".

The "event-chain" algorithm using the factorized Metropolis filter is described as follows (the current active particle $a = i$):

- A move of particle i is proposed, the factorized acceptance rate $P^{\text{acc}}_{\text{Metro}}(\delta E_{ij})$ is calculated with respect to all the other particles, and a group of random numbers $\{\eta_j \sim \text{unif}[0,1)\}$ is generated to determine the acceptance.

- If
$$\eta_j \leq P^{\text{acc}}_{\text{Metro}}(\delta E_{ij}),\ \forall j,$$
then this move is accepted; if there exists only one j such that
$$\eta_j > P^{\text{acc}}_{\text{Metro}}(\delta E_{ij}),$$
then this move is rejected, and j is the active particle in the next move.

We try to avoid multiple rejections at the same time, which means the step size should be small enough.

We consider the transition matrix of the event-chain algorithm, in the limit of an infinitesimal step ϵ, where

$$\mathrm{e}^{-\beta(\delta E_{ij}(\epsilon))} \sim 1 - \beta\delta E_{ij}(\epsilon).$$

(In the limit of matrix exponential, only the linear term determines the long-term properties, as shown in eq. (1.2.15).) For a general d- dimensional system of N particles, we need at least $2dN$ replicas. The transition matrix again consists of a shuffling matrix and a propagating matrix:

$$\begin{aligned}
&T_{\text{shuffle}}(\{\mathbf{x};\sigma_1,a_1\}\to\{\mathbf{x};\sigma_2,a_2\}) = \frac{1}{2Nd}, \forall\sigma_2\in\{\pm\mathbf{e}_1,\pm\mathbf{e}_2,\dots,\pm\mathbf{e}_d\}, \forall a_2\in\{1,2,\dots,N\};\\
&\qquad T^{\epsilon}_{\text{propagate}}(\{\mathbf{x};\sigma,a_1\}\to\{\mathbf{y};\sigma,a_2\});\\
&= \begin{cases} 1-\beta\sum_j \max[0,\delta E_{a_1 j}(\epsilon)] & \text{if } a_2 = a_1, y_{a_1} = x_{a_1}+\sigma\epsilon, y_i = x_i|_{i\neq a_1},\\ \beta\max[0,\delta E_{a_1 a_2}(\epsilon)] & \text{if } a_2\neq a_1, \mathbf{y}=\mathbf{x},\\ 0 & \text{if others;}\end{cases}\\
&\qquad \mathbf{T}^{\epsilon} = \alpha(\epsilon)\mathbf{T}_{\text{shuffle}} + [1-\alpha(\epsilon)]\mathbf{T}^{\epsilon}_{\text{propagate}}.
\end{aligned} \tag{2.3.4}$$

An optimal $\alpha(\epsilon)\sim\frac{\epsilon}{L}$ is estimated from previous discussions. This infinitesimal scheme preserves the conservation of probability and the global balance in the limit $\epsilon\to 0$, but it is somehow difficult to implement directly.

With the theory of the infinitesimal movement, we can then directly sample the distance in each step. Following E. Peters and G. de With [78], we define the *cumulative positive energy*:

$$E^+_{ij}(x_i - x_j) = \int_{r_0}^{x_i-x_j} \max(0, \frac{\mathrm{d}E_{ij}(r)}{\mathrm{d}r})\mathrm{d}r, \tag{2.3.5}$$

where (x_i, x_j) are the positions of the particle pair (i,j) (in one dimension), r_0 is a constant. The factorized Metropolis filter is then written as:

$$P^{\text{acc}}_{\text{FacMetro}}(x_i\to x_i+\delta x_i) = \prod_{j\neq i}\mathrm{e}^{-\beta\left(E^+_{ij}(x_i+\delta x_i - x_j) - E^+_{ij}(x_i-x_j)\right)}. \tag{2.3.6}$$

Inside each factor, we can directly sample ΔE^+_{ij}, and then find δx.

$$1 = \beta\int_0^\infty \mathrm{e}^{-\beta\Delta E^+}\mathrm{d}\Delta E^+ = \frac{1}{\beta}\int_0^\infty \mathrm{e}^{-\beta\Delta E^+}\frac{\mathrm{d}E^+}{\mathrm{d}x}\mathrm{d}x. \tag{2.3.7}$$

As $\frac{\mathrm{d}E^+}{\mathrm{d}x}$ is non-negative, $E^+(x)$ is monotonically non-decreasing, there exists an inverse function of $E^+(x)$

$$f(E) = \max\{x | E^+(x) = E\}.$$

Hence, using the idea of Bortz-Kalos-Lebowitz (BKL) algorithm [79], the event chain method in the continuous scheme is described as follows:

- In the current replica, the active particle i moves in the direction σ. For each other particle j that interacts with i, an energy fluctuation ξ_j is generated from the random number generator

  $$\xi_j = -\frac{1}{\beta} \log \mathrm{ran}(0,1], \tag{2.3.8}$$

  and its corresponds to a proposed move

  $$\delta x_i|_j = f(E_{ij} + \xi_j) - f(E_{ij}). \tag{2.3.9}$$

  After the particle i goes through the shortest displacement $\delta x_i|_k$, the first rejection is triggered, and the next active particle k is determined (the probability of triggering multiple particles simultaneously is extreme small):

  $$\delta x_i|_k = \min\{\delta x_i|_j, \forall j \neq i\}. \tag{2.3.10}$$

  This process continues and forms one "chain".

- Each "chain length" (total displacement of all particles) follows an exponential distribution. Once it reaches the limit, the particle stops (before a new event is triggered), and the lifting parameters (the active particle a and the direction σ) are shuffled.

This dynamics is named as "event-chain" algorithm as the rejection events form a chain in the simulation. In chapter 4 an illustration of the algorithm in the XY model is shown (Fig. 4.1).

2.4 Global Balance in Lifting and Stopping Times

In the algorithms of section 2.2 and 2.3, replicas are shuffled with a certain probability at each step. This leads to a Poisson process in the continuous limit. The transition matrix of this procedure is $\mathbf{T}_{\text{shuffle}}$ in eqs (2.3.1) and (2.3.4), which obeys the global-balance condition of eq. (2.1.1), and so does the other transition matrix $\mathbf{T}_{\text{propagate}}$.

For any two transition matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ following the global-balance condition in eq. (2.1.1), their normalized linear combinations and multiplications as new transition matrices, also preserve the same global-balance condition:

$$\begin{aligned} \boldsymbol{\pi} = \boldsymbol{\pi}[\alpha \mathbf{T}_1 + (1-\alpha)\mathbf{T}_2], \quad \text{where } \alpha \in (0,1), \\ \boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{T}_1\mathbf{T}_2. \end{aligned} \tag{2.4.1}$$

Irreducibility and aperiodicity are not specified in the equations: even the identity matrix $\mathbf{1}$ can be taken as a transition matrix, though it does not help in the mixing.

We consider a dynamics:

$$\boldsymbol{\pi}^{(\sum t_i)} = \boldsymbol{\pi}^{(0)} \mathbf{T}_{\text{propagate}}^{t_1} \mathbf{T}_{\text{shuffle}} \mathbf{T}_{\text{propagate}}^{t_2} \mathbf{T}_{\text{shuffle}} \dots \tag{2.4.2}$$

where $\{t_i\}$ is a sequence of stopping times that determines the "chain length" in the event-chain algorithm, following some configuration-independent distribution (e.g. the exponential distribution in all previous lifting schemes). For certain distributions of stopping times, the mixing process could be even faster with respect to the system, which we will present in section 3.2.1. The dynamics also applies to discrete time evolution, which we will discuss in chapter 3.

2.4.1 Weaker Condition: Infinite Chain

In the model of multiple one-dimensional random walkers, the transition matrix $\mathbf{T}_{\text{propagate}}$ preserves the global balance condition. However, it alone does not guarantee the irreducibility: it may only explore a one-dimensional subspace of Ω. The resampling of directions and active particles ($\mathbf{T}_{\text{shuffle}}$) provides better decorrelation and necessarily preserves the irreducibility.

In some special periodic systems with one-dimensional continuous random variables where the $\mathbf{T}_{\text{propagate}}$ is irreducible, the event chain is able to lead the system to its steady-state distribution in an infinite run. This applies to the harmonic model, the XY model [11] and many others, which we will show in chapter 4 and 5.

Nevertheless, there are examples that the "infinite chain" does not converge to the stationary distribution: in the one-dimensional hard-sphere model, the infinite event-chain will result in a Newtonian movement with elastic collisions, and the distance between particles is kept unchanged (as shown in section 3.1.3). With the "factor field" technique introduced in chapter 5, the irreducibility of an infinite chain is preserved in such systems.

2.4.2 Speed Limit of Irreversible Markov Chains

In order to achieve fast mixing, irreversible Markov chains including the probability-flow construction and lifting scheme, aim for the maximization of the net probability flow in all the states, or the "conductance"/"bottleneck ratio" of the graph associated to the state space [1] .

For a non-trivial subset $S \subset \Omega$, the stationary probability of S is defined as the "capacity":

$$\pi(S) = \sum_{i \in S} \pi_i.$$

With the transition matrix $\mathbf{T}$, the "ergodic flow" out of S (or the "edge measure" from S to $\bar{S}$) is defined as

$$F(S) \equiv \sum_{i \in S, j \in \bar{S}} \pi_i T_{ij},$$

[1] In this section we refer to *Approximate Counting, Uniform Generation and Rapidly Mixing* by A. Sinclair & M. Jerrum [80] and *Markov Chains and Mixing Times* by D. Levin, Y. Peres & E. Wilmer [36, section 7.2]. The notations of the same term are sometimes different in these two materials, we will mention both of them in the text.

where $\bar{S}$ is the complement set of S. It is self-evident that $F(S) = F(\bar{S})$. The "conductance"/"bottleneck ratio" of S is defined as

$$\Phi(S) \equiv F(S)/\pi(S),$$

and the global conductance/bottleneck ratio of the Markov chain (denoted by Φ) is the minimum $\Phi(S)$ of all possible S, which determines the slowest probability-flow exchange between all the states:

$$\Phi \equiv \min_{S\in\mathcal{E}(\Omega),\, \pi(S)\leq 1/2} \Phi(S).$$

This is the term we tried to maximize throughout this chapter with the lifting schemes, and it is closely related to the relaxation and mixing times of Markov chains. It is shown that

$$1-2\Phi \leq \lambda^* < 1-\frac{\Phi^2}{2},$$

for reversible Markov chains [81], and

$$t_{\text{mix}}(1/4) \geq \frac{1}{4\Phi},$$

for general Markov chains [36].

The global flow is limited by the conductance, as well as the smallest capacity of system

$$\pi_0 = \min_{x\in\Omega}\{\pi(x)\}.$$

The upper limit of mixing time in a Markov chain is bounded to a factor of $\mathcal{O}(\frac{1}{\pi_0})$ [81].

One example showing the importance of conductance is a "bottleneck" in a graph (as in Fig. 2.5 (a)): due to the probability-flow limit of the "bottleneck", any local Markov chain will encounter a slowing down, which is not solved by the lifting scheme.

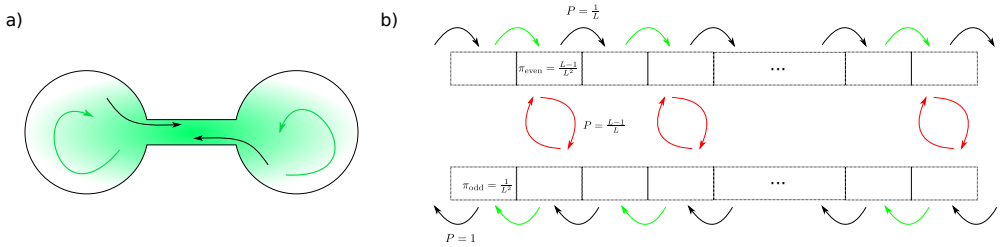


Figure 2.5: (a) An illustration of a "bottleneck" configuration. The thin path in between constrains the conductance of the graph. (b) Lifting in a non-uniform one-dimensional lattice $\mathbb{Z}_{2L}$, described by eq. (2.4.3). The odd sites ($x = 1, 3, \ldots$) play the roles of bottlenecks, and the main flows (red) are restricted to local circuits (hopping between replicas on the same site), resulting in a non-directional diffusive walk.

In an extreme case of the random walk on a graph with non-uniform weight as mentioned in section 2.2.1, a one-dimensional lattice $\mathbb{Z}_{2L}$ has the following steady state:

$$\pi(x) = \begin{cases} \frac{1}{L^2} & \text{for } x = 1, 3, \ldots, \\ \frac{L-1}{L^2} & \text{for } x = 2, 4, \ldots. \end{cases} \tag{2.4.3}$$

The probability flow will be mostly trapped on the sites $x = 2, 4, \ldots$ in the lifting scheme, and the lifted Markov chain will once again fall into a diffusive process (see Fig. 2.5 (b)).

Though these types of limits can not be overcome with local Markov chains, a well-designed lifting scheme with an optimal stopping rule has a general advantage over the conventional reversible methods [82]. In the following chapters of the thesis, we will continue to explore such algorithms .

2.5 Conclusion

In the present chapter we discussed the algorithms based on the global-balance condition, and presented lifting schemes which construct probability-flow circuits in state space by introducing a set of lifting variables. Such a scheme has a mixing time of $\mathcal{O}(L)$ in the one-dimensional single-particle random walk, and it is generalized to the "event-chain" algorithm in the multi-dimensional systems of multiple particles with interaction. The "factorized Metropolis filter" applied in the event-chain algorithm can accelerate the mixing process with a Boltzmann stationary distribution. In the end, we discussed the role of global conductance in Markov chains and its connection to lifting schemes.

In the remainder of the thesis (chapters 3, 4 and 5), we will apply the lifting scheme and the event-chain algorithm presented in this chapter, to hard-sphere systems, spin systems and soft-sphere systems, in order to study their mixing and dynamical scaling properties.

Chapter 3

Irreversible Markov Chains in Hard-sphere Models

In chapter 2 we introduced the "lifting" concept to one-dimensional sampling and to the Metropolis algorithm for canonical ensembles, and presented the "event-chain" algorithm for general systems. This method has been widely applied to physical systems, although there is no exact description of its mixing and equilibration time scales.

In this chapter, we will present an exact calculation of the mixing time of the event-chain Monte Carlo algorithm with special stopping times in the one-dimensional hard-sphere model. Related results also help characterize mixing properties of the lifted Metropolis algorithm and improve its speed. It leads us to a new variant of the event-chain algorithm as well, which is applicable to higher-dimensional hard-sphere models.

3.1 Global Balance in One-dimensional Hard Disks

We begin with a one-dimensional mono-disperse hard-sphere system of size L under periodic boundary condition with the N spheres of diameter d, as illustrated in Fig. 3.1 (a). In this dynamics no hopping over the neighbors is allowed (the order of spheres is kept unchanged). The positions of the centers of the spheres, $\{x_1, \dots, x_N\}$ in one configuration are N ordered random variables.

A mapping exists between the model of hard spheres on a circle of length L and the model of the same number of points on a smaller circle of size $\tilde{L} = (L - Nd)$ (see Fig. 3.1 (b)). There are $(N-1)$ degrees of freedom in this model other than a global shift. We may choose the free distance between 2 neighboring spheres $\{\delta_i, i = 1, \dots, N-1\}$ to describe the configuration, with the notation

$$\delta_i = x_i - x_{i-1} - d, \tag{3.1.1}$$

which also corresponds to the distance between the particle $i-1$ and i.

Assuming that the relaxation from a compact configuration (e.g. $\delta_i = 0, i = 1, \dots, N-1$) indicates the mixing time of a local algorithm, we introduce $u_i = \sum_{i=1}^{N/2} \delta_i$ as the "half-system distance" (its probability distribution is discussed in section 3.2.2). In

this chapter, the numerical evaluations of the so-defined mixing times are provided with the evaluation of half-system distance.

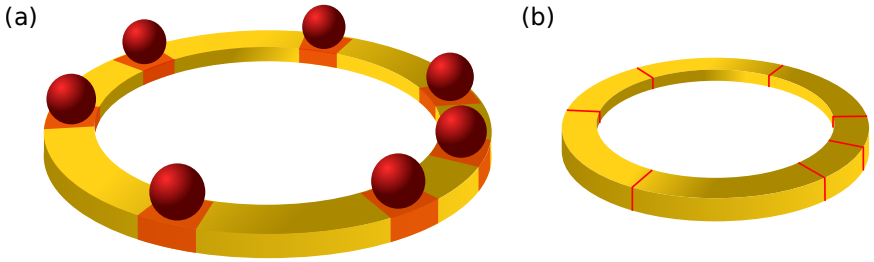


Figure 3.1: One-dimensional hard-sphere models with periodic boundary conditions. (a): N spheres of diameter d on a ring of length L. (b): N volume-less particles on a ring of length $\tilde{L} = (L - Nd)$. The dynamics are equivalent for both representations, as long as no hopping is allowed. [2]

The dynamics based on the global-balance condition concerns the conservative probability flows in the steady-state distribution, as indicated in section 2.1 (eq. (2.1.1) to 2.1.3). In hard-sphere system, every valid configuration (without overlapping) has equal statistical weight, and any invalid configuration has the probability of zero. We omit the normalization weight and take the probability of each valid configuration as

$$\pi(\{\delta_j\}) = 1,$$

in the steady-state distribution, and then focus on the relation between the transition matrix and the probability flows. [3]

We consider the probability flow $\mathcal{F}(\{\delta_j\})$ into a specific configuration $\{\delta_j\}$,

$$\mathcal{F}(\{\delta_j\}) = \sum_{i=1}^{N} \mathbf{P}(a=i)\mathcal{F}_i(\{\delta_j\}), \ \sum_{i=1}^{N} \mathbf{P}(a=i) = 1, \tag{3.1.2}$$

where a is the label of the active sphere (we use the same notation as in section 2.2), and $\mathcal{F}_i(\{\delta_j\})$ stands for the probability flow coming from an attempted move of the active sphere i to this configuration (on the condition $a = i$). (In the following, we omit the notation of $\{\delta_j\}$ when it cannot be confused with other configurations.)

With the notations of accepted and rejected flows in eqs (2.1.2) and (2.1.3), each $\mathcal{F}_i$ can be split into four parts, namely the forward/backward accepted/rejected flows (see Fig. 3.2),

$$\mathcal{F}_i = \sum_{\sigma=\pm 1} (\mathcal{A}_i^{\sigma} + \mathcal{R}_i^{\sigma}), \tag{3.1.3}$$

where σ is a random variable indicating the direction of the attempted move.

[2] This figure is taken from Fig. 1 in Publication II [17].

[3] The discussion of the following dynamics, including the sequential Metropolis algorithm, forward Metropolis algorithm, lifted forward Metropolis algorithm with and without restarts, is largely based on [18]. However, the notation is changed for a more general probability distribution of a step ϵ, and the probability flow set-up in the lifting scheme is consistent with chapter 2 but different from [18].

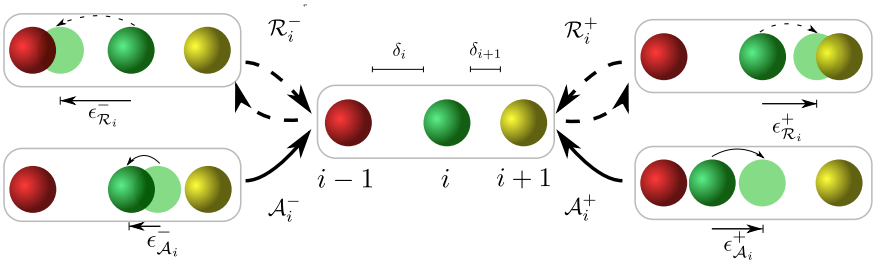


Figure 3.2: The probability flow into a configuration *via* an attempted move of the sphere i (green). $\mathcal{A}_i^{\pm}$ are the forward/backward accepted probability flows, and each of them is obtained by an accepted displacement of the sphere i. $\mathcal{R}_i^{\pm}$ stands for the forward/backward rejected probability flows respectively, and it is obtained by a rejected attempt of the sphere i, from the current configuration. δ_i indicates the free space between sphere $i-1$ and i. [5]

The proposal of each step ϵ follows a probability distribution, with the probability density $p(x)$. (We assume $p(0)$ is finite, so $\mathbf{P}(\epsilon = 0) = 0$. Generally $p(x)$ can be conditional on the current configuration and the active sphere as $p(x; \{\delta_j\}, a = i)$. In the discussion of the Metropolis algorithm, we assume it is independent of these two factors). For a given step of ϵ, we have the weights for the probability flows:

$$\begin{aligned} \mathcal{A}_i^+(\epsilon) &= \Theta(\delta_i - \epsilon), \quad \mathcal{A}_i^-(\epsilon) = \Theta(\delta_{i+1} + \epsilon), \\ \mathcal{R}_i^+(\epsilon) &= \Theta(\epsilon - \delta_{i+1}), \quad \mathcal{R}_i^-(\epsilon) = \Theta(-\epsilon - \delta_i), \end{aligned} \tag{3.1.4}$$

where Θ is the Heaviside function. The accepted flows are given by the integrals:

$$\mathcal{A}_i^+ = \int_0^\infty p(x)\mathcal{A}_i^+(x)\mathrm{d}x, \quad \mathcal{A}_i^- = \int_{-\infty}^0 p(x)\mathcal{A}_i^-(x)\mathrm{d}x, \tag{3.1.5}$$

and similarly for the rejected flows.

Based on these notations, a reversible Markov chain, for example the *heat-bath algorithm* (or the *Glauber dynamics*) in section 1.6.2 which randomly picks a sphere and places it between its neighbors with a uniform distribution, preserves the detailed-balance condition:

$$\mathbf{P}(a = i)p(x; \{\delta_j\}, a = i)\mathcal{A}_i(x; \{\delta_j\}) = \mathbf{P}(a = i)p(-x; \{\delta'_j\}, a = i)\mathcal{A}_i(-x; \{\delta'_j\}). \tag{3.1.6}$$

where $\{\delta'_j\}$ is the configuration obtained from $\{\delta_j\}$ after the move $(+x, a = i)$. The heat-bath algorithm has a mixing time[6] of $\mathcal{O}(N^3 \log N)$, and belongs to the "SEP class" in section 1.5.4.

In the following, we will discuss the global-balance algorithms for the system. These algorithms also have similar properties in the further applications to soft-sphere systems.

[5] This is a remake of Fig. 2 in [18].

[6] The result is not rigorously known. The mixing time $\mathcal{O}(N^3 \log N)$ is proven in fixed boundary condition, but it is proven between $\mathcal{O}(N^3)$ and $\mathcal{O}(N^3 \log N)$ in periodic boundary condition [51, 52]. The numerical result indicates the latter [18].

3.1.1 Sequential Algorithms

Metropolis Algorithm with Symmetric Proposal

From eqs (3.1.4) and (3.1.5), we see that

$$\begin{aligned}\mathcal{R}_i^- &= \int_{-\infty}^{0} p(x)\Theta(-x-\delta_i)\mathrm{d}x = \int_0^{\infty} p(-x)\Theta(x-\delta_i)\mathrm{d}x \\ &= \int_0^{\infty} p(-x)[1-\mathcal{A}_i^+(x)]\mathrm{d}x = \int_{-\infty}^{0} p(x)\mathrm{d}x - \tilde{\mathcal{A}}_i^+,\end{aligned} \tag{3.1.7}$$

where $\tilde{\mathcal{A}}_i^+ = \mathcal{A}_i^+$ if $p(x) = p(-x)$. (In the derivation of eq. (3.1.7), we used the identity equation: $\Theta(x-\delta)+\Theta(\delta-x)=1$.)
In other words, if $p(x)$ is *even* (that $p(x)=p(-x)$), we have

$$\mathcal{R}_i^{\mathrm{sym}-} + \mathcal{A}_i^{\mathrm{sym}+} = \mathbf{P}(\epsilon<0) = 1/2,\ \mathcal{R}_i^{\mathrm{sym}+} + \mathcal{A}_i^{\mathrm{sym}-} = \mathbf{P}(\epsilon>0) = 1/2.$$

Hence,

$$\mathcal{F}_i^{\mathrm{sym}} = \mathcal{R}_i^{\mathrm{sym}-} + \mathcal{A}_i^{\mathrm{sym}+} + \mathcal{R}_i^{\mathrm{sym}+} + \mathcal{A}_i^{\mathrm{sym}-} = 1. \tag{3.1.8}$$

According to eq. (3.1.2),

$$\mathcal{F}^{\mathrm{sym}} = \sum_{i=1}^{N} \mathcal{F}_i^{\mathrm{sym}} \mathbf{P}(a=i) = \sum_{i=1}^{N} \mathbf{P}(a=i) = 1, \tag{3.1.9}$$

so that the total probability flow is always 1, for all possible probability distribution of the active sphere $\mathbf{P}(a=i)$. It also implies a preservation of the global-balance condition for any sequence of the active sphere labels $\{a(t), t=1,2,\dots\}$, as long as the distribution of each step ϵ is symmetric. (Nevertheless there are other conditions to guarantee the irreducibility, e.g. each sphere will be picked infinite times in the process.)

In a special case that the active sphere is picked from 1 to N periodically, it is the *sequential Metropolis algorithm* that we mentioned in section 2.1. Empirically, it is faster than the reversible Metropolis algorithm, but only by a constant factor [74]. It is observed to mix in $\mathcal{O}(N^3 \log N)$ steps, and it belongs to the "SEP class" in section 1.5.4, just as the Glauber dynamics.

Sequential Heat-bath Algorithm

In the Glauber dynamics, the distribution of ϵ has dependence on the configuration:

$$p(x; a=i) = \begin{cases} \frac{1}{\delta_i+\delta_{i+1}} & \text{if } -\delta_{i+1} \le x \le \delta_i, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1.10}$$

We have

$$\begin{aligned}\mathcal{R}_i^{\mathrm{Glauber}-} &= \mathcal{R}_i^{\mathrm{Glauber}-} = 0, \\ \mathcal{F}_i^{\mathrm{Glauber}} &= \mathcal{A}_i^{\mathrm{Glauber}+} + \mathcal{A}_i^{\mathrm{Glauber}-} = 1.\end{aligned} \tag{3.1.11}$$

The total probability flow in eq. (3.1.2) equals 1, for any $\mathbf{P}(a=i)$. Hence a *sequential heat-bath algorithm* is valid as well, which also mixes in $\mathcal{O}(N^3 \log N)$ steps and belongs to the "SEP class" in section 1.5.4.

3.1.2 Forward Metropolis

If the $p(x)$ has a bias on either side, $\mathcal{F}_i = 1$ no longer holds true for a general configuration $\{\delta_i\}$, and an arbitrary choice of the active sphere may lead to $\mathcal{F} \neq 1$. Nevertheless, if $\mathbf{P}(a = i)$ is chosen wisely, the probability flows coming from different active spheres may compensate.

We now search for the appropriate flows satisfying eq. (3.1.2). It is deduced from eq. (3.1.4):

$$\begin{gathered} \mathcal{A}_i^+(\epsilon) + \mathcal{R}_{i-1}^+(\epsilon) = 1, \;\; \mathcal{A}_i^-(\epsilon) + \mathcal{R}_{i+1}^-(\epsilon) = 1, \\ \mathcal{A}_i^+ + \mathcal{R}_{i-1}^+ = \mathbf{P}(\epsilon > 0), \;\; \mathcal{A}_i^- + \mathcal{R}_{i+1}^- = \mathbf{P}(\epsilon < 0), \\ \mathcal{A}_i^+ + \mathcal{R}_{i-1}^+ + \mathcal{A}_i^- + \mathcal{R}_{i+1}^- = 1. \end{gathered} \tag{3.1.12}$$

Hence, if we take $\mathbf{P}(a = i) = 1/N, i = 1, \ldots, N$:

$$\mathcal{F}^{\text{equ}} = \frac{1}{N}\sum_{i=1}^{N}(\mathcal{A}_i^+ + \mathcal{R}_i^+ + \mathcal{A}_i^- + \mathcal{R}_i^-) = \frac{1}{N}\sum_{i=1}^{N}(\mathcal{A}_i^+ + \mathcal{R}_{i-1}^+ + \mathcal{A}_i^- + \mathcal{R}_{i+1}^-) = 1. \tag{3.1.13}$$

The total probability flow is 1 again. Similar to the previous discussion, global balance is preserved for any probability distributions $p(x)$ or any sequence of the probability distributions $\{p^{(t)}(x)\}$ ($p^{(t)}(x)$ is the distribution of ϵ at time t, and it is independent of the label of the active sphere i), as long as each sphere has equal probability to be activated. (Nevertheless there are other conditions to guarantee the irreducibility, e.g. $p(x)$ is not a countable sum of Dirac's δ functions.)

The *forward Metropolis algorithm* corresponds to the special case where the spheres only move in the positive direction ($\mathbf{P}(\epsilon < 0) = 0$, see Fig. 3.3). Empirically, its mixing time is $\mathcal{O}(N^{\frac{5}{2}})$ [18], and belongs to the "TASEP" class indicated in section 1.5.4.

As a completement to our discussion, the sequential variant in section 3.1.1 is not compatible with the forward Metropolis algorithm; in other words, a "sequential forward Metropolis algorithm" that chooses the active sphere sequentially and proposes it to move forward at each time does not preserve the global-balance condition, as the probability flow coming from the single active sphere i does not necessarily equal 1 in general cases:

$$\mathcal{F}_i = \mathcal{A}_i^+ + \mathcal{R}_i^+ + \mathcal{A}_i^- + \mathcal{R}_i^- \neq 1.$$

3.1.3 Lifted Forward Metropolis (without restarts)

Based on the discussion in chapter 2, eq. (3.1.12) is considered with the lifting scheme in section 2.3.1, where a rejection triggers a lifting in the replicas, namely the activation of the rejecting sphere. Similar to eq. (3.1.2), we divide the total probability flow into:

$$\mathcal{F} = \sum_{s=\pm 1}\sum_{i=1}^{N}\mathcal{F}(a = i, \sigma = s), \tag{3.1.14}$$

where $\mathcal{F}(a = i, \sigma = s)$ (it may be denoted as $\mathcal{F}(i, s)$ if it cannot be confused with the other probability flows) stands for the probability flow arriving at the same configuration ($\{\delta_j\}$), with i as the *current* active sphere and s as the current moving direction. It can also be viewed as the probability flow in the replica denoted by $(a = i, \sigma = s)$.

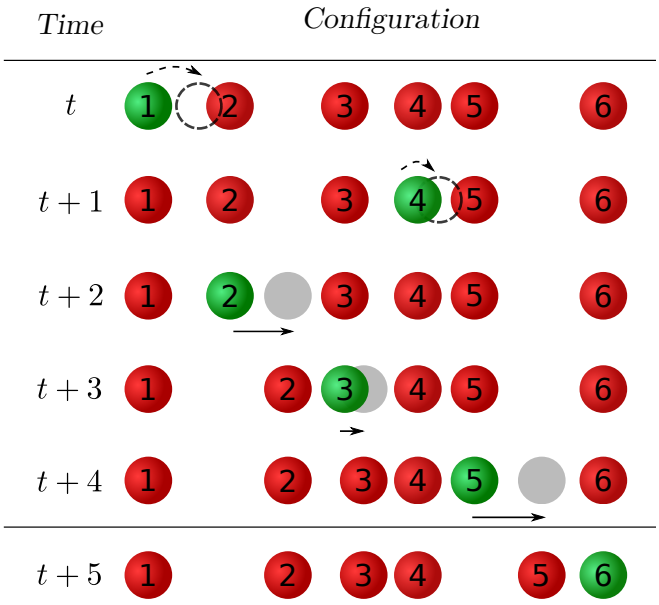


Figure 3.3: The scheme of forward Metropolis algorithm. At each time step, a random sphere (the green one) attempts to move forward. This move is rejected if it overlaps or goes beyond the next one (as at time t and $t+1$), and it is accepted otherwise (as at $t+2, t+3$ and $t+4$).

($\mathcal{F}(a=i)$ is different from $\mathcal{F}_i$ in sections 3.1.1 and 3.1.2, where i indicates the *previous* active sphere.)
In the lifting scheme without shuffles, σ is always a constant. Without loss of universality, in the following discussion we assume that $\sigma = 1$, $\mathbf{P}(\epsilon > 0) = 1$. Using the same notation from eq. (2.2.1)

$$\pi(\{\delta_j\}) = \sum_{i=1}^{N} \pi(\{\delta_j\}, a = i, \sigma = +1),$$

we assume that each replica has the same weight ($\pi(\{\delta_j\}, a = i) = 1/N$). As the rejection of the sphere $(a - \sigma)$ triggers the activation of the sphere a, we have

$$\mathcal{F}(i,1) = \pi(a=i)\mathcal{A}_i^+ + \pi(a=i-1)\mathcal{R}_{i-1}^+ = \frac{1}{N}\mathbf{P}(\epsilon > 0) = \pi(a=i), \tag{3.1.15}$$

the probability flows in the lifting scheme preserve global-balance condition as well.

This algorithm is known as *lifted forward Metropolis algorithm*, at each time step the active sphere i is proposed to move forward with a distance of ϵ. If the sphere i overlaps with the next, then this move is rejected and the next sphere $i+1$ is activated; otherwise the move is accepted, and the sphere i remains active. The process is illustrated in each single row of Fig. 3.4.

It is seen in the simulations that the *lifted forward Metropolis algorithm (without restarts)* mixes in a time of $\mathcal{O}(N^{5/2})$ [18]. Hence, similar to the forward Metropolis algorithm, it is expected to belong to the "TASEP class" in the discrete space.

[8] This figure is adapted from the Fig. 6 in Publication II [17].

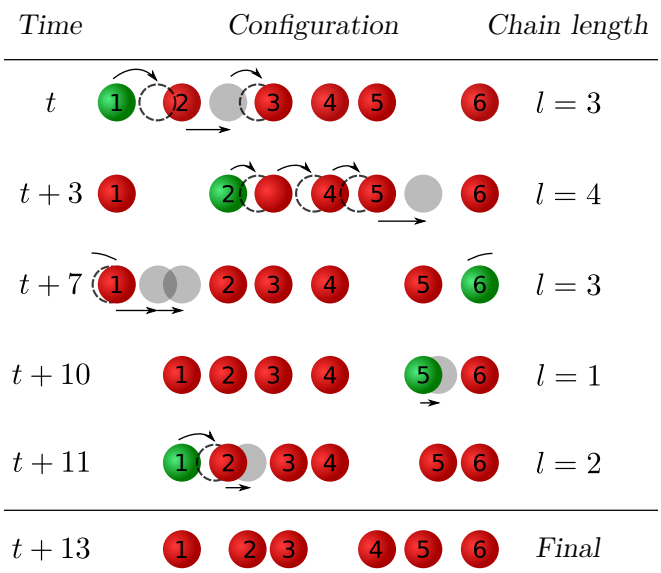


Figure 3.4: The scheme of the lifted forward Metropolis algorithm with restarts. At each time step a random sphere (the green one) attempts to move forward. This move is rejected and triggers the next sphere if it overlaps/goes beyond the next one (indicated by the curved arrows and dashed spheres), or accepted if not (indicated by the straight arrows and the shady spheres). The randomized stopping time for shuffle is denoted as "chain length" in the figure. [8]

Irreducibility of Lifted Forward Metropolis Algorithm

In the special case of infinitesimal move ϵ in the lifted forward Metropolis algorithm, the dynamics becomes a deterministic infinite event chain (mentioned in section 2.4.1), as a prime example for an algorithm that satisfies global-balance condition, but violates irreducibility. Nevertheless, the lifted forward Metropolis algorithm preserves irreducibility, even if ϵ has a cut-off ϵ_{max} in the tail of its distribution:

$$\mathbf{P}(\epsilon > \epsilon_{\text{max}}) = 0,\ 0 < \epsilon_{\text{max}} < \tilde{L}.$$

In the following, we will prove this argument with the model of volume-less particles, whose dynamics is identical to that of hard spheres. The positions of the particles $(x_1, \dots, x_N)$ follow the restriction $x_1 \leq \cdots \leq x_N \leq x_1 + \tilde{L}$.

We will firstly show that there exists a route to reach a "condensed" state ($x_1 = \cdots = x_N$) from any possible initial configuration $(x_1, \dots, x_N)$, *if ϵ has a continuous distribution between 0 and any finite ϵ_{max}*. Without loss of generality, we assume the first particle is active in the initial configuration, and the distance between the nearest neighbors is larger than ϵ_{max} (the route can be generalized without this restriction). In this route, the first one stops at the position $x_2 - \epsilon_{\text{max}}$, and the second one stops at the position $x_3 - \epsilon_{\text{max}}$, so on and so forth, in the end a configuration $(x_2 - \epsilon_{\text{max}}, \dots, x_N - \epsilon_{\text{max}}, x_1)$ is reached (which is $(x_1, x_2 - \epsilon_{\text{max}}, \dots, x_N - \epsilon_{\text{max}})$ after a translation), while the particle at x_1 is active. The new configuration is similar to the initial one, with all the other particles "moved backwards" to x_1. After finite similar

iterations, a configuration

$$(x_1, x_1, x_3' = x_3 - (x_2 - x_1), \ldots, x_N' = x_N - (x_2 - x_1))$$

is obtained. Again, the second one stops at $x_3' - \epsilon_{\text{max}}$, the third one stops at $x_4' - \epsilon_{\text{max}}$, so on and so forth, the one before the last stops at $x_N' - \epsilon_{\text{max}}$, and the last one stops at x_1: the positions of the other particles are shifted to x_1 by the same distance ϵ_{max} as before, and 3 particles meet at x_1 eventually. The similar procedure can be done for all the other $(N-3)$ particles, and in the end a condensed configuration is achieved.

On the other hand, due to the time reversibility of the dynamics shown in Fig. 3.5, any possible configuration is reachable from a condensed configuration, if it can collapse to a condensed one. For a lifted Metropolis process moving in the positive direction (as Fig. 3.5 (a)), its time reversal is also a lifted Metropolis process, with the same acceptance and rejection on each move but in the negative direction (as Fig. 3.5 (b)). If we apply a space reflection to the time reversal, a lifted Metropolis process moving in the positive direction is constructed (as Fig. 3.5 (c)), which has the initial state as the final state of (a) (reflected) and *vice versa*. Since the condensed configuration is invariant under reflection, its connectivity to any other states is verified by our first argument.

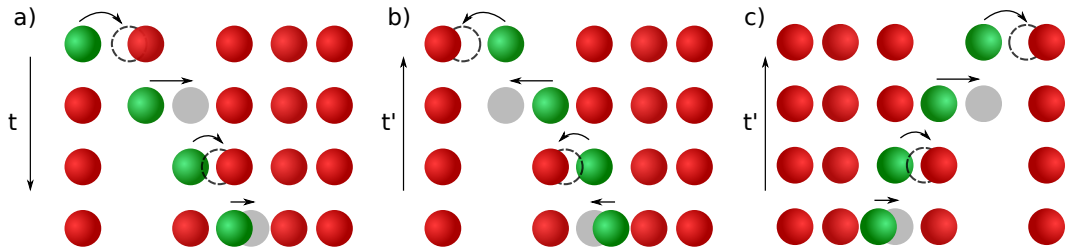


Figure 3.5: The route construction by time reversal. (a): A dynamics in lifted forward Metropolis algorithm moving in the positive direction. (b): The time reversal of (a), is a lifted forward Metropolis dynamics moving in the negative direction. (c): The reflection of (b), as a lifted forward Metropolis dynamics moving in the positive direction, with the final state as the initial one in (a) and the initial state as the final one in (a).

Using the condensed configuration as a special "node" in the graph of the states, we prove that each state is strongly connected to this configuration; then any two states are strongly connected, which indicates the irreducibility of the algorithm. Nevertheless, in the event-chain dynamics where $\epsilon_{\text{max}} = 0$, restarts are needed essentially. We will continue this discussion in the following.

3.1.4 Lifted Metropolis (with restarts) and Event-chain Algorithm

As we proposed in section 2.4, if there are multiple transition matrices preserving the global-balance condition, their linear combinations and multiplications will do the same.

Mixing time	Discrete	Continuous
$\mathcal{O}(N^3 \log N)$	SEP*[44]	Metropolis, sequential Metropolis, Glauber dynamics*[51, 52]
$\mathcal{O}(N^{5/2})$	TASEP*[50]	forward Metropolis[18], lifted forward Metropolis[18]
$\mathcal{O}(N^2 \log N)$	lifted TASEP [18] (w/restarts)	lifted forward Metropolis (w/restarts)[18, 17], event-chain algorithm (w/restarts)*,[18, 17] lifted sequential forward Metropolis (w/swaps & restarts)[17],
$\mathcal{O}(N^2)$		event-chain algorithm (w/swaps & restarts)* [17]

Table 3.1: The classification of algorithms according to their mixing time. Rigorous results are indicated with an asterisk (*).

The "shuffle" process of the active particle or the moving directions preserves the global-balance condition, as:

$$\pi(a,\sigma) = \sum_{i=1}^{N} \sum_{s=\pm 1} \pi(i,s) T_{\text{shuffle}}[(i,s) \to (a,\sigma)], \tag{3.1.16}$$

which works for any non-negative transition matrix $\mathbf{T}_{\text{shuffle}}$. We may choose

$$T_{\text{shuffle}}[(i,s) \to (i,s)] = 0,$$

to force the lifting parameters to change.

Global balance is preserved if the transition matrix of shuffling is mixed with those of the lifted Metropolis algorithm (see Fig. 3.4). In the lifted Metropolis algorithm with restarts, the mixing time is $\mathcal{O}(N^2 \log N)$, when $\epsilon \sim \tilde{L}/N$ and the "chain length" (the number of steps between two shuffles) is best chosen as $l \sim N$ [18].

In its continuous limit, the *event-chain algorithm with restarts* also has the mixing time of $\mathcal{O}(N^2 \log N)$, if the chain length $l \sim \tilde{L}$. These two dynamics belong to the "lifted TASEP class" mentioned in section 1.5.4.[9]

So far, we have explored a number of global-balance-based algorithms, and categorized them in Table 3.1 (it includes the algorithms we will introduce in section 3.2.3). In the following sections we will study these algorithms with a specific stopping rule, and analyze their convergence.

3.2 Mixing in One Dimension

In this section, we will analyze the mixing properties of the lifted dynamics. To simplify the problem, we will start from its continuous limit, the event-chain algorithm.

[9] These classes are specified in [18]. However, the existence of the definite classes is not clear. For example, the discrete and the continuous algorithms with swaps in section 3.2.3 have different time scale, although they are linked by a continuous change of parameters.

We designate the normalized partial sums of the sequence $\{\delta_i\}$ in eq. (3.1.1) by

$$s_i = \frac{\sum_{j=1}^{i} \delta_j}{L - Nd} = \frac{\sum_{j=1}^{i} \delta_j}{\tilde{L}}, \tag{3.2.1}$$

where $\{s_i\}$ corresponds to an ordered sequence of random variables on the interval $[0,1)$. (e.g. in Fig. 3.1 (b), the perimeter of the circle is 1.)

In equilibrium without ordering (all $(N-1)!$ permutations of $\{s_i\}$ are considered), every particle in Fig. 3.1 (b) has an independent uniform distribution on the domain. We base the following discussions on the stationary distribution.

3.2.1 Single-particle Case

In the simplest case where the exact calculation of the TVD is possible, we firstly consider the sampling of a single particle. In section 1.5.3 we discussed the random walk/Brownian walk of a single particle in a loop, and in section 2.2.1 we presented the lifted walk with an exponentially distributed "chain length". Here we will continue the calculation of the mixing time of a single one-dimensional walker in event-chain dynamics with a more general chain length.

The chain length is denoted by ϵ_i ($i = 1, 2, \ldots$, which is proportional to the stopping time), and different chains have the identical distribution (the probability density is written as $p_\epsilon(x)$ or simply $p(x)$) for their lengths.

Walk in a Fixed Direction

To begin with, we consider the total displacement in only one direction as the "infinite chain" in section 2.4.1 (hence $\mathbf{T}_{\text{shuffle}}(\sigma)$ in eq. (2.3.1) becomes an identity transformation). In this case, the dynamics is deterministic, as there is only one possible state at any given time. Optionally we use a randomized stopping rule to evaluate the probability distribution after a given number of stops (instead of a given period of time). The total displacement after the n^{th} chain is denoted by:

$$x_n = \sum_{j=1}^{n} \epsilon_j. \tag{3.2.2}$$

The probability density of the random variable x_n is denoted by $P_n(x)$.

For the sums of independent random variables, we can use the method of the characteristic functions in section 1.1.2. Here we use the Fourier transformation of $p(x)$ with an extra factor of 2π (so it matches the periodic domain of $[0,1)$):

$$f(t) = \mathcal{F}[p(x)] = \int_{-\infty}^{\infty} p(x) \mathrm{e}^{2\pi i x t} \mathrm{d}x = \mathbb{E}_\epsilon(\mathrm{e}^{2\pi i \epsilon t}). \tag{3.2.3}$$

The characteristic function of x_n is

$$f^n(t) = \mathcal{F}[P_n(x)] = \int_{-\infty}^{\infty} P_n(x) \mathrm{e}^{2\pi i x t} dx, \tag{3.2.4}$$

using eq. (1.1.7). To simplify eq. (3.2.4), we take $|f(t)| = |f(-t)| = A(t)$, $\arg[f(t)] = \theta(t)$. Since

$$|f(t)| \leq \int_{-\infty}^{\infty} |p(x)\mathrm{e}^{2\pi i x t}| dx = 1, \tag{3.2.5}$$

there is a restriction of the domain that $|A(t)| \leq 1$ ($A(t) = 1$ if and only if $p(x)$ is a delta function), $\theta(t) \in [0, 2\pi)$.
Without loss of generality, we assume that in the initial configuration, the particle is sited in $X_0 = 0$, and the corresponding probability density of the particle is

$$\pi^{(0)}(x) = \delta(x),$$

due to the translational symmetry of the system. As the domain is periodic on $[0, 1)$, the final position X_n after n stops, as a random variable, has the following relation with x_n:

$$\begin{aligned} X_n &= \mod(x_n, 1), \\ P_n^{\mathrm{mod}}(x) &= \sum_{m\in\mathbb{Z}} P_n(x+m). \end{aligned} \tag{3.2.6}$$

where $P_n^{\mathrm{mod}}(x)$ is the probability density of X_n. The normalization of probability is verified:

$$\begin{aligned} \int_{-\infty}^{\infty} P_n(x)dx &= \sum_{m\in\mathbb{Z}} \int_m^{m+1} P_n(x)dx = \sum_{m\in\mathbb{Z}} \int_0^1 P_n(x+m)dx \\ &= \int_0^1 \sum_{m\in\mathbb{Z}} P_n(x+m)dx = \int_0^1 P_n^{\mathrm{mod}}(x)dx = 1. \end{aligned} \tag{3.2.7}$$

After n steps, the total variance distance (TVD) between the current state and the steady state is

$$\mathrm{TVD}(n) = \int_0^1 \left| P_n^{\mathrm{mod}}(x) - 1 \right| \mathrm{d}x, \tag{3.2.8}$$

according to the definition of TVD in eq. (1.4.14)

The *Poisson Summation Formula* [83, Section 5.4] states that, for a real integrable function h:

$$\sum_{m=-\infty}^{+\infty} h(m) = \sum_{t=-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x)\mathrm{e}^{2\pi i t x} \mathrm{d}x. \tag{3.2.9}$$

By replacing $h(x)$ with $P_n(x+m)$, we have:

$$\sum_{k\in\mathbb{Z}} f^n(k)\mathrm{e}^{2\pi i k x} = \sum_{m\in\mathbb{Z}} P_n(x+m) = P_n^{\mathrm{mod}}(x), \tag{3.2.10}$$

which gives the Fourier expansion of $P_n^{\mathrm{mod}}(x)$. With the earlier notation, it equals:

$$\begin{aligned} P_n^{\mathrm{mod}}(x) &= \sum_{k\in\mathbb{Z}} f^n(k)\mathrm{e}^{2\pi i k x} = \sum_{k\in\mathbb{Z}} A^n(k)\mathrm{e}^{i(n\theta(k)+2\pi k x)} \\ &= 1 + \sum_{k\in\mathbb{N}^+} 2A^n(k)\cos\left(n\theta(k) + 2\pi k x\right) \\ &= 1 + \sum_{k\in\mathbb{N}^+} 2\mathrm{e}^{n\log A(k)}\cos\left(n\theta(k) + 2\pi k x\right), \end{aligned} \tag{3.2.11}$$

and

$$\mathrm{TVD}(n) = \int_0^1 \left| \sum_{k\in\mathbb{N}^+} 2\mathrm{e}^{n\log A(k)} \cos\left(n\theta(k) + 2\pi kx\right) \right| \mathrm{d}x, \tag{3.2.12}$$

according to eq. (3.2.8). In this equation, $A(k) < 1, \forall k \in \mathbb{N}^+$ is a necessary condition for exponential convergence. If $p(x)$ has a continuous distribution in a subset of the domain, the convergence is guaranteed; on the other hand, if $p(x)$ is a sum of a (countable) series of Dirac δ functions, the TVD will never decrease to zero.

If the phases are "random" for each k, the largest amplitude determines the fluctuation, which corresponds to the relaxation time and decays at the scale:

$$\tau = \max_{k\in\mathbb{N}^+} \left[-\frac{1}{\log A(k)} \right]. \tag{3.2.13}$$

Walk with Turning-back

We may now think of the process with turning-back after each chain, as discussed in section 2.2.1. In this case, we can split the total displacement as n chains in the positive direction and n' chains in the negative direction ($n' = n-1$ or n, depending on the total number of chains).

For the distribution of the backward displacement, its Fourier transformation is $f(-t)$. Still, we have:

$$\begin{aligned} f^n(t) f^{n'}(-t) &= \mathcal{F}\left[P_{n+n'}(x)\right], \\ \sum_{m\in\mathbb{Z}} P_{n+n'}(x+m) &= \sum_{k\in\mathbb{Z}} f^n(k) f^{n'}(-k) \mathrm{e}^{2\pi i k x} \\ = 1 + \sum_{k\in\mathbb{N}^+} 2A^{n+n'}(k) &\cos\left((n-n')\theta(k) + 2\pi kx\right), \end{aligned} \tag{3.2.14}$$

indicating that the TVD decays at the same speed as eq. (3.2.13).

Some Special Stopping Times

In the walk in a fixed direction, if the probability density function of ϵ is shifted by a constant:

$$p_\epsilon(x) = p_\epsilon^{\mathrm{old}}(x - x_0),$$

the convergence property will remain unchanged, due to the translational invariance of the TVD.

Based on this derivation, the distribution centered at zero would reduce the total displacement $|x_n|$ in eq. (3.2.2). In the following discussion, we will shift the center of the symmetric $p_\epsilon(x)$ to the origin for simplicity.

If $\epsilon \sim \mathrm{unif}[\frac{\lambda}{2}, -\frac{\lambda}{2})$, according to eqs (3.2.11) and (3.2.12), we have:

$$\begin{aligned} \sum_{k=-\infty}^{\infty} P_n^{\mathrm{unif}}(x+k) - 1 &= \sum_{k\in\mathbb{N}^+} 2\left[\frac{\sin(\pi k\lambda)}{\pi k\lambda}\right]^n \cos(2\pi kx), \\ \mathrm{TVD}_{\mathrm{unif}}(n;\lambda) &\approx \frac{2}{\pi}\left(\frac{\sin(\pi\lambda)}{\pi\lambda}\right)^n. \end{aligned} \tag{3.2.15}$$

It shows that with an increasing λ, the time scale would decay algebraically with oscillations (see Fig. 3.6 (a)). When $\lambda \in \mathbb{N}^+$, the target distribution of unif$[0,1)$ is reached within a single chain.

If $\epsilon \sim \mathcal{N}(0,\sigma^2)$, similarly we have

$$\begin{aligned}\sum_{k=-\infty}^{\infty} P_n^{\text{Gauss}}(x+k) - 1 &= \vartheta_3\left[\pi x, \mathrm{e}^{-2\pi^2 n\sigma^2}\right] - 1 \\ &= 2\sum_{k=1}^{\infty} \exp\left(-2k^2\pi^2 n\sigma^2\right)\cos\left(2k\pi x\right), \\ \text{TVD}_{\text{Gauss}}(n;\sigma) &\approx \frac{2}{\pi}\exp\left(-2\pi^2 n\sigma^2\right),\end{aligned} \tag{3.2.16}$$

which is already discussed in section 1.5.3. As shown in Fig. 3.6 (b), its first-order approximation describes the evolution of TVD very well.

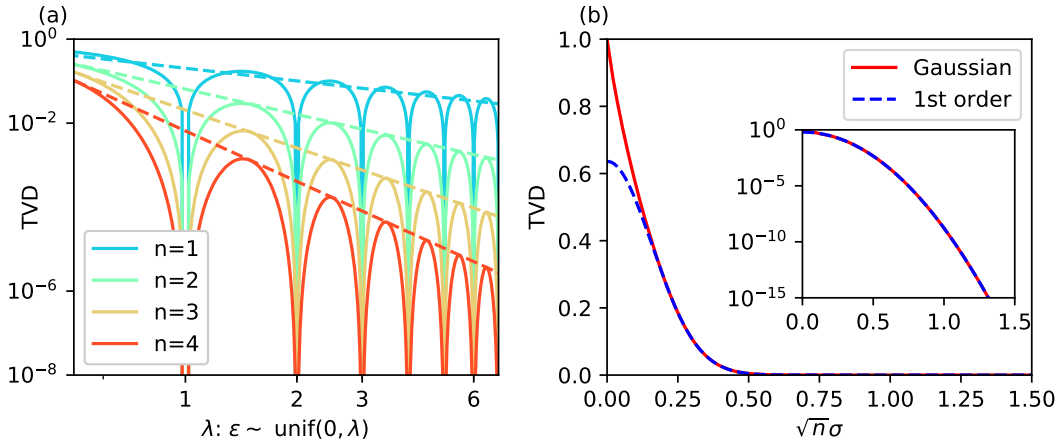


Figure 3.6: TVD for a single sphere on a ring with uniform and Gaussian distributions of displacements. (a): TVD after n displacements $\epsilon \sim$ unif$[0,\lambda)$. The TVD trivially vanishes for integer λ. Peaks decay as $\frac{2}{\pi}(\pi\lambda)^{-n}$ (for $n \to \infty$). (b): TVD for n Gaussian displacements with standard deviation σ, compared with its first-order approximation from the Jacobi ϑ function (see eq. (3.2.16)). The inset illustrates the good agreement of the approximation on a logarithmic scale.[11]

If $\epsilon \sim \text{Exp}(1/\lambda)$ ($p_\epsilon(x) = \lambda \mathrm{e}^{-x/\lambda}$), which corresponds to the Poisson process used in chapter 2,

$$A(k) = \frac{1}{\sqrt{1+(2\pi k\lambda)^2}}, \theta(k) = \tan^{-1}(2\pi k\lambda). \tag{3.2.17}$$

It shows a similar decay as eq. (3.2.15) in the uniform case with the growth of λ, but without oscillations.

Generally, for a continuous distribution of ϵ with a variance ~ 1, the mixing time is $\mathcal{O}(1)$. If the size of the system scales as $\tilde{L}$, ϵ should have a variance $\sim \tilde{L}^2$ for an efficient mixing.

[11] This figure is taken from the Fig. 5 in Publication II [17].

3.2.2 Multi-particle Case

According to the mapping between the positions of spheres and those of the particles, their dynamics are exactly the same when we consider $\{\delta_i\}$ (or $\{s_i\}$, see Fig. 3.1). In the event-chain dynamics, each collision will lead two particles to exchange their momentum (the lifting parameter of the active state).

As all the particles are volume-less and indistinguishable, we may assume that the positions and the lifting parameters of the particles are exchanged in one collision. Effectively, only one particle is sampled throughout one chain as illustrated in Fig. 3.7.

The distribution of the position of the "effective particle" is only dependent of that of the chain length ϵ, as we discussed in the single particle case. Hence, the mixing of a one-dimensional hard-sphere system is regarded as the independent mixing processes of different "effective particles" in the event-chain dynamics.

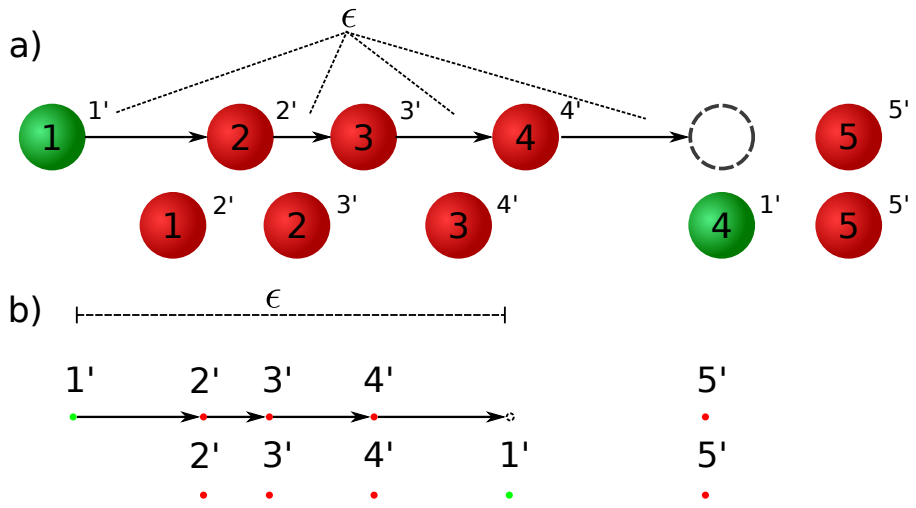


Figure 3.7: The process of one chain in the event-chain dynamics in (a) the hard-sphere model, and in (b) the model of volume-less particles. The dynamics of these two models map to each other. For the model of volume-less particles (b), if we look at the labels with prime (the "effective particle"), only one particle ($1'$) is effectively sampled in the process, while all the others stay at their former positions. It also means that in the hard-sphere model (a), only one disk ($1'$) is sampled in a chain.

"Perfect Sampling" of Multiple Particles

We begin with the simplest case, that the chain length ϵ follows the uniform distribution between $[0,1)$.[12] After one chain, one "effective" particle i (its position is denoted by as y_i) is perfectly sampled, having triggered at most $N-1$ events. If a random particle is picked after a chain is finished (which corresponds to the "shuffle" process

[12]For a more general distribution of ϵ, we can not obtain the exact result in the limit $N \to \infty$. Nevertheless, the discussion of $\epsilon \sim \text{unif}[0,1)$ provides a lower bound and an estimation of the mixing time for other distributions.

in section 2.3.1), then the problem matches perfectly to the coupon-collector problem in section 1.5.1.
In formal presentation, every particle is at the origin in the initial configuration. The initial probability density is a product of Dirac δ functions:

$$\pi^{(0)}(\{y_i\}) = \prod_{i=1}^{N} \delta(y_i), \quad \text{TVD}(0) = 1. \tag{3.2.18}$$

If there is at least 1 "effective" particle not moved, the TVD is still 1 due to the δ functions; once all the particles are moved, the TVD is reduced to 0. From the phase space $\Omega = [0,1)^N$, we take a subset

$$\mathcal{A} = \{\mathbf{y} \mid \exists\, i \text{ with } y_i = 0\}. \tag{3.2.19}$$

Using the equation eq. (1.4.14),

$$\|\pi^{(n)} - \pi\|_{\text{TV}} = \max_{\mathcal{A}' \subseteq \Omega} |\pi^{(n)}(\mathcal{A}') - 1| = \pi^{(n)}(\mathcal{A}), \tag{3.2.20}$$

which means all the particles are perfectly sampled, once probability density does not contain a Dirac δ function. Hence, there is a mapping between the strong stationary time (defined in section 1.5.2) of this system and the "first-complete-collection time" T_1 of coupon collection in section 1.5.1:

$$\begin{aligned} \|\pi^{(n)} - \pi\|_{\text{TV}} &= \mathbf{P}(T_1 > n) \\ &\sim 1 - \exp\left[-\exp\left(-\frac{n - N\log N}{N}\right)\right]. \end{aligned} \tag{3.2.21}$$

The mean collection time is

$$\mathbb{E}(T_1) \approx N(\log N + \gamma),$$

so the mixing time is $\mathcal{O}(N \log N)$ chains. As each chain triggers $(N/2)$ events on average and $(N-1)$ events at most, the mixing time counted in single events is $\mathcal{O}(N^2 \log N)$.

General Chain Length ϵ

For a more general distribution of the chain length ϵ, the "perfect sampling" can not be reached with finite chains. Eq. (3.2.20) is then rewritten as:

$$\|\pi^{(n)} - \pi\|_{\text{TV}} = \max_{\mathcal{A}' \subseteq \Omega} |\pi^{(n)}(\mathcal{A}') - 1| > \pi^{(n)}(\mathcal{A}) = \mathbf{P}(T_1 > n). \tag{3.2.22}$$

The lower bound of the mixing time is $\mathcal{O}(N \log N)$ chains.
According to the discussions in section 3.2.1, the TVD of a single ("effective") particle decreases exponentially with time (if $\text{var}(\epsilon) \sim 1$). We presume that the TVD is limited by the "effective particle" moved the fewest times. We assume that the number of times is denoted by m, then the problem is similar to an "m^{th}-complete-collection time" in section 1.5.1.

P. Erdös and A. Rényi provided an estimation of its tails [19],

$$\mathbf{P}(T_m > n) \sim \exp\left(-\Upsilon/(m-1)!\right), \tag{3.2.23}$$

with

$$\Upsilon = \exp\left[-\frac{n - N\log N - (m-1)N\log\log N}{N}\right]. \tag{3.2.24}$$

The mean m^{th}-complete-collection time is still $\mathcal{O}(N\log N)$ (if m is small compared to $\log N/\log\log N$):

$$\mathbb{E}(T_m) \sim N\log N + (m-1)N\log\log N.$$

The number of chains to pick each of the N particles at least m times only adds an $N\log\log N$ correction to the general $N\log N$ scale of chains. The mixing time is estimated $\mathcal{O}(N^2\log N)$ events as the "perfect sampling" case, based on the assumption of m^{th} complete collection.

This argument also applies to the lifted Metropolis with restarts. Due to the randomness introduced by the Metropolis move, we suggest that the distribution of the number of events in one chain has a standard deviation of αN, and the distribution of the step size is of the order $\frac{1}{\alpha N}$, where α is some constant larger than 1. The mixing-time evaluation in Fig. 3.11(c)(d) confirms our estimation that the mixing time is $\mathcal{O}(N\log N)$.

Heuristic Mixing-time Evaluation *via* the Mid-system Variance

For completeness, we compute the exact steady-state distribution of one of the N hard spheres (with ordering) on the ring. In view of the mapping of Fig. 3.1, we only need to calculate for volume-less particles in a ring of length 1.

We estimate the steady probability distribution of

$$u_i = \sum_{k=0}^{N/2} \delta_{i+k} = s_{i+N/2} - s_i, \tag{3.2.25}$$

which we used to examine the steady state in sections 3.1 and 3.2.3. As the system is periodic, u_i has the identical distribution for all the i (though they are not independent). In the following we will consider the probability distribution of u without specifying i.

Firstly we consider a group of random variables $\{y_i, i = 1,\ldots,N-1\}$. Each of them follows an independent uniform distribution unif[0,1). If we sort $\{y_i\}$ by its order, it has the same distribution as $\{s_i, i = 1,\ldots,N-1\}$. On the condition that $y_{\frac{N}{2}} = x$, the probability of having half of the rest of $\{y_i\}$ larger than x and the other half smaller than x is:

$$\begin{aligned}
&\mathbf{P}(y_{\frac{N}{2}} \text{ is in the middle} | y_{\frac{N}{2}} = x)\\
=&\binom{N-2}{N/2-1}\mathbf{P}(y_1 \le x)\cdots\mathbf{P}(y_{\frac{N}{2}-1} \le x)\mathbf{P}(y_{\frac{N}{2}+1} \ge x)\cdots\mathbf{P}(y_{(N-1)} \ge x)\\
=&\frac{(N-2)!}{[(N/2-1)!]^2}x^{(N/2-1)}(1-x)^{(N/2-1)}.
\end{aligned} \tag{3.2.26}$$

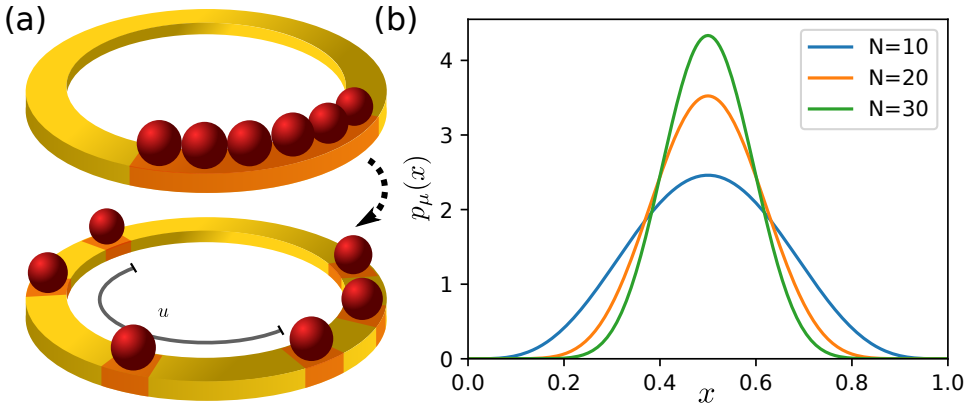


Figure 3.8: (a): An illustration of the process from a compact initial configuration to equilibrium. The "half-system distance" $u = x_{i+N/2} - x_i$ is defined as the free distance between the two spheres i and $i + N/2$. (b): The probability density function $p_u(x)$ of u. It is symmetric with respect to $x = 1/2$ ($\langle x \rangle = 1/2$), with $\mathrm{var}(u) = \frac{1}{4(N+1)}$. The width of the distribution narrows ($\sim \frac{1}{\sqrt{N}}$) as N increases.

Considering the permutation of y_i as $y_{\frac{N}{2}}$, the probability density function of u is calculated with a marginal distribution,

$$p_u(x) = \frac{(N-1)!}{[(N/2-1)!]^2}(x - x^2)^{(N/2-1)}, \tag{3.2.27}$$

and its mean and its variance are

$$\langle u \rangle = 1/2, \ \mathrm{var}(u) = \frac{1}{4(N+1)}. \tag{3.2.28}$$

The probability density function is shown in Fig. 3.8 (b).

In a compact initial configuration, all particles are restricted in a small part of the interval, $\langle (u_i - \frac{1}{2})^2 \rangle \sim 1$. In contrast, in the equilibrium condition/steady-state distribution, it is equal to $\frac{1}{4(N+1)}$. This provides us a good criterion in the evaluation of the mixing time by measuring $\langle (u_i - \frac{1}{2})^2 \rangle$. In the discussion of section 3.1, the mixing time of different derivations of Metropolis algorithm is evaluated in this way [18]. However, this method remains naïve: it provides only a lower bound and may not sufficiently guarantee convergence in general cases.

3.2.3 Swap Labeling in Algorithms

Based on the discussion of section 3.2.2, the logarithmic limit of the coupon collector can be overcome, if each "effective particle" is moved at most once before they are all moved. As the "effective particle" changes its label after a collision, we need to track the "effective particles" together with the positions.

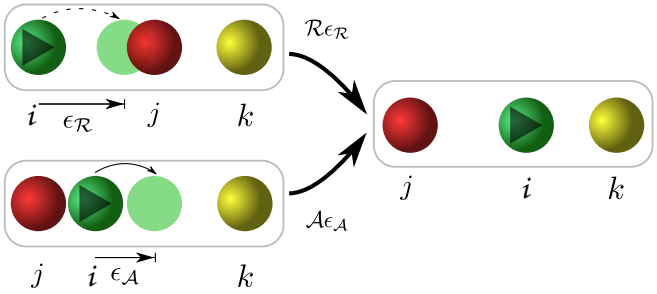


Figure 3.9: The probability flow into a configuration $(\{\delta_j\},\{P_j\}=\{j,i,k\},a=i)$ for a lifted forward Metropolis with swaps[14]. A rejected move, by a displacement $\epsilon_{\mathcal{R}}$ (upper case), entails a swap and contributes $\mathcal{R}_{\epsilon_{\mathcal{R}}}$. An accepted move, by a displacement $\epsilon_{\mathcal{A}}$ (lower case), contributes $\mathcal{A}_{\epsilon_{\mathcal{A}}}$. For any $\epsilon>0$, one of the flow is 0, and the other flows equals $\pi(\{\delta_j\})$.

Here we introduce a "swap" algorithm: each particle is marked with a "label" for the "effective particle", and any rejection of a proposed move will lead to a swap of the labels between the active particle and the one who rejects. In this scheme, each configuration is characterized by $(\{\delta_j\},\{P_j\})$, where $\{P_j\}$ stands for a permutation of labels. We assume that all these configurations have the same weight, and we may omit the notation of $\{P_j\}$ in the distribution for simplicity.

In the special case where $\epsilon>0$ as in Fig. 3.9 (it also applies to a generally distributed ϵ), we have

$$\mathcal{F}(\{\delta_j\}|a=i)=\int_0^\infty \mathrm{d}\epsilon p(\epsilon)\underbrace{[\mathcal{A}_\epsilon(\{\delta_j\}|a=i)+\mathcal{R}_\epsilon(\{\delta_j\}|a=i)]}_{=1\text{ (see Fig. 3.9)}}=1=\pi(\{\delta_j\}|a=i), \tag{3.2.29}$$

where $\mathcal{F}(\{\delta_j\}|a=i)$ is the probability flow in to the configuration $\{\delta_j\}$ on the condition that the active particle is i. Similar to the discussion in section 3.1.1, for any probability distribution of the labels of active particle $\{\mathbf{P}(a=i),i=1,2,\dots,N\}$, or for any sequence of the labels of the active particle $\{a(t),t=1,2,\dots\}$, we have

$$\begin{aligned}\mathcal{F}(\{\delta_j\})&=\sum_{a=i}^{N}\mathcal{F}(\{\delta_j\}|a=i)\mathbf{P}(a=i)\\&=\sum_{a=i}^{N}\pi(\{\delta_j\}|a=i)\mathbf{P}(a=i)=\pi(\{\delta_j\}),\end{aligned} \tag{3.2.30}$$

the global-balance condition is preserved.

In the special case that $\epsilon>0$ and the label is picked sequentially by the end of each chain:

$$\dots,\underbrace{i,i,\dots,i,i}_{\text{chain }n},\underbrace{(i+1),(i+1),\dots,(i+1),(i+1)}_{\text{chain }n+1},\underbrace{(i+2),(i+2),\dots,(i+2),(i+2)}_{\text{chain }n+2},\dots, \tag{3.2.31}$$

[14]This figure is taken from the Fig. 2 in Publication II [17].

we have a *sequential lifted forward Metropolis algorithm with swaps*, as illustrated in Fig. 3.10. In its continuous limit, the *event-chain algorithm with swaps* mixes the system in $\mathcal{O}(N)$ chains instead of $\mathcal{O}(N \log N)$. (In the special case that the chain length follows a uniform distribution on $[0,1)$, it mixes exactly in N chains.)

The sequential lifted forward Metropolis algorithm with swaps mixes faster than the (random) lifted forward Metropolis algorithm, as shown in Fig. 3.11. For a given discretization $2/\alpha$, the sequential lifted forward algorithm mimics the $\mathcal{O}(N^2)$ mixing of the sequential event-chain algorithm with swaps, that it decreases every N chains (N^2 steps). However, for large N it is unable to keep the TVD small enough after N chains (see Fig. 3.11 (a)), and it begins to cross over into $\mathcal{O}(N^2 \log N)$ mixing as the (random) lifted forward Metropolis algorithm. In order to reduce TVD, one needs more detailed discretization as shown in Fig. 3.11 (b). After all, this approach is not able to mix in the time of $\mathcal{O}(N^2)$ steps. It then belongs to the "lifted TASEP class" as the random version, with a mixing time of $\mathcal{O}(N^2 \log N)$ steps (they are summarized in Table 3.1).

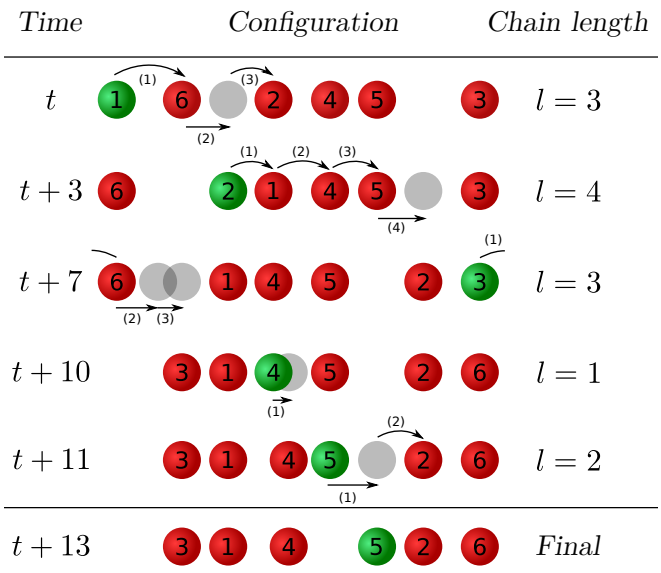


Figure 3.10: Sequential lifted forward Metropolis algorithm (with swaps). Configurations $\mathbf{x}_t, \ldots, \mathbf{x}_{t+13}$ sampled through five chains with active sphere $1, 2, \ldots, 5$ are shown. Chain lengths are $l_1 = 3, \ldots, l_5 = 2$. Each sphere displacement $\epsilon_t > 0$ is either accepted or rejected: if it is accepted, it remains active; if it is rejected, it induces a swap of the labels with the one who rejects. So the same label remains active throughout a chain.[16]

[16]This figure is taken from the Fig. 6 in Publication II [17].
[18]This figure is taken from the Fig. 7 in Publication II [17].

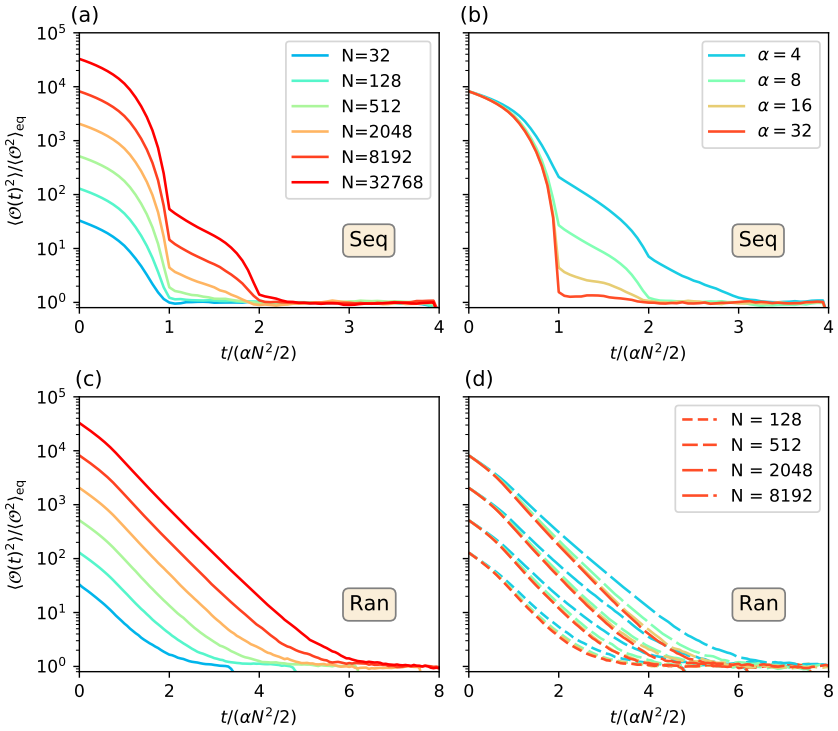


Figure 3.11: Crossover from the discrete lifted algorithm to the event-chain algorithm, *via* the variance of the mid-system distance u (see section 3.2.2), started from compact initial condition. Discrete step size with $\epsilon \sim \text{unif}[0, \frac{1}{N\alpha})$, and chain length $t \sim \text{unif}[\alpha, \alpha N)$ (a): Sequential lifted Metropolis algorithm with constant $\alpha = 10$ for different N: the cross-over from perfect sampling for small N at a time scale $\mathcal{O}(N^2)$ towards $\mathcal{O}(N^2 \log N)$ appears evident. (b): Sequential algorithm for $N = 8192$, with increasing α: $\mathcal{O}(N^2)$ mixing scale emerges for large α. (c): Random lifted Metropolis algorithm with $\alpha = 10$ for different N (legend as in (a)): $\mathcal{O}(N^2 \log N)$ mixing time scale. (d): Random lifted forward Metropolis algorithm: limited role of α (color code for α as in (b)).[18]

3.3 From One Dimension to Higher Dimensions

In section 2.3, we generalized the event-chain algorithm to higher dimensions, where it preserves irreducibility, aperiodicity and continues to satisfy the global-balance condition. Based on previous discussions, a randomized stopping rule can accelerate the mixing process in higher dimensions.

As presented before, there is no phase transition in the one-dimensional hard-sphere model. However, the two-dimensional case is much different: various phases appear in this model such as liquid, hexatic and solid phases, depending on the density [5].

3.3.1 Liquid, Solid and Hexatic Phases

There has been evidence of phase transition in the hard-disk model since 1962 [84]. There are at least 2 phases: a solid phase where each particle has the fixed neighbors, and a liquid phase where the neighbors of one particle can travel infinitely far after a long time. This transition happens in canonical ensembles of a fixed volume, which depends on the density instead of the temperature.

For a multiple particle system with an initial configuration of a lattice structure, we use the bases of lattice $\mathbf{R}_i$ to label the particles. The position of a particle can be divided to the lattice base $\mathbf{R}_i$ and a vector field of displacement $u(\mathbf{R}_i)$:

$$\mathbf{r}_i = \mathbf{R}_i + u(\mathbf{R}_i). \tag{3.3.1}$$

The behavior of the *positional correlation function* ($|\mathbf{k}| \sim 1/L$)

$$C_{\text{position}}(R) = \langle \exp\{i\mathbf{k}[\mathbf{r}(\mathbf{R}) - \mathbf{r}(0)]\}\rangle$$

determines the *position order* of the system, and similarly, for a given unit lattice vector $\mathbf{a}$, the *orientation order* is evaluated by the *orientational correlation function*:

$$C_{\text{orient}}(R) = \langle \exp\{i\mathbf{k}[\mathbf{r}(\mathbf{R}+\mathbf{a}) - \mathbf{r}(\mathbf{R})]\}\rangle\,.$$

For these correlation functions (denoted by $C(R)$ in general), in the limit that $\mathbf{R} \to \infty$ (ξ is the correlation length discussed in section 1.6.3):

- If $C(R) \sim \mathcal{O}(A + \mathrm{e}^{-R/\xi})$, where A is non-zero, it is *long-range* order.
- If $C(R) \sim \mathcal{O}(R^{-\nu})$, it is *quasi-long-range* order.
- If $C(R) \sim \mathcal{O}(\mathrm{e}^{-R/\xi})$, it is *short-range* order.

The long-range positional order is impossible in two dimensions, but the long-range orientational order is possible [85]. Later in 1972 [86], the two-dimensional phase transition was once again proposed, with more specified phases and more accurate descriptions, as shown in Table 3.2.

The quasi-long-range position order/long-range orientational order of the solid phase is broken by the pairs of dislocations as large as the system size, and the quasi-long-range orientational order is broken by the disclinations, as illustrated in Fig. 3.12.

Phase	Position order	Orientation order
Solid	quasi-long-range	long-range
Hexatic	short-range	quasi-long-range
Liquid	short-range	short-range

Table 3.2: The different phases in a two-dimensional mono-disperse model and their corresponding spatial correlation properties.

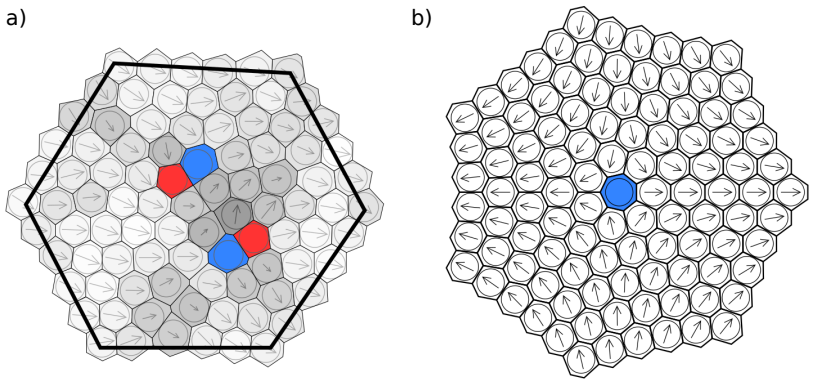


Figure 3.12: In the figures the red cell has 5 neighbors and the blue cell has 7 neighbors. The arrow inside each cell indicates its ψ^6 parameter. (a): a pair of dislocations that distort the local orientation. (b): a "negative" disclination, that the orientation winds by -2π around the blue cell. A "positive" one can be constructed similarly.[20]

An order parameter ψ^6 is introduced to characterize the phases. It is defined as

$$\psi_j^6 = \frac{1}{n_j} \sum_{\langle j,k \rangle} \mathrm{e}^{i6\phi_{j,k}}, \tag{3.3.2}$$

where n_j is the number of neighbors of the disk j, $\langle j,k \rangle$ indicates all the neighbors k with respect to the Voronoi cells, and $\phi_{j,k}$ is the angle of the vector from the center of j to k, with respect to the x-axis. This function has a modulus of 1 when the disk j has 6 neighbors equally spaced. The correlations of ψ^6 is density-dependent, especially near the critical point.

For a global observable Ψ^6 defined as the mean of ψ^6,

$$\Psi^6 = \frac{1}{N} \sum_{j=1}^{N} \psi_j^6, \tag{3.3.3}$$

it has different type of distribution in different phases. As shown in Fig. 3.14, the distribution is always symmetric. In the solid phase, the distribution forms a ring (or several detached clusters far from the centers), while in the liquid phase, it forms a circle concentrated in the center.

[20] This figure is taken from Fig. (1.11) and (1.12) in the thesis of E. Bernard [77].

According the KTHNY theory [86, 87, 88], there exists an infinite-order topological phase transition in the two-dimensional melting process, similar to the one in two-dimensional XY model that we introduce in chapter 4. Further research with event-chain algorithm identifies a first-order phase transition between the liquid phase and the hexatic phase in the hard-disk model [13], and liquid-hexatic coexistence in soft-disk melting [15].

In the following, we will apply the idea of "swapping" optimization of event-chain algorithm to two-dimensional hard disks near the transition point, and compare its result with that of the conventional method.

3.3.2 Swap Labeling in Two Dimensions

Based on the discussions in section 2.4, the event-chain algorithm is applied in general systems, with "shuffles" at the "stopping" time, as long as the "shuffle" process preserves the global balance (for example, sequentially picking the active labels). Hence, this label-swapping event-chain algorithm is also valid in higher dimensions.

We consider a lifted TASEP in a two-dimensional square lattice (see Fig. 3.13). It resembles the one-dimensional case in each row and each column. As each "effective" particle needs to be sampled in both dimensions, and the mixing time is at least twice as much as that of the one-dimensional case. In order not to miss any dimension, we put one label on each dimension, as discussed in section 3.2.3 to get over the coupon-collecting limitation.

Here we propose a general scheme for of the label-swapping event-chain algorithm: each disk is equipped with two labels $(i_1, i_2), i_1, i_2 = 1, \dots, N$, and each collision leads to a swap of the labels. A collision can trigger the swap of all the labels (then effectively only one label is needed, see Fig. 3.13 (a)), or only the swap of the label in the corresponding dimension, as in Fig. 3.13 (b). After one chain, the new chain will begin with the particle of the next "active label". The "active dimension" can be picked randomly or sequentially one after the other.

However, there is no more "perfect" choice of the stopping time as in the one-dimensional uniform case. Based on the efficiency analysis in [9, 77], we take the distribution of stopping time as unif$[-\sqrt{N}\lambda/2, \sqrt{N}\lambda/2)$, where λ is about three times the mean free space near the transition point.

In Fig. 3.14, we see that the distributions of Ψ^6 obtained *via* the conventional event-chain Monte Carlo, the event-chain Monte Carlo with single-labeling, and the one with double-labeling are the same, in agreement with previous studies [9, 77]. The autocorrelation of Ψ^6 indicates an acceleration by using single labeling and sequentially choosing the active dimensions (see Fig. 3.15). However, as the dynamics in high density is relatively slow, the increase in the mixing speed appears to be only a constant factor.

3.4 Conclusion

In this chapter we analyzed the dynamics of multiple variants of the Metropolis algorithm (including sequential Metropolis algorithm, forward Metropolis algorithm,

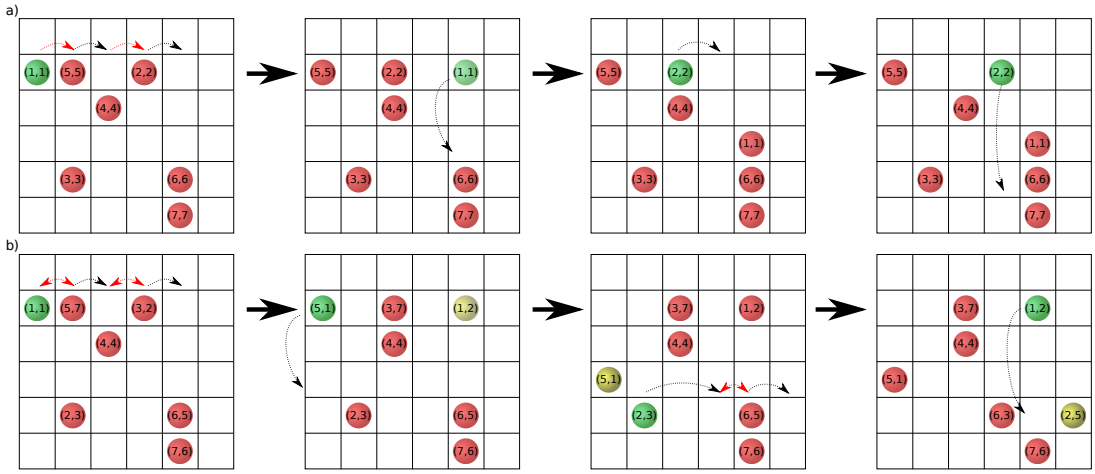


Figure 3.13: Two-dimensional label-swapping lifted TASEP. In each configuration, the green disk is the active one with a randomized chain length. The label-swapping is indicated with red arrows, and the active direction is picked sequentially one after the other. (a) The labels are swapped in both dimensions, then effectively we only need one label; (b) The labels are swapped only in the active dimension, the last moved disk is colored yellow.

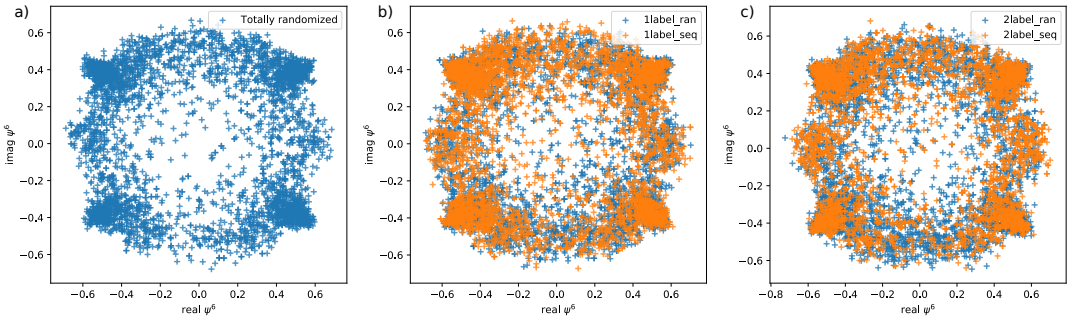


Figure 3.14: Scatter plot of Ψ^6 for 256 disks in a square under periodic boundary condition (the density is 0.700, near the liquid and the hexatic phases). (a) conventional event-chain algorithm, (b) single label method as in Fig. 3.13 (a), (c) two labels as in Fig. 3.13 (b). In (b) (c), the orange data points are provided by sequentially choosing the active dimension, and the blue data points are provided by random choosing the active dimension. We see that they all converge to the same distribution, and $\langle \Psi^6 \rangle = 0$ for a square box.

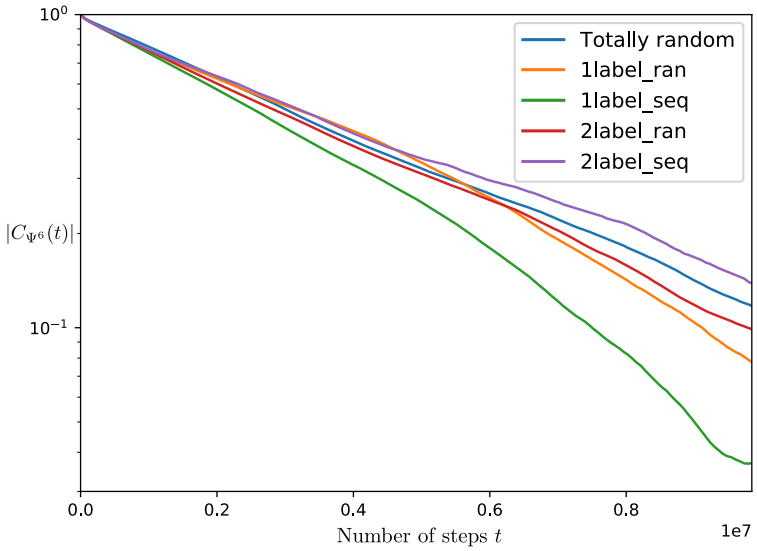


Figure 3.15: The absolute autocorrelation function of Ψ^6 for 256 disks in a square under periodic boundary condition (the density is 0.700, near the boundary of the liquid and hexatic phases). The green curve of the event-chain Monte Carlo with a single label and the active dimension sequentially chosen, appears faster than the others.

lifted forward Metropolis algorithm with and without restarts in Table 3.1) in the one-dimensional hard-disk model based on the global-balance condition.

We calculated the mixing times of a single particle in an interval, and generalized it to describe the mixing of a one-dimensional system of particles and hard spheres. In the case of "perfect sampling", the mixing time is directly related to the first-complete collection time in the coupon-collector problem. We also discussed the mixing time for a general distributed chain length, which provided a good estimation of the lifted forward Metropolis algorithm with restarts.

To overcome the logarithmic limit induced by the coupon-collector problem, we proposed a "swap" algorithm. It mixes in $\mathcal{O}(N^2)$ single steps in the event-chain algorithm, but when it is applied to the lifted sequential forward Metropolis algorithm, there is a cross-over of $\mathcal{O}(N^2 \log N)$ mixing time with the conventional methods.
We also generalized the "swapping" algorithm to two-dimensional hard-disk model. It is valid, and shows an acceleration slightly below the critical point.

Chapter 4

Event-chain Dynamics in Spin Models

As mentioned in chapter 2, the event-chain algorithm was extended from hard-core interactions to continuous potential, and sometimes an "infinite chain" can replace a sequence of finite event chains, to form an irreducible Markov-chain algorithm. However, the steady-state distributions and the dynamics become complex once again, and it is hard to evaluate the mixing time rigorously. Earlier work hinted at the possibility that the event-chain algorithm would lead to a significant speed-up of autocorrelations [11] and even had the potential for a decreased dynamical critical exponent [89].

In the present chapter we will start from the XY and Heisenberg spin models in two and three dimensions, focus on the different dynamics of patterns (spin wave and vortices) in event-chain Monte Carlo, analyze the Fréchet distribution of the maximum vortex–anti-vortex pair distance, and provide an estimation of the autocorrelation time. We believe that the dynamics of the spin waves here can be translated *mutatis mutandis* to the phonons in particle systems (as the fast dynamics is brought out in the simplified harmonic model).

We will understand the speed-up of the irreversible method to primarily act on the spin-wave degrees of freedom, whereas the relaxation of topological excitations (which are absent in one-dimensional models) remains unchanged. At low temperature, the topological excitations are strongly bounded. The speed-up of spin-wave excitations leads to improved dynamical exponents.

4.1 Event-chain Algorithm in Spin Systems

Spin systems and their second-order phase transitions have attracted people's attention since the discovery of Curie point in 1895 [90]. In the 1970s, with the development of Kosterlitz–Thouless theory [91], the infinite-order phase transition of the two-dimensional XY model started a new era of solid-state physics. Monte Carlo methods played an important role in exhibiting this phase transitions [92], and in characterizing the monopole excitations of the three-dimensional Heisenberg model [93].

4.1.1 Spin Models

The Hamiltonian of a spin model is given by:

$$H = -\sum_{1\le i<j\le N} J_{ij}\vec{S}_i\cdot\vec{S}_j, \tag{4.1.1}$$

where $\{J_{ij}\}$ are coupling constants, $\vec{S}_i$ represents the spin on the i^{th} site. Depending on the number of dimensions of each spin, the Hamiltonian can describe the Ising model $(O(1))$

$$\begin{aligned} H_{\text{Ising}} &= -\sum_{1\le i<j\le N} J_{ij}s_is_j, \\ s_k &\in \{-1,1\},\ k=1,\dots,N, \end{aligned} \tag{4.1.2}$$

the XY model $(O(2))$

$$\begin{aligned} H_{\text{XY}} &= -\sum_{1\le i<j\le N} J_{ij}\vec{S}_i\cdot\vec{S}_j, \\ \vec{S}_i = (S_{ix},S_{iy}) &= (\cos\phi_i,\sin\phi_i),\ 0\le\phi_i\le 2\pi, \end{aligned} \tag{4.1.3}$$

and the Heisenberg model $(O(3))$

$$\begin{aligned} H_{\text{Heisenberg}} &= -\sum_{1\le i<j\le N} J_{ij}\vec{S}_i\cdot\vec{S}_j, \\ \vec{S}_i = (S_{ix},S_{iy},S_{iz}) &= (\sin\theta_i\cos\phi_i,\sin\theta_i\sin\phi_i,\cos\theta_i),\ 0\le\theta_i\le\pi,\ 0\le\phi_i\le 2\pi, \end{aligned} \tag{4.1.4}$$

respectively.

The setting of J_{ij} determines the system's behavior: if $J_{ij}\ge 0$, the system is ferromagnetic, that all the spins tend to the same orientation at low temperature, resulting a non-zero magnetization[1]; if $J_{ij}\le 0$ the system is anti-ferromagnetic at low temperature (for non-frustrated cases, such as square and hexagonal lattices), with nearly zero magnetization; if J_{ij} is given randomly, the system may become a spin glass [95].

Magnetic systems usually have short-range interactions ($J_{ij}=o(r_{ij}^{-2})$). For such systems there is no long-range order in the one-dimensional Ising model [34], or in other continuous model with dimension lower than two [94, 85], which indicates no ferromagnetic phase transition. However, the two-dimensional case is a special one with the existence of a quasi-long-range order. It is broken into short-range order at high temperature, with the emergence of topological defects, resulting in an infinite-order phase transition [96, 91].

In the following we study the ferromagnetic model with the nearest-neighbor interaction ($J_{ij}=1$ on the neighboring sites i,j and 0 for the others). For example, in the XY model, the Hamiltonian can also be written as

$$H_{XY} = -\sum_{\langle i,j\rangle}\cos(\phi_i-\phi_j), \tag{4.1.5}$$

where $\langle i,j\rangle$ indicates a pair of neighbors.

[1] To be accurate, a stable ferromagnetic phase only exists in the dimension higher than two for the Ising model, or higher than three for the other continuous models, if the interaction decays fast enough, as a statement of Mermin-Wagner theorem [94].

4.1.2 Event-chain Algorithm in Spin Models

According to section 2.3.2, the event-chain algorithm (with the factorized Metropolis filter) in the XY model follows [11]:

- The simulation starts from an initial configuration, with an active site i;
- The thermal fluctuation of the interaction between i and its neighbor site j is given by
$$\xi_{ij} = -\frac{1}{\beta}\log \mathrm{ran}(0,1).$$
A proposed displacement is given by:
$$\Delta d_j = f^{-1}(E_{ij}+\xi_{ij}) - f^{-1}(E_{ij}), \tag{4.1.6}$$
where
$$\begin{aligned} f(d) &= \int_0^d \max[\sin(x),0]\mathrm{d}x \\ &= \begin{cases} 2\lceil \frac{d}{2\pi}\rceil & \text{if } d \geq 2\pi\lfloor\frac{d}{2\pi}\rfloor+\pi, \\ 2\lfloor\frac{d}{2\pi}\rfloor + 1 - \cos(d-2\pi\lfloor\frac{d}{2\pi}\rfloor) & \text{if } d < 2\pi\lfloor\frac{d}{2\pi}\rfloor+\pi, \end{cases} \end{aligned} \tag{4.1.7}$$
is a monotonic function of the energy and $f^{-1}(E)$ is its inverse function.
- $\Delta d_{j_{\min}} = \min\{\Delta d_j\}$ triggers the next event. We update ϕ_i with $\phi_i + \Delta d_{j_{\min}}$, and the site $j_{\min}$ is the next active site. Then the process goes back to step 2 and iterates, creating a chain as shown in Fig. 4.1.

In a general lifting algorithm, the direction of rotation and the active site need to be resampled after a period of time to ensure irreducibility (the number of events $\sim N$, as discussed in sections 3.1.4 and 3.2.2). The XY model is a special case that the event-chain Monte Carlo algorithm works well without restarting. We will discuss more about it in chapter 5.

This algorithm can be generalized to arbitrary coupling $\{J_{ij}\}$, including the cases of anti-ferromagnetic spin models and spin glasses. For the Heisenberg model, its Hamiltonian can be written with $\{\theta_i, \phi_i\}$:

$$\begin{aligned} H &= -J\sum_{\langle i,j\rangle} \cos(\theta_i)\cos(\theta_j) + \sin(\theta_i)\sin(\theta_j)\left[\cos(\phi_i)\cos(\phi_j)+\sin(\phi_i)\sin(\phi_j)\right] \\ &= -J\sum_{\langle i,j\rangle} \cos(\theta_i)\cos(\theta_j) + \sin(\theta_i)\sin(\theta_j)\cos(\phi_i-\phi_j). \end{aligned} \tag{4.1.8}$$

It can be regarded as an XY model with the coupling constants determined by $\{\theta_i\}$. In the event-chain Monte Carlo algorithm, eq. (4.1.8) can be treated in the same way as in eq. (4.1.5), by choosing θ's as the fixed parameters and ϕ's as the variables. The irreducibility of the algorithm is ensured by alternating the axes in cycles

$$(x,y,z) \to (y,z,x) \to (z,x,y) \to (x,y,z),$$

and the event-chain algorithm for Heisenberg model is constructed in the following [89]:

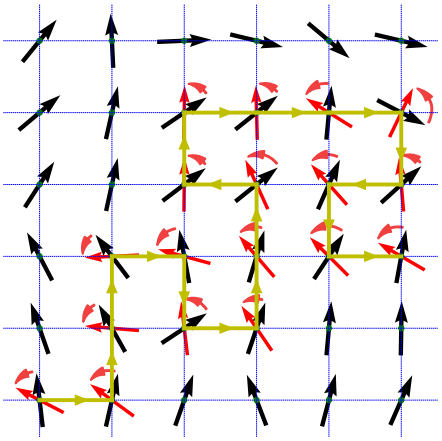


Figure 4.1: An illustration of the event-chain in the two-dimensional XY model. The earliest rejection determines the angle of rotation, and triggers the lifting of the active site. Rotation in only one direction without reshuffling the active site also guarantees irreducibility in this model.

- The coordinates (d1,d2,d3) is chosen from $\{(x,y,z),(y,z,x),(z,x,y)\}$. For a given set of coordinates (d1,d2,d3), $\{\theta_i,\phi_i\}$ are calculated as following:

$$\theta_i = \arccos s_{d3};\ \phi_i = \text{sign}(s_{d2})\arccos s_{d1}. \tag{4.1.9}$$

- By taking the coupling constants $J_{ij} = \sin(\theta_i)\sin(\theta_j)$ and the angles of spins as $\{\phi_i\}$, the event-chain algorithm for XY spin glass is applied on this model;
- Once a chain is finished, the process goes back to the first step, the coordinates (d1,d2,d3) is switched to another, and the event-chain algorithm continues.

4.1.3 The Dynamics of Susceptibility

In this complex system, an explicit calculation of spectral gap or TVD is inaccessible. To evaluate the speed of the algorithm, the autocorrelation time of a conventional order parameter is measured. For example the susceptibility χ in ferromagnetism is

$$\chi = \frac{||\vec{M}||^2}{N}, \tag{4.1.10}$$

with $\vec{M}$ being the magnetization of the system:

$$\vec{M} = \sum_{i=0}^{N} \vec{S}_i. \tag{4.1.11}$$

As a rotational invariant observable, the susceptibility χ usually has a longer autocorrelation time than the magnetization. In event-chain algorithm, the autocorrelation function of χ at $T \leq T_{\text{KT}}$ has two time scales [11], a fast decay followed by a much slower one, as shown in Fig. 4.2 (a). (The same behavior of the autocorrelation time is

found in Villain's model as well.) Also, the annealing process of a random state at low temperature varies much from sample to sample (see Fig. 4.2 (b)). These phenomena indicate a complicated dynamics.

In the following sections, we will start from the low-temperature approximation of the XY model, as a simple harmonic model, to explore its dynamics, and relate the second scaling to the vortices in the model.

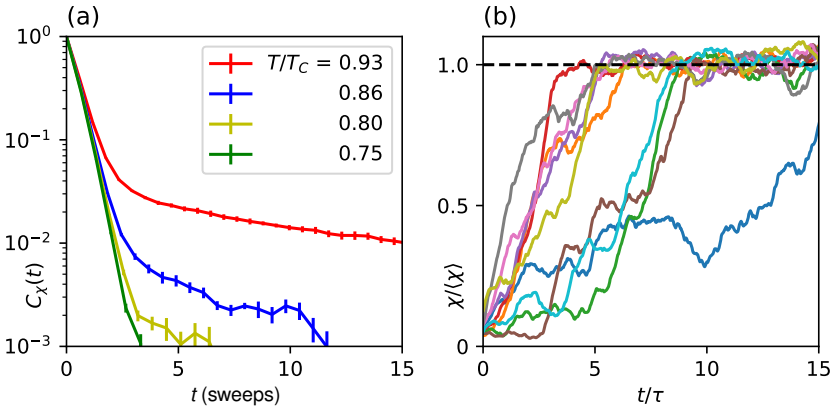


Figure 4.2: Time evolution of the spin susceptibility in the XY model on a 64×64 square lattice (time t measured in sweeps). (a) Susceptibility autocorrelations below T_{KT}. A 2-level scaling is clear in the autocorrelation function. (b) Individual susceptibility evolutions at $T/T_{\mathrm{KT}} = 0.93$ starting from random initial configurations (autocorrelation time τ) lead to large sample-to-sample fluctuations.[3]

4.2 The Dynamics of Spin Waves

4.2.1 Quadratic Approximation

The quadratic approximation is used in the limit of low temperature, where the higher-order terms in the polynomial expansion are neglected. For example, with the cos term in the XY model approximated by

$$\cos(\Delta\phi) \approx 1 - \frac{1}{2}\Delta\phi^2,$$

the quadratic Hamiltonian is:

$$\begin{aligned} H(\{\phi_i\}) &= -\sum_{\langle i,j\rangle} \cos(\phi_i - \phi_j) \\ &\sim \sum_{\langle i,j\rangle} \left(\frac{1}{2}(\phi_i - \phi_j)^2 - 1\right). \end{aligned} \tag{4.2.1}$$

[3]This is a remake of the Fig. 1 in the Publication I [12].

The partition function of this system is a high-dimensional Gaussian integral, and once the quadratic form is diagonalized, all the details in this model are revealed. For the Heisenberg model with the quadratic approximation, $\{\theta_i\}$ and $\{\phi_i\}$ are totally decoupled, so that they behave as two replicas of the XY model. In the following we will concentrate on the Hamiltonian in eq. (4.2.1).

The harmonic model, even on a general lattice (such as triangular or hexagonal ones) with long-range interaction, is exactly solvable using Fourier transformation.[4] Without loss of universality, we use the square lattice and the nearest neighbor interactions. From eq. (4.2.1), by omitting the constants, we have

$$\begin{aligned} H(\{\phi_i\}) &= \frac{1}{2} \sum_{\vec{r}\in \mathrm{Lat}, 1\le i\le d} [\phi(\vec{r}) - \phi(\vec{r}+\vec{e}_i)]^2 \\ &= \frac{1}{2} \sum_{\vec{k}\in \mathrm{FBZ}, 1\le i\le d} [2 - 2\cdot\cos(\vec{k}\cdot\vec{e}_i)]\hat{\phi}(\vec{k})\hat{\phi}(-\vec{k}) \\ &= \sum_{\vec{k}\in \frac{1}{2}\mathrm{FBZ}, 1\le i\le d} [1-\cos(\vec{k}\cdot\vec{e}_i)][\psi_1^2(\vec{k}) + \psi_2^2(\vec{k})] \\ &= \sum_{\vec{k}\in \frac{1}{2}\mathrm{FBZ}} [d - \sum_{1\le i\le d} \cos(\vec{k}\cdot\vec{e}_i)][\psi_1^2(\vec{k}) + \psi_2^2(\vec{k})] \\ &= \sum_{\vec{k}\in \frac{1}{2}\mathrm{FBZ}} \varepsilon(\vec{k})[\psi_1^2(\vec{k}) + \psi_2^2(\vec{k})] \end{aligned} \tag{4.2.2}$$

where "Lat" indicates all sites on the lattice, "FBZ" is the first Brilluoin Zone of the reciprocal lattice, and "$\frac{1}{2}$FBZ" is one half of the first Brilluoin Zone. The Fourier transformation is:

$$\hat{\phi}(\vec{k}) = \frac{1}{\sqrt{N}} \sum_{\vec{r}\in\mathrm{Lat}} \phi(\vec{r})\mathrm{e}^{-i\vec{k}\vec{r}},\ \phi(\vec{r}) = \frac{1}{\sqrt{N}} \sum_{\vec{k}\in\mathrm{FBZ}} \hat{\phi}(\vec{k})\mathrm{e}^{i\vec{k}\vec{r}}, \tag{4.2.3}$$

with the change of variables:

$$\begin{aligned} \psi_1(\vec{k}) &= \frac{1}{\sqrt{2}}(\hat{\phi}(\vec{k}) + \hat{\phi}(-\vec{k})) = \sqrt{\frac{2}{N}} \sum_{\vec{r}\in\mathrm{Lat}} \phi(\vec{r})\cos\vec{k}\vec{r}, \\ \psi_2(\vec{k}) &= \frac{1}{\sqrt{2}i}(\hat{\phi}(\vec{k}) - \hat{\phi}(-\vec{k})) = \sqrt{\frac{2}{N}} \sum_{\vec{r}\in\mathrm{Lat}} \phi(\vec{r})\sin\vec{k}\vec{r}, \\ \varepsilon(\vec{k}) &= d - \sum_{1\le i\le d} \cos(\vec{k}_i\vec{e}_i). \end{aligned} \tag{4.2.4}$$

The partition function is then:

$$Z = \int \prod_{\vec{k}\in\frac{1}{2}\mathrm{FBZ}} \mathrm{d}\psi_1(\vec{k})\mathrm{d}\psi_2(\vec{k})\mathrm{e}^{-\beta\varepsilon(\vec{k})[\psi_1(\vec{k})^2+\psi_2(\vec{k})^2]}. \tag{4.2.5}$$

[4] The following calculations are done in [97]. However, we write the terms more explicitly, which facilitates a numerically exact approach.

As we see, $\psi_1(\vec{k}), \psi_2(\vec{k})$ are independent variables in Z, and each of them follows a Gaussian distribution. However, there is a restriction on $\psi_1(\vec{k}), \psi_2(\vec{k})$ due to their symmetry in the first Brilluoin Zone. We rewrite eq. (4.2.2) as:

$$H(\{\phi_i\}) = \sum_{\vec{k}\in\frac{1}{2}\text{FBZ}} \varepsilon(\vec{k})[\psi_1^2(\vec{k}) + \psi_2^2(\vec{k})] \tag{4.2.6}$$

$$= \sum_{\vec{k}\in\text{reduced FBZ}} \varepsilon(\vec{k})[n_1(\vec{k})\psi_1^2(\vec{k}) + n_2(\vec{k})\psi_2^2(\vec{k})], \tag{4.2.7}$$

where "reduced FBZ" includes only one of $\vec{k}$ and $-\vec{k}$ for all $\vec{k} \in$ FBZ, unless $\vec{k} = \mod(-\vec{k}, 2\pi)$. In other words,

- if $\vec{k} \neq \mod(-\vec{k}, 2\pi)$, then $n_1(\vec{k}) = n_2(\vec{k}) = 1$;
- if $\vec{k} = \mod(-\vec{k}, 2\pi)$, then $n_1(\vec{k}) = 1/2, n_2(\vec{k}) = 0$.

The second case only includes the vertices of the first Brilluoin Zone, which is negligible compared with the first one. In general, we have

$$\left\langle \psi_i(\vec{k}_1)\psi_j(\vec{k}_2) \right\rangle = \delta_{i,j}\delta_{\vec{k}_1,\vec{k}_2} \frac{1}{2n_i(\vec{k}_1)\beta\varepsilon(\vec{k}_1)}. \tag{4.2.8}$$

As presented in section 1.6.3, the spatial correlation is closely related to the phase transition. It is given by:

$$g(\vec{r}) = \langle \vec{m}_0 \vec{m}_{\vec{r}} \rangle = \langle \cos(\phi(\vec{r}) - \phi(0)) \rangle = \exp\left(-\frac{1}{2} \left\langle [\phi(\vec{r}) - \phi(0)]^2 \right\rangle \right). \tag{4.2.9}$$

The right-hand side of eq. (4.2.9) is obtained as $[\phi(\vec{r}) - \phi(0)]$ follows Gaussian distribution:

$$\begin{aligned} \phi(\vec{r}) - \phi(0) &= \text{Re}\left(\frac{1}{\sqrt{N}} \sum_{\vec{k}\in FBZ} \hat{\phi}(\vec{k})(\exp\left(i\vec{k}\cdot\vec{r}\right) - 1)\right) \\ &= \sqrt{\frac{2}{N}} \sum_{\vec{k}\in\text{Reduced FBZ}} n_1(\vec{k})(\cos\vec{k}\cdot\vec{r} - 1)\psi_1(\vec{k}) + n_2(\vec{k})\sin\vec{k}\cdot\vec{r}\psi_2(\vec{k}). \end{aligned} \tag{4.2.10}$$

Its variance is calculated using eq. (4.2.8):

$$\langle [\phi(\vec{r}) - \phi(0)]^2 \rangle = \sum_{\vec{k}\in\text{Reduced FBZ}} \frac{n_1(\vec{k})(\cos\vec{k}\cdot\vec{r} - 1)^2 + n_2(\vec{k})(\sin\vec{k}\cdot\vec{r})^2}{N\beta\varepsilon(\vec{k})}. \tag{4.2.11}$$

In the continuous limit of the reciprocal lattice (infinite large lattice), the finite vertices are neglected, and the summation is replaced by an integral:

$$\begin{aligned} \langle [\phi(\vec{r}) - \phi(0)]^2 \rangle &= \int_{\vec{k}\in\text{Reduced FBZ}} \frac{n_1(\vec{k})(\cos\vec{k}\cdot\vec{r} - 1)^2 + n_2(\vec{k})(\sin\vec{k}\cdot\vec{r})^2}{N\beta\varepsilon(\vec{k})} \\ &= \int_{\vec{k}\in\text{FBZ}} \frac{2\sin^2(\vec{k}\cdot\vec{r}/2)}{N\beta(d - \sum_{1\le i\le d}\cos(\vec{k}_i\vec{e}_i))} \mathrm{d}^d\vec{k}. \end{aligned} \tag{4.2.12}$$

In the limit of large $\vec{r}$ and $\vec{k_{\min}}\vec{r} \sim 1$, the divergence of the integration appears when the dimension is lower than 3, as a basic conclusion of the Mermin–Wagner theorem [94]:

$$\langle [\phi(\vec{r}) - \phi(0)]^2 \rangle \propto \begin{cases} r & \text{if } d = 1, \\ \log r & \text{if } d = 2, \\ 1 & \text{if } d \geq 3. \end{cases} \tag{4.2.13}$$

Combining the result with eq. (4.2.9), we then find the spatial correlation has a short-range order in the one-dimensional case, a quasi-long-range order in the two-dimensional case, and a long-range order in the cases of dimension higher than three. After all, there does not exist any phase transition by reducing temperature.

This result was obtained by F. Wegner in 1967, but at the time it was not clear whether the behavior found in the harmonic model would translate to the XY model. The logarithmic correlations of XY model at low temperature were proven by J. Fröhlich and T. Spencer in 1981 [98]. The logarithm term in eq. (4.2.13) is of the same order of magnitude as the famous single-vortex contribution to the free energy, which we will then introduce in section 4.3.

4.2.2 Event-chain Dynamics in Spin-wave Mode

The spatial correlation provides a good approach in evaluating the dynamics. In the following, we consider the spatial correlation of the most distant sites as the slowest random variable in a local dynamics. We suppose there are two distant sites located at $\vec{0}$ and $\vec{r}$, at the time t_0, $\phi(\vec{0}, t_0) < \phi(\vec{r}, t_0)$. If a new configuration needs to fully decorrelate with the former one at the time t_0, the event $\phi(\vec{0}, t) > \phi(\vec{r}, t)$ should be achieved at least once.

In the event-chain algorithm, the step size is mainly determined by the temperature, if most of the sites are equilibrated (see section 5.1.3 for further details). In the case that $\phi(\vec{r}, t_0) - \phi(\vec{0}, t_0)$ is much larger the average step sizes, it will take a few moves on the site $\vec{0}$ until its $\phi(0) \sim \phi(\vec{r})$.

We think of an "out of equilibrium" state, with $\phi(\vec{0}, t_0) = \min_{\vec{R}}\{\phi(\vec{R}, t_0)\}$ and $\phi(\vec{r}, t_0) = \max_{\vec{R}}\{\phi(\vec{R}, t_0)\}$. Before $\phi(\vec{0})$ gets close enough to $\phi(\vec{r}, t_0)$, the sites with smaller ϕ tends to move more times than the ones with larger ϕ, as shown in Fig. 4.3. We call this dynamics "water-level" as it tends to fill the gap before ascending to the top. In the end, at a given temperature, the number of sweeps needed is

$$t \propto [\phi(\vec{r}, t_0) - \phi(\vec{0}, t_0)]. \tag{4.2.14}$$

In the harmonic model, the domain of ϕ is $(-\infty, \infty)$. With an event-chain dynamics of a fixed direction, each point in this domain can be reached at most once, while the irreducibility is guaranteed by the translational symmetry of the system. The mixing time can be infinite according to eq. (4.2.14), as the initial distortion can go to infinity. The autocorrelation time has the following behavior:

$$t_{\text{cor}}(L) \propto \sigma^{\text{eq}}(L) \propto \begin{cases} \sqrt{L} & \text{if } d = 1, \\ \sqrt{\log L} & \text{if } d = 2, \\ 1 & \text{if } d \geq 3, \end{cases} \tag{4.2.15}$$

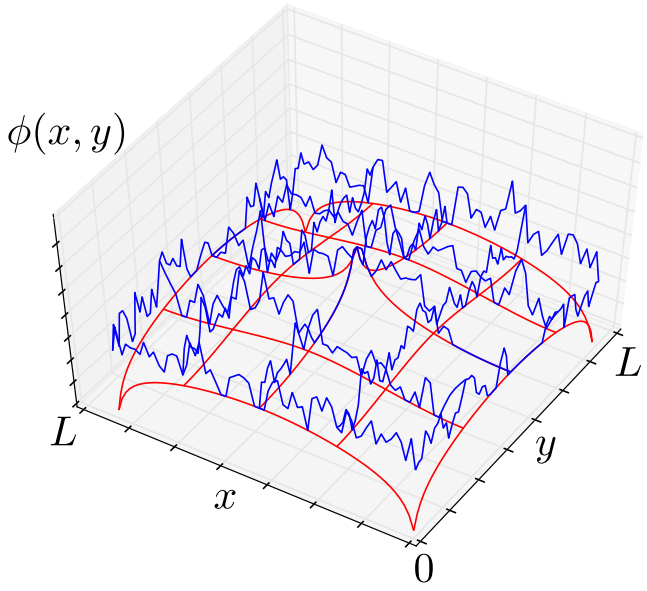


Figure 4.3: The evolution of the configuration in the event-chain algorithm in two-dimensional harmonic model. The distortion in the initial configuration (red) is much larger than the amplitude of thermal fluctuations. It can be seen that the event-chain dynamics tends to move the lower ones (on the corners) before the global fluctuation gets small enough.

where

$$\sigma^{\mathrm{eq}}(L) = \max_{\vec{r}\in\mathrm{Lat}} \sqrt{\left\langle [\phi(\vec{r}) - \phi(0)]^2 \right\rangle},$$

based on eqs (4.2.13) and (4.2.14). This relation is verified by the autocorrelation time of the "structure factor" of the lowest Fourier mode (see Fig. 4.4).

We use the square of the amplitude of the $\vec{k}$ mode as the observable:

$$\begin{aligned} |\hat{\phi}(\vec{k})|^2 = \hat{\phi}(\vec{k})\hat{\phi}(-\vec{k}) = \frac{1}{N}\left(\sum_{\vec{r}\in\mathrm{Lat}} \phi(\vec{r})\mathrm{e}^{-i\vec{k}\vec{r}}\right)\left(\sum_{\vec{r}\in\mathrm{Lat}} \phi(\vec{r})\mathrm{e}^{+i\vec{k}\vec{r}}\right) \\ = \frac{1}{N}\left(\sum_{\vec{r}\in\mathrm{Lat}} \phi(\vec{r})\cos\vec{k}\vec{r}\right)^2 + \frac{1}{N}\left(\sum_{\vec{r}\in\mathrm{Lat}} \phi(\vec{r})\sin\vec{k}\vec{r}\right)^2 = \frac{1}{2}[\psi_1^2(\vec{k}) + \psi_2^2(\vec{k})]. \end{aligned} \tag{4.2.16}$$

It shares a similar form with the structure factor in particle systems, in the limit of small $\vec{q}$:

$$S(\vec{q}) = \frac{1}{N}\left|\sum_a \mathrm{e}^{i\vec{q}\cdot\vec{R}_a}\right|^2 = \frac{1}{N}\left(\sum_a \cos\vec{q}\cdot\vec{R}_a\right)^2 + \frac{1}{N}\left(\sum_a \sin\vec{q}\cdot\vec{R}_a\right)^2. \tag{4.2.17}$$

As $\psi_i(\vec{k})$ follows the Gaussian distribution, the square of the amplitude follows χ_2^2

[6] This is a remake of the Fig. 7 in the Publication I [12].

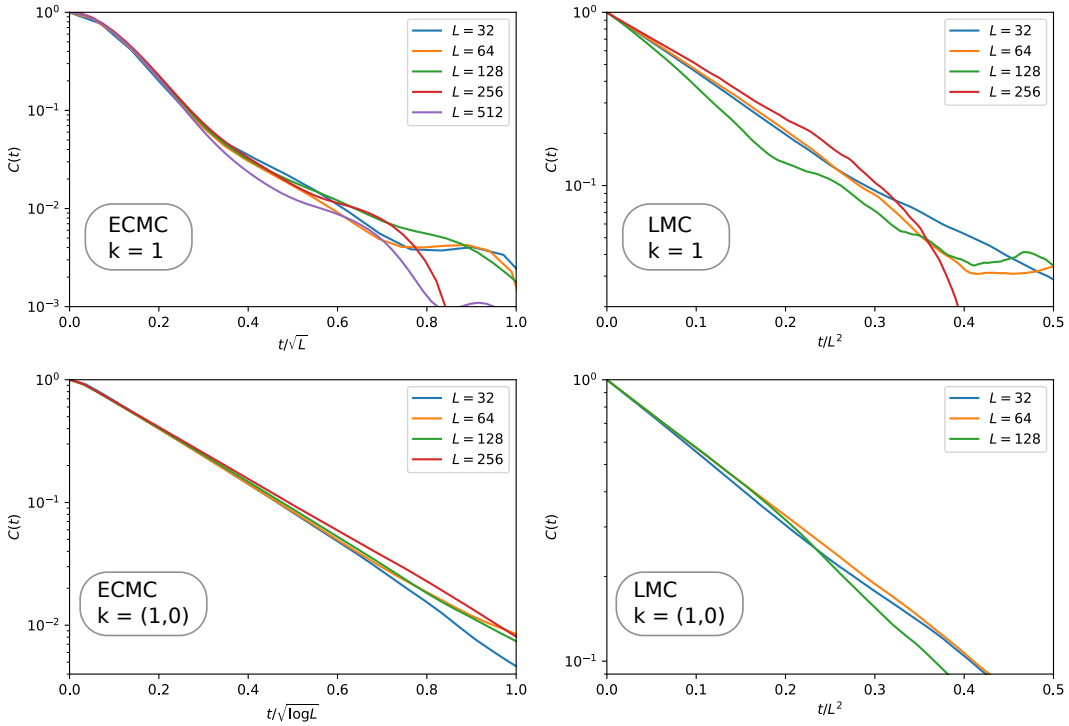


Figure 4.4: Equilibrium autocorrelation functions $C(t)$ of the squared amplitude $|\hat{\phi}(\vec{k})|^2$ of the lowest Fourier mode in the harmonic model for the event-chain algorithm (ECMC) and for local Monte Carlo (LMC). *Upper*: $C(t)$ for the Fourier mode $k = 1$ in 1D. *Lower*: $C(t)$ for the Fourier mode $k = (1,0)$ in 2D. Data are in agreement with the scaling of eq. (4.2.15).[6]

distribution (exponential distribution):

$$P_{|\hat{\phi}(\vec{k})|^2}(x) = \frac{1}{\sigma^2(\mathbf{k})} e^{-x/\sigma^2(\mathbf{k})}, \tag{4.2.18}$$

with its mean and variance

$$\begin{aligned} \left\langle |\hat{\phi}(\vec{k})|^2 \right\rangle &= \sigma^2(\mathbf{k}) = \frac{1}{2\beta\varepsilon(\mathbf{k})}, \\ \mathrm{var}(|\hat{\phi}(\vec{k})|^2) &= \sigma^4(\mathbf{k}) = \frac{1}{4\beta^2\varepsilon^2(\mathbf{k})}, \end{aligned} \tag{4.2.19}$$

where $\sigma(\mathbf{k}) = \frac{1}{\sqrt{2\beta\varepsilon}}$ is the variance of $\psi_1(\mathbf{k})$ and $\psi_2(\mathbf{k})$ (we take $n_1 = n_2 = 1$).
Although we can not calculate the time-dependence of the event-chain Monte Carlo algorithm for the harmonic model, we believe that the evolution of $|\hat{\phi}(\vec{k})|^2$ provides the slowest time scale. In Fig. 4.4, we see that the equilibrium auto-correlation time of the event-chain algorithm grows proportionally to $\sigma^{\mathrm{eq}}(L)$. In sharp contrast, the local Metropolis method has a diffusive $\mathcal{O}(L^2)$ autocorrelation time.

In the XY model at low temperature, there is a bounded phase difference between two neighbors

$$\Delta\phi \leq \pi,$$

then the distortion is bounded by the size of the system L. Hence, the mixing time of the spin-wave mode is expected to be:

$$t_{\text{mix}} = \mathcal{O}(L). \tag{4.2.20}$$

The equilibrium autocorrelation has the same time scale as the harmonic model. We relate the first fast scaling in Fig. 4.2 to the harmonic "spin-wave" mode.

4.2.3 From the Spin Wave to the Vibrations in Solids

The systems at low temperature and high pressure behave similarly, for example the mechanical dynamics of solids can also be approximated by the quadratic interaction. Now we consider the harmonic solid model, whose Hamiltonian on a lattice is

$$H(\{X_i\}) = \sum_{\langle i,j\rangle} \frac{k}{2}(\vec{X}_i - \vec{X}_j + \vec{r}_{ij})^2, \tag{4.2.21}$$

where $\vec{X}_i$ is the displacement of the particle i from the origin, $\vec{r}_{ij}$ is the mean displacement of equilibrium from site i to j, $\langle i,j\rangle$ indicates the closest neighbors. Assuming that $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$, the Hamiltonian is transformed into:

$$\begin{aligned} H(\{X_i\}) &= \sum_{\langle i,j\rangle} \frac{k}{2}(\vec{X}_i - \vec{X}_j + \vec{r}_{ij})^2 = \sum_{\langle i,j\rangle} \frac{k}{2}\left((\vec{X}_i - \vec{r}_i) - (\vec{X}_j - \vec{r}_j)\right)^2 \\ &= \sum_{\langle i,j\rangle} \frac{k}{2}(\vec{u}_i - \vec{u}_j)^2 = \sum_{\langle i,j\rangle,x} \frac{k}{2}(u_{i,x} - u_{j,x})^2. \end{aligned} \tag{4.2.22}$$

where $\vec{u}_i$ is the displacement of the particle i from the position $\vec{r}_i$. The components of $(\vec{u}_i - \vec{u}_j)^2$ on all the dimensions are decoupled in the Hamiltonian, as the quadratic approximation of Heisenberg model we discussed before [99].

The fast scaling in the event-chain method can be extended to a lattice model with phonon mode, and to even more general cases with certain optimization. We will continue this discussion in chapter 5.

4.3 Vortices in the Two-Dimensional XY Model

In the quadratic approximation of two-dimensional XY model, there is a quasi-long-range order of the field ϕ [97]. However, the existence of short-range correlations is proven mathematically in the two-dimensional XY model at high temperature [100]. The physics changed by the harmonic approximation is that the phase space of a circle $\mathbf{S}^1$ is replaced with the linear space $\mathbf{R}$, while the mapping between two spaces with different topological structures will involve singularities, such as the mapping from $\mathbf{R}^2$ (the coordinates) to $\mathbf{S}^1$ in two-dimensional XY model. Within field theory, a

singularity appears as a "vortex". For a closed region, we use the winding number (or "charge") q to measure the number of vortices:

$$2\pi q = \oint_{\partial \mathcal{C}} \vec{\nabla}\phi(\vec{r}) \cdot \mathrm{d}\vec{l}, \tag{4.3.1}$$

and in the discrete lattice:

$$q(\mathcal{C}) = \sum_{k \in \partial \mathcal{C}} [\mod(\phi_k - \phi_{k+1}, 2\pi) - \pi]/2\pi. \tag{4.3.2}$$

$q = +1$ corresponds to a vortex and $q = -1$ stands for an anti-vortex. On a square lattice, the position of vortices are restricted on a plaquette with four spins on the corners. We assume that the vortices are positioned in the centers of these plaquettes in our later discussion.

The energy of a single vortex diverges if the system is infinitely large ($\propto \log L$) or infinitely divisible ($\propto (-\log a)$). The inferior cut-off (the lattice parameter a) and the superior cut-off (the size of the system L) restrict the energy of the system with a single vortex:

$$\begin{aligned} E_v &= \frac{1}{2} J_R \int_a^L 2\pi r \frac{1}{r^2} \mathrm{d}r + E_c \\ &= \pi J_R \log \frac{L}{a} + E_c, \end{aligned} \tag{4.3.3}$$

where J_R is the renormalized coupling constant (which may change with the lattice type and temperature), E_c is the core energy that counts for the contribution from $r < a$.

In a similar model (Villain's model, see section 1.6.4 and [64]), the vortex mode and spin-wave mode can be separated in the Hamiltonian *via* a duality transformation. The "vortex part" of the Hamiltonian exhibits an interaction between vortices isomorphic to that of the two-dimensional Coulomb gas:

$$U_{ij}(\vec{r}) = -\pi J_R q_i q_j \log \frac{r}{a} + 2E_c, \tag{4.3.4}$$

where r is the distance between two vortices, q_1 and q_2 are their charges. Different from the case of a single free vortex, the energy of a pair is no longer bounded by the size of the system.

The Kosterlitz–Thouless theory on the two-dimensional XY model attributes the short-range correlation at high temperature to the appearance of free vortices. The free energy of a single vortex is:

$$\begin{aligned} F_v &= E_v - T S_v \\ &\sim \pi J_R \log \frac{L}{a} - T \log \frac{L^2}{a^2} \\ &= (\pi J_R - 2T) \log \frac{L}{a}. \end{aligned} \tag{4.3.5}$$

As the system size increases, the existence of free vortices is probable only if F_v stays around 0. This argument then predicts the critical temperature

$$T_{\mathrm{KT}} = \frac{\pi J_R}{2} \text{ or } \beta_{KT} = \frac{2}{\pi J_R}, \tag{4.3.6}$$

above which free vortices are likely to appear. This phase transition is known as an infinite-order phase transition, that the derivative of the free energy of all the orders are continuous [91], the same as the solid-hexatic phase transition in section 3.3.1.

4.3.1 Vortex-pair Distance

According to eq. (4.3.4), the energy of a vortex–anti-vortex pair is bounded by its size instead of the system size. In analogy with eq. (4.3.5), the free energy of a vortex–anti-vortex pair is:

$$\begin{aligned} F_{vp}(d) &= U(d) - TS_{vp} \\ &\sim \pi J_R \log \frac{d}{a} - 2T \log \frac{L}{a}. \end{aligned} \tag{4.3.7}$$

From the equation we see that, at any given temperature (even below the critical one), the size of the vortex–anti-vortex pair can go to infinity as the system size increases infinitely. A phenomenological relation between the vortex–anti-vortex pair's distance and its corresponding system size is deduced:

$$d \propto L^{\frac{2T}{\pi J_R}}. \tag{4.3.8}$$

On the other hand, the net charge of all the vortices and anti-vortices inside a system under the periodic boundary condition is zero,

$$\sum_{v_i \in C} q_{v_i} + \sum_{a_i \in C} q_{a_i} = 0,$$

as the integral in eq. (4.3.1) compensates on the periodic boundaries ∂C. The charge inside a plaquette of a two-dimensional square lattice is either ± 1 or 0 (which means $q_{v_i} = 1$ and $q_{a_i} = -1$), according to eq. (4.3.2). Based on these facts, we will consider a method to classify the N vortices and N anti-vortices to N vortex–anti-vortex pairs, on a two-dimensional square lattice under the periodic boundary condition.

The physical system at low temperature tends to minimize its energy by alternating the alignment of the surrounding spins, as we see in Fig. 4.5. The first-order approximation of the total energy is the energy of the vortex–anti-vortex pairs, according to eq. (4.3.4):

$$E(\{\vec{d}_i\}) = \sum_i^N U_{v_i,a_i}(d_i) = \pi J_R \sum_i^N \log \frac{d_i}{a} + 2NE_c. \tag{4.3.9}$$

Regarding the principle of minimum energy, we provide a simple yet practical algorithm, by minimizing the objective function ϵ:

$$\epsilon(\{v_i, a_{P_i}\}) = \sum_{i=1}^N \log |\vec{R}_{v_i} - \vec{R}_{a_{P_i}}|, \tag{4.3.10}$$

where $\{P_i\}$ is obtained by a permutation P on the sequence $\{1, \ldots, N\}$, $\vec{R}_{v_i}$ and $\vec{R}_{a_i}$ denote the positions of the vortex v_i and anti-vortex a_i. The optimal pairing is given by:

$$\{v_i, a_{p_i}\} = \arg\min_{P \in S_N} \epsilon(\{v_i, a_{P_i}\}), \tag{4.3.11}$$

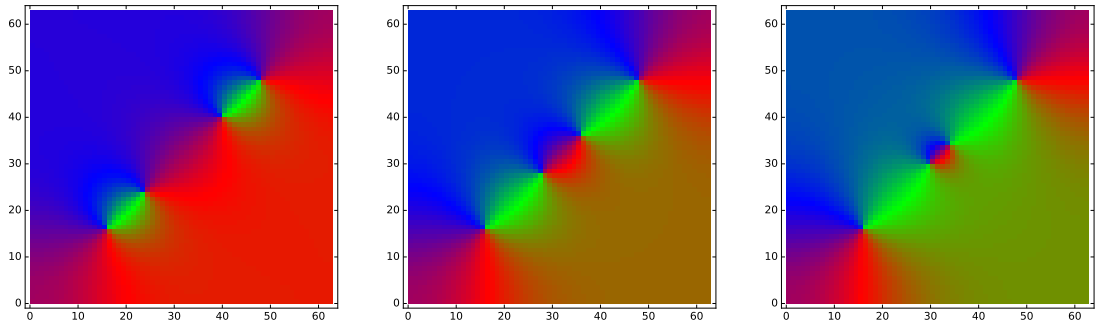


Figure 4.5: The configurations of two pairs of vortices. From left to right, the vortex and anti-vortex in the center get closer, and the orientation of the surrounding spins suggests that the pairing is changing from 2 symmetric ones to a large one and a small one. The color code is as in Fig. 4.6.

for all possible P from the permutation group S_N. The optimal assignment of the N vortex–anti-vortex pairs can be solved with the "Hungarian method" [101] for weighted matching problems. This method recursively reduces the size of the cost array and identifies the pairing of minimal ϵ, in the complexity of $\mathcal{O}(N^3)$. An example of the assignment is shown in Fig. 4.6.

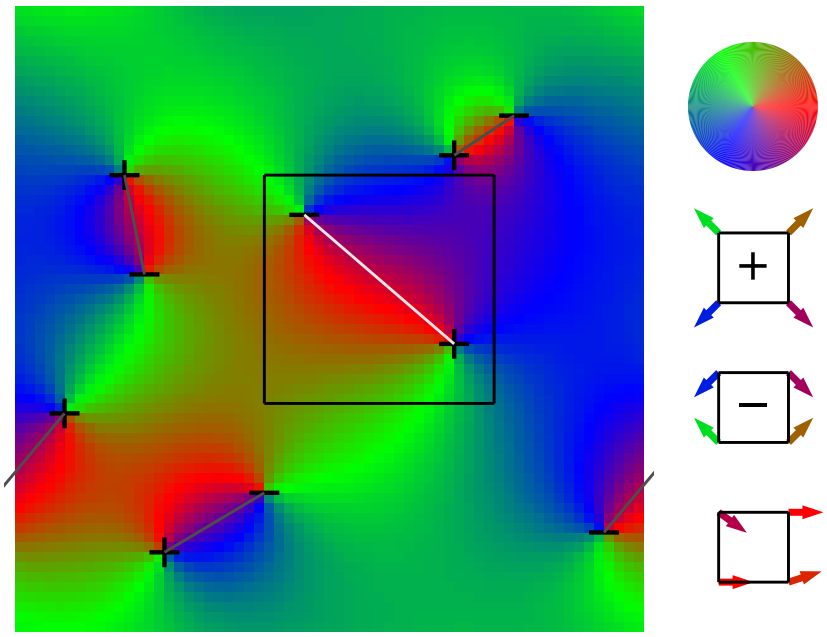
In the optimal assignment, the pair (v_i, a_{p_i}) of the largest separation defines the *max-distance* $d_{\max}$ of the configuration. It is noticed that the existence of large vortex–anti-vortex pairs and reduced susceptibility are strongly correlated (see Fig. 4.7). This is explained with the distortion of spin alignment by a vortex–anti-vortex pair: the larger pair it is, the wider domain in the configuration will be affected. In the following, we will use $d_{\max}$ to study the event-chain dynamics.

4.3.2 Maximum Vortex-pair Distribution

As an observable, the maximum vortex–anti-vortex pair distance has its own distribution in equilibrium, and autocorrelation time in event-chain dynamics. We again face the problem of defining the "worst" initial configuration for an empirical approach to the mixing time.

In analogy with the "out-of-equilibrium" study of the spin-wave mode in section 4.2.2, we begin with the evolution of the maximum vortex–anti-vortex pair distance in the extreme cases, such as Fig. 4.8 (a): in the initial configuration there are four widely separated vortices and anti-vortices (or a quadrupole) of size $\frac{L}{2}$. We measure the time τ_{conv} (counted in sweeps) when the susceptibility reaches its mean value at a given temperature, which can be seen as a "poor-man's" approximation to determining a mixing time. In Fig. 4.8 (b), at low temperature the mean time to reach equilibrium τ_{conv} is proportional to the square of the vortex–anti-vortex pair's size. Hence, the time to reach equilibrium of a vortex–anti-vortex pair in event-chain dynamics is proportional to its size's square (and independent of the system's size), as

[8] This is taken from the Fig. 2 in the Publication I [12].
[10] This is taken from the Fig. 3 in the Publication I [12].



(a) (b)

Figure 4.6: Vortices in the XY model. (a): A configuration with 5 vortices ("+") and 5 antivortices ("-"). The lines indicate matched vortex–anti-vortex pairs in the optimal assignment (see eq. (4.3.10)), and the length of the longest line (shown in white) equals the max-distance. A subsystem containing the max-distance pair is highlighted. (b) *from above*: The color code for the spin orientations, vortex, anti-vortex, and neutral plaquette configuration. [8]

$$\tau_{\text{conv}}(d) \propto d^2. \tag{4.3.12}$$

In the worst case, when the size of the vortex–anti-vortex pair is in the scale of the system L, it will take $\mathcal{O}(L^2)$ to annihilate. However, a vortex–anti-vortex pair as large as the system at a temperature below T_{KT} is rarely found since they are closely bound [91]; according to eq. (4.3.4), the distribution of its distance has a polynomial tail:

$$\begin{aligned} P(\vec{d}) &= \frac{1}{Z}\mathrm{e}^{-\beta E_p(d)} \\ &= \frac{1}{Z}\mathrm{e}^{-2\beta E_c}\,(d/a)^{-\pi\beta J_R} \propto d^{-\pi\beta J_R}, \end{aligned} \tag{4.3.13}$$

which means the probability decreases algebraically with the growth of the vortex–anti-vortex pair distance. This conclusion also applies to the maximum vortex–anti-vortex pair distance.

The maximum vortex–anti-vortex pair distance, as the maximum of a group of random variables, should have some max-stable property as mentioned in section 1.1.2. We consider a $L \times L$ system, which can be divided into n^2 practically "independent"

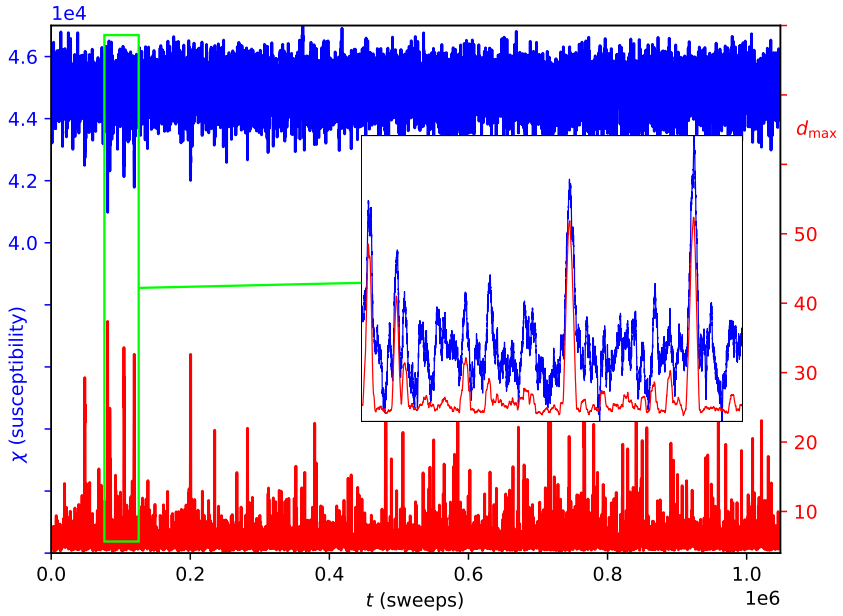


Figure 4.7: Time-evolution of the vortex–anti-vortex max-distance in the 384×384 XY model at $T/T_{\mathrm{KT}} = 0.933$ compared to that of the susceptibility (smoothed over a small time window). The inset illustrates the correlation between max-distance and susceptibility in greater detail. [10]

subsystems of size $L/n \times L/n$ ($L \gg na$). The max-distance of the large system is the maximum of n^2 "independent" max-distances of the smaller systems, if no vortex–anti-vortex pair is separated in the partition. (In the limit of infinite L, for any n, the ratio of being separated by the partition is going to zero.)
In other words, if

$$d_i \sim P(x;\beta,L), \text{ and } D = \max_{1\le i\le n^2}\{d_i\},$$

then

$$D \sim P(x;\beta,nL),$$

where D is the max-distance of the whole system, and d_i is the max-distance of each subsystem i.

We assume that the distribution of max-distance will converge to a stable one, in the limit of infinite L at a temperature below T_{KT}. Then it belongs to the extreme value distribution family discussed in section 1.1.2. The polynomial tail in eq. (4.3.13) leads to the Fréchet distribution, specifically:

$$f_{d_{\max}}(x) = \frac{\alpha}{s}\left(\frac{x}{s}\right)^{-1-\alpha} \exp\left[-\left(\frac{x}{s}\right)^{-\alpha}\right], \tag{4.3.14}$$

with its cumulative distribution

$$F_{d_{\max}}(x) = \exp\left[-\left(\frac{x}{s}\right)^{-\alpha}\right]. \tag{4.3.15}$$

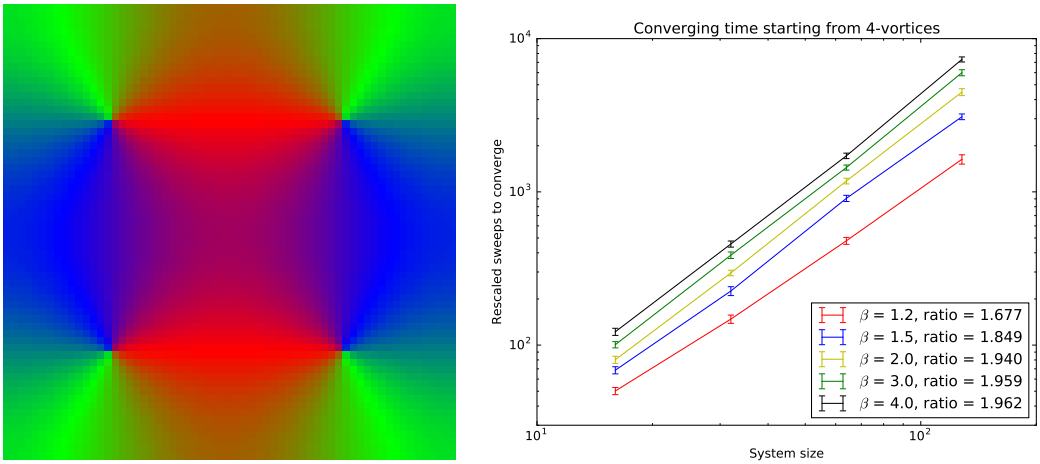


Figure 4.8: (a) The initial configuration used for the measurement of the τ_{conv} (which is similar to the mixing time). (b) The scaling of τ_{conv} (measured in sweeps) at different temperatures below T_{KT}. It is seen that the time grows as L^2 at low temperatures.

α describes the power-law decay of the max-distance distribution, and s sets its L-dependent scale. The minimum value of the max-distance is strictly 0, so there is no shift factor, and the scaling factor is proportional to $n^{1/\alpha} = L^{2/\alpha}$.
If $\alpha > 2$, the mean, the median and the mode of the Fréchet distribution grows slower than L, and the probability in the cut-off $\mathbf{P}(x > \sqrt{2}L)$ becomes negligible when L goes to infinity. Hence, we conjecture that the Fréchet distribution is the exact distribution of max-distance in the limit of large L, using eq. (4.3.10) for the objective function in the vortex–anti-vortex assignment.

Slightly below T_{KT}, the Fréchet distribution already provides a good fit for the max-distance distribution (see Fig. 4.9). By extracting the dependence of L from the scaling factor s, the Fréchet distribution of the d_{max} is then characterized by parameters $(\alpha, L^{2/\alpha}s_0)$, where α and s_0 are both dependent on the temperature, but not on the system size L (as shown in the inset of Fig. 4.9).

The scale parameter α of Fréchet distribution can be evaluated *via* $\langle d_{\text{max}} \rangle$, as

$$\langle d_{\text{max}}(L, \beta) \rangle \propto s(L, \beta) = s_0(\beta) L^{2/\alpha}. \tag{4.3.16}$$

The log–log plot of $d_{\text{max}} - L$ is a straight line, as shown in the inset of Fig. 4.10. By comparing eq. (4.3.16) to eq. (4.3.8), qualitatively we find:

$$\alpha \propto J_R/T. \tag{4.3.17}$$

The renormalized coupling constant J_R converges to a constant when $T \approx 0$, so the relation between α and $1/T$ is approximately linear at low temperature, as plotted in Fig. 4.10.

[12] This is a remake of the Fig. 4 in the Publication I [12].
[14] This is a remake of the Fig. 5 in the Publication I [12].

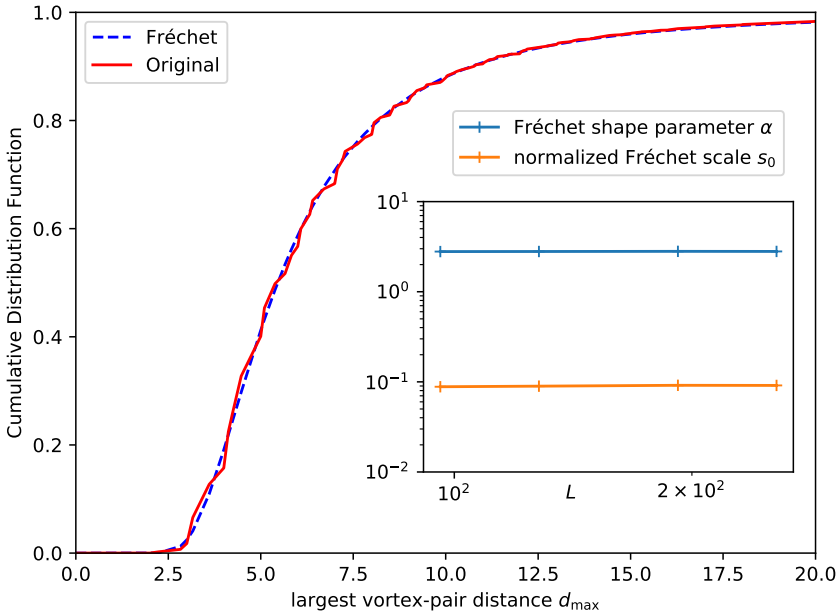


Figure 4.9: Cumulative max-distance distribution in the XY model at $L = 256$ and $T/T_{KT} = 0.965$ fitted with the Fréchet distribution in eq. (4.3.15) with $\alpha = 2.80$ and $s = 4.79$ ($s_0 = 0.0913$). The inset shows that the fitting parameters α and s_0 are independent of L.[12]

In the end, we note that for $\alpha = 2$, the distribution of d_{max} scales with $s \propto L$. This is observed for $T/T_{KT} \to 1^-$. Above T_{KT}, the Fréchet distribution no longer describes the maximum vortex–anti-vortex pair distribution, as its heavy tail is cut by the size of the system, but the distribution of the max-distance is still self-similar after rescaling.

4.3.3 Autocorrelation Time of the Vortex Mode

At low temperatures ($< T_{KT}$), d_{max} scales as $L^{2/\alpha} \ll L$ ($\alpha > 2$). According to our discussion in the beginning of the last section, it will take the time $\propto d_{max}^2$ to merge the vortex–anti-vortex pair in the event-chain dynamics. Hence, the autocorrelation time is expected to scale as $s^2 = L^{4/\alpha} s_0^2$. By combining the previous discussions, we have:

$$t_{cor} \sim L^{4/\alpha} \sim \begin{cases} L^2 & \text{for } T \to T_{KT}^-, \\ L^{\text{const } T} & \text{for } T \to 0. \end{cases} \tag{4.3.18}$$

The dynamical scaling parameter $z = 4/\alpha$ of the event-chain algorithm is thus connected to the scale parameter of a Fréchet distribution, and it is predicted to vanish in the zero-temperature limit. This relation $z = 4/\alpha$ is consistent with the autocorrelation of the susceptibility χ below the critical temperature (see Fig. 4.11). So far, we identify the second mode as the vortex mode.

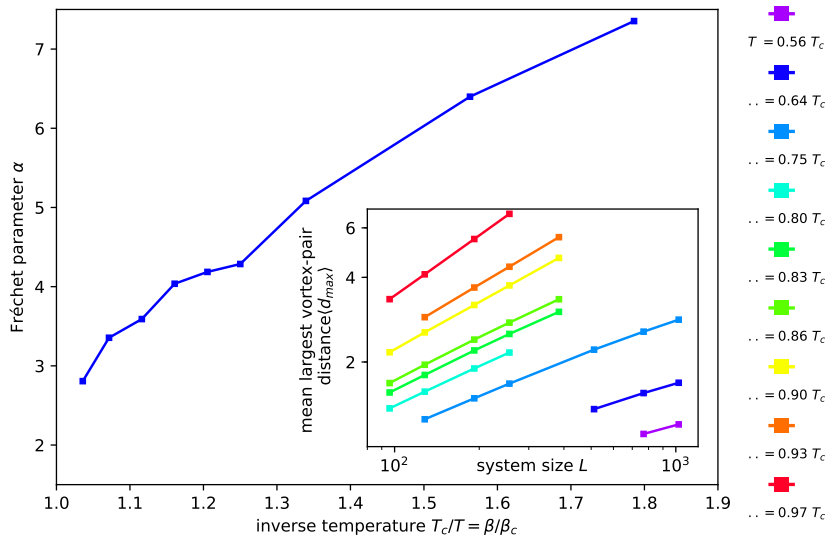


Figure 4.10: Scaling of the Fréchet parameter α with the inverse temperature β demonstrating that $\alpha \to \infty$ for $\beta \to \infty$ (Our theoretical model indicates $\alpha \propto \beta$ at low temperature). The inset shows $\langle d_{\max} \rangle$ with respect to system size L, from which α is obtained.[14]

The local Monte Carlo method takes a time of $\mathcal{O}(L^2)$ to recover from the least "preferable" state (which we presume to correspond to the mixing time) for both the spin-wave mode and the vortex mode. At low temperature the dynamical scaling of the local Monte Carlo is given by the harmonic model. The Wolff cluster algorithm has a much faster scaling, which we will discuss in section 4.4.3.

On the contrary, in the event-chain algorithm, the time to merge the largest vortex–anti-vortex pair (which corresponds to the mixing time in a rigorous context) is $\mathcal{O}(L^2)$, and the autocorrelation time is $\mathcal{O}(L^{4/\alpha}) > \mathcal{O}(\sqrt{\log L})$. It indicates that at any temperature $T > 0$, the autocorrelation exhibits two scales.

4.4 Other Topological Excitations

A vortex in a two-dimensional plane is arguably the simplest topological excitation. Many other topological excitations exist in field theory, and here we will only list a few of them that are frequently seen in condensed-matter theory, with special event-chain dynamics.

In the three-dimensional Heisenberg model, evidence shows that the autocorrelation time of event-chain dynamics is $\mathcal{O}(L)$ at the critical temperature [89] (and it decreases quickly to $\mathcal{O}(1)$ off the critical temperature). As we proved that the spin-wave mode in three-dimensional lattice has a correlation time $\mathcal{O}(1)$, the disagreement once again leads us to investigate the topological excitations.

[16] This is a remake of the Fig. 6 in the Publication I [12].

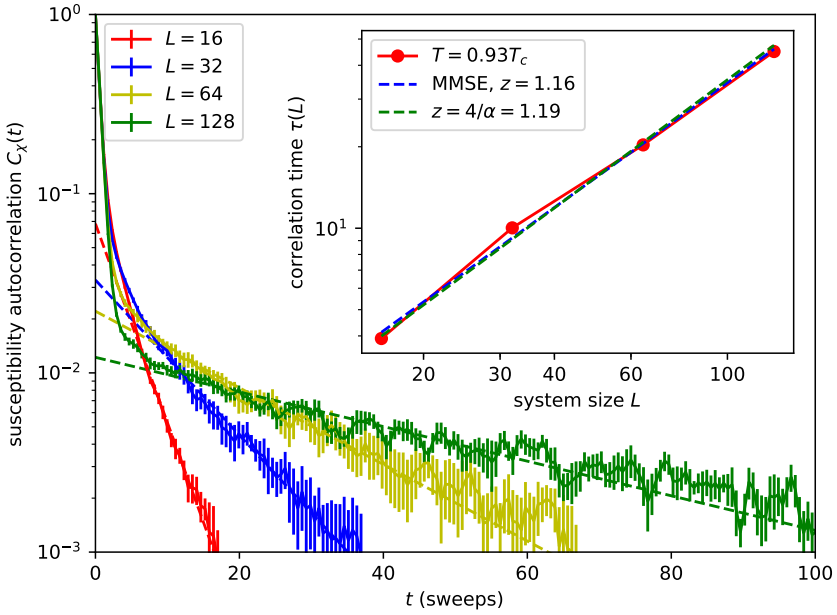


Figure 4.11: Susceptibility autocorrelation at $T = 0.93T_{\mathrm{KT}}$ for different system sizes in the two-dimensional XY model. The inset shows their equilibrium correlation time, compared with the scaling z obtained by Minimum Mean Square Error method and the fitting using $z(T) = 4/\alpha(T)$ from Fig. 4.10. The result of Fréchet theory of the vortex distribution agrees with the autocorrelation of the susceptibility.[16]

4.4.1 Monopoles in the 3D Heisenberg Model

Monopole is a topological defect in the mapping from $\mathbf{R^3}$ to $\mathbf{S}^2$. In field theory, the winding number/charge is defined as:

$$Q = \frac{1}{4\pi} \oint_{\partial C} \mathrm{d}^2 x \, (\vec{s} \cdot (\partial_1 \vec{s} \times \partial_2 \vec{s})), \tag{4.4.1}$$

where C is a bulk in R^3, ∂C is its surface. $\vec{s}(\vec{x})$ is an $O(3)$ field with $\vec{s}^2 = 1$, Q is the winding number/charge of the monopole enclosed in C. In the lattice model, the integral can be replaced by the summation on all the surfaces, and each surface is triangularized to multiple oriented components. Here, the charge inside each triangle $(\vec{s}_1, \vec{s}_2, \vec{s}_3)$ has a relation:

$$\begin{aligned} \mathrm{e}^{i2\pi q(\vec{s}_1,\vec{s}_2,\vec{s}_3)} &= \frac{1 + \vec{s}_1 \cdot \vec{s}_2 + \vec{s}_2 \cdot \vec{s}_3 + \vec{s}_3 \cdot \vec{s}_1 + i\vec{s}_1 \cdot (\vec{s}_2 \times \vec{s}_3)}{\sqrt{2(1 + \vec{s}_1 \cdot \vec{s}_2)(1 + \vec{s}_2 \cdot \vec{s}_3)(1 + \vec{s}_3 \cdot \vec{s}_1)}}, \\ |q(\vec{s}_1, \vec{s}_2, \vec{s}_3)| &= \frac{1}{2\pi} \arccos\left(\frac{1 + \vec{s}_1 \cdot \vec{s}_2 + \vec{s}_2 \cdot \vec{s}_3 + \vec{s}_3 \cdot \vec{s}_1}{\sqrt{2(1 + \vec{s}_1 \cdot \vec{s}_2)(1 + \vec{s}_2 \cdot \vec{s}_3)(1 + \vec{s}_3 \cdot \vec{s}_1)}} \right), \end{aligned} \tag{4.4.2}$$

where the sign of q is determined by

$$\mathrm{sign}(q(\vec{s}_1, \vec{s}_2, \vec{s}_3)) = \mathrm{sign}[\vec{s}_1 \cdot (\vec{s}_2 \times \vec{s}_3)].$$

The charge on the surface is proportional to the signed area of the spherical triangle with corners of $(\vec{s}_1, \vec{s}_2, \vec{s}_3)$, and their summation around a closed bulk is an integer. [102]

Evidence has shown an abundance of monopoles near the critical temperature [93]. On the other hand, the energy of a monopole–anti-monopole pair is proportional to its size [103, 104],

$$E(R) = \mathcal{O}(R).$$

It is very unlikely to have a large monopole–anti-monopole pair in equilibrium, as it would cost less energy by breaking them into two smaller pairs.

As a non-local topological defect, a large monopole–anti-monopole pair will not vanish easily with local algorithms. We prepare the initial configuration of four monopoles and four anti-monopoles (namely an octupole) and find the time to merge them is at least proportional to distance of a pair at low temperatures in the event-chain algorithm:

$$t_{\text{merge}}(d) \propto d.$$

It is faster than the vortex mode, as the interaction between monopoles is stronger. Hence, the mixing time from the most unfavorable configuration to equilibrium is $\mathcal{O}(L)$ sweeps.

4.4.2 Bloch Mode

In the previous discussion we mentioned that vortices appear in the mapping from $\mathbf{R^2}$ to $\mathbf{S}^1$. In the simulation with periodic boundary condition, another topological defect, which is called a "Bloch mode", merges in the mapping from $\mathbf{T^2}$ (two-dimensional lattice under periodic boundary condition, like a torus) to $\mathbf{S}^1$. It is illustrated in Fig. 4.12, and described with the equation:

$$\phi(x) = 2\pi \frac{x}{L}. \tag{4.4.3}$$

This mode is topologically non-trivial: if we select the contour ∂C in eq. (4.3.1) winding across the periodic boundary (e.g. from $x = L/2$ to $x = L$, across $x = 0$ and ending at $x = L/2$), the integration will give a non-zero charge. In other words, it is an extreme case of the vortex, which cannot be wiped out by a sequence of local moves.

The Bloch mode is a slow mode of the XY model in two and three dimensions, for both the event-chain algorithm and the local Monte Carlo method, although it would have to be considered for an analysis of the mixing time. Both algorithms take a time scale of $\mathcal{O}(L^2)$ at low temperature. However, this mode has too few degrees of freedom; it is too "clean" to appear at low temperature, while at higher temperature the vortex will destroy its alignment. Therefore, it plays less of a role in our discussions. This mode is unstable in Heisenberg model, because the extra dimension of θ permits it to bypass the barrier without any increase in the energy.

4.4.3 Comparison with Other Algorithms

Based on the discussions before, we summarize the dynamics of event-chain algorithm of 2D XY model in Table 4.1.

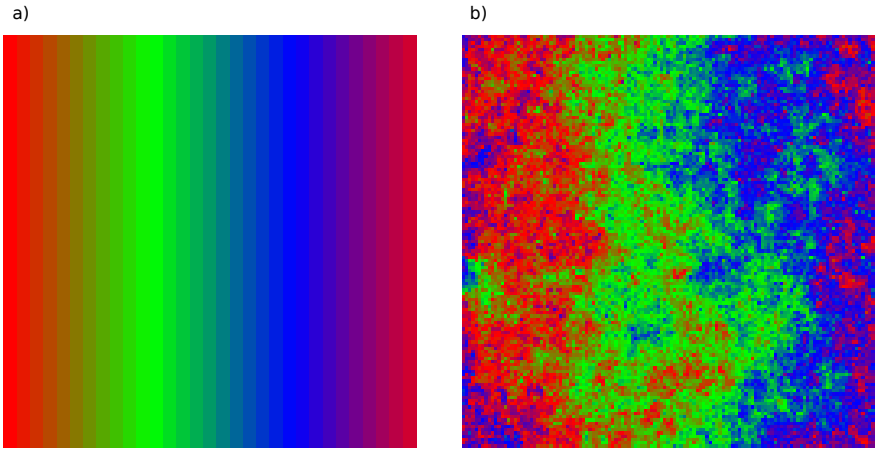


Figure 4.12: (a) An illustration of the Bloch mode in a two-dimensional XY model, according to eq. (4.4.3). (b) A relaxation from the Bloch mode in (a) using the event-chain algorithm at $T = 0.75T_{\mathrm{KT}}$. The pattern is changed very little. The color code is as in Fig. 4.6.

Algorithm	Mode	Mixing time	Autocorrelation time
ECMC	spin wave	$\mathcal{O}(L)$	$\mathcal{O}(\sqrt{\log L})$
	vortex	$\mathcal{O}(L^2)$	$\mathcal{O}(L^{4/\alpha})$
	Bloch mode	$\mathcal{O}(L^2)$	-
LMC	spin wave	$\mathcal{O}(L^2)$	$\mathcal{O}(L^2)$
	vortex[17]	$\mathcal{O}(L^2)$	$\mathcal{O}(L^{4/\alpha})$
	Bloch mode	$\mathcal{O}(L^2)$	-

Table 4.1: The time scaling of event-chain Monte Carlo and local Monte Carlo.

The mixing time and the autocorrelation time of the spin-wave mode in local Metropolis dynamics are $\mathcal{O}(L^2)$, as shown in Fig. 4.4. For the vortex mode, the mixing time also takes about $\mathcal{O}(L^2)$ sweeps. These lead to a fact that we cannot distinguish the vortex mode from the spin-wave mode in the autocorrelation.

The topological defect can evolve faster with non-local dynamics. For example, Wolff's cluster algorithm in section 1.6.4 [63] provides the vortices in two-dimensional XY model a chance to "hop" from one side of the cluster to another side by flipping it, resulting in a much shorter mixing time and equilibrium autocorrelation time:

$$t_{\mathrm{mix}}^{\mathrm{Wolff}} \sim \mathcal{O}(1),\ t_{\mathrm{cor}}^{\mathrm{Wolff}} \sim \mathcal{O}(1).$$

The fast dynamics also applies to the Bloch modes, and also to the monopole–anti-monopole pairs in three-dimensional Heisenberg model.

[17] However, we are not able to verify the autocorrelation time in the simulation of local Monte Carlo (as it is impossible to split the vortex mode from the spin-wave mode, if they have the same scaling). Here we put our conjecture.

4.5 Conclusion

In this chapter, we observed a large speed-up for the relaxation of spin-wave excitations ($z \sim 0$) of the event-chain algorithm compared to the local Monte Carlo algorithm ($z \sim 2$). We also studied the relaxation of the topological excitations, including vortex–anti-vortex pairs in the two-dimensional XY model and monopole–anti-monopole pairs in three-dimensional Heisenberg model. We identified the distribution of the maximum vortex–anti-vortex pair distance $d_{\max}$ as a Fréchet distribution. In the XY model, the potential of a vortex–anti-vortex pair is weak, which leads to a $d_{\max}^2$ relaxation time, and an L^2 mixing time at the worst case. Nevertheless, the autocorrelation time scale indicated by the Fréchet parameter $\alpha(T)$ is smaller, which goes to $\mathcal{O}(L^0)$ at the low-temperature limit. These results agree with the autocorrelation of the susceptibility in previous studies [11], as the spin-wave mode outspeeds the vortex mode, resulting in a two-scale phenomenon.

Furthermore, in the particle system, a fast-relaxing phonon mode in event-chain algorithm is expected in analogy with the spin-wave mode. In chapter 5 we will discuss such systems (e.g. Lennard-Jones systems), with the goal of achieving fast dynamics.

Chapter 5

Event-chain Algorithm with Factor Fields

In chapters 3 and 4, we studied two distinct Markov-chain dynamics of the event-chain algorithm in one-dimensional models. Depending on the precise algorithm, the hard-sphere system was found to have a mixing time of $\mathcal{O}(N^2 \log N)$ for an appropriately chosen distribution of chain length (see Table 3.1). On the other hand we have reasons to believe that the harmonic spin-wave/lattice-wave system has an autocorrelation time of $\mathcal{O}(N^{3/2})$ events (listed in Table 4.1). For a system with an interaction between these two classes (e.g. the Lennard-Jones potential, which reduces to the harmonic interaction at high density with a bounded attraction at long range), we also expect a dynamics between these two, $\mathcal{O}(N^{3/2})$ and $\mathcal{O}(N^2 \log N)$.

In the present chapter, we will begin with the lifting scheme of multiple particles and multiple types of interactions, introduce the optimization with a factor field, and analyze the behavior of the event-chain algorithm in the optimal dynamics. The final result will be that, although irreversible Markov chains accelerate mixing-time scales with respect to the reversible Markov chains, their behavior can be improved even further by tuning them to a special point corresponding to the harmonic model.

5.1 Factor Fields in Event-chain Dynamics

5.1.1 General Lifting Scheme of Interacting Systems

In chapter 4 we discussed spin models with nearest-neighbor interactions, such as the XY model and Heisenberg model. For a general system of N interacting particles, all the interactions are split into a set of factors $\mathcal{M}$ [1], that

$$\mathcal{M} = \{(I_M, T_M) | I_M \in \mathcal{P}(\{1, 2, \ldots, N\}), T_M \in \mathcal{T})\}, \tag{5.1.1}$$

where $\mathcal{P}$ is the power set of the indices (consisting of single ones, pairs, triplets and more), and $\mathcal{T}$ is the set of interaction types. For the total potential energy, we have

$$E(\{\mathbf{r}_i\}) = \sum_{M \in \mathcal{M}} E_M(\{\mathbf{r}_i : i \in I_M\}), \tag{5.1.2}$$

[1] This paragraph of discussion is largely based on the work by M. Faulkner et al. [16, Section II].

where E_M has the interaction of type T_M, depends on the factor indices I_M. According to the factorized Metropolis filter discussed in section 2.3.2, the acceptance rate of a move

$$P^{\text{acc}}_{\text{FacMetro}}(\{\delta E^+_M | M \in \mathcal{M}\}) = \prod_{M\in\mathcal{M}} \mathrm{e}^{-\beta(\delta E^+_M)} = \exp\left(-\beta \sum_{M\in\mathcal{M}} \delta E^+_M\right), \tag{5.1.3}$$

where δE^+ equals δE if it is positive, and equals zero otherwise, as previously defined in eq. (2.3.5). A rejection-activation lifting scheme is given: if the move is rejected by the factor of δE_M, then $i \in I_M$ can be chosen as the next active particle, which is uniquely determined if the factor comprises only two particles or spins (it can be generalized for the factor sets containing more elements).

5.1.2 One-dimensional Lennard-Jones Model

Now we consider a one-dimensional particle system with nearest-neighbor Lennard-Jones interactions. The potential is

$$V_{\text{LJ}}(r) = \frac{1}{r^{12}} - \frac{1}{r^6}, \tag{5.1.4}$$

where we choose the natural unit, for which the two terms on the right-hand side of eq. (5.1.4) indicate a repulsive potential at close range and an attractive potential at long range. The minimum energy is reached at $r_0 = 2^{\frac{1}{6}} \approx 1.122$, with $V_{\text{LJ}}(r_0) = -1/4$. The interaction is attractive if $r > r_0$, otherwise it is repulsive.

With the discussion of the section 5.1.1, the factor set is

$$\mathcal{M}_{\text{LJ}} = \{(\{i, i+1\}, \text{LJ}) : i \in 1, \ldots, N\}. \tag{5.1.5}$$

An alternative is more widely used

$$\mathcal{M}_{12+6} = \{(\{i, i+1\}, \text{LJ}_{12}), (\{i, i+1\}, \text{LJ}_6) : i \in 1, \ldots, N\}, \tag{5.1.6}$$

where LJ_{12} and LJ_6 stand for the repulsion and attraction respectively:

$$V_{\text{LJ}_{12}}(r) = \frac{1}{r^{12}}; V_{\text{LJ}_6}(r) = -\frac{1}{r^6}.$$

An active particle i moving in the positive direction will trigger the rejection with either the factor $((i, i+1), \text{LJ}_{12})$ or $((i, i-1), \text{LJ}_6)$ in $\mathcal{M}_{12+6}$, as illustrated in Fig. 5.1(a).

It is observed that in the special case where $L = Nr_0$, the event-chain Monte Carlo with the factor set $\mathcal{M}_{\text{LJ}}$ has an autocorrelation time of $\mathcal{O}(N^{3/2})$ single steps, the same as the spin-wave mode in section 4.2. However, if there is a slight change of the density (e.g. $L = 0.99Nr_0$), the autocorrelation time increases to $\mathcal{O}(N^2)$ single steps again. In the case of $L = 0.99Nr_0$ at high temperature $T \sim |V_{\text{LJ}}(r_0)|$, the event-chain algorithm is fast with the factor set $\mathcal{M}_{\text{LJ}}$ in eq. (5.1.5), and its autocorrelation time is $\mathcal{O}(N^2)$ single steps (a dynamical exponent $z = 1$ with restarts, see Fig. 5.1(b)), agreeing with the dynamics of the lifted forward Metropolis algorithm in one-dimensional

[3] This figure is taken from our submitted manuscript in Publication III [20].

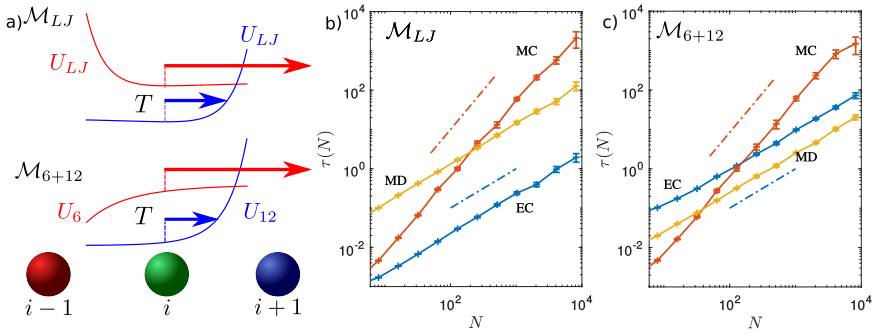


Figure 5.1: (*In this figure all measurements are taken in number of sweeps, so that one sweep* $\sim N$ *events.*) (a) illustrates the dynamics of the event-chain algorithm with the factor sets $\mathcal{M}_{LJ}$ and $\mathcal{M}_{6+12}$. (b) and (c) show the integrated autocorrelaton times of the structure factor (of the lowest Fourier component, see eq. (4.2.17)) as a function of system size N, for the one-dimensional periodic Lennard-Jones model ($\langle r \rangle = 0.99 r_0$). The autocorrelation times of reversible local Markov-chain dynamics (red), the event-chain Monte Carlo with restarts (blue) and molecular dynamics (yellow) are plotted. (b): $\beta|V_{\mathrm{LJ}}(r_0)| = 0.1$, the event-chain Monte Carlo with $\mathcal{M}_{LJ}$ (the result of $\mathcal{M}_{12+6}$ is similar). (v): $\beta|V_{\mathrm{LJ}}(r_0)| = 10$, the event-chain Monte Carlo with $\mathcal{M}_{12+6}$. In both cases, the scaling of the event-chain Monte Carlo and Molecular dynamics is $z = 1$, while in reversible local Markov-chain dynamics $z = 2$.[3]

hard-sphere model in section 3.1.4. At low temperature, the event-chain Monte Carlo with the factor set of $\mathcal{M}_{12+6}$ slows down largely, but keeps the scaling of $z = 1$ (see Fig. 5.1(c)). In the following, we will find the reason behind the different scalings in the event-chain dynamics.

5.1.3 Another Visit to the Harmonic-solid Model

As the one-dimensional Lennard-Jones model in the low-temperature limit is described by its quadratic approximation, we expect to recover the superior autocorrelation times discussed in section 4.2 in this model also. To this aim, we investigate the one-dimensional harmonic chain in section 4.2.3. The interaction between two particles is $\frac{k}{2}(\Delta x - a)^2$, where a is the lattice parameter. It could be either attractive ($\Delta x > a$) or repulsive ($\Delta x < a$).

In more general cases, we assume that the energy between two particles is

$$E_{\mathrm{harm}}(\Delta x; b) = \frac{k}{2}(\Delta x - b)^2, \tag{5.1.7}$$

where the equilibrated distance between particles is b. Then the total energy of a chain

of length L with N particles under the periodic boundary condition is

$$\begin{aligned} E_{\text{Harm}}(\{x_i\};b) &= \frac{k}{2}(x_2-x_1-b)^2+\cdots+\frac{k}{2}(x_N-x_{N-1}-b)^2+\frac{k}{2}(x_1+L-x_N-b)^2 \\ &= E_{\text{Harm}}(\{x_i\};0)-kbL+\frac{1}{2}Nkb^2 \\ &= E_{\text{Harm}}(\{x_i\};0)+\frac{1}{2}Nk(b-L/N)^2-kL^2/(2N). \end{aligned} \tag{5.1.8}$$

It means that the parameter b is irrelevant: any b simply shifts the ground-state energy, and the minimum is reached at $b=L/N$. In other words, different E_{Harm}'s have the same derivatives (the same force on each particle):

$$\frac{\partial E_{\text{Harm}}(\{x_i\};b)}{\partial x_i} = k(x_i-x_{i-1}-b)+k(x_{i+1}-x_i-b) = \frac{\partial E_{\text{Harm}}(\{x_i\};0)}{\partial x_i}, \tag{5.1.9}$$

which shows that the mechanical property of the model is unchanged. Both molecular-dynamics and reversible Monte Carlo algorithms are insensitive to b (as we will show, this is however not the case for the event-chain Monte Carlo approach). This argument also works for fixed boundary conditions.

In an extreme case, even if the interaction is purely attractive ($b=0$), the steady state will stay the same, which is verified by all Monte Carlo methods, including Metropolis, heat-bath and event-chain algorithms. Metropolis and Glauber dynamics are identical for any b according to eq. (5.1.9). However, in the event-chain algorithm, the dynamics is largely determined by the form of interactions. The speed and even the time scale would be changed by b.

The value of b firstly determines the step size. For a system in equilibrium ($x_{i+1}-x_i\approx L/N$) with $b'=b-L/N=0$, it is estimated that

$$\frac{1}{2}ks^2\approx T; \langle s\rangle \approx \sqrt{\frac{2T}{k}}, \tag{5.1.10}$$

where s is the step size. But in the case $b'\gg\langle s\rangle$,

$$\begin{gathered} \frac{1}{2}k\left((s+b')^2-b'^2\right)=\frac{k}{2}s^2+kb's\approx T, \\ s<T/kb'\ll T/k\,\langle s\rangle\,, \end{gathered} \tag{5.1.11}$$

as illustrated in Fig. 5.1.3. The mean step size $\langle s\rangle$ increases as $|b'|$ decreases, which leads to an acceleration of the speed.

On the other hand, we attribute the influence on the scaling to the nearly deterministic triggering. In the case that $b'<0$ (as $b=0$ in Fig. 5.1.3 (a)), the probability of triggering the left is much larger; as b' (b) increases, the left side is less biased, and finally when $b'=0$ ($b=L/N$), the lifting index on particle i has equal probability to move in both the directions, as shown in Fig. 5.1.3 (c). This is a key of fast dynamics which will be used in the general optimization. We will continue the discussion of the special motion of the lifting index in section 5.3.

[5]This figure is taken from our submitted manuscript in the Publication III [20].

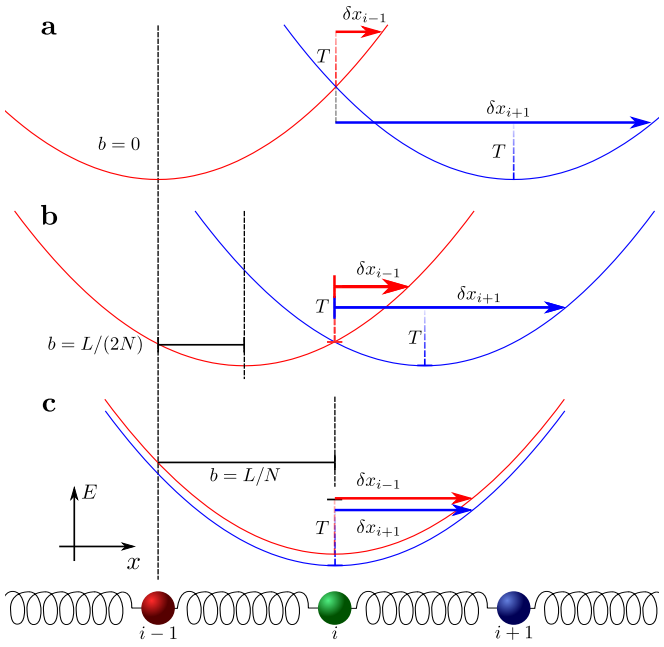


Figure 5.2: Triggering of particles attached to a harmonic string in the event-chain dynamics. The green particle is in the middle of the red and the blue. (a) shows the case of $b = 0$ (pure attraction), the red and blue curves indicate the harmonic interactions of the corresponding particles. In (b) b increases, and finally in (c) $b = L/N$, the harmonic interactions of the two corresponding particles overlap. The proposed moves are indicated by the solid colored arrows, with respect to the temperature T (vertical dashed lines). In (a), the displacement s in eq. (5.1.11) is the determined by the shorter of the two (the red, δx_{i-1}). In (b) s increases, and in (c) it reaches its maximum, with the same size of proposed moves on both sides. [5]

5.1.4 Factor Fields and Pair-wise Decomposition

The shift of harmonic model in eq. (5.1.8) is a compensation of linear terms. For more general models, a shift can be added similarly. We suppose the energy of the system is:

$$E(\{x_i\}) = V(x_1 - x_0) + V(x_2 - x_1) + \cdots + V(x_n - x_{n-1}). \tag{5.1.12}$$

In the periodic boundary condition we have $x_0 = x_n - L$, and in the fixed boundary condition x_0 (x_n) has a fixed value. The energy is rewritten as :

$$\begin{aligned} E(\{x_i\}) &= V(x_1 - x_0) + V(x_2 - x_1) + \cdots + V(x_n - x_{n-1}) \\ &\quad + h\left[(x_1 - x_0) + (x_2 - x_1) + \cdots + (x_n - x_{n-1})\right] - hL \\ &= \tilde{V}(x_1 - x_0) + \tilde{V}(x_2 - x_1) + \cdots + \tilde{V}(x_n - x_{n-1}) - hL, \end{aligned} \tag{5.1.13}$$

where $\tilde{V}(r) = V(r) + hr$, h is a constant, namely the "factor field". By choosing an appropriate h, one may eliminate the linear term in $V(r)$ near the ground state to have a fast dynamics.

We take the Lennard-Jones model for example. The potential is expanded to its quadratic term:

$$V_{\mathrm{LJ}}(r) = \frac{1}{r^{12}} - \frac{1}{r^6} = \left(\frac{1}{\Delta^{12}} - \frac{1}{\Delta^6}\right) + \frac{6(-2+\Delta^6)}{\Delta^{13}}(r-\Delta) + \left(\frac{78}{\Delta^{14}} - \frac{21}{\Delta^8}\right)(r-\Delta)^2 + \ldots. \tag{5.1.14}$$

By choosing the factor field as

$$h_{\mathrm{LJ}} = -\frac{6(-2+\Delta^6)}{\Delta^{13}}, \tag{5.1.15}$$

where $\Delta = L/N$, the linear term in the expansion is compensated in the simulations.

With this optimization, it is observed that the autocorrelation time of the structure factor in the Lennard-Jones chain is $\mathcal{O}(N^{3/2})$ single steps, the same as the scaling of the spin-wave. At higher temperature, the factor field replaced by the pressure helps the relaxation in more general cases. For its mechanism, we will explain in the following section.

5.2 Generalized Factor Fields

The factor field was initially introduced to compensate the linear term of interaction. However, without being necessarily equal to the linear term, it can still accelerate the dynamics largely in practice. We will begin with the example of hard spheres where the interaction has no linear term, and identify the optimal value of factor field.

5.2.1 Factor Fields in the Hard-sphere Model

This factor field optimization of eq. (5.1.13) applies to all one-dimensional interacting systems including the hard spheres. For N volume-less particles on a periodic interval $[0,1)$ with a given factor field h (we assume $L_{\text{free}} = 1$), the dynamics is simple:

- The proposed displacement of particle i follows an exponential distribution:

$$\delta x \sim \mathrm{Exp}(\beta h) \quad (\mathbb{E}(\delta x) = \frac{1}{\beta h}).$$

- if $x_{i+1} - x_i > \delta x$, x_i is updated to $x_i + \delta x$, and the particle $(i-1)$ is triggered; otherwise x_i is updated to x_{i+1}, and the particle $(i+1)$ is triggered.

We choose $h = N/\beta$ (which is also the pressure of the system [105]), so the step size is proportional to $1/N$.

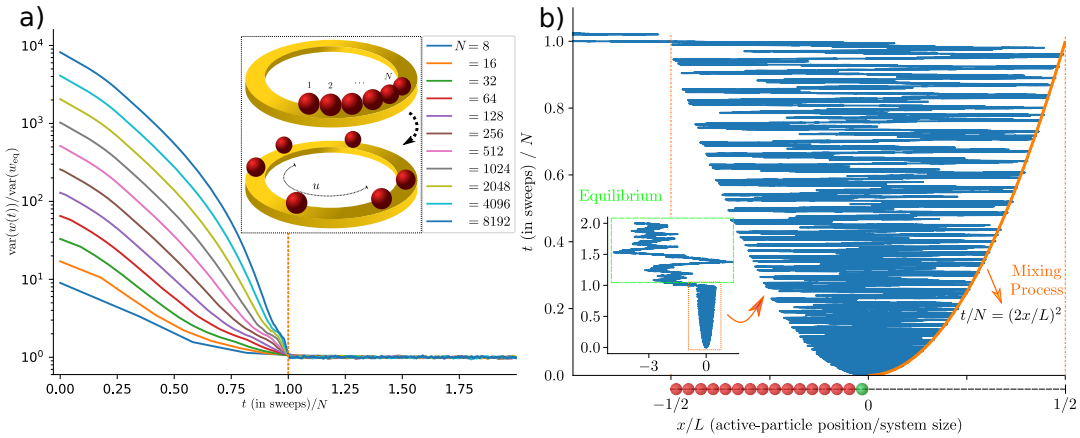


Figure 5.3: The mixing process in hard spheres. (a): The evolution of normalized u^2 ($u = x_{i+N/2} - x_i - 0.5$). The values of different systems all converge after about N^2 events, indicating a mixing time $(1 \pm 0.005)N^2$. The inset illustrates the initial configuration and the measurement of u. (b): The position of the active particle as a function of time with a compact initial configuration ($N = 4096, a = 1/(2N), L = 1$, the interval $[-1/2, 0]$ is fully occupied with disks as illustrated). The end of the system relaxes *via* oscillations of the active particle between the (right-most) free end and the (left-most) compact starting point. The left inset shows the movement of the active particle on a larger time scale, that after N^2 events the active particle performs a Lévy walk (see Fig. 5.4 (b) for details); the right curve shows the position of the active particle fitted to a parabola, which agrees with the N^2 scaling. The transition point between the different behaviors agrees with the mixing-time estimate in (a). [7]

This algorithm is different from the previous ones in section 3.1, as it is possible to trigger in both directions, and an infinite chain preserves irreducibility. The study of the structure factor again shows an autocorrelation time of $\mathcal{O}(N^{3/2})$ single steps, the same as those in the harmonic model and the Lennard-Jones model. Nevertheless, in the mixing process starting from a compact initial configuration, the estimation based on the "half-system distance"(see section 3.2.2) indicates a mixing time of N^2 single

[7] We record in sweeps (when the total displacement reaches a given amount) instead of events, because the configuration after an event is very special (e.g. a collision of particles) in the event-chain algorithm. These figures are taken from our submitted manuscript in the Publication III [20].

steps in Fig. 5.3 (a). It agrees with the estimation of the harmonic model with an upper bound of the difference between neighbors (see eq. (4.2.20)).
Moreover, in the investigation of the movement of the active particle, it is seen that the compact configuration expands to the full system as the active particle extends the boundary on the positive direction, and drags the others left behind with the factor field, which forms an oscillation between the two sides (see Fig. 5.3 (b)). In this optimal dynamics, the mean displacement of the active particle is zero, as predicted in section 5.1.3. After the mixing process, the active particle will move much longer in the same direction, as we will present in section 5.3.1.

5.2.2 Pressure as the Optimal Factor Field

We find the optimal factor field and the fast dynamics coincidentally in the hard spheres, by choosing the factor field as the *mean pressure* of the system. The dynamical scaling ($z = 1/2$ or $t_{\text{cor}} = \mathcal{O}(N^{3/2})$ in dimension one) is also seen in other models with the optimal factor field.

We conjecture that the optimal factor field is the *mean pressure* in all the systems, with which the active particle has equal probabilities to trigger its neighbors in both directions. The mean pressure in a canonical ensemble is:

$$\langle p(N,T)\rangle = -\frac{\partial F(N,L,T)}{\partial L}, \text{ where } F = -\frac{1}{\beta}\frac{\partial \log Z(N,T,L)}{\partial \beta}. \tag{5.2.1}$$

In the example of biased harmonic model, the pressure is

$$\langle p_{\text{harm}}(b)\rangle = k(b - L/N). \tag{5.2.2}$$

which equals the force exerted by the particle to a(n) (imaginary) boundary. The system with $b = L/N$ satisfies $p_{\text{harm}} = 0$. Similar calculations can be done for one-dimensional hard spheres and Lennard-Jones model. In more general cases, the virial expression provides a better clue.

Virial of the Pressure

In the canonical ensemble, the virial expression of the pressure is:

$$\beta p = \rho + \frac{1}{V}\left\langle \sum_{\langle i,j\rangle} (x_j - x_i)\frac{\beta \partial E_{ij}(\mathbf{r}_i - \mathbf{r}_j)}{\partial x_i} \right\rangle. \tag{5.2.3}$$

In the event-chain algorithms with lifting schemes, the average is taken from the replicas with different active particles. By summing up the lifting events in a chain, it is obtained that:

$$\beta p = \rho \left\langle \frac{x_{\text{final}} - x_{\text{initial}}}{l} \right\rangle_{\text{chains}}, \tag{5.2.4}$$

where x_{initial} and x_{final} indicate the coordinates of the active particle in the initial and final configuration, and l is the total length of the chain. The eq. (5.2.4) is deduced from hard-sphere models and examined with continuous potential [10], stating that

the mean displacement of the active particle is proportional to the pressure of the system.

By introducing the factor field $h = p$, the effect of pressure in eq. (5.2.4) is compensated: the mean displacement of the active particle $(x_{\text{final}} - x_{\text{initial}})$ is zero. This result agrees with the dynamics in Fig. 5.3 (b) and the argument of equal probabilities of triggering in all directions in section 5.1.3. Hence we conjecture $h = p$ is the optimal condition in general cases of event-chain dynamics.

5.3 Dynamics of Lifting Index in Event-chain Algorithm

The event-chain dynamics in hard spheres and the Lennard-Jones model with the optimal factor field accelerates the convergence even faster than the algorithms discussed in chapter 3. We attribute the dynamics to the "reversible triggering", that the neighbors in all the directions have the same probability to be activated. It shows an advantage over the "propagative triggering" (such as the algorithms in the TASEP and lifted TASEP classes in Table 3.1, whose lifting index always moves in the "forward" direction).

The movement of the lifting index was investigated in the multi-dimensional XY models by K. Kimura et al. in 2017, which exhibits some super-diffusive behavior at the critical temperature [106]. However, the mechanism was not well explained. As the same scaling of the autocorrelation time ($\mathcal{O}(N^{3/2})$) is seen in harmonic model and hard spheres (in section 5.2.1, we will then investigate the dynamics of the lifting index in general cases.

5.3.1 Triggering in One-Dimensional Models

The movement of the lifting index of the system is denoted by a random variable $s(t)$ (t indicates the number of events in an infinite chain):

$$s(t) = \begin{cases} +1 & \text{if } +x \text{ is triggered,} \\ -1 & \text{if } -x \text{ is triggered.} \end{cases} \tag{5.3.1}$$

For example, in the one-dimensional hard-sphere model (without the factor field), $s(t)$ is always 1; in the harmonic model ($b = L/N$ in eq. (5.1.8)), we have $\langle s(t) \rangle_t = 0$.

As shown in Fig. 5.4, the autocorrelation of $s(t)$ of shows an algebraic decay with time in the one-dimensional models with the optimized factor field:

$$\langle s(\tau) s(t+\tau) \rangle_\tau \propto t^{-2/3}. \tag{5.3.2}$$

This relation is verified both in the one-dimensional Lennard-Jones model and in the one-dimensional hard spheres (see Fig. 5.4 (a) (b)).

[9] This figure is taken from our submitted manuscript in the Publication III [20].

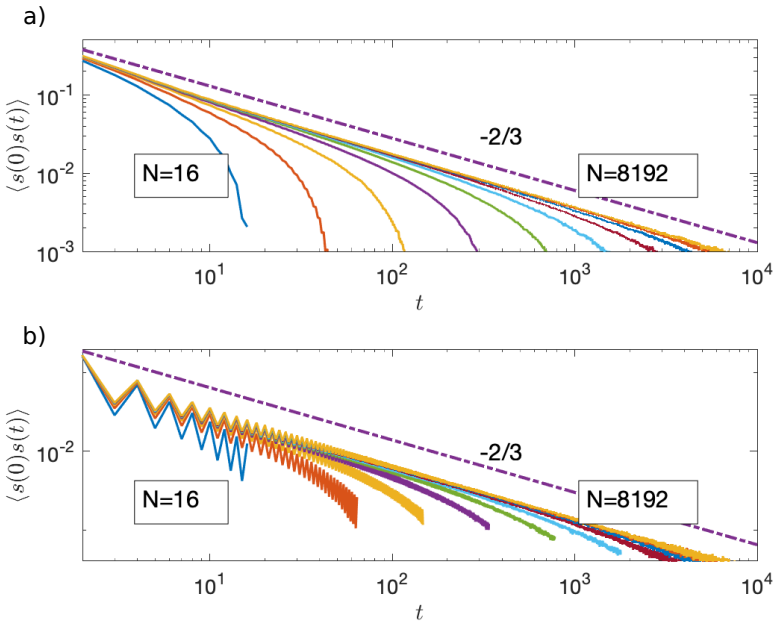


Figure 5.4: Autocorrelation functions of $s(t)$ in the event-chain Monte Carlo with the optimal factor field in one-dimensional models. (a): Lennard-Jones chain; (b): Hard-sphere chain. The autocorrelation decays algebraically with time ($C_s(t) \propto t^{-2/3}$).[9]

The total displacement of the lifting index is

$$S(t) = \sum_{t'=1}^{t} s(t').$$

More generally, if

$$\langle s(\tau)s(\tau+t)\rangle_\tau \propto t^{-\gamma},\ \gamma > 0 \text{ for } t \gg 1,$$

the mean-square displacement is

$$\left\langle S(t)^2 \right\rangle = \sum_{t'=1}^{t}\sum_{t''=1}^{t} \langle s(t')s(t'')\rangle \sim t^{2-\gamma}. \tag{5.3.3}$$

With the notation of Lévy walk in section 1.5.3, we find that the case $\gamma = 2/3$ corresponds to a super-diffusive process.

This super-diffusive characteristic is closely related to the fast relaxation of the event-chain algorithm. Similar to the "water-level" discussion in section 4.2.2, each particle needs to be activated at least once in order to fully decorrelate with its previous configuration. As $|S| \sim t^{1-\gamma/2}$, we need at least the time $T \sim N^{\frac{1}{1-\gamma/2}}$. By taking $\gamma = 2/3$, this conjecture implies a time scale of $N^{3/2}$, which agrees with the autocorrelation time of $\mathcal{O}(N^{3/2})$.

In another aspect, as we discussed in the "water-level" dynamics in section 4.2.2, the autocorrelation time ($\mathcal{O}(N^{3/2})$) is determined by the mean deviation of a particle

from its equilibrated position. The "half-system distance" in hard spheres plays the similar role, and its standard deviation is also proportional to $\sqrt{\tilde{L}}$ ($u \sim \sqrt{\tilde{L}}$ if $\tilde{L} \propto N$) as in the harmonic model, so the autocorrelation time is proportional to $N^{3/2}$.

A general argument of the instability of one-dimensional systems is made by R. Peierls in 1936 [107], that the mean deviation of a particle is always larger than $\mathcal{O}(1)$. Hence, the autocorrelation time of the event-chain dynamics in such systems is larger than $\mathcal{O}(N^1)$ single steps. (On the other hand, as the distortion is usually restricted by the system size, the autocorrelation hardly exceeds $\mathcal{O}(N^2)$.)

Return Probabilities

The number of times $n_l(t)$ that the original active particle is activated in t events is a random variable known as the *local time*, which is determined by the random walk of the lifting index. For example it follows a half-gaussian distribution for $n_l > 0$, with $\mathbb{E}(n_l) \propto \sqrt{t}$ in a Brownian walk [108][109].

In the event-chain dynamics with the optimal factor field, it is observed that the $\mathbb{E}(n_l) \propto t^{\gamma/2} = t^{1/3}$ as shown in Fig. 5.5, and the distribution is invariant under a rescaling (see the inset). This result agrees with our argument of the super-diffusion of the lifting index.

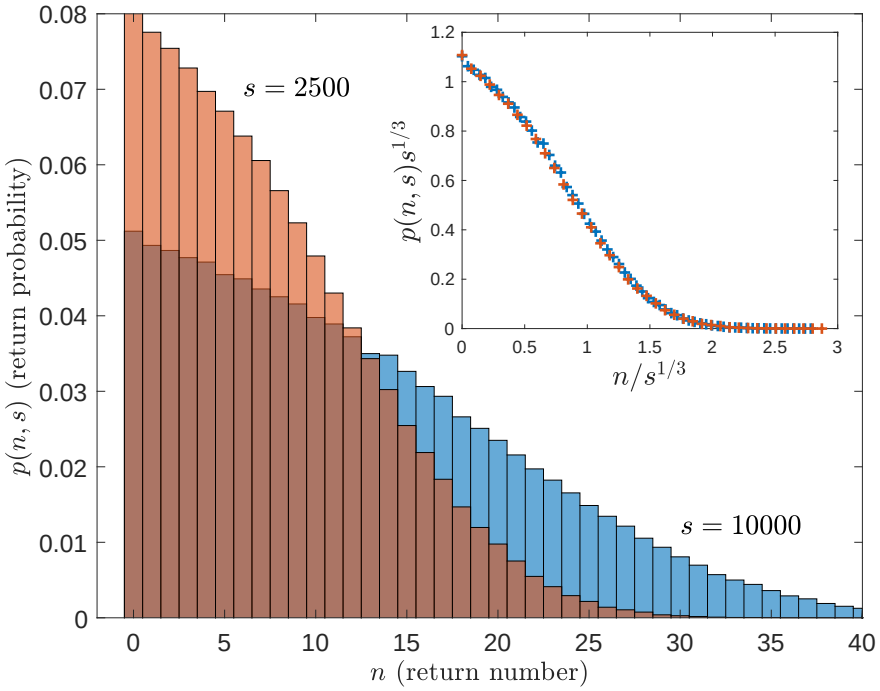


Figure 5.5: ECMC with factor field for a 1D Lennard-Jones system ($N = 8192$, $T/\epsilon = 1$). Probability $p(n;t)$ to return n times to the original active particle during t events (for $t = 2500$ and $t = 10000$). Mean and standard deviation of $p(n;t)$ both grow as $n^{\gamma/2} \sim t^{1/3}$ (see the inset). Inset: data collapse using scaling variables $p(n;t)t^{1/3}$ *vs.* $n/t^{1/3}$.[11]

5.3.2 Triggering in Two-Dimension Models

In the two-dimensional square lattice, the movement of the lifting index $s(t)$ takes 4 values. We assign the triggering in y directions with imaginary number:

$$s(t) = \begin{cases} \pm 1 & \text{if } \pm x \text{ is triggered,} \\ \pm i & \text{if } \pm y \text{ is triggered.} \end{cases} \tag{5.3.4}$$

The autocorrelation is a complex function, where the real part describes the random drifting, and the imaginary part indicates the turning/rotation. As shown in Fig. 5.6, the autocorrelation indicates $\gamma \sim 5/4$. Because $\langle S(t)S^*(t)\rangle$ has a lower bound when $\gamma > 1$ ($\langle s(t)s^*(t')\rangle\,|_{t'=t} = 1$), it will not spread slower than a diffusion:

$$\langle S(t)S^*(t)\rangle \sim t. \tag{5.3.5}$$

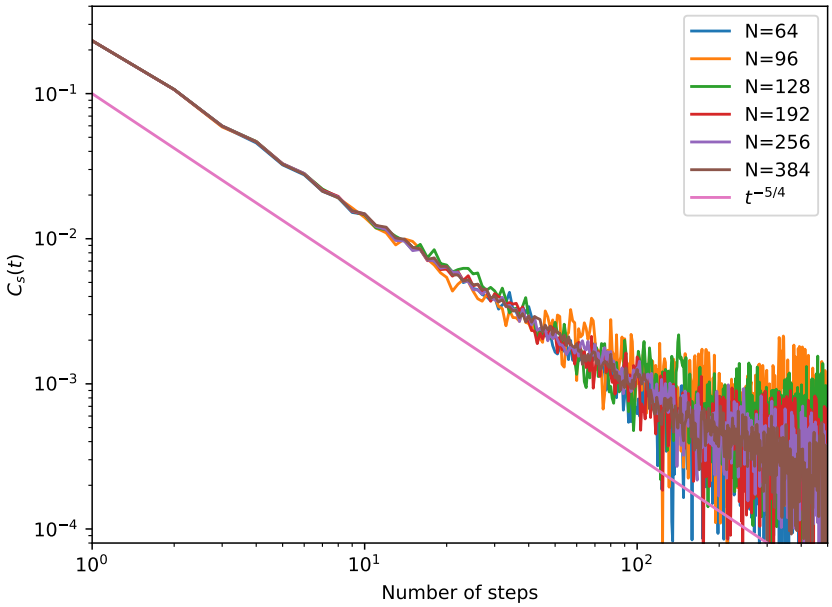


Figure 5.6: Autocorrelation function (absolute value) of $s(t)$ in two-dimensional harmonic model. We find $C_s(t) \propto t^{-5/4}$ for large t.

As $|S| \sim t^{1/2}$, at least the time $T \sim L^2 = N$ is needed for each site to be activated once. This argument is consistent with the autocorrelation time of $\mathcal{O}(N\sqrt{\log N})$ in Table 4.1 and section 4.2.2. The diffusive behavior of the lifting index agrees with the observation of the two-dimensional XY model below the critical temperature [106].

5.4 Conclusion

In this chapter, we combined the dynamics in chapters 3 and 4 to study one-dimensional systems as the Lennard-Jones model and again, the hard spheres. We analyzed their

[11] This figure is taken from our submitted manuscript in the Publication III [20].

differences and proposed a "factor field" which accelerates the relaxation of general systems in the event-chain dynamics. Furthermore, we attributed fast relaxation to the Lévy walk of the lifting index (of the active particle) in the event-chain dynamics, which is super-diffusive in dimension one and diffusive in dimension two. Its close relation to the critical phases [106] may provide ideas of critical dynamics and optimization of algorithms in higher dimensions.

General Conclusion

In the past 70 years, the Markov-chain Monte Carlo methods have been applied widely to many fields of studies. From the Metropolis-Hastings algorithm [4, 55] to the cluster algorithms [63], numerous methods based on the detailed balance were proposed with the goal of fast convergence. In the recent decades, with the "lifting scheme" [73], irreversible Markov chains began attracting attentions. Based on the global balance, they have proven to mix faster than the reversible Markov chains [82]. One of the irreversible Markov chains, the "event-chain" Monte Carlo algorithm, is applied in a wide range of physical systems, including hard spheres [9], particle systems [10, 16], spin models [11, 12], etc.

During my thesis, I began with the study of the event-chain algorithm two types of models, the hard-sphere systems [17] and spin systems [12]. They have very different dynamics in the mixing process from extreme conditions and in equilibrium: in the hard-sphere model, the event-chain dynamics strictly depends on the distribution of the chain lengths, and it has a mixing time of $\mathcal{O}(N^2 \log N)$ in one dimension; while in the spin-wave model, the dynamics has little dependence of the chain length, and exhibits an autocorrelation time of $\mathcal{O}(N^{3/2})$ in one dimension.

In the one-dimensional hard-sphere model, an exact calculation of the mixing time is related to the coupon-collector problem, based on randomized stopping times and "perfect sampling" on an interval. It provides theoretical estimations of the mixing times in some other cases, and leads to a general "swap" algorithm in both one and two dimensions.

In the two-dimensional XY model, a slow "vortex mode" is identified in comparison with the fast "spin-wave" mode. Following the Fréchet distribution, the maximum vortex–anti-vortex pair distance gives (what we believe) the slowest time scaling of the system. Other topological defects in the spin systems are also studied, although they play less of the role.

The fast dynamics of the harmonic (spin-wave) model provides a general method of optimization of the event-chain Monte Carlo by introducing an factor field, which considerably reduces the autocorrelation time in the one-dimensional systems such as Lennard-Jones model, and even hard spheres. The super-diffusive Lévy walk of the lifting index (the active particle) is related to the fast dynamics in one dimension.

Since the irreversible factorized paradigm was proposed, trials have been made [106] to explain its complex dynamical property [75]. This thesis succeeds in explaining the dynamics in several models and utilizing the special fast dynamics to optimize the others. However, there are some questions left open:

- The super-diffusion and oscillation dynamics in the event-chain algorithm with the optimal factor field may lead to certain universal groups of special dynamical scaling. To identify the groups, we need to find preciser formula as descriptions.
- The optimization using factor field in general systems of multiple dimensions is yet to study, when the "shuffle" of the active dimension is necessary. The optimal stopping rule will be explored.
- The connection between the diffusion of the lifting index and the fast relaxation in more general cases (e.g. in the multi-dimensional XY models [106]) needs a further clarification.
- A general optimization of the event-chain algorithm (or even a new algorithm) treating the topological defects (vortices and dislocations) without using nonlocal method is still a challenge.

We expect these problems would be solved in our future work.

Bibliography

[1] M. De Buffon; *Histoire naturelle générale et particulière*; (1753).

[2] J. von Neumann; *Various techniques used in connection with random digits*; in *The Monte Carlo Method,* Number 12 in National Bureau of Standards Applied Mathematics Series; pp. 36–38 (United States Government Printing Office, Washington, DC) (1951).

[3] N. Metropolis & S. Ulam; *The Monte Carlo Method*; Journal of the American statistical association **44**, pp. 335–341 (1949).

[4] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller & E. Teller; *Equation of State Calculations by Fast Computing Machines*; J. Chem. Phys. **21**, pp. 1087–1092 (1953) .

[5] W. Krauth; *Statistical Mechanics: Algorithms and Computations* (Oxford University Press, USA) (2006).

[6] B. L. Hammond, W. A. Lester & P. J. Reynolds; *Monte Carlo methods in ab initio quantum chemistry*; volume 1 (World Scientific) (1994).

[7] B. F. Manly; *Randomization, bootstrap and Monte Carlo methods in biology* (Chapman and Hall/CRC) (2006).

[8] C. Andrieu, N. de Freitas, A. Doucet & M. I. Jordan; *An Introduction to MCMC for Machine Learning*; Machine Learning **50**, pp. 5–43 (2003). ISSN 1573-0565.

[9] E. P. Bernard, W. Krauth & D. B. Wilson; *Event-chain Monte Carlo algorithms for hard-sphere systems*; Phys. Rev. E **80**, p. 056704 (2009) .

[10] M. Michel, S. C. Kapfer & W. Krauth; *Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps*; J. Chem. Phys. **140**, p. 054116 (2014) .

[11] M. Michel, J. Mayer & W. Krauth; *Event-chain Monte Carlo for classical continuous spin models*; EPL (Europhysics Letters) **112**, p. 20003 (2015) .

[12] Z. Lei & W. Krauth; *Irreversible Markov chains in spin models: Topological excitations*; EPL (Europhysics Letters) **121**, p. 10008 (2018) .

[13] E. P. Bernard & W. Krauth; *Two-Step Melting in Two Dimensions: First-Order Liquid-Hexatic Transition*; Phys. Rev. Lett. **107**, p. 155704 (2011) .

[14] M. Engel, J. A. Anderson, S. C. Glotzer, M. Isobe, E. P. Bernard & W. Krauth; *Hard-disk equation of state: First-order liquid-hexatic transition in two dimensions with three simulation methods*; Phys. Rev. E **87**, p. 042134 (2013) .

[15] S. C. Kapfer & W. Krauth; *Two-Dimensional Melting: From Liquid-Hexatic Coexistence to Continuous Transitions*; Phys. Rev. Lett. **114**, p. 035702 (2015) .

[16] M. F. Faulkner, L. Qin, A. C. Maggs & W. Krauth; *All-atom computations with irreversible Markov chains*; J. Chem. Phys. **149**, p. 064113 (2018) .

[17] Z. Lei & W. Krauth; *Mixing and perfect sampling in one-dimensional particle systems*; EPL (Europhysics Letters) **124**, p. 20003 (2018) .

[18] S. C. Kapfer & W. Krauth; *Irreversible Local Markov Chains with Rapid Convergence towards Equilibrium*; Phys. Rev. Lett. **119**, p. 240603 (2017) .

[19] P. Erdős & A. Rényi; *On a classical problem of probability theory*; (1961).

[20] Z. Lei, W. Krauth & A. C. Maggs; *Event-chain Monte Carlo with factor fields*; (2018); arXiv:1812.02494.

[21] A. Markov; *Extension of the law of large numbers to quantities, depending on each other (1906). Reprint.* Journal Électronique d'Histoire des Probabilités et de la Statistique [electronic only] **2**, pp. Article 10, 12 p., electronic only–Article 10, 12 p., electronic only (2006).

[22] A. Einstein; *Investigations on the Theory of the Brownian Movement* (Courier Corporation) (1956).

[23] A. N. Kolmogorov; *Foundations of the Theory of Probability: Second English Edition* (Courier Dover Publications) (2018).

[24] P.-S. Laplace; *Pierre-Simon Laplace Philosophical Essay on Probabilities: Translated from the fifth French edition of 1825 With Notes*; volume 13 (Springer Science & Business Media) (2012).

[25] A. N. Shiryaev; *Probability-1: Volume 1*; Graduate Texts in Mathematics 95; 3rd edition (Springer-Verlag New York) (2016); ISBN 978-0-387-72205-4, 978-0-387-72206-1.

[26] B. V. Gnedenko, A. N. Kolmogorov, K. L. Chung & J. L. Doob; *Limit Distributions for Sums of Independent Random Variables. Revised Edition* ; revised edition (Addison-Wesley) (1968).

[27] L. de Haan & A. Ferreira; *Extreme Value Theory: An Introduction* (Springer Science & Business Media) (2007).

[28] P. Lévy; *Sur certains processus stochastiques homogènes*; Compos. Math. **7**, pp. 283–339 (1939) ISSN 0010-437X; 1570-5846/e.

[29] A. Y. Khinchin; *Theory of correlation of stationary stochastic processes*; Uspekhi matematicheskikh nauk pp. 42–51 (1938).

[30] M. Fréchet; *Sur la loi de probabilite de l'écart maximum*; Ann. Soc. Math. Polon., 6, 93-116 (1927).

[31] R. A. Fisher & L. H. C. Tippett; *Limiting forms of the frequency distribution of the largest or smallest member of a sample*; Mathematical Proceedings of the Cambridge Philosophical Society **24**, pp. 180–190 (1928) .

[32] B. Gnedenko; *Sur la distribution limite du terme maximum d'une série aléatoire*; Annals of mathematics pp. 423–453 (1943) .

[33] G. Grimmett & D. Stirzaker; *Probability and random processes* (Oxford university press) (2001).

[34] E. Ising; *Beitrag zur theorie des ferromagnetismus*; Zeitschrift für Physik **31**, pp. 253–258 (1925) .

[35] C. D. Meyer; *Matrix analysis and applied linear algebra*; volume 71 (Siam) (2000).

[36] D. A. Levin, Y. Peres & E. L. Wilmer; *Markov Chains and Mixing Times* (American Mathematical Society) (2008).

[37] A. Sokal; *Monte Carlo methods in statistical mechanics: foundations and new algorithms*; in *Functional integration*, pp. 131–192 (Springer) (1997).

[38] D. Aldous & J. A. Fill; *Reversible markov chains and random walks on graphs, 2002. Unfinished monograph, recompiled 2014*; (2002).

[39] P. Diaconis; *The cutoff phenomenon in finite Markov chains*; Proceedings of the National Academy of Sciences **93**, pp. 1659–1664 (1996) .

[40] P. Flajolet, D. Gardy & L. Thimonier; *Birthday paradox, coupon collectors, caching algorithms and self-organizing search*; Discrete Applied Mathematics **39**, pp. 207 – 229 (1992). ISSN 0166-218X.

[41] D. Williams; *Probability with martingales* (Cambridge university press) (1991).

[42] N. Berestycki; *Lectures on mixing times* (Cambridge University) (2014).

[43] G. Pólya; *Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz*; Mathematische Annalen **84**, pp. 149–160 (1921).

[44] H. Lacoin; *The simple exclusion process on the circle has a diffusive cutoff window*; Ann. Inst. H. Poincaré Probab. Statist. **53**, pp. 1402–1437 (2017) .

[45] B. Morris; *The mixing time for simple exclusion*; Ann. Appl. Probab. **16**, pp. 615–635 (2006) .

[46] M. Gorissen, A. Lazarescu, K. Mallick & C. Vanderzande; *Exact Current Statistics of the Asymmetric Simple Exclusion Process with Open Boundaries*; Phys. Rev. Lett. **109**, p. 170601 (2012) .

[47] L.-H. Gwa & H. Spohn; *Six-vertex model, roughened surfaces, and an asymmetric spin Hamiltonian*; Phys. Rev. Lett. **68**, pp. 725–728 (1992) .

[48] A. Lazarescu & K. Mallick; *An exact formula for the statistics of the current in the TASEP with open boundaries*; J. Phys. A **44**, p. 315001 (2011) .

[49] T. Chou, K. Mallick & R. K. P. Zia; *Non-equilibrium statistical mechanics: from a paradigmatic model to biological transport*; Rep. Prog. Phys. **74**, p. 116601 (2011) .

[50] J. Baik & Z. Liu; *TASEP on a Ring in Sub-relaxation Time Scale*; J. Stat. Phys. **165**, pp. 1051–1085 (2016). ISSN 1572-9613.

[51] D. Randall & P. Winkler; *Mixing Points on an Interval*; in *Proceedings of the Seventh Workshop on Algorithm Engineering and Experiments and the Second Workshop on Analytic Algorithmics and Combinatorics, ALENEX /ANALCO 2005, Vancouver, BC, Canada, 22 January 2005,* pp. 218–221 (2005).

[52] D. Randall & P. Winkler; *Mixing Points on a Circle*; in *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques: 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th International Workshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22-24, 2005. Proceedings,* , edited by C. Chekuri, K. Jansen, J. D. P. Rolim & L. Trevisan; pp. 426–435 (Springer Berlin Heidelberg, Berlin, Heidelberg) (2005); ISBN 978-3-540-31874-3.

[53] L. Wasserman; *All of Statistics: A Concise Course in Statistical Inference*; Springer Texts in Statistics (Springer) (2003); ISBN 0387402721,9780387402727.

[54] J. W. Gibbs; *Elementary principles in statistical mechanics* (New York: Charles Scribner's sons; London: Edward Arnold) (1902).

[55] W. K. Hastings; *Monte Carlo sampling methods using Markov chains and their applications*; Biometrika **57**, pp. 97–109 (1970) .

[56] A. A. Barker; *Monte carlo calculations of the radial distribution functions for a proton-electron plasma*; Australian Journal of Physics **18**, pp. 119–134 (1965) .

[57] W. Janke; *Multicanonical Monte Carlo simulations*; Physica A: Statistical Mechanics and its Applications **254**, pp. 164 – 178 (1998). ISSN 0378-4371.

[58] P. C. Hohenberg & B. I. Halperin; *Theory of dynamic critical phenomena*; Rev. Mod. Phys. **49**, pp. 435–479 (1977) .

[59] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman & L. Troyansky; *Determining computational complexity from characteristic 'phase transitions'*; Nature **400**, p. 133 (1999) .

[60] L. Onsager; *Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition*; Phys. Rev. **65**, pp. 117–149 (1944) .

[61] R. H. Swendsen & J.-S. Wang; *Nonuniversal critical dynamics in Monte Carlo simulations*; Phys. Rev. Lett. **58**, pp. 86–88 (1987) .

[62] C. Fortuin & P. Kasteleyn; *On the random-cluster model: I. Introduction and relation to other models*; Physica **57**, pp. 536 – 564 (1972) .

[63] U. Wolff; *Collective Monte Carlo Updating for Spin Systems*; Phys. Rev. Lett. **62**, pp. 361–364 (1989) .

[64] J. Villain; *Theory of one-and two-dimensional magnets with an easy magnetization plane. II. The planar, classical, two-dimensional magnet*; Journal de Physique **36**, pp. 581–590 (1975) .

[65] W. Janke & K. Nather; *High-precision Monte Carlo study of the two-dimensional XY Villain model*; Phys. Rev. B **48**, pp. 7419–7433 (1993) .

[66] T. Obuchi & H. Kawamura; *Monte Carlo simulations of the three-dimensional XY spin glass focusing on chiral and spin order*; Phys. Rev. B **87**, p. 174438 (2013) .

[67] P. Kasteleyn; *The statistics of dimers on a lattice: I. The number of dimer arrangements on a quadratic lattice*; Physica **27**, pp. 1209 – 1225 (1961) .

[68] W. Krauth & R. Moessner; *Pocket Monte Carlo algorithm for classical doped dimer models*; Phys. Rev. B **67**, p. 064503 (2003) .

[69] C. Dress & W. Krauth; *Cluster algorithm for hard spheres and related systems*; J. Phys. A **28**, pp. L597–L601 (1995) .

[70] L. Santen & W. Krauth; *Absence of Thermodynamic Phase Transition in a Model Glass Former*; Nature **405**, pp. 550–1 (2000) .

[71] D. J. Earl & M. W. Deem; *Parallel tempering: Theory, applications, and new perspectives*; Phys. Chem. Chem. Phys. **7**, pp. 3910–3916 (2005) .

[72] L. Wang; *Discovering phase transitions with unsupervised learning*; Phys. Rev. B **94**, p. 195105 (2016) .

[73] P. Diaconis, S. Holmes & R. M. Neal; *Analysis of a nonreversible Markov chain sampler*; Ann. Appl. Probab. **10**, pp. 726–752 (2000) .

[74] C. J. O'Keeffe & G. Orkoulas; *Parallel canonical Monte Carlo simulations through sequential updating of particles*; J. Chem. Phys. **130**, p. 134109 (2009) .

[75] M. Michel; *Irreversible Markov chains by the factorized Metropolis filter: Algorithms and applications in particle systems and spin models*; Theses; Ecole Normale Superieure de Paris - ENS Paris (2016).

[76] K. S. Turitsyn, M. Chertkov & M. Vucelja; *Irreversible Monte Carlo algorithms for efficient sampling*; Physica D: Nonlinear Phenomena **240**, pp. 410 – 414 (2011). ISSN 0167-2789.

[77] E. Bernard; *Algorithms and applications of the Monte Carlo method: Two-dimensional melting and perfect sampling*; Theses; Université Pierre et Marie Curie - Paris VI (2011).

[78] E. A. J. F. Peters & G. de With; *Rejection-free Monte Carlo sampling for general potentials*; Phys. Rev. E **85**, p. 026703 (2012) .

[79] A. Bortz, M. Kalos & J. Lebowitz; *A new algorithm for Monte Carlo simulation of Ising spin systems*; Journal of Computational Physics **17**, pp. 10 – 18 (1975). ISSN 0021-9991.

[80] A. Sinclair & M. Jerrum; *Approximate counting, uniform generation and rapidly mixing Markov chains*; Information and Computation **82**, pp. 93 – 133 (1989). ISSN 0890-5401.

[81] A. Sinclair; *Improved Bounds for Mixing Rates of Markov Chains and Multicommodity Flow*; Combinatorics, Probability and Computing **1**, p. 351–370 (1992) .

[82] F. Chen, L. Lovász & I. Pak; *Lifting Markov Chains to Speed up Mixing*; Proceedings of the 17th Annual ACM Symposium on Theory of Computing p. 275 (1999) .

[83] M. J. Lighthill; *An introduction to Fourier analysis and generalised functions* (Cambridge University Press) (1958).

[84] B. J. Alder & T. E. Wainwright; *Phase Transition in Elastic Disks*; Phys. Rev. **127**, pp. 359–361 (1962) .

[85] N. D. Mermin; *Crystalline Order in Two Dimensions*; Phys. Rev. **176**, pp. 250–254 (1968) .

[86] J. M. Kosterlitz & D. J. Thouless; *Long range order and metastability in two dimensional solids and superfluids. (Application of dislocation theory)*; J. Phys. C **5**, pp. L124–L126 (1972) .

[87] B. I. Halperin & D. R. Nelson; *Theory of Two-Dimensional Melting*; Phys. Rev. Lett. **41**, pp. 121–124 (1978) .

[88] A. P. Young; *Melting and the vector Coulomb gas in two dimensions*; Phys. Rev. B **19**, pp. 1855–1866 (1979) .

[89] Y. Nishikawa, M. Michel, W. Krauth & K. Hukushima; *Event-chain algorithm for the Heisenberg model: Evidence for $z \simeq 1$ dynamic scaling*; Phys. Rev. E **92**, p. 063306 (2015) .

[90] C. Kittel & P. McEuen; *Introduction to solid state physics*; volume 8 (Wiley New York) (1996).

[91] J. M. Kosterlitz & D. J. Thouless; *Ordering, metastability and phase transitions in two-dimensional systems*; J. Phys. C **6**, pp. 1181–1203 (1973) .

[92] M. Hasenbusch; *The two-dimensional XY model at the transition temperature: a high-precision Monte Carlo study*; J. Phys. A **38**, pp. 5869–5883 (2005) .

[93] C. Holm & W. Janke; *Monte Carlo study of topological defects in the 3D Heisenberg model*; J. Phys. A **27**, pp. 2553–2563 (1994) .

[94] N. D. Mermin & H. Wagner; *Absence of Ferromagnetism or Antiferromagnetism in One- or Two-Dimensional Isotropic Heisenberg Models*; Phys. Rev. Lett. **17**, pp. 1133–1136 (1966) .

[95] M. Mézard, G. Parisi & M. Virasoro; *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*; volume 9 (World Scientific Publishing Company) (1987).

[96] J. M. Kosterlitz; *Phase Transitions in Long-Range Ferromagnetic Chains*; Phys. Rev. Lett. **37**, pp. 1577–1580 (1976) .

[97] F. Wegner; *Spin-ordering in a planar classical Heisenberg model*; Zeitschrift für Physik **206**, pp. 465–470 (1967) .

[98] J. Fröhlich & T. Spencer; *The Kosterlitz-Thouless transition in two-dimensional Abelian spin systems and the Coulomb gas*; **81**, pp. 527–602 (1981). ISSN 1432-0916.

[99] B. Jancovici; *Infinite Susceptibility Without Long-Range Order: The Two-Dimensional Harmonic "Solid"*; Phys. Rev. Lett. **19**, pp. 20–22 (1967) .

[100] M. Aizenman & B. Simon; *A comparison of plane rotor and Ising models*; Physics Letters A **76**, pp. 281 – 282 (1980). ISSN 0375-9601.

[101] C. H. Papadimitriou & K. Steiglitz; *Combinatorial Optimization: Algorithms and Complexity* (Prentice-Hall) (1982).

[102] B. Berg & M. Lüscher; *Definition and statistical distributions of a topological number in the lattice O(3) -model*; Nuclear Physics B **190**, pp. 412 – 424 (1981). ISSN 0550-3213.

[103] S. Ostlund; *Interactions between topological point singularities*; Phys. Rev. B **24**, pp. 485–488 (1981) .

[104] M. Lau & C. Dasgupta; *Role of topological defects in the phase transition of the three-dimensional Heisenberg model*; J. Phys. A **21**, pp. L51–L57 (1988) .

[105] L. Tonks; *The Complete Equation of State of One, Two and Three-Dimensional Gases of Hard Elastic Spheres*; Phys. Rev. **50**, p. 955 (1936) .

[106] K. Kimura & S. Higuchi; *Anomalous diffusion analysis of the lifting events in the event-chain Monte Carlo for the classical XY models*; EPL (Europhysics Letters) **120**, p. 30003 (2017) .

[107] R. Peierls; *On Ising's model of ferromagnetism*; Mathematical Proceedings of the Cambridge Philosophical Society **32**, pp. 477–481 (1936). ISSN 1469-8064.

[108] C. Banderier & M. Wallner; *Local time for lattice paths and the associated limit laws*; in *Proceedings of the 11th International Conference on Random and Exhaustive Generation of Combinatorial Structures (GASCom 2018), Athens, Greece, June 18-20, 2018. CEUR Workshop Proceedings,* , volume 2113, edited by L. Ferrari & M. Vamvakari; pp. 69–78 (CEUR Workshop Proceedings) (2018).

[109] W. Feller; *An introduction to probability theory and its applications. Vol. I*; Third edition (John Wiley & Sons Inc., New York) (1968).

Publications

Publication I: Irreversible Markov chains in spin models: Topological excitations

Publication I

Ze Lei and Werner Krauth
Europhysics Letters **121**, p. 10008 (2018)

This article studies the event-chain dynamics in spin models, specially the 2D XY model, and identifies a fast converging spin-wave mode and a slow vortex mode. Further discussions based on the distribution of the vortex and points out that the influence of vortex will be negligible at low temperature. (See chapter 4.)

A Letters Journal Exploring the Frontiers of Physics

January 2018

EPL, **121** (2018) 10008
doi: 10.1209/0295-5075/121/10008

www.epljournal.org

Irreversible Markov chains in spin models: Topological excitations

Ze Lei[1(a)] and Werner Krauth[1,2(b)]

[1] *Laboratoire de Physique Statistique, Département de physique de l'ENS, Ecole Normale Supérieure, PSL Research University, Université Paris Diderot, Sorbonne Paris Cité, Sorbonne Universités, UPMC Univ. Paris 06, CNRS 75005 Paris, France*

[2] *Department of Physics, Graduate School of Science, The University of Tokyo - 7-3-1 Hongo, Bunkyo, Tokyo, Japan*

received 24 November 2017; accepted in final form 21 February 2018
published online 9 March 2018

PACS 02.70.Tt – Justifications or modifications of Monte Carlo methods
PACS 75.10.Hk – Classical spin models
PACS 02.50.Ng – Distribution theory and Monte Carlo studies

Abstract – We analyze the convergence of the irreversible event-chain Monte Carlo algorithm for continuous spin models in the presence of topological excitations. In the two-dimensional XY model, we show that the local nature of the Markov-chain dynamics leads to slow decay of vortex-antivortex correlations while spin waves decorrelate very quickly. Using a Fréchet description of the maximum vortex-antivortex distance, we quantify the contributions of topological excitations to the equilibrium correlations, and show that they vary from a dynamical critical exponent $z \sim 2$ at the critical temperature to $z \sim 0$ in the limit of zero temperature. We confirm the event-chain algorithm's fast relaxation (corresponding to $z = 0$) of spin waves in the harmonic approximation to the XY model. Mixing times (describing the approach towards equilibrium from the least favorable initial state) however remain much larger than equilibrium correlation times at low temperatures. We also describe the respective influence of topological monopole-antimonopole excitations and of spin waves on the event-chain dynamics in the three-dimensional Heisenberg model.

Introduction. – Classical spin models have played a crucial role in the theory of critical phenomena and in the formulation of topological phases and their associated transitions. The analysis of vortices and their interactions in the two-dimensional XY model has led, in particular, to the development of the Kosterlitz-Thouless theory [1], which initiated the era of topology in condensed-matter physics. Likewise, spin models have been instrumental in the continued development of the Markov-chain Monte Carlo method, and especially in the invention of advanced sampling methods. Cluster Monte Carlo algorithms [2] were of prime importance to show that Kosterlitz-Thouless theory actually applied to the phase transition in the two-dimensional XY-model [3]. Monte Carlo methods also elucidated the role of topological excitations in other models, such as the three-dimensional Heisenberg model [4].

In recent years, irreversible Monte Carlo algorithms have increasingly come into focus. In these methods, the asymptotic steady state (reached in the long-time limit) still corresponds to thermodynamic equilibrium, but it is realized with non-zero probability flows. The event-chain Monte Carlo algorithm [5,6], in particular, implements the global balance condition in a maximally asymmetric way. It relies on the concept of lifted Markov chains [7]. Besides short-range and long-range particle systems [8–10], the event-chain algorithm applies to continuous spin models such as the 2D and 3D XY model [11,12] and the 3D Heisenberg model [13]. Improved convergence time scales were generally observed.

In this paper, we discuss the influence of topological excitations and of spin waves on the convergence of the event-chain algorithm, mostly concentrating on the two-dimensional XY model with its energy

$$E = -J \sum_{\langle i,j \rangle} \boldsymbol{S}_i \cdot \boldsymbol{S}_j, \tag{1}$$

with two-dimensional unit spins $\boldsymbol{S}_k = (S_k^x, S_k^y) = (\cos\phi_k, \sin\phi_k)$ on a square lattice with $N = L \times L$ sites. In eq. (1), the bracket $\langle , \rangle$ denotes nearest neighbors. A rotation-invariant observable in this model is the

(a) E-mail: ze.lei@ens.fr
(b) E-mail: werner.krauth@ens.fr

susceptibility:

$$\chi = \left(\sum_i \boldsymbol{S}_i \right)^2 . \tag{2}$$

Its autocorrelation function,

$$C_\chi(\tau) = \frac{\langle \chi(t+\tau)\chi(t) \rangle_t}{\langle \chi^2(t) \rangle_t}, \tag{3}$$

characterizes the time evolution of the system.

For the XY model, the event-chain algorithm (see [11]) rotates a given active spin $\boldsymbol{S}_i$ in positive sense in a sequence of infinitesimal moves until further rotation is vetoed through the factorized Metropolis algorithm [11]. More, precisely, the motion of the active spin $\boldsymbol{S}_i$ is influenced by all its neighbors $\boldsymbol{S}_j$:

- For each of the neighboring j, a single random number allows one to determine $\theta_{i,j}$, the hypothetical angle of rotation of $\boldsymbol{S}_i$ before it would get vetoed by spin $\boldsymbol{S}_j$.
- $k = \arg\min_{\{j\}} \{\theta_{i,j}\}$ determines the total rotation of $\boldsymbol{S}_i$, namely $\theta_{i,k}$, and the next active spin, namely $\boldsymbol{S}_k$.

(See ref. [11] for a full discussion, and for proof that this algorithm converges towards the Boltzmann distribution, although it only rotates spins in positive sense.) The event-chain algorithm violates the detailed-balance condition, but respects global balance. The latter is necessary to ensure convergence towards the equilibrium Boltzmann distribution. We also consider the harmonic approximation of the XY model [14], where in the energy of eq. (1) each term $\boldsymbol{S}_i \cdot \boldsymbol{S}_j = \cos(\phi_i - \phi_j)$ is approximated by $1 - \frac{1}{2}(\phi_i - \phi_j)^2$, and, finally, the three-dimensional Heisenberg model, where the spins $\boldsymbol{S}_i$ are three-dimensional unit vectors. The XY model features vortex excitations and it is the unbinding of vortex-antivortex pairs which take place at the critical temperature $T_c = 0.893J$. Below the critical temperature, however, the large-scale excitations of the XY model are spin waves. We will argue that the two-stage susceptibility autocorrelation at low temperature (see fig. 1(a)) corresponds in fact to the fast decay of spin waves under event-chain dynamics and to the slow decay of the vortex-antivortex pairs. For $T/T_c \to 0$, where vortices are tightly bound, the event-chain algorithm is asymptotically fast ($z \approx 0$), as we corroborate by simulations. However, the equilibrium correlations do not give the complete picture of the time behavior of the Markov chain under consideration. Indeed, one may study the relaxation to equilibrium after a quench from another temperature (typically from $T = \infty$ to $T < T_c$). Here, a wide spectrum of relaxation times becomes relevant, and equilibration can take much longer than the equilibrium correlation time τ (see fig. 1(b)). The quench dynamics is sensitive to the mixing time, which quantifies the approach towards equilibrium from the most unfavorable initial configuration [15]. Although the equilibrium correlations are described by a dynamical critical exponent $z \sim 0$ as $T/T_c \to 0$, we will argue that the mixing time remains at $z \sim 2$.

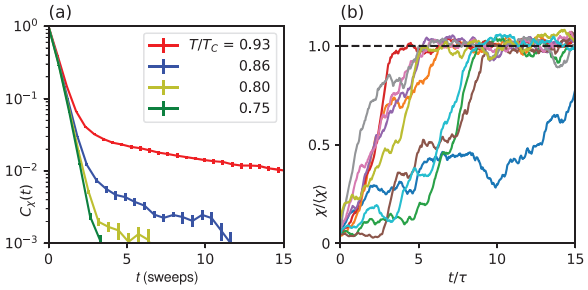


Fig. 1: (Color online) Time evolution of the spin susceptibility in the XY model on a 64×64 square lattice (time t measured in sweeps). Left: susceptibility autocorrelations below T_c. Right: individual susceptibility evolutions at $T/T_c = 0.93$ starting from random initial configurations (equilibrium autocorrelation time τ). Large sample-to-sample fluctuations are apparent.

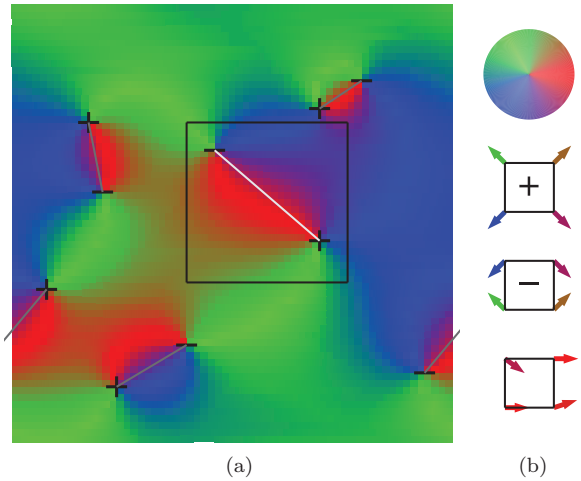


Fig. 2: (Color online) Vortices in the XY model. Left: configuration with 5 vortices ("+") and 5 antivortices ("−"). The lines indicate matched vortex-antivortex pairs in the optimal assignment (see eq. (4)), and the length of the longest line (shown in white) equals the max-distance. A subsystem containing the max-distance pair is highlighted. Right, from above: color code for the spin orientations, vortex, antivortex, and neutral plaquette configuration.

Vortex-antivortex pairs, max-distances. – For the XY model on a square lattice, vortices or antivortices, located on plaquettes delimited by four spins, are signalled by differences of neighboring spins that do not sum to zero when going around the plaquette in positive sense, but rather to 2π (vortex) or to -2π (antivortex) (see fig. 2(b)). With periodic boundary conditions, vortices and antivortices appear in pairs. In a configuration with n such pairs, the vortices $(v_1, v_2, \ldots, v_n)$ can be paired up with the antivortices $(a_{P_1}, a_{P_2}, \ldots, a_{P_n})$ according to one of the $n!$ permutations P. We suppose that the physically relevant pairing corresponds to the minimum of the Kosterlitz-Thouless vortex-antivortex-pair energy $\pi J_R \log(R/a) + 2E_c$ (we use a lattice parameter $a = 1$ in

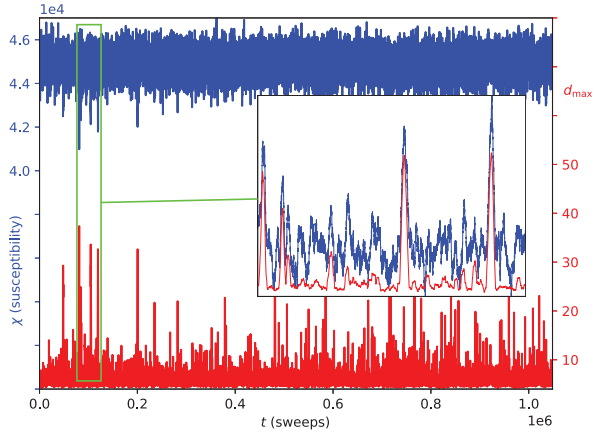


Fig. 3: (Color online) Time evolution of the vortex-antivortex max-distance in the 384×384 XY model at $T/T_c = 0.933$ compared to that of the susceptibility (smoothed over a small time window). The inset illustrates the correlation between max-distance and susceptibility in greater detail.

the later discussion), where the core energy E_c is the same for all configurations of n pairs and where the value of the renormalized stiffness J_R of Kosterlitz-Thouless theory does not influence the minimum [1]. We thus neglect interactions between different vortex-antivortex pairs. The proper association of each vortex v_i with its antivortex a_{P_i} defines an assignment problem (see fig. 2(a)) aimed at minimizing the objective function ϵ:

$$\epsilon(\{v_i, a_{P_i}\}) = \sum_{i=1}^{n} \log |\boldsymbol{R}_{v_i} - \boldsymbol{R}_{a_{P_i}}|. \tag{4}$$

The optimal assignment of the n vortex-antivortex pairs can be determined with a standard algorithm [16], known as the "Hungarian method" for weighted matching problems. This $\mathcal{O}(n^3)$ method recursively reduces the size of a cost array and identifies a pairing of minimal ϵ. In the optimal assignment, the pair (v_i, a_{P_i}) of largest separation defines the configuration's *max-distance* $d_{\max}$. Remarkably, the time evolution of the max-distance during a computation mimics that of the susceptibility (see fig. 3). Large vortex-antivortex pairs (indicated by $d_{\max} \gg a$) and small susceptibilities are particularly well correlated, and both persist on long time scales (see inset of fig. 3).

We suggest that at low temperature the max-distance length scale determines the relaxation time scale. To show this, we prepare initial configurations with only two vortex pairs arranged in a square of length $d_{\max} = d$ (such a configuration can be constructed with periodic boundary conditions). We then track the time needed for the susceptibility to reach the equilibrium value (within a few percent). At temperature $T \sim T_c$, the system quickly generates many vortices that screen the distribution of the initial scale. In contrast, at low temperature, vortex-antivortex pairs at distance d must approach each other before they can be annihilated. Indeed, we find that the time to converge the square-shaped configuration of fixed d is independent of the system size L, and proportional to $\mathcal{O}(d^2)$. Taking $d = \mathcal{O}(L)$, this implies that the mixing time τ_{mix} (the time to reach equilibrium from the most unfavorable initial condition [15]), is at least $\mathcal{O}(L^2)$.

For $L \to \infty$, the probability to have a vortex-antivortex pair spaced by $\boldsymbol{d}$ is

$$\begin{aligned} P(\boldsymbol{d}) &= \frac{1}{Z} \mathrm{e}^{-\beta E_p(d)} \\ &= \frac{1}{Z} \mathrm{e}^{-2\beta E_c} (d/a)^{-\pi\beta J_R} \\ &\propto d^{-\pi\beta J_R}, \end{aligned} \tag{5}$$

where E_p is the pair energy of Kosterlitz-Thouless theory [1]. Because of eq. (5), the distribution of the max-distance for n vortices must be polynomial for $d_{\max} \to \infty$. For $T/T_c \to 0$, the power-law exponent must diverge as the vortex-antivortex pairs are more and more tightly bound.

Fréchet distribution, vortex max-distance. – At temperatures below T_c, for $L \to \infty$, vortex-antivortex pairs are bound [1], so that the equilibrium max-distance $d_{\max}$ is much smaller than the system size, and its probability distribution $p(d_{\max})$ decays algebraically for large arguments (see eq. (5)). The $L \times L$ system can be divided into n^2 practically independent subsystems of size $L/n \times L/n$. The max-distance of the large system at scale L is the maximum of n^2 independent max-distances on a scale L/n. Extreme-value statistics [17] allows one to connect the distribution $p(d_{\max})$ at scale L with the one at L/n. It must correspond to the Fréchet distribution (with zero minimum value), specifically:

$$p(d_{\max}) = \frac{\alpha}{s} \left(\frac{d_{\max}}{s}\right)^{-1-\alpha} \exp\left[-\left(\frac{d_{\max}}{s}\right)^{-\alpha}\right] \tag{6}$$

with its cumulative distribution

$$P(d_{\max}) = \exp\left[-\left(\frac{d_{\max}}{s}\right)^{-\alpha}\right]. \tag{7}$$

Here, α describes the power-law decay of the max-distance distribution for large arguments (which is the same on scales L and L/n), and s sets its L-dependent scale. The maximum of N independent samples of a Fréchet distribution with parameters (α, s) is distributed following a Fréchet distribution with parameters $(\alpha, N^{1/\alpha} s)$. It then follows that the Fréchet distribution of the max-distance in a system of size L must be described by parameters $(\alpha, L^{2/\alpha} s_0)$, where both α and s_0 depend on β, but not on L, for large L. Slightly below T_c already, the Fréchet distribution provides an excellent fit for the max-distance distribution and the fitting parameters α and s_0 are indeed independent of L for a given temperature (see fig. 4). Also, we note that for $\alpha = 2$, the distribution of $d_{\max}$ scales with $s \propto L$. This is observed for $T/T_c \to 1^-$. At low temperatures, we observe $\alpha \propto 1/T$ (see fig. 5) in agreement with eq. (5).

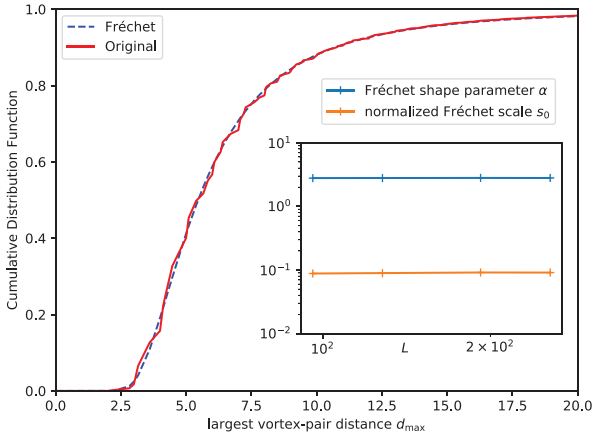


Fig. 4: (Color online) Cumulative max-distance distribution in the XY model at $L = 256$ and $T/T_c = 0.965$ fitted with the Fréchet distribution with $\alpha = 2.80$ and $s = 4.79$ ($s_0 = 0.0913$). The inset illustrates that the fitting parameters α and s_0 are independent of L.

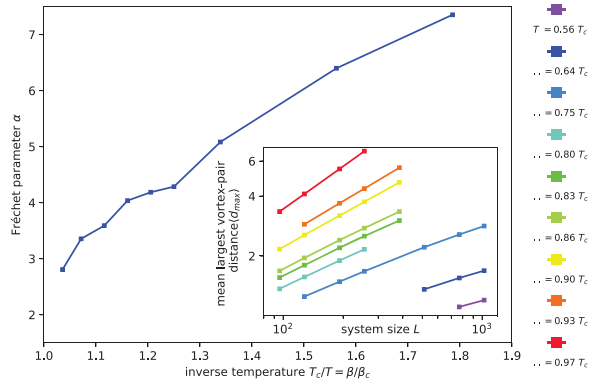


Fig. 5: (Color online) Scaling of the Fréchet parameter α with inverse temperature β demonstrating that $\alpha \to \infty$ for $\beta \to \infty$ (Our theoretical model indicates $\alpha \propto \beta$ at low temperature). The inset shows $\langle d_{\max} \rangle$ *vs.* system size L, from which α is obtained.

Below T_c, $d_{\max}$ scales as $\sim L^{2/\alpha} \ll L$ as $\alpha > 2$ (which means $d_{\max}/L^{2/\alpha}$ is an L-independent quantity) and we expect the equilibrium correlation time to scale with $s^2 = L^{4/\alpha} s_0^2$:

$$\tau_{\text{corr}}^{\text{vortex}} \sim L^{4/\alpha} \sim \begin{cases} L^2, & \text{for } T \to T_c^-, \\ L^{\text{const}T}, & \text{for } T \to 0. \end{cases} \tag{8}$$

The effective dynamical scaling parameter $z = 4/\alpha$ of the event-chain algorithm is thus connected to the scale parameter of a Fréchet distribution and it is predicted to vanish in the zero-temperature limit.

This relation between the scaling factor z and the Fréchet parameter α in eq. (8) is consistent with the scaling of the autocorrelation of the susceptibility below the critical temperature, as in fig. 6.

Harmonic model, spin waves. – The ansatz of eq. (8) for the equilibrium correlations only describes the

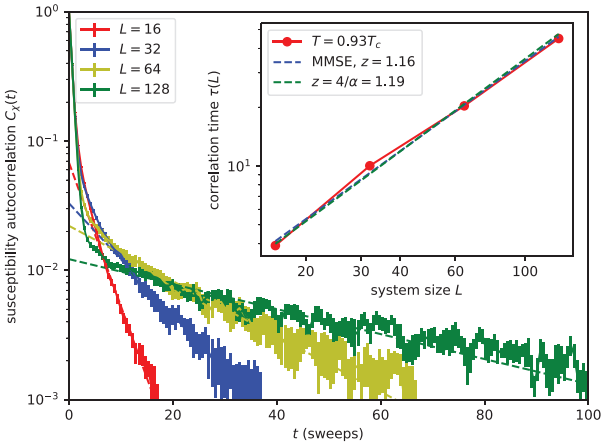


Fig. 6: (Color online) Susceptibility autocorrelation below the critical temperature ($T = 0.93T_c$) for different system sizes in the XY model. The inset shows their correlation time, compared with the scaling z obtained by Minimum Mean Square Error method and the fitting using $z(T) = 4/\alpha(T)$ from fig. 5. Note the agreement between the Fréchet theory of the vortex distribution and the relaxation of the susceptibility.

relaxation of topological excitations, parametrized by the max-distance. We now consider spin waves which, below T_c, are the dominant large-distance excitations for local Monte Carlo dynamics, where they take $\mathcal{O}(L^2)$ sweeps to relax. In the event-chain algorithm, they relax in $\mathcal{O}(L^0)$ sweeps, so that our ansatz is indeed consistent. To show this, we study the harmonic model, an approximation to the XY Hamiltonian, where the spin variables ϕ interact as follows:

$$E = \frac{J}{2} \sum_{\langle i,j \rangle} (\phi_i - \phi_j)^2. \tag{9}$$

This model is exactly solved by taking Fourier modes as the independent variables [14]. The two-dimensional harmonic model has algebraically decaying spin correlations with an exponent that approaches zero as $T/T_c \to 0$. From the exact solution of the harmonic model, it follows that the difference of ϕ on sites distant by $\mathcal{O}(L)$ is on a scale

$$\sigma^{\text{eq}}(L) \propto \begin{cases} \sqrt{L}, & \text{if } d = 1, \\ \sqrt{\log L}, & \text{if } d = 2, \\ 1, & \text{if } d \geq 3. \end{cases}$$

The event-chain algorithm for the harmonic model can only increase the value of ϕ_i. We find that in one sweep ($\mathcal{O}(N)$ events), the mean value $\langle \phi_i \rangle$ of a configuration increases by $\mathcal{O}(1)$. The correlation time of the algorithm is reached when the mean increase per site is on the order of the equilibrium correlation σ^{eq}. This implies the relation

$$\tau_{\text{corr}}^{\text{harm}} \sim \sigma^{\text{eq}}(L). \tag{10}$$

Equation (10) predicts a dynamical scaling exponent of 1/2 for the 1D harmonic model, and an exponent $z = 0$ in higher dimensions. This fast dynamical scaling, in sharp

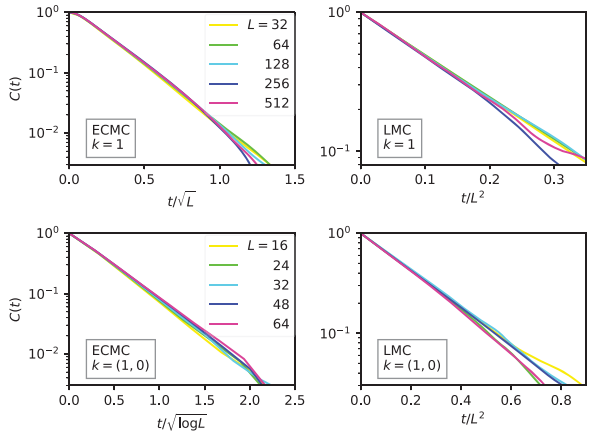


Fig. 7: (Color online) Equilibrium auto-correlation functions $C(t)$ of the lowest Fourier modes in the harmonic model for the event-chain algorithm (ECMC) and for local Monte Carlo (LMC). Top: $C(t)$ for the Fourier mode $k = 1$ in 1D. Bottom: $C(t)$ for the Fourier mode $k = (1, 0)$ in 2D. Data are in agreement with the scaling of eq. (10).

contrast to the behavior of the local Metropolis algorithm (with $z \sim 2$) is verified for the autocorrelation times for Fourier modes with small $\mathbf{k}$ (see fig. 7).

In the XY model below T_c, the two types of excitations generate two time scales for the equilibrium autocorrelation function of the event-chain algorithm. This corresponds to what is observed in the susceptibility, where we thus associate the fast initial decay with spin waves ($\tau_{\text{corr}}^{\text{harm}} \sim \text{const}$), and the slow decay with topological excitations (vortex–antivortex pairs, $\tau_{\text{corr}}^{\text{vortex}} \sim L^{\text{const}T}$ at low temperature and $\tau_{\text{corr}}^{\text{vortex}} \sim L^2$ for $T/T_c \to 1^-$) (see fig. 1).

Monopoles, Bloch waves. – Topological excitations also play a prominent role in other spin models, for example the 3D Heisenberg model. Low-temperature excitations in that model can also be described by spin waves in addition to topological excitations. In the event-chain algorithm, spin waves again come with a dynamical critical exponent ~ 0. Heisenberg-model monopoles and antimonopoles are again located on the dual lattice, and they can be identified using a well-defined algorithm [18].

In the 3D Heisenberg model, monopole-antimonopole pairs proliferate near the critical point [4]. Their excitation energy increases with the separation d as $\mathcal{O}(d)$ [19,20]. The event-chain algorithm, at low temperature, again moves each monopole or antimonopole by $\mathcal{O}(1)$ per sweep. From initial configurations with pairs separated on a scale $\mathcal{O}(L)$, we find that relaxation towards equilibrium takes $\mathcal{O}(L)$ sweeps (rather than $\mathcal{O}(L^2)$, as for the XY model). Configurations with widely separated pairs play no role at low temperature, and the spin waves are again treated efficiently in the event-chain algorithm, so that $z \to 0$ for $T/T_c \to 0$. Nevertheless, the mixing time scale for the approach to equilibrium from an unfavorable configuration is $\mathcal{O}(L)$ sweeps.

Finally, there are other types of topological excitations, besides the point-like ones (vortices, monopoles) discussed here. Bloch modes, in the XY model with periodic boundary conditions, correspond for example to a state in which the spins rotate by 2π as one coordinate, say x, goes from 0 to L. Bloch waves are a slow mode in the event-chain algorithm for the XY model (but not in the Heisenberg model), and stable on a time scale $\mathcal{O}(L^2)$ at low temperature in both 2D and 3D.

Conclusions. – In this paper, we exhibited a considerable speed-up for the relaxation of spin-wave excitations of the event-chain algorithm compared to the local Monte Carlo algorithm. Indeed, in the harmonic model, which has only spin waves, the event-chain algorithm equilibrates in a constant number of sweeps for $d > 1$, whereas the local algorithm equilibrates with $z \sim 2$. We have also studied the relaxation of topological excitations, namely the vortex-antivortex pairs in the 2D XY model and the monopole-antimonopole pairs in the Heisenberg model. In the XY model, below the critical temperature, vortex-antivortex pairs are bound, and we parametrize this binding with a single parameter, the max-distance $d_{\max}$ that can be computed with a combinatorial-optimization algorithm. We find that the probability distribution of $d_{\max}$ is a Fréchet distribution (with zero minimum value). In the XY model, the vortex-antivortex potential is very weak, leading to a $d_{\max}^2$ relaxation time and, at worst, an L^2 mixing time. However, equilibrium-correlation time scales are much smaller. In the event-chain algorithm, these vortex-antivortex excitations are no longer concealed by the spin waves, and they in fact constitute the slowest dynamical modes. It is thus found to have a smaller dynamical exponent than the local Monte Carlo algorithm for all temperatures below T_c. In particle systems, we likewise expect the fast relaxation of phonon modes (which, in analogy to the spin waves of this paper, are also described by a harmonic model) to be key to the success of the event-chain algorithm at high densities [8]. However, the fundamental difference between mixing times (needed to reach equilibrium from the most unfavorable initial condition) and equilibrium correlation times (needed to move to a new independent configuration from an equilibrium starting configuration) appears clearly [15]. It will certainly have to be taken into account in applications.

* * *

We thank Youjin Deng for helpful discussions in the initial stages of this work and Cris Moore for useful suggestions.

REFERENCES

[1] Kosterlitz J. M. and Thouless D. J., *J. Phys. C*, **6** (1973) 1181.
[2] Wolff U., *Phys. Rev. Lett.*, **62** (1989) 361.
[3] Hasenbusch M., *J. Phys. A*, **38** (2005) 5869.

[4] Holm C. and Janke W., *J. Phys. A*, **27** (1994) 2553.
[5] Bernard E. P., Krauth W. and Wilson D. B., *Phys. Rev. E*, **80** (2009) 056704.
[6] Michel M., Kapfer S. C. and Krauth W., *J. Chem. Phys.*, **140** (2014) 054116.
[7] Diaconis P., Holmes S. and Neal R. M., *Ann. Appl. Probab.*, **10** (2000) 726.
[8] Bernard E. P. and Krauth W., *Phys. Rev. Lett.*, **107** (2011) 155704.
[9] Kapfer S. C. and Krauth W., *Phys. Rev. Lett.*, **114** (2015) 035702.
[10] Harland J., Michel M., Kampmann T. A. and Kierfeld J., *EPL*, **117** (2017) 30001.
[11] Michel M., Mayer J. and Krauth W., *EPL*, **112** (2015) 20003.
[12] Kimura K. and Higuchi S., *EPL*, **120** (2017) 30003.
[13] Nishikawa Y., Michel M., Krauth W. and Hukushima K., *Phys. Rev. E*, **92** (2015) 063306.
[14] Wegner F., *Z. Phys.*, **206** (1967) 465.
[15] Levin D. A., Peres Y. and Wilmer E. L., *Markov Chains and Mixing Times* (American Mathematical Society) 2008.
[16] Papadimitriou C. H. and Steiglitz K., *Combinatorial Optimization: Algorithms and Complexity* (Prentice-Hall) 1982.
[17] De Haan L. and Ferreira A., *Extreme Value Theory: An Introduction* (Springer Science & Business Media) 2007.
[18] Berg B. and Lüscher M., *Nucl. Phys. B*, **190** (1981) 412.
[19] Lau M. and Dasgupta C., *J. Phys. A*, **21** (1988) L51.
[20] Ostlund S., *Phys. Rev. B*, **24** (1981) 485.

Publication II: Mixing and perfect sampling in one-dimensional particle systems

Publication II

Ze Lei and Werner Krauth
Europhysics Letters **124**, p. 20003 (2018)

This article analytically evaluates the mixing time of the event-chain algorithm in one-dimensional hard-sphere model, and relate it to a coupon-collector problem. It also confirms a former conjecture of the dynamics of the lifted Metropolis algorithm [18]. Furthermore, a "swap" modification of the event-chain algorithm and the lifted Metropolis algorithm is proposed, which helps overcome the limit of coupon-collector. This algorithm is also valid in higher dimensions. (See chapter 3.)

epl A Letters Journal Exploring the Frontiers of Physics

October 2018

EPL, **124** (2018) 20003 www.epljournal.org
doi: 10.1209/0295-5075/124/20003

Mixing and perfect sampling in one-dimensional particle systems

Ze Lei[a] and Werner Krauth[b]

Laboratoire de Physique Statistique, Département de Physique de l'ENS, Ecole Normale Supérieure, PSL Research University, Université Paris Diderot, Sorbonne Paris Cité, Sorbonne Universités, UPMC Univ. Paris 06, CNRS - 75005 Paris, France, and
Max-Planck-Institut für Physik komplexer Systeme - Nöthnitzer Str. 38, 01187 Dresden, Germany

received 18 June 2018; accepted in final form 12 October 2018
published online 19 November 2018

PACS 02.70.Tt – Justifications or modifications of Monte Carlo methods
PACS 02.50.Ng – Distribution theory and Monte Carlo studies

Abstract – We study the approach to equilibrium of the event-chain Monte Carlo (ECMC) algorithm for the one-dimensional hard-sphere model. Using the connection to the coupon-collector problem, we prove that a specific version of this local irreversible Markov chain realizes perfect sampling in $\mathcal{O}(N^2 \log N)$ single steps, whereas the reversible local Metropolis algorithm requires $\mathcal{O}(N^3 \log N)$ single steps for mixing. This confirms a special case of an earlier conjecture about $\mathcal{O}(N^2 \log N)$ scaling of mixing times of ECMC and of the lifted forward Metropolis algorithm, its discretized variant. We also prove that sequential ECMC (with swaps) realizes perfect sampling in $\mathcal{O}(N^2)$ single events. Numerical simulations indicate a cross-over towards $\mathcal{O}(N^2 \log N)$ mixing for the sequential forward swap Metropolis algorithm, that we introduce here. We point out open mathematical questions and possible applications of our findings to higher-dimensional models.

Sampling, mixing, perfect sampling, stopping rules. – Ever since the 1950s [1], Markov-chain Monte Carlo (MCMC) methods have ranked among the most versatile approaches in scientific computing. Monte Carlo algorithms strive to sample a probability distribution π. For an N-particle system in statistical mechanics, with particle $i \in \{1, \dots, N\}$ described by coordinates x_i, sampling π corresponds to generating configurations $\mathbf{x} = \{x_1, \dots, x_N\}$ distributed with the Boltzmann probability $\pi(\mathbf{x}) \propto \exp[-\beta E(\mathbf{x})]$, where E is the system energy and β the inverse temperature. For problems where $\mathbf{x}$ lies in a high-dimensional discrete or continuous space Ω, this sampling problem can usually not be solved directly [2,3].

MCMC consists instead in sampling a probability distribution $\pi^{(t)}$ that evolves with a time $t = 0, 1, 2, \dots$. The initial probability distribution, at time $t = 0$, $\pi^{(t=0)}$, can be sampled directly. Often, it is simply composed of a single configuration, so that $\pi^{(0)}$ is a δ-function on an explicitly given configuration $\{x_1, \dots, x_N\}$. In the limit $t \to \infty$, the distribution $\pi^{(t)}$ approaches the target probability distribution $\pi = \lim_{t\to\infty} \pi^{(t)}$. The key challenges in MCMC are the development of fast algorithms and the estimation of the time scale on which $\pi^{(t)}$ becomes close to π, for any initial distribution $\pi^{(0)}$. This program has met with considerable success in some models of statistical physics, as for example in the local Glauber dynamics in the two-dimensional Ising model [4,5].

The difference between two (normalized) probability distributions π and $\tilde{\pi}$ can be quantified by the total variation distance (TVD) [6,7],

$$\|\tilde{\pi} - \pi\|_{\text{TV}} = \frac{1}{2} \int_\Omega |\tilde{\pi}(\mathbf{x}) - \pi(\mathbf{x})| \mathrm{d}\mathbf{x} \tag{1}$$

$$= \int_{\mathcal{S}^+} [\tilde{\pi}(\mathbf{x}) - \pi(\mathbf{x})] \mathrm{d}\mathbf{x} \tag{2}$$

$$= \max_{\mathcal{A}\subseteq\Omega} \left| \int_{\mathcal{A}} [\tilde{\pi}(\mathbf{x}) - \pi(\mathbf{x})] \mathrm{d}\mathbf{x} \right|$$

$$= \max_{\mathcal{A}\subseteq\Omega} |\tilde{\pi}(\mathcal{A}) - \pi(\mathcal{A})|, \tag{3}$$

which satisfies $0 \le \|\tilde{\pi} - \pi\|_{\text{TV}} \le 1$. (Here, Ω is partitioned into two subsets $\mathcal{S}^+$ and $\mathcal{S}^-$, with $\tilde{\pi}(\mathbf{x}) \ge \pi(\mathbf{x}), \forall \mathbf{x} \in \mathcal{S}^+$ and analogously for $\mathcal{S}^-$. The integral over $\mathcal{S}^+$ in eq. (2) equals the corresponding one over $\mathcal{S}^-$, because of the normalization of probability, and it accounts for the prefactor of $\frac{1}{2}$ in eq. (1). Equation (3) reaches its maximum when $\mathcal{A} = \mathcal{S}^+$ or $\mathcal{S}^-$.) The mixing time of a Monte Carlo algorithm is defined as the time t after which the TVD (with $\tilde{\pi} \equiv \pi^{(t)}$ in eq. (3)) is smaller than a given threshold ϵ,

[a] E-mail: ze.lei@ens.fr
[b] E-mail: werner.krauth@ens.fr

for any initial distribution $\pi^{(0)}$. Although it is of great conceptual importance, the TVD cannot usually be computed. Because of this difficulty, practical simulations often carry systematic uncertainties that are difficult to quantify. Also, heuristic convergence criteria abound for the approach towards equilibrium in MCMC [3,8,9]. They most often involve time-correlation functions of observables, rather than the probability distribution itself (as in eqs. (1) and (3)).

In rare cases, MCMC algorithms allow for the definition of a stopping rule (based on the concept of a strong stationary time [6]), that yields a simulation-dependent time $t_{\rm ex}$ at which the configuration is sampled *exactly* from the distribution π. The value of $t_{\rm ex}$ depends on the realization of the Markov chain (that is, the sampled moves and, ultimately, the drawn random numbers). Nevertheless, the distribution of $t_{\rm ex}$ has an exponential tail, so the overall TVD still decays exponentially. For example, to randomize the configuration of N coins, at each time a random coin may be picked up and tossed with equal probability on either face. The time $t_{\rm ex}$ is when each coin was tossed at least once so that the reached configuration is exactly random. This example is equivalent to the "coupon-collector" problem for a model of N coupons. Now, at each time, one random coupon is marked. At the "coupon-collector" time, $t_{\rm ex}$, all the coupons have been marked. This problem will be discussed in detail later. The time $t_{\rm ex}$ is related to the mixing time [6]. Stopping rules exist for quite intricate models, as for the Ising model, using the coupling-from-the-past framework [3,10].

Most Markov-chain Monte Carlo algorithms are reversible: They satisfy the detailed-balance condition. Prominent examples are the Metropolis and the heat-bath algorithms [1,3]. In recent years, however, irreversible MCMC methods based on the global-balance condition have shown considerable promise [11–15]. In these algorithms, $\pi^{(t)}$ approaches π for long times, but the net probability flow no longer vanishes. One of them, the event-chain Monte Carlo (ECMC) algorithm [13,14], has proven useful for systems ranging from hard-sphere models [16] to spin systems [17], polymers [18,19] and to long-range interacting ensembles of molecules, such as water [20], where the Coulomb interaction plays a dominant role [21]. However, no exact results were known for the mixing behavior of ECMC, except for the case of a single particle [22].

In the present paper, we rigorously establish ECMC mixing times and stopping rules of the model of N hard spheres on a one-dimensional line with periodic boundary conditions (a circle). Reversible MCMC algorithms for this model and its variants were analyzed rigorously [23,24] and irreversible MCMC algorithm were discussed in detail [15]. The 1D hard-sphere model and reversible and irreversible MCMC algorithms are closely related to the symmetric exclusion process (SEP) on a periodic lattice [25] and to the totally asymmetric simple exclusion process (TASEP) [26–28]. For ECMC, an algorithm that

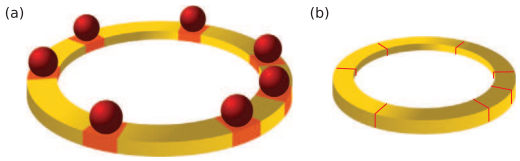


Fig. 1: (Color online) 1D hard-sphere model with periodic boundary conditions. (a) N spheres of diameter d on a ring of length L. (b) N point particles on a ring of length $L_{\rm free} = L - Nd$. Configurations and local MCMC algorithms are equivalent for both representations.

is closely related to the lifted Metropolis algorithm [15], we compute the TVD in a special case, and obtain the mixing times as a function of the parameter ϵ. We confirm the $\mathcal{O}(N^2 \log N)$ single-step mixing time that had been conjectured on the basis of numerical simulations [15]. Furthermore, we obtain a stopping rule for ECMC. We moreover present sequential variants of the forward Metropolis algorithm and the ECMC algorithm. For the latter, we prove an $\mathcal{O}(N^2)$ exact-sampling result that seems however not to generalize to the discretized version of the algorithm.

Hard spheres in 1D, reversible Monte Carlo. – The mixing and convergence behavior of Markov chains for particle systems has been much studied. For hard spheres in 2D and above, no rigorous results exist for the mixing times of local Markov chains. In the past, there was considerable controversy about the nature of the phase transition [16]. The mixing times of non-local MCMC algorithms could be clarified only at low densities [29,30]. We thus restrict our attention to the 1D hard-sphere model with periodic boundary conditions.

The 1D hard-sphere model can be represented as N spheres of diameter d on a line of length L with periodic boundary conditions (on a ring, see fig. 1(a)). A valid configuration a of N spheres has unit statistical weight $\pi(a) = 1$. Spheres do not overlap: The distance between sphere centers, and in particular between neighboring spheres, is larger than d. Each configuration of N hard spheres is equivalent to a configuration of N point particles on a ring of length $L_{\rm free} = L - Nd > 0$ (see fig. 1(b)).

We only consider *local* Markov chains, where moves of sphere i are accepted or rejected based on the position of i's neighbors only. To do so, we implement locality by rejecting a move of sphere i if the displacement would generate an overlap, but also if sphere i would hop over one of its neighbors. In this way, any local Monte Carlo move of spheres on a circle corresponds to an equivalent move in the point-particle representation (for which there are no overlaps and moves are rejected only because they represent a hop over a neighbor). The dynamics of both models is thus the same. In this paper, following [15], we count time steps in single displacements. This is more convenient because in our algorithms, new displacements build on previous ones (they form chains). Although we will study Markov chains that relabel spheres, we consider only

the relaxation of quantities that can be expressed through the unlabeled distances between neigboring spheres. This excludes the mixing in permutation space of labels or the self-correlation of a given sphere with itself (or another labeled sphere) at different times.

Detailed balance consists in requiring

$$\pi(a)p(a \to b) = \pi(b)p(b \to a), \tag{4}$$

where $p(a \to b)$ is the conditional probability to move from configuration a to configuration b and $\pi(a)p(a \to b)$ is the equilibrium probability flow from a to b. Equation (4) states that in equilibrium, the net probability flow between any two configurations vanishes. The heat-bath algorithm is a local reversible MCMC algorithm. At each time step, it replaces a sampled sphere i randomly in between its neighbors. The heat-bath algorithm mixes in at least $\mathcal{O}(N^3)$ and at most $\mathcal{O}(N^3 \log N)$ single steps [24]. Numerical simulations favor the latter possibility ($\mathcal{O}(N^3 \log N)$) [15]. For the one-dimensional hard-sphere model on a line without periodic boundary condition, the mixing time of $\mathcal{O}(N^3 \log N)$ single steps (corresponding to $\mathcal{O}(N^2 \log N)$ "sweeps" of N steps) is rigorously proven [23].

Analogous to the heat-bath algorithm, the reversible Metropolis algorithm also satisfies the detailed-balance condition: At each time step, a randomly chosen sphere i attempts a move by ϵ taken from some probability distribution. The move is rejected if the proposed displacement ϵ is larger than the free space in the direction of the proposed move ($x_{i_+} - x_i - d$ for $\epsilon > 0$) or behind it ($x_i - x_{i_-} - d$ for $\epsilon < 0$) (where we suppose that i_+ is the right-hand-side neighbor of i, etc., and imply periodic boundary conditions). In the point-particle model, the equivalent move is rejected if the particle would hop over one or more of its neighbors and is accepted otherwise. Rigorous results for mixing times are unknown for the Metropolis algorithm, but numerical simulations clearly identify mixing in $\mathcal{O}(N^3 \log N)$ single steps as for the heat-bath algorithm [15]. In the discrete 1D hard-sphere model on the circle with L sites and N particles, the Metropolis algorithm is implemented in the so-called simple exclusion process (SEP), where at each time step, a randomly chosen particle attempts to move with equal probability to each of its two adjacent sites. The move is rejected if that site is already occupied. The mixing time of the SEP is $\sim (4\pi^2)^{-1} N L^2 \log N$ single steps (for $L \geq 2N$) [25].

From the forward Metropolis to the event-chain algorithm. – Irreversible Monte Carlo algorithms violate the detailed-balance condition of eq. (4) but instead satisfy the weaker global-balance condition

$$\sum_b \pi(b)p(b \to a) = \pi(a). \tag{5}$$

Together with the easily satisfiable aperiodicity and irreducibility conditions [6], the global-balance condition ensures that the steady-state solution corresponds to the probability π, but without necessarily cancelling the net probability flow $\pi(a)p(a \to b) - \pi(b)p(b \to a)$ between configurations a and b (cf. eq. (4)). Here, we take up the forward Metropolis algorithm studied earlier, in a new variant that involves swaps. This allows us to arrive at an exact mixing result.

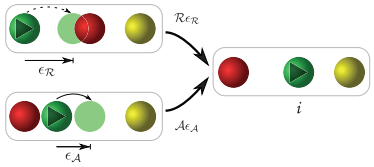


Fig. 2: (Color online) Probability flow of the forward swap Metropolis algorithm into a configuration (a, i) (the active sphere i is shown in green). A rejected sphere move, by a displacement $\epsilon_{\mathcal{R}}$ (upper case), entails a swap and contributes $\mathcal{R}_{\epsilon_{\mathcal{R}}}$. An accepted sphere move, by a displacement $\epsilon_{\mathcal{A}}$ (lower case), contributes $\mathcal{A}_{\epsilon_{\mathcal{A}}}$. For any ϵ, one of the probability flows equals one, and the other zero $[(\mathcal{R}_\epsilon, \mathcal{A}_\epsilon) \in \{(0,1),(1,0)\}]$.

In the forward swap Metropolis algorithm[1], at each time step, a randomly chosen sphere i attempts to move by a random displacement ϵ with a predefined sign (that for clarity, we take to be positive). If the move is rejected (the displacement ϵ does not yield a valid hard-sphere configuration), the sphere swaps its label with the sphere responsible for the rejection (see the upper move in fig. 2). Else, if the displacement ϵ is accepted, the sphere i simply moves forward (see the lower move in fig. 2). The total probability flow into a configuration (a, i) (that is, the N-sphere configuration a with the active sphere i) is

$$\mathcal{F}(a,i) = \int_0^\infty \mathrm{d}\epsilon p(\epsilon) \underbrace{[\mathcal{A}_\epsilon(a,i) + \mathcal{R}_\epsilon(a,i)]}_{=1 \text{ (see fig. 2)}} = 1 = \pi(a), \tag{6}$$

so that the algorithm satisfies global balance. The swap allows both the rejected and the accepted moves into the configuration (a, i) to involve the sphere i only. The forward swap Metropolis algorithm is equivalent (up to relabeling) to the forward Metropolis algorithm treated earlier if at each time step the active sphere i is sampled randomly (see fig. 3). The mixing time of this algorithm was conjectured to be $\mathcal{O}(N^{5/2})$ single steps, based on numerical simulations [15]. This agrees with the proven mixing-time scale of the totally asymmetric simple exclusion process (TASEP) [28].

The forward swap Metropolis algorithm satisfies global balance for any choice of the sphere i and any step-size distribution $p(\epsilon)$. This implies that the active-sphere index i need not be sampled randomly for the algorithm to remain valid. This distinguishes it from the forward Metropolis algorithm (without the swaps) treated in previous work [15]. In particular, the sphere i remains active

[1] The forward Metropolis algorithm introduced earlier [15] did not feature swaps.

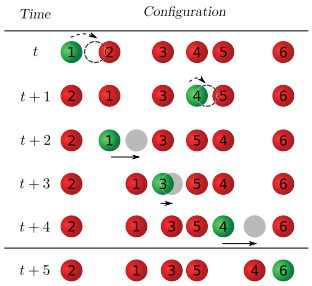


Fig. 3: (Color online) Forward swap Metropolis algorithm, with configurations $\mathbf{x}_t, \ldots, \mathbf{x}_{t+5}$. The active sphere is sampled randomly at each time step so that the swaps merely relabel the spheres.

for several chains in a row. The forward swap Metropolis algorithm, run with the following sequence of active spheres:

$$\ldots, \underbrace{i, i, \ldots, i, i}_{\text{chain } n}, \underbrace{j, j, \ldots, j, j}_{\text{chain } n+1}, \underbrace{k, k, \ldots, k, k}_{\text{chain } n+2}, \ldots, \qquad (7)$$

is equivalent to the lifted forward Metropolis algorithm studied earlier [15], if the active spheres $i, j, k, \ldots$ in eq. (7) are sampled randomly. The algorithm naturally satisfies the global-balance condition, and again, each individual move attempts a displacement by a distance ϵ sampled from a given distribution $p(\epsilon)$ that vanishes for negative ϵ, and the chain lengths (number of repetitions of $i, j, k, \ldots$) $n, n+1, \ldots$ are sampled from a distribution. Numerical simulations have lead to the conjecture that this algorithm mixes in $\mathcal{O}(N^2 \log N)$ single steps [15], that is, much faster than the local reversible Markov chains.

ECMC is a general algorithm formulated in continuous time and space. In its hard-sphere version [13], only a single sphere moves, in a fixed direction. This "active" sphere continues to move until it is stopped (rejected) by another sphere (which indicates a "lifting"), and the rejecting sphere becomes the active one. This process repeats until the total displacement reaches a given "chain length" ℓ. The algorithm has been very successful [16] in higher dimensions. The sequence of "lifting" events forms the eponymous event chain. ECMC is the continuous-time limit of the lifted forward Metropolis algorithm, with the simultaneous limits $\epsilon \to 0$ and $l \to \infty$, but $(\langle\epsilon\rangle l) \to \ell$, where the chain length ℓ on the scale L_{free}, is again sampled from a given probability distribution. In the point-particle representation of fig. 1(b), one "event" chain of the ECMC algorithm simply moves the active particle i from its initial position x_i to $x_i + \ell$. It advances the time as $t \to t + \ell$, and increments the number of chains as $n \to n + 1$. The number of ECMC events (the number of changes of the active sphere) then grows approximately as $\sim (\ell/L_{\text{free}})N$. When $\ell \sim \text{unif}(0, L_{\text{free}})$, this places particle i at a random position on a ring. For this special uniform distribution of chain lengths, a perfect sample is clearly obtained once all particles were at least once picked as the active particle. This situation will now be analyzed in terms of the coupon-collector problem (see [31,32]).

For the ECMC with $\ell \sim \text{unif}(0, L_{\text{free}})$, the TVD can be expressed by the probability that at least one particle has never been picked as an active particle of a chain. Without restriction, we suppose that the initial configuration is the compact state $\mathbf{x} = \{0, 0, \ldots, 0\}$. We also measure time in the number of chains n (which means $t(n) = (\langle\ell\rangle/L_{\text{free}})Nn$ number of events). In eq. (3), the set $\mathcal{A}$ is

$$\mathcal{A} = \{\mathbf{x} \mid \exists\ i \text{ with } x_i = 0\}. \qquad (8)$$

Also, clearly, $\pi^{(n)}(\mathcal{A})$ equals the probability that at least one particle has never been picked as an active particle of a chain, whereas $\pi(\mathcal{A}) = 0$, as it is a lower-dimensional subset of Ω. This then becomes the coupon-collector problem introduced before. The expected time to have them all marked at least once is

$$\langle n_1\rangle = NH_N = N\log N + \gamma N + \frac{1}{2} + \mathcal{O}(1/N), \qquad (9)$$

where $H_N = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{N}$ is the N-th harmonic number and $\gamma = 0.5772\ldots$ is the Euler-Mascheroni constant. The tail distribution (the probability that the complete-collection time n_1 is larger than a given time $T = N\log N + cN$) is smaller than the sum of probabilities for each single coupon to not have been marked from the beginning:

$$\begin{aligned} P(n_1 > N\log N + cN) &< N(1 - \frac{1}{N})^{N(\log N + c)} \\ &= N\exp(-\log N - c) = \exp(-c). \end{aligned} \qquad (10)$$

The naive estimation of eq. (10) was improved [31] by removing the intersection of events:

$$P(n_1 < N\log N + cN) \sim \exp[-\exp(-c)]. \qquad (11)$$

This can be generalized to multiple complete collections (see eq. (14) and fig. 4).

From eqs. (1), (3), and (11), we obtain, for $N \to \infty$,

$$||\pi^{(n)} - \pi||_{\text{TV}} \sim 1 - \exp\left[-\exp\left(-\frac{n - N\log N}{N}\right)\right]. \qquad (12)$$

Rather than computing the difference between $\pi^{(n)}$ and π at a fixed number n of chains, one can simply run ECMC (with $\ell \sim \text{unif}(0, L_{\text{free}})$) until the time t_{ex} at which chains with all of the N particles as active ones have completed. The expected number of chains is given by eq. (9), and in both ways, we see that mixing takes place after $\mathcal{O}(N\log N)$ chains (corresponding to $\mathcal{O}(N^2\log N)$ single events), confirming, for a special distribution of ℓ, an earlier conjecture [15]. The discussed mixing behavior of ECMC can more generally be obtained for distributions $\ell \sim \text{unif}(c, c + L_{\text{free}})$ with arbitrary (and even with negative) c. In our special case, choosing $c = -L_{\text{free}}/2$ would lead to the smallest number of individual events. In view

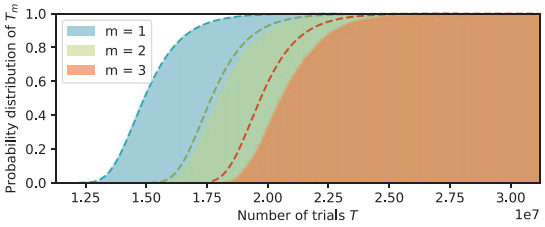


Fig. 4: (Color online) Cumulative probability of the coupon collector problem ($m = 1$), and of the m-coupon collector problem for $m = 2$ and $m = 3$. Numerical simulations (the colored histograms) for $N = 2^{20}$ particles are compared to the asymptotic tail probability of eq. (14) (the dashed curves).

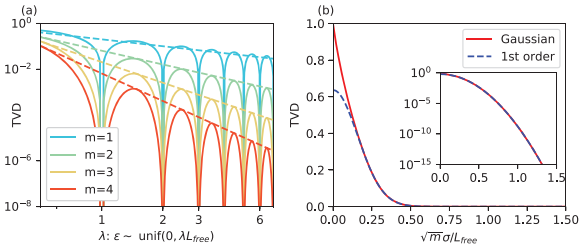


Fig. 5: (Color online) TVD for a single sphere on a ring with uniform and Gaussian distributions of $\ell \equiv \epsilon$. (a) TVD after m displacements $\epsilon \sim \text{unif}(0, \lambda L_{\text{free}})$. The TVD trivially vanishes for integer λ. Peaks decay as $\frac{2}{\pi}(\pi\lambda)^{-m}$ (for $m \to \infty$). (b) TVD for m Gaussian displacements with standard deviation σ, compared with its first-order approximation from the Jacobi ϑ function (see eq. (20)). The inset illustrates the good agreement of the approximation on a logarithmic scale.

of the practical applications of ECMC, it appears important to understand whether this dependence on the distribution of ℓ (rather than on its mean value) has some relevance for the simulation of discrete 1D models, and whether it survives in higher dimensions, and for continuous (non-hard-sphere) potentials.

We next consider more general distributions, namely the uniform distribution $\ell \sim \text{unif}(0, \lambda L_{\text{free}})$, as well as the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, where μ is the mean value and σ the standard deviation. Again, particles are effectively independent and we conjecture the mixing time (which can now never lead to perfect sampling) to be governed by the particle which has moved the least number, m, of times. This is equivalent to the generalization of m-complete-collection in the coupon-collector problem [31], whose tail probability is given by

$$P(n_m < T) \sim \exp(-\Upsilon/(m-1)!) \tag{13}$$

with

$$\Upsilon = \exp\left[-\frac{T - N\log N - (m-1)N\log\log N}{N}\right] \tag{14}$$

(see fig. 4). This means that the number of chains to collect each of the N coupons at least m times only adds an $N \log\log N$ correction to the general $N \log N$ scale of chains.

To gain intuition about the general mixing process with an arbitrary distribution of the chain length, we now compute the TVD for the single-particle problem (for which $\ell \equiv \epsilon$). For simplicity, we set $L_{\text{free}} = 1$ (measure the displacements in units of L_{free}). Because of periodic boundary conditions, particle positions x are defined only modulo 1. Its probability distribution after m chains is therefore given by the sum over the different windings k:

$$\pi^{(m)}(x) = \sum_{k=-\infty}^{+\infty} p_m(x+k), \tag{15}$$

where $p_m(x)$ is the distribution of the sum of m chain lengths. The TVD for chain lengths $\ell_i \sim \text{unif}(0, \lambda)$, as discussed, equals the one for $\ell_i \sim \text{unif}(-\lambda/2, \lambda/2)$. $p_m^{\text{unif}}(x)$ then follows the distribution using the characteristic function:

$$p_m^{\text{unif}}(x) = \int_{-\infty}^{\infty} \mathrm{d}t \mathrm{e}^{-2\pi i t x} \left[\frac{\sin(\pi\lambda t)}{\pi\lambda t}\right]^m. \tag{16}$$

Using the Poisson summation formula and subtracting the equilibrium distribution $\pi(x) = 1$, we find

$$\sum_{k=-\infty}^{\infty} p_m^{\text{unif}}(x+k) - 1 = \sum_{k\in\mathbb{N}^+} 2\left[\frac{\sin(\pi k\lambda)}{\pi k\lambda}\right]^m \cos(2\pi k x).$$

The TVD for chain lengths $\ell_i \sim \text{unif}(0, \lambda)$ thus satisfies

$$\begin{aligned}\|\pi^{(m)} - \pi\|_{\text{TV}} &= \int_0^1 \mathrm{d}x \left|\sum_{k\in\mathbb{N}^+}\left[\frac{\sin(\pi k\lambda)}{\pi k\lambda}\right]^m \cos(2\pi k x)\right| \\ &\sim \frac{2}{\pi}\left|\frac{\sin(\pi\lambda)}{\pi\lambda}\right|^m \qquad (\text{for } m \to \infty). \end{aligned} \tag{17}$$

The TVD trivially vanishes for integer λ (see fig. 5(a)). Its peaks decay as $\frac{2}{\pi}(\pi\lambda)^{-m}$.

For Gaussian-distributed chain lengths $\ell_i \sim \mathcal{N}(\mu, \sigma^2)$, the sum of m chains is distributed as

$$\sum_{i=1}^{m} \ell_i \sim \mathcal{N}(m\mu, m\sigma^2). \tag{18}$$

With ϑ_3 the Jacobi theta function, we now have

$$\begin{aligned}\sum_{k=-\infty}^{\infty} p_m^{\text{Gauss}}(x+k) - 1 &= \vartheta_3\left[\pi(x+\mu), \mathrm{e}^{-2\pi^2 m\sigma^2}\right] - 1 \\ &= 2\sum_{k=1}^{\infty} \exp(-2k^2\pi^2 m\sigma^2)\cos\left[2k\pi(x+m\mu)\right]. \end{aligned} \tag{19}$$

The TVD for the distribution of eq. (18) satisfies

$$\begin{aligned}\|\pi^{(m)} - \pi\|_{\text{TV}} &= \int_0^1 \mathrm{d}x \left|\sum_{k=1}^{\infty} \exp(-2k^2\pi^2 m\sigma^2)\cos(2k\pi x)\right| \\ &\sim \frac{2}{\pi}\exp(-2\pi^2 m\sigma^2) \qquad (\text{for } m\sigma^2 \to \infty) \end{aligned} \tag{20}$$

(see fig. 5(b)).

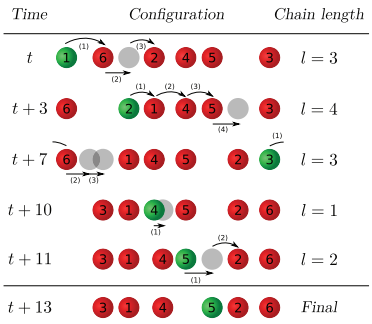


Fig. 6: (Color online) Sequential lifted forward Metropolis algorithm (with swaps). Configurations $\mathbf{x}_t, \ldots, \mathbf{x}_{t+13}$ sampled through five chains with active sphere $1, 2, \ldots, 5$ are shown. Chain lengths are $l_1 = 3, \ldots, l_5 = 2$. Each sphere displacement $\epsilon_t > 0$ is either accepted or, if rejected, it induces a swap, so that the same sphere remains active throughout a chain.

Both for the uniform and the Gaussian distribution, the single-sphere TVD decreases exponentially with the number m of displacements (which are equivalent to single-particle chains). We expect the same behavior for the N-sphere problem, where m is now the number of chains for the m-complete-collection problem.

Sequential forward Metropolis, sequential ECMC. – ECMC, with randomly sampled initial spheres and a standard deviation of the chain-length distribution $\sigma \sim L_{\text{free}}$, mixes in $\mathcal{O}(N^2 \log N)$ events (corresponding to $\mathcal{O}(N \log N)$ chains). In the label-switching framework of ECMC, each chain consists in advancing the particle i by a distance ℓ times, and both the ECMC and the forward Metropolis versions are correct. Instead of sampling the active sphere for each chain, so that the coupon-collector aspect necessarily brings in the $\log N$ factor in the scaling of mixing times, we may also sequentially increment the active-sphere index for each chain (see fig. 6):

$$\ldots, \underbrace{i, \ldots, i}_{\text{chain } i}, \underbrace{i+1, \ldots, i+1}_{\text{chain } i+1}, \underbrace{i+2, \ldots, i+2}_{\text{chain } i+2}, \ldots, \qquad (21)$$

(where particle numbers are implied modulo N). Sequential ECMC, with a distribution $\ell_i \sim \text{unif}(0, L_{\text{free}})$ produces an exact sample in $\mathcal{O}(N^2)$ events (corresponding to exactly N chains).

Evidently, the analysis of eqs. (17) and (20) can be applied to the sequential ECMC with distributions such as $\text{unif}(0, \lambda L_{\text{free}})$ and, more generally, distributions with $\sigma \sim L_{\text{free}}$. After each "sweep" of chains, the TVD factorizes, and we expect mixing to take place after $\mathcal{O}(N)$ chains (corresponding to $\mathcal{O}(N^2)$ events).

ECMC is the limit of the lifted forward Metropolis algorithm, and the sequential ECMC the limit of the sequential lifted forward Metropolis algorithm for step sizes much smaller than the mean free space between spheres ($\langle \epsilon \rangle = L_{\text{free}}/(2N\alpha)$ with $\alpha \gg 1$). For a given discretization $2/\alpha$, and for small N, the sequential lifted forward algorithm

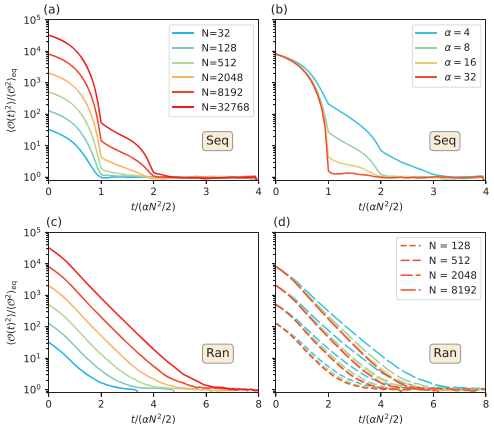


Fig. 7: (Color online) Crossover from the discrete lifted algorithm to ECMC, *via* the variance of the mid-system distance $x_{i+N/2} - x_i$ for ordered x_i, started from compact initial condition (see [15]). Discrete step size with $\epsilon \sim \text{unif}(0, L_{\text{free}}/N/\alpha)$, and chain length $l \sim \text{unif}\{\alpha, \alpha N\}$ (a) Sequential lifted Metropolis algorithm with constant $\alpha = 10$ for different N: The cross-over from perfect sampling for small N at a time scale $\mathcal{O}(N^2)$ towards $\mathcal{O}(N^2 \log N)$ appears evident. (b) Sequential algorithm for $N = 8192$, with increasing α: $\mathcal{O}(N^2)$ mixing scale emerges for large α. (c) Random lifted Metropolis algorithm with $\alpha = 10$ for different N (legend as in (a)): $\mathcal{O}(N^2 \log N)$ mixing time scale (conjectured earlier [15]). (d) Random lifted forward Metropolis algorithm: Limited role of α (color code for α as in (b)).

mimics the $\mathcal{O}(N^2)$ mixing of the sequential ECMC, but for large N, it seems to cross over into $\mathcal{O}(N^2 \log N)$ mixing (see fig. 7(a)). $\mathcal{O}(N^2)$ mixing also emerges at fixed N for large α (see fig. 7(b)). (This is obtained using the heuristic mid-system variance $x_{i+N/2} - x_i$ for ordered x_i, see [15].) In contrast, the random lifted forward Metropolis algorithm shows $\mathcal{O}(N^2 \log N)$ mixing (see fig. 7(c)), as discussed earlier [15]. This scaling is little influenced by the discretization (see fig. 7(d)). It thus appears that the $N \to \infty$ limit of the sequential lifted forward Metropolis algorithm does not commute with the small-discretization limit $\alpha \to \infty$.

Conclusions. – In this paper we have proven that for 1D hard spheres, ECMC with a uniform distribution of chain length $\ell \sim [0, L_{\text{free}}]$, with $L_{\text{free}} = L - Nd$ realizes a perfect sample in $\mathcal{O}(N^2 \log N)$ single steps (events) that correspond to $\mathcal{O}(N \log N)$ chains. This confirms, in a special case, an earlier conjecture [15] for the mixing time of ECMC. This also proves that ECMC can be much faster than local reversible Monte Carlo algorithms. We computed the TVD but also indicated a stopping rule for the time t_{ex} after which the configuration is exactly in equilibrium. We have also provided numerical evidence that the $N^2 \log N$ mixing prevails for other distributions of ℓ, namely for the uniform distribution $\text{unif}(0, \lambda L_{\text{free}})$ and the Gaussian, and used the coupon-collector problem to justify this approximation. We have furthermore discussed a sequential ECMC algorithm which mixes in $\mathcal{O}(N^2)$ events.

This algorithm uses "particle swaps", but it remains local. The discrete version of this algorithm (the sequential lifted forward Metropolis algorithm) crosses over, for large N, to $\mathcal{O}(N^2 \log N)$ mixing.

The lessons from our analysis of 1D hard-sphere systems are threefold. First, irreversible Markov chains can be proven to mix on shorter time scales than reversible algorithms. Second, the speed of ECMC was proven to depend on the whole distribution of the chain lengths ℓ, but to be independent of its mean value. Third, sequential-update algorithms (that remain valid in higher dimensions) can be shown to mix on faster time scales than random-update versions. It will be important to understand whether the design and the analysis of ECMC algorithms in higher-dimensional particle models will be influenced by these lessons, that were overlooked in all previous work on ECMC.

* * *

We thank Florent Krzakala and Sebastian C. Kapfer for helpful discussions. WK acknowledges support from the Alexander von Humboldt Foundation.

REFERENCES

[1] Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H. and Teller E., *J. Chem. Phys.*, **21** (1953) 1087.
[2] Devroye L., *Non-Uniform Random Variate Generation* (Springer, New York) 1986.
[3] Krauth W., *Statistical Mechanics: Algorithms and Computations* (Oxford University Press) 2006.
[4] Martinelli F., *Lectures on Glauber Dynamics for Discrete Spin Models*, in *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXVII - 1997*, edited by Bernard P. (Springer, Berlin, Heidelberg) 1999, pp. 93–191.
[5] Lubetzky E. and Sly A., *Commun. Math. Phys.*, **313** (2012) 815.
[6] Levin D. A., Peres Y. and Wilmer E. L., *Markov Chains and Mixing Time* (American Mathematical Society) 2008.
[7] Diaconis P., *J. Stat. Phys.*, **144** (2011) 445.
[8] Berg B. A., *Markov Chain Monte Carlo Simulations and Their Statistical Analysis: With Web-based Fortran Code* (World Scientific) 2004.
[9] Landau D. and Binder K., *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press) 2013.
[10] Propp J. G. and Wilson D. B., *Random Struct. Algorithms*, **9** (1996) 223.
[11] Turitsyn K. S., Chertkov M. and Vucelja M., *Physica D*, **240** (2011) 410.
[12] Fernandes H. C. and Weigel M., *Comput. Phys. Commun.*, **182** (2011) 1856.
[13] Bernard E. P., Krauth W. and Wilson D. B., *Phys. Rev. E*, **80** (2009) 056704.
[14] Michel M., Kapfer S. C. and Krauth W., *J. Chem. Phys.*, **140** (2014) 054116.
[15] Kapfer S. C. and Krauth W., *Phys. Rev. Lett.*, **119** (2017) 240603.
[16] Bernard E. P. and Krauth W., *Phys. Rev. Lett.*, **107** (2011) 155704.
[17] Lei Z. and Krauth W., *EPL*, **121** (2018) 10008.
[18] Kampmann T. A., Boltz H.-H. and Kierfeld J., *J. Chem. Phys.*, **143** (2015) 044105.
[19] Harland J., Michel M., Kampmann T. A. and Kierfeld J., *EPL*, **117** (2017) 30001.
[20] Faulkner M. F., Qin L., Maggs A. C. and Krauth W., *J. Chem. Phys.*, **149** (2018) 064113.
[21] Kapfer S. C. and Krauth W., *Phys. Rev. E*, **94** (2016) 031302.
[22] Diaconis P., Holmes S. and Neal R. M., *Ann. Appl. Probab.*, **10** (2000) 726.
[23] Randall D. and Winkler P., *Mixing Points on an Interval*, in *Proceedings of the Seventh Workshop on Algorithm Engineering and Experiments and the Second Workshop on Analytic Algorithmics and Combinatorics, ALENEX/ANALCO 2005, Vancouver, BC, Canada, 22 January 2005* (SIAM) 2005, pp. 218–221.
[24] Randall D. and Winkler P., *Mixing Points on a Circle*, *Lect. Notes Comput. Sci.*, Vol. **3624** (Springer, Berlin, Heidelberg) 2005, pp. 426–435.
[25] Lacoin H., *Ann. Inst. H. Poincaré Probab. Stat.*, **53** (2017) 1402.
[26] Gwa L.-H. and Spohn H., *Phys. Rev. Lett.*, **68** (1992) 725.
[27] Chou T., Mallick K. and Zia R. K. P., *Rep. Prog. Phys.*, **74** (2011) 116601.
[28] Baik J. and Liu Z., *J. Stat. Phys.*, **165** (2016) 1051.
[29] Wilson D. B., *Random Struct. Algorithms*, **16** (2000) 85.
[30] Kannan R., Mahoney M. W. and Montenegro R., *Rapid mixing of several Markov chains for a hard-core model*, in *14th Annual ISAAC, Lect. Notes Comput. Sci.*, Vol. **2906** (Springer, Berlin, Heidelberg) 2003, pp. 663–675.
[31] Erdos P. and Rényi A., *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **6** (1961) 215.
[32] Blom G., Holst L. and Sandell D., *Problems and Snapshots from the World of Probability* (Springer, New York) 1994.

Publication III: Event-chain Monte Carlo with factor fields

Publication III

Ze Lei, Werner Krauth and Anthony C. Maggs
Arxiv: 1812.02494 , submitted tp Phys. Rev. E.

We study one-dimensional interacting-particle models using the event-chain Monte Carlo algorithm. We show that a slowdown of the method (rooted in the mismatch of potential differences) is overcome by the introduction of factor fields. The method is motivated for the harmonic chain, and further tested for Lennard-Jones interacting particles and for hard spheres in one diemension, by choosing the optimal factor field as the *pressure* of the system. Generalizations to higher dimensions are outlined. (See chapter 5.)

Event-chain Monte Carlo with factor fields

Ze Lei,[1,*] Werner Krauth,[1,†] and A. C. Maggs[2,‡]

[1] *Laboratoire de Physique Statistique, Ecole Normale Supérieure / PSL Research University, UPMC, Université Paris Diderot, CNRS, 24 rue Lhomond, 75005 Paris, France*

[2] *CNRS UMR7083, ESPCI Paris, PSL Research University, 10 rue Vauquelin, 75005 Paris, France*

(Dated: December 7, 2018)

We study the dynamics of one-dimensional (1D) interacting particles simulated with the event-chain Monte Carlo algorithm (ECMC). We argue that previous versions of the algorithm suffer from a mismatch in the factor potential between different particle pairs (factors) and show that in 1D models, this mismatch is overcome by factor fields. ECMC with factor fields is motivated, in 1D, for the harmonic model, and validated for the Lennard-Jones model as well as for hard spheres. In 1D particle systems with short-range interactions, autocorrelation times generally scale with the second power of the system size for reversible Monte Carlo dynamics, and with its first power for regular ECMC and for molecular-dynamics. We show, using numerical simulations, that they grow only with the square root of the systems size for ECMC with factor fields. Mixing times, which bound the time to reach equilibrium from an arbitrary initial configuration, grow with the first power of the system size.

I. INTRODUCTION

The dynamics of physical systems plays an important role in numerous fields of science. The study of dynamics aims at elucidating equilibrium and non-equilibrium phenomena, including correlation functions, coarsening dynamics after a quench, and manifestly non-equilibrium phenomena such as turbulence. In computational statistical physics, Markov chain Monte Carlo (MCMC) [1–3] and molecular-dynamics (MD) algorithms [4] are often employed to generate equilibrated samples and to determine thermodynamic averages and correlations. The non-equilibrium aspect then consists in characterizing the approach to equilibrium from an arbitrary, atypical, initial condition. This is quantified by the mixing time [2], an important figure of merit for MCMC. The other important time scale characteristic of a physical system is the autocorrelation times of the underlying Markov process given by the inverse gap of the transition matrix.

In reversible Markov chains (as used in the vast majority of MCMC algorithms), the requirement that the long-time steady state corresponds to thermodynamic equilibrium is expressed through the detailed-balance condition, which assures that all the net probability flows vanish in equilibrium. In recent years, however, irreversible Markov chains have been found to show considerable promise [5–8]. They feature a steady state with non-vanishing net probability flows if a weaker global-balance condition is satisfied. Global balance corresponds to an incompressibility condition in configuration space: the steady-state flows into each configuration sum to the flows out of it.

An example of an irreversible Markov chain is the event-chain algorithm (ECMC) [9, 10]. This algorithm has been successfully applied to many problems from hard-sphere and soft-sphere melting [11–13] to spin models [14, 15] and quantum-field theory [16]. In this paper, we study the relaxation times of ECMC in one-dimensional (1D) models of N particles with local interactions [17, 18]. We analyze in detail the relaxation of both Lennard-Jones and hard-sphere models, study the statistical properties of ECMC trajectories and show how to greatly accelerate known algorithms by the introduction of a "factor field", which compensates the system pressure, P, without influencing physical properties.

A. Characteristic times of Markov chains

Irreversible MCMC algorithms can be faster than their reversible counterparts. A particularly interesting case is the 1D hard-sphere model of N spheres (rods). For this model, the local heat-bath algorithm mixes in $\mathcal{O}\left(N^3 \log N\right)$ moves [19] on an interval with fixed boundary conditions. The mixing time for the same model with periodic boundary conditions is between $\mathcal{O}\left(N^3\right)$ and $\mathcal{O}\left(N^3 \log N\right)$ [20]: Simulations favor the latter [17]. The reversible Metropolis algorithm has a similar mixing behavior. Various local irreversible Markov chains mix in $\mathcal{O}\left(N^{5/2}\right)$ moves (forward Metropolis algorithm [17]), in $\mathcal{O}\left(N^2 \log N\right)$ steps (lifted forward Metropolis algorithm and ECMC [17, 18]) and even $\mathcal{O}\left(N^2\right)$ single moves with a re-labeling ECMC [18].

The lifted forward Metropolis algorithm in continuous space with infinitesimal movements constitutes ECMC. For hard spheres, it is deterministic without restarts, but then mixes in $\mathcal{O}\left(N^2 \log N\right)$ steps at randomized stopping times [18]. Although ECMC is irreversible under a transformation of the time $t \mapsto -t$, under the combined transformation of times and positions $(t, x) \mapsto (-t, -x)$,

* ze.lei@ens.fr
† werner.krauth@ens.fr
‡ anthony.maggs@espci.fr

the dynamics runs backwards in time. The irreducibility of the lifted forward Metropolis algorithm can be shown using this time-reversal property. It may also explain why ECMC is typically as fast as MD.

Previous work has also explored the autocorrelation times (rather than the mixing times) under ECMC dynamics in the D-dimensional harmonic-solid model, of which the equilibrium properties can also be obtained exactly (see [21]). In 1D, the dynamic exponent of the autocorrelations under ECMC dynamics takes the strikingly low value of $z = 1/2$, corresponding to an equilibrium autocorrelation time involving $\mathcal{O}(N^{3/2})$ moves or $\tau \sim \mathcal{O}(N^{1/2})$ sweeps. This is $N^{1/2}$ times smaller (faster) than the best autocorrelation time found in the hard-sphere system.

The present paper starts from the similarity between the dynamics of the 1D harmonic-solid model and that of the Lennard-Jones model at low temperature T. We generalize this favorable scaling of the harmonic model to all T in the Lennard-Jones model as well as the hard-sphere model. We expect that this concept can also be generalized for higher-dimensional models [22, 23].

B. 1D particle systems, algorithms

We consider a 1D system of N particles $i \in \{1, \ldots, N\}$ with $x_i < x_{i+1}$ on an interval of length L with periodic boundary conditions in N and in L. In the reversible local Metropolis algorithm, at each iteration, a randomly chosen particle i is proposed to move as $x_i \to x_i + \mathrm{ran}[-\epsilon, \epsilon]$, where ran is a random number uniformly distributed between $-\epsilon$ and ϵ. For hard spheres, the move is accepted if the new sphere position does not lead to overlaps with spheres $i-1$ and $i+1$, and in addition does not induce a change of the ordering. In the presence of a potential U, the move is accepted with probability $\min(1, \exp(-\Delta U/T))$, where ΔU is the change in potential for the proposed move. The amplitude ϵ is chosen to maximize the speed of the method. In the heat-bath algorithm the distribution of the particle i is fully resampled in the potential of its neighbors at each time step.

ECMC, for one-dimensional hard spheres [9, 17], consists in moving spheres in a chain-like manner. Up to a restart time, sphere i moves with unit velocity until it collides with sphere $i+1$, at which moment it stops, and sphere $i+1$ moves forward. For each of the subsequent "chains" (the displacements between restarts), the starting sphere is randomly chosen, and the length of the chain (the time until the next restart) is sampled from a distribution on the length scale L. For a more general interaction potential, ECMC breaks up the total system potential up into separate "factor potentials", each of which is treated independently [10, 24]. A factor potential provides for a randomized stopping time. For a given move involving particle i, the smallest stopping time of all factors provides the next event time. The next particle to move is determined through a lifting scheme [22, 25] from the factor triggering the event. With potentials more general than hard spheres, restarts are no longer required to ensure irreducibility of ECMC.

In ECMC, path statistics in equilibrium and pressure P are linked by

$$P/T \propto \frac{1}{t} \left\langle x_{i(t)} - x_{i(0)} \right\rangle , \tag{1}$$

where $x_i(t)$ is the position of the particle that is active at time t [10, eq.(20)]. Eq. (1) holds for all time intervals t. It is very convenient as an unbiased estimator of the pressure, and has been much used [13]. The factor fields of the present paper will allow us to exactly compensate the pressure without affecting the physical properties of the system, and lead to greatly accelerated ECMC methods.

In many-particle simulations, MD algorithms generally feature smaller relaxation time scales than MCMC methods. In essence this is because momentum conservation (present in MD, but absent in MCMC) allows for faster transport of inhomogeneities in the velocity and position fields ([26, 27]). In our comparisons with ECMC, MD simulations are performed using the leapfrog or Verlet algorithm coupled to a Langevin thermostat for the velocity. The integration time step δt is adjusted by finding the stability limit of the integrator, then reducing δt by an order of magnitude. Inverse error analysis shows that the effective Hamiltonian is close to that of the original model, with a systematic shift of $O(\delta t^2)$ in the effective Hamiltonian [28]. We choose the strength of coupling to the Langevin thermostat so that the longest wavelength mode is close to critically damped.

We concentrate our measurements on the dynamics of the structure factor of the lowest Fourier coefficient

$$S(q) = \frac{1}{N} \left| \sum_{j=1}^{N} \mathrm{e}^{iq \cdot x_j} \right|^2 , \tag{2}$$

with $q = 2\pi/L$, which is sensitive to large-scale motion of particles. The integrated autocorrelation times τ of $S(\frac{2\pi}{L})$ are measured in "sweeps", that is, a constant time interval for all N particles in MD, N attempted displacements in MCMC, or N events in ECMC (in comparing the methods, we compensate for the different implementation speeds of a sweep in MD, MCMC, ECMC). We use the blocking method [29] to quantify the algorithm speed.

C. ECMC for harmonic interactions

We first consider a harmonic potential with a minimum at a separation b between neighboring particles:

$$U_{\mathrm{harm}}(x_{i+1} - x_i; b) = \frac{k}{2} \left[(x_{i+1} - x_i) - b \right]^2 , \tag{3}$$

where periodic boundary conditions for the particle separation $x_{i+1} - x_i$ are taken. They are also implied for

the particle indices. The total potential of the system of a fixed length L is

$$U_{\rm harm}(\{x_i\};b) = \frac{k}{2}\sum_{i=1}^{N}(x_{i+1}-x_i-b)^2 \quad (4)$$

$$= U_{\rm harm}(\{x_i\};0) - kbL + \frac{1}{2}Nkb^2, \quad (5)$$

where periodic boundary conditions in N and L are again understood. Because of the periodic boundary conditions, the choice of the equilibrium separation b simply shifts the ground-state potential, without changing the stationary distribution and equilibrium correlations. Nevertheless, the ground-state potential is dependent on L and it determines some thermodynamic properties, such as the pressure:

$$P_{\rm harm}(b) = k(b - L/N). \quad (6)$$

The system with $b = L/N$ satisfies $P_{\rm harm} = 0$.

In a periodic system, MD and the reversible Metropolis algorithm are strictly independent of b, as they only rely on the forces (identical derivatives of eqs (4) and (5) with respect to the x_i) or potential differences between configurations. However, the explicit form of the pairwise interaction influences the ECMC dynamics, as the factor potentials are treated independently. One such factor potential may thus contain the single term $U_{\rm harm}(x_{i+1} - x_i; b)$ with its explicit dependence on b. In the following we consider such factors between all neighboring pairs of particles. For $b = 0$, the harmonic interactions on particle i from its neighbors is attractive if $x_{i-1} < x_i < x_{i+1}$. It implies that for an active particle i with a positive displacement, the particle $i-1$ is likely to trigger the next event in ECMC (and to be the next active particle) (see Fig. 1a). The displacement δx per event is:

$$\delta x \sim \frac{T}{k(L/N-b)}, \text{ if } T \ll \frac{k}{2}(L/N-b)^2. \quad (7)$$

As b increases, the triggering probability is less biased and the displacement gets larger, and eventually reaches the maximum:

$$\delta x \sim \sqrt{\frac{2T}{k}}, \text{ when } b = L/N, \quad (8)$$

with symmetric triggering probabilities in both directions (see Fig. 1c).

At low T, the displacement per event in eq. (7) is much smaller than that in eq. (8). We expect that the case $b = L/N$ leads to larger amplitude movements of the active particle i, at the same time the transfer of activity is equally often toward $i+1$ and toward $i-1$, and characterizes the detailed ECMC dynamics. The case $b = L/N$ indeed gives rise to the exceptionally fast dynamics, characterized by $z = 1/2$ [21]. The aim of the present paper is to generalize this fast relaxation to arbitrary potentials.

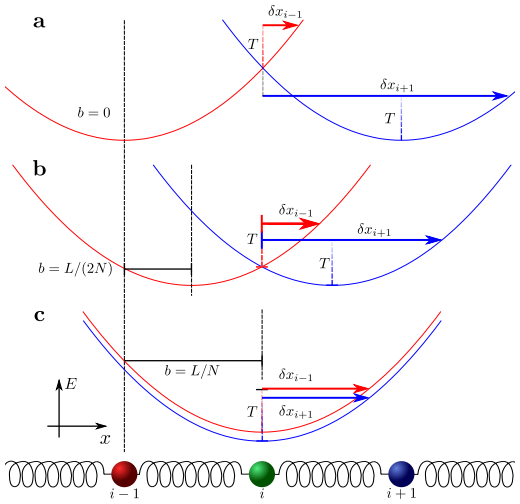


FIG. 1. ECMC dynamics in the harmonic model. Particle i is active, and the red and blue curves indicate the interactions with particles $i-1$ and $i+1$, respectively. The proposed moves are indicated by solid colored arrows, with respect to the temperature T (vertical dashed lines). **a**: For $b = L/N$, the interractions with particles $i-1$ and $i+1$ overlap. **b**: For $b = L/(2N)$, the potential with $i-1$ proposes a smaller displacement than $i+1$. Particle $i-1$ is more likely to be activated. **c**: For $b = 0$, the moves proposed by $i-1$ and by $i+1$ are balanced.

II. LINEAR LENNARD-JONES MODEL WITH ECMC

We study now the Lennard-Jones potential

$$U_{\rm LJ}(r) = \frac{1}{r^{12}} - \frac{1}{r^6}, \quad (9)$$

where $r = x_{i+1} - x_i$, with periodic booundary conditions. The minimum is $U_{\rm LJ}(r_{\rm min}) = -1/4$ for $r_{\rm min} = 2^{1/6}$, which sets a typical potential scale of the system $\epsilon = |U_{\rm LJ}(r_{\rm min})|$. In previous work [22] we proposed multiple factor sets within ECMC. Here we take into consideration two distinct factor sets $\mathcal{M}_{LJ}$ and $\mathcal{M}_{6+12}$: the former groups the terms $1/r^6$ and $1/r^{12}$ into a single Lennard-Jones factor, while the latter treats them separately as two factors which independently trigger events.

As the particles always move in the positive direction (x_i is always increasing), the active particle i, with the factor set $\mathcal{M}_{6+12}$, will either trigger the particle $(i+1)$ by the repulsive contribution $1/r^{12}$ or the particle $(i-1)$ by the attractive contribution $1/r^6$. The factor set $\mathcal{M}_{LJ}$ can lead to a trigger from $(i+1)$ or $(i-1)$ since the Lennard-Jones interaction has both increasing and decreasing branches.

In the following, we will show that the large-scale dynamics of ECMC are very sensitive to the choice of factor sets, even if all choices lead to the same equilibrium state. Good choices are crucial in the creation of efficient algorithms.

A. Simulations of 1D Lennard-Jones models

We simulate a slightly compressed ($P > 0$) linear Lennard-Jones model with periodic boundary conditions with average separation between particles equal to $\Delta = 1.06 < r_{\min}$ and use the reversible local Metropolis MCMC method, MD, as well as ECMC with the factor set $\mathcal{M}_{LJ}$ (see Fig. 2a). Metropolis MCMC is asymptotically the slowest method for $N \to \infty$: the autocorrelation time (measured in sweeps) increases as N^z with $z = 2$ characteristic of the diffusion of density fluctuations. MD is better behaved, due to the propagative compressional waves which more efficiently sample long-wavelength modes. MD is however disadvantaged by the necessity of using a small integration time step δt to stably explore the dynamics. The result from ECMC is very favorable, we see a low dynamic exponent ($z = 1$) combined with a small prefactor in the scaling: the algorithm makes a large leap (without systematic errors) for each iteration.

However, ECMC can also be less efficient than MD, in certain implementations (see Fig. 2b). Here we use the factor set $\mathcal{M}_{6+12}$, at low T. Here an analogous phenomenon occurs to that displayed in Fig. 1, in a form which is amplified by the splitting of the $1/r^6$ and $1/r^{12}$ contributions to the potential. The algorithm advances with the use of steps which are too small to efficiently explore the local environment. This slowdown of ECMC at low T was pointed out previously [30]. We now study analytically the Lennard-Jones interaction in eq. (9) at low T, and make contact with the harmonic model in order to eliminate this slowdown.

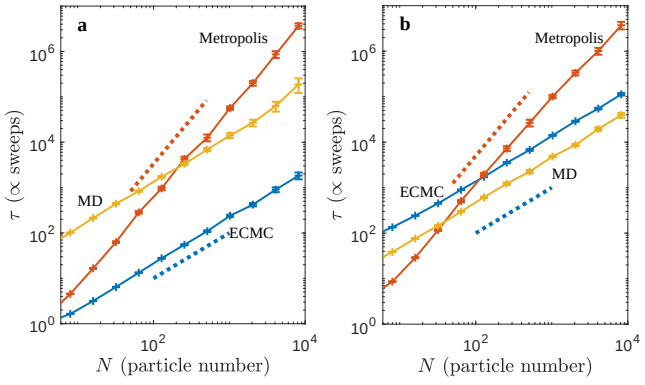


FIG. 2. Equilibrium autocorrelation times τ vs. system size N, for the 1D periodic Lennard-Jones model. Reversible local Markov-chain dynamics, ECMC with restarts and MD. **a**: $T/\epsilon = 10$, combined factors, $\mathcal{M}_{LJ}$. **b**: $T/\epsilon = 0.1$, separate factors, $\mathcal{M}_{6+12}$. Scalings $\tau \sim N$ and $\tau \sim N^2$ are indicated with dotted lines.

A straightforward expansion of the potential $U_{\rm LJ}(r)$ of eq. (9) to second order around a generic position $r = \Delta$

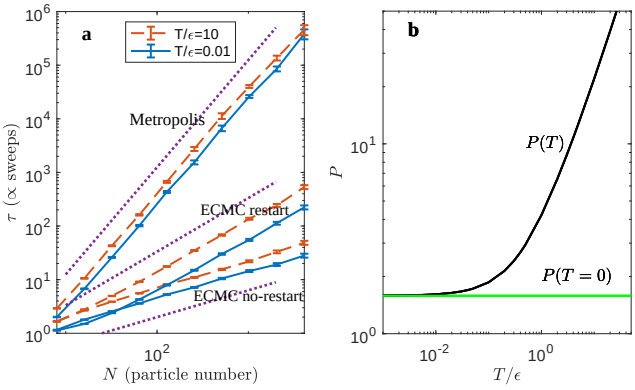


FIG. 3. Equilibrium autocorrelation timess τ vs. system size N, for the 1D periodic Lennard-Jones system with factor fields. **a**: Reversible local Metropolis dynamics and ECMC with and without restarts, at high T, $T/\epsilon = 10$ and low T, $T/\epsilon = 0.01$. Scalings $\tau \sim N^{1/2}$, $\tau \sim N$, and $\tau \sim N^2$ are indicated with dotted lines. **b**: Pressure $P(T)$ and its $T \to 0$ limit (see eq. (13)).

yields

$$U_{\rm LJ}(r) = U_{\rm LJ}(\Delta) - h_{\rm LJ}(\Delta)(r-\Delta) \tag{10}$$

$$+ \frac{1}{2}k_{\rm LJ}(\Delta)(r-\Delta)^2 + \dots . \tag{11}$$

This validates the (obvious) fact that the 1D Lennard-Jones model, in the limit $T \to 0$, is described by a harmonic model with, in analogy to eq. (5), a "stiffness"

$$\frac{1}{2}k_{\rm LJ}(\Delta) = \left(\frac{78}{\Delta^{14}} - \frac{21}{\Delta^8}\right) \tag{12}$$

and a linear coefficient

$$h_{\rm LJ}(\Delta) = -6(-2+\Delta^6)/\Delta^{13}. \tag{13}$$

Summed over the N pairs $(i, i+1)$ (with periodic boundary conditions), the constant ($U_{\rm LJ}(\Delta)$) and the first-order term in eq. (11) are without incidence on the constant-volume thermodynamics and the stationary distribution.

Analogously to eq. (5), we may add a temperature-dependent factor-field interaction

$$U_{\rm LJ}^{\rm fact} = h_{\rm LJ}^{\rm fact}(T)\sum_{i=1}^{N}(x_{i+1}-x_i) \tag{14}$$

to the total Lennard-Jones potential $\sum_i U_{\rm LJ}(x_{i+1}-x_i)$. The model defined by $U_{\rm LJ} + U_{\rm LJ}^{\rm fact}$ differs from the model given by $U_{\rm LJ}$ alone (in the presence of periodic boundary conditions), as the two have different pressures. Nevertheless, the samples obtained from the two associated Boltzmann distributions are the same, and therefore also all probability distributions and correlation functions at constant L. We choose a factor field to exactly compensate the linear term in the interaction in the limit $T \to 0$:

$$h_{\rm LJ}^{\rm fact}(T) = h_{\rm LJ}(\Delta) \quad \text{(for small } T\text{)}. \tag{15}$$

This clearly eliminates the inefficiencies of ECMC at low temperatures. In the model defined by eq. (15), the pressure P vanishes as $T \to 0$. Because of the connection between the pressure and the path statistics expressed in eq. (1), the ECMC trajectories are then without a drift term, and the expected displacement vanishes. As we now confirm numerically, we can speed up ECMC at arbitrary T by adopting a factor potential that exactly compensates for P.

We start by performing a set of short simulations to measure P from eq. (1) (see Fig. 3b). The function $P(T)$, thus obtained, recovers the $T \to 0$ limit. We then fix the value of the factor field in longer simulations to characterize the dynamics (see Fig. 3a). Indeed, both at low and at high T, ECMC remains efficient, and the dynamical exponent $z = 1/2$ corresponds to the harmonic model for $b = L/N$. This was tested for temperatures as high as $T/\epsilon = 10$ where the interactions for Lennard-Jones particles are dominated by the short-ranged repulsive core. The ansatz $h^{\text{fact}} = P$ for the factor field thus holds at temperatures at which the harmonic approximation of the potential no longer applies. Maximum efficiency is found for ECMC without restarts only: restarting the chain after $\sim N$ events leads to a larger dynamic exponent.

For the Lennard-Jones system, ECMC with factor fields requires finding roots to the equations

$$\frac{1}{r^{12}} - \frac{1}{r^6} \pm Pr = \Delta U. \tag{16}$$

We use the iterative Halley method [31], a higher-order generalization of Newton's method. It has the advantage of stability when starting an iteration near a stationary point of the function eq. (16). We start the iteration with a guess obtained with one of two methods. For small ΔE we make a harmonic approximation to the left-hand side of eq. (16). For large ΔE, the starting point is approximated as a root to the equation $1/r^{12} = \Delta E$. The iteration converges to machine precision within three iterations. The relative speeds shown in Fig. 3a account for this slow, iterative step through an appropriate proportionality factor for each algorithm. Alternatives to root finding may including thinning methods (as used in the cell-veto algorithm [32]) which compare rates derived from eq. (16) to an analytically tractable bound.

B. Extensions: Alternative factor sets

ECMC allows for many other choices for factors and also for lifting schemes. We may generalize the factor field method to the $\mathcal{M}_{6+12}$ factor set by introducing one factor field each for $1/r^6$ and for $1/r^{12}$ interactions (checking the convergence of the method for multiple correlation functions), but we did not explore fully the optimal choice of the two factor fields. We also studied factor sets which contain all the interactions of the model. This scheme is particularly interesting because the active particle i simultaneously explores the potential due to both $i-1$ and $i+1$, without the need for an explicit factor field. Again, this scheme uses an iterative solver. The full system factors also require a more sophisticated lifting scheme – generalizations of the "inside first" and "outside first" methods [22]. Particles with positive factor derivatives and particles with negative factor derivative are aligned in index order (see for instance [22, Fig. 10]). The lifting dynamics corresponds to the alignment of factors vertically. In such schemes, factors contain $\mathcal{O}(N)$ terms. Efficient alignment of the lifting diagram requires the use of a tree structure for bookkeeping with an effort $\mathcal{O}(\log(N))$. We found this method however to be less efficient than the factor field, and so do not report further on speed measurements.

III. FACTOR FIELDS FOR 1D HARD SPHERES

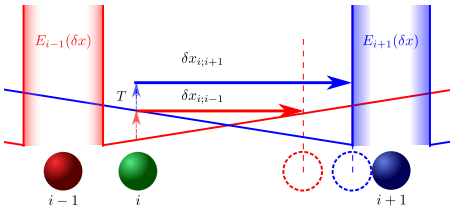


FIG. 4. ECMC dynamics with factor fields (indicated by inclined straight lines) for 1D hard spheres. The moves of the active sphere i proposed by the factors with $i+1$ and $i-1$ are indicated by horizontal arrows and dashed sphere positions. In the optimal dynamics, the slopes of the factor fields equal $\pm P$ in eq. (17).

To illustrate the generality of factor fields, we now consider their application to the 1D hard-sphere model, where the potential can no longer be expanded as a power series (as in eq. (11)). Nevertheless, the model has a well-defined pressure, that is computed from the partition function $Z = (L - N\sigma)^N$. This gives for the free energy, at temperature T, $F = -T \log Z = -NT \log L_{\text{free}}$, with $L_{\text{free}} = L - N\sigma$ and therefore [33]

$$P = \frac{NT}{L_{\text{free}}}. \tag{17}$$

A. Implementation, autocorrelation times

The implementation of ECMC with factor fields for hard spheres does not require numerical root finding: an active particle i, moving to the right, generates two possible events, a hard-sphere collision with the particle $i+1$ or else a trigger due to the factor field of particle $i-1$ (see Fig. 4). The latter path length is sampled from an

exponential distribution

$$\rho(x) = \frac{P}{T} \mathrm{e}^{-xP/T}. \tag{18}$$

The smaller of the two proposed displacements yields the next event, and it defines the lifting, as the new active particle is the one that has triggered the event. Irreducibility is guaranteed in the dynamics with an infinite event chain, and restarts are no longer needed, unlike for hard-sphere ECMC without the factor field.

We study the autocorrelation time in sweeps (see Fig. 5a) and compare with the reversible local Metropolis algorithm as well as ECMC without a factor field. Again, we note the acceleration brought by the addition of a factor field with a dynamic exponent $z = 1/2$, just as for the linear Lennard-Jones and the harmonic models. Non-optimal factor fields slow down the dynamics of the longest wavelength modes, an effect which becomes stronger for larger N (see Fig. 5b).

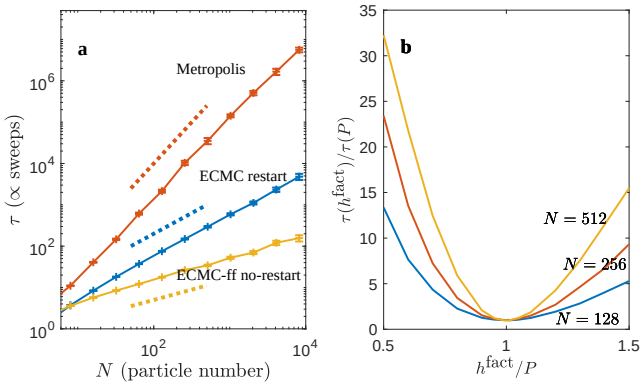


FIG. 5. Autocorrelation time τ (in sweeps) for 1D hard spheres. **a**: Reversible local Metropolis MCMC, ECMC (with restarts) and without factor field and ECMC (without restarts) with optimal factor field, h^{fact}. Scalings $\tau \sim N^{1/2}$, $\tau \sim N$, and $\tau \sim N^2$ are indicated with dotted lines. **b**: τ from ECMC *vs.* factor field (ECMC, without restarts).

B. Evaluation of mixing times

We have so far considered the equilibrium autocorrelation time, which is only one of the two relevant measures for the speed of an algorithm; it measures the time to move from one configuration (taken in equilibrium) to another independent one. The mixing time, in contrast, considers the time it takes to reach a first equilibrium configuration from an arbitrary non-equilibrium state. The scaling with N of the equilibrium autocorrelation time and of the mixing time differs for many MCMC algorithms in 1D particle systems (see [2] for a mathematical discussion of mixing times and equilibrium autocorrelation times, and [18] for a discussion in the context of ECMC.)

To determine the mixing time for the hard-sphere model with factor fields, we use a discretized version of the smallest Fourier coefficient of the structure factor in eq. (2), namely the variance var(w) of the "half-system distance"

$$w = x_{i+N/2} - x_i - N\sigma/2 \tag{19}$$

from a compact initial configuration [17] where var(w) $\propto N^2$ to the (exactly known) equilibrium value, which is $\propto N$. Tracking the variance signals a mixing time very close to N sweeps, a value that we conjecture to be exact (see Fig. 6). This is a faster scaling than the $\mathcal{O}(N \log N)$ sweep mixing-time behavior of ECMC with restarts (without factor field) [18].

Relaxation occurs in the following manner from a compact configuration: First, the active particle is driven to the right end of the system which over-relaxes (see Fig. 6). This drives the activity back into the bulk, to the boundary with the compact interior. A series of cycles of increasing amplitude relaxes the end of the system with penetration into the compact region following a law in $\sqrt{t}$. We note that the mixing time is longer than the equilibrium autocorrelation time (see the discussion in Section IV).

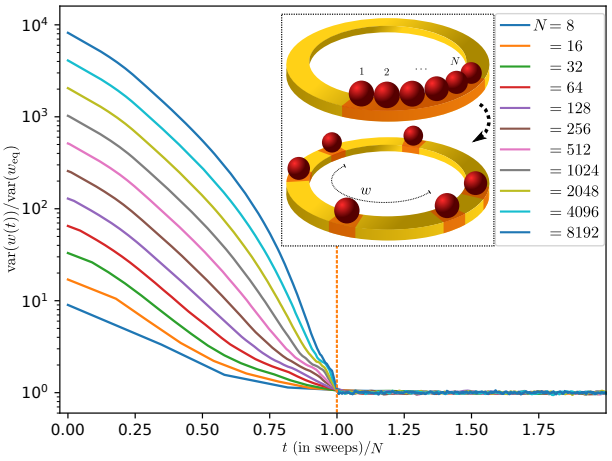


FIG. 6. Variance of the half-system distance w (see eq. (19)) *vs.* time for various N (hard-sphere model with factor fields (without restarts)). The observable relaxes to its equilibrium value at the mixing time $(1.000 \pm 0.005) \times N$ sweeps for the hard-sphere model with factor fields (without restarts).

IV. ACTIVE-PARTICLE DYNAMICS

The choice of factor fields, even if it is without incidence on spatial correlation functions and thermodynamic properties at constant L, strongly influences the ECMC dynamics. In this section, we consider the large-scale motion of the particle $i(t)$ that is active at time t, in order to probe how the exponent $z = 1/2$ arises from the

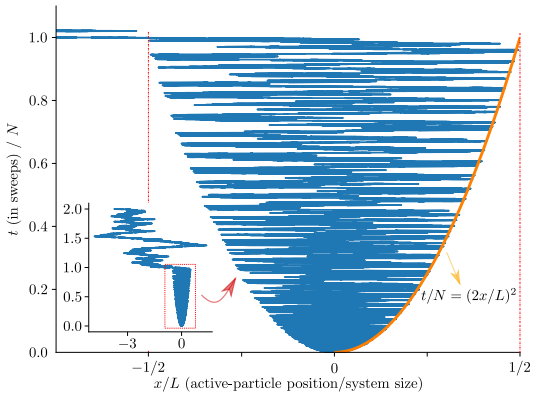


FIG. 7. Position x_i of the active sphere i *vs.* time t ($N = 4096, \sigma = L/(2N)$). At $t = 0$, the interval $[-L/2, 0]$ is close packed, the interval $[0, L/2]$ is empty. The physical extent expands through oscillations, growing as $\sqrt{t}$, and reaches $[x_1, x_N] \simeq L$ at $t \simeq N$. The inset illustrates the position of x_i (without periodic wrapping) on a larger time interval.

local active-particle dynamics. It is convenient to take into consideration discrete "event times" $s = 0, 1, 2, \ldots$, rather than the continuous time t of the Markov process. Because of the ordering of indices, we have $i(s+1) = i \pm 1$, with "event steps" $u(s)$ as $i(s+1) = i(s) + u(s)$ and $u(s) \in \{-1, 1\}$. It follows from eq. (1) that $\langle u \rangle > 0$ and $\langle u \rangle < 0$ for $P > 0$ and $P < 0$, respectively, which means that the ECMC trajectory is described by a forward drift (for $P > 0$) or a backward drift (for $P < 0$). With a factor field equal to P, the drift terms vanish, and ECMC trajectories feature positive and negative event steps (liftings $i(s+1) = i(s) + 1$ and $i(s+1) = i(s) - 1$) with equal probabilities [34]. To better characterize the time series $u(s)$ in this case, for both hard spheres and Lennard-Jones particles, we compute the event-step autocorrelation $\langle u(0)u(s) \rangle$ (see Fig. 8). We find that for large N, the autocorrelation decays as a power law:

$$\langle u(0)u(s) \rangle \sim s^{-\gamma}. \tag{20}$$

(This scaling applies on times shorter than those required to explore the whole system. On longer time scales the correlation in eq. (20) decays exponentially.)

The active particle at event time s (without periodic wrapping) is given by

$$i(s) = i(s = 0) + \sum_{s'=1}^{s} u(s'). \tag{21}$$

We now follow a trajectory which starts with $i(s = 0) = 0$. For vanishing long-range correlations in the event steps $u(s)$, the motion of the activity, characterized by the second moment of $i(s)$, would be diffusive ($\langle i^2(s) \rangle \sim s$. Rather, we find for large s, using eq. (20) with $\gamma < 1$:

$$\langle i^2(s) \rangle = \sum_{s'=1}^{s} \sum_{s''=1}^{s} \langle u(s')u(s'') \rangle \sim s^{2-\gamma}. \tag{22}$$

The position of the active particle is thus characterized by super-diffusive behavior. The observed value $\gamma = 2/3$ (see Fig. 8) implies

$$\langle i^2 \rangle \sim s^{4/3} \quad \text{or} \quad |i| \sim s^{2/3}. \tag{23}$$

The dynamics of the active particle has long-time memory for $N \to \infty$. The trajectories contain long runs separated by changes of the direction of motion, so that the average motion is undirected, as required by eq. (1).

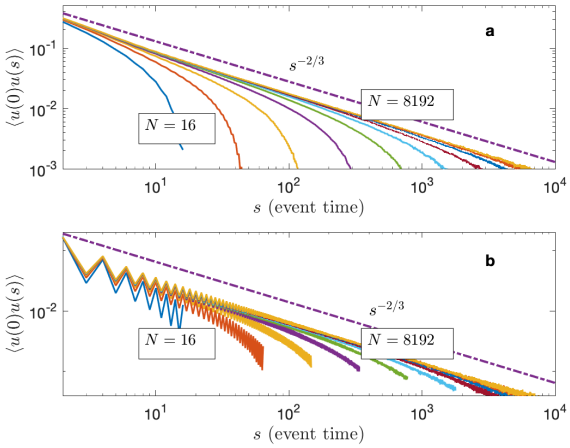


FIG. 8. Equilibrium autocorrelation of event steps $u \in \{-1, 1\}$ with event time s for ECMC with factor field (no restarts). **a**: 1D Lennard-Jones model shows monotonic decay. **b**: 1D hard spheres display oscillatory behavior with a power-law envelope. The scaling $\langle u(0)u(s) \rangle \sim s^{-2/3}$ is indicated with dotted lines (see eq. (20)).

A. Scaling for the active-particle dynamics

The discrete event time $s = 0, 1, 2, \ldots$ grows with the time t of the Markov process (that we measure in sweeps) as $s \propto Nt$. The same argument applies to the autocorrelation event time, in events, and the autocorrelation time τ, in sweeps, $s_{\text{auto}} \propto N\tau(N)$. The super-diffusive motion constrains the dynamic exponent z which relates complexity to system size:

$$s_{\text{auto}} \sim N^{(1+z)}. \tag{24}$$

A configuration can decorrelate from its previous history only if the super-diffusive walk visits each sphere at least once. Thus we require:

$$|i(\tau)| \sim {s_{\text{auto}}}^{1-\gamma/2} \sim N^{(1+z)(1-\gamma/2)} \geq N, \tag{25}$$

implying that $z \geq \gamma/(2-\gamma)$. If we take $\gamma = 2/3$, we find $z \geq 1/2$ compatible with the autocorrelation scaling reported previously for the harmonic model [18], and also compatible with the data in Figs 2 and 5.

A supplementary physical hypothesis of perfect local equilibration during the ECMC motion leads to a definite prediction for γ: the equalibrium fluctuations in particle separation in a system section of length $|i|$ increase as

$$\Delta x_p \sim |i|^{1/2}. \tag{26}$$

After s events the active label visits particles in a volume $|i| \sim s^{1-\gamma/2}$, so that on average each particle moves

$$\Delta x_\gamma = \frac{s}{|i|} \sim s^{\gamma/2} \sim |i|^{\gamma/(2-\gamma)} \tag{27}$$

times. If we assume that the motion of the particles is comparable to that required to resample the internal states of the section of length $|i|$ we find $\Delta x_p \sim \Delta x_\gamma$ so that $\gamma = 2/3$, and $z = 1/2$.

For this mechanism to work, the correlated random motion of the active particle must behave in a special way: both the mean and the standard deviation of the distribution of Δx_γ must have identical scaling with s. (If only the mean increases the spheres will be displaced uniformly without re-equilibrating the internal degrees of freedom.)

B. Active-particle return probabilities

The distribution of eq. (27)) allows for a rapid decay of autocorrelation functions. We consider the dynamics of a particle which is active at time $s = 0$. This particle can only move forward a large distance if the active label returns to it frequently, that is, if for many times s, one has $i(s) = i(s = 0)$. We thus study in greater detail the returns to the origin of the active label, in the presence of factor fields.

We generate an equilibrated configuration of the Lennard-Jones system and from the signal $i(s)$ calculate the distribution of the number n of returns to the origin within s events (see Fig. 9). For Brownian walks of length s, n is related to the "local time" [35, 36], and the local-time distribution $p(n, s)$ is half-gaussian defined for $n > 0$. In ECMC, the probability $p(n, s)$ of returns of the active-particle label to the origin (which gives the number of forward steps) is also maximum at zero, and decays monotonically with n. The mean and standard deviation of the number of steps drawn from such a distribution grow in the same way with s (see inset of Fig. 9). Even though the whole system moves forward in an ECMC simulation the dynamics is spatially heterogeneous. Widely separated particles move forward with different numbers of steps so that the internal modes of the system are efficiently resampled as is needed for the ansatz in eq. (27) to apply.

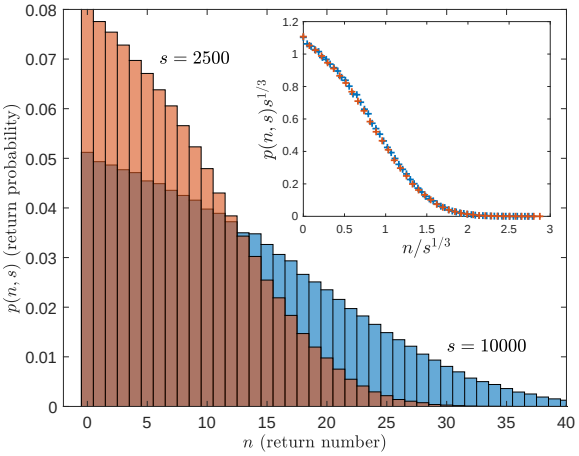


FIG. 9. ECMC with factor fields for a 1D Lennard-Jones system ($N = 8192$, $T/\epsilon = 1$). Probability $p(n, s)$ to return n times to the original active particle during s events (for $s = 2500$ and $s = 10000$). Mean and standard deviation of $p(n, s)$ both grow as $s^{\gamma/2} \sim s^{1/3}$ (see eq. (27)). Inset: data collapse using scaling variables $p(n, s)s^{1/3}$ *vs.* $n/s^{1/3}$.

V. CONCLUSIONS

We have compared in detail the dynamics of three simulation methods (reversible MCMC, MD and ECMC) for 1D systems with local interactions. We have shown that in many situations ECMC displays the same dynamic scaling ($z = 1$) as molecular dynamics. Both are asymptotically faster than the diffusive behavior found in MCMC ($z = 2$). With a good choice of factors, ECMC is much faster than MD, since it does not need to use a small integration time step to stably explore configurations. Furthermore, unlike MD, ECMC is exact to machine precision, as it is free from time-discretization errors.

Generalizing from the 1D harmonic model, we map 1D systems onto thermodynamically equivalent systems at zero pressure with periodic boundary conditions. This leads to further acceleration of ECMC for both smooth and discontinuous potentials. We have found in this case a remarkably low dynamic exponent ($z = 1/2$), better than MD. This acceleration is associated with a modification of the dynamics of the event steps (as a consequence of eq. (1)). Rather than displaying directed motion, the signal $i(s)$ becomes super-diffusive and optimally explores local density fluctuations, being driven forward in regions of high density, and back in regions of low density. A scaling hypothesis predicts a super-diffusive law of the form $\langle i^2(s) \rangle \sim s^{4/3}$ for the dynamics of the active label as well as an explanation for the emergence of the exponent $z = 1/2$.

There is a clear interest in generalizing these results to higher-dimensional models. Already a two and three-dimensional harmonic model has been shown [18] to display accelerated convergence in the ECMC algorithm. In

geometries which remain fixed, such as the XY model or fixed harmonic networks (without disorder) it appears possible to implement generalized factor fields. With fluctuating neighbor relations, for instance in a fluid, the generalization of factor fields will represent an interesting challenge [37].

ACKNOWLEDGMENTS

W.K. acknowledges support from the Alexander von Humboldt Foundation.

[1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1092, 1953.

[2] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*. American Mathematical Society, 2008.

[3] W. Krauth, *Statistical Mechanics: Algorithms and Computations*. Oxford University Press, 2006.

[4] B. J. Alder and T. E. Wainwright, "Phase Transition for a Hard Sphere System," *J. Chem. Phys.*, vol. 27, pp. 1208–1209, 1957.

[5] K. S. Turitsyn, M. Chertkov, and M. Vucelja, "Irreversible Monte Carlo algorithms for efficient sampling," *Physica D: Nonlinear Phenomena*, vol. 240, no. 4–5, pp. 410 – 414, 2011.

[6] Y. Sakai and K. Hukushima, "Dynamics of One-Dimensional Ising Model without Detailed Balance Condition," *Journal of the Physical Society of Japan*, vol. 82, no. 6, p. 064003, 2013.

[7] Y. Sakai and K. Hukushima, "Eigenvalue analysis of an irreversible random walk with skew detailed balance conditions," *Phys. Rev. E*, vol. 93, p. 043318, 2016.

[8] J. Bierkens and A. Bouchard-Côté and A. Doucet and A. B. Duncan and P. Fearnhead and T. Lienart and G. Roberts and S. J. Vollmer, "Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains," *Statistics & Probability Letters*, vol. 136, pp. 148–154, 2018.

[9] E. P. Bernard, W. Krauth, and D. B. Wilson, "Event-chain Monte Carlo algorithms for hard-sphere systems," *Phys. Rev. E*, vol. 80, p. 056704, 2009.

[10] M. Michel, S. C. Kapfer, and W. Krauth, "Generalized event-chain monte carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps," *J. Chem. Phys.*, vol. 140, no. 5, p. 054116, 2014.

[11] E. P. Bernard and W. Krauth, "Two-Step Melting in Two Dimensions: First-Order Liquid-Hexatic Transition," *Phys. Rev. Lett.*, vol. 107, p. 155704, 2011.

[12] 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," *Phys. Rev. E*, vol. 87, p. 042134, 2013.

[13] S. C. Kapfer and W. Krauth, "Two-Dimensional Melting: From Liquid-Hexatic Coexistence to Continuous Transitions," *Phys. Rev. Lett.*, vol. 114, p. 035702, 2015.

[14] M. Michel, J. Mayer, and W. Krauth, "Event-chain Monte Carlo for classical continuous spin models," *EPL*, vol. 112, no. 2, p. 20003, 2015.

[15] Y. Nishikawa, M. Michel, W. Krauth, and K. Hukushima, "Event-chain algorithm for the heisenberg model: Evidence for $z \simeq 1$ dynamic scaling," *Phys. Rev. E*, vol. 92, p. 063306, Dec 2015.

[16] M. Hasenbusch and S. Schaefer, "Testing the event-chain algorithm in asymptotically free models," *Phys. Rev. D*, vol. 98, p. 054502, 2018.

[17] S. C. Kapfer and W. Krauth, "Irreversible Local Markov Chains with Rapid Convergence towards Equilibrium," *Phys. Rev. Lett.*, vol. 119, p. 240603, 2017.

[18] Z. Lei and W. Krauth, "Mixing and perfect sampling in one-dimensional particle systems," *EPL*, vol. 124, no. 2, p. 20003, 2018.

[19] D. Randall and P. Winkler, "Mixing Points on an Interval," in *Proceedings of the Seventh Workshop on Algorithm Engineering and Experiments and the Second Workshop on Analytic Algorithmics and Combinatorics, ALENEX /ANALCO 2005, Vancouver, BC, Canada, 22 January 2005*, pp. 218–221, 2005.

[20] D. Randall and P. Winkler, "Mixing Points on a Circle," in *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques: 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th International Workshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22-24, 2005. Proceedings* (C. Chekuri, K. Jansen, J. D. P. Rolim, and L. Trevisan, eds.), pp. 426–435, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[21] Z. Lei and W. Krauth, "Irreversible Markov chains in spin models: Topological excitations," *EPL*, vol. 121, p. 10008, 2018.

[22] M. F. Faulkner, L. Qin, A. C. Maggs, and W. Krauth, "All-atom computations with irreversible Markov chains," *J. Chem. Phys.*, vol. 149, no. 6, p. 064113, 2018.

[23] A. C. Maggs, "Multiscale Monte Carlo Algorithm for Simple Fluids," *Phys. Rev. Lett.*, vol. 97, p. 197802, 2006.

[24] E. A. J. F. Peters and G. de With, "Rejection-free Monte Carlo sampling for general potentials," *Phys. Rev. E*, vol. 85, p. 026703, 2012.

[25] J. Harland, M. Michel, T. A. Kampmann, and J. Kierfeld, "Event-chain Monte Carlo algorithms for three- and many-particle interactions," *EPL*, vol. 117, no. 3, p. 30001, 2017.

[26] B. J. Alder and T. E. Wainwright, "Decay of the Velocity Autocorrelation Function," *Phys. Rev. A*, vol. 1, pp. 18–21, 1970.

[27] J. P. Wittmer, P. Polińska, H. Meyer, J. Farago, A. Johner, J. Baschnagel, and A. Cavallo, "Scale-free center-of-mass displacement correlations in polymer melts without topological constraints and momentum conservation: A bond-fluctuation model study," *J. Chem. Phys.*, vol. 134, no. 23, pp. 234901–234901, 2011.

[28] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations; 2nd ed.* Dordrecht: Springer, 2006.

[29] H. Flyvbjerg and H. G. Petersen, "Error estimates on averages of correlated data," *J. Chem. Phys.*, vol. 91, no. 1, pp. 461–466, 1989.

[30] Y. Hu and P. Charbonneau, "Clustering and assembly dynamics of a one-dimensional microphase former," *Soft Matter*, vol. 14, no. 20, pp. 4101–4109, 2018.

[31] T. R. Scavo and J. B. Thoo, "On the Geometry of Halley's Method," *The American Mathematical Monthly*, vol. 102, no. 5, pp. 417–426, 1995.

[32] S. C. Kapfer and W. Krauth, "Cell-veto Monte Carlo algorithm for long-range systems," *Phys. Rev. E*, vol. 94, p. 031302, 2016.

[33] L. Tonks, "The Complete Equation of State of One, Two and Three-Dimensional Gases of Hard Elastic Spheres," *Phys. Rev.*, vol. 50, p. 955, 1936.

[34] K. Kimura and S. Higuchi, "Anomalous diffusion analysis of the lifting events in the event-chain Monte Carlo for the classical XY models," *EPL*, vol. 120, no. 3, p. 30003, 2017.

[35] W. Feller, *An introduction to probability theory and its applications. Vol. I.* Third edition, New York: John Wiley & Sons Inc., 1968.

[36] C. Banderier and M. Wallner, "Local time for lattice paths and the associated limit laws," in *Proceedings of the 11th International Conference on Random and Exhaustive Generation of Combinatorial Structures (GASCom 2018), Athens, Greece, June 18-20, 2018. CEUR Workshop Proceedings* (L. Ferrari and M. Vamvakari, eds.), vol. 2113, pp. 69–78, CEUR Workshop Proceedings, 2018.

[37] D. C. Wallace, *Thermoelastic Theory of Stressed Crystals and Higher-Order Elastic Constants*, vol. 25, pp. 301–404. New York and London: Academic Press, 1970.

Résumé

Cette thèse étudie la chaîne de Markov irréversible dans les systèmes de spin et les systèmes de particules, explique théoriquement leurs spécialités dynamiques, propose une amélioration des méthodes de Monte Carlo en ce qui concerne les propriétés systématiques.

Les deux premiers chapitres examinent la théorie des probabilités, la chaîne de Markov et la méthode de Monte Carlo. La chaîne de Markov irréversible, avec le schéma de «lifting» et le filtre Metropolis factorisé, augmente la vitesse de mélange à une échelle supérieure dans de nombreux modèles.

Le troisième chapitre étudie le modèle de la sphère dure. À partir du résultat exact obtenu à partir du modèle unidimensionnel dans la limite continue, l'algorithme de «event-chain» est lié au problème du collecteur de coupons dans l'évaluation du temps de mélange. Un algorithme séquentiel de «event-chain» est proposé pour accélérer le processus de mélange, ce qui est également valable dans les cas plus généraux de dimensions supérieures. Pour les algorithmes Metropolis plus généraux avec «lifting», leur croisement avec l'algorithme de «event-chain» est discuté.

Le quatrième chapitre présente la dynamique de la chaîne de Markov irréversible pour les modèles de spin continu utilisant le filtre de Metropolis, en présence d'excitations topologiques. La nature locale de la dynamique de la chaîne de Markov conduit à un mode lent de vortex et à un mode rapide d'onde de spin dans le modèle XY. La corrélation à l'équilibre est quantifiée pour varier de $z \approx 2$ à la température critique à $z \approx 0$ à la limite de basse température, et l'influence respective sur le modèle tridimensionnel de Heisenberg est également décrite.

Le cinquième chapitre, basé sur la connaissance des deux chapitres précédents, propose une optimization du filtre de Metropolis pour les modèles de particules généraux, en introduisant un champ auxiliaire. Les simulations sur une chaîne unidimensionnelle de Lennard-Jones montrent une accélération évidente en tant que mode de spin-wave. D'autres études vérifient le comportement super-diffusif de l'algorithme de «event-chain», ce qui peut expliquer la vitesse haute de la dynamique.

Abstract

This thesis studies the irreversible Markov chain in the spin systems and particle systems, theoretically explains their dynamical specialties, proposes an improvement to the Monte Carlo methods with respect to the systematic properties.

The first two chapters review the probability theory, Markov chain and Monte Carlo method. The irreversible Markov chain, with the "lifting" scheme and factorized Metropolis filter, increases the mixing speed at a higher scale in many models.

The third chapter studies the hard sphere model. From the exact result obtained from the one-dimensional model in the continuous limit, the "event-chain" algorithm is related to the coupon-collector problem, in the evaluation of mixing time. A sequential "event-chain" algorithm is proposed to accelerate the mixing process, which is also valid in more general cases of higher dimensions. For more general Metropolis algorithms with "lifting", their crossover with the "event-chain" algorithm is discussed.

The fourth chapter presents the dynamics of the irreversible Markov-chain for continuous spin models using Metropolis filter, in the presence of topological excitations. The local nature of the Markov-chain dynamics leads to a slow vortex mode and a fast spin-wave mode in the XY model. The equilibrium correlation varies from $z \approx 2$ at the critical temperature to $z \approx 0$ at the low temperature limit, and the respective influence on three-dimensional Heisenberg model is also described.

The fifth chapter, based on the knowledge of the previous two chapters, proposes an optimization of Metropolis filter for general particle models, by introducing an auxiliary field. Simulations on one dimensional Lennard-Jones chain exhibit an obvious acceleration as the spin-wave mode. Further studies verify a super-diffusive behavior of the "event-chain" algorithm, which may explain the high speed of the dynamics.

Résumé

Cette thèse étudie les chaînes de Markov irréversibles pour les modèles de spin et les systèmes de particules. Il analyse leur convergence vers l'équilibre dans un certain nombre de cas et propose de nouveaux algorithmes aux propriétés améliorées. Les deux premiers chapitres examinent certains aspects de la théorie des probabilités et de la théorie des chaînes de Markov. Une chaîne de Markov irréversible particulière, utilisant le concept de « lifting » et employant le filtre factorisé Metropolis, est discutée. On constate qu'il accélère le mélange dans de nombreux modèles. Le troisième chapitre étudie les chaînes de Markov irréversibles pour le modèle 1D à sphère-dure. Nous obtenons un résultat exact dans un cas de « event-chain » continu. Nous associons ce problème et ses solutions au problème des « collecteurs de coupons ». Il a été prouvé qu'un nouvel algorithme séquentiel « event-chain » accélérait encore le mélange, et reste également valable dans les dimensions supérieures. Pour les algorithmes Metropolis plus généraux avec « lifting », leur croisement avec l'algorithme « event-chain » est discuté. Le quatrième chapitre présente la dynamique des chaînes de Markov irréversibles pour les modèles de spin continu utilisant le filtre factorisé Metropolis, en présence d'excitations topologiques. La nature locale de la dynamique de la chaîne de Markov conduit à un mode lent de « vortex » et à un mode rapide de « l'onde de spin » dans le modèle 2D de XY. La corrélation à l'équilibre varie de $z \approx 2$ à $T = T_c$ à $z \approx 0$ à $T \approx 0$. Le cas du modèle 3D de Heisenberg est également décrit. Le cinquième chapitre propose une optimisation du filtre factorisé Metropolis pour les modèles de particules généraux, en introduisant un champ moléculaire de compensation. Les simulations sur des chaînes de Lennard-Jones unidimensionnelles indiquent une accélération considérable. On pense qu'un comportement super-diffusif de l'algorithme «chaîne d'événements» explique la vitesse élevée de la dynamique.

Mots clés

Méthode de Monte Carlo, Chaîne de Markov irréversible, Modèle XY, Excitations topologiques, Modèle des disques durs, Modèle de Lennard-Jones

Abstract

This thesis studies irreversible Markov chains for spin models and particle systems. It analyzes their convergence towards equilibrium in a number of cases, and it proposes new algorithms with improved properties. The first two chapters review some aspects of probability theory, and of the theory of Markov chains. A particular irreversible Markov chain, making use of the "lifting" concept and employing the factorized Metropolis filter is discussed, which speeds up the mixing in many models. The third chapter studies irreversible Markov chains for the 1D hard-sphere model. We obtain an exact result in a continuous "event-chain" case. We relate this problem, and its solutions, to the coupon-collector problem. A new sequential "event-chain" algorithm further accelerates mixing, and remains valid in higher dimensions. For general Metropolis algorithms with "lifting", their crossover with the "event-chain" algorithm is discussed. The fourth chapter presents the dynamics of the irreversible Markov-chains for continuous spin models using the factorized Metropolis filter, in the presence of topological excitations. The local nature of the dynamics leads to a slow "vortex" mode and a fast "spin-wave" mode in the 2D XY model. The correlation varies from $z \approx 2$ at $T = T_c$ to $z \approx 0$ at $T \approx 0$. The case of the 3D Heisenberg model is also described. The fifth chapter proposes an optimization of the factorized Metropolis filter for general particle models, by introducing a compensating molecular field. Simulations on one-dimensional Lennard-Jones chains indicate a considerable acceleration. A super-diffusive behavior of the "event-chain" algorithm is thought to explain the speed of dynamics.

Mots clés

Monte Carlo method, Irreversible Markov chain, XY model, Topological excitations, Hard-disk model, Lennard-Jones model