

Irreversible Markov chains by the factorized Metropolis filter : algorithms and applications in particle systems and spin models

Manon Michel

▶ **To cite this version:**

Manon Michel. Irreversible Markov chains by the factorized Metropolis filter : algorithms and applications in particle systems and spin models. Statistical Mechanics [cond-mat.stat-mech]. Université Paris sciences et lettres, 2016. English. NNT : 2016PSLEE039 . tel-01394204v2

HAL Id: tel-01394204

https://theses.hal.science/tel-01394204v2

Submitted on 21 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

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

Préparée au laboratoire de physique statistique
à l'École Normale Supérieure de Paris

Irreversible Markov Chains by the Factorized Metropolis Filter: Algorithms and Applications in Particle Systems and Spin Models

Ecole doctorale n°564

École doctorale Physique en Île de France EDPIF

Spécialité Physique

COMPOSITION DU JURY :

Mme. DIJKSTRA, Marjolein
Universiteit Utrecht, rapporteur

M. MOULINES, Éric
Télécom Paritech, rapporteur

M. MAGGS, Anthony
ESPCI, examinateur

M. MOORE, Cristopher
Santa Fe Institute, examinateur

M. ZAMPONI, Francesco
ENS Paris, examinateur

M. KRAUTH, Werner
ENS Paris, directeur de thèse

Soutenue par Manon **MICHEL**
le 17 octobre 2016

ENS ÉCOLE NORMALE SUPÉRIEURE

Département de Physique
École normale supérieure

PSL RESEARCH UNIVERSITY PARIS

École Doctorale de Physique de la Région Parisienne - EDPIF 564
École normale supérieure





Laboratoire de physique statistique

Thèse de doctorat

Irreversible Markov chains by the factorized Metropolis filter: Algorithms and applications in particle systems and spin models

---

présentée par Manon Michel

pour obtenir le titre de docteur de l'École normale supérieure

Soutenue le 17 octobre 2016 devant le jury composé de:

Marjolein Dijkstra	Rapporteur
Werner Krauth	Directeur de thèse
Anthony Maggs	Examinateur
Cristopher Moore	Examinateur
Éric Moulines	Rapporteur
Francesco Zamponi	Examinateur

À mes parents, Hélène et Max,
pour leur affection et soutien tout au long de ces années.

Remerciements

Je remercie Werner Krauth pour la confiance qu'il m'a apportée en acceptant de diriger ce travail doctoral, pour m'avoir guidée et encouragée pendant plusieurs années et pour avoir relu avec beaucoup de soin ce manuscrit. Je remercie également Florent Krzakala pour avoir accepté d'être le parrain de cette thèse. Je remercie Marjolein Dijkstra, Éric Moulines, Anthony Maggs, Cristopher Moore et Francesco Zamponi d'avoir accepté d'évaluer ce travail, en particulier les deux premiers pour en être les rapporteurs.

Pour les collaborations fructueuses et les nombreuses discussions scientifiques, je remercie Sebastian Kapfer, avec lequel c'est un véritable plaisir de travailler ou de participer à des conférences. Je remercie également les différentes personnes avec lesquelles j'ai pu collaborer: Johannes Mayer pour les systèmes de spins XY, Yoshihiko Nishikawa et Koji Hukushima pour les systèmes de spins Heisenberg.

Je remercie également les professeurs qui m'ont poussée dans mon intérêt pour la recherche; Michel Faye pour tous ses efforts pour nous transmettre l'intuition de la physique dans ce qu'elle a d'essentiel, qu'elle soit cachée dans des trébuchets ou des transits de Vénus; feu Prebagaran Mottou, pour sa patience à l'égard de mes questions, ses examens toujours plus exigeants à base d'aventures de Bernard et Bianca et ses cours ambitieux qui me plaisaient tant; Jean-Pierre Sanchez pour la rigueur et les bases solides qui ne m'ont plus quittée; Nicolas Schlosser pour son enthousiasme, le ski et simplement pour avoir cru en moi; feu Jean-Daniel Bloch pour son authenticité et ses fameux jets de craie; Gérard Mantin, merci de m'avoir tant envoyée au tableau et de ne pas avoir accepté moins que ce que vous me croyiez capable, je ne pourrais vous offrir assez de Michokos pour vous remercier; Dominique Zann pour sa disponibilité, sa patience et pour m'avoir fait douter entre la physique et la chimie, c'est dire; Stéphane Olivier pour ses cours incroyables, les expériences avec le Michelson et sa patience face à mon cahier de brouillon. Une fois arrivée à l'ENS, je remercie Jean-François Allemand et Frédéric Chevy pour avoir créé un environnement d'apprentissage très stimulant. Enfin, je remercie les différents professeurs de physique statistique que

j'ai eu la chance de pouvoir écouter sur les bancs de l'École: Michel Bauer, Bernard Derrida, Emmanuel Trizac, Stéphane Fauve, Alain Comtet, Werner Krauth et Henk Hilhorst.

Ces dernières années, j'ai eu la chance d'enseigner à l'université Pierre et Marie Curie à des élèves motivés et charmants. Je remercie ainsi Jean-Louis Cantin, Christophe Prigent et Frédéric Daigne pour m'avoir fait confiance dans cette tâche, pendant laquelle j'ai tant appris.

Je remercie Éric Pérez et Jorge Kurchan, directeurs successifs du laboratoire de physique statistique, pour m'avoir accueillie au sein de cette institution et pour avoir fait en sorte que je puisse mener ma tâche de représentant des doctorants librement et efficacement. Je remercie également Zaïre Dissi pour son aide technique et pour avoir sauvé mon premier ordinateur de la peste bubonique. Pour leur bienveillance et patience, je remercie Marie Gefflot, Annie Ribaudeau, Nora Sadoui et Benoît Paulet. Sans leur assistance dans les démarches administratives, ces années auraient été encombrées de problèmes moins scientifiques et auraient compté moins de moments conviviaux autour des cafés du jeudi.

Pour avoir créé une atmosphère de travail joviale et humaine, je remercie tout d'abord mes compagnons les plus proches dans cette aventure de Monte Carlo: Tommaso Comparin, le grand sage du laboratoire, Ze Lei, le grand pourvoyeur de snack chinois, et Juliane Klamser, le nouveau soleil de l'équipe. Je souhaite à ces deux derniers beaucoup de succès pendant leur doctorat et à Tommaso du courage pour la fin de la rédaction. Je remercie tout aussi chaleureusement les autres doctorants et postdoctorants avec lesquels j'ai eu le plaisir de partager ces moments de notre jeunesse: Étienne Bernard et Swann Piatecki, les grand frères et anciens membres de notre équipe; Isabelle Motta, Benjamin Ravaux, Fabien Souris et Romain Lhermerout avec lesquels j'ai eu le plaisir d'organiser le séminaire PhD & PostDoc, la journée PhD & PostDoc et bien sûr les randonnées ! Quentin Feltgen et Nariaki Sakai, qui ont repris le flambeau de l'organisation du séminaire PhD & PostDoc; Tridib Sadhu pour la musique et la gentillesse de laisser son appartement se faire envahir de doctorants lors de nos fameux dîners; Anirudh Kulkarni pour sa bonne humeur et son oreille attentive; mais aussi les autres compagnons de ses moments joyeux, Diego Contreras pour son amour des vélos et du voyage, Volker Pernice pour les cafés et les couvertures, Lorenzo Posani pour son énergie et son humour, Samuel Poincloux pour les promenades dans Lyon et les tricots, Riccardo Rossi pour sa force tranquille et sa rapidité au pot et Arthur Prat-Carrabin pour ses conseils et sa bonne humeur; je n'oublie pas celui qui a été l'un des doctorants incontournables du laboratoire, Tommaso Brotto, merci de t'être occupé de toutes ces tâches que les autres préféraient éviter et surtout d'avoir cimenté le groupe des doctorants; je remercie également ma partenaire des heures de travail nocturnes, Gaïa Tavoni; je remercie enfin Kris Van Houcke pour sa tempérance et sa sagesse et Félix Werner pour nos discussions sur la randonnée. Je remercie une seconde fois ceux qui ont accepté de relire cette thèse, Tridib Sadhu, Anirudh Kulkarni et Juliane Klamser. Merci pour votre patience à l'égard de mes tournures un brin françaises.

Je remercie tous mes amis qui ont su m'offrir des intermèdes heureux, avec plus ou moins de péripéties, entre deux phases de recherche; les nimpros bien sûr: Tiphaine, Pierre, Sévan, Paul, Nicolas, Anne-Sophie, Guillaume, Marion, Violaine, Tania, Julian, Antoine, François, Fathi, Marie, Samuel, Guillaume, Cassandre, Félix, Victor et Anouck; les physiciens: Claude, Stéphanie, Gabriel, Éric, Lucie et Laure; les patineurs: Marine, Sary et Guillaume; les cavaliers: Manon, Mathilde et Gérald; les amis tout simplement: Swann, Katia et Hélène. Merci de m'honorer de votre amitié.

Finalement, je remercie ma famille et en particulier ceux qui ont toujours été là, mes parents, Hélène et Max. Merci de vous être tant souciés de mon épanouïssement et merci de m'avoir appris non pas la facilité mais la persévérance et l'effort. Merci de votre aide dans les dernières heures de rédaction, les juments et moi-même vous remercions tout particulièrement pour les soins que vous leur avez prodigués, quand je n'étais pas disponible. Je remercie également Alix pour avoir partagé tant de soirées, d'histoires et de kilomètres avec moi.

General introduction

In nature, there exist many systems composed of a great number of objects interacting with each other, such as atomic gases, flock of birds, neural network or economic systems. Composed of many degrees of freedom, those systems exhibit interesting collective phenomena, even if the microscopic interactions are only local. In particular, during second-order phase transitions, the correlations between the objects appear on an infinite scale and the system moves as one [1]. Statistical mechanics offers a powerful framework to describe such systems, by assuming that, at equilibrium, each system visits in a stochastic way every one of its microscopic states during its evolution [2–6]. Building on the theory of probabilities, equilibrium statistical mechanics describes complex systems by enumerating all their possible microscopic configurations and the probability to find the system in a particular one. The successes of this approach extend now well beyond physics, from classical to quantum physics and astrophysics, and reach a broad range of topics from biology and neuroscience to social sciences or finance, with direct applications, such as deep learning and big data.

The description of a system by its configuration space and their probabilities corresponds analytically to solving high-dimensional integrals, which often turns out to be an impossible task. The advent of the Monte Carlo method [7–9] has revolutionized the computation of these high-dimensional integrals, and therefore the understanding of complex systems [10–12]. This method runs a Markov chain on the configuration space. As the chain explores this space, it samples the configurations with the correct probability, once it reaches the stationary regime, called equilibrium. It then computes the needed integrals as a configurational average over the samples. Deriving also from probability theory and used with success in the same various range of domains from physics [13, 14] to biology [15] and computer science [16], it is now hard to differentiate the tool from the discipline it was meant to study. But, in spite of celebrated past successes, the Monte Carlo method faces today new challenges. The seminal paradigm introduced by Metropolis et al., (1953) [17] imposes strict rules, that rely on the rejection of moves and often lead to the diffusive dynamics of the random

walk. On complex energy landscapes with numerous metastable states (e.g. in spin glasses) or for probability distributions exhibiting a large number of bottlenecks (e.g. in dense hard spheres), the equilibration time becomes often larger than the available simulation time, while reaching equilibrium is a crucial condition for retrieving the correct probability distribution. The extension of the Metropolis algorithm by Hastings (1970) [18] led to the development of new methods which are used with success in spin systems [19, 20], but are efficient in specific systems only. In order to reduce the diffusive dynamics of the local update Metropolis algorithm, algorithms producing global moves were designed [21–25]. The concept of lifting was then created to unify these approaches [26–28].

During my PhD at Laboratoire de physique statistique at École Normale Supérieure, I worked on the conception and study of irreversible Markov-chain Monte Carlo algorithms and their application to particle and spin systems. This thesis reviews the work done under the direction of Werner Krauth. I had the great pleasure to collaborate with Werner Krauth, Sebastian Kapfer (soft spheres and elasticity), Johannes Mayer (XY spins model), Yoshihiko Nishikawa and Koji Hukushima (Heisenberg spins). As this work is at the junction between physics and mathematics, particular attention has been paid to make this thesis accessible to both communities. In his lecture notes on statistical mechanics [29], R. P. Feynman introduced the Boltzmann distribution and the expected value of an observable over it by stating that *'this fundamental law is the summit of statistical mechanics, and the entire subject is either the slide-down from this summit, as the principle is applied to various cases, or the climb-up to where the fundamental law is derived and the concepts of thermal equilibrium and temperature T is clarified. We will begin by embarking on the climb.'*. In a more humble way, this thesis offers the climb-up to the fundamental concepts of statistical mechanics and, from there, a path to slide down by using stochastic numerical computations by Monte Carlo method, instead of traditional deterministic calculations.

First, Chapter 1 introduces essential concepts of statistical mechanics and Markov chains. Then, Chapter 2 reviews the different Monte Carlo methods and the limitations encountered by traditional schemes, in particular around phase transitions, where dynamical slowing down appears. The rest of the thesis is dedicated to the results that I obtained during my thesis and that are summarized in three publications [30–32], which are attached in this thesis, see Publication 1, Publication 2 and Publication 3, and two manuscripts in preparation [33, 34]. Chapter 3 presents a general scheme for rejection-free irreversible Markov chains, that relies on the new factorized Metropolis filter. These irreversible factorized Metropolis algorithms draw on the lines of previous works on upgrading the Metropolis dynamics [24, 25] and build on the lifting framework developed in mathematics [26–28]. These algorithms violate the Metropolis paradigm, as they do not obey detailed balance, they proceed by global moves in a persistent way and are implemented using infinitesimal steps. The applications of this irreversible factorized Metropolis scheme, its implementation and performance in soft-sphere and continuous spin systems are discussed in Chapter 4.

Contents

General introduction **11**

Contents **13**

1 Markov processes **17**
1.1 Fluctuations and statistical ensembles 18
1.1.1 Maxwell-Boltzmann distribution 18
1.1.2 Classical statistical mechanics and ensemble averages 20
1.1.3 Stochastic processes 23
1.2 Markov processes and master equation 26
1.2.1 General properties of a Markov process 27
1.2.2 Stationary and homogeneous Markov processes 28
1.2.3 Master equation 31
1.3 Convergence in Markov processes 33
1.3.1 Ergodic and mixing properties 33
1.3.2 Mixing in Markov chains 35

2 Monte Carlo Method **43**
2.1 Direct sampling for high-dimensional integration 44
2.1.1 Weak and strong laws of large numbers and the Monte Carlo method 45
2.1.2 Inversion sampling 47
2.1.3 Acceptance-rejection method 50
2.1.4 Importance sampling 52
2.2 Markov-chain Monte Carlo method 53
2.2.1 Principles of the Markov-chain Monte Carlo method 54
2.2.2 Metropolis algorithm 56
2.2.3 A priori probabilities 57

2.2.4 Heat-bath algorithm . . . 58
2.3 Convergence and error . . . 60
2.3.1 Initialization bias . . . 60
2.3.2 Uncorrelated samples . . . 61
2.3.3 Correlated samples and autocorrelation times . . . 62
2.3.4 Scaling of autocorrelation times around a phase transition . . . 64
2.4 Cluster algorithms . . . 67
2.4.1 Fortuin-Kasteleyn transformation . . . 67
2.4.2 Swendsen-Wang algorithm . . . 69
2.4.3 Wolff algorithm . . . 69
2.4.4 Cluster algorithms for spin glasses . . . 70
2.5 Non-local algorithms . . . 71
2.5.1 Geometric cluster algorithm . . . 72
2.5.2 Jaster algorithm . . . 73

3 Irreversible factorized Metropolis algorithm 75
3.1 The Lifting Framework . . . 77
3.1.1 Lifting for the unidimensional random walk . . . 77
3.1.2 General case . . . 79
3.1.3 Lifting for two particles . . . 84
3.2 The Factorized Metropolis filter . . . 87
3.2.1 Definition of the Factorized Metropolis filter . . . 88
3.2.2 Implementation . . . 90
3.2.3 The factorized Metropolis filter in the lifting framework, beyond ± replicas . . . 92
3.3 Continuous-time scheme . . . 93
3.3.1 Maximal global-balance by infinitesimal steps . . . 93
3.3.2 Event-driven approach: Event sampling . . . 96
3.3.3 Observable averaging . . . 101
3.4 Infinite chains . . . 101

4 Applications of the irreversible factorized Metropolis algorithm 107
4.1 Soft-sphere systems . . . 108
4.1.1 Melting in soft spheres . . . 108
4.1.2 Performance . . . 109
4.2 Pressure computation . . . 111
4.3 Continuous classical spin systems . . . 114
4.3.1 Planar rotator spin systems . . . 115
4.3.2 Heisenberg spin systems on a three-dimensional lattice . . . 122

General conclusion 125

Bibliography 127

Publications **139**
Publication 1 139
Publication 2 150
Publication 3 158

CHAPTER 1

Markov processes

In 1738, D. Bernoulli published *Hydrodynamica* [35] that dealt with hydrodynamics, also known as fluid mechanics, and was organized around the idea of conservation of energy. Even in modern physics, the atomistic nature of fluids is ignored in hydrodynamics. A fluid is described using a mesoscopic scale which is based on the possibility of subdividing the fluid into infinitesimal elements. The averaged properties of these elements are those of the macroscopic material, such as pressure, density or temperature. Such an assumption asks for a good separation of scales between the macroscopic and microscopic level. This mesoscopic description assumes the conservation of mass, energy and momentum. This link between a mesoscopic description and conservation laws appears also in statistical mechanics.

In the same book, D. Bernoulli also laid the basis for the kinetic theory of gases by arguing that gases consist of a great number of molecules moving in all directions and that the macroscopic notion of pressure and heat find their origin in their motions. It is almost two centuries later that the existence of atoms and molecules was proven by the study of the Brownian motion by A. Einstein in 1905 [2] and observed experimentally by J. Perrin in 1908 [36], as the Brownian motion of pollen grains or dust particles in a liquid results directly from collisions with the liquid molecules. In 1857, R. Clausius [37] introduced a more sophisticated version of the theory, with translational, rotational and vibrational degrees of freedom. In 1859, J. C. Maxwell [3, 4] formulated the Maxwell distribution of molecular velocities. Finally, in 1871, L. Boltzmann [5, 6] generalized Maxwell's distribution to the Maxwell-Boltzmann distribution, which initiated statistical mechanics.

A. Einstein described the Brownian motion as a memoryless stochastic process, a Markov process. This stochastic description reflects the fact that microscopic information is converted over time into fluctuations within the system on the macroscopic scale [38–40]. The loss of the information about the initial conditions, as the system is able

to reach the equilibrium, will be of particular interest, especially in the case of Markov processes, that are at the core of the Markov-chain Monte Carlo method.

In this chapter, the ensemble average over the Maxwell-Boltzmann distribution is introduced in Section 1.1.1. The equivalence of time average and ensemble average, the foundation of equilibrium statistical mechanics, is then discussed in Section 1.1.2, along with the description of a system by means of a stochastic process in Section 1.1.3. Among stochastic processes, we will focus on the memoryless Markov processes, in Section 1.2, and discuss their speed of convergence to equilibrium in Section 1.3.

1.1 Fluctuations and statistical ensembles

A gap exists between the reversible laws of classical mechanics on the microscopic scale and the irreversible phenomenological laws on the macroscopic scale. Computing the exact variations of the positions and momenta of a large number of atoms appears to be an impossible task, even if it is in principle possible. Macroscopic observables, however, may behave smoothly and be described by simple laws, and such, without knowing the microscopic details. This section addresses how statistical mechanics solves this issue by reformulating drastically any single varying function of time $f(t)$ by an ensemble of functions, i.e. a stochastic process $\{f_1, f_2, f_3, \dots\}$. All averages over some time interval $\bar{f}$ are then replaced by averages over the ensemble $\langle f \rangle$.

1.1.1 Maxwell-Boltzmann distribution

We introduce here the basic laws of statistical mechanics, first for a quantum mechanical system, for simplicity, as it can be decomposed into a mixture of eigenvectors of the Hamiltonian H. A more extended review can be found in [29, chapter 1].

We consider a system with a Hamiltonian H, whose eigenvector $|i\rangle$ of energy E_i, called eigenstate, follows

$$H|i\rangle = E_i|i\rangle. \tag{1.1}$$

Systems of interest to statistical mechanics, e.g. often composed of $\mathcal{O}(10^{23})$ particles, exhibit an extremely large degeneracy of energy levels. We denote by $\Omega(E, V, N)$ the number of eigenstates with energy E of a system of N particles in a volume V.

The basic, but non trivial, assumption of statistical mechanics is the equiprobability of any eigenstate of $\Omega(E, V, N)$. It can be justified by considering a system S weakly coupled to a heat bath HB for a long time, as is done in [29]. The system is in thermal equilibrium, i.e. all the fast phenomena have happened but the slow ones, e.g. the erosion of the enclosure of the gas, have not. We consider two different states of the system of same energy $E_r = E_s$. If the system is in state r, any extremely small perturbation will cause the system to go into a different state of essentially the same energy, such as s. Reciprocally, the same holds true if the state is in s. As the system is in contact with the heat bath for a long time, one would expect states of equal energy

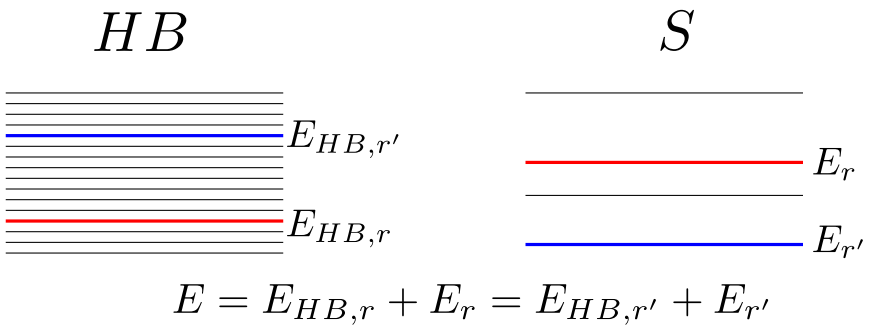


Figure 1.1: Quasi-continuous energy levels of the heat bath HB in comparison to the energy levels of the system S.

to be equally likely. The equiprobability holds, if the system is ergodic, as is discussed in Section 1.3.1. Finally, if two states of the same energy are equally probable, the probability P of a state having energy E_r is a function of the energy E_r only.

Now we write T the total system composed of HB and S, of total energy $E = E_{HB} + E_S$. As HB is a large heat bath, its energy E_{HB} is large, $E_{HB} \gg E_S$, and the possible energy levels of HB are assumed to be quasi-continuous in comparison to the energy levels of S, as illustrated in Fig. 1.1. The probability $P(E_r)$ that S is in state r and that $E_S = E_r$ is proportional to the number of ways S can have that energy, i.e. the number of states of H that allows T to have the correct energy E, i.e. $|\Omega_{HB}(E - E_r)|$. So that,

$$\frac{P(E_r)}{P(E_{r'})} = \frac{|\Omega_{HB}(E - E_r)|}{|\Omega_{HB}(E - E_{r'})|} = e^{\log(|\Omega_{HB}(E-E_r)|) - \log(|\Omega_{HB}(E-E_{r'})|)}. \tag{1.2}$$

As $E_r \ll E$, $\beta(E) = d\log(|\Omega_{HB}(E)|)/dE|$ is almost constant in the range under consideration[1] and

$$\begin{aligned} \frac{P(E_r)}{P(E_{r'})} &= e^{-\beta(E_r - E_{r'})} \\ P(E_r) &\propto e^{-\beta E_r}. \end{aligned} \tag{1.3}$$

β is called the inverse temperature, as the temperature T is conventionally defined as $\beta = 1/k_b T$, where k_b is the Boltzmann's constant. We define the canonical partition function, Z, as

$$Z = \sum_i e^{-\beta E_i} = \mathrm{Tr}(\exp(-H/k_b T)), \tag{1.4}$$

where Tr denotes the trace of the operator. We also define the Boltzmann weight, i.e. the probability of a state i, $\pi(i) = e^{-\beta E_i}/Z$. The thermal or ensemble average of an observable θ is then,

[1] The assumption made here is equivalent to the one that the heat bath has a quasi-continuous spectrum in the region considered and for which there is no particular characteristic energy, [29, p. 3].

$$\begin{cases} \langle\theta\rangle = \frac{1}{Z}\sum_i \langle i|\theta|i\rangle e^{-\beta E_i} \\ \qquad = \frac{\mathrm{Tr}(\theta\exp(-\beta H))}{\mathrm{Tr}(\exp(-\beta H))}. \end{cases} \tag{1.5}$$

In the following, we will be concerned with the classical limit of Eq. 1.4 for N particle systems. In the classical limit, we have

$$\mathrm{Tr}(\exp(-\beta H)) \approx \mathrm{Tr}(\exp(-\beta U)\exp(-\beta K)), \tag{1.6}$$

with U the potential part and K the kinetic part of H. The sum over the states can then be replaced by an integration over all coordinates $\boldsymbol{r}$ and momenta $\boldsymbol{p}$, leading to the classical expression of the partition function,

$$Z = \frac{1}{h^{dN}N!}\int d\boldsymbol{p}^N d\boldsymbol{r}^N \exp\left(-\beta\left(\sum_j p_j^2/(2m_j) + U(\boldsymbol{r}^N)\right)\right), \tag{1.7}$$

the $N!$ term comes from the fact that permuting two similar particles does not change the observable properties of the macroscopic system and that the thermodynamic quantity of entropy has to be at a maximum in a closed system in equilibrium [41, 42]. The classical expression of the ensemble average of an observable θ is

$$\langle\theta\rangle = \frac{1}{Z}\int d\boldsymbol{p}^N d\boldsymbol{r}^N \theta(\boldsymbol{r}^N,\boldsymbol{p}^N)\exp\left(-\beta\left(\sum_j p_j^2/(2m_j) + U(\boldsymbol{r}^N)\right)\right). \tag{1.8}$$

Monte Carlo methods, see Chapter 2, have been designed to compute averages like in Eq. 1.8. But, unlike in molecular dynamics [43], Monte Carlo methods only focus on producing configurations of the positions of particles, because the particle velocities form an independent sampling problem. As particle velocities follow the Maxwell distribution, they are sampled easily, see Section 2.1.2.2. Thus, for an observable θ depending only on the set ζ of positions of the particles, we can rewrite Eq. 1.8 after renormalizing Z as,

$$\langle\theta\rangle = \frac{1}{Z}\int d\zeta\theta(\zeta)\exp\left(-\beta U(\zeta)\right). \tag{1.9}$$

Thus far, we have discussed the ensemble average of an observable θ. But the experimental average of an observable is actually its time average. The next section discusses how both averages can be considered equivalent, as long as the system is ergodic.

1.1.2 Classical statistical mechanics and ensemble averages

Let us consider a fluid consisting of N classical atoms in a volume V. One microscopic state $\zeta(t)$ of the fluid is described by the $6N$ independent position $\boldsymbol{r_j}$ and momentum

$\boldsymbol{p_j}$ coordinates of each atom. We wish to study the density of the fluid at a distance r from a given atom i, $\rho_i(r)$. In most experiments, the measurement of precise and instantaneous values $\rho_i(r, \zeta(t))$ is not possible, as it depends on the coordinates $\boldsymbol{r_j}$ of all particles j. For example, the pressure that the fluid will exert on a piston is, as a matter of fact, time-averaged by the inertia of the latter. It is then considered that the result of a measurement is the time average $\bar{\rho}_i(r)$ of $\rho_i(r)$. Each measurement of a macroscopic observable at a time t_0 is considered to take actually a certain interval of time t to be realized. During this interval, $\zeta(t)$ changes according to Newton's laws of motion and $\rho_i(r, \zeta(t))$ with it. The time average,

$$\frac{1}{t}\int_{t_0}^{t_0+t} \rho_i(r, \zeta(t'))dt', \tag{1.10}$$

is constant, if it is independent of the initial condition, i.e. t_0, and the length of the interval, i.e. t, given it is long enough. We consider this to be the case and will discuss this issue later in Section 1.3.1. One can argue that the macroscopic interval of time t for the measurement is extremely large from the microscopic point of view and one may take the limit $t \to \infty$,

$$\bar{\rho}_i(r) = \lim_{t\to\infty} \frac{1}{t}\int_{t_0}^{t_0+t} dt' \rho_i(r, t'). \tag{1.11}$$

As Eq. 1.11 does not depend on the initial condition, it is possible to rewrite it as an average over a collection of initial conditions,

$$\bar{\rho}_i(r) = \frac{\sum_{\text{initial conditions}} \lim_{t\to\infty} \frac{1}{t}\int_{t_0}^{t_0+t} dt' \rho_i(r, t')}{\text{Total number of initial conditions}}. \tag{1.12}$$

Experimentally, Eq. 1.12 can be realized by observing a large number of copies of the atomic fluid (Gibbs's approach [44]) or by observing only one and the same copy of the system on successive intervals, given the intervals are long enough so they are independent from one another (time approach). Following [11, p.16], we consider the limiting case of averaging over the set Ω of all possible initial conditions $\zeta(0)$,

$$\frac{\sum_{\text{initial conditions}} f(\zeta(0))}{\text{Total number of initial conditions}} \to \int_\Omega d\zeta(0)\pi(\zeta(0))f(\zeta(0)), \tag{1.13}$$

with f an arbitrary function of the initial coordinates $\zeta(0) = \{\boldsymbol{r_i}(0), \boldsymbol{p_i}(0)\}$. $\pi(\zeta(0))$ is the probability to start the averaging at $\zeta(0)$. For instance, if we realize the experiment of measuring one copy of the system at intervals, then the starting point of the measuring interval can be at any time and the initial conditions can be any state the system will visit. Therefore $\pi(\zeta(0)) = \pi(\zeta)$. In the case where the system is ergodic, as previously assumed by the independence from the initial conditions, the system visits during its evolution all states ζ of the phase space Γ for a given energy. The probability weight $\pi(\zeta)$ can then be understood as the probability to find the system in the state ζ during its evolution.

Using the equivalence of Eq. 1.13, the time average Eq. 1.11 is now equivalent to the ensemble average,

$$\bar{\rho}_i(r) = \langle \rho_i(r) \rangle = \int_\Omega d\zeta \pi(\zeta) \rho_i(r, \zeta), \tag{1.14}$$

where $\langle \cdot \rangle$ identifies with Eq. 1.9.

Eq. 1.14 is the main version of the ergodic hypothesis. $\bar{\rho}_i(r)$ describes then the equilibrium, without depending on the initial state. As the initial conditions do not matter, only the quantities that are conserved do, hence the link between mesoscopic description and conservation laws. Section 1.3.1 will address with more care the ergodic hypothesis. The systems studied later in Chapter 4 are ergodic, but not all systems are.

If one wishes to compute the average of a function, it is possible to either use the time average, as in molecular dynamics [43] or the ensemble average, as in the Monte Carlo method, see Chapter 2. Both approaches are equivalent, as shown by Eq. 1.14, and have been extremely versatile. For particle systems, the molecular dynamics solves the classical equations of motion and computes the trajectories of the particles from collision to collision with other particles. The Monte Carlo method builds a stochastic process that evolves on the configuration space and samples thus in a simpler way the different configurations needed for the ensemble average.

If time and ensemble averages are equivalent, given the system is ergodic, dynamical properties are still controlled by Newton's equations of motion. For a closed and isolated system[2], the microscopic deterministic motion is represented as a trajectory in phase space Γ. After a time t, each point $\zeta \in \Gamma$ is mapped in a unique and determined way into another point $\zeta^t = \phi(\zeta, t) \in \Gamma$. The initial probability density evolves in phase space Γ according to Liouville's equation. The deterministic trajectory generated this way is different from the probabilistic trajectories generated in stochastic processes. However, we are interested here on a coarse-grained picture and not on the full atomistic description. For such mesoscopic scale, only a subset of the degrees of freedom matters. The microscopic variables that evolve on a much faster time scale act then as a heat bath, like it was assumed for Eq. 1.11. It then induces stochastic transitions among the relevant and slower mesoscopic variables. In the case of a separation of time scales, the Liouville equation that controls the temporal evolution of the initial distribution over Γ can be reduced to a Markovian master equation, Eq. 1.56, and the stochastic process is then Markovian, see Section 1.2.

The stochastic nature does not come from setting a probability density $P(x)$ as the initial state. For instance, the irregular motion of a Brownian particle in a bath cannot be linked to a probability distribution of some initial state. This irregularity is actually the work of the bath molecules in the vicinity and is linked to all the microscopic variables describing the total system that have been eliminated in order to get an equation for only the Brownian particle. It is the elimination of the bath variables that allows one to establish the stochastic process of Brownian motion [40, p. 55-57].

[2]No energy or particle exchange

1.1.3 Stochastic processes

In Section 1.1.2, the equivalence between time and ensemble averages was presented. It led to the mesoscopic stochastic description, that contains both the deterministic laws and the fluctuations. Such a description consists of eliminating a large number of microscopic variables, so that a small set of mesoscopic variables obeys approximately an independent set of deterministic equations. The deterministic nature of those equations, as the hydrodynamics equations for instance, is indeed only approximated, as the now hidden microscopic variables introduce fluctuation terms [38]. The fluctuations, acting like a heat bath, change the mesoscopic variables into stochastic processes.

Following [40], a stochastic process $Y = (X_{t_1}, X_{t_2}, ...X_{t_n})$ is an ordered collection of a stochastic (or random) variable X, e.g. the ensemble of positions of one atom at successive times. The random variable X is defined by a set of possible values Ω, called the configuration space, and by a probability distribution (or density) $P(x)$ over this set, with

$$\begin{aligned} P(x) &\geq 0 \\ \int_\Omega P(x)dx &= 1. \end{aligned} \tag{1.15}$$

The probability that X has a value between x and $x + dx$ is $P(x)dx$. From Eq. 1.15, it is possible to define the average of any function f, defined on the same configuration space Ω, as in Eq. 1.14,

$$\langle f \rangle = \int_\Omega f(x)P(x)dx. \tag{1.16}$$

For multivariate distributions $\boldsymbol{X} = (X_1, X_2, ..., X_n)$, as for the atomic fluid with the whole set of position and momentum coordinates of all atoms, the probability distribution $P(\boldsymbol{x})$ can be understood as the joint probability distribution of the n variables $X_1, ..., X_n$. A marginal distribution $P(x_{i_1}, ..., x_{i_s})$ on a subset of $s < n$ variables $X_{i_1}, ..., X_{i_s}$, may be defined as

$$P(x_{i_1}, ..., x_{i_s}) = \int P(x_1, ..., x_n) \prod_{j \notin \{i_1, ..., i_s\}} dx_j. \tag{1.17}$$

For instance, in the atomic fluid, the marginal distribution of the position coordinates is

$$P(\boldsymbol{r_1}, ..., \boldsymbol{r_N}) = \int P(\boldsymbol{r_1}, ..., \boldsymbol{r_N}, \boldsymbol{p_1}, ..., \boldsymbol{p_N}) \prod_{j=0}^{N} d\boldsymbol{p_j}. \tag{1.18}$$

The average of the density function $\rho_i(r)$ is then

$$\langle \rho_i(r) \rangle = \int_\Omega \rho_i(r) P(\boldsymbol{r_1}, ..., \boldsymbol{r_N}) \prod_{j=0}^{N} d\boldsymbol{r_j}. \tag{1.19}$$

One can define also the mth moment of X,

$$\langle X^m \rangle = \int_\Omega x^m P(x) dx, \tag{1.20}$$

and the variance or dispersion $\mathrm{var}(X) = \sigma_X^2$,

$$\mathrm{var}(X) = \sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2, \tag{1.21}$$

that will be useful in discussing errors in Monte Carlo Method in Section 2.3. The mth-moment of X exists if and only if the integral $\int_\Omega |x^m| P(x) dx$ converges. If the mth-moment does not exist, neither does every nth-moment, $n \geq m$. If the mth-moment does exist, so does every nth-moment, $n \leq m$.

The joint distribution of a set of s variables, while attributing fixed values to the remaining $(r-s)$ variables, is called the conditional probability distribution of $X_{i_1}, ..., X_{i_s}$, conditional on $X_{j_1}, ..., X_{j_{r-s}}$ having the prescribed values $x_{j_1}, ..., x_{j_{r-s}}$ and denoted as $P(x_{i_1}, ..., x_{i_s} | x_{j_1}, ..., x_{j_{r-s}})$. Following the definition of the marginal and the conditional probability distribution, Bayes' rule is then expressed by,

$$P(x_1, ..., x_r) = P(x_{j_1}, ..., x_{j_{r-s}}) P(x_{i_1}, ..., x_{i_s} | x_{j_1}, ..., x_{j_{r-s}}) \tag{1.22}$$

or

$$P(x_{i_1}, ..., x_{i_s} | x_{j_1}, ..., x_{j_{r-s}}) = \frac{P(x_1, ..., x_r)}{P(x_{j_1}, ..., x_{j_{r-s}})}. \tag{1.23}$$

It leads to the concept of statistical independence of two sets. If the set of r variables can be subdivided in two sets $(X_{i_1}, ..., X_{i_s})$ and $(X_{j_1}, ..., X_{j_{r-s}})$, so that P_r factorizes,

$$P(x_1, ..., x_r) = P_s(x_{i_1}, ..., x_{i_s}) P(x_{j_1}, ..., x_{j_{r-s}}), \tag{1.24}$$

then the two sets are called statistically independent and the factor P_s (respectively P_{r-s}) identifies with the conditional probability density $P_{s|r-s}$ (respectively $P_{r-s|s}$). Basically, this means that fixing the values of one set brings no extra information for the possible values in the other set, as they do not affect each other. This will be particularly useful to devise the factorized Metropolis filter that leads to rejection-free irreversible Markov chains, as discussed in Section 3.2.

As in Eq. 1.20, one can define the moments of a multivariate distribution,

$$\langle X_1^{m_1} .. X_r^{m_r} \rangle = \int_\Omega x_1^{m_1} ... x_r^{m_r} P(x_1, ..., x_r) dx_1 ... dx_r. \tag{1.25}$$

We are interested in stochastic processes $Y = \{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$ that are ordered according to the time t, as illustrated in Fig. 1.1.3. A realization of the process $y(t)$ is an ordered collection $\{x_t\}$, where x_t are successive realizations of X. In statistical

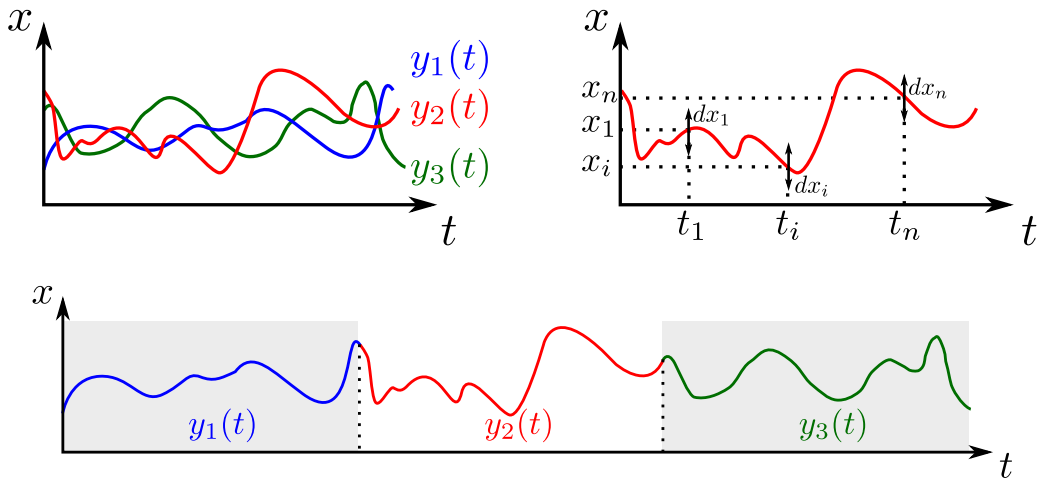


Figure 1.2: A stochastic process Y is defined as an ordered collection of a random variable X. Here it is ordered along the time t and Y stands for the successive positions $\{x_t\}$ of a Brownian particle. **Top left:** A stochastic process is the ensemble of all its realizations $y_i(t)$, where $y_i(t)$ is a sequence of realization x of X. **Top right:** One can define the probability $P_n(x_1,t_1;x_2,t_2;\ldots;x_n,t_n)dx_1\ldots dx_n$ for a realization to go through the n windows centered on (x_i,t_i) and of width dx_i. **Bottom:** For self-averaging processes, which can be decomposed on long time intervals that do not affect one another, the stochastic process Y is simply the ensemble of all the intervals $y_i(t)$.

mechanics, the stochastic process Y itself is understood as the ensemble of these realizations $y(t)$, like an ensemble of a large number of copies of an experiment. The ensemble average $\langle\cdot\rangle$ is

$$\begin{aligned}\langle Y(t_i)\rangle &= \int y(t_i)\mathcal{P}(y)dy \\ \langle Y(t_i)\rangle &= \int_\Omega xP_1(x,t_i)dx.\end{aligned} \tag{1.26}$$

Here the functional $\mathcal{P}$ is the probability distribution of y, $\int\mathcal{P}(y)dy = 1$. It can be rewritten as the joint probability distribution P_n of observing a certain realization $\{x_{t_1},x_{t_2},\ldots,x_{t_n}\}$ as

$$P_n(x_1,t_1;x_2,t_2;\ldots;x_n,t_n) = \langle\delta(Y(t_1)-x_1)\delta(Y(t_2)-x_2)\ldots\delta(Y(t_n)-x_n)\rangle. \tag{1.27}$$

A stochastic process is completely defined by the complete set of P_n, as shown by the Kolmogorov existence theorem [45, 46]. Also P_n does not change on interchanging two pairs (x_k,t_k) and (x_l,t_l). Considering the atomic fluid, the ensemble average of the density $\langle\rho_i(r)\rangle$ is obtained by considering all the possible realizations $\rho_i(r,t)$, leading to Eq. 1.14.

The nth moment, from n values $t_1,...,t_n$ of the time t, can also be defined,

$$\langle Y(t_1)...Y(t_n)\rangle = \int_\Omega x_1\ldots x_nP_n(x_1,t_1;\ldots;x_n,t_n)dx_1\ldots dx_n. \tag{1.28}$$

The normalized autocorrelation C_Y function will be particularly useful,

$$\begin{aligned} C_Y(t_1,t_2) &= \frac{\langle Y(t_1) - \langle Y(t_1)\rangle\rangle\langle Y(t_2) - \langle Y(t_2)\rangle\rangle}{\sqrt{\mathrm{var}(Y(t_1))\mathrm{var}(Y(t_2))}} \\ &= \frac{\langle Y(t_1)Y(t_2)\rangle - \langle Y(t_1)\rangle\langle Y(t_2)\rangle}{\sqrt{\mathrm{var}(Y(t_1))\mathrm{var}(Y(t_2))}}. \end{aligned} \tag{1.29}$$

From Eq. 1.28, one can define the stationary property. The statistical property of a stationary process does not change with time, i.e. all the nth moments of a stochastic process are not affected by any shift τ in time,

$$\langle Y(t_1+\tau)...Y(t_n+\tau)\rangle = \langle Y(t_1)...Y(t_n)\rangle. \tag{1.30}$$

A stationary stochastic process has therefore a time-independent average and its autocorrelation function $C_Y(t_1,t_2)$ of Eq. 1.29 depends only on the difference of time $|t_1 - t_2|$, in particular $C_Y(t_1,t_2) = C(t_2,t_1)$,

$$C_Y(t) = \frac{\langle Y(t)Y(0)\rangle - \langle Y\rangle^2}{\langle Y^2\rangle - \langle Y\rangle^2}. \tag{1.31}$$

This leads to the definition of an autocorrelation time τ_{exp}, so that $C_Y(|t_1 - t_2| > \tau_{\mathrm{exp}})$ is negligible, see Section 1.3.2.4 and Section 2.3.3. In practice, stationary processes are only an approximation; they arise by considering analytically the right time scale, so that the evolution of the microscopic variables is averaged, whereas the macroscopic variation is still negligible.

We discussed how the equivalence of time and ensemble average allows for a reformulation of the problem in terms of a stochastic process. From the moment the system exhibits a self-averaging property, which is the case in particular for a stationary system, the irregularly varying function can be cut into a collection of long time intervals. These intervals are an ensemble of realizations of the stochastic process. It is necessary that there is no influence of an interval on the next one. If this is the case, the process is called ergodic, see Section 1.3.1.

1.2 Markov processes and master equation

Before discussing Markov processes developed within Markov-chain Monte Carlo methods in Section 2.2, we address more generally in this Section the Markovian description of physical systems. As seen in Section 1.1.2, if the separation of scales between microscopic and macroscopic variables is good enough, the underlying stochastic process has a memory of only its immediate past and is said to be Markovian. Experimentally, the Markov property is only approximate. For a Brownian particle for instance, if the previous displacement was a large one, the probability that the current displacement is also large is slightly higher. The autocorrelation time of the velocity, even if small, is not strictly zero and a large displacement is more likely just after a

large one. But, as the experimental sampling of positions is made on a coarse-grained time scale longer than the autocorrelation time of the velocity, see Section 1.1.2, the process appears Markovian.

The Markovian nature depends on the number of variables used to describe the process. For example, if a stochastic process described by r variables is Markovian, it is not necessary that the stochastic process constituted by a subset of s variables from the previous ones is Markovian. The information brought by the s variables may not be sufficient to predict the future configuration, not even the configuration restricted to these s variables.

However, if a process is not Markovian, it is sometimes possible to extend the configuration space by additional variables and create a Markovian process. These additional variables describe explicitly the information that was previously implicit and hidden in the past values of the variables. This extension of the configuration space will be one of the key elements of the rejection-free Monte Carlo scheme, see Section 3.1.

As discussed in Section 1.1.2, the goal is to find a small set of variables preserving the Markovian nature. Experimentally, it is possible to do so for most many-body systems, but it remains an approximate description restricted to a macroscopic, coarse-grained level. This reduction is usually called *contraction* or *projection* and the justification of this approximation is still a debate at the most fundamental level of statistical mechanics.

1.2.1 General properties of a Markov process

Markov processes (or Markov chains) exhibit the Markov property: The conditional probability distribution of future states depends only upon the present state. For any set of n successive times (i.e. $t_1 < t_2 < ... < t_n$),

$$P(x_n, t_n | x_1, t_1;\ x_2, t_2;\ ...;\ x_{n-1}, t_{n-1}) = P(x_n, t_n | x_{n-1}, t_{n-1}). \tag{1.32}$$

A Markov process is fully determined by the initial probability distribution $P(x_1, t_1)$ and the transition probability $P(x_2, t_2 | x_1, t_1) = p((x_1, t_1) \to (x_2, t_2))$. Using only these functions, the complete hierarchy can be retrieved,

$$P_3(x_1, t_1;\ x_2, t_2;\ x_3, t_3) = P_1(x_1, t_1) P(x_2, t_2 | x_1, t_1) P(x_3, t_3 | x_2, t_2). \tag{1.33}$$

Eq. 1.33 written for two times t_1 and t_2 is simply the Bayes rule, Eq. 1.23,

$$P_2(x_1, t_1;\ x_2, t_2) = P_1(x_1, t_1) P(x_2, t_2 | x_1, t_1). \tag{1.34}$$

Integrating Eq. 1.34 over x_2 and dividing by $P_1(x_1, t_1)$ leads to the conservation condition for all x_1, t_1, t_2,

$$1 = \int P(x_2, t_2 | x_1, t_1) dx_2 \tag{1.35}$$

Equivalent to,

$$P_1(x_1,t_1) = \int P_1(x_1,t_1)P(x_2,t_2|x_1,t_1)dx_2 \tag{1.36}$$

Integrating now Eq. 1.34 over x_1 leads to the necessary condition for all x_2, t_1, t_2,

$$P_1(x_2,t_2) = \int P(x_2,t_2|x_1,t_1)P_1(x_1,t_1)dx_1. \tag{1.37}$$

Eq. 1.37 is known as the global-balance condition in the Markov-chain Monte Carlo method, in Chapter 2, Section 2.2. By analogy with the mass flow in hydrodynamics, the instantaneous probability flow $\phi_{x_1\to x_2}(t_1,t_2)$ between two configurations is defined as

$$\phi_{x_1\to x_2}(t_1,t_2) = P_1(x_1,t_1)P(x_2,t_2|x_1,t_1). \tag{1.38}$$

As for the mass flow, the probability flow is composed of a velocity term $P_1(x_2,t_2|x_1,t_1)$ and a mass term $P_1(x_1,t_1)$ and obeys a continuity condition, Eq. 1.36. Following this analogy, the global-balance condition Eq. 1.37 and the conservation condition Eq. 1.36, when combined, yield the incompressibility of the probability flows,

$$\begin{gathered}\underbrace{\sum_{x'\in\Omega} P_1(x',t_1)P(x,t_2|x',t_1)}_{\text{All flows into configuration } x} = \underbrace{\sum_{x'\in\Omega} P_1(x,t_1)P(x',t_2|x,t_1)}_{\text{All flows out of configuration } x} \text{ for all } x, x' \text{ in } \Omega. \\ \sum_{x'\in\Omega} \phi_{x\to x'}(t_1,t_2) = \sum_{x'\in\Omega} \phi_{x'\to x}(t_1,t_2)\end{gathered} \tag{1.39}$$

Finally, a relationship between the transition probabilities can be obtained by integrating Eq. 1.33 over x_2. Considering that the initial probability distribution is arbitrary, we obtain, for any $t_1 < t_2 < t_3$,

$$P_1(x_3,t_3|x_1,t_1) = \int P(x_3,t_3|x_2,t_2)P(x_2,t_2|x_1,t_1)dx_2, \tag{1.40}$$

Eq. 1.40 is known as the Chapman-Kolmogorov equation. Any two nonnegative function $P(x_1,t_1)$ and $P(x_2,t_2|x_1,t_1)$ following the set of equations (Eq. 1.37, Eq. 1.40) or equivalently the set (Eq. 1.37, Eq. 1.35) define uniquely a Markov process. Even if a Markov process is specified only by P_1 and P_2, one needs however all the P_n to demonstrate the Markovian nature of a process.

1.2.2 Stationary and homogeneous Markov processes

Markov processes that are stationary, as defined in Eq. 1.30, are of special interest, in particular for describing a system in equilibrium and the fluctuations within. Consider a system described by a Markov process $Y(t)$. When the system reaches equilibrium,

$Y(t)$ becomes stationary. In particular, $P_1(x,t) = \pi(x)$ is independent of time and is the equilibrium distribution of the quantity Y, as described by statistical mechanics, Section 1.1.1. For simplicity, we consider here discrete-time Markov chains. The continuous-time Markov chains will be presented in Section 1.2.3.

1.2.2.1 Transfer Matrix

The transition probabilities of a stationary Markov process only depend on the time difference. Once the initial probability distribution is known, the Markov chain is then completely characterized by a stochastic matrix $\mathbf{T}$, called the transfer matrix. The matrix $\mathbf{T}$ has for elements all the transition probabilities,

$$\begin{aligned} P(X_{t+1} = x' | X_t = x) &= T(x,x') \\ P(X_{t+\tau} = x' | X_t = x) &= T^\tau(x,x'), \end{aligned} \tag{1.41}$$

with $(x,x') \in \Omega^2$ and $\tau = t_2 - t_1$. The element $T^\tau(x_1,x_2)$ is the probability that a chain on a state x reaches x' in τ steps. The matrix $\mathbf{T}^\tau$ is simply the τ-fold matrix product of $\mathbf{T}$. We can also define the stationary probability flow between two states x, x',

$$\begin{aligned} \phi_{x\to x'} &= \pi(x)T(x,x') \\ \phi^\tau_{x\to x'} &= \pi(x)T^\tau(x,x'). \end{aligned} \tag{1.42}$$

The Chapman-Kolmogorov equation Eq. 1.40 becomes a simple matrix product, with $\tau, \tau' > 0$,

$$\mathbf{T}^{\tau+\tau'} = \mathbf{T}^\tau \mathbf{T}^{\tau'}. \tag{1.43}$$

Eq. 1.41 is not restricted to positive τ [40]. As the probability $P(X_{t+\tau} = x', X_t = x)$ is symmetric, i.e. $P(X_{t+\tau} = x', X_t = x) = P(X_t = x, X_{t+\tau} = x')$, from Eq. 1.41, we can write,

$$\begin{aligned} \pi(x)T^\tau(x,x') &= \pi(x')T^{-\tau}(x',x), \\ \phi^\tau_{x\to x'} &= \phi^{-\tau}_{x'\to x}. \end{aligned} \tag{1.44}$$

Eq. 1.44 applies for any stationary Markov process, among those, physical systems at equilibrium, without additional assumption. It is simply the time reversal.

For a stationary process, the conservation condition and the global-balance condition can be simplified as,

$$\text{Conservation condition: } \int T(x,x')dx' = 1 \text{ for all } x \text{ in } \Omega, \tag{1.45}$$

$$\text{Global-balance condition: } \int \pi(x')T(x',x)dx' = \pi(x) \text{ for all } x \text{ in } \Omega. \tag{1.46}$$

From Eq. 1.45, the matrix $\mathbf{T}$ has a right-eigenvector $\mathbf{1}$ of eigenvalue 1 and, from Eq. 1.46, $\mathbf{T}$ has the stationary distribution π as a left-eigenvector of eigenvalue 1. The spectrum

of the transfer matrix contains important information about the properties of convergence and relaxation of the Markov chain, as will be discussed further in Section 1.3.2.

The global balance in Eq. 1.46 is automatically fulfilled, when the stronger condition of the detailed balance Eq. 1.47, also called reversibility condition, is satisfied,

$$\text{Detailed-balance condition: } \underbrace{\pi(x)T(x,x')}_{\text{Flow } x\to x'} = \underbrace{\pi(x')T(x',x)}_{\text{Flow } x'\to x} \text{ for all } x \text{ and } x' \text{ in } \Omega. \quad (1.47)$$

Note that Eq. 1.44 is not the detailed-balance condition, which has $+\tau$ on the right-hand side. Enforcing the detailed-balance condition requires an additional physical justification. Eq. 1.47 is equivalent to the joint probability being symmetric, i.e. $P_2(X_{t-1} = x, X_t = x') = P_2(X_{t-1} = x', X_t = x)$. Hence the name reversible Markov chain is used for any chain that obeys detailed balance. From now on, we only consider $\tau > 0$ to avoid confusion between time reversal and time reversibility.

1.2.2.2 Extraction of a subensemble

Stationary Markov processes will prove particularly useful in Chapter 2, regarding the extraction of a particular subensemble. A subensemble of a stochastic process Y is the subset of all realizations $y(t)$ that obey a given condition. For instance, they all take the value x_0 at time t_0. This defines a new, non-stationary Markov process Y^* for $t > t_0$,

$$\begin{aligned} P_1^*(x_1,t_1) &= T^{t_1-t_0}(x_0,x_1) \\ P^*(x_2,t_2|x_1,t_1) &= T^{t_2-t_1}(x_1,x_2). \end{aligned} \quad (1.48)$$

More generally, one could extract a subensemble where at a given time t_0 the values of the realization $y(t)$ are distributed according to a given probability distribution $p(x_0)$,

$$\begin{aligned} P_1^*(x_1,t_1) &= \int dx_0 p(x_0) T^{t_1-t_0}(x_0,x_1) \\ P^*(x_2,t_2|x_1,t_1) &= T^{t_2-t_1}(x_1,x_2). \end{aligned} \quad (1.49)$$

Although they are extracted from a stationary process, these processes are non-stationary, given $p \neq \pi$, as a time t_0 is singled out. The probability flows near t_0 are not the stationary ones. However, the transition probabilities still depend only on the time interval alone, as it is the same as the underlying stationary process. Such non-stationary Markov processes whose transition probabilities only depend on time difference are called homogeneous processes or Markov process with stationary transition probability.

For a stationary process, extracting such subensemble physically means that one prepares the system in a certain non-equilibrium state, x_0 and expects that after a long time the system returns to equilibrium, i.e.,

$$P_1^*(x_1,t_1) \xrightarrow[t\to\infty]{} \pi(x_1). \quad (1.50)$$

Moreover, as any initial state x_0, or more generally any initial probability distribution $p(x_0)$, is arbitrary, it is expected that,

$$T^{t_1-t_0}(x_0, x_1) \xrightarrow[t_1\to\infty]{} \pi(x_1). \tag{1.51}$$

Quantifying the asymptotic behavior in Eq. 1.50 and Eq. 1.51 is a major problem in dealing with Markov processes and is discussed in Section 1.3. The convergence of Markov processes will be at the heart of the Monte Carlo methods, see Chapter 2 and Chapter 3, that relies on Markov processes to simulate ensemble average, as in Eq. 1.14.

1.2.3 Master equation

The Chapman-Kolmogorov equation Eq. 1.43 can be rewritten as the master equation, in the limit of vanishing time difference τ', i.e. continuous time, as a differential equation. The master equation has the advantage of being more convenient for mathematical operations but, most of all, it has a more direct physical interpretation.

1.2.3.1 Derivation of the master equation

For simplicity, we consider a homogeneous Markov process. From Eq. 1.37, $T^{\tau'}(x_1, x_2)$ can be rewritten for small τ' [40, 47] as

$$T^{\tau'}(x_1, x_2) = (1 - a_0(x_1)\tau')\delta(x_1 - x_2) + \tau' W(x_1, x_2) + o(\tau'), \tag{1.52}$$

with $W(x_1, x_2)$ being the transition probability per unit of time from x_1 to x_2. $W(x_1, x_2)$ is nonnegative. $1 - a_0\tau'$ is the probability that no transition takes place during τ', with,

$$a_0(x_1) = \int W(x_1, x_2)dx_2, \tag{1.53}$$

as $\int T^{\tau'}(x_1, x_2)dx_2 = 1$, following Eq. 1.45. Inserting Eq. 1.52 into the Chapman-Kolmogorov equation Eq. 1.43,

$$T^{\tau+\tau'}(x_1, x_3) = T^{\tau}(x_1, x_3)(1 - a_0(x_1)\tau') + \tau' \int T^{\tau}(x_1, x_2)W(x_2, x_3)dx_2. \tag{1.54}$$

Eq. 1.54 is then divided by τ', the limit $\tau' \to 0$ is taken and a_0 is replaced by Eq. 1.53. A differential form of the Chapman-Kolmogorov equation is then obtained,

$$\frac{\partial T^{\tau}(x_1, x_3)}{\partial t} = \int dx_2 \left[T^{\tau}(x_1, x_2)W(x_2, x_3) - T^{\tau}(x_1, x_3)W(x_3, x_2)\right], \tag{1.55}$$

Considering x_1 as an initial state, $P(x, t = 0) = \delta(x - x_1)$, it is usually written as

$$\frac{\partial P(x, t)}{\partial t} = \int dx' \left[P(x', t)W(x', x) - P(x, t)W(x, x')\right], \tag{1.56}$$

This is the famous master equation. It can be understood as a *continuity* equation. The master equation contains indeed the fact that the total probability is conserved,

$\int_x P(x,t)dx = 1$ at all times t, and also that any probability 'lost' for a state x in the transition to a state x' is actually gained in the probability of x'. This 'incompressibility' is the condition that Markov chains in the Markov-chain Monte Carlo method have to follow in order for the scheme to be correct, see Section 2.2.

As shown by its derivation, the master equation is always the result of an approximation and relies on the existence of the transition rates $W(x,x')$. The quantity $W(x,x')\Delta t$ is the probability for a transition from x to x' during a short time Δt and it is computed, for a given system, by means of any available approximation method that is valid for short times. Unlike the Liouville equation or Schrödinger equation, the master equation is mesoscopic and is also irreversible, as the evolution of a system towards the equilibrium. The property that all solutions of the master equation tend to the stationary solution or one of the stationary solutions will be discussed, as the veracity of Eq. 1.51, in the next Section 1.3.

1.2.3.2 Transfer Matrix for continuous-time Markov chain

It is useful to arrange the transition rate $W(x,x')$ into a matrix $\mathbf{W}$ so that,

$$\frac{\partial P(x,t)}{\partial t} = \boldsymbol{P}(t)\mathbf{W}. \tag{1.57}$$

This is achieved by

$$\mathbf{W}(x,x') = W(x,x') - \delta(x-x')\sum_{x''} W(x,x''). \tag{1.58}$$

It is noteworthy that $\mathbf{W}$ has π as a right-eigenvector but of eigenvalue 0. Once the stationary regime is reached and $T^{\infty}(x_1,x_3) = \pi(x_3)$, the right-hand side of the master equation Eq. 1.56 is indeed 0, as it is the global-balance condition Eq. 1.46. It is consistent with the fact that, once the equilibrium is reached, the system follows a reversible evolution and $\frac{\partial P(x,t)}{\partial t} = 0$.

From Eq. 1.57, we can define the transfer matrix $\mathbf{T}^t$ [48, p. 25][49, Chapter 2,3] as,

$$T^t(x,x') = \exp(\mathbf{W}t)(x,x'), \tag{1.59}$$

it is easy to check that $\pi\mathbf{T}^t = \pi$.

Given a discrete-time chain $\mathbf{T}$, one can define the continuous-time chain $\tilde{\mathbf{T}}$, by defining the transition rates $W(i,j) = T(i,j)$, for $i \neq j$. Many quantities are unchanged by the passage from the discrete-time chain to the continuous-time chain, as the stationary distribution π. This would be helpful to devise the continuous-time irreversible Markov chains that are used in the irreversible factorized Metropolis paradigm, as discussed in Chapter 3.

1.3 Convergence in Markov processes

As discussed in Section 1.1.2 and Section 1.2, the ergodic hypothesis is at the basis of statistical mechanics, as it is needed for the equivalence between time and ensemble averages, Eq. 1.14. The ergodic theory quickly became a whole research field in mathematics. Beyond statistical mechanics, it is now an even more important issue because of its importance in diverse fields like celestial mechanics (regarding the stability of the solar system) or chemistry (stability of isolated excited molecules).

In this Section, we discuss the ergodic hypothesis and the mixing property. In particular, the capacity of a system to reach equilibrium from a non equilibrium initial condition in Section 1.3.1 and the convergence behavior of Markov chains that may describe those systems in Section 1.3.2.

1.3.1 Ergodic and mixing properties

If a system is in equilibrium at energy E, long-time averages of any observable should not depend on the initial state, as is assumed in Eq. 1.10. In principle, different initial conditions will give different average values for the observable, following the Liouville equation. However, if the system passes close to nearly all the states compatible with conservation of energy, then the initial condition does not matter. The justification of this hypothesis is not trivial and is the subject of the ergodic theory.

Although ergodicity has been proven for finite systems of hard spheres [50, 51], this is not the case for other finite systems. For instance, glasses and metastable phases are not ergodic in practice and harmonic solids are simply not ergodic. Another important example is a system in a phase with broken symmetries. They are not ergodic, as they stay in one separate region in phase space. Even systems which present small perturbations, which should spread equally the initial energy among the different modes, can be non ergodic. For instance, E. Fermi, J. R. Pasta, S. Ulam and M. Tsingou simulated a one-dimensional chain of atoms, interacting through an anharmonic potential. They did not manage to observe thermalization that the anharmonicity was expected to bring [52]. It was discovered later that introducing perturbations is not a sufficient condition to ensure ergodicity. In spite of the perturbations, trajectories in the phase space may remain quasiperiodic, as they follow invariant tori which are linked to conserved quantities. For a certain numbers of initial conditions, the system is unable to explore the full phase space Γ. This result is known as the KAM theorem, after Kolmogorov [53], Arnold [54] and Moser [55].

The difference between ergodic and non-ergodic systems is of importance only for small systems that are not coupled to a heat bath. We will take a pragmatic position and assume from now on the ergodic hypothesis for two reasons. First, we are interested in systems with strong interactions between their particles, so that there are no integrable constants of motion other than functions of the energy E [56], as in the case of hard spheres [50, 51]. Second, we are most concerned here with the possibility of a

system to reach equilibrium from a given initial density ρ_0 on the phase space. Ergodicity is not a sufficient condition to ensure that a probability distribution that is initially localized will evolve into the equilibrium distribution. If the system does so, it exhibits the stronger property of mixing and, as a consequence, is ergodic [57, p. 25][56]. For a system in state ζ at t, we denote $\phi_{-t}(\zeta)$ as its initial state. The mixing property is that, for every pair of functions f and g whose squares are integrable on Ω,

$$\lim_{t\to\infty}\int_\Omega f(\zeta)g(\phi_{-t}(\zeta))d\zeta = \frac{\int_\Omega f(\zeta)d\zeta \int_\Omega g(\zeta)d\zeta}{\int_\Omega d\zeta} \tag{1.60}$$

If $g = \rho_0$ in Eq. 1.60, it shows that $\int_\Omega d\zeta f(\zeta)\rho_0[\phi_{-t}(\zeta)]$ will approach the equilibrium value of f for large t. [50, 51] showed that hard spheres are mixing. The mathematical definition of mixing was introduced in 1932 by J. von Neumann [58], who also played a great role in the development of the Monte Carlo method (see Chapter 2). This was later developed by E. Hopf [59, 60]. However, the concept goes back to J. W. Gibbs [44, p. 144] who introduced it by the analogy of stirring an incompressible liquid containing a certain amount of coloring matter. Gibbs understood that the density probability $\rho(t)$ on a phase space of a mixing system approaches its limit not in the usual pointwise sense, but in a coarse-grained sense: The average of $\rho(t)$ over a region R of Ω approaches the equilibrium limit as $t \to \infty$ for each R.

A proper approach to equilibrium should not be possible for any finite mechanical system, due to the recurrence theorem of Poincaré stating that every such system eventually returns arbitrarily close to its initial state [61]. It can be argued that, when an ensemble of systems is considered, the individual systems may return to their initial state, as required by the Poincaré theorem, but this will happen at different times for different systems, so that, at any particular time, only a very small fraction of the systems in the ensemble are close to their initial states. The more general reason of the irrelevance of the Poincaré theorem in mixing systems is that the motion of the phase point is very unstable [56]. States that start very close to each other become widely separated as time progresses. The recurrence time is then extremely sensitive to the initial conditions. This type of instability appears to be characteristic of real physical systems and the work of unavoidable small external perturbations. It leads to irreversibility: If we reverse the velocities of every particle, the system would not necessarily return or even come close to its initial state. This instability is also found in molecular dynamics simulations, although of different nature. The numerical simulation of the equations of motion from time 0 to time t followed by the backwards simulation from time t to 0 will lead to a new state, as the numerical integration is unstable to small rounding-off errors made during the computation. They play the same role as external perturbations in real systems.

Eq. 1.60 can also be interpreted as requiring that, as equilibrium is reached with $t \to \infty$, every correlation function such as $\langle f(\zeta)g[\phi_{-t}(\zeta)]\rangle$ approaches a limit $\langle f\rangle\langle g\rangle$. This translates to the following: if Q and R are arbitrary regions of Ω and the initial distribution is uniform over Q, then the fraction of the distribution on R at time t will approach a limit as t goes to ∞, this limit being the fraction of the area of Ω occupied by

R. This criterion for equilibrium will be particularly useful to assess the convergence in Markov-chain Monte Carlo method, see Section 2.3.

1.3.2 Mixing in Markov chains

As discussed in Section 1.2.2 and Section 1.2.3, a major issue in Markov processes is to prove their mixing behavior, Eq. 1.50, and to assess how fast a chain $\mathbf{T}$ can go from an initial probability distribution μ to its stationary distribution π. Regarding Markov chains, the ergodic and even mixing properties are more simply demonstrated. From this point on, we will consider homogeneous Markov chains, see Section 1.2.2.2, as are used in Chapter 2 and Chapter 3. Also, for simplicity, we will consider chains on finite discrete Ω. Much of the theory for general configuration spaces derives from the discrete case by replacing sums by integrals and matrices by kernels, although the proofs are often considerably harder [62].

1.3.2.1 Ergodic Theorem for Markov chains

A Markov chain $\mathbf{T}$ is said to be irreducible if, for any two states $x, x' \in \Omega$, there exists τ such that $T^\tau(x, x') > 0$. This property simply requires that any state can be reached from any other one. Following the Perron-Frobenius theorem, an irreducible Markov chain possesses a stationary distribution, which is unique [63, p. 12-14]. Furthermore it is proven by the Ergodic Theorem that, for any irreducible Markov chain, the time average is equal to the ensemble average over the stationary distribution, for any initial distribution μ, [63, p. 58]. The probability $\pi(x)$ to find the chain in x is equal to the asymptotic proportion of time the chain spends in a state x, which is equivalent to the average time the chain takes to start from x and return to x,

$$\pi(x) = \frac{1}{\langle \tau_x^{H+} \rangle_x}, \tag{1.61}$$

with $\langle \cdot \rangle_x$ the average on the ensemble of realizations of the chain with x as the initial state and τ_x^{H+} the positive hitting time for a given state $x \in \Omega$, i.e. the first strictly positive time at which the chain visits x,

$$\tau_x^{H+} = \min \left(t \geq 1 : X_t = x \right). \tag{1.62}$$

For any states x and x', the expectation of the positive hitting time of state x' for an irreducible chain starting from x is always finite.

Eq. 1.61 can be proven by considering a Markov chain and its initial state z. When the chain visits z again at t, the section from 0 to $t-1$ is stored, and so on after each visit in z. It gives a collection of realizations of the chain starting in z of average length $\langle \tau_z^{H+} \rangle_z$. Averaging the number of visits in x on this collection gives Eq. 1.61 [63, p. 20][48, p. 28].

1.3.2.2 Convergence Theorem

Even if a Markov chain is ergodic in theory, in practice it could take an infinite time for the chain to decorrelate from the initial distribution and reach the stationary probability distribution π. Thus it is important to understand how the decay to equilibrium behaves.

As we are interested in the speed of mixing of a chain, another simple property is required for the completion of the Convergence theorem for chains with discrete time. A chain $\mathbf{T}$ is called aperiodic if all states have a period of 1. The period of a state x is the greatest common divisor, noted gcd, of the set of times $\mathcal{T}(x) = \{\tau \geq 1; T^\tau(x,x) > 0\}$. For an irreducible chain, the period of the chain is defined to be the period which is common to all states, since, if $\mathbf{T}$ is irreducible, $gcd\ \mathcal{T}(x) = gcd\ \mathcal{T}(x')$ for all $(x,x') \in \Omega$. It follows that if $\mathbf{T}$ is irreducible and aperiodic, there is an integer r such that $T^r(x,x') > 0$ for any $(x,x') \in \Omega$ [63].

To monitor the rate of convergence, the total variation distance between the initial probability distribution, μ, and the stationary distribution π is defined as the maximum difference between the probabilities assigned to a single event A by the two distributions,

$$d(t) = ||\mu_t - \pi||_{TV} = \max_{A \subset \Omega} |\mu_t(A) - \pi(A)|. \tag{1.63}$$

Eq. 1.63 can actually be rewritten as [63, p. 48],

$$||\mu_t - \pi||_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu_t(x) - \pi(x)|. \tag{1.64}$$

We consider now an irreducible Markov chain $\mathbf{T}$ on finite set Ω, which is moreover aperiodic, in case of discrete time steps. The chain $\mathbf{T}$ has the stationary distribution π. Then, the Convergence Theorem stipulates that there exist constants $\alpha \in [0,1)$ and $C > 0$, such that,

$$d(t) = \max_{x \in \Omega} ||T^t(x,\cdot) - \pi||_{TV} \leq C\alpha^t. \tag{1.65}$$

It proves an exponential decay to equilibrium for $\mathbf{T}$ and justifies Eq. 1.50 and Eq. 1.51.

1.3.2.3 Mixing time

In order to characterize the exponential decay of Eq. 1.65, the mixing time τ_{mix} is defined as the time needed for the current distribution $T^t(x,\cdot)$ to be close enough to π according to a parameter ϵ,

$$\tau_{\text{mix}}(\epsilon) = \min\{t | d(t) \leq \epsilon\}. \tag{1.66}$$

The time τ_{mix} is the time needed by the chains to reach equilibrium from the worst initial condition. For $\epsilon = 1/e$, $d(t) = \mathcal{O}\left(e^{-t/\tau_{\text{mix}}(1/e)}\right)$. Computing τ_{mix} directly is often not possible, but the spectral properties of $\mathbf{T}$ can be used for bounding τ_{mix}.

We already know that π is one of the eigenvectors of $\mathbf{T}$ and corresponds to the eigenvalue 1. For simplicity, we assume for now that $\mathbf{T}$ is reversible, i.e. follows detailed balance Eq. 1.47. Then $\mathbf{T}$ is diagonalizable. We consider for now the discrete-time case. The matrix $\mathbf{T}$ being stochastic, all its eigenvalues λ are bounded $|\lambda| \leq 1$. As $\mathbf{T}$ is aperiodic and irreducible, -1 is not an eigenvalue. As $\mathbf{T}$ is irreducible, the largest eigenvalue λ_1, equal to 1, and the corresponding eigenvector, π, are unique, according to the Perron-Frobenius theorem. We can enumerate the eigenvalues of $\mathbf{T}$ as $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n$ and the corresponding eigenvectors as $\{\psi_i\}$, that are orthonormal with respect to the inner product [63, p.153],

$$\langle f, g \rangle_\pi = \sum_{x \in \Omega} f(x) g(x) \pi(x). \tag{1.67}$$

Considering matrix powers, we obtain the spectral representation formula for discrete time steps [48, p. 73],

$$T_{ij}^t = \sum_{m=1}^{n} \lambda_m^t \pi_j \psi_{im} \psi_{jm}, \tag{1.68}$$

For continuous time steps, the eigenvalues of $-\mathbf{W}$ are all positive and are ordered as $0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n$. The spectral representation formula is [48, p. 73],

$$T_{ij}^t = \sum_{m=1}^{n} \exp(-\lambda_m t) \pi_j \psi_{im} \psi_{jm} \tag{1.69}$$

Eq. 1.69 can be obtained by Eq. 1.68 by going to the limit of infinitesimal time steps in a discrete-time chain. It leads to a Poisson distribution and the relationship $\lambda_m^c = 1 - \lambda_m^d$, c is for continuous time and d for discrete time.

An interesting consequence of the spectral representation is that, considering the simpler case of continuous time, there exists i such that,

$$T_{ii}^t - \pi_i \sim c_{ii} e^{-\lambda_2 t} \text{ as } t \to \infty, \tag{1.70}$$

where $c_{ii} = \pi_i \psi_{i2}^2 > 0$. A more general formulation is, for i, j such that $c_{ij} = \pi_j \psi_{i2} \psi_{j2} \neq 0$,

$$T_{ij}^t - \pi_j = c_{ij} e^{-\lambda_2 t} + o(e^{-\lambda_2 t}) \text{ as } t \to \infty, \tag{1.71}$$

Thus the eigenvalue λ_2 can be interpreted as the asymptotic rate of convergence to the stationary distribution. It is then natural to define a relaxation time τ_{rel} as,

$$\begin{aligned} \tau_{\text{rel}} &= 1/\lambda_2 && \text{for a continuous time chain} \\ \tau_{\text{rel}} &= 1/(1-\lambda^*) && \text{for a discrete time chain,} \end{aligned} \tag{1.72}$$

with $\lambda^* = \max\{|\lambda_m| < 1\} = \max(\lambda_2, |\lambda_n|)$. It is usual to define the absolute spectral gap $\gamma^* = 1 - \lambda^* > 0$, so that $\tau_{\text{rel}} = 1/\gamma^*$. It is noteworthy that the spectral representation formula for continuous time is completely monotonic.

The relaxation time plays an important role in the variance of an observable f and its average for a stationary process, which is needed to compute the error in Markov-chain Monte Carlo method, see Section 2.3. If we extract a subensemble of the stationary chain $\mathbf{T}$ so that it starts in x with probability $\pi(x)$, then [63, p. 155],

$$\begin{aligned} \mathrm{var}\left(\sum_y T^t(x,y)f(y)\right) &\leq \lambda^{*2t}\mathrm{var}(f) \leq e^{-2t/\tau_{\mathrm{rel}}}\mathrm{var}(f) && \text{for a discrete-time chain} \\ \mathrm{var}\left(\sum_y T^t(x,y)f(y)\right) &\leq e^{-2\lambda^* t}\mathrm{var}(f) = e^{-2t/\tau_{\mathrm{rel}}}\mathrm{var}(f) && \text{for a continuous-time chain} \end{aligned} \tag{1.73}$$

By the Convergence Theorem Eq. 1.65, we know that $\sum_y T^t_{xy} f(y) \to \langle f \rangle$ for any initial configuration $x \in \Omega$, i.e. the function $\mathbf{T}^t f$ approaches a constant function. We can now make a quantitative statement about this approach, if $t \geq \tau_{\mathrm{rel}}$, then the standard deviation of $\mathbf{T}^t f$ is bounded by $1/e$ times the standard deviation of f. Equality in Eq. 1.73 is achieved when f is the eigenvector ψ_{i^*} corresponding to λ^*.

Within the Monte Carlo method, we will be interested in the average of f, see Chapter 2. We note the running average of f, $S(t) = \sum_{s=0}^{t-1} f(X_s)$ in discrete time and $S(t) = \int_0^t f(X_s)ds$ in continuous time and σ^2 the limit of the variance $\lim_{t\to\infty} t^{-1}\mathrm{var}S_t$. We have, for a reversible chain [48, p. 134-135],

$$0 < \sigma^2 \leq 2\tau_{\mathrm{rel}}\mathrm{var}(f), \tag{1.74}$$

with $\mathrm{var}(f)$ the variance of f over the distribution π, as defined in Eq. 1.21. Bounds on $S(t)$ can also be derived.

The time τ_{rel} gives information on how long one should wait in order to retrieve the correct average of f, when the initial condition is an equilibrated configuration x, whereas τ_{mix} gives that information for a non-equilibrated configuration. Using τ_{rel}, it is however possible to get an upper and lower bounds for τ_{mix}, [63, p. 155][63, p. 268],

$$\begin{aligned} (\tau_{\mathrm{rel}} - 1)\log\left(\frac{1}{2\epsilon}\right) \leq \tau_{\mathrm{mix}}(\epsilon) \leq \log\left(\frac{1}{\epsilon\pi_{\min}}\right)\tau_{\mathrm{rel}} \text{ for discrete time} \\ \tau_{\mathrm{rel}}\log\left(\frac{1}{2\epsilon}\right) \leq \tau_{\mathrm{mix}}(\epsilon) \leq \log\left(\frac{1}{\epsilon\pi_{\min}}\right)\tau_{\mathrm{rel}} \text{ for continuous time} \end{aligned} \tag{1.75}$$

with $\pi_{\min}$ the minimum of π. One should note that this bounding is not always efficient to determine τ_{mix}. For instance, the cut-off phenomenon shows the limitation of Eq. 1.75 [64, 65]. A system exhibiting a cut-off will fall straight into equilibrium for t slightly higher than τ_{mix}, whereas for t lower than τ_{mix} the system is far away from equilibrium. In terms of total variation distance, Eq. 1.63, $d(t)$ will be nearly 1 for $t < \tau_{\mathrm{mix}}$ and will drop to 0 for $t > \tau_{\mathrm{mix}}$. In such cases, $\tau_{\mathrm{mix}} \gg \tau_{\mathrm{rel}}$, as there is no cut-off for Markov chains with $\tau_{\mathrm{mix}}/\tau_{\mathrm{rel}}$ bounded [63, p. 252], and it is not possible to infer τ_{mix} from τ_{rel}.

1.3.2.4 Relaxation time and autocorrelation times

Computing explicitly the eigenvalues of $\mathbf{T}$ is often no trivial task, all the more for a Markov chain devised on the space of all the possible configurations of a N-particle or N-spin systems, as considered in Section 2.2. Therefore, it is useful to link τ_{rel} to the correlations of the observables in practice, when the chain is in equilibrium. An observable f has an autocorrelation function C_f, as defined in Eq. 1.31, that decays exponentially for large t ($\sim e^{-t/\tau_{\text{exp}}(f)}$). We can define the exponential autocorrelation time $\tau_{\text{exp}}(f)$ [66, p. 6] as

$$\tau_{\text{exp}}(f) = \lim_{t\to\infty} \sup \frac{t}{-\log |C_f(t)|}, \tag{1.76}$$

and

$$\tau_{\text{exp}} = \sup_f \tau_{\text{exp}}(f). \tag{1.77}$$

The time τ_{exp} is the relaxation time of the slowest mode in the system. For reversible chains, τ_{exp} relates to τ_{rel} as follows, [48, 66],

$$\begin{aligned} \tau_{\text{exp}} &= -1/\log \lambda^* \leq \tau_{\text{rel}} \qquad \text{in discrete time} \\ \tau_{\text{exp}} &= \tau_{\text{rel}} \qquad\qquad\qquad\quad \text{in continuous time} \end{aligned} \tag{1.78}$$

More generally, for non-reversible chains, [66, 67] proved that in discrete time, τ_{exp} relates to the spectral radius of $\mathbf{T} - \Pi$, with $(\Pi f)_i = \sum_j \pi_j f_j, \forall i$, according to the generalization of the spectral radius formula [67].

As the time scales τ_{exp} and τ_{rel} reflect the slowest mode, it is useful to define the integrated autocorrelation time $\tau_{\text{int}}(f)$. The time $\tau_{\text{int}}(f)$ contains indeed the information given by all the different time scales that are present in C_f. These different time scales reflect the different eigenvalues of $\mathbf{T}$, outside the highest λ^*. The time $\tau_{\text{int}}(f)$ is defined [66] as

$$\tau_{\text{int}}(f) = \frac{1}{2} \sum_{t=-\infty}^{t=\infty} C_f(t), \tag{1.79}$$

so that we have more quantitative information than just upper and lower bounds for the variance of $\bar{f} = \frac{1}{N}\sum_{i=1}^{N} f_i$,

$$\begin{aligned} \text{var}(\bar{f}) &= \frac{\text{var}(f)}{N} \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{k}{N}\right) C_k(f) \\ &\approx \frac{2\tau_{\text{int}}(f)\text{var}(f)}{N} \leq \frac{1}{N}(2\tau_{\text{rel}})\text{var}(f). \end{aligned} \tag{1.80}$$

The times $\tau_{\text{int}}(\theta)$ and $\tau_{\text{exp}}(\theta)$ are different from each other and it was proved that $\tau_{\text{int}}(\theta) \leq \tau_{\text{exp}}(\theta)$ in realistic models [67]. The two characteristic times are equal, up to small corrections for small $\tau_{\text{int}}(\theta)$, only in the case where the autocorrelation function $C(k)$ is a pure exponential. Section 2.3 will discuss more thoroughly the different roles

of τ_{int} and τ_{exp} in assessing errors in the Markov-chain Monte Carlo method.

Finally, most of the results of this Section, summarized in Tab. 1.1, extend to infinite Ω. The matrix arithmetic in the finite case extends to the countable case without any problem, as do the concepts of irreducibility and aperiodicity [63]. Moreover, we will be concerned in Chapter 3 and Chapter 4 with irreducible chains whose stationary probability distribution π exists and is known. Then, the chain is positive recurrent, i.e. the average time to return to a visited state is finite, and $\lim_{t\to\infty} ||\mathbf{T}^t(x,\cdot) - \pi||_{TV} = 0$ [63, p. 279]. For continuous spaces, like the positions of spheres in a box, if it is measurable, it is possible to define a Markov kernel K that is the equivalent of $\mathbf{T}$ [68]. For reversible Markov chains, spectral techniques can again be tried [62].

Conclusion

In this chapter, we presented the equivalence of time and ensemble average, which is at the basis of statistical mechanics. It was then discussed how it is useful to describe physical systems by Markov processes. We then introduced the fundamental and tricky problem of having quantitative information on the convergence of Markov chains. The next chapter will deal with the reverse situation of knowing transition rates and retrieving the stationary distribution from them. In this new setting, the stationary distribution, i.e. the Maxwell-Boltzmann distribution, is known and we want to get equilibrated samples from it. These samples will allow us to compute the ensemble averages of macroscopic observables, e.g. pressure, orientation, etc., that describe physical systems at equilibrium. As we will see, the key to the Markov-chain Monte Carlo method is to devise special Markov chains, that may then be different from the physical Markov process used to describe the system, but that exhibit fast convergence properties. Fast algorithms are particularly needed around phase transitions where the correlation length becomes infinite and the autocorrelation times scale with the size of the system.

Towards equilibrium	From a non-equilibrated state	From an equilibrated state
Mixing	τ_{mix} Eq. 1.66	$\tau_{\text{rel}} \quad = 1/\lambda^* \ \mathbf{C}$ Eq. 1.72 $\quad = 1/(1-\lambda*) \ \mathbf{D}$

In equilibrium	Slowest mode	Observable f
Asymptotic decay of autocorrelation	$\tau_{\text{exp}} \quad = \tau_{\text{rel}} \ \mathbf{C}$ Eq. 1.77, 1.78 $\quad = -1/\log \lambda^* \leq \tau_{\text{rel}} \ \mathbf{D}$	$\tau_{\text{exp}}(f)$ Eq. 1.76
Asymptotic Variance ($\lim_{N\to\infty} \text{var}(\bar{f})/\text{var}(f)$)	$2\tau_{\text{rel}}/N$ Eq. 1.72, 1.74	$2\tau_{\text{int}}(f)/N$ Eq. 1.79, 1.80

$$\boxed{\tau_{\text{mix}} \succ \tau_{\text{rel}} \geq \tau_{\text{exp}} \geq \tau_{\text{int}}}$$

Table 1.1: Different time scales rule the approach of a Markov chain towards the equilibrium, e.g. the stationary regime. These characteristic times are linked together by bounding relationships, that may prove more or less informative according to the situation ($a \succ b$ stands for $ca \geq b$, with $c > 0$). In the following, we will be particularly concerned with assessing the autocorrelation times τ_{exp} and τ_{int}. Here, **C** stands for continuous-time chains and **D** stands for discrete-time chains.

CHAPTER 2

Monte Carlo Method

The beginning of stochastic computation takes place in the XVIIIth century with Buffon's needle problem, the earliest problem of geometric probability. In 1733, the French naturalist asked the question of the probability that a needle of a certain length will lend on a line of the parquet flooring, made of equally spaced and parallel planks [69, p. 43-45] and solved it in 1777 [70, p. 100-104]. Even if it was not Buffon's prime interest, the needle problem can be considered as the first Monte Carlo computation.

It was two centuries later, in the mid XXth century, that the idea of replacing deterministic calculation by stochastic computation really gained the interest of the scientific community. In the 1930s, E. Fermi experimented with the Monte Carlo method regarding neutron diffusion. More importantly, it was in the 1940s that the first version of modern Monte Carlo methods was born at the Los Alamos National Laboratory, along with its code name given by N. Metropolis to protect the secrecy of the project [7–9] (for historical review, [71, 72]). In 1946, the mathematician S. Ulam proposed to use random experiments to find the average distance travelled by a neutron before a collision with an atomic nucleus or how much energy would be released after a collision. Physicists were struggling to do so by using conventional and deterministic methods. J. von Neumann, convinced of the importance of the breakthrough, programmed the ENIAC computer to carry out Monte Carlo calculations and developed the use of the middle-square method to generate pseudorandom numbers and speed up the simulations [73].

Finally, Metropolis et al. [17] brought in 1953 a key contribution, which started fifty years of intense use of Markov-chain Monte Carlo methods. The Metropolis algorithm is commonly regarded as one of the most important algorithms, see Tab. 2.1. Recent and famous examples in statistical physics are the melting of hard disks [13, 74], where Monte Carlo simulations ended a debate of fifty years, the computation of configurational entropy in granular systems [75], investigating the Edwards hypothesis

1. **1946: J. von Neumann, S. Ulam and N. Metropolis, Monte Carlo method**
2. 1947: G. Dantzig, simplex method for linear programming
3. 1950: M. Hestenes E. Stiefel and C. Lanczos, Krylov subspace iteration method
4. 1951: A. Householder, decompositional approach to matrix computations
5. 1957: J.Backus, Fortran optimizing compiler
6. 1959-61: J. G. F. Francis, QR algorithm
7. 1962: T. Hoare, Quicksort
8. 1965: J. Cooley, fast Fourier transform
9. 1977: H. Ferguson and R. Foreade, integer relation detection algorithm
10. 1987: L. Greengard and V. Rokhlin, fast multipole algorithm

Table 2.1: Top Ten algorithms of the xx$^{\text{th}}$ century in chronological order according to editors J. Dongarra and F. Sullivan [88].

of equiprobability of jammed states [76], path integrals Monte Carlo for simulating bosons [77], notably Helium liquid [14], or bold diagrammatic Monte Carlo for strongly interacting fermions [78]. First successful in statistical physics, the Monte Carlo method soon became popular in a wide range of different sciences, such as cosmology with Hamiltonian Monte Carlo [79], chemistry [80, 81], economics [82–84], neuroscience [15] and artificial intelligence [16]. More generally, the Monte Carlo method provides a platform for solving multidimensional Bayesian problems [85–87]. However, naively implementing the Metropolis algorithm is bound to fail around phase transitions and for systems possessing complex energetic landscapes, as spin glasses, where metastable states are numerous.

In this chapter, we introduce the concept of stochastic sampling by Monte Carlo method, first the direct sampling in Section 2.1, and second, the Markov-chain Monte Carlo method in Section 2.2. We then address in Section 2.3 the limitations of the Markov-chain Monte Carlo method as the samples are no more independent, but correlated. Finally, we exhibit the solutions to those issues brought by the cluster algorithms for specific systems in Section 2.4 and Section 2.5.

2.1 Direct sampling for high-dimensional integration

S. Ulam's original idea at the base of the Monte Carlo method is to *"change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations"* [72]. As long as a problem can be given a probabilistic interpretation, it can be modeled using random numbers. As statistical physics formulates irregular phenomena as a stochastic ensemble, see Section 1.1, it leads to solving high

dimensional integrals Section 1.1.1 or the corresponding differential equations (see Section 1.2.3), which is analytically a hard task. Following Ulam's idea and relying on Bernoulli's law of large numbers, a high dimensional integral is now understood as an average over a collection of random configurations, that follow the correct probability distribution.

2.1.1 Weak and strong laws of large numbers and the Monte Carlo method

As demonstrated by the law of large numbers, the average of an observable θ over a probability distribution π, e.g. the Boltzmann distribution, written as the integral Eq. 2.1, may be estimated without bias by its empirical average over random and independent samples drawn from the same distribution π, as in Eq. 2.2. Such a collection of samples is said to be independent and identically distributed, *i.i.d.*, in particular they have all the same mean and variance. Obtaining such a collection is commonly referred to as direct sampling or static Monte Carlo. The estimator of the empirical mean is called the Monte Carlo estimator (for a more thorough study of variance and error in Monte Carlo method, see Section 2.3).

$$\langle\theta\rangle = \int d\boldsymbol{x}\, \theta(\boldsymbol{x})\pi(x), \tag{2.1}$$

$$\langle\theta\rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{\text{sample } i=1}^{N} \theta(\boldsymbol{x_i}^{\text{sampled}}). \tag{2.2}$$

The weak law of large numbers, commonly referred to as *LLN*, states that, for large N, the empirical average $\bar{\theta}$ of i.i.d. θ_i converges in probability to its expected value $\langle\theta\rangle$, i.e. for any $\epsilon > 0$,

$$\lim_{N\to\infty} P(|\bar{\theta} - \langle\theta\rangle| \geq \epsilon) = 0. \tag{2.3}$$

The weak LLN is a simple application of the Bienaymé - Chebyshev inequality to the random variable $\bar{\theta}$ for random variable θ_i with a finite variance. The Bienaymé - Chebyshev inequality states indeed that for a random variable θ_i and ϵ a positive real number,

$$P(|\theta_i - \langle\theta\rangle| \geq \epsilon) \leq \frac{\text{var}(\theta)}{\epsilon^2}. \tag{2.4}$$

However, a finite variance is only a sufficient condition for the weak LLN to hold [89, p. 76]. The strong law of large numbers states that $\bar{\theta}$ converges almost surely to $\langle\theta\rangle$, given $\langle\theta\rangle$ is finite. It means that the events where $\bar{\theta}$ does not converge towards $\langle\theta\rangle$ have a probability zero,

$$P(\lim_{N\to\infty} \bar{\theta} = \langle\theta\rangle) = 1. \tag{2.5}$$

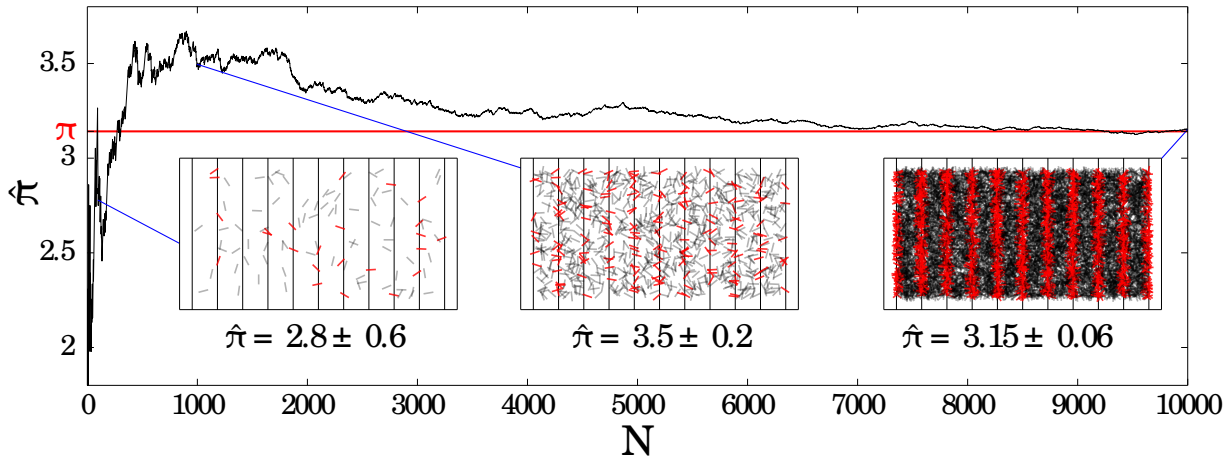


Figure 2.1: Evolution of the value of Buffon's estimator $\hat{\pi}$, Eq. 2.7, for a ratio $r = 0.33$, as the number N of needles increases. After, $N = 10^4$ throws, $\hat{\pi} = 3.115 \pm 0.06$. The insets show configurations at $N = 100$, $N = 1000$ and $N = 10000$, where needles hitting a line are represented in red, grey otherwise.

Monte Carlo methods obey both the weak and strong LLN.

Buffon's problem is a realization of the LLN. In this problem, a needle is thrown on the floor and the hit observable θ (1 if the needle crosses a planck line (= hit), 0 otherwise) is updated at each throw. This experiment can be reproduced numerically by generating the needle's position (x, y) and angle ϕ randomly, as shown on Fig. 2.1.1. The theoretical value of $\langle\theta\rangle$,

$$\langle\theta\rangle = \frac{2r}{\pi}, \tag{2.6}$$

with $r = l/d$ the ratio between the length l of the needle and the distance between lines d; We consider the case $r \leq 1$. This is estimated by $\bar{\theta}$,

$$\bar{\theta} = \frac{1}{N}\sum_{i=0}^{N}\theta_i. \tag{2.7}$$

This estimator has a variance of,

$$\sigma_{\bar{\theta}} = \frac{2r}{N\pi}\left(1 - \frac{2r}{\pi}\right), \tag{2.8}$$

as the total number of hits follows a binomial distribution of parameters N and $p = 2r/\pi$.

It is noteworthy that $\pi = 2r/\langle\theta\rangle$ may be estimated by the unbiased $\hat{\pi}$, also called Buffon's estimator,

$$\hat{\pi} = \frac{2r}{\bar{\theta}(r)} = \frac{2r}{\frac{1}{N}\sum_{i=0}^{N}\theta_i}, \tag{2.9}$$

whose variance is [90],[91, p.7],

$$\sigma_{\hat{\pi}} = \frac{\pi^4}{4r^2}\sigma_{\bar{\theta}} = \frac{\pi^2}{2N}\left(\frac{\pi}{r} - 2\right). \tag{2.10}$$

The subtlety of the Monte Carlo method is then to find the suitable sampling, i.e. the best way to create the collection of i.i.d. variables. The first sampling that was implemented was the direct sampling. Direct sampling, as in Buffon's problem, generates samples independently and directly from the desired distribution, unlike the Markov-chain Monte Carlo method, as will be discussed in Section 2.2.

Numerical sampling is strongly linked to the problem of generating random numbers from a computer. From now on, we assume that there exists a perfect uniform $[0,1]$ random variables generator, henceforth called $\mathbf{ran}(0,1)$. More generally, $\mathbf{ran}(a,b)$ stands for a uniform generator of random variables on the interval $[a,b]$. Historically, several methods were developed to solve the issue of using uniform distributions of random numbers to generate the desired nonuniform distributions [92]. We focus in the next Sections on two of the most important ones, the inversion sampling (see Section 2.1.2) and the acceptance-rejection method (see Section 2.1.3).

2.1.2 Inversion sampling

2.1.2.1 Method

Also known as inverse transform sampling, the first technique proposed by S. Ulam [72] is to form the inverse function of the cumulative function F of the wanted continuous distribution f, Eq. 2.11,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} dx' f(x') = F(x - dx) + f(x)dx. \tag{2.11}$$

The function F takes values in $[0,1]$ and is at least weakly monotonic. It allows us to define its inverse F^{-1} as,

$$F^{-1}(v) = \inf\left\{x' | F(x') \geq v,\ 0 < v < 1\right\}. \tag{2.12}$$

When F is strictly monotonic, i.e. $f(x) > 0, \forall x$, the inverse function F^{-1} is directly,

$$x = F^{-1}(v),\ 0 < v < 1. \tag{2.13}$$

If v is drawn from $\mathbf{ran}(0,1)$, then $F^{-1}(v)$ follows the distribution F and conversely, if X follows the distribution F, $F(X)$ follows $\mathbf{ran}(0,1)$ [92, p. 28].

The discrete version of the inversion method is the tower sampling method. From the discrete probability distribution $\{f_1, f_2, \ldots, f_n\}$, we construct the cumulative distribution F_k as,

$$\begin{aligned} F_k &= \sum_{i=1}^{k} f_i, k \geq 1 \\ F_0 &= 0. \end{aligned} \tag{2.14}$$

As for the continuous case, a random number ν is drawn from $\mathbf{ran}(0,1)$. It samples the state k, which satisfies

$$F_{k-1} < \nu \leq F_k. \tag{2.15}$$

The condition Eq. 2.15 is equivalent to the inverse definition Eq. 2.12 and Eq. 2.13 of the continuous case. In practice, finding $k \in \{1, \dots n\}$ fulfilling Eq. 2.15 may be sped up by using binary search, which runs at worst in logarithmic time with n. The target value ν is compared to the middle element of the array $\{1, \dots n\}$. If not equal, the search continues on the half of the array where the target can lie, until it is successful.

This inversion method can be interpreted as a sample transformation, stressing interestingly the equivalence between integration and sampling [12]. Any change of variables in an integral can be translated into a change on how to generate random numbers. The goal of the inversion method is to transform the integral over the flat uniform distribution $\mathbf{ran}(0,1)$ into the integral of the wanted distribution f defined on interval $[a,b]$. This integral transformation,

$$\int_0^1 d\nu \to \int_a^b dx f(x), \tag{2.16}$$

is indeed equivalent to the sample transformation,

$$\{\text{sample } \nu |\ \mathbf{ran}(0,1)\} \to \{\text{sample } x |\ f(x)\}, \tag{2.17}$$

which follows from the change of variable,

$$d\nu = K dx f(x), \tag{2.18}$$

with K a constant. It leads to,

$$\nu = KF(x) + K', \tag{2.19}$$

with K' a constant. We are back to the formulation of Eq. 2.12and Eq. 2.13.

In particular, this sample transformation method will be used to sample the maximum energy increase in the irreversible factorized Metropolis paradigm discussed in the next Chapter 3.

2.1.2.2 Inversion method for sampling Gaussian random numbers

The sample transformation is particularly useful for producing Gaussian random numbers y following the probability distribution,

$$\pi(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right], \tag{2.20}$$

with μ and σ two parameters coding respectively for the mean and variance of the distribution. By the change of variables $x = (y-\mu)/\sigma$, one can always come back to the normal distribution,

$$\pi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]. \tag{2.21}$$

Sampling Gaussian numbers described by Eq. 2.21 can be done by generating a large amount of independent random numbers from the uniform distribution $\mathbf{ran}(-1/2, 1/2)$. Their sum, rescaled by a factor $1/12$ (which comes from the variance of $\mathbf{ran}(-1/2, 1/2)$), will follow the normal distribution, as proved by the central limit theorem, but still remains an approximation.

It is also possible to use a multidimensional sample transformation to generate without approximation those Gaussian numbers. Considering the probability distribution of two independent Gaussian random numbers x and y,

$$\int_{-\infty}^{+\infty} \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} \int_{-\infty}^{+\infty} \frac{dy}{\sqrt{2\pi}} e^{-x^2/2} = \int_{-\infty}^{+\infty} \frac{dxdy}{2\pi} e^{-(x^2+y^2)/2}, \tag{2.22}$$

we introduce polar coordinates ($x = r\cos\phi, y = r\sin\phi, dxdy = rdrd\phi$),

$$\int_{-\infty}^{+\infty} \frac{dxdy}{2\pi} e^{-(x^2+y^2)/2} = \int_0^{2\pi} \frac{d\phi}{2\pi} \int_0^{\infty} rdre^{-r^2/2} = \int_0^{2\pi} \frac{d\phi}{2\pi} \int_0^{\infty} due^{-u}. \tag{2.23}$$

On the right-side of Eq. 2.23, the first integral is sampled by $\phi = \mathbf{ran}(0, 2\pi)$ and the second integral by $r = \sqrt{2u}$, with $u = -\log(\mathbf{ran}(0,1))$, leading to the sampling of x and y. More generally, sampling d independent Gaussian random numbers is equivalent to sampling a random point in a d-dimensional unit sphere.

This equivalence is closely linked to the Maxwell distribution of velocities in a gas [12, p. 103]. Considering N spheres of mass m in a box of dimension d, a set of velocities $\{v_1 \dots v_N\}$ is legal if it corresponds to the correct value of the kinetic energy K,

$$K = \frac{1}{2} m \sum_{i=1}^{N} {v_i}^2. \tag{2.24}$$

Applying the equiprobability principle, see Section 1.1, all legal configurations have the same unnormalized weight $\pi = 1$, illegal configurations have a null weight. Each velocity has d components, so that a legal set of velocities correspond to a point on a dN-dimensional sphere of radius $r = \sqrt{(2K/m)}$. As $K/(dN)$ is the mean kinetic energy per degree of freedom, it is equal to $\frac{1}{2}k_bT = 1/2\beta$. Sampling a random point

on the hypersphere is then the same as sampling dN Gaussian numbers with a variance of $1/m\beta$. For $d = 2$, the probability distribution of a single component of the velocity is then,

$$\pi(v_x)dv_x = \sqrt{\frac{m\beta}{2\pi}} \exp\left(-\frac{1}{2}\beta m v_x^2\right). \tag{2.25}$$

The Maxwell-distribution agrees well with the velocity distribution obtained by molecular dynamics [12]. As velocities form an independent sampling problem of random points on the surface of a hypersphere, it is natural that velocities disappear from a Monte Carlo simulation.

Finally, it is noteworthy that the possibility to directly simulate Gaussian numbers is also at the core of the path integral method for Quantum Monte Carlo. Using the Lévy construction [93], one can directly sample paths contributing to the density matrix of a free particle or for a harmonic oscillator.

2.1.3 Acceptance-rejection method

The second method is the acceptance-rejection method, also called *hit-or-miss*, developed also by J. von Neumann [72]. It is particularly useful when it is difficult to form the inverse function needed in the inversion sampling method, notably for multidimensional Hamiltonian encountered in physics.

The acceptance-rejection method's idea is to randomly generate a configuration that is then accepted or not, depending on its probability weight. Buffon's problem can be interpreted as an acceptance-rejection sampling. The position and angle of the needle are generated randomly and this configuration is accepted, if the needle hits a floor line. The sum in Eq. 2.7 is simply equal to the total number of valid configurations.

The acceptance-rejection method can be easily applied to particle systems in physics, for which the inversion method fails. For a N hard-sphere system of radius σ in a hypercube L^d, the position of each sphere is generated by sampling d random numbers $\mathbf{ran}(0, L)$. Once the configuration is generated, its validity is checked by looking for overlap between two spheres or with the hypercube's boundary, if the boundary condition is not periodic. Such sampling allows for the recovery of the partition function (here with periodic boundary condition),

$$Z = \int_0^L \left\{d^d \boldsymbol{r_i}\right\}_{1\leq i\leq N} \prod_{i,j} H(|\boldsymbol{r_i} - \boldsymbol{r_j}| - 2\sigma), \tag{2.26}$$

with H being the Heaviside function. Fig. 2.1.3 illustrates the sampling of valid configurations for two hard disks in a square with periodic boundary. This sampling is actually equivalent to the sampling of a point in the free volume that is left, when the position of the first sphere is fixed.

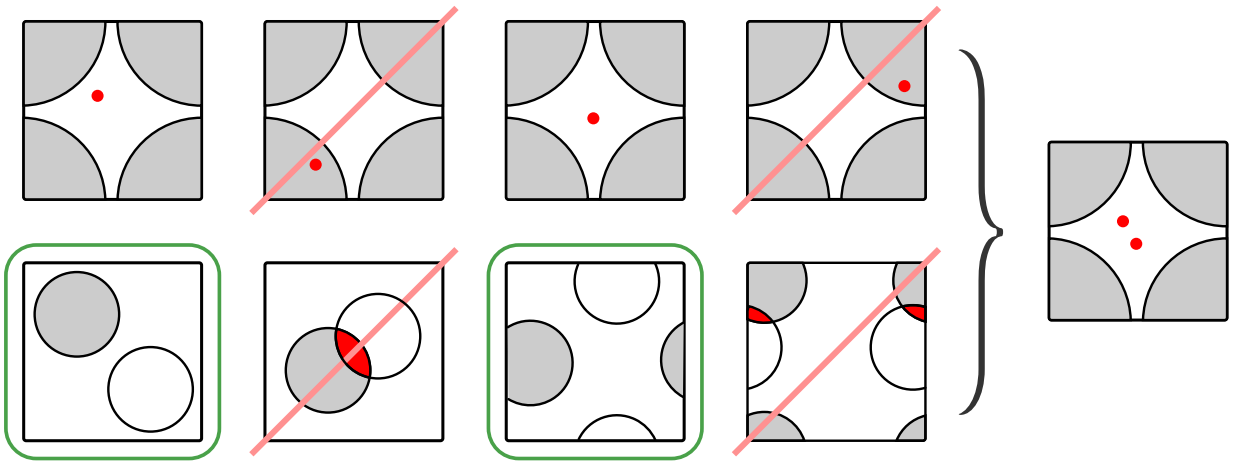


Figure 2.2: Direct sampling by acceptance-rejection method of the correct configurations for a two-hard-disks system with periodic boundary. Configurations are produced randomly by drawing four random numbers (one for each of the coordinate of each sphere). **Bottom row:** represents the sampled configurations, that are either accepted (green contour) or rejected (red slash). **Top row:** represents each configuration in the phase space (white area). In this example, only two configurations are accepted.

Acceptance-rejection sampling can also be used to sample more general probability distribution than distributions with equiprobably valid configurations, as in hard spheres or Buffon's problem. For instance, we want to generate samples following the distribution f, defined on the interval $[a, b]$ and whose normalizing constant is not known. To do so, we draw a box extending from a to b and from 0 to y_0 with $y_0 > \sup_{x\in[a,b]} f(x)$. Random points are then uniformly sampled in this box and considered valid when they fall below f.

Fig. 2.1.3 exhibits such sampling by acceptance-rejection method of a general distribution. It comes down to the sampling of the free volume of a two-hard-disks gas. As for Buffon's problem, the hit observable Θ (1 if the configuration is valid, 0 otherwise) follows a binomial law of parameter $n = 1, p = \int_a^b dx f(x)/(y_0|b-a|)$. Noting $I = \int_a^b dx f(x)/(y_0|b-a|)$, it follows that

$$\begin{aligned} \langle\Theta\rangle &= I \\ \mathrm{var}(\Theta) &= I(1-I). \end{aligned} \tag{2.27}$$

It is noteworthy that the variance of Θ is 0, when $I = 1$, meaning the sampling box area is exactly the same as the area under the curve of f. This is the case only when the points are directly sampled from f.

A limitation of the acceptance-rejection sampling is that the same amount of time, in the case of a uniform random number generation, is spent sampling high probability configurations as the low probability ones. First, the proposal distribution for the generation of random numbers does not have to be the uniform **ran** distribution but

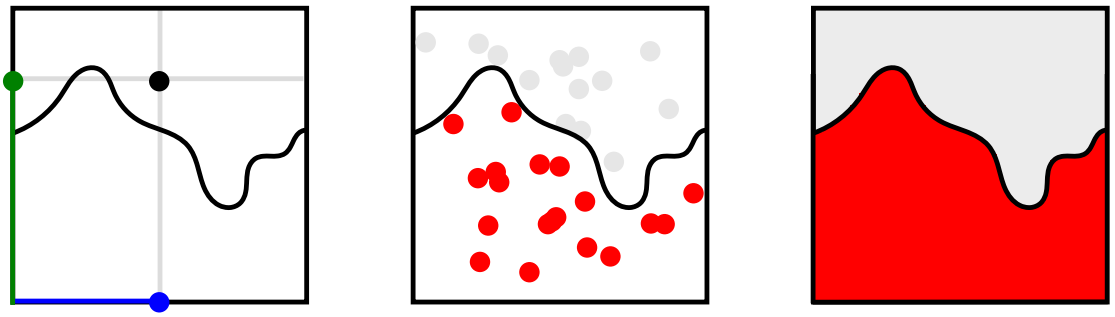


Figure 2.3: Direct sampling by acceptance-rejection method for a general distribution f on an interval $[a, b]$. To achieve such a sampling, the maximal value of f needs to be known. **Left:** For each single random point, two random numbers (green and blue spheres) are drawn uniformally in the square box. **Center:** The area under the curve f is sampled by registering a hit, $\Theta_i = 1$ (red), each time a random point is under the curve, otherwise $\Theta_i = 0$ (grey). **Right:** As the number of samples N grows, the estimator $\bar{f} = \frac{1}{N}\sum_{i=0}^{N}\Theta_i$ converges to $\langle f \rangle = \int_a^b dx f(x)$, with a variance $\mathrm{var}(\bar{f}) = \frac{1}{N}I(1-I)$, with I the ratio between the red area and the total box area.

only needs to be easily sampled from, such as by inversion sampling, and to be as least as high as the distribution f in any point from the interval $[a, b]$. A smarter procedure is then to use importance sampling, as explained below.

2.1.4 Importance sampling

Even if it is clear that the acceptance-rejection method with a box support will converge to the correct value, as stipulated by LLN, this method is inefficient for sampling averages over a Boltzmann distribution as Eq. 1.9. Most of the simulation time is indeed wasted at computing configurations with very small Boltzmann weights. Importance sampling addresses this issue by allowing us to sample many points in regions where the Boltzmann weight is large and few elsewhere.

As previously, we consider the probability distribution function f on the interval $[a, b]$, noting its integral $I = \int_a^b f(x)dx$. I can be rewritten as

$$I = \int_a^b \frac{f(x)}{g(x)} g(x)dx, \tag{2.28}$$

with g a nonnegative probability distribution function that now controls the distribution of the sampling points in the interval $[a, b]$. Moreover, we choose g so that it is the derivative of G a nonnegative, non-decreasing function with $G(a) = 0$ and $G(b) = 1$. I can then be written as,

$$I = \int_a^b \frac{f(x)}{g(x)} dG(x) = \int_0^1 \frac{f(G^{-1}(x))}{g(G^{-1}(x))} dx \tag{2.29}$$

I can now be estimated by generating N random uniform numbers $x_i = \mathbf{ran}(0,1)$,

$$I = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N} \frac{f(G^{-1}(x_i))}{g(G^{-1}(x_i))}. \tag{2.30}$$

We note $\hat{I}$ the estimate of I by the sum in Eq. 2.30. The variance $\sigma^2_{\hat{I}}$ of this estimator is [11] ,

$$\sigma^2_{\hat{I}} = \frac{1}{N}\left[\left\langle \left(\frac{f}{g}\right)^2 \right\rangle - \left\langle \frac{f}{g} \right\rangle^2\right] \tag{2.31}$$

Eq. 2.31 reveals the importance of the choice of g. If g is proportional to f, then the variance of the estimator $\hat{I}$ simply vanishes. If g is only a constant, as for a square sampling box, then the variance becomes large. Although importance sampling methods have proven powerful in some domains, it is not possible as of today to construct a transformation, as shown in Eq. 2.28 and Eq. 2.29, that will generate points with a probability distribution proportional to the Boltzmann factor. A necessary condition for achieving such a sampling is to compute analytically the partition function of the system, Eq. 1.7. However, if such computation would be possible for the system under study, there would hardly be any need for a Monte Carlo simulation.

Direct sampling's advantage lies in the statistical independence of the samples, as the configurations are produced without any correlations between them. However, producing configurations from high-dimensional probability distributions in such a manner is no trivial task. In most cases, the rate of rejection increases exponentially with the number of dimensions. It leads to prohibitively long simulation times, a problem known as the *curse of dimensionality* [94]. Smart methods, like importance sampling, have been devised to address such issue. But, unfortunately, they lose their power in most cases encountered in physics, notably because of the impossibility to compute the partition function. As will be explained in the next Section 2.2, the Markov-chain Monte Carlo method brings a solution to this problem, by changing conceptually the way samples are proposed. Now a stochastic process, in practice always chosen Markovian, see Section 1.1.3 and Section 1.2, is simulated on the configuration space Ω, so that its stationary distribution is π. Markov-chain Monte Carlo method is also called the dynamic Monte Carlo method.

2.2 Markov-chain Monte Carlo method

In 1953, Metropolis et al. [17], and later the theoretical justification of Wood and Parker [95] and the generalization of Hastings [18], solves the partition function and high rejection issues, which was encountered by direct sampling. The Metropolis algorithm

provides indeed the right probability distribution of the configurations, from the moment one can compute a function f proportional to the density π. Therefore the issue of the computation of the partition function, or any normalizing factor, vanishes.

To do so, a sequence of samples is now obtained by running a Markov chain. Hence the name of Markov-chain Monte Carlo method. Such algorithms can be then understood as a Markov process following its own Master equation, see Section 1.2, with a special choice of transition probabilities, see Section 2.2.2, which depend only on ratios of the form $\pi(x)/\pi(x')$ with x and x' two configurations and π the Boltzmann probability distribution. A Markov-chain Monte Carlo method is then equivalent to inventing a stochastic time evolution for a given system.

Unlike samples produced by direct sampling, samples originating from a Markov-chain Monte Carlo method are thus correlated and assessment of errors asks for more care, see Section 2.3. The time evolution does not need, however, to correspond to any real physical dynamics; rather, the dynamics is to be chosen on the basis of its computational efficiency. The situation is then different than in molecular dynamics simulations [43]. In molecular dynamics, the microcanonical ensemble is sampled by numerically solving Newton's equations of motion. The extension to averages over the canonical ensemble is ensured by coupling the system to a heat bath [11, p. 139]. Particles of the system can undergo collisions with the heat bath, during which they gain a new velocity, drawn from the Maxwell-Boltzmann distribution, see Section 2.1.2.2. The mixing of Newtonian dynamics and stochastic collisions changes the simulation into a Markov process. It is however possible to keep the simulation deterministic by using an extended Lagrangian instead of introducing a heat bath [11, p. 147].

2.2.1 Principles of the Markov-chain Monte Carlo method

In the Markov-chain Monte Carlo method, a Markov chain is devised on the space state Ω, such that the chain is irreducible, aperiodic and its unique steady-state distribution is precisely the wanted probability distribution π, see Section 1.3.2. Under the conditions of irreducibility and aperiodicity, it can be proven that the chain follows the strong law of large numbers, see Section 2.1.1, and the central limit theorem, see Section 2.3.3 [66]: long-time average of any observable f converges with probability 1 to its average on the stationary distribution π and with fluctuations of size proportional to the square-root of the simulation time. Below are reviewed the necessary conditions that the matrix $\mathbf{T}$ of the Markov chain has to fulfill, in order to produce the correct stationary distribution.

As explained in Section 1.2, a homogeneous Markov chain is fully determined by the initial probability distribution and the transition probabilities contained in the transfer matrix $\mathbf{T}$. In the Markov-chain Monte Carlo method, the initial probability distribution is in most cases a Dirac distribution centered on a valid initial state.

The transition probabilities need to ensure that the Markov chain is irreducible, see Section 1.3.2.1. Furthermore, the transition probabilities must satisfy the conservation

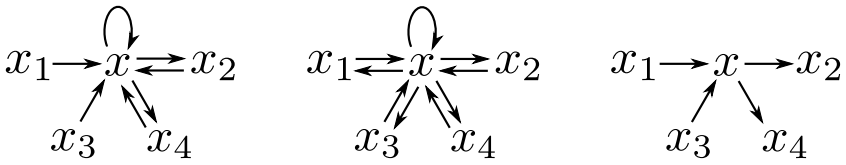


Figure 2.4: Representation of the probability flows (arrows) between a given configuration x and its neighbors $\{x_i\}$. Global balance must always be fulfilled in order to retrieve the correct probability distribution. Detailed balance and maximal global balance are particular cases of global balance. Detailed balance imposed a symmetric equilibrium of flows between any two configurations, whereas maximal global balance offers an out-of-equilibrium scheme where all flows are irreversible.

condition Eq. 1.37 and the global-balance condition Eq. 1.46, so that the stationary distribution is π. Usually in Markov-chain Monte Carlo method, as the conservation condition is trivially fulfilled, the global-balance condition actually refers to the combination of both equations into the incompressibility condition Eq. 1.39.

As seen in Section 1.2.2.1, the global-balance condition is a necessary condition and is fulfilled the moment the sufficient condition of detailed balance is satisfied. As illustrated in Fig. 2.2.1, it is useful to define the extreme case of flows satisfying the maximal global-balance condition, i.e. flows that are strictly irreversible on the configuration space. The rules followed by the Markov chain with transition probabilities, or rate for continuous-time chain, $p(x \to x')$ are, for all x in Ω,

	Transition probabilities	Stationary probability flows
Global balance (Necessary)	$\sum_{x'} \pi(x)p(x \to x') = \sum_{x'} \pi(x')p(x' \to x)$	$\sum_{x'} \phi_{(x\to x')} = \sum_{x'} \phi_{(x'\to x)}$
Detailed balance (Sufficient)	$\forall x', \pi(x)p(x \to x') = \pi(x')p(x' \to x)$	$\forall x', \phi_{(x\to x')} = \phi_{(x'\to x)}$

(2.32)

Detailed-balance Markov chains that have the uniform probability as the stationary distribution are characterized by symmetric transfer matrix $\mathbf{T}$. Reciprocally, if a transfer matrix $\mathbf{T}$ is symmetric, detailed balance is satisfied and the uniform distribution is the stationary distribution. For continuous systems with detailed balance, it is sometimes possible to transform the coordinates until the stationary distribution is the uniform one. The matrix $\mathbf{T}$ is then symmetric. For discrete systems, states may be broken into degenerate sub-states, in order to retrieve a uniform stationary distribution. These transformations are actually the ones used for the inversion sampling and tower sampling method, see Section 2.1.2, where samples following the distribution π are transformed so they follow the uniform distribution **ran**.

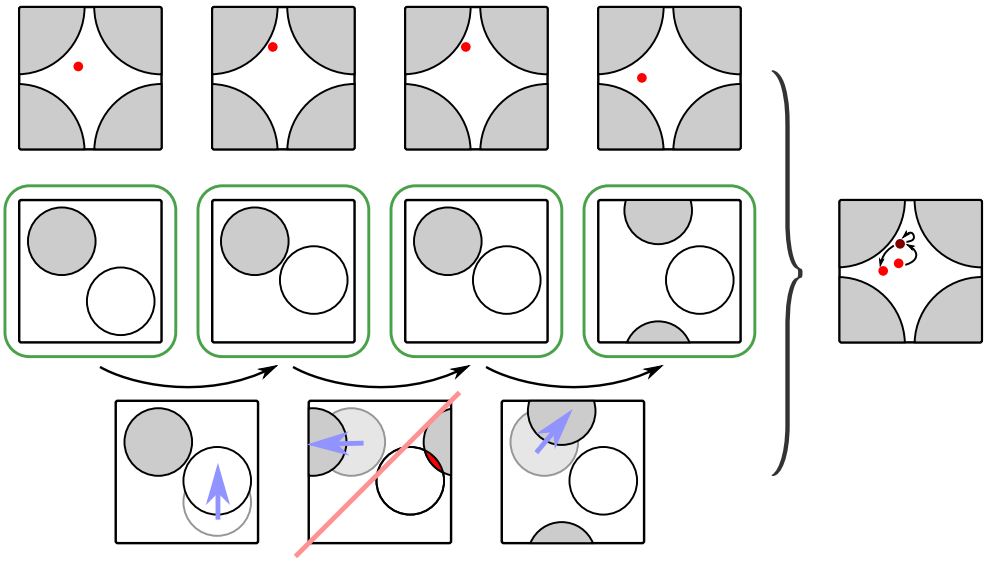


Figure 2.5: Metropolis sampling of the correct configurations for a two-hard-spheres system. The key contribution of the Metropolis paper [17] was the idea that instead of choosing configurations randomly, then weighting them with $\exp(-\beta E)$, we choose now configurations with a probability $\exp(-\beta E)$ and weight them evenly.

2.2.2 Metropolis algorithm

Historically, the traditional Metropolis algorithm obeys the symmetric detailed-balance condition. Its generality and ease of use has opened the way to a large class of detailed-balance Markov-chain Monte Carlo algorithms. But, the symmetry and locality enforced by the detailed-balance condition often lead to dynamics similar to a diffusion process, see Section 2.3. However we will review, in Section 2.4, detailed-balance algorithm that avoid this issue, notably by using efficient asymmetric a priori probabilities described in Section 2.2.3. On the other hand, algorithms satisfying only global balance display out of equilibrium dynamics and are often quicker than Markov chains fulfilling the detailed balance, as will be explained in Chapter 3.

Starting in a valid state x, the Metropolis algorithm picks out a configuration x' in a uniform and random manner. In most cases, the choice of x' is limited to a finite range of configurations close to the initial one x. The probability to choose x from x' is the same as choosing x' from x. For hard-sphere systems, a random sphere will be updated by a random vector δ, so that $|\delta| < \delta_0$. The move from x to x' is then accepted with the acceptance probability, hereafter called the Metropolis filter,

$$p^{\text{acc}}_{\text{Metro}}(x \to x') = \min\left(1, \frac{\pi(x')}{\pi(x)}\right). \tag{2.33}$$

The Markov chain is irreducible, as there is always a positive probability of going from a state s to a state x'. Detailed balance is also easily checked,

$$\pi(x)p^{\text{acc}}_{\text{Metro}}(x \to x') = \min\left(\pi(x), \pi(x')\right) = \pi(x')p^{\text{acc}}(x' \to x). \tag{2.34}$$

As exhibited in Fig. 2.2.2 for a two hard-sphere system, if the move $x \to x'$ is accepted, the chain goes from the state x to the state x' and the state x' is stored. If the move is rejected, the chain remains on x and the state x is stored again. The same process is then repeated from the current state.

Acceptance-rejection method in direct sampling produces uniformly random configurations and weights them thereafter according to π. In the Metropolis algorithm, all valid configurations are weighted the same way, but they are produced with a probability π. Rejections now result in the repetition of configurations of higher probabilities, which is consistent with the picture of probability as being equivalent to the time spent by the Markov chain in the respective configuration, see Section 1.3.

Even if it was not stated as such in the original paper [17], the Metropolis algorithm realizes a Markov process, described by a Master equation leading to the same physical probability distribution $\boldsymbol{\pi}$, see Section 1.2.3. The global-balance condition Eq. 1.46 is then simply the Master equation in the stationary regime, where $d\pi^t(x)/dt = 0$. However, as will be discussed later on, algorithms with transition probabilities no longer equivalent to the Metropolis ones are not a faithful representation of the Master equation of the physical system anymore.

2.2.3 A priori probabilities

A generalization of the Metropolis algorithm is due to Hastings [18], which introduces a priori probabilities $\mathcal{A}$. The move from x to x' is now proposed according to the a priori probability $\mathcal{A}(x \to x')$. The transition probability $p(x \to x')$ to actually make the move $x \to x'$ is now divided into two independent components, the probability of considering the move from x to x', $\mathcal{A}(x \to x')$ and the probability $p^{\text{acc}}(x \to x')$ of accepting it,

$$p(x \to x') = \mathcal{A}(x \to x') \cdot p^{\text{acc}}(x \to x') \tag{2.35}$$

The Metropolis algorithm can be seen as a special case, where $\mathcal{A}$ is symmetric and uniform for all possible moves. Therefore $p(x \to x')$ is equivalent to $p^{\text{acc}}(x \to x')$.

The detailed-balance condition can be now written as,

$$\frac{p^{\text{acc}}(x \to x')}{p^{\text{acc}}(x' \to x)} = \frac{\pi(x')}{\mathcal{A}(x \to x')} \frac{\mathcal{A}(x' \to x)}{\pi(x)}, \tag{2.36}$$

and the generalized Metropolis-Hastings acceptance probabilities,

$$p^{\text{acc}}_{\text{Metro}}(x \to x') = \min\left(1, \frac{\pi(x')}{\mathcal{A}(x \to x')} \frac{\mathcal{A}(x' \to x)}{\pi(x)}\right). \tag{2.37}$$

From Eq. 2.37, one can see that a symmetric a priori probability, as a sphere or square in the continuous space around x [12], disappears from the acceptance probability,

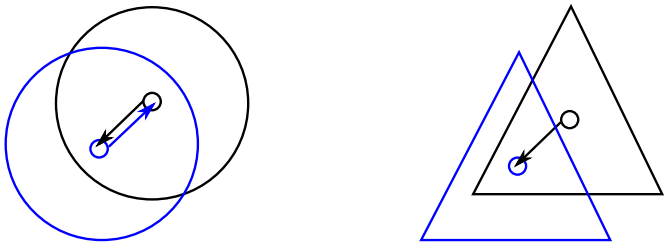


Figure 2.6: **Left:** Symmetric a priori probabilities; there is the same probability to propose a move from the blue position to the black position than to move from the black position to the blue one. Symmetric a priori probabilities disappear from the generalized Metropolis-Hastings acceptance probability expression, Eq. 2.37. **Right:** Asymmetric a priori probabilities; it is not possible to propose the move from the blue position to the black one, whereas one can a priori move from the black to the blue position. However, the latter move is not accepted in a detailed-balance scheme, as written in Eq. 2.37.

as $\mathcal{A}(x \to x')/\mathcal{A}(x' \to x) = 1$, and so cannot bring any drastic change. On the other hand, an asymmetric a priori probability, as a triangle on the continuous space, adds a degree of freedom in the construction of the Markov process, see Fig. 2.6. A priori probabilities are the equivalent of importance sampling, see Section 2.1.4, in a Markov-chain Monte Carlo scheme. This can be used to lessen the rate of rejection and allows for bigger moves. For instance, in the case where $\pi(x)$ is large and $\pi(x')$ is small, proposing the move $x \to x'$ will lead most of the time to a rejection. With an asymmetric a priori probability, it is possible to propose this move less frequently and avoid wasting simulation time. This idea of proposing moves that would be most likely successful is a key concept in the elaboration of the irreversible factorized Metropolis paradigm, presented in Chapter 3.

As mentioned in [12], an interesting case is when $\mathcal{A}(x \to x') = \pi(x')$ and $\mathcal{A}(x' \to x) = \pi(x)$, leading to a generalized Metropolis acceptance probability of one. Such a situation is equivalent to direct sampling, Section 2.1 and is at the core of cluster algorithms, see Section 2.4. The former rejections of not valid configurations are now replaced by the rejections of not valid moves. Even if the prohibitive limitations of direct sampling, in particular importance sampling, apply in this formalism, a priori probabilities can play a great role for systems or subsystems where direct sampling can be nearly implemented. A priori probabilities can be then understood as the equivalent in numerical computation of perturbation theory in theoretical physics.

2.2.4 Heat-bath algorithm

An alternative to the Metropolis algorithm is the Glauber dynamics or heat-bath algorithm, named after the thermalizing of subsystems with their environment. This algorithm is more famous under the name of Gibbs sampler in statistics [85, 86]. In this dynamics, moves are proposed through a priori probabilities that implement di-

rect sampling for subsystems. Glauber dynamics was first developed by Barker (1965) [96] for a study of a proton-electron plasma.

We consider a Markov chain devised on a system of interacting particles of inverse temperature β. The chain is now in a configuration C_1 of energy E_1 and probability weight $\pi(C_1)$. A configuration C_2 of energy E_2 and probability weight $\pi(C_2)$ is constructed by choosing a random particle j and updating its coordinates by a random displacement vector $\boldsymbol{\delta}$, such that $|\boldsymbol{\delta}| < \delta_0$. We note the difference in energies between the two configurations ΔE. The next configuration in the Markov chain will be,

$$\left.\begin{aligned} C_1 \text{ with probability } p_{C_1} &= \frac{\pi(C_2)}{\pi(C_1)+\pi(C_2)} = \frac{e^{-\beta E_1}}{e^{-\beta E1}+e^{-\beta E2}} = \frac{1}{1+e^{-\beta\Delta E}} \\ C_2 \text{ with probability } p_{C_2} &= \frac{\pi(C_1)}{\pi(C_1)+\pi(C_2)} = \frac{e^{-\beta E_2}}{e^{-\beta E1}+e^{-\beta E2}} = \frac{1}{1+e^{\beta\Delta E}} \end{aligned}\right\} p_{C_1}+p_{C_2}=1. \tag{2.38}$$

Such a scheme is detailed-balance, hence correct. But, unlike in the Metropolis scheme, there is no consideration about the sign of ΔE. Moreover, it is noteworthy that $p(C_1 \to C_2) = p_{C_2} = p(C_2 \to C_2)$, so that the initial configuration has no effect on the move. Therefore the name *heat-bath*, as the scheme acts like a thermalization of a spin with its local environment.

This scheme can be rewritten as follows: First choose a subsystem $\{1,2\}$ by picking $\boldsymbol{\delta}$. Second, sample directly C_1 with probability $\pi_{\{1,2\}}(C_1)$ or C_2 with $\pi_{\{1,2\}}(C_2)$. This is easily made by drawing a single random number $\nu = \mathbf{ran}(0,1)$. If $\nu < \pi_{\{1,2\}}(C_1)$, then C_1 is sampled, C_2 otherwise.

It is also possible to interpret the Glauber dynamics as a Metropolis scheme with a priori probabilities,

$$p(C_1 \to C_2) = P(\boldsymbol{\delta})p_{C_2} = \mathcal{A}(C_1 \to C_2)p^{\text{acc}}(C_1 \to C2) \tag{2.39}$$

with,

$$\mathcal{A}(C_1 \to C_2) = P(\boldsymbol{\delta})p_{C_2} = \frac{\pi(C_2)P(\boldsymbol{\delta})}{\pi(C_1)+\pi(C_2)}. \tag{2.40}$$

$\mathcal{A}(C_1 \to C_2)$ is actually equal to the probability weight of C_2, once rescaled to the subsystem $\{1,2\}$. As explained in Section 2.2.3, this is equivalent to direct sampling in the subsystem. The acceptance probabilities are indeed always equal to 1. From Eq. 2.37, as $P(\boldsymbol{\delta}) = P(-\boldsymbol{\delta})$,

$$\begin{aligned} p^{\text{acc}}_{\text{Metro}}(C_1 \to C_2) &= \min\left(1, \frac{\pi(C_2)}{\mathcal{A}(C_1 \to C_2)}\frac{\mathcal{A}(C_2 \to C_1)}{\pi(C_1)}\right) \\ &= \min\left(1, \frac{\pi(C_2)}{\frac{\pi(C_2)P(\boldsymbol{\delta})}{\pi(C_1)+\pi(C_2)}}\frac{\frac{\pi(C_1)P(-\boldsymbol{\delta})}{\pi(C_1)+\pi(C_2)}}{\pi(C_1)}\right) \\ &= 1. \end{aligned} \tag{2.41}$$

It is possible to generalize the Glauber dynamics to scheme where the probability $P(\cdot \to \{1,2\})$ to pick the subsystem $\{1,2\}$ is not the same from C_1 or C_2. The probabilities are now [18],

$$p_{C_2} = \frac{\pi(C_2)P(2 \to \{1,2\})}{\pi(C_1)P(1 \to \{1,2\}) + \pi(C_2)P(2 \to \{1,2\})}. \tag{2.42}$$

$P(\cdot \to \{1,2\})$ can be understood as an additional a priori probability.

The Glauber dynamics can easily be generalized thanks to the tower sampling method, see Section 2.1.2, to scheme with discrete subsystems with more than 2 configurations. Likewise, the generalization to continuous subsystems can be made using inversion method, see Section 2.1.2. The Glauber dynamics is particularly used in spin systems, where it is easy to pick subsystems by picking a random spin.

2.3 Convergence and error

Monte Carlo method, though it supplies quasi-exact predictions, remains a stochastic method. Moreover, samples produced through Markov-chain Monte Carlo are correlated, as, although Markov chains do converge to the stationary distribution π, Eq. 1.65, they do so following more or less long time scales (see Section 1.3.2). It is then required to carefully assess the statistical error for a set of samples, which is no trivial task [66, 97].

Errors come from two distinct sources. First, the initialization bias describes the correlations that a sample may have with the initial configuration used by the algorithm. One should wait long enough, so that the initialization bias disappears and the system is thermalized (see Section 2.3.1). Second, the ensemble average of an observable f is often estimated by taking the average over the set of samples and this estimation comes with an error (see Section 2.3.2), all the more important as the samples are correlated in Markov chains (see SecErrorAutocorr). One has then to wait long enough between two thermalized samples, so that they are slightly correlated. However, around second-order phase transitions, the time needed to produce two independent samples diverges with the system size because of the critical slowing-down phenomenon (see Section 2.3.4).

2.3.1 Initialization bias

We saw in Section 1.3.2.3 that the initialization bias is ruled by the time needed to achieve thermalization, i.e. τ_{mix}, Eq. 1.66. Computing τ_{mix} is however a very hard problem and it is usual to set an upper bound on τ_{mix}, Eq. 1.75, by using the relaxation time τ_{rel}, Eq. 1.72. The relaxation time gives information on the initialization bias when thermalization is already achieved, i.e. on the correlations between two equilibrated samples.

As the upper bound of Eq. 1.75 depends on the probability of the least likely state, it is common for it to be very large and little information can be extracted from it. Therefore, in practice, the initialization time, during which all states are discarded, is usually taken to be $20\tau_{\text{rel}}$ or $20\tau_{\text{exp}}$, with τ_{exp} the autocorrelation time of the slowest mode, Eq. 1.77, that is very close to τ_{rel}, Eq. 1.78. However, this criterion appears artificial. Another strategy consists in running simulations from two opposite initial configurations. Typically, around a phase transition, a good choice is a disordered configuration and an ordered one. If we define

$$\bar{d}(t) = \max_{i,j} ||\mathbf{T}^t(i,\cdot) - \mathbf{T}^t(j,\cdot)||_{TV}, \tag{2.43}$$

then

$$d(t) \leq \bar{d}(t) \leq 2d(t), \tag{2.44}$$

with $d(t) = \max_{x\in\Omega} ||T^t(x,\cdot) - \pi||_{TV}$ the distance defined in Eq. 1.65. When the simulations converge to the same configuration after a time t, one can take t as a good estimation of τ_{mix}, following Eq. 2.44.

The coupling of Markov chains formalizes this strategy. Chains are started from every possible initial condition. If two chains move to the same state during the same time step, they merge into a single chain. When all the chains have merged, i.e. coupled, all correlations with the initial configuration are lost. Thus, the time at the final coupling τ_{coup} is always larger than the mixing time τ_{mix}. Coupling methods became very popular with the *Coupling From The Past* method developed by Propp and Wilson [98]. If this method suppresses the problem of estimating correlation times, it is only applicable to specific systems. It indeed requires to simultaneously follow a large number of Markov chains and is sensitive to the chaotic properties of the Markov chain, known as damage spreading [99]. Such features make the coupling time τ_{coup} explode, whereas τ_{mix} remains small in comparison. In particular, the chaotic behavior was observed in dense hard-sphere systems and spin glasses at low temperatures [99].

2.3.2 Uncorrelated samples

For a given physical system, we want to assess by a Monte Carlo method the expectation value $\langle\theta\rangle$ of an observable θ. To do so, θ can be estimated by the estimator $\bar{\theta}$, the usual mean over a set of N sequential samples $\{\theta_j\}$,

$$\bar{\theta} = \frac{1}{N}\sum_{j=1}^{N}\theta_j. \tag{2.45}$$

The difference between θ and $\bar{\theta}$ is that the former is an ordinary number, whereas the latter is a random one, which fluctuates around θ and should come with an error $\mathcal{E}_{\bar{\theta}}$. According to the Bienaymé - Chebyshev inequality Eq. 2.4, the error can be estimated through the variance of $\bar{\theta}$. In practice, the variance, Eq. 2.46, of those fluctuations is not estimated through the repetition of the whole Monte Carlo simulations, but is rather estimated through the distribution of the individual samples θ_j.

$$\sigma_{\bar{\theta}}^2 = \mathcal{E}_{\bar{\theta}}^2 = \langle \bar{\theta}^2 \rangle - \langle \bar{\theta} \rangle^2. \tag{2.46}$$

Injecting the definition of $\bar{\theta}$,

$$\begin{aligned} \sigma_{\bar{\theta}}^2 &= \frac{1}{N} \left\langle \sum_{i=1}^{N} \theta_i \frac{1}{N} \sum_{j=1}^{N} \theta_j \right\rangle - \left\langle \frac{1}{N} \sum_{i=1}^{N} \theta_i \right\rangle \left\langle \frac{1}{N} \sum_{j=1}^{N} \theta_j \right\rangle \\ \sigma_{\bar{\theta}}^2 &= \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \theta_i \theta_j \rangle - \frac{1}{N^2} \sum_{i,j=1}^{N} \langle \theta_i \rangle \langle \theta_j \rangle . \end{aligned} \tag{2.47}$$

If the N samples are obtained through direct sampling, see Section 2.1 or, more generally, are independent, then $\langle \theta_i \theta_j \rangle = 0$ for $i \neq j$ and $\sigma_{\bar{\theta}}^2$ simply relates to the variance of the individual samples $\sigma_{\theta_j}^2 = \langle \theta_j^2 \rangle - \langle \theta_j \rangle^2$. If we assume equilibrium, the individual variances do not depend on time i and

$$\sigma_{\bar{\theta}}^2 = \sigma_{\theta_j}^2 / N. \tag{2.48}$$

Whatever the form of the distribution $P(\theta_i)$ is, the distribution of the mean value is Gaussian, by the central limit theorem, at least for uncorrelated data in the asymptotic regime of large N. It is noteworthy that $P(\theta_i)$ is often close to Gaussian because the θ_i are usually averaged over many degrees of freedom, e.g. θ is the magnetization of a system.

Under the assumption of a Gaussian distribution for the mean $\bar{\theta}$, 68% of all simulations would yield a mean value in the range $[\bar{\theta} - \sigma_{\bar{\theta}}, \bar{\theta} + \sigma_{\bar{\theta}}]$. This one-sigma squared error $\mathcal{E}_{\bar{\theta}}^2$ is the one usually being quoted along the mean value. The two-sigma interval groups 95.4% of the simulations and the three-sigma intervals 99.7%.

2.3.3 Correlated samples and autocorrelation times

In Markov-chain Monte Carlo, samples are correlated and $\langle \theta_i \theta_j \rangle \neq 0$ for $i \neq j$. The simpler relation Eq. 2.48 does not hold anymore. Separating the diagonal terms and the now non-null off-diagonal terms leads to

$$\sigma_{\bar{\theta}}^2 = \frac{1}{N^2} \sum_{i=1}^{N} \left(\langle \theta_i^2 \rangle - \langle \theta_i \rangle^2 \right) + \frac{1}{N^2} \sum_{i,j \neq 1}^{N} \left(\langle \theta_i \theta_j \rangle - \langle \theta_i \rangle \langle \theta_j \rangle \right) . \tag{2.49}$$

Assuming the samples have been collected once the Markov chain hit the stationary regime, averages are now invariant to time translation, see Section 1.1.3, and only the time difference between two samples θ_i and θ_j is relevant. Identifying $\sigma_{\theta_i}^2$, Eq. 2.49 becomes,

$$\sigma_{\bar{\theta}}^2 = \frac{\sigma_{\theta_i}^2}{N} + \frac{2}{N^2} \sum_{k=1}^{N-1} \left(\langle \theta_1 \theta_{k+1} \rangle - \langle \theta_1 \rangle \langle \theta_{k+1} \rangle \right) (N - k) . \tag{2.50}$$

As was done in Section 1.3.2.4, the integrated autocorrelation time $\tilde{\tau}_{\text{int}}(\theta)$ is introduced in order to have a clearer understanding of the impact of the correlations on the variance of the estimator, as

$$\sigma^2_{\bar{\theta}} = \frac{\sigma^2_{\theta_i}}{N} 2\tilde{\tau}_{\text{int}}(\theta), \tag{2.51}$$

with

$$\tilde{\tau}_{\text{int}}(\theta) = \frac{1}{2} + \sum_{k=1}^{N} C(k) \left(1 - \frac{k}{N}\right), \tag{2.52}$$

where $C(k)$ is the normalized autocorrelation function, Eq. 1.31.

When going to large time separations k, the autocorrelation function $C(k)$ decays as an exponential with the characteristic time $\tau_{\text{exp}}(\theta)$, Eq. 1.76 Section 1.3.2.4. This exponential decay allows us to approximate the proper integrated autocorrelation time $\tilde{\tau}_{\text{int}}(\theta)$ by the commonly used integrated autocorrelation time $\tau_{\text{int}}(\theta)$ [67], defined in Eq. 1.79, from the moment the length of the simulation $N \gg \tau_{\text{exp}}(\theta)$, which is actually a primary condition for producing a meaningful set of samples.

Finally, one can define an effective number of samples N_{eff}, which corresponds to the equivalent number of independent samples achieving the same variance as the N correlated ones,

$$N_{\text{eff}} = \frac{N}{2\tau_{\text{int}}(\theta)}. \tag{2.53}$$

The smaller the integrated autocorrelation time, the better the statistics, as, at only every $2\tau_{\text{int}}(\theta)$, the samples are approximately uncorrelated.

In Section 1.3.2.3, the relationship between the eigenvalues of the transfer matrix $\mathbf{T}$ of a Markov chain and its exponential autocorrelation time was presented in Eq. 1.78. In practice, it is however impossible to directly compute the second highest eigenvalue of $\mathbf{T}$. Also, getting dynamic quantities as correlation times from a numerical simulations come at a much higher cost than sampling some static quantities.

A running estimator of $\tau_{\text{int}}(\theta)$ can be computed by using the estimator of the autocorrelation function $\hat{C}(k)$,

$$\hat{C}(k) = \frac{\frac{1}{N-k}\sum_{i=1}^{N-k} \theta_i \theta_{i+k} - \left(\frac{1}{N}\sum_{i=1}^{N} \theta_i\right)^2}{\frac{1}{N}\sum_{i=1}^{N} \theta_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} \theta_i\right)^2}. \tag{2.54}$$

As k increases, the number of samples available to compute the mean values of $\theta_i \theta_{i+k}$ decreases. This leads to the divergence of the variance of $\hat{C}(k)$ with increasing k. Therefore it is necessary to define an upper limit k_{max} in order to retrieve a correct estimation for $\tau_{\text{int}}(\theta)$ and, as a consequence, a correct error estimate,

$$\hat{\tau}_{\text{int}}(\theta)(k_{\max}) = \frac{1}{2} + \sum_{k=1}^{k_{\max}} \hat{C}(k). \tag{2.55}$$

The value of $k_{\max}$ is a compromise between the systematic error of truncating the sum in Eq. 2.55 and the statistical error brought by the diverging variance of $\hat{C}(k)$. A current procedure is to cut the summation once $k_{\max} \geq 6\hat{\tau}_{\text{int}}(\theta)(k_{\max})$ [66]. This leads to the *a priori* error estimate,

$$\mathcal{E}_{\tau_{\text{int}}(\theta)} = \tau_{\text{int}}(\theta)\sqrt{\frac{2(2k_{\max}+1)}{N}} \approx \tau_{\text{int}}(\theta)\sqrt{\frac{12}{N_{\text{eff}}}}. \tag{2.56}$$

Instead of truncating the sum after $k_{\max}$, it is possible to approximate the tail end of $C(k)$ by a single exponential with $\tau_{\text{exp}}(\theta)$ as the characteristic time. It leads to an expression that may be used to estimate $\tau_{\text{int}}(\theta)$ or $\tau_{\text{exp}}(\theta)$,

$$\hat{\tau}_{\text{int}}(\theta)(k_{\max}) = \tau_{\text{int}}(\theta) - c\exp(k_{\max}/\tau_{\text{exp}}(\theta)), \tag{2.57}$$

where c is a constant.

Finally, the error can also be estimated through tools like binning analysis or the more elaborate jackknife [97].

When dealing with thermalized, but correlated samples extracted from a Markov-chain Monte Carlo simulation, the error is ruled by the integrated autocorrelation time τ_{int}, which has the exponential autocorrelation time τ_{exp} as its upper bound. The time scale τ_{exp} gives information on the time needed to decorrelate from a thermalized sample.

In practice, in order to have a squared error of $\epsilon^2\sigma^2_{\theta_i}$, one should run a simulation based on the Markov-chain Monte Carlo method over a time span of $\tau_{\text{mix}} + 2\tau_{\text{int}}\epsilon^2$. Unfortunately, autocorrelation times are linked to the autocorrelation length in the system and as a system approaches a critical point and its phase transition, the large spatial correlations arise along with temporal correlations. At the critical point itself, the temporal correlations and their characteristic time scales diverge.

It is noteworthy that it is possible to produce uncorrelated samples by Markov-chain Monte Carlo method, by using a *perfect sampling* method, which rely on the coupling of Markov chains, like the *Coupling From The Past* method [98].

2.3.4 Scaling of autocorrelation times around a phase transition

It is possible to give to a Markov-chain Monte Carlo process a dynamic interpretation using the Master equation, see Section 1.2.3. This approach is useful to link the statistical errors, as discussed in Section 2.3.3, with the dynamic correlation functions of the appropriate stochastic model. One can then understand what actually the slowest variables are. For instance, conservation laws may cause long temporal correlations in the system. Thus, the statistical inefficiency near second-order phase transitions,

i.e. continuous phase transitions, reflects critical slowing down [1]. Near first- order transitions, i.e. discontinuous phase transitions, rather than critical slowing down, metastability and hysteresis may arise [100].

2.3.4.1 First-order phase transitions

First-order transitions are quite common in nature [100]. For instance, the ordinary melting of water is a first-order transition. First-order transition are characterized by discontinuities in the order parameter, like the jump experienced by the magnetization M, when the transition between the paramagnetic and ferromagnetic phases occurs at the transition point T_c. This discontinuity is caused by the possibility of two, or more, phases coexisting. For instance, for the melting transition in the NVT ensemble, the two coexisting phases are the ordered solid and the disordered liquid.

Unlike for second-order phase transitions, see Section 2.3.4.2, the correlation length in the coexisting pure phases is finite and the autocorrelation time does not diverge in the pure phases. However, for a system of dimension D, configurations exhibiting a mixed phase made from coexisting phases contain interfaces between the phases, which bring an extra energy σL^{D-1}, with σ the interface tension and L^{D-1} the surface of the interfaces. In addition to the energy of the two coexisting phases, that is equal to the energy of the pure phases, the interface tension leads to an extra interfacial Boltzmann weight of $\exp(-2\sigma L^{D-1})$, the factor 2 accounting for the topological constraint of periodic boundary. Due to this factor, for large system sizes, it may take a long time before the system flips from one phase to the other. The autocorrelation time associated to this flipping mode scales as the inverse of the suppression factor,

$$\tau_{\exp} \propto \exp(2\sigma L^{D-1}). \tag{2.58}$$

As the autocorrelation time now grows exponentially with the system size, this behaviour is called supercritical slowing down. As Eq. 2.58 find its roots in the probability distribution itself and not in the locality or diffusivity of the proposed moves, cluster algorithms are of no help. A new approach is needed, as with the use of multicanonical simulations [101]. On the other hand, studying the decay of metastable states through nucleation and growth by the dynamic interpretation of Monte Carlo is an interesting problem and the basis for a broad range of kinetic Monte Carlo studies of stochastic processes.

2.3.4.2 Second-order phase transitions

The characteristic feature of second-order transition is a divergent spatial correlation length ξ, associated with divergent temporal correlations, at the transition point β_c, [1, 102, 103]. The divergence of the time scale, leading to the critical slowing down, is determined in part by the nature of the conservation laws. This divergence reflects the scale invariance, at the core of renormalization group treatments and of the universality concept. At the critical temperature β_c, thermodynamic fluctuations are expected to

be on all length scales, implying power-law singularities in thermodynamics functions, such as the correlation length,

$$\xi = \xi_0^{\pm} |1 - T/T_c|^{-\nu} + \text{correction terms}, \tag{2.59}$$

where $\xi_0^{\pm}$ is the critical amplitude on the high- and low-temperature side of the transition respectively and ν is a critical exponent. In spin systems, similar singularities of the specific heat, magnetization and susceptibility define the critical exponents α, β and γ respectively. The critical exponents are related to each other through scaling relations and only two of them are considered independent.

The autocorrelation time τ_{exp} is linked to the correlation length ξ, as, qualitatively, the information brought by locally updating a state by a step of the Markov chain has to propagate over the correlation volume before obtaining a new statistically independent state. So, close to β_c, we have,

$$\tau_{\text{exp}} \propto \xi^z \propto |1 - t/T_c|^{\nu z}. \tag{2.60}$$

Here z is the dynamical critical exponent and is independent from the other critical exponents. It depends on the algorithm used. For a local Markov-chain Monte Carlo procedure, as the Metropolis, Section 2.2.2, or heat-bath algorithm, Section 2.2.4, the information is expected to travel diffusively, from nearest neighbor to nearest neighbor, through a random walk in the configuration space, which requires on average ξ^2 steps to propagate over a distance proportional to ξ. We then expects z to be approximately 2 and numerical estimates for the Ising model yield $z \approx 2.15$ in two dimensions and $z \approx 2.05$ in three dimensions, see Tab. 2.2. This scaling of the autocorrelation time with the diverging correlation length leads to the critical slowing-down phenomenon and the dramatic reduction of the level of accuracy attainable in practice, when close to a critical point. In any numerical simulation, the system is, however, of finite size and the correlation length does not actually diverge, as all other quantities, but equals the linear size L of the system, as illustrated in Fig. 2.7 for a two-dimensional Ising system. Therefore thermodynamic scaling laws are replaced by finite-size scaling, [10, 11], as $|1 - t/T_c| \propto \xi^{-1/\nu} \to L^{-1/\nu}$ and for autocorrelation time,

$$\tau_{\text{exp}} \propto L^z, \tag{2.61}$$

and the autocorrelation times scale with the system size. Far away from β_c, however, ξ is finite and the autocorrelation times do not depend on the system size. It is noteworthy that, near a critical point, the common assumption that τ_{exp} and τ_{int} are of the same order of magnitude may not be true. Replacing time by space, τ_{exp} is the analogue of a correlation length, whereas τ_{int} is the analogue of a susceptibility.

Eq. 2.61 is the reason that developing non-local algorithms is important. Instead of flipping a spin or moving a single sphere, a step of the Markov chain should consist of moving a cluster of spins or spheres in a consistent way, to escape from the local and diffusive behavior leading to $z \sim 2$. Different cluster algorithms were then developed, see Section 2.4, and, even if z varies from one algorithm to the other, in most cases, z

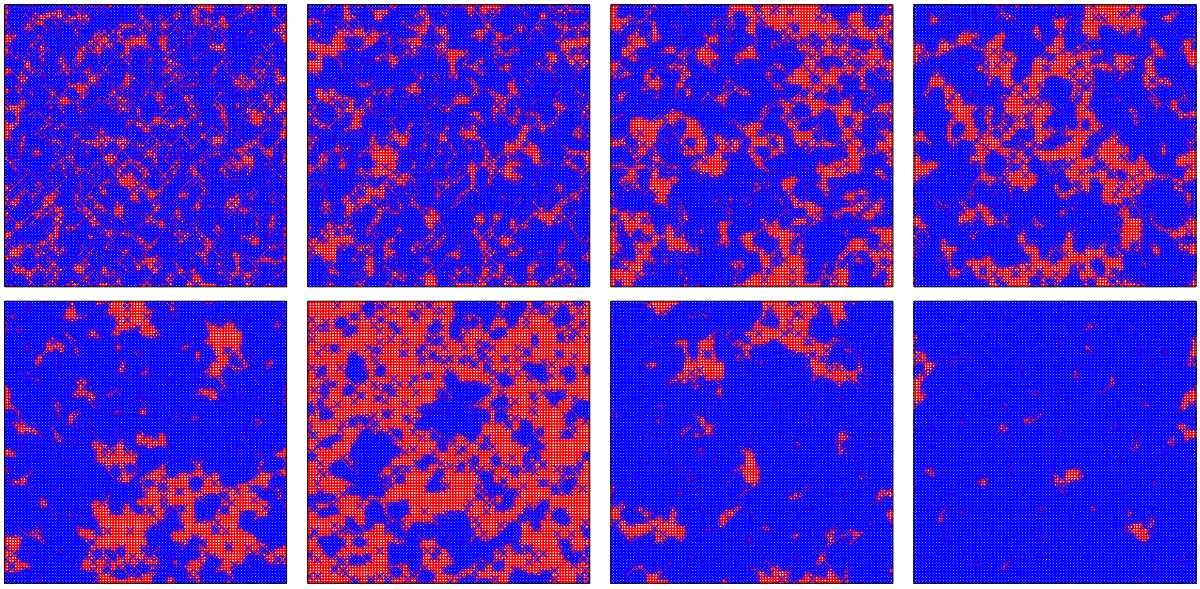


Figure 2.7: Configurations of a system of 100×100 Ising spins on a $2D$ lattice, from high temperatures (upper left) to just below the critical temperature. Spins up are coded in red, Spins down in blue. From **top left** to **bottom right**: $\beta/\beta_c = 0.5$, 0.7, 0.85, 0.9, 0.95, 0.98, $\gtrsim 1.0$, 1.05. The critical region is characterized by large spatial correlations, as the system undergoes a second-order phase transition [104].

is smaller than 1, see Tab. 2.2. Applied to the $2D$ Ising model, some algorithms even achieve $z \approx 0$, leading to a logarithmic divergence and a fast mixing. However, one should always check that the proportionality constant in Eq. 2.61 is not too large.

2.4 Cluster algorithms

Any Markov-chain Monte Carlo algorithm brings correlations between samples that hinder the statistical accuracy of the estimators of observables. As these correlations are characterized by the integrated and exponential autocorrelation times, it is then a crucial issue to develop algorithms that keep autocorrelation times as short as possible. It is in the late 1980s that the critical slowing down was not seen anymore as inescapable, as it is in experiments. Cluster and non-local algorithms, as presented in the next Section 2.4, managed to get past this limit by proposing new rules for the construction of the Markov chain.

2.4.1 Fortuin-Kasteleyn transformation

To beat critical slowing down in spin systems on lattice, the idea was to propose flips of correlated-spin clusters instead of single-spin flips. C. M. Fortuin and P. W. Kasteleyn [110, 111] started this paradigm shift by mapping a ferromagnetic Potts model onto a corresponding percolation model, where states are produced by throwing down

Algorithm	Ising 2D			Ising 3D		
	z	Obs.	Ref.	z	Obs.	Ref.
Metropolis	2.1667(5)	(M, τ_{exp})	[105, 106]	2.055(10)	(M, τ_{rel})	[107]
Swendsen-Wang	0.35(1)	(E, τ_{exp})	[19]	0.75(1)	(E, τ_{exp})	[19]
	0.27(2)	(E, τ_{int})	[108]	0.50(3)	(E, τ_{int})	[108]
	0.20(2)	(χ, τ_{int})	[108]	0.50(3)	(χ, τ_{int})	[108]
	$\mathcal{O}(\log L)$	(M, τ_{exp})	[109]			
Wolff	0.26(2)	(E, τ_{int})	[108]	0.28(2)	(E, τ_{int})	[108]
	0.13(2)	(χ, τ_{int})	[108]	0.14(2)	(χ, τ_{int})	[108]

Table 2.2: Dynamical critical exponents z, as defined in Eq. 2.61, for Ising spin systems in two dimensions and three dimensions for the local Metropolis algorithm and the cluster algorithms of Swendsen-Wang and Wolff. For each value of z, the studied time is indicated. M stands for magnetization, E energy and χ susceptibility. Different observables may yield quite different values for z. In 3D, the Wolff algorithm is clearly more efficient than the Swendsen-Wang algorithm, whereas, in 2D, they appear equivalent.

particles or bonds in an uncorrelated manner. Thanks to the Fortuin-Kasteleyn transformation, the standard Monte Carlo dynamics, which exhibits slow critical relaxation, is now mapped into one where such behavior does not exist.

Following the notations of [10], the partition function of the q-state Potts model is,

$$Z = \sum_{\sigma_i} e^{-\beta J \sum_{i,j} (\delta_{\sigma_i \sigma_j} - 1)}. \tag{2.62}$$

The transformation replaces each pair of interacting Potts spins on the lattice by a bond on an equivalent lattice with probability,

$$p = 1 - e^{-\beta J \delta_{\sigma_i \sigma_j}}, \tag{2.63}$$

which is only non-zero if the spins σ_i and σ_j are in the same state. This leads to the construction of a lattice with bonds connecting some spin sites. One then produces a set of clusters of spins of same state, with different sizes and shapes. The partition function Z can then be rewritten as,

$$Z = \sum_{\text{bonds}} p^b (1-p)^{(N_b - b)} q^{N_c}, \tag{2.64}$$

where b is the numbers of bonds, N_b is the total numbers of possible bonds and N_c the total number of clusters.

This equivalence between the Potts and percolation problems was exploited to devise new Monte Carlo methods for spin systems as showed in the next Section 2.4.2 and Section 2.4.3.

2.4.2 Swendsen-Wang algorithm

R. H. Swendsen and J. S. Wang [19] were the first to use the Fortuin-Kasteleyn transformation for developing Monte Carlo algorithms. As in the Markov-chain Monte Carlo method, one starts with a valid initial spin configuration. Then, one constructs the bonds through the lattice with the probability p of Eq. 2.63. Each cluster is then randomly assigned a new spin value. The bonds are erased and the procedure is repeated. Such a procedure allows for non-local moves on the energy.

At each move, all the spins are flipped. Detailed balance is fulfilled, since the spin flip is involutary, as, applied twice, it is the identity. It is easy to see that the spin flip is rejection-free, as the cluster construction can be interpreted as drawing from an a priori probabilities $\mathcal{A}(C_1 \to C_2)$, with C_1 the initial configuration and C_2 the corresponding configuration after the clusters are flipped. Following Eq. 2.64, the construction ensures that $\mathcal{A}(C_1 \to C_2) = \pi(C_2)$, leading to a rejection-free scheme, see Section 2.2.3.

As the probability p to draw a bond between two sites depends on the temperature, the resulting cluster distribution varies strongly with the temperature. At very high temperatures, p is small and the clusters will be quite small. On the contrary, at low temperature, all neighboring spins with the same state will end up in the same cluster and the system will tend to oscillate back and forth between similar structures. So in both cases, this new procedure does not improve by far a standard Monte Carlo approach. But, near a critical point, the distribution of clusters will be quite diverse and the different configurations produced by cluster flipping will end up substantially different from one another. Therefore critical slowing down is reduced, as the reduction in the scaling of the characteristic time with the linear size of the lattice shows: from the diffusive value of $z \sim 2$ for Metropolis single-site spin flip to a value of about 0 ($\mathcal{O}(\log)$) in two dimensions and ~ 0.5 in three dimensions, see Tab. 2.2.

2.4.3 Wolff algorithm

One of the issues encountered by the Swendsen-Wang algorithm is that all clusters are treated in the same way, independently of their sizes. As a consequence, a significant effort is spent addressing small clusters, whereas they do not contribute to the critical slowing down. U. Wolff [20] proposed an alternative detailed-balance algorithm, based also on the Fortuin-Kasteleyn theorem, that grows and flips single clusters sequentially. Eliminating partially the constraint of dealing with small clusters, the performance of the Wolff algorithm is generally better than the one presented by the Swendsen-Wang method, see Tab. 2.2, and is simpler to implement.

The Wolff algorithm starts from a valid spin configuration. A single spin is then randomly chosen and bonds are then drawn to its nearest-neighbors of same state with

the probability p of Eq. 2.63. Then, the same bond drawing procedure is applied to any sites that are now connected with the initially chosen spin and so on. The process stops when no new bonds are formed and the entire cluster is then flipped.

Apart from being simpler, the Wolff single-cluster algorithm shows better results than the Swendsen-Wang multiple-cluster algorithm, especially in $3D$, see Tab. 2.2. On average, the single-cluster method leads to flips of larger clusters. Some care is nonetheless necessary with the definition of the unit of time, since the number of flipped spins varies from cluster to cluster. This number depends crucially on temperature since the average cluster size automatically adapts to the correlation length. It is then usual to define one sweep as the time unit. A sweep corresponds to $V/\langle C\rangle$ single cluster steps, with $\langle C\rangle$ denoting the average cluster size, so that, after one sweep, V spins are flipped on average. This definition allows for direct comparison with the Swendsen-Wang algorithm or the Metropolis algorithm.

For classical continuous spin models, as will be considered in Section 4.3.1, the spin-flip operation is generalized into a reflection with respect to a hyperplane orthogonal to a random vector $\boldsymbol{r}$ [20],

$$R(\boldsymbol{r})\sigma_x = \sigma_x - 2(\sigma_x \cdot \boldsymbol{r})\boldsymbol{r}. \tag{2.65}$$

The reflection is involutory, as $R \circ R = \mathbb{1}$. The cluster flipping procedure is now replaced by first choosing a random vector $\boldsymbol{r}$ and a random lattice site x as the starting point of the cluster, that is flipped according to Eq. 2.65. Then all the nearest-neighbors y are checked, i.e. the bond $x - y$ is activated with probability,

$$p(\sigma_x, \sigma_y) = 1 - \exp(\min(0, \beta\sigma_x \cdot (1 - R(\boldsymbol{r})\sigma_y))) \tag{2.66}$$

If the bond is activated, σ_y is flipped. Iteratively, all bonds leading to yet unflipped neighbors of newly flipped spins are investigated, until the process stops. For continuous XY spins, the Wolff algorithm achieves a scaling $z \lesssim 0.1$, [20, 112]. Simulations implementing the Wolff algorithm clarified that the two-dimensional XY model undergoes a Kosterlitz-Thouless transition [113, 114]. However, if the Wolff algorithm can be implemented for the three-dimensional XY spin glass model, it loses all its efficiency [115, 116].

2.4.4 Cluster algorithms for spin glasses

Instead of having the same coupling constant J for all pairs of spins, spin glass systems possess a set of couplings $\{J_{ij}\}$ that is drawn from a random distribution. Spin glasses describe, for instance, materials where ferromagnetic impurities are coupled over intermediate distances, leading to couplings that may be more or less repulsive or attractive. Spin glass systems are disordered systems and they exhibit a large number of metastable states, that are bound to hinder the speed of local simulations. Hence there is a need for fast algorithms.

Both cluster algorithms can also be applied to spin glass models [117–119]. The antiferromagnetic interactions can be handled by emulating frozen spins as quenched, non-interacting impurities. However, there was no improvement in performance to be noted, in part because of the strong frustration at play. Another reason for this lack of efficiency lies in the fact that the speed of cluster algorithms in ferromagnetic Ising systems comes from their ability to switch quickly between the two ground states, thanks to large updates in magnetization at each step. However, such large strides do not facilitate moves between the large number of metastable states in spin glass systems.

In spite of these issues, Houdayer (2001) [120] introduced a successful rejection-free cluster algorithm, tailored to work for two-dimensional Ising spin glasses. The speedup was of several orders of magnitude but, to be efficient, the method asks for temperature not close to zero or a percolation threshold of more than 50%, which is not reached in the three-dimensional Ising spin glass for example. For the low-temperature regime, only small clusters are produced and, for systems with a percolation threshold under 50%, clusters span the entire system. The percolation threshold can be artificially increased, by diluting the lattice for instance [121], but such a method is highly dependent on the problem to be studied.

Recently, building on [120], [122] devised a method to produce isoenergetic cluster moves, where clusters of small sizes or clusters spanning the entire system are avoided, by restricting the cluster moves to temperatures where cluster percolation is hampered by the interplay of frustration and temperature. This method achieves an important speedup in thermalization, that appears to improve with the system size. The thermalization was judged based on the agreement of the internal energy per spin to the internal energy computed from the link overlap for Gaussian couplings, [123]. However, the cluster moves of [120] and [122], as they are isoenergetic, still need to implement single-spin flips, in order to be irreducible. To speed this process, both methods use parallel tempering [124], also known as replica exchange, which consists of running copies of the system at different temperatures and of exchanging them based on the Metropolis filter, Eq. 2.33. The goal is to make configurations at high temperatures available to the simulations at low temperatures and vice versa.

2.5 Non-local algorithms

Fluids or continuous particle systems, notably in two dimensions where the particular physics of Kosterlitz and Thouless transitions emerges, see Section 4.1, encounter also critical slowing-down phenomena. Moreover, at high densities, simpler cage phenomena may occur where moving a single particle is nearly impossible, as any traditional Metropolis move gets rejected. Also, away from any phase transition, a fluid in the NVT ensemble will present slow relaxation of long wavelength density fluctuations, because of the conservation of the density. This issue of long-time tails is known as the hydrodynamic slowing down [1].

A new class of methods for particle systems was created in order to generalize the drastic reduction of critical slowing down of cluster algorithms on spin systems, see Tab. 2.2. As for spins, the idea is to go beyond single particle move, but to produce instead persistent and consistent moves of several particles. Such moves do not happen on a local scale anymore and are not connected to the physical process of single-particle motion. Clusters are built from a symmetric operation that leaves the Hamiltonian invariant if applied to the entire configuration or applied twice. The challenge here is to find such an operation in particle systems.

2.5.1 Geometric cluster algorithm

The first attempt to generalize cluster algorithms to hard core systems was proposed in [125]. The challenge arises by the vanishing of symmetries that spin lattices offer in comparison to the continuum space of a fluid. For example, a spin flip is its own inverse. Therefore, flipping a spin twice will bring it back to its initial state, whereas updating a particle by the same displacement $\boldsymbol{\delta r}$ twice will not achieve the same result. In a rejection-free scheme, where the transition probability comes down to the a priori probability, symmetric a priori probabilities are necessary to fulfill detailed balance.

The geometric cluster algorithm [125], also called the pivot algorithm, builds such symmetric a priori probabilities. Considering a system of N hard spheres in configuration C_1, the algorithm is as follows,

- A pivot, i.e. a point, is chosen randomly in the box.
- A new configuration $\tilde{C}_1$ is obtained by carrying out a point reflection for all particles in C_1 according to that pivot. This operation is its own inverse.
- By superimposing C_1 and $\tilde{C}_1$, clusters of particles are identified as groups of particles that overlap.
- One can then exchange the positions of the particles of a given cluster between C_1 and $\tilde{C}_1$, without risking any overlap. Such exchange is done with a probability $1/2$ for each cluster. It leads to two new configurations $C_1 \to C_1'$ and $\tilde{C}_1 \to \tilde{C}_1'$.
- The new configuration C_1' is stored and serves as a new starting point of the scheme. $\tilde{C}_1'$ is discarded.

The superposition yielding N clusters offer 2^N possible new configurations. As the process is symmetric, detailed balance is satisfied, while moves are non-local. Moreover, irreducibility follows from the irreducibility of the local algorithm. This method was applied with success to binary [126] and polydisperse mixtures [127], where it avoids the jamming problem, in which a very large fraction of all trial moves is rejected because of overlaps. The algorithm was generalized to soft potentials in [128] by constructing the clusters according to probabilities that take into account the energy differences between the reflected and old positions, as is done by cluster algorithms in spin systems, Eq. 2.63.

However, an important limitation of this algorithm is that, for large densities that exceed the percolation threshold of the combined configurations, all particles will be in the same cluster, leading to a simple rotation of the system. The algorithm can then no longer be seen as mixing. Geometric cluster algorithms are then powerful methods, but only for some systems. This situation is similar to what happens with cluster algorithms for spin systems, see Section 2.4.4. For instance, the geometric cluster algorithm did not improve the notoriously difficult simulations of monodisperse hard disks, which was solved by an irreversible factorized Metropolis algorithm, the event-chain algorithm [13], as will be presented in Chapter 3.

2.5.2 Jaster algorithm

Another non-local algorithm for hard core particles was proposed in [129]. It displaces chains of particles, but is not rejection-free, because of the following avalanche issue. All particles of a chain are moved in the same direction with equal displacements. The chain is first started by the random choice of a particle i. This particle i is moved by a random vector $\boldsymbol{\delta r}$. If the move brings no overlap, it is accepted. If it brings an overlap with only one particle j, said particle is moved as was i, along $\boldsymbol{\delta r}$, and the same procedure is applied. If two or more particles are involved in the overlap, moving the collided particles along $\boldsymbol{\delta r}$ would create an avalanche, where the number of moving spheres may increase at each collision. In order to satisfy detailed balance, the move is rejected and the procedure is restarted from the beginning with the choice of a random sphere. It is possible to set the maximal number $n_{\max}$ of particles that are involved in the chains.

It was expected from the construction of the algorithm that short chains will be produced in disordered domains, while, in ordered domains where the positional correlation length is large, long chains will be built. The stability of the ordered regions is responsible for the large autocorrelation times in two-dimensional hard disk systems. This non-local algorithm brings a speedup of a factor of around 5 for the autocorrelation times, but it still scales the same way with the system size, as the traditional local Metropolis does [129]. However, in [129], it was shown that for systems of 4096 particles, chains with more than 16 particles are a very rare occurrence; as in a local Metropolis algorithm, a large proportion of moves is rejected.

Another version of this algorithm was also proposed in [129], in order to create longer chains. The particle i is now moved along a random direction $\boldsymbol{v}$ to the nearest collision point with one of the remaining particles. This particle is then moved in the same direction $\boldsymbol{v}$ and so on until the total number of particles moved in the chain adds up to a random number drawn beforehand between 1 and $n_{\max}$. At this point, the last particle is placed with uniform probability between its initial position and the collision point. In order to satisfy the detailed-balance condition, the new configuration is accepted with the probability $\min(1, \ell_{\text{end}}/\ell_{\text{start}})$, ℓ_{end} being the length between the initial position and collision point of the last particle and ℓ_{start} being the length between the final position of the first particle and the collision point in the $-\boldsymbol{v}$ direction. Such a

rule ensures indeed that a chain along v and its *return* chain along $-v$ are of the same probability. In [24], the event-chain algorithm works in a similar way, but, instead of setting the maximal number of particles involved in the chain, it sets the global displacement ℓ travelled along the chain. This is the key element in order to implement a global-balance scheme.

It appears clearly here that enforcing detailed balance comes with rejected moves and local, diffusive dynamics. In spite of moving clusters of particles, Jaster's algorithm [129] is unable to upgrade from a local and diffusive dynamics and to reduce the critical slowing down.

Conclusion

In this Chapter, we discussed the efficiency of the Monte Carlo method to simulate physical systems. In particular, the Markov-chain Monte Carlo method allows us to solve the issues brought about by the impossibility to compute the partition function and by the high dimensionality of the systems under study. In physical terms, the Markov-chain Monte Carlo method relies on inventing a stochastic time evolution for a given system. However, this time evolution does not need to correspond to any real physical dynamics. On the contrary, the dynamics is to be treated like any other numerical algorithm, on the basis of its computational efficiency. The need for fast algorithms becomes all the more dire, as, around critical points, critical slowing down drastically reduces the statistical accuracy for local algorithms.

The cluster algorithms devised for spin systems exhibit large speedups in comparison to local algorithms, like the Metropolis or heat-bath algorithm. Unfortunately, these speedups do not generalize to spin glasses or antiferromagnetic spin systems. In particle systems, different methods were tried to emulate cluster algorithms, in order to produce non-local and consistent moves of a cluster of particles. However, fulfilling the detailed-balance condition leads to powerful methods, but which are specialized to some systems, or to the drastic reduction of the probability for any large cluster to form, because of the avalanche process.

In the next Chapter, we introduce a new paradigm for Markov-chain Monte Carlo. This paradigm relies on the factorized Metropolis filter and the lifting concept. It produces an irreversible Markov chain based on non-local moves, which are rejection-free. Its generality of use allows one to implement it for particle systems as well as for spin systems.

CHAPTER 3

Irreversible factorized Metropolis algorithm

The Markov-chain Monte Carlo method in statistical physics has progressed far from the original local-move detailed-balance Metropolis algorithm [17]. In spin systems, the Fortuin-Kasteleyn transformation allows to set for Metropolis-Hastings algorithms a priori probabilities that equal the probability weights of the target configurations. These cluster algorithms have proven powerful in thermalizing certain spin systems and reduce the critical slowing-down phenomenon to nearly zero in lattice models and ferromagnetic continuous models [19, 20, 113, 114]. Although cluster algorithms can be implemented for a wide range of models, they are efficient only in a few of them. An important example is their difficulty to thermalize anti-ferromagnetic spin systems or spin glasses. In particle systems, different methods try to emulate cluster algorithms by producing non-local moves [125, 128] and proved to be efficient for polydisperse mixtures [126, 127]. However, as for spin systems, the fast mixing observed for those systems does not easily generalize. At the time, it seems like a single method could not be highly optimized and completely general at the same time.

In 2011, Bernard et al. [13] solved the issue of the famously difficult problem of the melting of hard disks. They relied on a new algorithm, the *event-chain* algorithm [24], that produces also non-local moves, but not detailed-balance ones. By only fulfilling the weaker condition of global-balance while avoiding the avalanche problem, clusters of particles can now be of large size, in comparison to [129]. Global-balance algorithms that break detailed-balance converge in general faster, as they introduce persistence between subsequent moves and reduce the diffusive nature of the Markov chain on small and intermediate time scales. Notable examples are guided random walks [21], hybrid Monte Carlo [22, 25] and overrelaxation [23]. The Markov chain lifting framework [26–28] unifies these concepts by augmenting the physical configuration space with auxiliary variables that determine the a priori probabilities.

Lifted Markov chains have already been applied to spin models [130, 131] but not

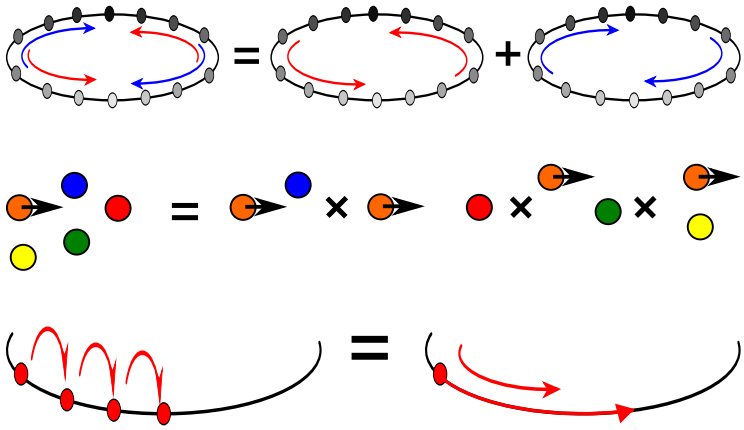


Figure 3.1: Irreversible Markov chains allow for constructing a rejection-free and maximal global-balance scheme, that is at the core of the irreversible factorized Metropolis algorithm. This new paradigm is coordinated around three principles: the lifting framework where the configuration space is transformed into two replicas thanks to an extra variable (**Top**), the factorized filter for generalization to multidimensional potentials (**Middle**) and finally the implementation with infinitesimal steps that leads to a continuum of valid configurations (**Bottom**).

to continuum systems. Drawing on the line of [24, 25], we generalized the event-chain scheme of hard-sphere systems to general systems by introducing a factorized version of the Metropolis filter that allows to treat a multidimensional potential as a product of independent unidimensional potential. Using the lifting framework, it is then possible to implement irreversible Markov chains in a maximal global balance scheme. The algorithm is rejection-free, even for moves between configurations of different energies. This new approach does not rely on any symmetries of the Hamiltonian to build symmetric a priori probabilities, like cluster algorithms do, and is not hindered by forcing the algorithm to obey the detailed-balance condition, which may lead to a high rate of rejection. Finally, this new irreversible factorized Metropolis paradigm is easily implemented in a large range of systems, where it achieves important speedups in comparison to the local Metropolis algorithm, as will be discussed in Chapter 4. This method gains its generality from replacing the general driver of most Monte Carlo algorithms, the local Metropolis filter.

In this chapter, we will demonstrate how the new factorized Metropolis filter leads to a maximal global-balance and rejection-free algorithm for general potentials. As illustrated in Fig. 3.1, the rejection-free irreversible factorized Metropolis paradigm relies on three key concepts: First, moves are proposed in a persistent way; Second, multidimensional potential are reduced to a collection of independent unidimensional ones; Third, moves are infinitesimal. The irreversible Markov chains are described by using the lifting concept, see Section 3.1, that is easily generalized thanks to the factorized filter to particle systems or any multidimensional potentials, see Section 3.2. Then, we will discuss how the use of infinitesimal steps allows for a continuum of

valid configurations, while producing a maximal balance of probability flows, see Section 3.3. We will review how moves are computed using an event-driven scheme [132], see Section 3.3.2, and ensemble averages of observables are retrieved from uniformly sampling configurations from the continuum of valid ones, see Section 3.3.3. Finally, the generalization leading to infinite chain will be discussed in Section 3.4.

3.1 The Lifting Framework

Section 2.4 and Section 2.5 discussed how an important acceleration can be gained by producing non local moves, in particularly around the transition point. As explained in Section 2.4.2, cluster algorithms can be interpreted as a detailed-balance Markov-chain Monte Carlo method with special a priori probabilities that enforce global moves. As the goal is to derive a general scheme that does not depend on the symmetry of the Hamiltonian, persistency is introduced by directly setting the a priori probabilities such that the direction of moves will be persistent and thus produce consistent sequential moves.

Adding persistency will however break the Markovian nature of the process. In order to recover it, it is necessary to add an additional variable to describe the chain, as explained in Section 1.2. Following this idea, a general framework was developed in [26–28] for speeding up Markov-chain Monte Carlo algorithms by lifting. A lifting of a Markov chain $\mathbf{T}$ on the configuration space Ω is a larger chain $\mathbf{T_L}$ obtained by extending each physical state x of Ω by an additional variable σ that fixes the a priori probability $\mathcal{A}((x,\sigma) \to (x',\sigma))$. It is then simple to upgrade a reversible Markov chain, obeying detailed-balance, into an irreversible Markov chain, obeying only the weaker condition of global balance. As observed by [26] and explored further by [27], lifting can lead to a substantially shorter mixing time of the Markov chain, Eq. 1.66, and, as a consequence, faster algorithms.

3.1.1 Lifting for the unidimensional random walk

As illustrated in Fig. 3.2, we consider a unidimensional random walk with periodic boundary on the unidimensional lattice $\Omega = \{1, 2, \ldots, n\}$, with stationary probability π. The random walk is described by the detailed-balance matrix $\mathbf{T}$,

$$T_{kl} = \begin{cases} \frac{1}{2} p^{\text{acc}}(k \to k-1) & \text{for } l = k-1 \\ \frac{1}{2} p^{\text{acc}}(k \to k+1) & \text{for } l = k+1 \\ 1 - \frac{1}{2}\left(p^{\text{acc}}(k \to k-1) + p^{\text{acc}}(k \to k+1)\right) & \text{for } l = k \end{cases} \tag{3.1}$$

with,

$$p^{\text{acc}}(k \to l) = \min\left(1, \frac{\pi(l)}{\pi(k)}\right) \tag{3.2}$$

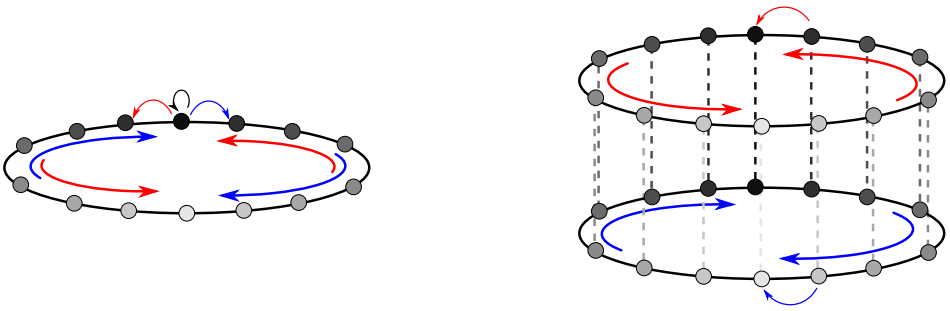


Figure 3.2: **Left:** Random walk on a loop with a probability on each location to move right, left or to stay at the current location, Eq. 3.1. **Right:** Extension of the configuration space Ω by addition of a lifting variable $\sigma \in \{\leftarrow, \rightarrow\}$. Two replicas of the initial physical configuration space are then created, $(\Omega, \leftarrow)$ where the chain only moves to the left, $(\Omega, \rightarrow)$ where the chain only moves to the right. If the move is rejected physically in $(\Omega, \rightarrow)$, a lifting move (dashed lines) takes place towards $(\Omega, \leftarrow)$, and conversely. During this lifting move, the physical configuration does not change, only the lifting variable is updated.

T realizes a Metropolis scheme, where, at each time step, it is proposed with probability $\frac{1}{2}$ to move from k to $k+1$ or to $k-1$. The acceptance probability of the move is the Metropolis filter, Eq. 2.33.

The space Ω is now extended into two replicas $(\Omega, \leftarrow)$ and $(\Omega, \rightarrow)$ thanks to the additional lifting variable $\sigma \in \{\leftarrow, \rightarrow\}$. In the lifted scheme, σ sets the a priori probabilities so that,

$$\text{Physical moves} \begin{cases} \mathcal{A}^L\left(\overrightarrow{k} \to \overrightarrow{k-1}\right)) = 0 & \mathcal{A}^L\left(\overleftarrow{k} \to \overleftarrow{k-1}\right) = p^{\text{acc}}(k \to k-1) \\ \mathcal{A}^L\left(\overrightarrow{k} \to \overrightarrow{k+1}\right) = p^{\text{acc}}(k \to k+1) & \mathcal{A}^L\left(\overleftarrow{k} \to \overleftarrow{k+1}\right) = 0 \end{cases}$$
$$\text{Lifting moves} \begin{cases} \mathcal{A}^L\left(\overrightarrow{k} \to \overleftarrow{k}\right) = 1 - p^{\text{acc}}(k \to k+1) & \mathcal{A}^L\left(\overleftarrow{k} \to \overrightarrow{k}\right) = 1 - p^{\text{acc}}(k \to k-1). \end{cases} \tag{3.3}$$

The acceptance probability of a move in the lifted scheme is 1. The scheme is rejection-free, as the rejected moves have been transformed into lifting moves. The detailed-balance condition is violated, as in $(\Omega, \leftarrow)$ (resp. $(\Omega, \rightarrow)$), only the physical moves $k \to k-1$ (resp. $k \to k+1$) are proposed, apart from the lifting moves $\overleftarrow{k} \to \overrightarrow{k}$ (resp. $\overrightarrow{k} \to \overleftarrow{k}$) between replicas. The global balance condition, Eq. 1.46, is still obeyed though, making the scheme correct,

$$\phi(\overrightarrow{k-1} \to \overrightarrow{k}) + \phi(\overleftarrow{k} \to \overrightarrow{k}) = \frac{\pi(k)}{2} = \phi(\overrightarrow{k} \to \overrightarrow{k+1}) + \phi(\overrightarrow{k} \to \overleftarrow{k}), \tag{3.4}$$

with the flows IN,

$$\phi(\overrightarrow{k-1} \to \overrightarrow{k}) + \phi(\overleftarrow{k} \to \overrightarrow{k}) = \frac{\pi(k-1)}{2} p^{\text{acc}}(k-1 \to k) + \frac{\pi(k)}{2}(1 - p^{\text{acc}}(k \to k-1)) = \frac{\pi(k)}{2}, \tag{3.5}$$

and the flows OUT,

$$\phi(\overrightarrow{k} \to \overrightarrow{k+1}) + \phi(\overrightarrow{k} \to \overleftarrow{k}) = \frac{\pi(k)}{2} p^{\text{acc}}(k \to k+1) + \frac{\pi(k)}{2}(1 - p^{\text{acc}}(k \to k+1)) = \frac{\pi(k)}{2} \tag{3.6}$$

If π is the uniform distribution, then τ_{mix} is reduced from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$ [26]. Indeed, if π is the uniform distribution $\mathbf{ran}(1,n)$, then no lifting move will take place and the random walk will cover the whole set Ω by moving step by step without ever turning back.

The square-root reduction of the mixing time is the best possible, while only using lifting [27]. If π is the stationary distribution of the initial scheme, it is still the stationary distribution of the lifted scheme, but needs to be renormalized (here a factor $1/2$, as they are two possible lifted states for the same physical state). From now on, as in Markov-chain Monte Carlo we are sampling a distribution proportional to the physical one, we will not mention this normalization and refer to both the stationary distributions of the reversible and lifted scheme as π.

3.1.2 General case

3.1.2.1 Two replicas

We saw how powerful the lifting of a configuration space was in a simple unidimensional random walk, achieving a square root reduction of the mixing time. We now consider a Markov chain $\mathbf{T}$ devised on a space state Ω, which obeys the detailed-balance condition and converges to the equilibrium distribution π. Each state is duplicated into two replicas, marked by the lifting variable $\sigma = \pm$. New transition probabilities are introduced within each of the replicas, so that $\mathbf{T} = \mathbf{T}^{(+)} + \mathbf{T}^{(-)}$, with every $T_{ij}^{(\pm)}$ positive. The chains $\mathbf{T}^{(\pm)}$ break detailed balance but satisfy the skew detailed balance condition,

$$\pi_i T_{ij}^{(+)} = \pi_j T_{ji}^{(-)} \forall i,j. \tag{3.7}$$

The lifting moves between the same physical states of each replica are represented in the total transfer matrix $\mathbf{T_L}$ by the interreplica matrices $\mathbf{\Lambda}^{(\pm,\mp)}$, that are positive and diagonal,

$$\mathbf{T_L} = \begin{pmatrix} \mathbf{T}^{+} & \mathbf{\Lambda}^{(+,-)} \\ \mathbf{\Lambda}^{(-,+)} & \mathbf{T}^{-} \end{pmatrix}. \tag{3.8}$$

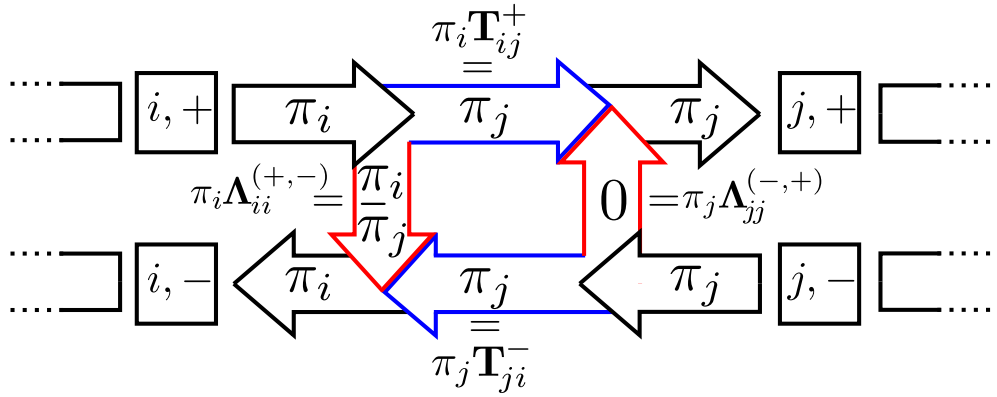


Figure 3.3: Probability flows for a two replicas maximal global-balance scheme. The probability flows are represented by the arrows, black codes for the total flows IN and OUT of each configurations, blue codes for physical moves and red for the lifting moves. Here we consider $\pi_i < \pi_j$ and set $T^+_{ij} = T_{ij} = \pi_j/\pi_i$ and $T^-_{ji} = T_{ji} = 1$. T^+_{ji} and T^-_{ij} are 0 to achieve a maximal global-balance scheme. The interreplica term $\mathbf{\Lambda}^{(+,-)}_{ii}$ compensates the compressibility introduced in the probability flows by the transition probabilities T^+_{ij} and T^-_{ji}.

The global-balance condition, and by it the convergence towards π, is enforced by the choice of the interreplica matrices $\mathbf{\Lambda}^{(\pm,\mp)}$. They compensate any compressibility of the probability flows that the lifting of the configuration space may have introduced with the matrices $\mathbf{T}^{\pm}$. In general, multiple choices of $\mathbf{\Lambda}^{(\pm,\mp)}$ are possible, as the global-balance condition on $\mathbf{T_L}$ only fixes the difference $\mathbf{\Lambda}^{(+,-)} - \mathbf{\Lambda}^{(-,+)}$,

$$\begin{cases} \sum_j \pi_j T^{\pm}_{ji} + \pi_i \mathbf{\Lambda}^{(\mp,\pm)}_{ii} &= \pi_i \\ \sum_j \pi_i T^{\pm}_{ij} + \pi_i \mathbf{\Lambda}^{(\pm,\mp)}_{ii} &= \pi_i. \end{cases} \tag{3.9}$$

Using the skew detailed-balance condition, Eq. 3.7, it leads to

$$\begin{cases} \mathbf{\Lambda}^{(\pm,\mp)}_{ii} - \mathbf{\Lambda}^{(\mp,\pm)}_{ii} = \sum_j \left(T^{\pm}_{ij} - T^{\mp}_{ij} \right), \\ T^{\pm}_{ii} = 1 - \sum_{j \neq i} T^{\pm}_{ij} - \mathbf{\Lambda}^{(\pm,\mp)}_{ii} \end{cases} \tag{3.10}$$

To produce only the necessary flows, the minimum $\mathbf{\Lambda}^{(\pm,\mp)}$ satisfying the difference in Eq. 3.10 are obtained with

$$\mathbf{\Lambda}^{(\pm,\mp)}_{ii} = \max \left(0, \sum_j \left(T^{\pm}_{ij} - T^{\mp}_{ij} \right) \right), \tag{3.11}$$

as $\mathbf{\Lambda}^{(\pm,\mp)} \geq 0$. Thus the interreplica flow goes only in one direction. For the same reason, we also set $T^{\pm}_{ii} = 0$, so that,

$$\mathbf{\Lambda}^{(\pm,\mp)}_{ii} = 1 - \sum_{j \neq i} T^{\pm}_{ij} \tag{3.12}$$

In this scheme, all the physical rejected moves are transformed into the interreplica flows, as was observed for the unidimensional random walk in Section 3.1.1.

Finally, the maximal global-balance scheme is achieved, if it is also enforced that, for a given pair of states i, j, probability only flows in one direction in each replica,

$$\begin{cases} T^+_{ij} = T_{ij} & T^+_{ji} = 0 \iff \text{in } (\Omega, +), \text{ only } i \to j \text{ moves} \\ T^-_{ij} = 0 & T^-_{ji} = T_{ji} \iff \text{in } (\Omega, -), \text{ only } j \to i \text{ moves} \end{cases} . \tag{3.13}$$

As shown in Fig. 3.3, the probability flows on the extended configuration space are then restricted to the necessary amount to obey global balance, which is the case for the example of the unidimensional random walk of Section 3.1.1.

Originally, in [26], the scheme counts one more degree of freedom, the flip rate θ. In Ω^+, after proposing the move $(i, +) \to (i+1, +)$, the chain is in $(i+1, +)$, if the move is accepted according to the probability p^{acc} adapted from the reversible scheme, or in $(i, -)$ otherwise. The extra step is to propose to change the lifting variable with the probability θ. Thus, after a step, the chain can be in,

$$(i, +) \to \begin{cases} (i, +) \text{ with probability } (1 - p^{\text{acc}})\theta \\ (i, -) \text{ with probability } (1 - p^{\text{acc}})(1 - \theta) \\ (i+1, +) \text{ with probability } p^{\text{acc}}(1 - \theta) \\ (i+1, -) \text{ with probability } p^{\text{acc}}\theta. \end{cases} \tag{3.14}$$

As we are interested in irreversible scheme, θ is set to 0.

By duplicating the configuration space by an auxiliary variable σ, sequential moves are persistent, until the value of σ changes. Detailed balance is broken but the global-balance condition is fulfilled, by switching periodically from one replica to the other. In a maximal global-balance scheme, if the chain is on Ω^+, it will keep moving in the direction set by $+$, until there is a lifting move (rejection in the physical space), where it will then switch to the replica Ω^- by updating the lifting variable but keeping the same physical configuration. Therefore, there is no formal rejection on the extended configuration space, as rejection are now replaced by a lifting move.

This 2-replica scheme was used by [130] for the mean-field Ising model,

$$E(\{s_k\}) = -\frac{J}{2N} \sum_{k,k'} s_k s'_k = -\frac{J}{2N} M(\{s_k\})^2, \tag{3.15}$$

with the magnetization M,

$$M(\{s_k\}) = \sum_k s_k. \tag{3.16}$$

The replicas correspond respectively to the increasing of the magnetization $M \to M - 2$ by flipping down spins and to the decreasing of the magnetization $M \to M + 2$ by flipping up spins. As the energy is proportional to M^2 and configurations are completely

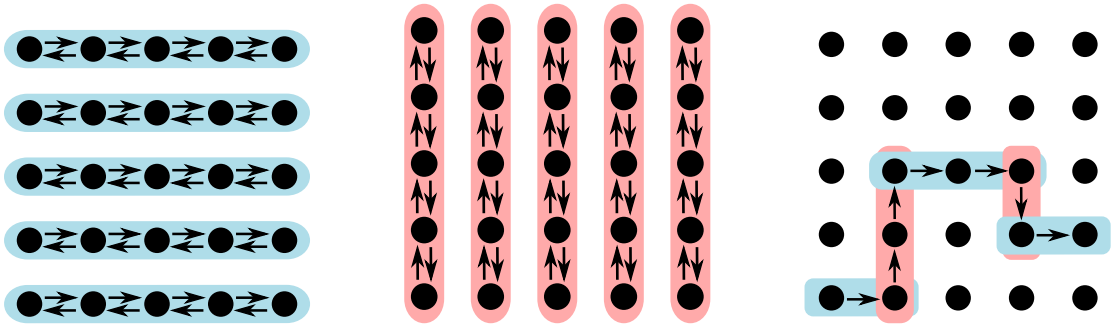


Figure 3.4: For a Markov chain on a bidimensional lattice, the configuration space can be partitioned into the set of horizontal lines $\{L_{xj}\}$ (**Left**) and into the set of vertical lines $\{L_{yj}\}$ (**Center**). On $\{L_{xj}\}$, states are updated along the $\pm x$ directions and on $\{L_{yj}\}$, states are updated along the $\pm y$ directions. Finally, the collection of the partitions $\{L_{xj}\}$ and $\{L_{yj}\}$ allows to reach any state y from a state x, as there is always a path $x_1, \ldots, x_n$ connecting the two states, such that any two sequential intermediate states x_i and x_{i+1} are in a common line L (**Right**).

determined by M, the situation then is similar to the $1D$ random walk in Section 3.1.1 and the lifting scheme brings a square-root reduction of the exponential autocorrelation time [130] and of the integrated autocorrelation times of the magnetization [131].

The situation is different in the two-dimensional Ising model with nearest-neighbor interactions, as simulated in Fig. 2.7,

$$E(\{s_k\}) = -J \sum_{\langle k,k' \rangle} s_k s'_k. \tag{3.17}$$

The magnetization does not characterize fully a configuration anymore. No clear reduction of the critical slowing down appears, but a constant speedup of 5 for the integrated autocorrelation times of the energy and magnetization is observed [131].

3.1.2.2 $d \times 2$ replicas

The lifting framework for 1×2 replicas extends naturally for $d \times 2$ replicas for d-dimensional distributions. It is necessary to be able to partition Ω into ordered lines, corresponding to the direction along which a state is updated, see Fig. 3.4. There should be one partition per dimension, leading to a set of partitions $\{i\}$, each corresponding to the linearly ordered lines $\{L_{ij}\}$, that correspond to the direction set by i. Furthermore, Ω must be connected in the sense that for each x, y in Ω, there is a path $x_0 = x, x_1, \ldots, x_l = y$ such that each pair x_i, x_{i+1} are in a common line. For the euclidean space $\mathbb{R}^d$, a natural choice is to set the partition i as the set of the lines parallel to the i-th coordinate. Such a partition is presented in Fig. 3.4 for the case of a bidimensional lattice, but the natural choice may not be the simplest, as illustrated in Fig. 3.5, where the bidimensional grid is transformed into a unidimensional grid.

The scheme proposed in [26] on the extended space $\Omega \times \{-1, +1\}^d$ is:

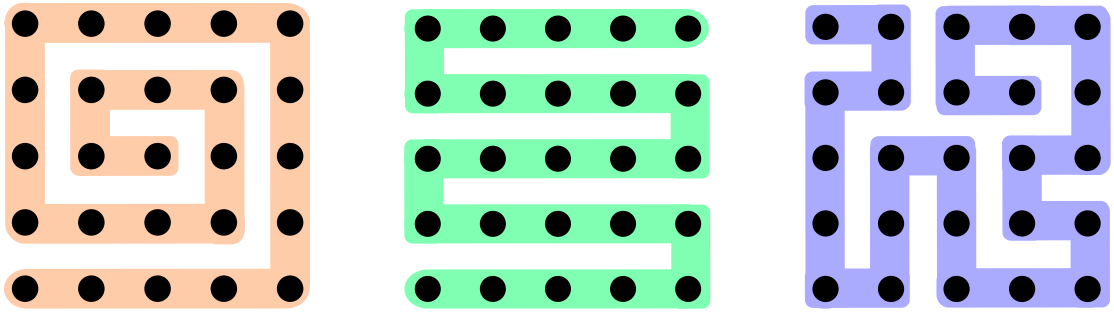


Figure 3.5: For the bidimensional lattice, there are better collections of partitions than the one presented in Fig. 3.4. With these partitions, the problem becomes again unidimensional and there is no need to resample the direction i. However, creating such collection is only possible when one has some knowledge on the stationary distribution π. If some of the states of the grid has a probability weight equal to zero, then those collections will not be irreducible.

- Choose one of the directions i with a probability w_i.
- Update the physical state x according to $z(i) = z \in \{+, -\}$. The proposed move is $x \to x + z\delta_i$. If the move is accepted, the chain moves to the state $(x + z\delta_i,\ z(i) = z)$. Otherwise, it moves to $(x,\ z(i) = -z)$.
- Propose a flip of probability θ_i. If a flip is sampled, the new state is $(x + z\delta_i,\ z(i) = -z)$ or $(x,\ z(i) = z)$, according to the acceptance or rejection of the physical move.
- Update the set $\{z(i)\}$ to the new value of its i-th component $z(i)$.

Partitioning Ω allows one to reduce the multidimensional problem into unidimensional ones and to treat each one of them as is done in the 2-replicas case. In practice, the simulation of the chain does not require the lines to be constructed explicitly, see Chapter 4, only that it is possible to move from the current point on a line to its successor or predecessor.

The set $\{z(i)\}$ acts as the memory of the process. It allows for persistent moves, but in a delayed way, as one needs to wait for a resampling of the same direction i to see its effect. We will be concerned with complex systems of particles, where one can easily fall into a dynamic trap. For instance, in a hard-disk systems, disks often form cages that trap other disks. Suppressing these cages is achieved by moving in a consistent way the prisoner disk and the jailer disks. As explained in Publication 1 [30], to solve this issue, we do not resample the direction at each moves, unlike [26]. We define instead a lifting variable $\sigma = (i, z)$, where i is the direction of the move and $z = \pm$ its sense. There is no set $\{z(i)\}$. When a rejection occurs, z is updated to $-z$. To change the direction i, the lifting variable is resampled after a fixed number of steps, in analogy with molecular dynamics [43], where velocities may be resampled

every now and then. This does not introduce any bias and leads to ballistic moves of several particles. Thus, there is no need to keep and update a set $\{z(\sigma_i)\}$, which can reach an considerable size when considering a large number N of particles in several dimensions and where z will take dN values, see Section 3.2.3. Finally, implementing a resampling of the lifting variable at each step is simply impossible, while using infinitesimal steps, which are crucial in order to avoid the multirejection problem, see Section 3.3.

The difficulty with implementing lifting for multidimensional distributions is that appropriate sets of lines of direction, i.e. appropriate lifting variables in our scheme, must be found, preferably ones which will be effective in eliminating diffusive behavior. For instance, for a discrete grid, it is easy to define lines, but if the distribution is non zero only on a connected subset of the grid, these lines might not be effective in eliminating diffusive behavior. The factorized Metropolis filter, see Section 3.2, will be the key to the construction of the collection of partitions for any multidimensional potential, notably in continuous space.

3.1.3 Lifting for two particles

The implementation to a unidimensional two-particles system with periodic boundary follows easily. Two spheres interacting through a hardcore potential or a soft potential ($\propto r^{-n}$) is indeed only a unidimensional problem, the degree of freedom being the interdistance r between the two spheres. But even for a simple problem as two soft spheres, there are two ways to extend the configuration space by lifting:

- fixed-sphere scheme: Following [26], the lifting variable $\sigma = (i, z)$ sets the sphere i that moves and the direction $z = \pm x$. The direction z changes when there is a rejection. The moving sphere is resampled after a given number of steps, as shown on Fig. 3.6.

- The new fixed-direction scheme we propose: The lifting variable $\sigma = z$ sets only the sphere that moves, $z = i$, and the direction is fixed, $+$. When a rejection occurs, the other sphere j moves next and $z = j$, see Fig. 3.7. The direction of move does not need to be resampled for the algorithm to be irreducible on the configuration space Ω.

As illustrated in Fig. 3.4 and Fig. 3.5, both schemes are correct and actually equivalent from the point of view of the sampling of the interdistance r. On the space Ω of the coordinates of the spheres, however, the fixed-direction scheme is a maximal global-balance one, whereas the fixed-sphere scheme is not as efficient as forth-and-back moves are produced, as illustrated in Fig. 3.8. The maximal global-balance scheme is generalized with ease to two-particles systems in higher dimensions. The lifting variable just needs to set the direction x_i of the moves, but the moves are still proposed

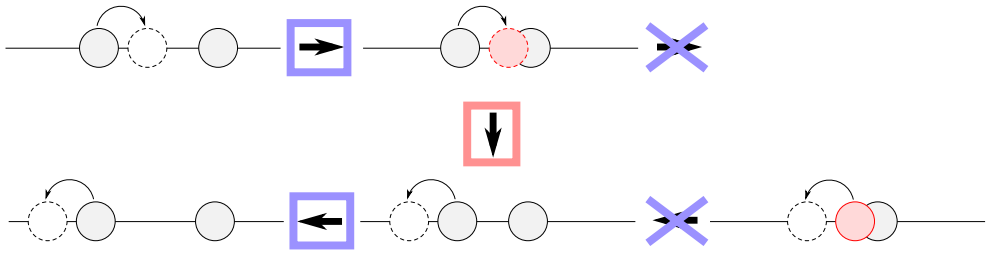


Figure 3.6: Lifting for two hard spheres: The lifting variable fixes the direction and the moving sphere. The moving sphere is updated by physical moves (blue box arrow) until it is rejected (blue cross arrow) and a lifting move (red box arrow) happens. The direction changes. In this scheme, every configuration has one predecessor and one successor, making the scheme correct. Configurations of hard spheres are indeed equiprobable, from the moment they are valid, i.e. there is no overlap.

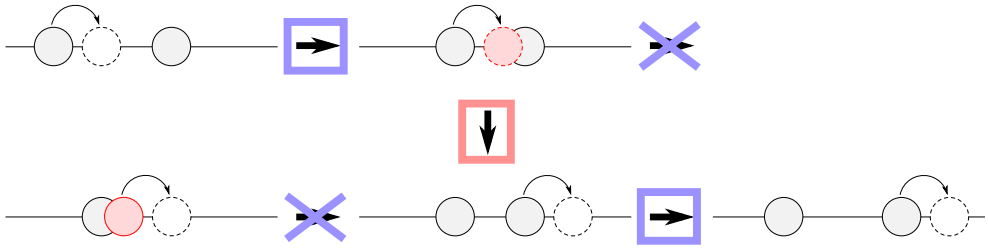


Figure 3.7: Lifting for two hard spheres: The lifting variable sets only which sphere to move. The direction is fixed. When a rejection happens, the next sphere becomes the moving sphere. Every configuration has also one predecessor and one successor.

only in the $+x_i$ sense, see Fig. 3.9. The direction is resampled uniformly over $\{+x_i\}$ after a fixed time.

For two soft spheres interacting by the potential $E(r)$, maximal global-balance condition is verified, as probability flows are produced only in one direction and Eq. 3.9 is obeyed: We consider a configuration $\{\boldsymbol{r_1}, \boldsymbol{r_2}, \sigma = \{1, +x\}\}$, where the lifting variable σ fixes the moves so that the position of the sphere 1 is updated along $+x$, the balance of flows is,

$$\text{Flows OUT} \begin{cases} \text{Physical flow: } \pi(\{\boldsymbol{r_1}, \boldsymbol{r_2}\}) \min\left(1, \dfrac{\pi(\{\boldsymbol{r_1} + \delta\boldsymbol{x}, \boldsymbol{r_2}\})}{\pi(\{\boldsymbol{r_1}, \boldsymbol{r_2}\})}\right) \\ \text{Lifting flow: } \pi(\{\boldsymbol{r_1}, \boldsymbol{r_2}\}) \left(1 - \min\left(1, \dfrac{\pi(\{\boldsymbol{r_1} + \delta\boldsymbol{x_1}, \boldsymbol{r_2}\})}{\pi(\{\boldsymbol{r_1}, \boldsymbol{r_2}\})}\right)\right) \\ \hline \text{Total flow OUT: } \pi(\{\boldsymbol{r_1}, \boldsymbol{r_2}\}) \end{cases} \tag{3.18}$$

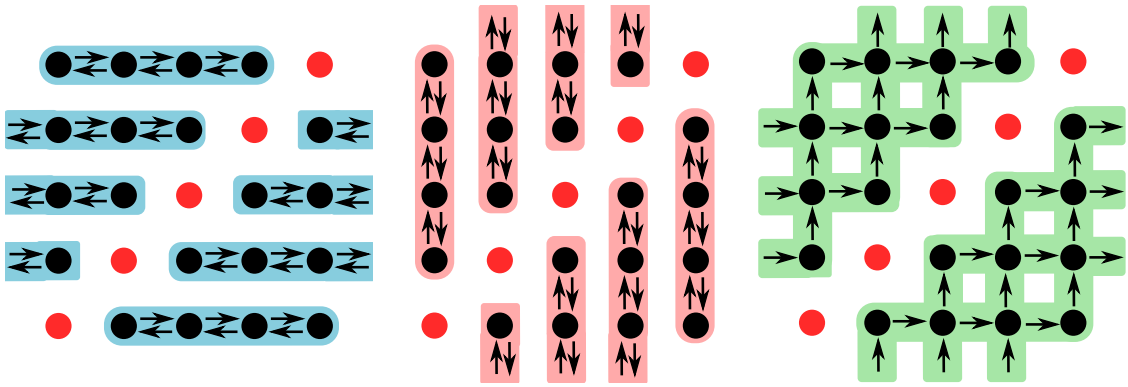


Figure 3.8: Configuration space for two hard spheres on an unidimensional discrete lattice with periodic boundary. The position x_1 of the sphere 1 is represented by the x axis on the bidimensional lattice, the position x_2 of the sphere 2 by the y axis. The two hard spheres cannot be on the same site, i.e. $x_1 = x_2$, the forbidden configurations are in red. The two schemes, as illustrated in Fig. 3.6 and Fig. 3.7, have a different collection of independent lines along which the configuration is updated. **Left and Center**: For the first scheme, there are two partitions of lines (blue: lines which updates sphere 1; red: lines which updates sphere 2) . A sphere is moved back and forth on a line until the lifting variable is resampled. The scheme is not maximal global-balanced. **Right:** The second scheme needs only one partition (green), so there is no need to resample. The scheme is maximal global-balanced.

$$\text{Flows IN} \begin{cases} \text{Physical flow: } \pi(\{\boldsymbol{r_1} - \boldsymbol{\delta x}, \boldsymbol{r_2}\}) \min\left(1, \dfrac{\pi(\{\boldsymbol{r_1}, \boldsymbol{r_2}\})}{\pi(\{\boldsymbol{r_1} - \boldsymbol{\delta x}, \boldsymbol{r_2}\})}\right) \\ \text{Lifting flow: } \pi(\{\boldsymbol{r_1}, \boldsymbol{r_2}\}) \left(1 - \min\left(1, \dfrac{\pi(\{\boldsymbol{r_1}, \boldsymbol{r_2} + \boldsymbol{\delta x}\})}{\pi(\{\boldsymbol{r_1}, \boldsymbol{r_2}\})}\right)\right) \\ \hline \text{Total flow IN: } \pi(\{\boldsymbol{r_1}, \boldsymbol{r_2}\}) \end{cases} \tag{3.19}$$

As the potential E only depends on the interdistance $r = |\boldsymbol{r_1} - \boldsymbol{r_2}|$, the energy of the configurations $(\{\boldsymbol{r_1}, \boldsymbol{r_2} + \delta x\})$ and $(\{\boldsymbol{r_1} - \boldsymbol{\delta x}, \boldsymbol{r_2}\})$ are the same, making the Boltzmann weight π the same.

The generalization to N spheres is however more tricky, as finding a collection of partition for the space Ω for a maximal global-balance scheme is not as straightforward anymore. The avalanche problem, discussed in Section 2.5.2, is a serious issue for hard spheres, as several overlaps leads to a conflict on which one of the collided spheres to move next. For soft spheres, the problem is even more complicated, as, when a rejection occurs because of a too big global energy increase, choosing randomly another sphere than the one being moved is not correct. The probability flows will indeed not be the same, as the lifting flow corresponds to the rejection flow by all the spheres and compensates in this way, see Fig. 3.3. It is possible to solve these problems by

giving up on the maximal global-balance scheme and by producing backwards moves, but in the next Section 3.2, I present how the factorized Metropolis filter allows for the definition of lifting variables corresponding to independent lines along which the states are updated in a maximal global-balance fashion.

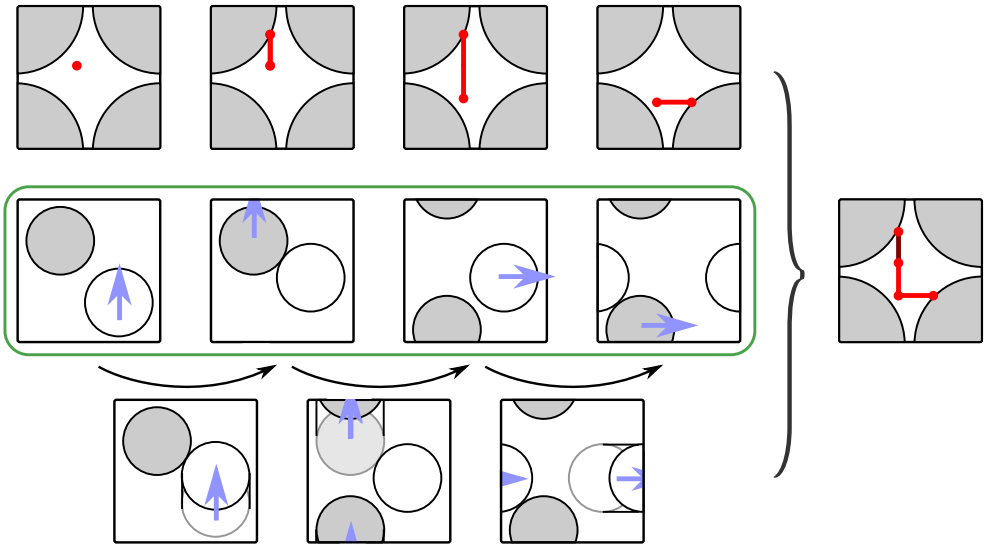


Figure 3.9: Lifting for two spheres in two dimensions. **Top row:** The volume of valid configurations is represented in white on the space of coordinates of the interdistance vector $\boldsymbol{r}$, the sampled states are represented in red. **Middle row:** The lifting variable sets the moving sphere and the direction $+x$ or $+y$. When the physical move is rejected, the collided sphere moves in turn in the same direction. After a given amount of steps, the direction is resampled uniformly between $+x$ and $+y$. **Bottom row:** The scheme here is implemented using an event-driven approach, making every visited configuration valid, see Section 3.3.

3.2 The Factorized Metropolis filter

An important problem encountered in previous methods, see Section 2.5.2, while trying to implement persistency in Monte Carlo method consists in the arising of the avalanche issue in hard spheres or the difficulty to find a general rule to simply implement the global-balance conditions in soft-sphere systems for instance. One solution to correct the scheme is to implement a detailed-balance condition, where a global move is accepted or not given its total energy variation. The development of the lifting framework for upgrading reversible schemes to nonreversible ones, Section 3.1, does not allow one to construct a general maximal global-balance filter. It indeed adapts the acceptance probabilities from the reversible chains. Therefore, the solution lies in developing a reversible filter, that will lead to a maximal global-balance scheme, once lifted. Moreover, the filter should be as general as the Metropolis filter, Eq. 2.33, is for detailed-balance schemes.

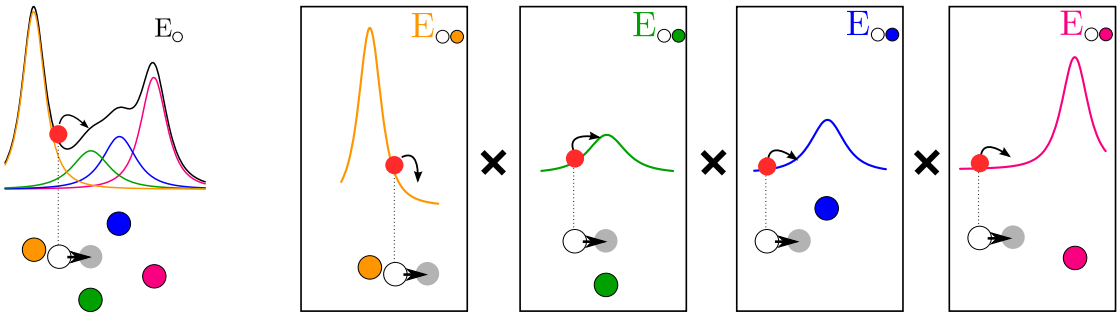


Figure 3.10: Instead of accepting or rejecting a move according to the global change in energy (**Left**), the factorized filter is written as an independent products of the different energy factors (**Right**: here the energy factors are the pairwise interactions). Sampling such filter is easily made by looking at the acceptance or rejection of every independent energy factor. Doing so allows to get a direct information on which factor is rejecting the move or not.

In this section, we discuss how the general factorized Metropolis filter, Publication 1 [30], once it is combined with the lifting approach, leads to maximal global-balance schemes.

3.2.1 Definition of the Factorized Metropolis filter

The factorized Metropolis filter separates a multidimensional potential into independent factors, reducing it to a product of unidimensional potentials, see Fig. 3.10. The factors will serve as the backbone for the construction of the partition of Ω by lifting variables, as explained in Section 3.1.2.2.

Two potential factors are said to be independent, if the probabilities for their own increase to reject a move are independent. As the probability of a physical state of energy E is the Boltzmann weight $\exp(-\beta E)$, up to the normalizing constant of the partition function, Eq. 1.7, two potential factors ΔE_1 and ΔE_2 are then independent if the global potential change ΔE between two states ζ and ζ' can be written as their sum,

$$\begin{aligned} E(\zeta') - E(\zeta) = \Delta E &= \Delta E_1 + \Delta E_2 \\ \exp(-\beta \Delta E) &= \exp(-\beta \Delta E_1)\exp(-\beta \Delta E_2). \end{aligned} \tag{3.20}$$

The factorized Metropolis filter then articulates around these factors. The acceptance probability to accept the move $\zeta \to \zeta'$ is

$$\begin{aligned} p^{\text{acc}}_{\text{Fact}}(\zeta \to \zeta') &= \min(1, \exp(-\beta \Delta E_1)) \text{ x } \min(1, \exp(-\beta \Delta E_2)) \\ &= p^{\text{acc}}_{\text{Metro}}(\Delta E_1) p^{\text{acc}}_{\text{Metro}}(\Delta E_2), \end{aligned} \tag{3.21}$$

making both potential factors independent in their acceptance or rejection, as defined in Eq. 1.24, whereas it is not the case in the standard Metropolis filter,

$$p^{\text{acc}}_{\text{Metro}}(\zeta \to \zeta') = \min(1, \exp(-\beta(\Delta E_1 + \Delta E_2))). \tag{3.22}$$

As the factorized Metropolis filter can be interpreted as a product of Metropolis acceptance probabilities, it exhibits the same symmetry as the Metropolis filter, Eq. 2.33. Therefore, when the a priori probabilities are symmetric, the factorized Metropolis filter obeys the detailed-balance condition,

$$\begin{aligned}
\pi(\zeta)\mathcal{A}(\zeta \to \zeta')p^{\text{acc}}_{\text{Fact}}(\zeta \to \zeta') &= \pi(\zeta')\mathcal{A}(\zeta' \to \zeta)p^{\text{acc}}_{\text{Fact}}(\zeta' \to \zeta) \\
\pi(\zeta)p^{\text{acc}}_{\text{Metro}}(\Delta E_1)p^{\text{acc}}_{\text{Metro}}(\Delta E_2) &= \pi(\zeta')p^{\text{acc}}_{\text{Metro}}(-\Delta E_1)p^{\text{acc}}_{\text{Metro}}(-\Delta E_2) \\
\pi(\zeta)\exp(-\beta\Delta E_1)\exp(-\beta\Delta E_2) &= \pi(\zeta'),
\end{aligned} \tag{3.23}$$

so the correct stationary distribution π is retrieved, as $\pi(\zeta')/\pi(\zeta) = \exp(-\beta\,(\Delta E_1 + \Delta E_2))$.

More generally, a multidimensional potential E defined on a set of variables $\{x_i\}$ can be factorized into the potential factors corresponding to its partial derivatives,

$$dE(\zeta) = \sum_i \frac{\partial E}{\partial x_i} dx_i \tag{3.24}$$

and

$$p^{\text{acc}}_{\text{Fact}}(\zeta \to \zeta') = \prod_i \min\left(1, \exp\left(-\beta \int_\zeta^{\zeta'} \frac{\partial E}{\partial x_i} dx_i\right)\right). \tag{3.25}$$

The factorization along each coordinates is always possible but it may not be the most physical one and may fail to be the most efficient. The problem is similar to the choice of a collection of partitions for Ω in the lifting framework, see Section 3.1.2.2. For instance for a pairwise potential, as commonly encountered in a system of N spheres or spins, see Fig. 3.10, a factorization can be the decomposition of the potential E on its pairwise components E_{ij}, as,

$$\begin{aligned}
E(\zeta) &= \sum_{\langle i,j \rangle} E_{ij}(\zeta) \\
\Delta E(\zeta \to \zeta') &= \sum_{\langle i,j \rangle} \Delta E_{ij}.
\end{aligned} \tag{3.26}$$

The pair energy E_{ij} can also be decomposed on each coordinate axis. However, as the a priori probabilities in the lifting framework will only be non zero for moves according to a given axis set by the lifting variable, there is no need to actually decompose the pairwise factors into each one of their coordinates subfactors. Thus, if we consider a system of N particles described respectively by one vector $\boldsymbol{r_i}$ and interacting through a pairwise potential $E(\{\boldsymbol{r_i}\}) = \sum_{\langle i,j \rangle} E_{ij}(\boldsymbol{r_{ij}})$, with $\boldsymbol{r_{ij}} = \boldsymbol{r_j} - \boldsymbol{r_i}$, the factorized filter simply writes itself as,

$$\begin{aligned}
p^{\text{acc}}_{\text{Fact}}(\boldsymbol{r_i} \to \boldsymbol{r_i'}) &= \prod_j p^{\text{acc}}_{\text{Met}}(\boldsymbol{r_{ij}} \to \boldsymbol{r_{ij}'}) \\
&= \prod_j \min\left(1, \exp\left[-\beta\Delta E_{ij}(\boldsymbol{r_{ij}} \to \boldsymbol{r_{ij}'})\right]\right).
\end{aligned} \tag{3.27}$$

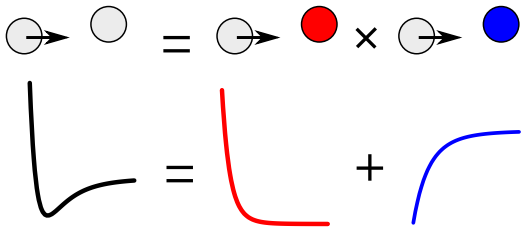


Figure 3.11: Factorization of the Lennard-Jones potential (black) on its attractive (blue) and repulsive (red) parts. It can be interpreted as substituting two ghosts particles to the actual one. One of the ghost particles codes for the attractive interaction, the other for the repulsive one. For a given move, only one of the ghost particles may reject the moves.

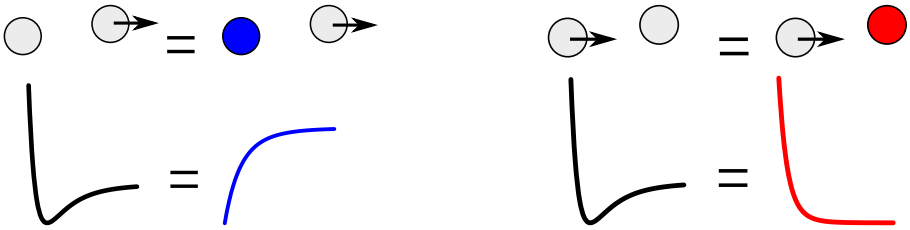


Figure 3.12: Factorization of the Lennard-Jones potential on its attractive (blue) and repulsive (red) parts. **Left:** As the moving sphere gets further away from the other sphere, only the attractive part can reject the move. **Right:** As the moving sphere gets closer, the global potential reduces in that case to its repulsive part, as the attractive part always accepts the move.

Decomposing a potential E into factors can then be useful to separate different components of E, instead of only factorizing along the coordinates. For instance, the factorization can be used to separate the attractive $\propto 1/r^6$ and repulsive $\propto 1/r^{12}$ components in the Lennard-Jones potential, see Fig. 3.11. For a given move, only one of the two components will participate in the total $p^{\text{acc}}_{\text{Fact}}$, see Fig. 3.12. Eq. 3.25 can indeed be rewritten as,

$$p^{\text{acc}}_{\text{Fact}}(\zeta \to \zeta') = \prod_i \exp\left[-\beta \max\left(0, \delta E_i(\zeta \to \zeta')\right]\right., \tag{3.28}$$

noting δE_i the factors of E. Thus, only the positive factors, translating an increase of energy, participate in the value of $p^{\text{acc}}_{\text{Fact}}$, as negative factors always gives 1. Repulsive and attractive factors are always of opposite signs and, therefore, only one of the two does participate.

3.2.2 Implementation

Unlike in the Metropolis filter, every singled out factor of the potential has an independent probability to reject or accept the move. There are two ways of sampling the

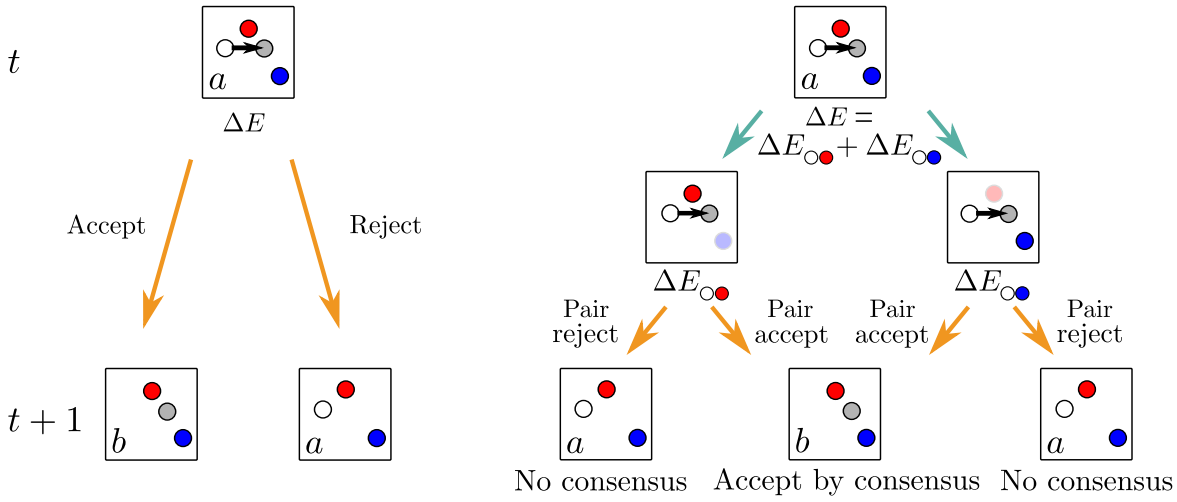


Figure 3.13: At time t, a move is proposed from configuration a to configuration b. **Left:** Traditional Metropolis move: following $p^{\text{acc}}_{\text{Met}}$, it is either accepted, leading to the configuration b at time $t+1$ or rejected, leading to configuration a at time $t+1$. **Right:** Consensus rule imposed by the factorized filter: The change of energy between configuration a and configuration b is treated independently for each factor of the energy, here the pair energies. If both factors accept the move, the move is accepted. If one of the factor rejects the move, the move is rejected.

acceptance or the rejection of a proposed move $\zeta \to \zeta'$,

- One computes the global value of $p^{\text{acc}}_{\text{Fact}}(\zeta \to \zeta')$ and draws a random number $\mathbf{ran}(0,1)$ to sample if the move is accepted, see the left panel of Fig. 3.13.

- Capitalizing on the formulation based on a product of probabilities of independent events, one computes for each factor δE_i the respective probability $p^{\text{acc}}_{\text{Met}}(\delta E_i)$ and draws a random number $\mathbf{ran}(0,1)$ that will decide if i accepts the move or not. This is the consensus rule, as is illustrated in Fig. 3.13. The move will be indeed accepted if all independent factors accept it, see the right panel of Fig. 3.13.

The consensus rule asks for more random numbers, but it allows to get a direct information on which factor δE_i is rejecting the move or not. Information that will be extremely useful to implement a maximal global-balance scheme, as was done for two particles, see Section 3.1.3. More generally, as $p^{\text{Fact}}_{\text{acc}} \leq p^{\text{Met}}_{\text{acc}}$, the interest of the factorized filter is limited for detailed-balance scheme, as it induces more rejection than the Metropolis one. For instance, the rate of rejections is about 50% higher for a bidimensional system of N soft spheres in a volume V, with a pairwise interaction $\propto 1/r^{12}$ on a range of density $\rho = N/V \in [0.8, 1.2]$, Publication 1 [30]. For hard-sphere system, the factorized and standard Metropolis filter agree, as the acceptance probability of a move is always one (no overlap created by the move) or zero (overlap created by the move) in both filter.

3.2.3 The factorized Metropolis filter in the lifting framework, beyond $\pm$ replicas

The factorized filter allows to generalize the maximal global balance described in Section 3.1.3 from two to N particles in d dimensions, as the multidimensional potential is factorized into independent unidimensional problems. We assume for now that multiple rejections are not produced, this will be addressed in the Section 3.3.

The irreversible factorized scheme is as follows: A sphere i is updated according to a direction $+x_k$. When the move of the sphere i is rejected by the potential factor of the pair energy with a sphere j, see Fig. 3.13, the sphere j is then moved in turn on the same axis $+x$. At some point, the direction is resampled.

In the lifting framework, the lifting variable $\sigma^{\text{fact}} = (x_k,\ z = i)$ sets the moving sphere i and the direction of the moves along $+x_k$. When there is a rejection of the physical move by a sphere j, the lifting variable is updated to $(x_k,\ z = j)$. After a fixed number of steps, the lifting variables is resampled, so that the chain is irreducible and explores all the directions. The resulting scheme is presented in Fig. 3.14.

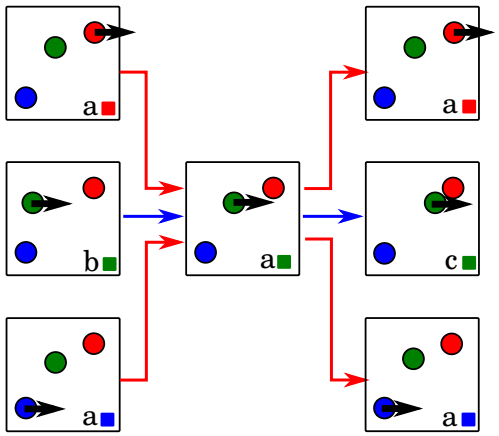


Figure 3.14: Maximal global-balance condition for a three-sphere system. The physical configurations a, b and c are extended by the lifting variable fixing the moving sphere (■: red sphere moving, ■: blue sphere moving and ■: green sphere moving) and the direction, here $+x$. The physical moves are represented by blue arrows and the lifting moves by red ones. When a physical move is rejected, a lifting move happens instead. The physical configuration is unchanged but the lifting variable is updated to the sphere rejecting the increase in pair energy. It is the sphere that is moved next.

As explained in Section 3.1.3, this maximal global-balance scheme differs from the $d \times 2$-replicas lifting of [26], see Section 3.1.2.2. In the $d \times 2$-replicas formalism of [26], there is one partition $(i,\ x_k)$ of independent lines per dimension and per sphere. A state will be updated, as a sphere moves back and forth along a direction. These moves are set by the lifting variable $\sigma^{\pm} = (i,\ x_k,\ z = \pm)$. The back-and-forth moves set by $z = \pm$ ensure that the global-balance condition is obeyed.

In the present scheme, there is only one partition x_k per dimension. The lifting

variable component z takes its value in the set of spheres $\{i\}$. Thanks to the factorized filter, the global-balance condition is fulfilled, as illustrated in Fig. 3.14 and computed in Eq. 3.35 and Eq. 3.36.

The factorized filter generalizes the lifting framework to schemes with $d \times N$-replicas. The set of potential factors needs to be chosen with care, in order to realize the most efficient scheme without backward moves. This choice is actually quite natural in pairwise potential, where the factors identify with the pair energies $E_{ij}(|\boldsymbol{r_i} - \boldsymbol{r_j}|)$, which are symmetric in $\boldsymbol{r_i}$ and $\boldsymbol{r_j}$, so that the change of E_{ij} is the same, if $\boldsymbol{r_i}$ is updated by $-\boldsymbol{\delta}$ (used in the scheme $z = \pm$) or if $\boldsymbol{r_j}$ is updated by $\boldsymbol{\delta}$ (used in the scheme $z \in \{i\}$). The next Section 3.3 addresses in more details the requirements for the factors; the need to implement infinitesimal steps to avoid multiple rejection is also discussed.

3.3 Continuous-time scheme

In Section 3.2.3, we did not address the problem of multiple rejections. As for now, the scheme illustrated in Fig. 3.14 does not solve this problem. Even if the filter is factorized, one can imagine easily that a move with a big enough energy increase could trigger several factors. In those conditions, in order to obey global balance, one would need to enforce again backwards moves. In this Section, we introduce infinitesimal steps into the scheme, making the event of two factors being triggered at the same step an event of asymptotic probability zero. We explain how to implement an event-driven approach and how to average observables on the continuum of valid configurations between two lifting moves.

3.3.1 Maximal global-balance by infinitesimal steps

The global-balance condition enforces incompressible flows. The lifting flows compensate for the compressibility generated by the irreversible physical moves, see Fig. 3.15. If two factors reject a move, as it may be the case if implementing finite moves δx, the compressibility cannot then be counterbalanced by a single lifting flow. Infinitesimal steps dx ensures that only one factor at a time can cause a rejection, meaning that,

$$p^{\text{acc}}_{\text{fact}}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx}) = \prod_j p^{\text{acc}}_{ij} \approx 1 - \sum_j (1 - p^{\text{acc}}_{ij}) \prod_{k \neq j} p^{\text{acc}}_{ik} \approx 1 - \sum_j (1 - p^{\text{acc}}_{ij}). \tag{3.29}$$

For $p^{\text{acc}}_{ij}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx}) = 1 - \epsilon_{ij} \xrightarrow[dx \to 0]{} 1$, Eq. 3.29 is true at first order. Noting $\epsilon \geq \epsilon_{ij}$, we have $\forall \epsilon_{ij}$

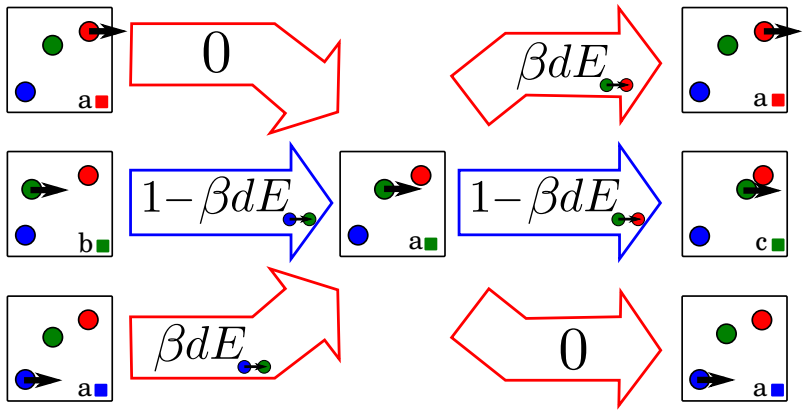


Figure 3.15: Probability flows for an irreversible Markov chain with the factorized Metropolis filter, implemented with infinitesimal steps in a three-sphere system with repulsive interactions.The physical configurations a, b and c are extended by the lifting variable fixing the moving sphere (■: red sphere moving, ■: blue sphere moving and ■: green sphere moving) and the direction, here $+x$. The physical flows are represented by the blue arrows and the lifting flows by the red arrows. The value represented in each arrow correspond to the flow value divided by the probability weight of the central configuration a. Lifting moves only happen when the energy is increasing during the physical moves, i.e. when the moving sphere goes closer to another.

$$\begin{aligned} p^{\text{acc}}_{\text{fact}}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx}) &= \prod_j p^{\text{acc}}_{ij} \\ &= 1 - \sum_{\text{subset}P} \prod_{j\in P}(1 - p^{\text{acc}}_{ij}) \prod_{k\notin P} p^{\text{acc}}_{ik} \\ &= 1 - \sum_{\text{subset}P} \prod_{j\in P} \epsilon_{ij} \prod_{k\notin P}(1 - \epsilon_{ik}) \\ &= 1 - \sum_j (1 - p^{\text{acc}}_{ij}) \prod_{k\neq j} p^{\text{acc}}_{ik} + o(\epsilon). \\ &= 1 - \sum_j (1 - p^{\text{acc}}_{ij}) + o(\epsilon). \end{aligned} \tag{3.30}$$

The element ϵ_{ij} identifies with the energy increase of the factor δE_{ij} caused by the move dx,

$$\begin{aligned} p^{\text{acc}}_{ij}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx}) &= \exp\left[-\beta \max\left(0, dE_{ij}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx})\right)\right] \\ &= 1 - \beta \max\left(0, dE_{ij}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx})\right) \xrightarrow[dx\to 0]{} 1. \end{aligned} \tag{3.31}$$

The total acceptance probability is then,

$$p^{\text{acc}}_{\text{fact}}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx}) = \prod_j (1 - \beta dE^+_{ij}) = 1 - \sum_j \beta dE^+_{ij} = \prod_j (1 - p^{\text{lift}}_{ij}), \tag{3.32}$$

noting $dE_{ij}^{+} = \max\left(0, dE_{ij}\right)$. Either the physical move is accepted or a lifting move takes place. The lifting probability per factor is,

$$p_{ij}^{\text{lift}}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx}) = 1 - p_{ij}^{\text{acc}}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx}) = \beta dE_{ij}^{+}, \tag{3.33}$$

When proposing a move, only the factors that code for an energy increase can reject the move with a probability $p_{ij}^{\text{lift}} > 0$. After a physical rejection, a lifting move happens between the replicas i and j. In a N particle system, it corresponds to moving the sphere j instead of the sphere i. The lifting moves only flows into the replica, where the move decreases the energy. It is useful to note that,

$$p_{ij}^{\text{lift}}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx}) + p_{ij}^{\text{lift}}(\boldsymbol{r_j} \to \boldsymbol{r_j} + \boldsymbol{dx}) = \beta dE_{ij}^{+} + \beta(-dE)_{ij}^{+} = \beta dE_{ij}. \tag{3.34}$$

We consider a system of N spheres, interacting by the potential $E = \sum_{\langle i,j \rangle} E_{ij}(\boldsymbol{r_{ij}})$. We note ζ the configuration $\{\boldsymbol{r_1}, \ldots, \boldsymbol{r_i}, \ldots, \boldsymbol{r_N}\}$ and $\zeta(i + dx) = \{\boldsymbol{r_1}, \ldots, \boldsymbol{r_i} + \boldsymbol{dx}, \ldots, \boldsymbol{r_N}\}$. As illustrated in Fig. 3.15, the flows IN and the flows OUT of a configuration $\{\zeta(i), \sigma = (i, +x)\}$ are equal to $\pi(\zeta)$, achieving maximal global balance,

$$\text{Flows OUT} \begin{cases} \text{Physical flow: } \pi(\zeta) p_{\text{fact}}^{\text{acc}}(\zeta(i) \to \zeta(i + dx)) \\ \qquad = \pi(\zeta) \left(1 - \sum_j \beta dE_{ij}^{+}\right) \\ \text{Lifting flow: } \pi(\zeta)(1 - p_{\text{fact}}^{\text{acc}}(\zeta(i) \to \zeta(i + dx))) \\ \qquad = \pi(\zeta) \sum_j p_{ij}^{\text{lift}} = \pi(\zeta) \left(\sum_j \beta dE_{ij}^{+}\right) \\ \hline \text{Total flow OUT: } \pi(\zeta) \end{cases} \tag{3.35}$$

$$\text{Flows IN} \begin{cases} \text{Physical flow: } \pi(\zeta(i - dx)) p_{\text{fact}}^{\text{acc}}(\zeta(i - dx) \to \zeta(i)) \\ \qquad = \pi(\zeta) p_{\text{fact}}^{\text{acc}}(\zeta(i) \to \zeta(i - dx)) \\ \qquad = \pi(\zeta) \left(1 - \sum_j \beta \left(-dE_{ij}\right)^{+}\right) \\ \text{Lifting flow: } \sum_j \pi(\zeta) p_{ij}^{\text{lift}}(\zeta(j) \to \zeta(j + dx)) \\ \qquad = \pi(\zeta) \sum_j \beta \left(-dE_{ij}\right)^{+} \\ \hline \text{Total flow IN: } \pi(\zeta) \end{cases} \tag{3.36}$$

When a lifting move happens, a flow $\pi(\zeta)\beta dE$ is redirected from one replica to the other. For the maximal global-balance scheme to be correct, this flow needs to

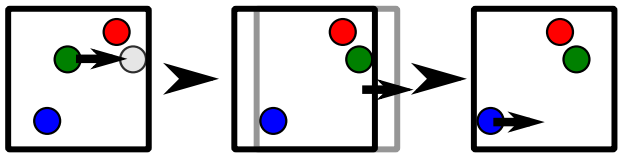


Figure 3.16: For wall boundary, it is possible to conserve the maximal global-balance scheme by factorizing the hard core interaction of the wall with each particle. When a particle hits the wall, the wall is then moved forward, until it hits another particle, which is then moved in turn.

compensate the physical flow. If the physical flow is equal to $\pi(\zeta)(1-\beta dE)$, the scheme is correct. Therefore the replica reached by lifting is the one where the moves are decreasing the energy, as it was previously increased in the former replica.

The factorized filter allows to propose another replica than just the one where moves are proposing backwards. This filter matches indeed in a global manner the compressibility of the physical flows by the lifting flows. It relies, however, on the symmetry of the potential factor δE on two variables $x_i \neq x_j$, so that $\delta E(+dx_i) = -\delta E(-dx_j)$. It is the case for pairwise potentials, that only depend on the difference $x_i - x_j$. If it is not possible to decompose the potential E into factors exhibiting such symmetry, for instance for a three-body interactions, a general solution is to implement backwards moves. However, the factorized filter allows to separate the symmetric components of the potential from the non symmetric ones, so that the symmetric parts are treated with a maximal global-balance scheme and the non symmetric ones with only a global-balance scheme. The global-balance scheme on the non symmetric components can actually be improved from the back-and-forth scheme, as explained in Section 3.4.

Finally, the maximal global-balance scheme is correct, even if the system has not a periodic boundary. If the boundary is a wall, the distance between the sphere hitting the wall and the wall itself will have to increase on the next move. Instead of making the sphere go backward, it may be better to actually make the wall go forward, i.e. all particles go backwards, until it hits another particle on the other side, which will be moved in turn. As exhibited in Fig. 3.16, this last scheme avoids to implement directly backwards moves and is maximal global-balanced. The same applies to particles trapped in an external potential. When the external potential rejects a physical move, it is moved forward until it is rejected by another particle, which will be the one being moved next.

3.3.2 Event-driven approach: Event sampling

3.3.2.1 Sampling a lifting event

In practice, infinitesimal moves require a scaling of the physical time, as in one unit of time, an infinite number of physical moves take place. However, the number of

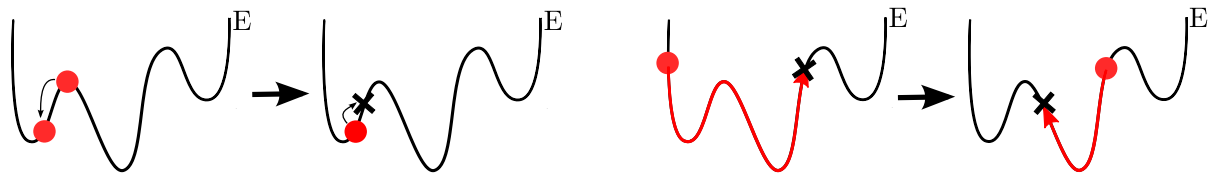


Figure 3.17: **Left:** detailed-balance Markov chains with local and finite moves. **Right:** Irreversible Markov chains with an event-driven implementation producing a continuum of valid configurations.

lifting moves remains finite. In an event-driven approach, the algorithmic complexity can be then made to scale with the number of lifting moves. Instead of sampling the acceptance of every infinitesimal physical moves, we directly sample the moment of a lifting move, called an event, as was done in [24], following [132]. Between two events, there is an infinity of Monte Carlo steps, giving a continuous trajectory on valid configurations, as shown in Fig. 3.17.

For an one-dimensional system with probability distribution π, potential E and a lifting variable updating the potential along the trajectory $s \to s + ds$, the probability $p_{\text{event}}(0 \to n)$ to accept n subsequent physical moves and then to reject the $(n+1)$-st physical move is,

$$p_{\text{event}}(0 \to n) = p_{\text{acc}}(0 \to 1) p_{\text{acc}}(1 \to 2) \dots p_{\text{acc}}(n-1 \to n)(1 - p_{\text{acc}}(n \to n+1)). \tag{3.37}$$

The jth term in this expression is $\min(1, \pi_j/\pi_{j-1})$, where $\pi_j = \pi(s_j = s + jds)$. Considering first the case where π_j is monotonically decreasing with j, this gives,

$$p_{\text{event}}(0 \to n) = \frac{\pi_{n-1}}{\pi_0}\left(1 - \frac{\pi_n}{\pi_{n-1}}\right) = \frac{-1}{\pi_0}\frac{\partial \pi}{\partial s}\Big|_{s=s_n} ds. \tag{3.38}$$

This probability is normalized, writing s_{event} the value of s at which the event of the rejection of the physical moves happens, by sample transformation (Section 2.1.2),

$$\frac{-1}{\pi_0}\int_0^\infty \frac{\partial \pi}{\partial s}\Big|_{s=s_{\text{event}}} ds_{\text{event}} = \frac{1}{\pi_0}\int_0^{\pi_0} d\pi_{\text{event}} = 1. \tag{3.39}$$

This integral is directly sampled by, see Section 2.1.2,

$$\begin{aligned} \pi_{\text{event}} &= \mathbf{ran}(0, \pi_0) \\ \pi_{\text{event}}/\pi_0 &= \mathbf{ran}(0,1), \end{aligned} \tag{3.40}$$

which is equivalent to the following sampling of the energy increase,

$$\Delta E(s_{\text{event}}) = -[\log \mathbf{ran}(0,1)]/\beta. \tag{3.41}$$

Simply sampling π uniformly between 0 and the present value, π_0 (equivalently, ΔE from its exponential distribution) thus yields the event time, s_{event}, see Fig. 3.18.

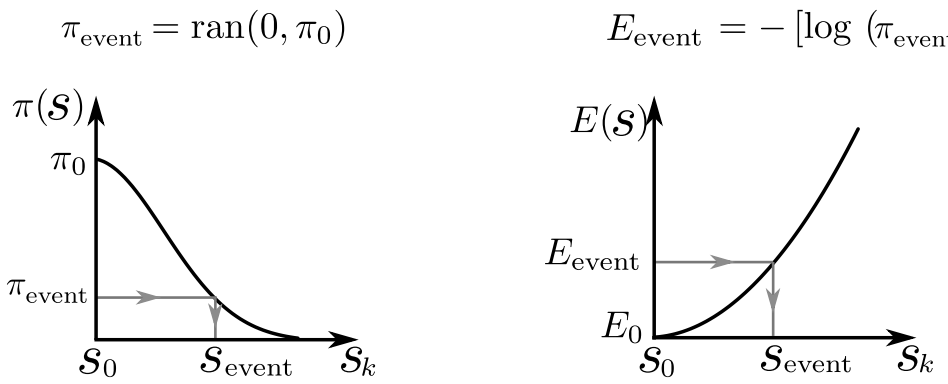


Figure 3.18: Event-driven implementation for a monotonic probability distribution π. From a starting point $s_k = s_0$ of weight π_o and energy E_0, s_k is updated by infinitesimal moves until $s_k = s_{\text{event}}$. The lifting event is sampled as $\pi_{\text{event}} = \mathbf{ran}(0, \pi_0)$, equivalent to an energy increase $\Delta E(s_{\text{event}}) = -[\log \mathbf{ran}(0,1)]/\beta$.

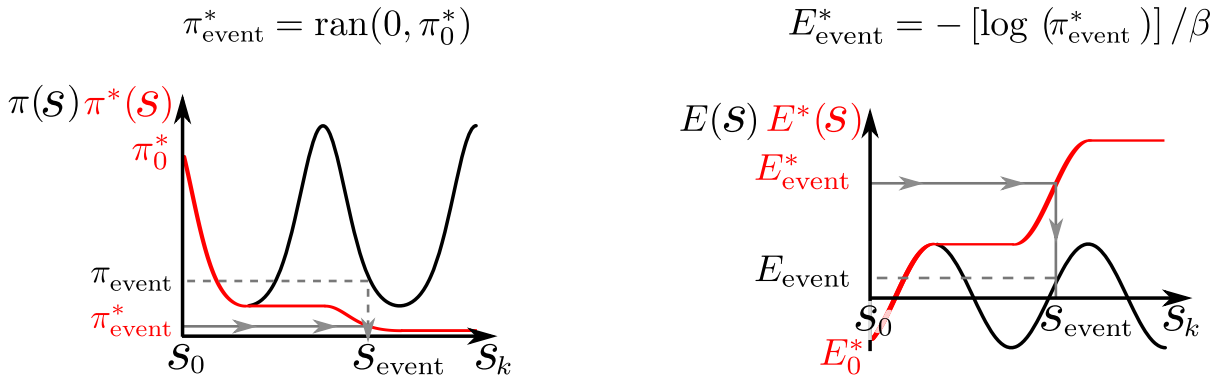


Figure 3.19: Event-driven implementation for a general probability distribution π. From a starting point $s_k = s_0$ of weight π_o and energy E_0, s_k is updated by infinitesimal moves until $s_k = s_{\text{event}}$. The lifting event is sampled as $\Delta E^*_{\text{event}} = [-\log \mathbf{ran}(0,1)]/\beta$.

For a non-monotonic probability distribution, all negative energy increments correspond to an acceptance probability 1, and disappear from Eq. 3.37. The sampling of the total energy increment in Eq. 3.41 turns into the sampling of only the positive energy increments. As shown in Fig. 3.19, this can be expressed as a function ΔE^*, constructed only from the positive increments of the energy E [25],

$$\Delta E^*(s_0 \to s_f) = \int_{s_0}^{s_f} \max\left(0, \frac{\partial E}{\partial s}\right) ds = \int_{s_0}^{s_f} dE^+(s). \tag{3.42}$$

It is noteworthy that this integral depends on the path followed by the updating of s. The event time s_{event} is then sampled by drawing a random number $\mathbf{ran}(0,1)$ and finding s_{event} so that,

$$\Delta E^*_{\text{event}} = \Delta E^*(s_0 \to s_{\text{event}}) = -[\log \mathbf{ran}(0,1)]/\beta. \tag{3.43}$$

For multidimensional potentials, the acceptance probability of each potential factor

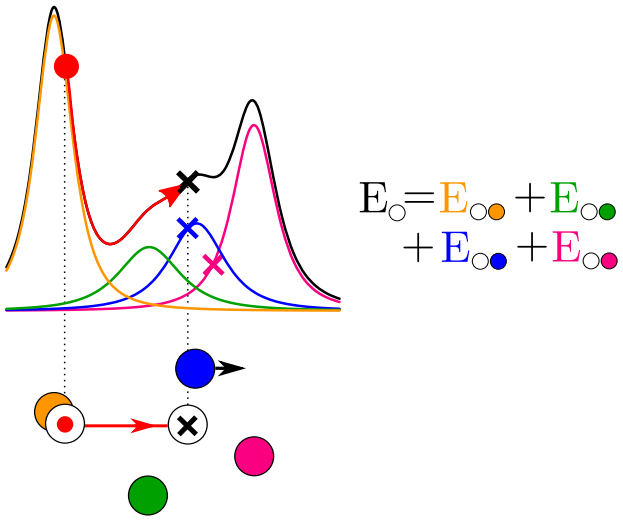


Figure 3.20: Event sampling for complex systems: A lifting event is sampled for each factor, here each pair energy, according to Eq. 3.43. Here the earliest lifting event involves the blue sphere. The orange sphere is then updated until the event time and the blue sphere will be the next one to be moved. The lifting event may not always involve the closest sphere. The energy increase sampled in Eq. 3.43 for the interaction between the orange and green spheres is too large to be achieved by a single approach of the orange sphere towards the green one.

δE_{ij} is independent. A lifting event is then sampled for each factor, according to the function ΔE^*_{ij}, which stores all positive values of δE_{ij}. The earliest event is the one actually realized, as illustrated on Fig. 3.20. The increment ΔE^*_{ij} can be interpreted as an energy increase made possible by the thermal fluctuations.

In this event-driven implementation, the fixed number of steps after which the direction of the moves and the moving spheres are resampled is replaced by a global displacement ℓ, referred to as the chain length: The distances travelled by the spheres between two lifting events are summed up. When their sum reaches ℓ, the lifting variable (direction and sphere) is uniformly resampled. The scheme of [26], see Section 3.1.2.2, would be impossible to implement while using infinitesimal steps, as the lifting variable would be resampled at every infinitesimal step. Along a given chain, moves share the same direction, as only the component z of the lifting variable is changed, when a lifting event occurs. The collection of chains can be interpreted as a collection of finite-time realizations of the Markov chain.

3.3.2.2 Practical computation of lifting events

It is possible that Eq. 3.43 yields a time event $s_{\text{event}} = \infty$, for example when the potential factor δE_{ij} is always zero, while the state is updated according to $+ds$ or when the thermal excitation ΔE^*_{ij} is too large and cannot be reached. Solving Eq. 3.43

can be nontrivial in general. Many potentials are, however, central potentials that depend on the distance between particles. If the pair potential consists of several terms, e.g. attractive and repulsive terms, it may be convenient to further factorize the potential, as was explained for the Lennard-Jones potential, see Section 3.2.1. This decomposition can lead to a higher event rate than the full potential. For instance, for the Lennard-Jones potential, the mean free path between events is reduced at most by half. In return, the decomposition simplifies the implementation of the irreversible factorized Metropolis algorithm, as a potential can be decomposed on monotonic components.

We now discuss how to solve Eq. 3.43 for a soft-sphere potential $E \propto 1/r^n$, first for a system without boundary. A sphere i is moved along the direction $\boldsymbol{s}$. Lifting events only happen when the energy is increased by the proposed physical moves, which is translated in Eq. 3.43 by the fact that only the positive energy increase matters. As E is monotonic on its rising part, i.e. as the spheres get closer, Eq. 3.43 amounts to sampling the energy increase $\Delta E^*_{ij} = -\log \mathbf{ran}(0,1)/\beta$, as in Eq. 3.41. It fixes the event time of the pair energy,

$$E^{\text{event}}_{ij} = E_{ij}(r_{ij}(s_{\text{event}})) = E^*_{ij}(0) + \Delta E^*_{ij}, \tag{3.44}$$

with,

$$E^*_{ij}(0) = \begin{cases} E_{ij}(r_{ij}(s=0)) & \text{if the initial pair energy derivative is positive.} \\ 0 & \text{otherwise} \end{cases} \tag{3.45}$$

and thus the interdistance at the event time is

$$r^{\text{event}}_{ij} = E^{-1}_{ij}(E^{\text{event}}_{ij}). \tag{3.46}$$

If Eq. 3.46 does not have a solution because E^{event}_{ij} exceeds any possible value, no lifting event involving the pair $\{ij\}$ happens, as for the green and orange spheres in Fig. 3.20. The physical moves are always accepted by j. If Eq. 3.46 does have a solution, s_{event} is then inferred from Eq. 3.46 by taking the positive root of $r^{\text{event}}_{ij} = |\boldsymbol{r_i} + \boldsymbol{s}_{\text{event}} - \boldsymbol{r_j}|$. Again, if no such root exists, no lifting event happens. In case of no lifting event, the sphere i is displaced until the finite chain length ℓ is reached. This situation is particular to some systems, for instance, it does not arise in spin systems.

In a periodic box of length L, a particle can pass by the same lifting partner more than once. The potential is no longer monotonic but periodic, as it is the case in Fig. 3.19. When the particle travels over a full box size, the energy increase is $\Delta E^{\max}_{ij}$. To solve Eq. 3.43, one first slices off the n full box translations allowed by the sampled total energy increase ΔE^*_{ij}, leaving a value $\Delta E^{*,f}_{ij}$,

$$\Delta E^*_{ij} = \Delta E^{*,f}_{ij} + n\Delta E^{\max}_{ij}, \tag{3.47}$$

and the distance s_f corresponding to $\Delta E_{ij}^{*,f}$ is calculated as above. Finally,

$$s_{\text{event}} = s_f + nL. \tag{3.48}$$

Even if solving Eq. 3.43 in periodic boundary does not complicate the implementation, it is possible to tune the fixed total displacement ℓ, so that the moving sphere can only interact with one periodic image of the other spheres or by introducing a lifting event of the moving sphere with itself after a displacement of half the box for instance.

3.3.3 Observable averaging

As the continuum of configurations between two lifting events is correctly sampled from π, one has to draw a uniform subset from this continuum, in order to retrieve the ensemble average of an observable. This may be done either by outputting the configuration at the end of a chain or during the chain itself. This problem is similar to the one of drawing a time at which to resample the lifting variable without introducing bias, as it also cannot be done at each infinitesimal step.

The lifting variable is resampled at the precise moment when the sequential displacements sum up to a length ℓ. At this point, the instantaneous configuration can be saved for ensemble averaging. This way, no bias is introduced as configurations are output at regular intervals, independently of the occurrence of the events. In constrast, it is incorrect to sample configuration at the event time. Configurations at an event time follow a different probability distribution, as, for instance in a hard-sphere system, these configurations always contain two touching spheres. For a given lifting variable updating the state by ds, the configurations output at the lifting events follow the probability distribution $|\partial\pi/\partial s ds|$, as from Eq. 3.38,

$$P(\text{event in state } (x,s)) = \pi(x) p_{\text{event}}(x \to x + ds) = \left[-\frac{\partial \pi}{\partial s}\Big|_x ds \right]^+ . \tag{3.49}$$

Therefore a Dirac function appears for the hard spheres (two touching spheres), as it is the derivative of the Heaviside distribution, Eq. 2.26.

The chain length ℓ can be either fixed beforehand or drawn randomly before starting a chain. Another method consists in introducing a small constant probability for terminating a chain in each infinitesimal move. This would effectively lead to a Poisson process and an exponentially distributed random ℓ.

3.4 Infinite chains

In Section 3.2, we explained how the factorized Metropolis filter allows for the implementation of a maximal global-balance scheme for a given reversible Markov

chain. The factorized Metropolis filter allows indeed for the chain to explore without resampling the complete partition of a lifting variable for a given dimension. It is however possible to link the dimensions by a detailed-balance scheme and so to avoid resampling the lifting variable [33].

We consider two soft spheres in a two-dimensional box with periodic boundary and interacting by the potential E. The lifting variable σ sets the direction $+x$ or $+y$ and the moving sphere 1 or 2. When $\sigma = (+x, i)$, the sphere i is updated according to $+x$, until a rejection occurs. The moving sphere is then the sphere j and $\sigma = (+x, j)$. If the initial state is (x_1, y_1, x_2, y_2), one will explore all the states $(\cdot, y_1, \cdot, y_2)$. But, in order to reach the configurations with other values in the y-coordinate than y_1 and y_2, one will need to resample σ. If a rejection happens in $(\zeta, \sigma = (i, +x))$, it would be indeed not correct to propose a lifting move $(\zeta, \sigma = (i, +x)) \to (\zeta, \sigma = (i, +y))$, as the lifting flow from $(\zeta, \sigma = (i, +x))$ is $\pi(\zeta)[\beta \partial E/\partial x]dx$ and the flow needed to compensate the physical flow into $(\zeta, \sigma = (\square, +y))$ is $\pi(\zeta)[\beta \partial E/\partial y]$, $\square$ notes for the sphere whose update by $+dy$ will decrease the energy, i.e $[\partial E/\partial y] > 0$. The two flows are not necessary equal and implementing such moves will not be correct.

The solution proposed in [33] is to accept or not the interdimensional lifting moves by the probability,

$$p^{\text{Lift}}((\zeta, \sigma = (i, +x)) \to (\zeta, \sigma = (\square, +y))) = \min\left(1, \frac{|\partial E/\partial y|}{|\partial E/\partial x|}\right) \tag{3.50}$$

Eq. 3.50 is only defined when a lifting move is sampled from $((\zeta, \sigma = (i, +x))$. If the interdimensional lifting move is not accepted, the usual lifting move takes place. The flows are then again correctly balanced, as the Metropolis filter used in Eq. 3.50 enforced detailed-balance with the correct flows as the stationary distribution. Except in situations where it is possible for no lifting move to happen, which is impossible in the common studied models of N particles or spins, there is no need to resample the lifting variable anymore and one can run a chain without ever stopping it.

For hardcore potential, the energy partial derivative in Eq. 3.50 is replaced by the partial derivative of the interdistance r. The derivative $-\beta d_s E(\zeta) = d\pi/ds|_\zeta$ reflects indeed the local increase or decrease in the density of valid states around ζ when making a move $+ds$, the similar quantity for hardcore models is the local decrease or increase of the interdistance r, as the number of valid states identifies with the geometric distance between two collisions.

Implementing interdimensional lifting moves is particularly useful to study anisotropic particles that possess the extra degree of freedom of rotation. Considering N bidimensional hardcore dimers of diameter σ, each dimer i is described by the position of its center (x, y) and its angle ϕ, the center of each monomer i_1, i_2 is then at $(x \pm \frac{\sigma}{2}\cos(\phi), x \pm \frac{\sigma}{2}\sin(\phi))$. The lifting variable sets the moving dimer i and the direction of the move, chosen in $\{+x, +y, +\phi, -\phi\}$. Both sense of rotations are indeed

needed as, when the dimer i collides with the dimer j after a rotation $+d\phi$, rotating j by $+d\phi$ may not separate the dimer. But more importantly, rotating the dimer j, even in the rotation sense that will separate both dimers, may not lead to the same increase in the local density of valid states, as will be yielded by rotating i by $-d\phi$, see Fig. 3.21. If we consider, without lack of generality, that the two monomers colliding are i_1 and j_1, then,

$$\begin{cases} \dfrac{\partial r_{i_1 j_1}}{\partial \phi_i} = \dfrac{\sigma}{2 r_{i_1 j_1}} \left[d \sin(\phi_{ij} - \phi_i) + \dfrac{\sigma}{2} \sin(\phi_i - \phi_j) \right] \\ \dfrac{\partial r_{i_1 j_1}}{\partial \phi_j} = \dfrac{-\sigma}{2 r_{i_1 j_1}} \left[d \sin(\phi_{ij} - \phi_j) + \dfrac{\sigma}{2} \sin(\phi_i - \phi_j) \right]. \end{cases} \tag{3.51}$$

where $x_i - x_j = d\cos(\phi_{ij})$ and $y_i - y_j = d\sin(\phi_{ij})$.

The symmetry is broken by the term $\sin(\phi_{ij} - \phi_.)$. If one implements a scheme where j is updated next, the chain will end up trapped dynamically, as dimers will be in contact with no way out. This dynamical trap arises from the fact that, by rotating the collided dimer next, the chain will step by step flow into configurations less and less likely, until reaching configurations where dimers are touching, see Fig. 3.21. To correctly update the next dimer j after a collision with the dimer i, the interdimensional move $(\zeta,\ \sigma\ =\ (i,\ +\phi)) \ \rightarrow\ (\zeta,\ \sigma\ =\ (j,\ \pm\phi))$ has a probability of,

$$\begin{aligned} p^{\text{Lift}}((\zeta, \sigma = (i, +\phi)) \rightarrow (\zeta, \sigma = (j, \pm\phi))) &= \min\left(1, \frac{|\partial r_{i_1 j_1} / \partial \phi_j|}{|\partial r_{i_1 j_1} / \partial \phi_i|}\right) \\ &= \min\left(1, \frac{|d \sin(\phi_{ij} - \phi_j) + \frac{\sigma}{2} \sin(\phi_i - \phi_j)|}{|d \sin(\phi_{ij} - \phi_i) + \frac{\sigma}{2} \sin(\phi_i - \phi_j)|}\right) \end{aligned} \tag{3.52}$$

The variable $\pm$ takes the value that separates the dimer, as lifting move takes place towards replicas proposing moves to more likely configurations.

The interdimensional lifting move allows us to produce persistent rotations of dimers, instead of always rotating the same dimer back and forth. However the scheme is only global balanced and not maximal global balanced. When simulating molecules, atoms may not interact the same way. For instance, in a fluid of water molecules, the hydrogen atoms of one molecule will be more attracted towards the oxygen of another, and not to its other hydrogen atoms. As shown in Fig. 3.22, the factorized Metropolis filter can be used to separate the molecule-molecule interaction into the simpler atom-atom interactions, as can be done for the Lennard-Jones potential, see Fig. 3.11. Rotations of molecules can also be implemented with interdimensional lifting moves, in order to export a persistent move on a cluster of molecules.

More generally, interdimensional lifting moves can be used in case of a non symmetric potential, like a three-body potential. In a three-body interaction, it is not pos-

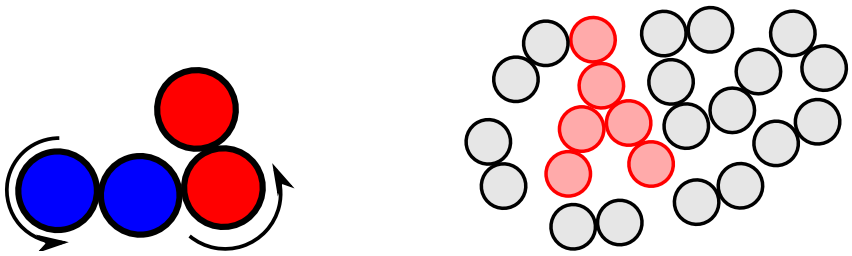


Figure 3.21: **Left:** if a dimer i collides with a dimer j by rotation, rotating the dimer j next is not correct if the flows are not rescaled with Eq. 3.52. **Right:** if implemented without taking care of rescaling the probability flows with Eq. 3.52, hardcore systems will get blocked in dynamical traps, as here in red; rotating one of the dimers leads to an immediate collision from one of the other two

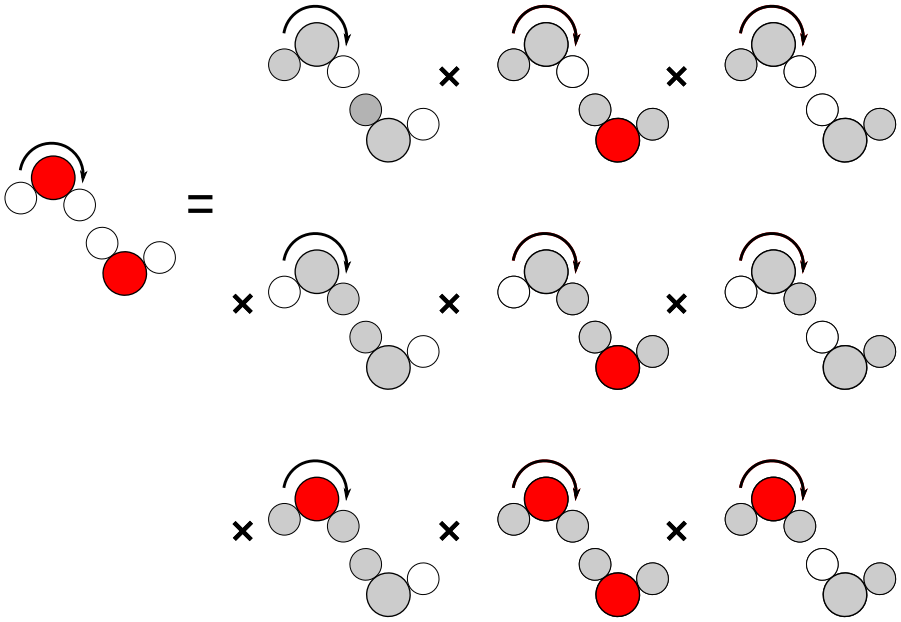


Figure 3.22: Factorization of the global interaction molecule-molecule into the simpler atom-atom interactions. Atoms not interacting in the considered factor are represented in grey. Red stands for an oxygen atom and white for an hydrogen one.

sible to segment the potential into pairwise components. Thus, while implementing the irreversible factorized scheme, if the three-body factor dE_{123} is the one rejecting the moves, the moving particle 1 has to go backwards. However, we can implement a scheme that allows to move one of the other two particles 2, 3 involved in the three-body interaction. One chooses randomly one particles i of the two particles and accepts the lifting move $(1, +s) \to (i, +s)$ proposes with the probability,

$$p^{\text{acc}}((1, +s) \to (i, +s)) = \min\left(1, \frac{|\partial E_{123}/\partial s_i|}{|\partial E_{123}/\partial s_1|}\right), \tag{3.53}$$

or one can also implement maximal global balance by implementing a tower sampling scheme, see Section 2.1.2.

Conclusion

The main theoretical result of this chapter is the factorized Metropolis filter. Combined with the lifting framework and an event-driven implementation, it offers a general paradigm for constructing maximal global-balance Markov-chains for a large class of systems. Irreversible factorized Metropolis algorithms produce chains of persistent and non-local moves and generate a continuum of valid samples of the equilibrium distribution. The factorized filter allows to generalize lifting with two replicas $z = \pm$ to a scheme with as much as N replicas $z \in \{1, ..., N\}$, with N being the number of particles. Both the original hard-sphere event-chain algorithm [24] and the hybrid Monte Carlo algorithm of [25] are particular cases of irreversible factorized Metropolis algorithms.

Factorized Metropolis algorithms are general, simple to use and can be parallelized [133, 134]. It is moreover possible to implement interdimensional lifting moves, leading to the rotation of clusters of particles or persistent moves for n-body interactions in a global-balance scheme. At the difference of many of the common acceleration methods for Markov chains, as the parallel tempering [124], the factorized Metropolis filter acts at the core of the Markov chain, in the socalled *Monte Carlo driver*. In Chapter 4, we will discuss how using such irreversible scheme improves greatly the thermalization in particles as spin systems.

CHAPTER 4

Applications of the irreversible factorized Metropolis algorithm

Numerical simulations of systems with many degrees of freedom play an important role as a non-perturbative approach to statistical physics, notably for studying equilibrium phase transitions. Phase transitions are however often characterized by dynamical slowing down, including critical slowing down in second-order transitions (Section 2.3.4.2), nucleation process in first-order transitions (Section 2.3.4.1), and glassy behavior in complex systems (Section 2.4.4). In these situations, Monte Carlo algorithms with local updates such as the Metropolis algorithm take an extremely long time to sample equilibrium states. For instance, in case of a second-order transition, the autocorrelation time scales quadratically with the linear size of the system, see Eq. 2.61. In a glassy system, the acceptance rate is small. Even if an event-driven implementation [132] is implemented to prevent from wasting simulation time into sampling rejected moves, the dynamics is still dominated by isolated deep local minima. The system follows a *futile* dynamics, where, after climbing out of a potential well, it slides all the way down in energy at the next step [12, 320].

In Chapter 3, we presented how to implement a maximal global-balance scheme relying on the factorized Metropolis filter and the lifting framework. These algorithms are referred to as irreversible factorized Metropolis Monte Carlo (IFMMC). In this Chapter, we will study the performance of IFMMC in a wide variety of systems, from soft spheres to spin glasses, in order to test the generality of a possible acceleration of the thermalization. Other algorithms that improve the efficiency in some systems do not indeed perform as well in others, as seen in Section 2.4.4 and Section 2.5.

In particle systems, good results were obtained in hard-sphere systems, where it was possible to design IFMMC without the justification of the factorized filter, [24]. We test the algorithms in soft-sphere systems, as addressed in Section 4.1, Publication 1 [30]. In particular, we discuss how the factorized Metropolis filter allows us to directly compute the pressure in particle systems, without any additional computation,

see Section 4.2. In spin systems, as discussed in Section 2.4, efficient algorithms already exist but they suffer a dramatic loss of their power in glassy systems, which display numerous metastable states. We review the implementation of IFMMC for classical and continuous spin systems, see Section 4.3. First for XY spins, with ferromagnetic interactions, see Section 4.3.1.2, and with random couplings, see Section 4.3.1.3, Publication 2 [31]. Finally, we study the performance of IFMMC in Heisenberg spin systems, see Section 4.3.2, Publication 3 [32].

4.1 Soft-sphere systems

The debate on the scenario of the melting of a bidimensional particle systems was caused in part by the dynamical slowing down related to phase transitions. It indeed leads to prohibitive large simulation times for the local Metropolis algorithm, if one uses a local Metropolis algorithm.

The factorized Metropolis filter combined with the lifting concept, offers a general framework for implementing irreversible Markov-chain Monte Carlo for general multi-dimensional potentials. We test its performance in bidimensional soft-sphere systems, Publication 1[30].

4.1.1 Melting in soft spheres

Two-dimensional particle systems with short-range interactions has been shown to form solids [43], albeit they do not exhibit long positional order [135], because of the importance of fluctuations. Two-dimensional solids are characterized by long-range orientational and quasilong-range positional orders, where positional correlation functions decay algebraically. When the system is liquid, positional and orientational order are both short ranged and the correlation functions are characterized by an exponential decay.

As the solid melts, it may go through an intermediate hexatic phase, characterized by short-range positional but quasi-long range orientational order. However, the scenario of the melting transition has been heavily debated, whether it follows, as in three dimensions, an one-step first order scenario between the liquid and the solid, without the intermediate hexatic phase, or whether it exhibits the two-steps scenario of Kosterliz, Thouless [136, 137], Halperin, Nelson [138, 139] and Young [140] (KTHNY) transition, where the hexatic phase separates the liquid phase from the solid ones by continuous transitions.

In hard-sphere systems, Bernard et al. (2011) [13] showed that bidimensional hard spheres melt through first a continuous transition from solid to hexatic and second through a first-order transition from hexatic to liquid. It was thanks to the event-chain algorithm [24] that the simulations could thermalize in a reasonable time, as it can bring in large systems of 10^6 particles a speedup of as much as two orders of magnitude in comparison to a local update scheme [74].

4.1.2 Performance

We studied a two-dimensional system of N particles interacting with a truncated pairwise power-law potential,

$$\begin{cases} E_{ij}(r) = \tilde{E}(\min(r, r_c)) \\ \tilde{E}(r) = \epsilon(\sigma/r)^n. \end{cases} \tag{4.1}$$

with $r_c = 1.8\sigma$, σ being the particle diameter. This potential describe a broad range of interactions, such as the dipole interaction in magnetic colloids [141] and Lennard-Jones particles, once decomposed into repulsive (n = 6) and attractive (n = 12) factors, see Section 3.2.1.

IFMMC is implemented by separating the global potential into independent factors identifying with the pair energies,

$$p^{\text{acc}}_{\text{Fact}}(\boldsymbol{r_i} \to \boldsymbol{r_i} + \boldsymbol{dx}) = \prod_{ij} \exp(-\beta d E^{+}_{ij}). \tag{4.2}$$

The configuration is updated by specifying the moving sphere and the direction. A sphere is updated in a persistent way along $+x$ or $+y$, until a lifting move is sampled. The sphere involved in the lifting is updated next. Lifting events are determined as is detailed in Section 3.3.2.2.

We compare the relative performance of the algorithms by computing the autocorrelation time for the slowest mode, see Section 2.3. As was previously done in particle systems [24, 74], we consider the global orientational order parameter Ψ_6 to exhibit the slow relaxation of the system,

$$\Psi_6 = \frac{1}{N}\sum_{i,j} \psi_{ij} = \frac{1}{N}\sum_{i,j} \frac{A_{ij}}{A_i} \exp(6i\theta_{ij}), \tag{4.3}$$

where θ_{ij} is the angle of the bond vector between particles i and j against a fixed axis and A_{ij}/A_i the contribution of particle j to particle i's Voronoi cell perimeter [142, 143], see Fig. 4.1.

As for any other observable, the autocorrelation function of Ψ_6 decays exponentially on a timescale τ_6. Ψ_6 being a global observable, we assume that τ_6 is representative for the relaxation time for dense liquid states close to the freezing point, located at $\rho\sigma^2 = 1$ for $n = 12$ and $\rho\sigma^2 = 0.89$ for $n = 1024$, where $\rho\sigma^2$ is the dimensionless density, $N\sigma^2/V$. Moreover, by symmetry, Ψ_6 has an average of zero, so that it prevents from missing a long time scale, even if we do not know the full probability distribution $\pi(\Psi_6)$. The autocorrelation of Ψ_6 then is simply,

$$C_{\Psi_6} = \frac{\langle \Psi_6(0)\Psi_6(t)\rangle}{\langle \Psi_6(0)\Psi_6(0)\rangle} \tag{4.4}$$

Both algorithms were tuned to their optimal parameters: IFMMC: A chain spans half the system volume; LMC : A move is proposed by a random vector sampled

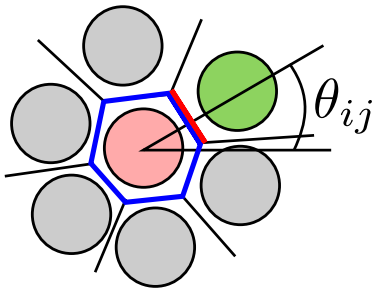


Figure 4.1: Local orientational order ψ_{ij} of the disk i (red) and the disk j (green). The nearest neighbors of i are determined by the Voronoi tesselation (Delaunay triangulation [144]): The disk i and j are neighbors if the midpoint of the line connecting their centers is closer to i and j than to any other sphere. The angle θ_{ij} is determined according to a fixed axis. The ratio A_{ij}/A_i corresponds to the ratio of the Voronoi facet length between i and j (red facet) over the total perimeter of the Voronoi cell around i.

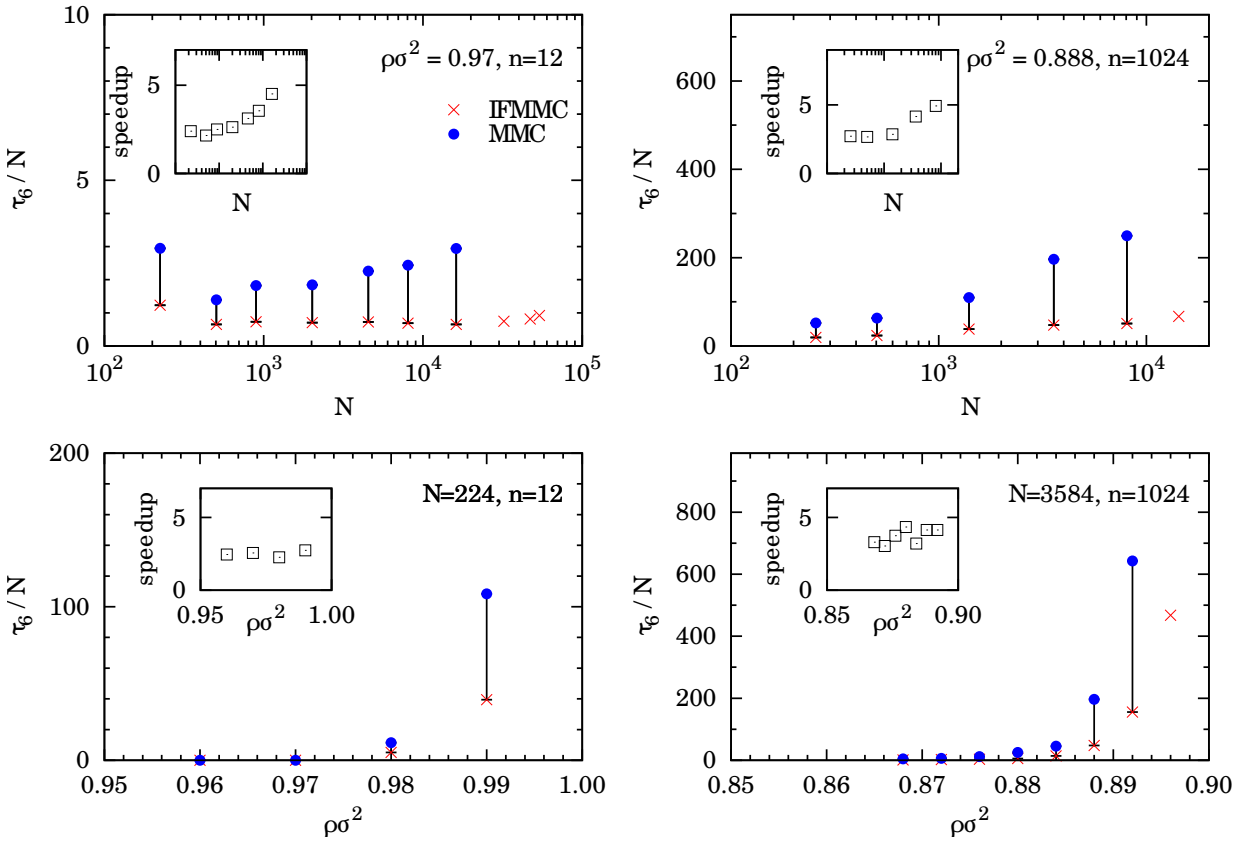


Figure 4.2: Autocorrelation time τ_6 of the global orientational parameter Ψ_6, Eq. 4.3 for IFMMC and LMC for a soft-disk system of N particles at inverse temperature $\beta = 1$. τ_6 is measured in CPU time. For large system sizes, LMC does not thermalize within the simulation time and the distribution $|\Psi_6|$ is not stationary, even though the simulation time exceeds τ_6 by a factor of 100. Therefore these values of τ_6 were not determined for these sizes.

from a disk of radius 0.16 σ. implementing IFMMC requires to sample a random number per pair energy, whereas LMC only needs one per step. However, the main computational workload occured for evaluating the potentials, not for the generation of random numbers. Therefore, one event of IFMMC is implemented in the same time as one attempted step of LMC. The time unit used to compare both algorithms is the CPU time, which corresponds to a global displacement of $100N\sigma$ in IFMMC and to $1000N$ attempted moves in LMC, where moves are proposed by a random vector sampled from a disk of radius 0.16 σ. We define speedup as the ratio of τ_6 in LMC over τ_6 in IFMMC.

As shown in Fig. 4.2, it appeared that IFMMC relaxes a few times quicker than MCMC for small system sizes in CPU units. In the region of study, the speedup does not seem to depend strongly on the density ρ, whereas it increases with the system size. This is consistent with what was observed in hard-sphere systems, [24, 74]. As LMC did not equilibrate in the allotted simulation time, we could not investigate further. However, IFMMC is expected to present the same increasing speedup with N for larger systems, all the more as there is no important difference in the speedup evolution between soft spheres $n = 12$ and quasi-hard spheres $n = 1024$. Finally, IFMMC was used with success in [145] for determining the melting scenario of soft spheres for $3 \leq n \leq 1024$, where the existence of the intermediate hexatic phase was observed for each n. Systems up to 10^6 particles were simulated.

4.2 Pressure computation

By producing a continuum of valid configurations and sampling lifting moves with a rate proportional to dE_{ij}^{+}, it is possible to compute directly the pressure and the stress tensor as a byproduct of the simulation without any additional computations, Publication 1 [30]

In order to obtain the equation of state in the NVT ensemble for a particle system, the pressure P must be computed. P is a dependent variable, defined as,

$$\beta P = \frac{\partial \log Z}{\partial V}. \tag{4.5}$$

The pressure is usually retrieved through the virial theorem [146, Sec. 2.2], either by averaging the virial or by integration of the product of the static pair-correlation function $g(r)$ and the pair potential $E_{ij}(r)$.

Direct averaging is not possible for hard-sphere interactions, because the potential is singular, and $g(r)$ has to be computed through a discrete approximation and extrapolated to the contact value to obtain P. Moreover, the dominant contributions to P come from close pairs, $r \approx \sigma$ for a soft power-law potential, even when the potential does not present singularities. These contributions are poorly sampled in the canonical ensemble. Finally, evaluating the virial during the simulation comes with extra computation, as the forces need to be evaluated.

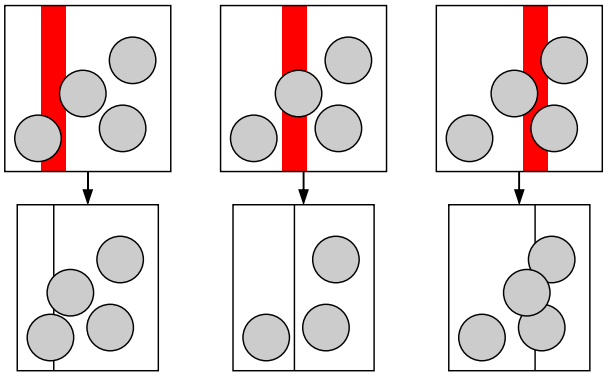


Figure 4.3: Virtual rift volume changes by random removal of an infinitesimal strip from a hard-sphere configuration. **Left:** A successful removal but the bonds are compressed. **Center:** Elimination of a particle (ideal gas pressure). Right: Generation of an overlap (excess pressure). The left and right cases become indistinct for soft interactions.

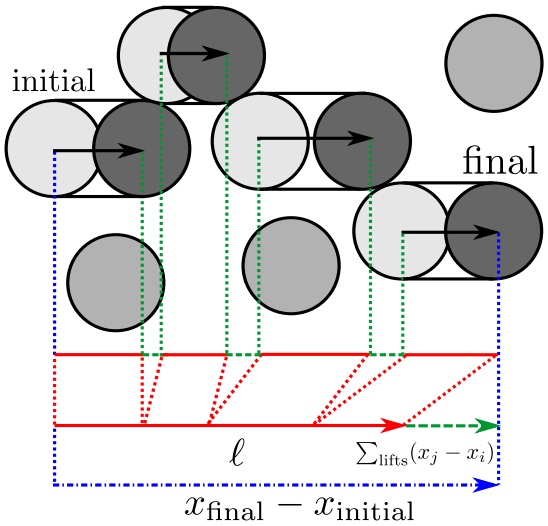


Figure 4.4: Direct computation of the pressure: The excess pressure is derived from the ratio of excess displacement (green dashed lines, $\sum_{\text{lifts}}(x_j - x_i)$) and the chain displacement ℓ (red solid line). For isotropic systems, only the distance between the final and initial particle $x_{\text{final}} - x_{\text{initial}}$ (blue dashed-dotted arrow) has to be recorded.

The situation is quite different for simulations using an event-driven molecular dynamics [43]. The virial pressure is directly related to the collision rate, which is a trivial byproduct of the computation [43],

$$\beta P = \rho - \frac{\beta \rho m}{2 T_{\text{sim}}} \sum_{\text{collisions}} b_{ij}, \tag{4.6}$$

where $\rho = N/V$ the particle number density, T_{sim} the total simulation time, m the mass of a particle and $b_{ij} = (\boldsymbol{r_i} - \boldsymbol{r_j})(\boldsymbol{v_i} - \boldsymbol{v_j})$, with $\boldsymbol{r_{i,j}}$, $\boldsymbol{v_{i,j}}$ the positions and velocities of the colliding particles.

In IFMMC, we access dE_{ij} through the lifting events, whose rates are directly related to P, as the collision rate is. It is then possible to derive, independently from the virial theorem, the pressure P in terms of displacement lengths. We consider a soft-particle system, but the results are also valid for hard particles, using arguments in [147], which connect the pressure to the stochastic geometry of the admissible configurations.

In order to compute Eq. 4.5, we consider virtual rift-volume changes, which allow to retrieve the ratio $Z(V - dV)/Z(V)$. They are produced by removing a randomly located strip of size dL_x x dL_y from the system, as is illustrated in Fig. 4.2. By considering all the possible position of the strips, this procedure leads to sampling all N particles configurations in the smaller box and thus yields the new partition function $Z(V - dV)$.

While going from $Z(V)$ to $Z(V - dV)$, the configurations need to be reweighted as the bonds between particles are compressed, whereas some configurations are illegal. For an isotropic system, the virial expression is then recovered,

$$\beta P = \rho + \frac{1}{V}\left\langle \sum_{\langle i,j \rangle} |x_j - x_i| \beta \frac{\partial E_{ij}(\boldsymbol{r_i} - \boldsymbol{r_j})}{\partial x_i} \right\rangle, \tag{4.7}$$

where $\langle \cdot \rangle$ is the usual canonical average. The first term corresponds to the ideal-gas pressure and comes from the illegal configurations with less than N particles, as some particles were in the removed strips. The second term is the non-ideal contribution to the pressure, as it comes from the reweighting, consequence of the compressed bonds. The probability that a bond is traversed by the removed strip is accounted for by $|x_i - x_j|$ and $\beta \partial E_{ij}/\partial x_i$ is equal at first order to the ratio of the Boltzmann weights $\pi(r'_{ij})/\pi(r_{ij})$. In hard spheres, the compressed bonds are replaced by overlaps. By symmetry, as $dE_{ij} = dE^+_{ij} + -(-dE_{ij})^+ = (dx_i \partial E/\partial x_i)^+ - (dx_j \partial E/\partial x_j)^+$,

$$\begin{aligned} \beta P &= \rho + \frac{1}{V}\left\langle \sum_i \sum_j (x_j - x_i)\beta \left[\frac{\partial E_{ij}}{\partial x_i}\right]^+ \right\rangle \\ &= \rho + \frac{N}{V}\left\langle \sum_j (x_j - x_i)\beta \left[\frac{\partial E_{ij}}{\partial x_i}\right]^+ \right\rangle, \end{aligned} \tag{4.8}$$

as spheres are indistinguishable.

We saw in Eq. 3.49 in Section 3.3.3 that the probability distribution followed by the event configuration is exactly $-\partial\pi/\partial s \; ds = \beta\pi\partial E/\partial s \; ds$, with s the lifting variable. In Eq. 4.8, the second term is actually the average of the distance $x_j - x_i$ on the event configurations related to the lifting variable x. Thus, replacing the ensemble average by the average over the simulation time, Eq. 4.8 can be rewritten as,

$$\beta P = \rho + \frac{N}{V}\frac{1}{S} \sum_{\text{Lifts } (i \to j)} (x_j - x_i). \tag{4.9}$$

with S the global displacement made during the simulation. We can segment the global displacement into the chains of length ℓ,

$$\beta P = \rho + \rho \left\langle \frac{1}{\ell} \sum_{\text{Lifts } (i \to j)} (x_j - x_i) \right\rangle_{\text{chains}}. \tag{4.10}$$

Denoting that $(x_{\text{final}} - x_{\text{initial}}) = \ell + \sum_{\text{Lifts}} (x_j - x_i)$, with x_{initial} the initial position of the first particle of the chain and x_{final} the final position of the last particle of the chain (adjusted for periodic boundary if necessary), see Fig. 4.2, we finally obtain,

$$\beta P = \rho \left\langle \frac{x_{\text{final}} - x_{\text{initial}}}{\ell} \right\rangle_{\text{chains}}. \tag{4.11}$$

Thus, it is sufficient to know the beginning and end configurations of the chains to compute the pressure P directly from the simulation.

In the ideal gas, no lifting move takes place, as there is no interaction. Eq. 4.11 reduces to the ideal gas pressure ρ. The excess displacement $(x_j - x_i)$ is positive for repulsive potentials, whereas it is negative for attractive potentials. For a potential as the Lennard-Jones potential, the excess displacements caused by the decomposed attractive and repulsive parts add up to the correct pressure, see Publication 1 [30].

For anisotropic systems where the collision rates can depend on the direction of the chains, the derivation presented in Fig. 4.2, where strips are removed in a direction normal to the chain direction is supplemented with an analogous result for transverse rifts, where strips are aligned with the chain. It leads in d dimensions to the more general formula,

$$\beta P = \rho + \frac{\rho}{d\ell} \left\langle \sum_{\text{Lifts } (i \to j)} \frac{(\boldsymbol{r}_j - \boldsymbol{r}_i{}^2)}{x_j - x_i} \right\rangle_{\text{chains}}, \tag{4.12}$$

where x is the coordinate parallel to the chain direction. More generally, the full stress tensor $\boldsymbol{u}$ can be obtained as an average of the outer product of the interparticle vector $\boldsymbol{r}_{ij}$,

$$\beta \boldsymbol{u} = -\rho \mathbf{1} - \frac{\rho}{d\ell} \left\langle \sum_{\text{Lifts } (i \to j)} \frac{\boldsymbol{r}_{ij} \boldsymbol{r}_{ij}^T}{x_j - x_i} \right\rangle_{\text{chains}}, \tag{4.13}$$

where $\mathbf{1}$ is the identity matrix in d dimensions.

4.3 Continuous classical spin systems

Spin models are of widespread interest in physics and neighboring sciences, as proved by the recent boom of their implementation, notably for solving inference problems, in neuroscience (neurons interacting through synapses), machine learning (Hopfield network) and economics (speculative agents in a market). In spite of their microscopic simplicity, they exhibit rich macroscopic behavior.

They also constitute one of the major testbeds for algorithms, notably around phase transitions. In order to avoid the slowing down, a wide variety of algorithms, such as the Swendsen–Wang [19] and the Wolff [20] cluster algorithms, and the multicanonical method [148] and the exchange Monte Carlo method [124] based on extended ensembles, have been proposed. All these algorithms are detailed-balance. Recently, algorithms specific to spin systems and breaking the detailed-balance condition but not the global-balance condition have been designed [130, 131, 149] and it has been shown that they improve the sampling efficiency in some cases. In this Section, we review the performance of IFMMC in ferromagnetic as well as glassy systems.

The $O(n)$ model was first introduced by H. E. Stanley [150]. In this model, n-component classical spins $\boldsymbol{S_i}$ are placed on the vertices of a lattice of dimension d. They interact through pairwise interactions E_{kl} with their $2d$ nearest neighbors. The Hamiltonian of the $O(n)$ model is,

$$E = \sum_{\langle k,l \rangle} E_{kl} = -\sum_{\langle k,l \rangle} J_{kl} \boldsymbol{S}_k \cdot \boldsymbol{S}_l - \sum_k \boldsymbol{h}_k \cdot \boldsymbol{S}_k. \tag{4.14}$$

The vector h_k corresponds to an external local magnetic field and is fixed to 0 from this point on. J_{kl} specifies the coupling strength for the pair of spins $\boldsymbol{S_k}, \boldsymbol{S_l}$. If all $J_{kl} = 1$, the system is said to be ferromagnetic. If the couplings $\{J_{kl}\}$ are drawn from a random distribution, the system is called a spin glass and possesses a complex energetic landscape, where numerous metastable states can be found. Important quantities to describe a spin system are the magnetization $\boldsymbol{M}$,

$$\boldsymbol{M} = \sum_k \boldsymbol{S_k}, \tag{4.15}$$

and the susceptibility χ,

$$\chi = \frac{|\sum_k \boldsymbol{S_k}|^2}{N} = \frac{|\boldsymbol{M}|^2}{N}, \tag{4.16}$$

For these systems, the factorized filter is given by,

$$p_{\text{acc}}^{\text{fact}} = \prod_{\langle k,l \rangle} p_{\text{acc}}^{kl} = \prod_{\langle k,l \rangle} \exp(-\beta \Delta E_{kl}^{+}) = \prod_{\langle k,l \rangle} \exp(-\beta \left[-J_{kl} \boldsymbol{S}_k \cdot \boldsymbol{S}_l\right]^{+}). \tag{4.17}$$

In the following, we will discuss the implementation and performance of IFMMC first in Section 4.3.1 in the $O(2)$-model, commonly called the XY model, for ferromagnetic couplings, Section 4.3.1.2, and random couplings, Section 4.3.1.3 and in the $O(3)$ model, usually referred to as the Heisenberg model, Section 4.3.2.

4.3.1 Planar rotator spin systems

The classical continuous two-dimensional spin model, also called classical XY model or classical rotor model, describes spins $\boldsymbol{S}_i$ evolving in two dimensions, which are then characterized by a single angle ϕ_i. In the absence of a local field, the potential described in Eq. 4.14 comes down to,

$$E = -\sum_{\langle k,l \rangle} J_{kl} \cos(\phi_k - \phi_l). \tag{4.18}$$

For nearest-neighbor interactions and for ferromagnetic interactions, the XY model exhibits the interesting physics of the Kosterlitz-Thouless transition [137, 151]. The XY model is commonly used to modelize two-dimensional systems that possess order parameters with the same symmetry, as for superfluid helium or hexatic liquid

crystals. The Kosterlitz-Thouless transition indeed stands out by the absence of symmetry breaking and its unbinding of topological defects (vortex, dislocation, etc.) as the system evolve from the low-temperature ordered phase to the high-temperature disordered phase.

The nature of the XY model transition was still highly controversial [152], in part because of the poor performance of the traditional LMC. It is by the use of the Wolff algorithm (see Section 2.4.3, we will now referred to it as SFC for single flip cluster), that the debate was clarified [113, 114] and the transition temperature $\beta = 1.1199$ determined up to five significant digits [153, 154]. Moreover SFC was decisive in achieving a detailed quantitative description of the XY model [155, 156].

As discussed in Section 2.4.4, SFC and its variants fail to really improve the thermalization in spin glasses, notably because of the large number of metastable states. The problem is particularly acute in the case of the three-dimensional XY spin glass model, where the algorithm loses all its power [115, 116]. For this much studied spin glass model, our understanding today still needs efficient algorithms to clarify the position and nature of the transition. This situation resembles the one of the XY model before the revolution triggered by the cluster algorithms. In this context, we implemented IFMMC for both a ferromagnetic XY model and the XY spin glass model, Publication 2 [31].

While implementing IFMMC, the configuration space is extended by a lifting variable, written as $\overset{\curvearrowright}{k}$. The variable $\overset{\curvearrowright}{k}$ singles out the spin k as the spin being updated along $+\phi$, $\phi_k \to \phi_k + d\phi$. If the move is accepted, the lifting variable for the next time step $t+1$ is again $\overset{\curvearrowright}{k}$. If the physical move is rejected, a lifting move takes place and the lifting variable is passed on to the spin l of the pair that rejected the move, and the physical configuration is unchanged.

In the event-driven approach, the lifting variable being set to $\overset{\curvearrowright}{k}$, the angle ϕ_k now rotates clockwise until the event, i.e. a lifting move, is produced through a rejection by a neighbor l. The lifting variable is updated to $\overset{\curvearrowright}{l}$, ϕ_l rotates clockwise, etc. The event is sampled through the sampling of the positive energy increase for each neighbor l, as explained in Section 3.3.2,

$$\begin{aligned}
\Delta E^*_{kl}(\phi_{\mathrm{l,event}}) &= -[\log \mathbf{ran}(0,1)]/\beta \\
\Delta E^*_{kl}(\phi_{\mathrm{l,event}}) &= \int_{\phi_{\mathrm{k},0}}^{\phi_{\mathrm{l,event}}} \left[\frac{\partial E_{kl}}{\partial \phi_k}\right]^+ d\phi_k \\
\Delta E^*_{kl}(\phi_{\mathrm{l,event}}) &= \int_{\phi_{\mathrm{k},0}}^{\phi_{\mathrm{l,event}}} \left[-J_{kl}\frac{\partial \cos(\phi_k - \phi_l)}{\partial \phi_k}\right]^+ d\phi_k.
\end{aligned} \tag{4.19}$$

To solve Eq. 4.19 and determine $\phi_{\mathrm{k,event}}$, any $2|J_{kl}|$ increment, corresponding to a 2π rotation of ϕ_k, are sliced off $\Delta E^*_{kl}(\phi_{\mathrm{l,event}})$, as the potential is periodic, see Section 3.3.2.2 and Fig. 3.19. After the withdrawal of the n full rotations, it leaves a remaining ΔE^f_{kl},

$$E^*_{\mathrm{init}} + \Delta E^f_{kl} = -|J_{kl}|\cos(\phi_{\mathrm{l,event}} - (\phi_l + \phi_0) - 2n\pi), \tag{4.20}$$

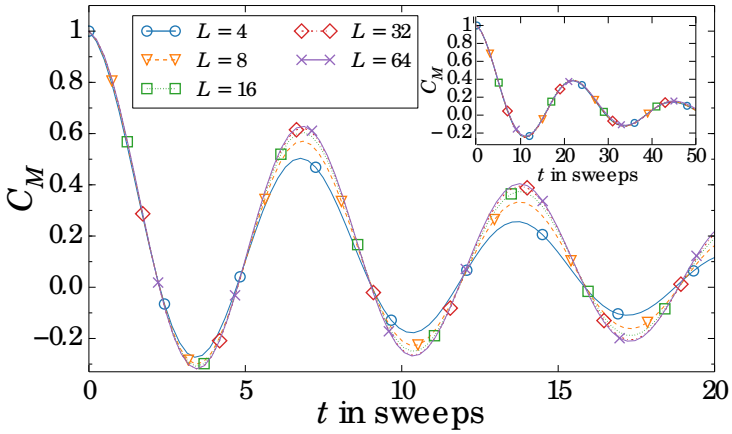


Figure 4.5: Autocorrelation function of magnetization $C_M(t)$ at the critical temperature for various system sizes. The inset shows the spin autocorrelation function of a trivial algorithm that only performs global rotations in spin space along the two axes.

where

$$E^*_{\text{init}} = \begin{cases} E_{kl} & \text{if the initial pair energy derivative is positive} \\ -|J_{kl}| & \text{otherwise} \end{cases}$$
$$\phi_0 = \begin{cases} 0 & \text{if } J_{kl} \geq 0 \\ \pi & \text{otherwise.} \end{cases} \tag{4.21}$$

The event time $\phi_{\text{l,event}}$ is sampled for every neighbor l and the event taking place is the one with the smallest angle.

Spins are always updated only along one direction and there is no need to resample the lifting variable to ensure irreducibility on the other space coordinates. We found that the efficiency was not increased by halting and restarting the simulation after fixed displacements Publication 2 [31]. Observables are still averaged by outputting a configuration after a global displacement ℓ, Section 3.3.3.

4.3.1.1 Choice of a slow variable for a spin system

To evaluate the correlation time and the dynamical critical exponent for IFMMC, one must pay attention to the irreversible nature of the underlying Markov chain. During one chain, spins all rotate in the same sense, and the system undergoes quasi-global rotations. This results in fast oscillations of the magnetization $\boldsymbol{M}$ and a quick decay of its autocorrelation function that is insensitive to the system size (see Fig. 4.5), and even to the temperature. However, this effect is also visible for a trivial algorithm, which simply performs global rotations (see the inset of Fig. 4.5). The trivial algorithm satisfies global balance, but its correlation time is infinite, as it does not relax the energy. A similar effect appears in IFMMC for particle systems [24], that likewise is not characterized by the mean net displacement of particles.

4.3.1.2 Ferromagnetic XY model on a bidimensional lattice

We study the performance of IFMMC in comparison to SFC and LMC for a bidimensional XY system with periodic boundary. We suppose the susceptibility χ of the system to be the slow variable monitoring the convergence dynamics, as explained in Section 2.3. Thermalization is ensured by checking agreement with [153]. This choice is motivated in particular because χ is not sensitive to global rotations, see Section 4.3.1.1. To estimate these convergence properties, we consider at the critical point $\beta = 1.1199$ [153] the susceptibility autocorrelation function,

$$C_\chi(t) = \frac{\langle \chi(t'+t)\chi(t')\rangle - \langle\chi\rangle^2}{\langle\chi^2\rangle - \langle\chi\rangle^2}. \tag{4.22}$$

As discussed in Section 2.4.3, time is measured in sweeps: For IFMMC, one sweep corresponds to $\sim N$ lifting events while for LMC, one sweep corresponds to $\sim N$ attempted moves. For SFC, a sweep denotes $\sim N$ spins added to clusters. The complexity and the CPU times used per sweep are roughly comparable.

Fig. 4.6 shows the autocorrelation function for the XY model at its critical point, obtained from very long single runs of the algorithms. For LMC and SFC, the decay of the susceptibility autocorrelation function can be described by a single time scale, while for IFMMC, it is well described by two time scales:

$$C_\chi(t) \simeq \begin{cases} \exp(-t/\tau^{\mathrm{LMC}}) & \text{(LMC)} \\ \exp(-t/\tau^{\mathrm{SFC}}) & \text{(SFC)} \\ A_0 \exp(-t/\tau_0^{\mathrm{IFMMC}}) + A_1 \exp(-t/\tau_1^{\mathrm{IFMMC}}) & \text{(IFMMC)} \end{cases}. \tag{4.23}$$

IFMMC is indeed first described by a fast time scale, portraying the rapid decay of the correlation function to $C_\chi \sim 0.1$ on a timescale τ_0^{IFMMC} of about 5 sweeps. A slow mode τ_1^{IFMMC} then sets in, presenting a $z \lesssim 2$ scaling ($\tau_1^{\mathrm{IFMMC}} \sim L^z$, with $N = L^z$, Section 2.3.4.2). As shown in Fig. 4.7, τ_1^{IFMMC} is an order of magnitude smaller than τ^{LMC}. Together with the initial rapid decrease, this makes IFMMC about one hundred times faster than LMC. However, its dynamical scaling exponent appears to be $z \lesssim 2$, as for LMC.

In particle systems, IFMMC also shows initial ballistic behavior, but then crosses over into slower decay [133]. A possible explanation for such behavior is that, after a few sweeps of rapid decay, the system undergoes a quasi-global rotation and a diffusive dynamics appears. Implementing the lifting framework is indeed not always bound to lead to a square-root reduction of the mixing time [27, 157] and it may be possible that the emergence of quasi-global rotations effectively induces some reversibility. Still, the speedup obtained is of two orders of magnitude.

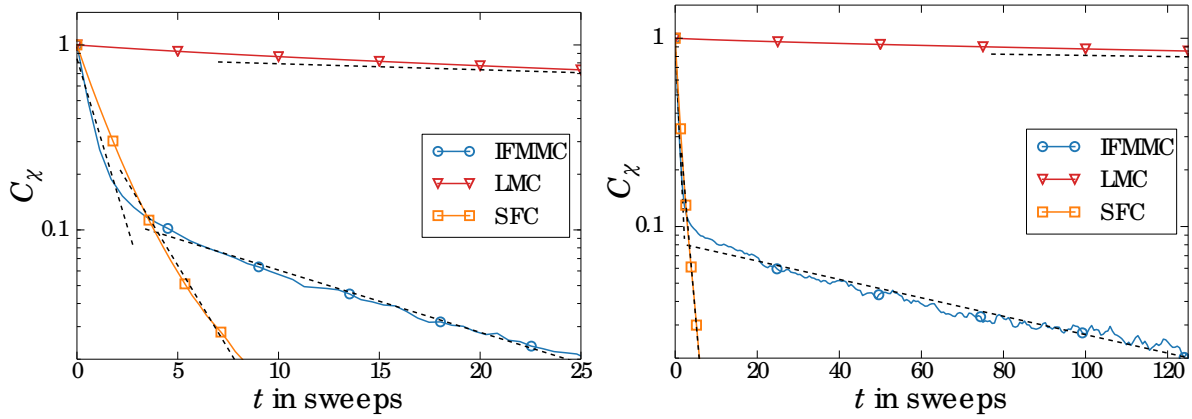


Figure 4.6: Autocorrelation function $C_\chi(t)$ for the two-dimensional XY model at the critical point $\beta = 1.1199$ for LMC (*red, triangle*), IFMMC (*blue, circle*), and SFC (*yellow, square*). Exponential fits (*black, dotted*) are as in Eq. 4.23. **Left:** $N = 32^2$. **Right:** $N = 128^2$. *Right:* Scaling of the autocorrelation time τ with the system size. Both LMC (*red, triangle*) and the slow scale of IFMMC (*dark blue, circle*) are compatible with a dynamical scaling exponent $z \sim 2$, $z_{\mathrm{IFMMC}} \leq z_{\mathrm{LMC}}$, as for soft spheres. Both the fast scale of IFMMC (*light blue, diamond*) and SFC (*yellow, square*) are compatible with $z \sim 0$. *Right Inset:* Speedup of IFMMC with respect to LMC *vs.* L.

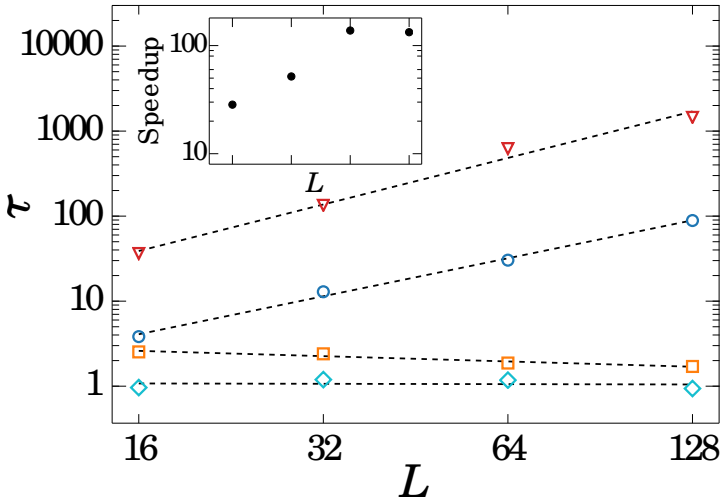


Figure 4.7: Scaling of the autocorrelation time τ with the system size for the two-dimensional XY model at the critical point $\beta = 1.1199$. Both LMC (*red, triangle*) and the slow scale of IFMMC (*dark blue, circle*) are compatible with a dynamical scaling exponent $z \sim 2$. Both the fast scale of IFMMC (*light blue, diamond*) and SFC (*yellow, square*) are compatible with $z \sim 0$. **Inset:** Speedup of IFMMC with respect to LMC *vs.* L.

4.3.1.3 XY spin glass model on a three-dimensional lattice

Spin glasses are key models of statistical mechanics. They possess both strong frustration and randomness, yielding a slow dynamics. The Edwards-Anderson (EA) model of spin glass where spins are located on each site of a lattice and interact via random coupling $\{J_{kl}\}$ was first proposed in [158]. The infinite-range or mean-field version of the EA model was solved thanks to the replica-symmetry breaking [159–161]. However, the behavior of finite range EA spin glass in three dimensions is still not understood. In particular, at low temperature, the phase diagram of the three-dimensional XY spin glass model is still being debated, regarding particularly a possible existence of separate spin-glass and chiral-glass phases [116, 124, 162, 163]. This possible spin-chirality decoupling means that the chirality orders at a temperature higher than the spin, with an intermediate chiral-glass phase, where only the chirality, Eq. 4.25, exhibits a glassy long-range order. We implemented IFMMC for the XY spin glass model on a cubic lattice, where the nearest-neighbor coupling constants J_{kl} are drawn from a Gaussian normal distribution of zero mean and unit variance, Publication 2 [31].

Following [115], we consider the chiral overlap between two independent systems, (1) and (2), with identical coupling constants,

$$p_\kappa = \frac{1}{N}\sum_{p=1}^{N} \kappa^{(1)}_{p\perp\mu}\kappa^{(2)}_{p\perp\mu}, \tag{4.24}$$

with $\kappa^{(i)}_{p\perp\mu}$ being the chirality at a plaquette p, perpendicular to the axis μ, defined as,

$$\kappa^{(i)}_{p\perp\mu} = \frac{1}{2\sqrt{2}}\sum_{(i,j)\in p} \mathrm{sgn}(J_{ij})\sin(\phi_i - \phi_j). \tag{4.25}$$

The sum $\sum_{(i,j)\in p}$ is taken over the four bonds encircling the plaquette p in clockwise order. By construction, p_κ is a symmetric function about zero.

At high temperature, the autocorrelation function of the chiral overlap for IFMMC yields a size independent speedup factor of ~ 5, in comparison to LMC, as shown in the inset in Fig. 4.8.

As the phase diagram of the XY spin glass model is still not agred upon, we consider $\beta = 3.636$, a temperature where the spin glass transition may take place [116] or, otherwise, which is below and near the transition [162, 163]. Without using parallel tempering [124], IFMMC clearly outperforms LMC. In one third of the samples for a system size of $N = 6^3$ and after 10^6 sweeps, IFMMC exhibits indeed a symmetrical chiral overlap, see Fig. 4.9, showing a full configuration space exploration, whereas LMC does not. For larger systems, the speed up of IFMMC in comparison to LMC seems to increase, but already for 10^3 systems, IFMMC no longer equilibrates at $\beta = 3.636$ after 10^6 sweeps.

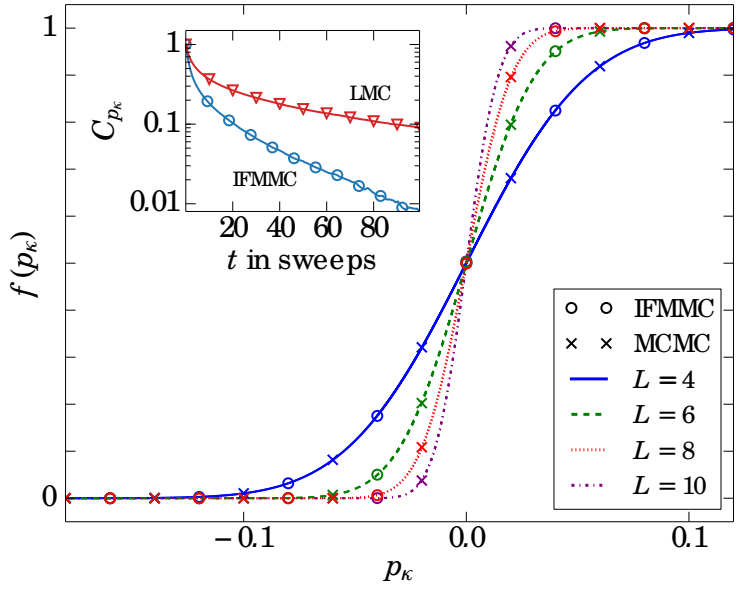


Figure 4.8: Cumulative distribution of the chiral overlap p_κ for the three-dimensional XY spin glass model for $N = 4^3, 6^3, 8^3, 10^3$ at $\beta = 1.5$, in the high-temperature phase (single samples). **Inset:** Autocorrelation function $C_{p_\kappa}(t)$ for $N = 6^3$ from LMC (red, triangle) and IFMMC (blue, circle).

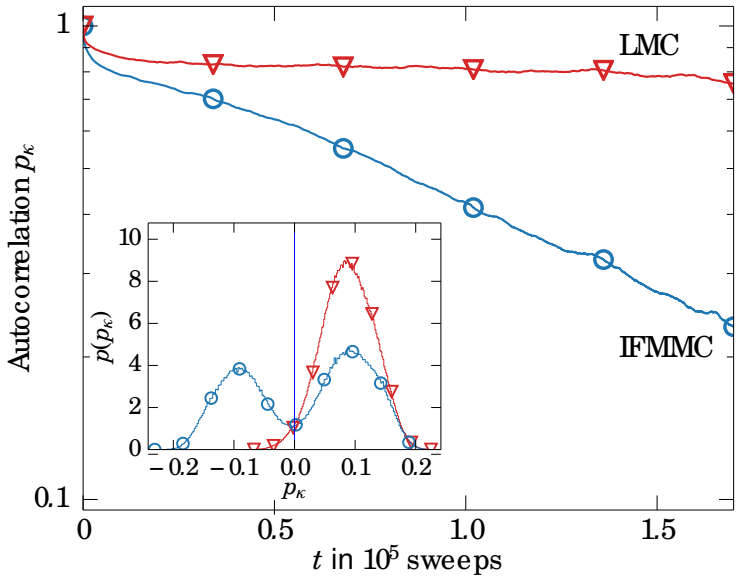


Figure 4.9: Chiral overlap autocorrelation function from IFMMC (blue, circle) and LMC (red, triangle) at $\beta = 3.636$ for a given sample at $N = 6^3$. **Inset:** Distributions of p_κ after 10^6 sweeps for the two algorithms in the same sample. Note the nearly symmetric distribution for IFMMC.

In conclusion, we obtain clear acceleration for the two-dimensional ferromagnetic XY model, as for the three-dimensional XY spin glass model. Unlike cluster methods, IFMMC appears very general. In Section 4.3.2, we discuss a reduction of the critical slowing down obtained for an $O(3)$ spin system.

4.3.2 Heisenberg spin systems on a three-dimensional lattice

Finally we implemented IFMMC for an $O(3)$ spin system, Publication 3 [32]. In a three-dimensional spin model, or commonly called Heisenberg spin model, the spins $\boldsymbol{S}_i$ are described by three components. The lifting variable $(k, \boldsymbol{v})$ needs to determine which spin, k, is rotating but also around which axis, $\boldsymbol{v}$. When a lifting event happens between spin k and spin l, the lifting variable is updated from $(k, \boldsymbol{v})$ to $(l, \boldsymbol{v})$.

In practice, for a fixed rotation axis $\boldsymbol{v}$, IFMMC algorithm for the Heisenberg model reduces to the one of the XY model: With $(\phi_{\boldsymbol{v},k}, \theta_{\boldsymbol{v},k})$ the spherical coordinates of a spin k in a system where the z-axis is aligned with $\boldsymbol{v}$, the pair energy E_{kl} between spins k and l is,

$$E_{kl} = -J' \cos(\phi_{\boldsymbol{v},k} - \phi_{\boldsymbol{v},l}) + K, \tag{4.26}$$

with

$$\begin{aligned} J' &= J \sin\theta_{\boldsymbol{v},k} \sin\theta_{\boldsymbol{v},l}, \\ K &= -J \cos\theta_{\boldsymbol{v},k} \cos\theta_{\boldsymbol{v},l}. \end{aligned}$$

Both J' and K depend only on the polar angle $\theta_{\boldsymbol{v}}$ and both parameters remain unchanged along the chain. The azimuthal-angle dependence in Eq. 4.26 is $\propto \cos(\phi_{\boldsymbol{v},k} - \phi_{\boldsymbol{v},l})$, as in the XY model. The sampling of the next event is then the same as Eq. 4.20, but with a renormalized J'. In order to be irreducible, only two different linearly independent axes are necessary. The chain is stopped after the chain reaches the length ℓ, the last configuration is stored for observable averaging and the lifting variable is resampled.

The dynamical critical exponent of LMC for the Heisenberg model on a three-dimensional lattice was estimated from the autocorrelation function of the magnetization $\boldsymbol{M}$ as $z = 1.96(6)$[164]. The overrelaxation algorithm [165, 166] seems to give $z \simeq 1.10$ [164], which was obtained from the autocorrelation function of the magnetization, and the SFC algorithm is believed to yield a value close to zero: $z \gtrsim 0$, a value obtained from the susceptibility autocorrelation function [167].

To evaluate the dynamical performance of IFMMC at the critical temperature $\beta = 0.693$, we consider the susceptibility χ, as for the XY model in Section 4.3.1, but also the energy density e. Both χ and e are insensitive to global rotations, see Section 4.3.1.1. As shown in Fig. 4.10, the autocorrelation functions both of the energy density and of the susceptibility are well approximated as a single exponential,

$$C_\chi(t) = \exp(-t/\tau), \tag{4.27}$$

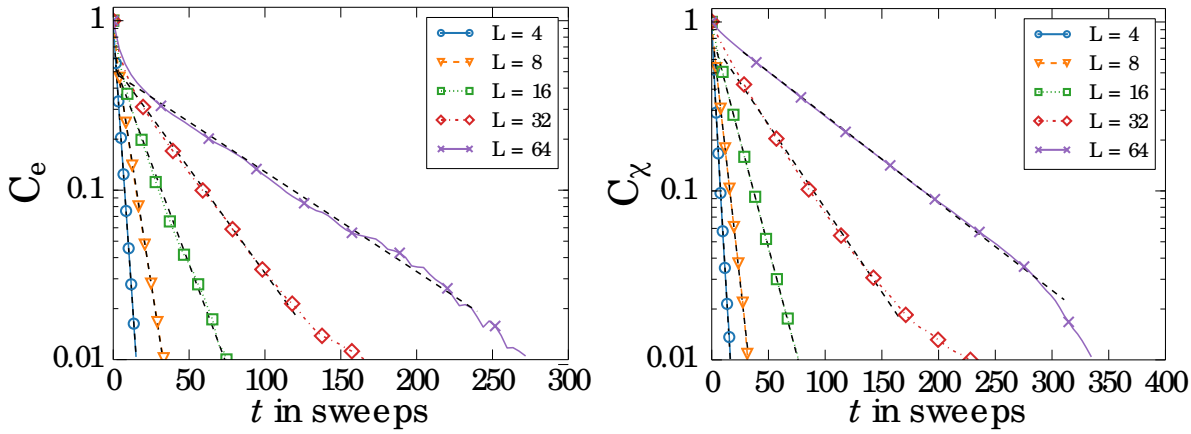


Figure 4.10: Autocorrelation functions and time constants of IFMMC for the three-dimensional Heisenberg model at its critical point $\beta = 0.693$. **Left:** Energy density autocorrelation function C_e for system sizes $4^3, 8^3, \ldots, 64^3$. **Right:** Susceptibility autocorrelation function C_χ for IFMMC for the three-dimensional Heisenberg system sizes $4^3, 8^3, \ldots, 64^3$.

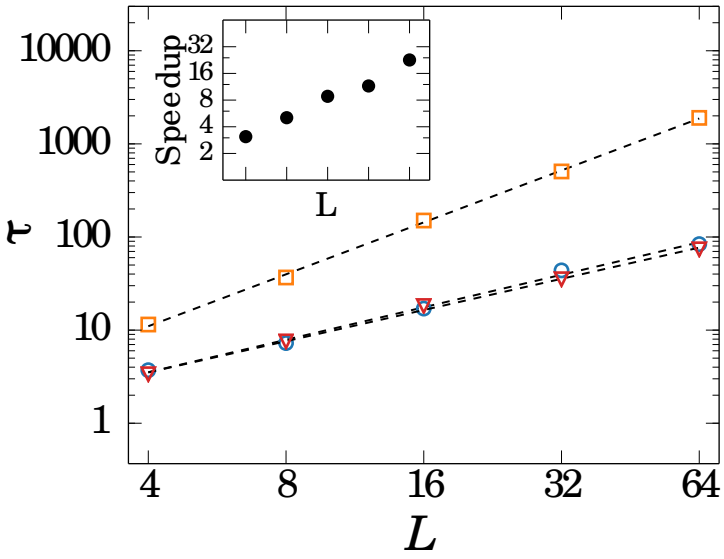


Figure 4.11: Scaling in the three-dimensional Heisenberg model at its critical point $\beta = 0.693$ of the autocorrelation time τ_χ (resp. τ_e) of the susceptibility χ (resp. energy density e) with system size L for IFMMC (*blue circles*) (resp. (*red triangles*)) and of the autocorrelation time of the susceptibility for LMC (*yellow squares*). Error bars are smaller than the markers size. **Inset:** Speedup for the susceptibility χ in comparison to LMC for system sizes $4^3, 8^3, \ldots, 64^3$.

on essentially the same timescales. Furthermore, the finite-size behavior of the autocorrelation times indicates a $z \simeq 1$ dynamical scaling, as exhibited in Fig. 4.11. This z value is significantly less than for LMC and very similar to the one obtained for overrelaxation methods, although the $z \simeq 0$ value of the cluster algorithm is not reached.

The earliest application of lifting [26], the motion of a particle on a one-dimensional n-site lattice with periodic boundary conditions, see Section 3.1.1, already featured the decrease of the dynamical scaling exponent from $z = 2$ to $z = 1$ (the reduction of the mixing time from $\propto n^2$ to $\propto n$). This reduction is made possible by the irreversibility of the Markov chain. It was pointed out that the square-root decrease of the critical exponent was the optimal improvement [27]. A clear reduction of the dynamical exponent z from a quadratic to a linear scaling had not been observed yet in soft-sphere systems and XY spin systems, although speedups appear to increase with the system size, as was shown in hard-sphere systems [74]. The three-dimensional Heisenberg model seems to be a first such IFMMC application with a lowered critical dynamical exponent $z \approx 1$. Our observation relies on the hypotheses that the energy and the susceptibility are indeed slow variables, and that the observed decay of the autocorrelation function continues for larger times. However, in Fig. 4.10, a crossover from $z = 1$ to $z = 2$ as it was observed for XY-model after ~ 5 sweeps in Fig. 4.6, appears unlikely to arise after hundreds of sweeps. The dynamical critical exponent $z \approx 1$ represents a maximal improvement with respect to the $z \approx 2$ of LMC, supposing again that the theorems of [27] apply to infinitesimal Markov chains.

Conclusion

In this Chapter, we compared the performances in terms of autocorrelation times reached by implementing the factorized Metropolis filter in an irreversible lifting scheme. Clear accelerations were obtained in soft spheres, as in the ferromagnetic XY model. Albeit no reduction of the critical slowing down was observed, definite evidence for a speedup that increases with system size was obtained. In the three-dimensional ferromagnetic Heisenberg system, a square-root reduction was shown, rendering the scaling of the autocorrelation time linear with the system size L. As shown by [27], this is the best reduction possible if implementing a lifting scheme. Moreover, the factorized filter allows us to retrieve directly and without any additional cost the pressure P of a particle system, an observable that was formerly very costly to compute.

The good performance of IFMMC in these different systems underlines the generality of the irreversible factorized Metropolis approach, that is already being used with success by the community in three-dimensional hard spheres [168], polymers [134, 169], self-avoiding walks [170], chiral magnets [171] and even in machine learning [172].

General conclusion

Since the beginning of the Markov-chain Monte Carlo method in the 1950s, numerical simulations have become a fundamental tool of statistical mechanics and beyond, acting like a real telescope of the mind. However fast mixing algorithms are still needed, notably for the study of particles and glassy spin systems. The seminal Metropolis algorithm [17] with local updates is limited in such systems and a wide variety of algorithms has been designed to counteract the slowing down linked to phase transitions.

In ferromagnetic spin systems, cluster algorithms [19, 20] suppress the slowing down but, even if they can be generalized, their efficiency is only observed in specific systems. The situation repeated itself in particle systems, where algorithms updating in one step clusters of particles were created [125, 128, 129]. They also accelerate the thermalization, but such improvement is also restricted to some systems.

During my thesis, I built on the lifting framework in mathematics [26, 27] and on previous attempts to set up irreversible Markov chains algorithms in physics, [24, 25], to design a general procedure for implementing an irreversible scheme out of a reversible ones [30]. My main theoretical result is the factorized Metropolis filter, which allows to treat any multidimensional potential as a collection of unidimensional ones. As it replaces the Metropolis filter at the most fundamental level of the method, the factorized Metropolis filter is simple to use and easy to generalize to any distribution.

The irreversible factorized scheme exhibits a clear acceleration in soft-sphere systems and bidimensional XY spin systems, and, most of all, a square-root reduction of the critical slowing down in Heisenberg spin systems [30–32]. As its performances are similar in a broad class of systems [134, 168–172], the irreversible factorized scheme appears to be more general than the highly optimized algorithms that only work in very specific circumstances.

For decades, stochastic calculations were seen as practical challenges, but whose theoretical grounds were set in stone. Although the irreversible factorized paradigm may offer a new starting point, some questions are still open. One of the main issues is

to push the effort in understanding the mixing process in lifting for continuous space. The goal is to understand why, in the ferromagnetic XY model, we observed the two-time-scale evolution, while in the ferromagnetic Heisenberg model, there is only one scale, that shows an important reduction of the critical slowing down. As discussed in [27] for chains on a discrete space, the square-root reduction of the mixing time is the best possible. Finally, an interesting challenge lies in the possibility to implement a maximal global-balance scheme for interdimensional lifting moves, where only a detailed-balance scheme is for now proven correct.

Bibliography

[1] P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Reviews of Modern Physics*, 49:435, 1977.

[2] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 17(8):549–560, 1905.

[3] J. C. Maxwell. Illustrations of the dynamical theory of gases. Part I. On the motions and collisions of perfectly elastic spheres. *Philosophical Magazine*, 4th series(19):19–32, 1860.

[4] J. C. Maxwell. Illustrations of the dynamical theory of gases. Part II. On the process of diffusion of two or more kinds of moving particles among one another. *Philosophical Magazine*, 4th series(20):21–37, 1860.

[5] L. Boltzmann. Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. *Wiener Berichte*, 58:517–560, 1868.

[6] L. Boltzmann. Einige allgemeine Sätze über Wärmegleichgewicht. *Wiener Berichte*, 63(2):679–711, 1871.

[7] J. von Neumann and S. Ulam. Random ergodic theorems. *Bulletin of the American Mathematical Society*, 51:660, 1945.

[8] S. Ulam, R. D. Richtmyer, and J. von Neumann. Statistical Methods in Neutron Diffusion. *Los Alamos Scientific Laboratory Report*, LAMS-551, 1947. This reference contains the von Neumann letter discussed in the present manuscript.

[9] N. Metropolis and S. Ulam. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

[10] D. P. Landau and K. Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, Cambridge, 2000.

[11] D. Frenkel and B. Smit. *Understanding Molecular Simulation - From Algorithms to Applications*. Academic Press, San Diego, 1996.

[12] W. Krauth. *Statistical Mechanics Algorithms and Computations*. Oxford University Press, 2006.

[13] E. P. Bernard and W. Krauth. Two-Step Melting in Two Dimensions: First-Order Liquid-Hexatic Transition. *Phys. Rev. Lett.*, 107:155704, 2011.

[14] D. M. Ceperley. Path integrals in the theory of condensed helium. *Reviews of Modern Physics*, 67(2):279–355, 1995.

[15] J. S. Coggan, T. M. Bartol, E. Esquenazi, J. R. Stiles, S. Lamont, M. E. Martone, D. K. Berg, M. H. Ellisman, and T. J. Sejnowski. Evidence for Ectopic Neurotransmission at a Neuronal Synapse. *Science*, 309(5733):446–451, 2005.

[16] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for Machine Learning. *Machine Learning*, 50:5–43, 2003.

[17] N. Metropolis, A. W. Rosenbluth, M. N Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[18] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[19] R. H. Swendsen and J.-S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86, 1987.

[20] U. Wolff. Collective Monte Carlo Updating for Spin Systems. *Physical Review Letters*, 62(4):361–364, 1989.

[21] P. Gustafson. A guided walk Metropolis algorithm. *Statistics and Computing*, 8(4):357–364, 1998.

[22] S. Duane, A. D. Kennedy, Pendleton B. J., and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.

[23] R. M. Neal. Suppressing Random Walks in Markov Chain Monte Carlo Using Ordered Overrelaxation. *Technical Report No. 9508, Departments of Statistics, University of Toronto*, 1995.

[24] E. P. Bernard, W. Krauth, and D. B. Wilson. Event-chain Monte Carlo algorithms for hard-sphere systems. *Physical Review E*, 80:056704, 2009.

[25] E. A. J. F. Peters and G. de With. Rejection-free Monte-Carlo sampling for general potentials. *Physical Review E*, 85:026703, 2012.

[26] P. Diaconis, S. Holmes, and R. M. Neal. Analysis of a nonreversible Markov chain sampler. *Annals of Applied Probability*, 10(3):726, 2000.

[27] F. Chen, L. Lovász, and I. Pak. Lifting Markov chains to speed up mixing. *Proc. 31st Annual ACM Symp. on Theory of Computing*, pages 275–281, 1999.

[28] H. C. Andersen and P. Diaconis. Hit and run as a unifying device. *Journal de la société française de statistique*, 148(4):5–28, 2007.

[29] R. P. Feynman. *Statistical Mechanics: A Set of Lectures*. W. A. Benjamin, 1972.

[30] M. Michel, S. C. Kapfer, and W. Krauth. Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps. *Journal of Chemical Physics*, 140:054116, 2014.

[31] M. Michel, J. Mayer, and W. Krauth. Event-chain Monte Carlo for classical continuous spin models. *EPL*, 112(2):20003, 2015.

[32] Nishikawa Y., M. Michel, W. Krauth, and K. Hukushima. Event-chain algorithm for the Heisenberg model: Evidence for z$\approx$ 1 dynamic scaling. *Physical Review E*, 92:063306, 2015.

[33] M. Michel and W. Krauth. Infinite Event Chain Monte Carlo. *Manuscript in preparation*.

[34] M. Michel, S. Kapfer, and W. Krauth. Elasticity of N-particle systems in Event Chain Monte Carlo. *Manuscript in preparation*.

[35] D. Bernoulli. *Hydrodynamica, sive De viribus et motibus fluidorum commentarii*. sumptibus Johannis Reinholdi Dulseckeri, 1738.

[36] J. Perrin. Mouvement brownien et grandeurs moléculaires. *Annales de Chimie et de Physique*, 18(8):5–114, 1909.

[37] R. Clausius. Ueber die Art der Bewegung, welche wir Wärme nennen. *Annalen der Physik*, 176(3):353–379, 1857.

[38] R. Kubo. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29:255–284, 1966.

[39] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer-Verlag Berlin Heidelberg, second edition edition, 1989.

[40] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier North Holland, third edition edition, 2007.

[41] R. H. Swendsen. Statistical mechanics of colloids and Boltzmann's definition of the entropy. *American Journal of Physics*, 74:187, 2006.

[42] D. Frenkel. Why colloidal systems can be described by statistical mechanics: some not very original comments on the Gibbs paradox. *Molecular Physics*, 112(17):2325–2329, 2014.

[43] B. J. Alder and T. E. Wainwright. Studies in Molecular Dynamics. II. Behavior of a Small Number of Elastic Sphere. *Journal of Chemical Physics*, 33:1439, 1960.

[44] J. W. Gibbs. *Elementary Principles in Statistical Mechanics*. Charles Scribner's Sons, 1902.

[45] A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, volume 2. Springer Verlag, Berlin, 1933.

[46] A. N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Publishing, New York, 1950.

[47] L. E. Reichl. *A modern course in statistical physics*. Wiley, 2nd edition edition, 1998.

[48] D. Aldous and J. A. Fill. *Reversible Markov Chains and Random Walks on Graphs*. 2002. Unfinished monograph, recompiled 2014.

[49] J. R. Norris. *Markov Chains*. Cambridge University Press, 1998.

[50] Y. Sinai. Dynamical systems with elastic reflections. *Russian Mathematical Surveys*, 25:137, 1970.

[51] N. Simànyi. Proof of the Boltzmann-Sinai ergodic hypothesis for typical hard disk systems. *Inventiones Mathematicae*, 154:123, 2003.

[52] E. Fermi, J. R. Pasta, and S. Ulam. *Studies of Nonlinear Problems*. Document Los Alamos 1940, 1955.

[53] A. Kolmogorov. On conservation of conditionally periodic motions for a small change in Hamilton's function. *Dokl. Akad. Nauk SSSR*, 98:527–530, 1954.

[54] V. Arnold. Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian. *Uspehi Mat. Nauk*, 18(5(113)):13–40, 1963.

[55] J. Möser. On invariant curves of area-preserving mappings of an annulus. *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II*, pages 1–20, 1962.

[56] J. L. Lebowitz and O. Penrose. Modern ergodic theory. *Physics Today*, 26(2):23–29, 1973.

[57] V. I. Arnold and A. Avez. *Ergodic Problems of Classical Mechanics*. New York: Benjamin, 1968.

[58] J. von Neumann. Zur Operatorenmethode In Der Klassischen Mechanik. *Annals of Mathematics*, 33:587, 1932.

[59] E. Hopf. On Causality, Statistics and Probability. *J. Math, and Phys.*, 13:51, 1934.

[60] E. Hopf. *Ergoden Theorie*. Springer Berlin, 1937.

[61] H. Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta mathematica*, 13, 1890.

[62] P. Diaconis. The Markov chain Monte Carlo revolution. *Bulletin of the American Mathematical Society*, 46(2):179–205, 2008.

[63] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. Providence, R.I. American Mathematical Society, 2009.

[64] D. Bayer and P. Diaconis. Trailing the Dovetail Shuffle to its Lair. *The Annals of Applied Probability*, 2(2):294–313, 1992.

[65] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences of the United States of America*, 93(4):1659–1664, 1996.

[66] A. D. Sokal. *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*. Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne, 1989.

[67] A. D. Sokal and L. E. Thomas. Exponential convergence to equilibrium for a class of random-walk models. *Journal of Statistical Physics*, 54(3-4):797–828, 1989.

[68] E. Nummelin. *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press, 1984.

[69] Buffon. *Histoire de l'Académie Royale des Sciences*. Imprimerie royale (Paris), 1733.

[70] G.-L. Leclerc comte de Buffon. *Histoire Naturelle, Générale et Particulière : supplément. Tome Quatrième.* Imprimerie royale (Paris), 1777.

[71] N. Metropolis. The Beginning of the Monte Carlo Method. *Los Alamos Science*, 15:125, 1987.

[72] R. Eckhardt. Stan Ulam, John von Neumann and the Monte Carlo Method. *Los Alamos Science, Special issue*, 15:131–137, 1987.

[73] J. von Neumann. Various techniques used in connection with random digits. In A. S. Householder, G. E. Forsythe, and H. H. Germond, editors, *Monte Carlo Method*, pages 36–38. National Bureau of Standards Applied Mathematics Series, 12, 1951.

[74] M. Engel, J. A. Anderson, S. C. Glotzer, M. Isobe, E. P. Bernard, and W. Krauth. Hard-disk equation of state: First-order liquid-hexatic transition in two dimensions with three simulation methods. *Physical Review E*, 87:042134, 2013.

[75] S. Martiniani, K. J. Schrenk, Stevenson J. D., D. J. Wales, and D. Frenkel. Turning intractable counting into sampling: Computing the configurational entropy of three-dimensional jammed packings. *Physical Review E*, 93:012906, 2016.

[76] S. F. Edwards and R. B. S. Oakeshott. Theory of powders. *Physica A*, 157(3):1080–1090, 1989.

[77] J. A. Barker. A quantum-statistical Monte Carlo method; path integrals with boundary conditions. *Journal of Chemical Physics*, 70(6):2914–2911, 1970.

[78] K. Van Houcke, F. Werner, E. Kozik, N. Prokof'ev, Svistunov B., M. J. H. Ku, A. T. Sommer, L. W. Cheuk, A. Schirotzek, and M. W. Zwierlein. Feynman diagrams versus Fermi-gas Feynman emulator. *Nature Physics*, 8:366–370, 2012.

[79] J. Jasche and F. S. Kitaura. Fast Hamiltonian sampling for large-scale structure inference. *Monthly Notices of the Royal Astronomical Society*, 407(1):29–42, 2010.

[80] D. T. Gillespie. A General Method of Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[81] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[82] P. Boyle, M. Broadie, and P. Glasserman. Monte Carlo Methods for Security Pricing. *Journal of Economic Dynamics & Control*, 21(8-9):1267–1321, 1997.

[83] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2004.

[84] T. Gerstner and P. E. Kloeden, editors. *Recent Developments in Computational Finance: Foundations, Algorithms and Applications*. Singapore:World Scientific, 2013.

[85] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[86] A. E. Gelfand and A. F. M. Smith. Sample Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

[87] L. Tierney. Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 2(4):1701–1728, 1994.

[88] J. Dongarra and S. Sullivan. Guest Editors Introduction to the top 10 algorithms. *Comput. Sci. Eng.*, 2(1):22–23, 2000.

[89] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics, 2004.

[90] N. Mantel. An Extension of the Buffon Needle Problem. *The Annals of Mathematical Statistics*, 24(4):674–677, 1953.

[91] H. Solomon. *Geometric Probability*, volume 28. SIAM, 1978.

[92] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

[93] P. Lévy. Sur certains processus stochastiques homogènes. *Composition Mathematica*, 7:283–339, 1940.

[94] R. E. Bellman. *Dynamic programming*. Princeton Universtiy Press, 1957.

[95] W. W. Wood and F. R. Parker. Monte Carlo Equation of State of Molecules Interacting with the Lennard-Jones Potential. *The Journal of Chemical Physics*, 27:720–733, 1957.

[96] A. A. Barker. Monte Carlo Calculations of the Radial Distribution Functions for a Proton-Electron Plasma. *Australian Journal of Physics*, 18(2):119–134, 1965.

[97] W. Janke. Statistical Analysis of Simulations: Data Correlations an d Error Estimation. In J. Grotendorst, D. Marx, and A. Muramatsu, editors, *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms - Lecture Notes*, volume 10, pages 423–445. NIC Series, 2002.

[98] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random structures and Algorithms*, 9(1-2):223–252, 1996.

[99] E. P. Bernard, C. Chanal, and W. Krauth. Damage spreading and coupling in Markov chains. *EPL*, 92(6):60004, 2010.

[100] K. Binder. Theory of first-order phase transitions. *Reports on Progress in Physics*, 50(7):783, 1987.

[101] W. Janke. Multicanonical Monte Carlo Simulations. *Physica A*, 254:164, 1998.

[102] L. van Hove. Correlations in Space and Time and Born Approximation Scattering in Systems of Interacting Particles. *Physical Review*, 95:249, 1954.

[103] L. van Hove. Time-Dependent Correlations between Spins and Neutron Scattering in Ferromagnetic Crystals. *Physical Review*, 95:1374, 1954.

[104] L. Onsager. Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition. *Physical Review*, 65:117, 1944.

[105] M. P. Nightingale and H. W. Blöte. Dynamic Exponent of the Two-Dimensional Ising Model and Monte Carlo Computation of the Subdominant Eigenvalue of the Stochastic Matrix. *Physical Review Letters*, 76(24):4548, 1996.

[106] M. P. Nightingale and H. W. Blöte. Monte Carlo computation of correlation times of independent relaxation modes at criticality. *Physical Review B*, 62(2):1089, 2000.

[107] N. Ito, K. Hukushima, K. Ogawa, and Y. Ozeki. Nonequilibrium Relaxation of Fluctuations of Physical Quantities. *Journal of the Physical Society of Japan*, 69:1931–1934, 2000.

[108] U. Wolff. Comparison between cluster Monte Carlo algorithms in the Ising model. *Physics Letters B*, 228(3):379–382, 1989.

[109] D. Heermann and A. Burkitt. System size dependence of the autocorrelation time for the Swendsen-Wang Ising model. *Physica A*, 162(2):210–214, 1990.

[110] P. W. Kasteleyn and C. M. Fortuin. Phase transitions in lattice systems with random local properties. *Journal of the Physical Society of Japan Supplement*, 26s:11–14, 1969.

[111] P. W. Kasteleyn and C. M. Fortuin. On the random-cluster model : I. Introduction and relation to other models. *Physica*, 57(4):536–564, 1972.

[112] R. G. Edwards and A. Sokal. Dynamic critical behavior of Wolff's collective-mode Monte Carlo algorithm for the two-dimensional $O(n)$ nonlinear σ model. *Physical Review D*, 40(4):1374–1377, 1989.

[113] W. Janke and K. Nather. High-precision Monte Carlo study of the two-dimensional XY Villain model. *Physical Review B*, 48:7419, 1993.

[114] M. Hasenbusch, M. Marcu, and K. Pinn. High precision renormalization group study of the roughening transition. *Physica A*, 208:124, 1994.

[115] H. Kawamura and M. S. Li. Nature of the Ordering in the Three-Dimensional XY. *Physical Review Letters*, 87:187204, 2001.

[116] T. Obuchi and H. Kawamura. Monte Carlo simulations of the three-dimensional XY spin glass focusing on chiral and spin order. *Physical Review B*, 87:174438, 2013.

[117] S. Liang. Application of cluster algorithms to spin glasses. *Physical Review Letters*, 69(14):2145–2148, 1992.

[118] V. Cataudella, G. Franzese, M. Nicodemi, and A. Coniglio. Critical clusters and efficient dynamics for frustrated spin models. *Physical Review Letters*, 72(10):1541–1544, 1994.

[119] F. Matsubara, A. Sato, O. Koseki, and T. Shirakura. Cluster Heat Bath Algorithm in Monte Carlo Simulations of Ising Models. *Physical Review Letters*, 78:3237–3240, 1997.

[120] J. Houdayer. A cluster Monte Carlo algorithm for 2-dimensional spin glasses. *The European Physical Journal B*, 22(4):479–484, 2001.

[121] T. Jörg. Cluster Monte Carlo Algorithms for Diluted Spin Glasses. *Progress of Theoretical Physics Supplements*, 157:349–352, 2005.

[122] Z. Zhu, A. J. Ochoa, and H. G. Katzgraber. Efficient Cluster Algorithm for Spin Glasses in Any Space Dimension. *Physical Review Letter*, 115:077201, 2015.

[123] H. G. Katzgraber, M. Palassini, and A. P. Young. Monte Carlo simulations of spin glasses at low temperatures. *Physical Review B*, 63:184422, 2001.

[124] K. Hukushima and K. Nemoto. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.*, 65:1604, 1996.

[125] C. Dress and W. Krauth. Cluster algorithm for hard spheres and related systems. *Journal of Physics A: Mathematical and General*, 28(23):L597–L601, 1995.

[126] A. Buhot and W. Krauth. Numerical Solution of Hard-Core Mixtures. *Physical Review Letter*, 80(17):3787–3790, 1998.

[127] L. Santen and W. Krauth. Absence of thermodynamic phase transition in a model glass former. *Nature*, 405:550–551, 2000.

[128] J. Liu and E. Luitjen. Rejection-Free Geometric Cluster Algorithm for Complex Fluids. *Physical Review Letter*, 92(3):035504, 2004.

[129] A. Jaster. An improved Metropolis algorithm for hard core systems. *Physica A*, 264(1-2):134–141, 1999.

[130] K. S. Turitsyn, M. Chertkova, and M. Vucelja. Irreversible Monte Carlo algorithms for efficient sampling. *Physica D*, 240(4-5):410–414, 2011.

[131] H. C. M. Fernandes and M. Weigel. Non-reversible Monte Carlo simulations of spin models. *Computer Physics Communications*, 182(9):1856–1859, 2011.

[132] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz. A new algorithm for Monte Carlo simulation of Ising spin systems. *Journal of Computational Physics*, 17(1):10–18, 1975.

[133] S. C. Kapfer and W. Krauth. Sampling from a polytope and hard-disk Monte Carlo. *Journal of Physics: Conference Series*, 454:012031, 2013.

[134] T. A. Kampmann, H. H. Boltz, and J. Kierfeld. Parallelized event chain algorithm for dense hard sphere and polymer systems. *Journal of Computational Physics*, 281:864, 2015.

[135] N. D. Mermin and H. Wagner. Absence of Ferromagnetism or Antiferromagnetism in One- or Two-Dimensional Isotropic Heisenberg Models. *Physical Review Letters*, 17:1307, 1966.

[136] J. M. Kosterlitz and D. J. Thouless. Long Range Order and Metastability in Two–Dimensional Solids and Superfluids. *J. Phys. C*, 5:L124, 1972.

[137] J. M. Kosterlitz and D. J. Thouless. Ordering, Metastability and Phase Transitions in Two–Dimensional Systems. *J. Phys. C*, 6:1181, 1973.

[138] B. I. Halperin and D. R. Nelson. Theory of Two-Dimensional Melting. *Physical Review Letters*, 41(Theory of Two-Dimensional Melting):121, 1978.

[139] D. R. Nelson and B. I. Halperin. Dislocation-mediated melting in two dimensions. *Physical Review B*, 19:2457, 1979.

[140] A. P. Young. Melting and the vector Coulomb gas in two dimensions. *Physical Review B*, 19:1855, 1979.

[141] K. Zahn, J. M. Méndez-Alcaraz, and G. Maret. Hydrodynamic Interactions May Enhance the Self-Diffusion of Colloidal Particles. *Physical Review Letters*, 79:175, 1997.

[142] K. J. Strandburg. Two-dimensional melting. *Reviews of Modern Physics*, 60:161, 1988.

[143] W. Mickel, S. C. Kapfer, G. E. Schröder-Turk, and K. Mecke. Shortcomings of the bond orientational order parameters for the analysis of disordered particulate matte. *Journal of Chemical Physics*, 138:044501, 2013.

[144] B Delaunay. Sur la sphère vide. À la mémoire de Georges Voronoï. *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Es- testvennykh Nauk*, 6:793–800, 1934.

[145] S. C. Kapfer and W. Krauth. Two-Dimensional Melting: From Liquid-Hexatic Coexistence to Continuous Transitions. *Physical Review Letters*, 114:035702, 2015.

[146] Hansen. J. P. and I. R. McDonald. *Theory of Simple Liquids*. Academic, London, 3rd edition, 2006.

[147] R. J. Speedy. Pressure of hard-sphere systems. *The Journal of Physical Chemistry*, 92(7):2016–2018, 1988.

[148] B. A. Berg and T. Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Physical Review Letter*, 68:9, 1992.

[149] H. Suwa and S. Todo. Markov Chain Monte Carlo Method without Detailed Balance. *Phys. Rev. Lett*, 105:120603, 2010.

[150] H. E. Stanley. Dependence of Critical Properties upon Dimensionality of Spins. *Physical Review Letters*, 20:589–592, 1968.

[151] J. M. Kosterlitz. The critical properties of the two-dimensional XY model. *J. Phys. C*, 7:1046, 1974.

[152] E. Seiler, I. O. Stamatescu, A. Patrascioiu, and V. Linke. Critical behaviour, scaling and universality in some two-dimensional spin models. *Nucl. Phys. B*, 305:623, 1988.

[153] M. Hasenbusch. The two dimensional XY model at the transition temperature: A high precision Monte Carlo study. *Journal of Physics A: Mathematical and General*, 38:5869–5883, 2005.

[154] Y. Komura and Y. Okabe. Large-scale Monte Carlo simulation of two-dimensional classical XY model using multiple GPUs. *J. Phys. Soc. Jpn.*, 81:113001, 2012.

[155] J. V. José, L. P. Kadanoff, S. Kirkpatrick, and D. R. Nelson. Renormalization, vortices, and symmetry-breaking perturbations in the two-dimensional planar model. *Physical Review B*, 16:1217, 1977.

[156] D. J. Amit, Y. Y. Goldschmidt, and G. Grinstein. Renormalisation group analysis of the phase transition in the 2D Coulomb gas, Sine-Gordon theory and XY-model. *J. Phys. A: Math. Gen*, 13(2):585, 1980.

[157] M. Hildebrand. Rates of Convergence of the Diaconis-Holmes-Neal Markov Chain Sampler with a V-shaped Stationary Probability. *Markov Processes and Related Fields*, 10:687–704, 2004.

[158] S. F. Edwards and P. W. Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5:965, 1975.

[159] G. Parisi. Infinite Number of Order Parameters for Spin-Glasses. *Physical Review Letters*, 43:1754, 1979.

[160] G. Parisi. A sequence of approximated solutions to the S-K model for spin glasses. *Journal of Physics A: Mathematical and General*, 13:L115, 1980.

[161] G. Parisi. The order parameter for spin glasses: a function on the interval 0-1. *Journal of Physics A: Mathematical and General*, 13:1101, 1980.

[162] J. H. Pixley and A. P. Young. Large-scale Monte Carlo simulations of the three-dimensional XY spin glass. *Physical Review B*, 78:014419, 2008.

[163] F. Romá and D. Domínguez. Critical behavior of spin and chiral degrees of freedom in three-dimensional disordered XY models studied by the nonequilibrium aging method. *Physical Review B*, 89:024408, 2014.

[164] P. Peczak and D. P. Landau. Dynamical critical behavior of the three-dimensional Heisenberg model. *Physical Review B*, 47:14260, 1993.

[165] M. Creutz. Overrelaxation and Monte Carlo simulation. *Physical Review D*, 36:515, 1987.

[166] F. R. Brown and T. J. Woch. Overrelaxed heat-bath and Metropolis algorithms for accelerating pure gauge Monte Carlo calculation. *Physical Review Letters*, 58:2394, 1987.

[167] C. Holm and W. Janke. Critical exponents of the classical three-dimensional Heisenberg model: A single-cluster Monte Carlo study. *Physical Review B*, 48:936, 1993.

[168] M. Isobe and W. Krauth. Hard-sphere melting and crystallization with event-chain Monte Carlo. *J. Chem. Phys.*, 143, 2015.

[169] T. A. Kampmann, H.-H. Boltz, and J. Kierfeld. Monte Carlo simulation of dense polymer melts using event chain algorithms. *Journal of Chemical Physics*, 143:044105, 2015.

[170] H. Hu, X. Chen, and Y. Deng. Irreversible Markov chain Monte Carlo algorithm for the self-avoiding walk. *arXiv:1602.01671*, 2016.

[171] Y. Nishikawa and K. Hukushima. Phase transitions and ordering structures of a model of a chiral helimagnet in three dimensions. *Physical Review B*, 94:064428, 2016.

[172] A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. The Bouncy Particle Sampler: A Non-Reversible Rejection Free Markov chain Monte Carlo Method. *arXiv:1510.02451*, 2016.

Publications

Publication 1: Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps

Manon Michel, Sebastian C. Kapfer and Werner Krauth, *J. Chem. Phys.*, **140**, 054116 (2014).

This article presents how to construct rejection-free and maximal global-balance Markov-chain Monte Carlo algorithms by implementing the factorized Metropolis filter with lifted Markov chains. The article describes the additional possibility offered by the factorized Metropolis filter, as the direct computation of pressure or the decomposition of a non monotonic potential into monotonic parts. Finally, it analyzes the performance of the irreversible factorized Metropolis algorithm, generalization of the event-chain algorithm, in bidimensional soft spheres.

THE JOURNAL OF CHEMICAL PHYSICS **140**, 054116 (2014)

Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps

Manon Michel,[a] Sebastian C. Kapfer,[b] and Werner Krauth[c]
Laboratoire de Physique Statistique, Ecole Normale Supérieure, UPMC, Université Paris Diderot, CNRS, 24 rue Lhomond, 75005 Paris, France

(Received 7 October 2013; accepted 16 January 2014; published online 7 February 2014)

In this article, we present an event-driven algorithm that generalizes the recent hard-sphere event-chain Monte Carlo method without introducing discretizations in time or in space. A factorization of the Metropolis filter and the concept of infinitesimal Monte Carlo moves are used to design a rejection-free Markov-chain Monte Carlo algorithm for particle systems with arbitrary pairwise interactions. The algorithm breaks detailed balance, but satisfies maximal global balance and performs better than the classic, local Metropolis algorithm in large systems. The new algorithm generates a continuum of samples of the stationary probability density. This allows us to compute the pressure and stress tensor as a byproduct of the simulation without any additional computations. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4863991]

I. INTRODUCTION

Markov-chain Monte Carlo (MCMC) methods in statistical physics have progressed far from the original local-move, detailed-balance Metropolis algorithm.[3] On the one hand, intricate non-local cluster moves have met with great success in lattice models.[4,5] To a lesser extent, continuum systems of hard spheres have in recent years also benefitted from non-local moves,[6–9] building on earlier work.[10–12] On the other hand, extensions of the classic detailed balance condition have allowed to construct Markov chains that converge faster. These algorithms introduce persistence between subsequent moves and reduce the diffusive nature of the Markov chain on small and intermediate length and time scales. Notable examples are guided random walks,[13] hybrid Monte Carlo,[14,15] and overrelaxation.[16] The Markov chain lifting framework[17–19] unifies these concepts by augmenting the physical configuration space with auxiliary variables that resemble the momentum in Newtonian time evolution and in molecular dynamics (MD).[20] Lifted Markov chains have already been applied to spin models,[23,24] but not to continuum systems.

The present article draws on the above lines of research. As a main theoretical result, we introduce a factorized version of the Metropolis filter (acceptance rule) that is well suited for the simulation of N-particle systems with pair-potential interactions. Combined with the concept of infinitesimal Monte Carlo moves, this filter allows us to construct a rejection-free event-chain Monte Carlo (ECMC) algorithm that breaks detailed balance yet satisfies maximal global balance. This algorithm builds upon a recent insightful hybrid Monte Carlo scheme.[15] By virtue of infinitesimal displacements of particles, our algorithm produces a continuum of configurations that all sample the equilibrium distribution. Samples are obtained efficiently using an event-driven algorithm. For hard spheres, the events correspond to hard-sphere collisions, and the new algorithm reduces to the hard-sphere event-chain algorithm.[6] For general pair interactions, our algorithm replaces the hard-sphere collisions by pairwise collisions, whose collision distance is resampled after each event from the pair potential. Finally, the continuum of ECMC samples permits to directly compute the pressure and the stress tensor in the NVT ensemble at no extra computational cost.

II. BALANCE CONDITIONS, FACTORIZED METROPOLIS FILTER

For an MCMC algorithm to converge to the stationary distribution, it must satisfy the global balance condition for the stationary flows $\varphi_{a\to b}$ from configuration a to b: the total flow into a configuration a must equal the total flow out of it,

$$\sum_b \varphi_{b\to a} = \sum_c \varphi_{a\to c} = \pi(a), \tag{1}$$

where $\pi(a)$ is the statistical weight of configuration a, e. g., given by a Boltzmann factor. The flow must also satisfy an ergodicity requirement.[21] The global balance, Eq. (1), is in particular satisfied by the *detailed balance* condition which equates the flows between any two configurations a and b:

$$\varphi_{a\to b} = \varphi_{b\to a} \tag{2}$$

(see Fig. 1). We will be concerned with algorithms satisfying *maximal global balance*, where flow between two configurations is unidirectional and flows from a to a (that is, rejections) are avoided: if $\varphi_{a\to b} > 0$, then $\varphi_{b\to a} = 0$. In this case, probability does flow back nonlocally from b to a. In the particle systems that we consider, this happens *via* the periodic boundary conditions.

MCMC methods commonly rely on the Metropolis algorithm, which enforces detailed balance of the flows between

[a]Electronic mail: manon.michel@ens.fr
[b]Electronic mail: sebastian.kapfer@ens.fr
[c]Electronic mail: werner.krauth@ens.fr

 © 2014 AIP Publishing LLC

Reuse of AIP Publishing content is subject to the terms: https://publishing.aip.org/authors/rights-and-permissions. Downloaded to IP: 129.199.121.207 On: Mon, 08 Feb 2016 19:51:24

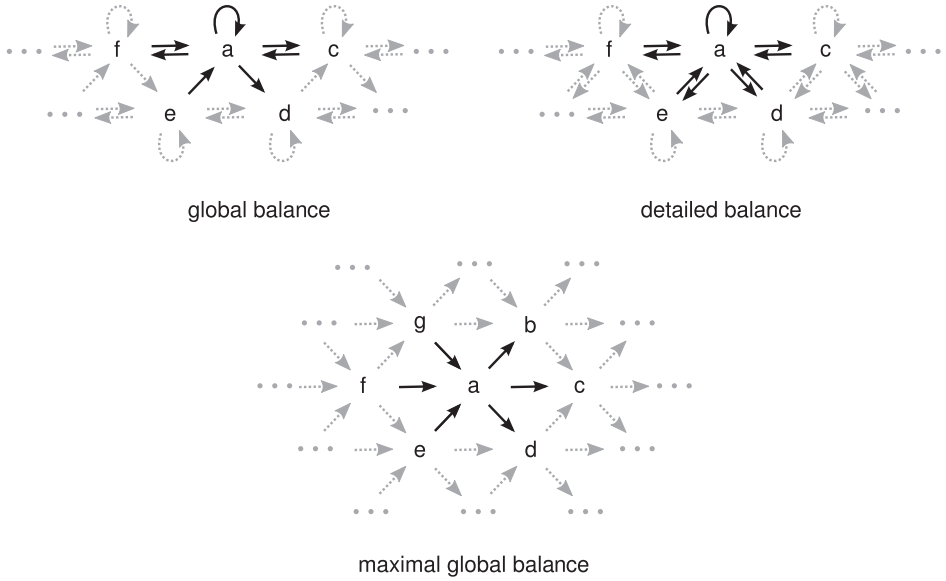


FIG. 1. Balance conditions for probability flow in Markov-chain Monte Carlo. Arrows represent stationary flows of equal magnitude. Top left: Global balance, a necessary condition for the convergence towards equilibrium. The total flow $\sum_c \varphi_{c\to a}$ into any configuration a must equal the total flow $\sum_c \varphi_{a\to c}$ out of it. The loops $\varphi_{a\to a}$, etc., correspond to rejected moves. Top right: Detailed balance, the net flow between any two configurations is zero, $\varphi_{a\to b} = \varphi_{b\to a}$. Bottom: Maximal global balance, $\varphi_{a\to b} > 0$ implies $\varphi_{b\to a} = 0$, the flow $\varphi_{a\to a}$ vanishes.

a and b as follows:

$$\varphi_{a\to b} = \mathcal{A}_{a\rightleftarrows b} \min(\pi(a), \pi(b)). \tag{3}$$

In our algorithm, the *a priori* probability $\mathcal{A}$ is symmetric and amounts to zero or a global constant that we drop for simplicity. Equation (3) is manifestly symmetric in $\pi(a)$ and $\pi(b)$ so that, by construction, $\varphi_{a\to b} = \varphi_{b\to a}$. Since $\varphi_{a\to b} = \pi(a)p(a \to b)$, with p the acceptance probability, Eq. (3) is equivalent to the well-known Metropolis filter,

$$p(a \to b) = \min\left(1, \frac{\pi(b)}{\pi(a)}\right), \tag{4}$$

that has been implemented in countless computer programs.

In statistical physics, the weight of a configuration a is often given by the Boltzmann factor $\pi(a) = \exp(-\beta E(a))$, where $E(a)$ is the energy of a and β is the inverse temperature, which we set to one for the majority of this article. Using the abbreviation

$$[x]^+ := \max(0, x) \quad (\geq 0), \tag{5}$$

we can write the Metropolis filter of Eq. (4) as

$$p(a \to b) = \min(1, \exp(-\Delta E)) = \exp(-[\Delta E]^+), \tag{6}$$

where $\Delta E = E(b) - E(a)$. This corresponds to the acceptance probability of a proposed move, whereas the rejection probability is $1 - p = 1 - \exp(-[\Delta E]^+)$.

We now consider an N-particle system with pair interactions $E = \sum_{\{i,j\}} E_{ij}$, where i and j, in our applications, label particles in D-dimensional space, but could also refer to spins or other degrees of freedom. The sum runs over all unordered pairs $\{i, j\}$ of particles. For such a system, the Metropolis filter has always been used as

$$p^{\text{Met}}(a \to b) = \min\Big(1, \exp\Big(-\sum_{\{i,j\}} \Delta E_{ij}\Big)\Big)$$
$$= \exp\Big(-\Big[\sum_{\{i,j\}} \Delta E_{ij}\Big]^+\Big). \tag{7}$$

In the present article, however, we introduce a *factorized Metropolis filter*,

$$p^{\text{fact}}(a \to b) = \prod_{\{i,j\}} \min(1, \exp(-\Delta E_{ij}))$$
$$= \exp\Big(-\sum_{\{i,j\}} [\Delta E_{ij}]^+\Big), \tag{8}$$

which also fulfills detailed balance by respecting the same flow symmetries as the standard Metropolis filter, as can be seen by applying the identity $\pi(a)/\pi(b) = \exp(\sum_{\{i,j\}} [\Delta E_{ij}]^+ - [-\Delta E_{ij}]^+)$ to Eq. (8). The conventional and the factorized Metropolis filter agree in the hard-sphere case[1] and (trivially) for $N = 2$. They differ whenever terms of opposite sign appear on the rhs of Eq. (8), i.e., for general interactions and $N > 2$. The factorization increases the rate of rejections in a detailed-balance MC algorithm. We find that for the soft-sphere interactions considered in this article, the rate of rejections is about 50% higher (soft spheres with $n = 12$, $\rho = 0.8 \ldots 1.2$, with a step size of 0.1 in units of the particle diameter). However, the factorization yields the acceptance probability as a product of independent pair interaction terms. This will be the key to derive a rejection-free lifted MCMC algorithm for general N-particle systems.

Reuse of AIP Publishing content is subject to the terms: https://publishing.aip.org/authors/rights-and-permissions. Downloaded to IP: 129.199.121.207 On: Mon, 08 Feb 2016 19:51:24

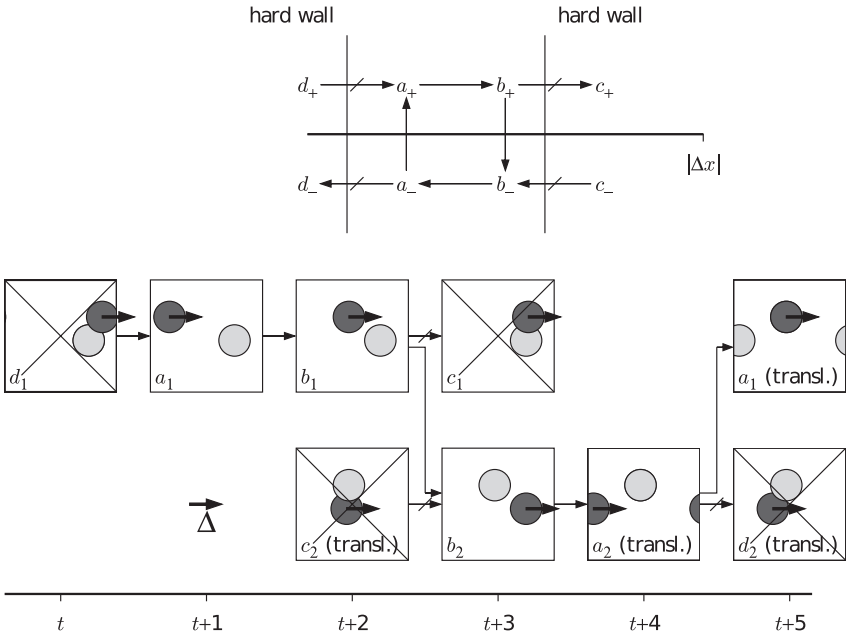


FIG. 2. Upper: Discrete one-dimensional system ($L = 2$) with constant probabilities $\pi(a) = \pi(b)$ and hard-wall conditions $\pi(c) = \pi(d) = 0$. The lifting variable $h = \pm 1$ corresponds to the direction of motion of the particle. Lower: Equivalent lifting algorithm for two hard spheres with finite displacement $\boldsymbol{\Delta}$ (the values of the lifting variable ± 1 are replaced by $\{1, 2\}$). The forbidden particle moves $b_1 \to c_1$ and $a_2 \to d_2$ trigger lifting moves. Maximal global balance is satisfied by moving in the $+x$ direction only. The equivalence between one-dimensional motion and two-particle dynamics in a constrained direction carries over to arbitrary pair potentials.

III. LIFTING: 1D SYSTEMS AND TWO PARTICLES IN A BOX

We now introduce the concept of Markov chain lifting in a simple setting, which we later generalize to interacting particle systems. We consider a one-dimensional discrete system with configurations a and stationary weights $\pi(a)$ (i.e., $a \in \{\Delta, 2\Delta, \ldots, L\Delta\}$). For moves sampled uniformly from $\{-\Delta, \Delta\}$, the standard Metropolis filter of Eq. (4) satisfies detailed balance $\varphi_{a\to a+\Delta} = \varphi_{a+\Delta\to a}$ $\forall a$. The stationary distribution π is sampled in the limit of infinite running times.

Lifting,[17–19] in this example, consists in duplicating each configuration a with a momentum-like variable into two configurations $a_\pm = (a, h = \pm 1)$. The lifting variable determines the next proposed move, which would in ordinary Metropolis MC be sampled from a prior distribution: For a_+, only the particle move $a \to a + \Delta$ is proposed, and for a_-, only $a \to a - \Delta$. For flow balance, we introduce *lifting moves* $a_+ \to a_-$ and $a_- \to a_+$ which take effect if the particle move is rejected, as summarized in the diagram,

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & (a-\Delta)_+ & \xrightarrow{\varphi_0} & a_+ & \xrightarrow{\varphi_1} & (a+\Delta)_+ & \longrightarrow & \cdots \\
 & & \uparrow\downarrow & & \uparrow\downarrow & & \uparrow\downarrow & & \\
 & & {\scriptstyle [\varphi_1-\varphi_0]^+} & & {\scriptstyle [\varphi_0-\varphi_1]^+} & & & & \\
\cdots & \longleftarrow & (a-\Delta)_- & \xleftarrow{\varphi_0} & a_- & \xleftarrow{\varphi_1} & (a+\Delta)_- & \longleftarrow & \cdots
\end{array} \tag{9}$$

where the flows φ_0 and φ_1 are given by

$$\varphi_0 = \min[\pi(a), \pi(a - \Delta)], \tag{10}$$

$$\varphi_1 = \min[\pi(a), \pi(a + \Delta)]. \tag{11}$$

We take the weights of the lifted configurations to be the same as the weights of the original configurations, $\pi(a_\pm) = \pi(a)$, adjusting the constant of normalization. This rejection-free MCMC algorithm satisfies maximal global balance, as only one of the two flows $\varphi_{a_+\to a_-}$ and $\varphi_{a_-\to a_+}$ can be non-zero. In the physical variables a, however, rejections are still present.

We now consider uniform stationary probabilities ($\pi(a) =$ constant) and impose hard-wall boundary conditions in our one-dimensional discrete model. The lifting flows are non-zero only for a rightmoving particle at $a = L\Delta$, and for a leftmoving particle at $a = \Delta$. For these configurations, the lifting flow equals the entire incoming flow, and the particle reverses its direction. One can show that the lifted algorithm visits all sites in $\mathcal{O}(L)$ steps, rather than in $\mathcal{O}(L^2)$ steps for the Metropolis algorithm.[2,22]

As demonstrated in Fig. 2, the discrete one-dimensional system with hard walls corresponds to two D-dimensional hard spheres that are constrained to move only, say, along the x direction, in a box with periodic boundary conditions. The new lifting variable i now indicates the moving sphere, and the hard wall turns into a no-overlap condition for the two spheres. Although the spheres only move towards the right, the algorithm satisfies maximal global balance due to flows across the periodic boundaries. Ergodicity for the unconstrained two-sphere problem in a D-dimensional box is achieved by resampling, after a fixed number of steps, the moving particle $i \in \{1, 2\}$ and the direction of motion $\boldsymbol{\Delta} \in \{+\boldsymbol{e}_x\Delta, +\boldsymbol{e}_y\Delta\}$ (for the example of two hard disks in a periodic box). The sequence of moves between resampling is referred to in the following as an *event chain*.

Reuse of AIP Publishing content is subject to the terms: https://publishing.aip.org/authors/rights-and-permissions. Downloaded to IP: 129.199.121.207 On: Mon, 08 Feb 2016 19:51:24

IV. INFINITESIMAL MOVES AND EVENT-CHAIN ALGORITHM FOR INTERACTING MANYPARTICLE SYSTEMS

We now extend the discussion of Sec. III to N-particle systems, first for hard spheres, and then for particles with arbitrary pairwise potentials. The idea to indicate the moving particle and its "momentum" by lifting variables generalizes trivially to the N-particle case. Special care is, however, required to preserve global balance, and we show that the factorized Metropolis filter can be used to achieve maximal global balance in the infinitesimal-move limit. We then implement this scheme efficiently in an event-driven MCMC algorithm.

Lifted configurations are now specified by the N hard-sphere centers $(\boldsymbol{r}_1, \dots, \boldsymbol{r}_N)$, the moving sphere i, and its direction of motion $\boldsymbol{\Delta}$. For concreteness, we focus on the positive x direction, $\boldsymbol{\Delta} = +\boldsymbol{e}_x\Delta$, as before. A particle move is

$$a_i = (\boldsymbol{r}_1, \dots, \boldsymbol{r}_i, \dots, \boldsymbol{r}_N) \to b_i = (\boldsymbol{r}_1, \dots, \boldsymbol{r}_i + \boldsymbol{\Delta}, \dots, \boldsymbol{r}_N). \tag{12}$$

This algorithm violates global balance because it generates configurations with multiple overlaps, see Fig. 3. In the presence of a multiple overlap, it is impossible to define flows that satisfy the global balance condition. Multiple overlaps vanish, and maximal global balance is recovered, for infinitesimal moves $|\boldsymbol{\Delta}| \to 0$: In that limit, the factorized Metropolis filter identifies a unique collision partner, with probability one, since no two particles are at the same distance from the moving particle. The collision partner then inherits the lifting variable and moves forward in the next step. By a succession of infinitesimal steps that add up to a finite chain displacement ℓ, this reproduces the hard-sphere event-chain algorithm.[6] Of course, the infinitesimal-move algorithm is not implemented naively through a fine discretization, but rather by identifying the next lifting event, and then advancing the moving disk to contact.

We now generalize the infinitesimal-move event-chain algorithm to arbitrary pair potentials, using the factorized Metropolis filter, Eq. (8). For general interactions, the energy change between configurations a_i and b_i which differ by an infinitesimal displacement $\mathrm{d}x_i$ of particle i is

$$\mathrm{d}E = E(b) - E(a) = \sum_{j(\neq i)} \frac{\partial E_{ij}(\boldsymbol{r}_j - \boldsymbol{r}_i)}{\partial x_i} \mathrm{d}x_i = \sum_{j(\neq i)} \mathrm{d}E_{ij}, \tag{13}$$

where $\mathrm{d}E_{ij}$ is the pairwise energy change, and E_{ij} the pair potential. According to the factorized Metropolis filter, the move is rejected with probability

$$1 - p^{\mathrm{fact}}(a_i \to b_i) = 1 - \exp\Big(-\sum_{j(\neq i)} [\mathrm{d}E_{ij}]^+\Big) = \sum_{j(\neq i)} [\mathrm{d}E_{ij}]^+. \tag{14}$$

Remarkably, for infinitesimal displacements, the rejection probability is a sum of pair terms, while the individual terms $[\mathrm{d}E_{ij}]^+$ normally neither add up to the total energy change $\mathrm{d}E$ nor to $[\mathrm{d}E]^+$. We use the terms in Eq. (14) as the probabilities for lifting moves

$$p^{\mathrm{lift}}(a_i \to a_j) = [\mathrm{d}E_{ij}]^+ \quad \forall j \neq i, \tag{15}$$

and obtain a rejection-free, maximal global balance MCMC algorithm. Fig. 4 illustrates that the total flow into the configuration a_i equals the total flow out of configuration a_i, satisfying the global balance condition Eq. (1). Explicitly, the lifting flows in the example of Fig. 4 are

$$\varphi_{a_2 \to a_1} = \pi(a)[\mathrm{d}E_{21}]^+,$$
$$\varphi_{a_3 \to a_1} = \pi(a)[\mathrm{d}E_{31}]^+,$$

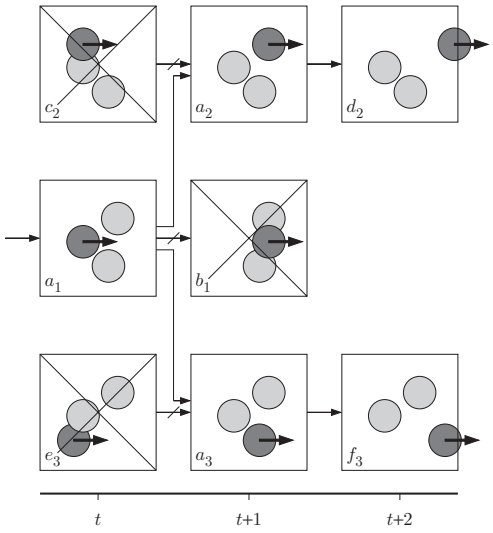


FIG. 3. Multiple overlaps for hard disks (weight π = constant for the non-overlapping physical configurations a, d, f). The violation of the global balance condition is caused by the multiple overlap in configuration b (overlap of disk 1 with both disks 2 and 3): the flow into all legal configurations a, d, and f must be equal, while the illegal (crossed-out) configurations b, c, and e generate zero flow. The multiple overlap disappears, and global balance is again satisfied, in the limit $|\boldsymbol{\Delta}| \to 0$.

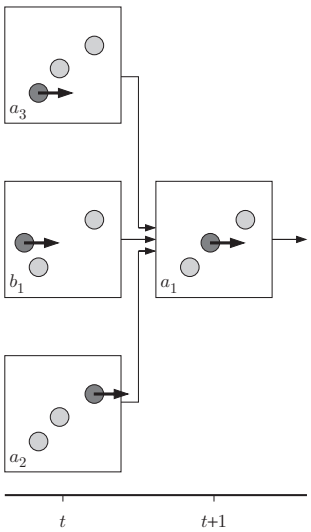


FIG. 4. Maximal global balance for N particles with arbitrary pair interactions (infinitesimal step, factorized Metropolis filter of Eq. (8)). Flow into a_1 is due to $N-1$ lifting moves (here, for $N=3$, $a_2 \to a_1$ and $a_3 \to a_1$) and to a particle move (here, $b_1 \to a_1$). For infinitesimal steps, the flow into a_1 equals $\pi(a)$ (see text), thus balancing the flow $\pi(a)$ out of a_1 and satisfying global balance, Eq. (1).

Reuse of AIP Publishing content is subject to the terms: https://publishing.aip.org/authors/rights-and-permissions. Downloaded to IP: 129.199.121.207 On: Mon, 08 Feb 2016 19:51:24

and the particle move flow:

$$\varphi_{b_1\to a_1} = \varphi^{\text{fact}}_{a\to b} = \pi(a)(1 - [\mathrm{d}E_{21}]^+ - [\mathrm{d}E_{31}]^+).$$

Indeed, Eqs. (14) and (15) define a rejection-free *infinitesimal* MCMC algorithm with maximal global balance.

In order to implement this algorithm efficiently, we choose an event-based approach. As mentioned, the factorized Metropolis filter ensures that no two lifting events can occur in the same infinitesimal timestep. Thus, every interaction of the moving particle with another can be treated independently of other interactions. Further, for fixed partner j, the lifting probabilities at successive timesteps are independent, and they vanish if the pair potential decreases during displacement. Following the Bortz-Kalos-Lebowitz (BKL) algorithm,[15,25] we determine the displacement until the first lifting move occurs by sampling a uniform random number Υ_{ij} from (0, 1] which determines the admissible energy increase until lifting, $E^*_{ij} = -\ln \Upsilon_{ij}$. The displacement until lifting s_{ij} is then found from

$$E^*_{ij} = \int_0^{E^*_{ij}} [\mathrm{d}E_{ij}]^+ = \int_0^{s_{ij}} \left[\frac{\partial E_{ij}(\boldsymbol{r}_j - \boldsymbol{r}_i - s\boldsymbol{e}_x)}{\partial s}\right]^+ \mathrm{d}s. \tag{16}$$

If this equation lacks a solution due to the shape of the interaction potential, or due to a large thermal excitation E^*_{ij}, no lifting event is generated, $s_{ij} = \infty$. While solving Eq. (16) can be nontrivial in general, we give a fast method for the most usual pair potentials below. The smallest of the $N - 1$ independent $\{s_{ij}\}_{j\neq i}$ determines the lifting move $i \to j^*$ which will actually take place, advancing the moving disk by $\min_{j\neq i}(s_{ij}) = s_{ij^*}$ in the $+\boldsymbol{e}_x$ direction and changing the lifting variable to j^*. We also make sure that the total displacement in a single event chain equals the chain displacement ℓ, which usually requires truncating the final event. After the end of the chain, the lifting variables are resampled. Each chain thus consists in an infinite succession of infinitesimal moves that add up to the chain displacement ℓ. Alternatively, we could introduce a small constant probability for terminating a chain in each infinitesimal move. This would effectively lead to exponentially distributed random ℓ, and is also a valid MCMC algorithm.

We conclude with some practical remarks on solving Eq. (16) for model potentials that occur in practice. Many model potentials are central potentials, $E_{ij}(\boldsymbol{r}_{ij}) = E_{ij}(r_{ij})$. If the pair potential consists of several terms, e.g., attractive and repulsive terms, it may be convenient to treat them separately by a further factorization of the Metropolis filter, and decompose the lifting probabilities, $[\mathrm{d}E^{\text{attr}}_{ij}]^+ + [\mathrm{d}E^{\text{rep}}_{ij}]^+$, where E^{attr} and E^{rep} are the attractive and repulsive parts of the pair potential. This decomposition can lead, however, to a higher event rate than the full potential. For instance, the Lennard-Jones potential reduces to two soft-sphere interactions, one attractive and one repulsive, and the mean free path between events is reduced at most by half in comparison to the case without decomposition. In return, the decomposition greatly simplifies the Monte Carlo program.

We will thus focus here on the case of soft-sphere potentials which are monotonous. A lifting move can only be generated if the moving particle is in the rising part of the pair potential. In this case, solving Eq. (16) amounts to sampling the energy increase $E^*_{ij} = -\log \Upsilon_{ij}$, with Υ_{ij} a uniform random number from (0, 1], which fixes the interaction energy $E^{\text{lift}}_{ij} = E_{ij}(r_{ij}) + E^*_{ij}$, and thus the interparticle distance $r^{\text{lift}}_{ij} = E^{-1}_{ij}(E^{\text{lift}}_{ij})$ at the lifting move. (If E^{lift} exceeds any possible value of the interaction potential, there is no lifting move generated.) The admissible displacement s_{ij} for the i, j particle pair is then the positive root of $r^{\text{lift}}_{ij} = |\boldsymbol{r}_j - \boldsymbol{r}_i - s_{ij}\boldsymbol{e}_x|$. Again, if no such root exists, no lifting move is generated, and particle i will pass particle j. Using this method, and the decomposition into attractive and repulsive terms, Eq. (16) is thus easily computable for a large range of potentials.

In systems with periodic boundary conditions, for very long chains, a particle can pass by the same collision partner more than once: The pair potential no longer is monotonous. This is most easily avoided by tuning the chain displacement so that the moving sphere can only interact with one periodic image of the other spheres or by introducing a lifting move of the moving sphere with itself after a displacement of half the box. This move does not change the statistics of the following events. After it, the next event is computed as usual.

V. SPEEDUP WITH RESPECT TO METROPOLIS MONTE CARLO

We now compare the performance of the generalized ECMC algorithm with Metropolis Monte Carlo (MMC). As an application, we consider a two-dimensional system of N particles interacting with a truncated pairwise power-law potential, $E_{ij}(r) = \tilde{E}(\min(r, r_c))$, with $\tilde{E}(r) = \epsilon(\sigma/r)^n$ and $r_c = 1.8\sigma$, σ being the particle diameter. This potential includes important physical interactions such as the dipole interaction in magnetic colloids,[26] hard disks ($n \to \infty$) and Lennard-Jones particles, once decomposed into a repulsive soft-sphere interaction ($n = 12$) and an attractive one ($n = 6$). In comparison to MMC, the ECMC algorithm uses more random numbers, one per interaction term, whereas Metropolis uses one per step. In our implementations, however, the main computational workload is the evaluation of the potentials, not the generation of random numbers (using the Mersenne Twister). One event of the ECMC algorithm is thus implemented in the same amount of time as one attempted step of MMC (3.2×10^9 steps/h in MMC, 1.5×10^9 events/h in ECMC). We compare the performance of the algorithms in terms of the central processing unit (CPU) time used (see Fig. 5 for details).

As estimate of the relative performance of the algorithms we consider as in other recent work[6,8,9] the autocorrelation time τ_6 of the global orientational order parameter Ψ_6,

$$\Psi_6 = \frac{1}{N} \cdot \sum_{i,j} \frac{A_{ij}}{A_i} \exp(6\mathrm{i}\theta_{ij}), \tag{17}$$

where θ_{ij} is the angle of the bond vector between particles i and j against a fixed axis and A_{ij}/A_i the contribution of particle j to particle i's Voronoi cell perimeter.[27,28] With Ψ_6 being a global observable, we assume that its autocorrelation time τ_6 is representative for the mixing time for dense liquid states close to the freezing point, located at $\rho\sigma^2 = 1$ for $n = 12$, and

Reuse of AIP Publishing content is subject to the terms: https://publishing.aip.org/authors/rights-and-permissions. Downloaded to IP: 129.199.121.207 On: Mon, 08 Feb 2016 19:51:24

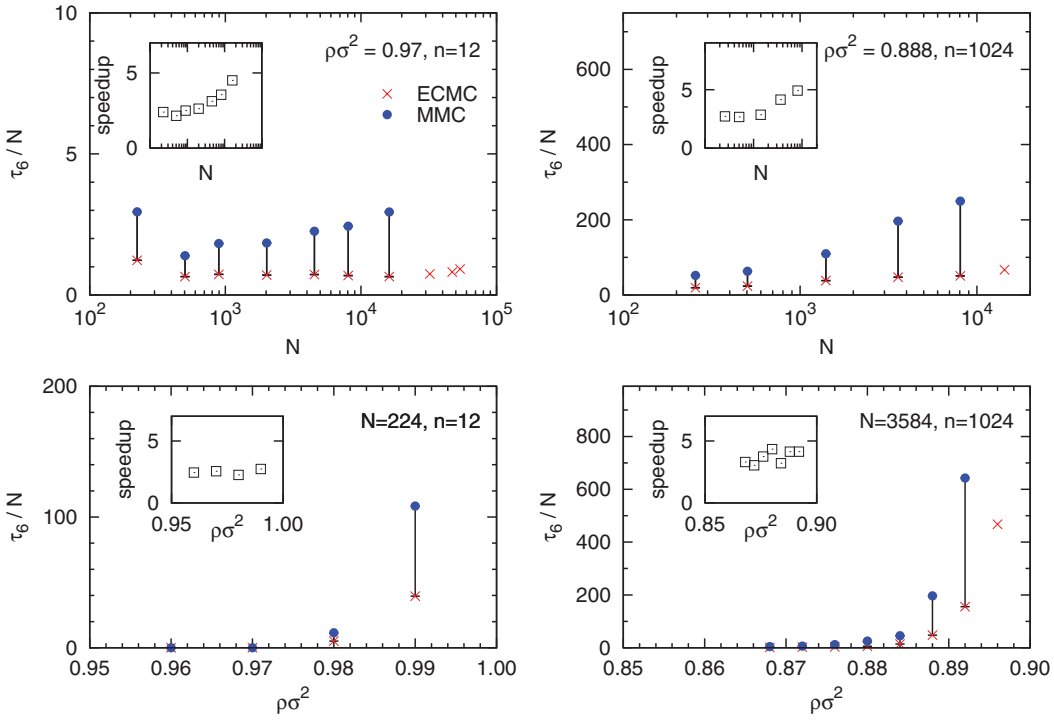


FIG. 5. Autocorrelation times τ_6 of the Ψ_6 orientational order parameter for event-chain Monte Carlo (ECMC) and Metropolis Monte Carlo (MMC), for soft-disk systems of N particles at inverse temperature $\beta = 1$. τ_6 is measured in arbitrary though comparable amounts of CPU time: One unit of CPU time for ECMC is a displacement of $T = 100N\sigma$, chain displacement $\ell = 0.025\sqrt{N}\sigma$, spanning thus about half the system volume. Event chains in $+x$ and $+y$ alternate every $\Delta T = N\sigma/2$ of displacement. One unit of CPU time for Metropolis consists of $1000N$ moves, where a move is the attempt to displace a particle by a random vector sampled from a disk of radius 0.16σ.

$\rho\sigma^2 \approx 0.89$ for harder disks with $n = 1024$, where $\rho\sigma^2$ is the dimensionless density, $N\sigma^2/V$.

We find that in terms of CPU time, in small systems, ECMC mixes a few times quicker than MMC. We tuned both algorithms to their optimal parameters (see Fig. 5 for details). Speedup, defined as the ratio $\tau_6(\text{MMC})/\tau_6(\text{ECMC})$, is, in the region of study, not found to be a strong function of density (Fig. 5, bottom row). For increasing system size, however, the speedup increases (Fig. 5, top row). An increase of speedup with system size has also been found for hard-sphere systems, where it approaches two orders of magnitude in large systems of 10^6 particles.[8] We thus expect that the generalized ECMC algorithm has similar characteristics with respect to system size as the hard-sphere ECMC algorithm. For very large systems, the MMC algorithm does not equilibrate within the allotted simulation time: The distribution of $|\Psi_6|$ is not yet stationary, even though the simulation time exceeds the τ_6 by a factor of 100. Thus, we have not determined τ_6 for these systems.

VI. DIRECT PRESSURE COMPUTATION

In order to obtain the equation of state in the NVT ensemble for the particle system under study, the pressure P must be computed. Usually, the pressure is obtained using the virial theorem (see Sec. 2.2 of Ref. 30), either by averaging the virial, or by integration of the product of the static pair correlation function $g(r)$ and the pair potential $E_{ij}(r)$. Direct averaging is not possible for hard-sphere interactions, since the potential is singular. It is thus required to compute a discrete approximation of $g(r)$ and extrapolate it to the contact value to obtain P. Even for non-singular steep potentials, the approach via $g(r)$ is bothersome, since the dominant contributions to P come from close pairs (for the family of power-law potentials, $r \approx \sigma$), which is poorly sampled in the canonical ensemble. Finally, evaluation of the virial during the simulation implies extra computation for evaluating the forces. By contrast, in hard-sphere event-driven molecular dynamics the virial pressure is directly related to the collision rate, which is a trivial byproduct of the computation:[20]

$$\beta P = \rho - \frac{\beta\rho m}{2T_{\text{sim}}N} \sum_{\text{collisions}} b_{ij}, \tag{18}$$

where $\rho = N/V$ is the particle number density, T_{sim} the total simulation time, m the mass of a particle, and $b_{ij} = (\boldsymbol{r}_i - \boldsymbol{r}_j)(\boldsymbol{v}_i - \boldsymbol{v}_j)$, with $\boldsymbol{r}_{i,j}$, $\boldsymbol{v}_{i,j}$ the positions and velocities of the colliding particles. In the following we show that in ECMC, the rate of lifting moves is, just like the collision rate in event-driven MD, directly related to the pressure. We give an elementary derivation independent of the virial theorem for the soft-particle case. The results are, however, also valid for hard particles and can be derived using arguments by Speedy,[29] which connect the pressure to the stochastic geometry of the admissible configurations.

In order to compute the pressure $\beta P = \partial \ln Z/\partial V$, we consider virtual *rift volume changes* effected by removing a randomly located strip of size $\mathrm{d}L_x \times L_y$ from the system (see Fig. 6). By considering all positions of the strip, this procedure yields all N-particle configurations in the

Reuse of AIP Publishing content is subject to the terms: https://publishing.aip.org/authors/rights-and-permissions. Downloaded to IP: 129.199.121.207 On: Mon, 08 Feb 2016 19:51:24

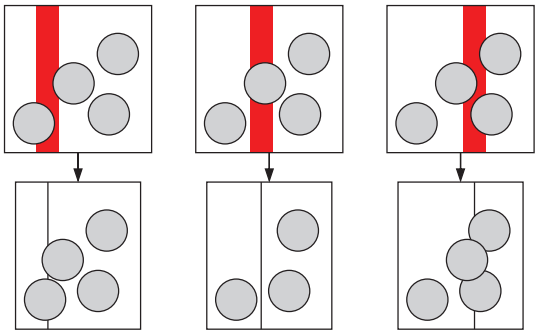


FIG. 6. Virtual rift volume changes by random removal of an infinitesimal strip from a hard-sphere configuration. Left: A successful removal. Center: Elimination of a particle (ideal gas pressure). Right: Generation of an overlap (excess pressure). The left and right cases become indistinct for soft interactions.

smaller simulation box, and thus the new partition function $Z(V + \mathrm{d}V)$. For isotropic systems, we recover the virial expression,

$$\beta P = \rho + \frac{1}{V}\left\langle \sum_{\{i,j\}} (x_j - x_i) \frac{\beta \partial E_{ij}(\boldsymbol{r}_i - \boldsymbol{r}_j)}{\partial x_i} \right\rangle, \qquad (19)$$

where $\langle \cdot \rangle$ is the canonical average. The first term of the rhs in Eq. (19) is due to particles located in the removed strip, which lead to illegal configurations with less than N particles (Fig. 6, center). This term yields the ideal-gas pressure. The non-ideal contribution to the pressure results from changes in the Boltzmann weight due to compressed bonds, with $(x_j - x_i)$ accounting for the probability of a bond traversing the removed strip. In hard spheres, this term is produced by particle overlap (Fig. 6, right panel). Replacing the canonical average in Eq. (19) by an average in the lifted canonical ensemble, $\langle \cdot \rangle_k := N^{-1} \sum_k \langle \cdot \rangle$, where k is the lifting variable, one of the sums collapses and yields a factor of N; we recover the probabilities for a lifting move from $i \to j$ and $j \to i$. An ECMC simulation will reproduce the lifted canonical average and thus yields an unbiased estimator of the pressure. Summing up all the lifting events in a chain (see Fig. 7), we obtain

$$\beta P = \rho \cdot \left\langle \frac{x_{\text{final}} - x_{\text{initial}}}{\ell} \right\rangle_{\text{chains}}, \qquad (20)$$

where $\langle \cdot \rangle_{\text{chains}}$ is the average over event chains, x_{initial} is the position of the first particle before the effects of the chain, and x_{final} the position of the last particle after, adjusted for periodic boundaries if necessary. Thus, it suffices to know the beginning and end of event chains to compute the pressure. Explicitly,

$$x_{\text{final}} - x_{\text{initial}} = \ell + \sum_{\text{lifts}} (x_j - x_i), \qquad (21)$$

where x_i and x_j are the positions of the moving particle i and of the hit particle j, respectively, *at lifting*, see Fig. 7. In the ideal gas, there are no lifting moves and Eq. (20) reduces to the ideal gas pressure. The excess displacement $(x_j - x_i)$ can be negative for an interaction potential with attractive components, such as Lennard-Jones. If the potential is decomposed into attractive and repulsive parts as outlined in Sec. IV, individual excess displacements for the two potentials also add

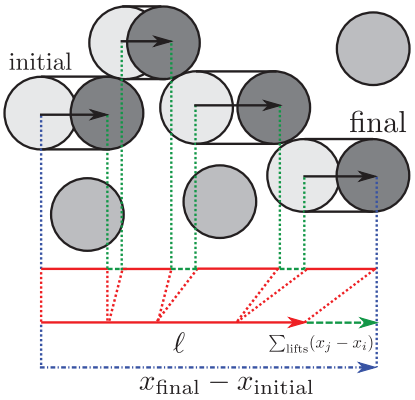


FIG. 7. Direct computation of the pressure: The excess pressure is derived from the ratio of excess displacement (green dashed lines, $\sum_{\text{lifts}}(x_j - x_i)$) and the chain displacement ℓ (red solid line). For isotropic systems, only the distance between the final and initial particle $x_{\text{final}} - x_{\text{initial}}$ (blue dashed-dotted arrow) has to be recorded.

up to the correct pressure. As evidenced by Table I, the results obtained from ECMC via Eq. (20) agree with the conventional virial approach. Since no extra computation is required, the procedure via the excess displacement in ECMC is more efficient than the virial approach, in particular for steep potentials.

Finally, one might be interested in anisotropic systems where the collision rates can depend on the direction of the event chains. In this case, the derivation presented for longitudinal rifts (removing strips normal to the chain direction) is supplemented with an analogous result for transverse rifts (removing strips aligned with the event chain), which leads in D dimensions to the full pressure,

$$\beta P = \rho + \left\langle \frac{\rho}{D\ell} \sum_{\text{lifts}} \frac{(\boldsymbol{r}_j - \boldsymbol{r}_i)^2}{x_j - x_i} \right\rangle_{\text{chains}}, \qquad (22)$$

where x is the coordinate parallel to the chain direction. More generally, the full stress tensor τ can be computed as an average of the dyadic product of the interparticle distance $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$ at collision:

$$\beta \tau = -\rho \mathbf{1} - \left\langle \frac{\rho}{\ell} \sum_{\text{lifts}} \frac{\boldsymbol{r}_{ij} \boldsymbol{r}_{ij}^{\mathrm{t}}}{x_j - x_i} \right\rangle_{\text{chains}}, \qquad (23)$$

where $\mathbf{1}$ is the identity matrix in D dimensions.

TABLE I. Comparison of pressure computed using the virial expression and from excess displacement in ECMC Eq. (20), for repulsive soft and hard sphere (HS) interactions, and for the Lennard-Jones (LJ) potential, at $\beta = 1$. Pressures and densities are nondimensionalized, $\beta P \sigma^2$ and $\rho \sigma^2$. The deviations given are standard errors from 10 independent simulations each. For LJ, the potential was decomposed into attractive and repulsive parts.

n	N	$\rho\sigma^2$	Virial pressure	ECMC pressure
12	2^{14}	0.990	14.4369 ± 0.0058	14.4267 ± 0.0038
48	2^{14}	0.860	8.753 ± 0.011	8.7565 ± 0.0023
48	2^{14}	0.888	9.441 ± 0.025	9.429 ± 0.027
1024	2^{14}	0.888	9.174 ± 0.028	9.1679 ± 0.0026
∞ (HS)	2^{16}	0.888	9.1667 ± 0.0073	9.1723 ± 0.0064
12, 6 (LJ)	2^{14}	0.888	1.44833 ± 0.00031	1.447623 ± 0.000045

Reuse of AIP Publishing content is subject to the terms: https://publishing.aip.org/authors/rights-and-permissions. Downloaded to IP: 129.199.121.207 On: Mon, 08 Feb 2016 19:51:24

VII. CONCLUSION

In the present article, we have generalized the event-chain Monte Carlo algorithm from hard spheres to particle systems interacting with arbitrary pair potentials, such as Lennard-Jones liquids or soft disks. The resulting algorithm is faster than conventional Metropolis Monte Carlo, with the gap in performance increasing with the system size. It is based on the lifting concept, and relies on a new factorization of the Metropolis filter, applied to infinitesimal Monte Carlo moves, to achieve maximal global balance. The infinitesimal moves are implemented efficiently in an event-based algorithm using the BKL approach. The algorithm generates a continuum of samples of the equilibrium distribution. This has allowed us to derive the pressure and the stress tensor in the NVT ensemble directly from the simulation without any additional computation. Even though presented in periodic boundary conditions, the algorithm also applies to nonperiodic systems, by introducing chains in the $-x$ and $-y$ direction to render it ergodic.

Infinitesimal moves permit to apply the framework of lifted Markov chains to the interacting particles problem, since they define uniquely the next event, while satisfying global balance. By subdivision into infinitesimal moves, both the original hard-sphere event-chain algorithm[6] and the hybrid MC algorithm of Peters and de With[15] are revealed to be lifting algorithms. Lifting improves mixing in large, strongly correlated systems, since clusters of particles are displaced in a cooperative way. It is thus applicable to packing problems, glassy systems, etc., particularly, as its dynamics are fundamentally different from the MMC or MD case. As the original event-chain algorithm, it can be parallelized.[9] We expect the algorithm to extend to complex fluids, with particles possessing internal degrees of freedom, to path integral (quantum) Monte Carlo, and other sampling problems.

ACKNOWLEDGMENTS

The authors thank E. A. J. F. Peters and M. Engel for useful discussion and comments on a previous version of the manuscript.

[1]The standard Metropolis filter and its factorized variant agree also for discretized interaction potentials with infinitesimal moves, as considered in Ref. 15.
[2]As it stands, the lifting algorithm is deterministic. To make it aperiodic, small lifting flows and also small on-site probabilities must be added to Eq. (9). This can be done[18] preserving the mixing time $\propto L$ for the lifting algorithm, whereas local MCMC algorithms mix in $\propto L^2$ steps. This mathematical result has been considerably generalized.[22]
[3]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[4]R. H. Swendsen and J. Wang, Phys. Rev. Lett. **58**, 86 (1987).
[5]U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[6]E. P. Bernard, W. Krauth, and D. B. Wilson, Phys. Rev. E **80**, 056704 (2009).
[7]E. P. Bernard and W. Krauth, Phys. Rev. Lett. **107**, 155704 (2011).
[8]M. Engel, J. A. Anderson, S. C. Glotzer, M. Isobe, E. P. Bernard, and W. Krauth, Phys. Rev. E **87**, 042134 (2013).
[9]S. C. Kapfer and W. Krauth, J. Phys.: Conf. Ser. **454**, 012031 (2013); arXiv:1301.4901.
[10]A. Jaster, Physica A **264**, 134 (1999).
[11]J. Liu and E. Luijten, Phys. Rev. Lett. **92**, 035504 (2004).
[12]C. Dress and W. Krauth, J. Phys. A **28**, L597 (1995).
[13]P. Gustafson, Stat. Comput. **8**, 357 (1998).
[14]S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[15]E. A. J. F. Peters and G. de With, Phys. Rev. E **85**, 026703 (2012).
[16]R. M. Neal, Technical Report No. 9508, Department of Statistics, University of Toronto, 1995.
[17]F. Chen, L. Lovász, and I. Pak, in *Proceedings of the Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, Atlanta, GA, 1999), p. 275.
[18]P. Diaconis, S. Holmes, and R. M. Neal, Ann. Appl. Probab. **10**, 726 (2000).
[19]H. C. Andersen and P. Diaconis, Journal de la société française de statistique **148**, 5 (2007).
[20]B. J. Alder and T. E. Wainwright, J. Chem. Phys. **33**, 1439 (1960).
[21]W. Krauth, *Statistical Mechanics: Algorithms and Computations* (Oxford University Press, 2006).
[22]M. Hildebrand, Markov Proc. Rel. Fields **10**, 687 (2004).
[23]K. S. Turitsyn, M. Chertkova, and M. Vucelja, Physica D **240**, 410 (2011).
[24]H. C. M. Fernandes and M. Weigel, Comput. Phys. Commun. **182**, 1856 (2011).
[25]A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. **17**, 10 (1975).
[26]K. Zahn, J. M. Méndez-Alcaraz, and G. Maret, Phys. Rev. Lett. **79**, 175 (1997).
[27]K. J. Strandburg, Rev. Mod. Phys. **60**, 161 (1988).
[28]W. Mickel, S. C. Kapfer, G. E. Schröder-Turk, and K. Mecke, J. Chem. Phys. **138**, 044501 (2013); arXiv:1209.6180.
[29]R. J. Speedy, J. Phys. Chem. **92**, 2016 (1988).
[30]J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic, London, 2006).

Reuse of AIP Publishing content is subject to the terms: https://publishing.aip.org/authors/rights-and-permissions. Downloaded to IP: 129.199.121.207 On: Mon, 08 Feb 2016 19:51:24

Publication 2: Event-chain Monte Carlo for classical continuous spin models

Manon Michel, Johannes Mayer and Werner Krauth, *EPL*, **112**, 20003 (2015).

This article discusses the implementation of the irreversible factorized Metropolis algorithm in systems of bidimensional ferromagnetic XY spins and in three-dimensional XY spin glasses. It demonstrates a $100-$fold speedup of the irreversible factorized Metropolis algorithm in respect to a local Monte Carlo method.

A Letters Journal Exploring the Frontiers of Physics

October 2015

EPL, **112** (2015) 20003
doi: 10.1209/0295-5075/112/20003

www.epljournal.org

Event-chain Monte Carlo for classical continuous spin models

Manon Michel[(a)], Johannes Mayer and Werner Krauth[(b)]

Laboratoire de Physique Statistique, Ecole Normale Supérieure / PSL Research University, UPMC, Université Paris Diderot, CNRS - 24 rue Lhomond, 75005 Paris, France

received 28 August 2015; accepted in final form 15 October 2015
published online 2 November 2015

PACS 02.70.Tt – Justifications or modifications of Monte Carlo methods
PACS 75.10.Hk – Classical spin models
PACS 75.10.Nr – Spin-glass and other random models

Abstract – We apply the event-chain Monte Carlo algorithm to classical continuum spin models on a lattice and clarify the condition for its validity. In the two-dimensional XY model, it outperforms the local Monte Carlo algorithm by two orders of magnitude, although it remains slower than the Wolff cluster algorithm. In the three-dimensional XY spin glass model at low temperature, the event-chain algorithm is far superior to the other algorithms.

Copyright © EPLA, 2015

Introduction. – Classical and quantum spin models are of fundamental interest in statistical and condensed-matter physics. Spin models are also a crucial test bed for computational algorithms.

An important representative is the model of continuous two-dimensional classical spins of fixed length (rotators) on a two-dimensional lattice. Thirty years ago, the existence and nature of the phase transition in this two-dimensional XY model were highly controversial [1]. The substitution of the traditional local Monte Carlo (LMC) algorithm [2] by Wolff's spin flip cluster (SFC) algorithm [3] then quickly allowed to clarify that this model indeed undergoes a Kosterlitz-Thouless transition [4,5], whose temperature is now known to five significant digits [6,7]. SFC has played a decisive role in understanding the physics of the XY model [8–10], and in arriving at its detailed quantitative description.

SFC and its variants can be implemented for a wide range of models, but they are efficient only in a few of them. Particularly frustrating is the case of the three-dimensional XY spin glass model, where the algorithm loses all its power [11,12]. For this much studied spin glass model, our understanding today resembles the one of the XY model before the revolution triggered by the cluster algorithms. Clearly, there still is a great need for more powerful algorithms for classical and quantum spin models.

Today's Markov-chain Monte Carlo algorithms generally follow the conventional paradigm based on three principles: 1) Each move represents a finite change of the configuration. It is independent of the previous move, and depends only on the configuration itself. 2) The algorithm satisfies the detailed-balance condition. 3) The decision whether a proposed move is accepted is based on the change in energy, using the Metropolis acceptance rule or the heat-bath condition [2,13].

In the present work, we show that the novel event-chain Monte Carlo (ECMC) paradigm [14–16], that has already been very successful in particle systems [17–20], can also be applied to the XY model and to the XY spin glass model. The paradigm breaks all three principles of the conventional Markov-chain scheme: Moves are infinitesimal rather than finite, although an event-driven scheme allows to recover finite displacements [16]. In one-dimensional systems, the moves do not change with time. In multidimensional systems, moves persist on long time scales. This is achieved within the Markov-chain scheme through additional "lifting" variables [15,21]. In addition, ECMC violates the detailed balance and only satisfies the weaker global balance condition (cf. [22–26]). Finally, the decision on future moves is based on the change in pair energies, rather than on the change in total energy. This is achieved by replacing the standard Metropolis algorithm by its recently introduced factorized variant [15].

For the two-dimensional XY model at the critical point, we find that ECMC is about 100 times faster than LMC, although the presence of a slow time scale in autocorrelation functions makes that it is not as fast as SFC. In the low-temperature phase of the three-dimensional XY spin

[(a)] E-mail: manon.michel@ens.fr
[(b)] E-mail: werner.krauth@ens.fr

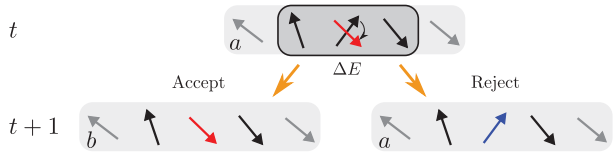


Fig. 1: (Colour on-line) LMC move for the one-dimensional XY model. Upper panel: configuration at time t and proposed displacement $\Delta\phi$ of a randomly chosen spin, corresponding to an energy change ΔE. Lower panel: possible configurations at time $t+1$. The proposed move is accepted with probability $\min(1, \exp(-\beta\Delta E))$ (left) and rejected otherwise (right).

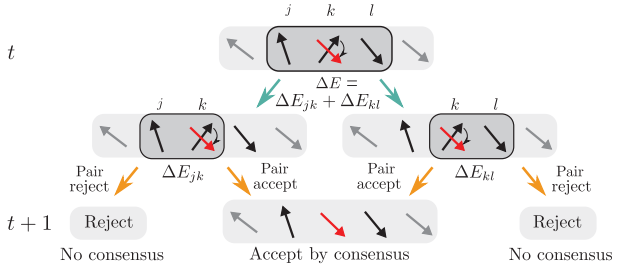


Fig. 2: (Colour on-line) Factorized Metropolis move. Upper panel: configuration at time t and proposed displacement $\Delta\phi$ of a randomly chosen spin k. Middle panel: factorization into pairs (j,k) and (k,l). In the factor (j,k), the move is pair-accepted with probability $\min(1, \exp(-\beta\Delta E_{jk})$, etc. Lower panel: possible configurations at time $t+1$; the proposed move is either accepted by consensus (*i.e.* accepted independently by each pair) or else rejected.

glass model, where SFC is known to be inefficient, ECMC clearly outperforms LMC.

From local Monte Carlo to the "event-chain" algorithm. – In the two-dimensional ferromagnetic XY model of spins $\boldsymbol{S}_k = (S_k^x, S_k^y) = (\cos\phi_k, \sin\phi_k)$ on a lattice with sites $i = 1, \ldots, N$, and with an energy

$$E = -\sum_{\langle i,j\rangle} J_{ij}\boldsymbol{S}_i \cdot \boldsymbol{S}_j = \sum_{\langle i,j\rangle} \underbrace{[-J_{ij}\cos(\phi_i - \phi_j)]}_{E_{ij}}, \qquad (1)$$

the coupling constants J_{ij} are all equal to one. The sum $\langle i,j\rangle$ goes over nearest neighbors on the lattice. We refer to the E_{ij} as "pair energies". The XY model on a two-dimensional square lattice undergoes a phase transition at inverse temperature $\beta = 1.1199$, see ref. [6].

In LMC, one proposes at each time step t a finite move from a configuration a to a configuration b (a rotation by a finite angle $\Delta\phi$ of a spin k), as sketched in fig. 1. To satisfy detailed balance [13], k is randomly chosen at each time step, and $\Delta\phi$ is sampled from a symmetric distribution around zero, so that $\Delta\phi$ arises with the same probability as $-\Delta\phi$. The proposed move corresponds to an energy change $\Delta E = E_b - E_a$ in eq. (1), and it is accepted with probability

$$p_{\rm acc}^{\rm Met} = \min(1, \exp(-\beta\Delta E)). \qquad (2)$$

The exponential in this equation corresponds to the ratio π_b/π_a of the Boltzmann weights of the configurations.

Practically, the move is accepted, and the configuration updated to b, if a uniform random number between 0 and 1 satisfies $\mathrm{ran}(0,1) < p_{\rm acc}^{\rm Met}$ (see [13]). Otherwise, the configuration at time $t+1$ is the same as the one at time t, namely a.

The recently introduced factorized algorithm [15] also satisfies the detailed-balance condition. In this method, the energy-based Metropolis acceptance probability is replaced by a factorized form which separately depends on the pair-energy changes:

$$p_{\rm acc}^{\rm fact} = \prod_{\langle k,l\rangle} p_{\rm acc}^{kl} = \prod_{\langle k,l\rangle} \min(1, \exp(-\beta\Delta E_{kl})). \qquad (3)$$

The proposed move $a \to b$ is accepted with this probability. The factorized algorithm always has a smaller acceptance rate than the conventional one, $p_{\rm acc}^{\rm fact} \le p_{\rm acc}^{\rm Met}$ (this will however turn out not to be a problem in ECMC). To implement eq. (3), one might use a single random number and accept the move if $\mathrm{ran}(0,1) < p_{\rm acc}^{\rm fact}$. We rather accept the move if several independent random numbers satisfy $\mathrm{ran}_{kl}(0,1) < p_{\rm acc}^{kl}$ for all pairs k,l. In other words, a move is accepted only if it is pair-accepted by all pairs k,l. This consensus rule is illustrated in fig. 2. We note that the factorization in eq. (3) relies on the possibility to cut the Hamiltonian into independent pieces. The factorization may also be used to separate different components of the inter-particle potential, as for example the $1/r^6$ and $1/r^{12}$ pieces in the Lennard-Jones potential [15,19].

The ECMC combines the factorized Metropolis probability with the "lifting" concept of Diaconis *et al.* [21] and with the idea of infinitesimal displacements [15]. The term "lifting" refers to the extension of the physical configuration by an additional variable that fixes the proposed move. Written as $\overset{\curvearrowright}{k}$, it singles out the spin k as the only one that can move, as $\phi_k \to \phi_k + \Delta\phi$ (see fig. 3). If the move is accepted, the lifting variable for the next time step $t+1$ is again $\overset{\curvearrowright}{k}$. If the physical move is rejected, a lifting move takes place and the lifting variable is passed on to the spin l of the pair that rejected the move, and the physical configuration is unchanged. In both cases, the value of $\Delta\phi$ is used again. Note that for infinitesimal $\Delta\phi$, the acceptance probabilities of the physical moves approach one and the rejection probabilities approach zero. Multiple rejections are totally suppressed, and the choice of $\overset{\curvearrowright}{l}$ is unique [15]. At each time step, either a lifting move or a physical move takes place, and ECMC is thus formally rejection-free.

ECMC satisfies the global balance condition in the XY model, as we now show: For simplicity, we consider only two spins and concentrate on a configuration d (see fig. 3). This configuration can only be reached through a lifting move from a or through a physical move from b. The global-balance condition [13] states that the flow into

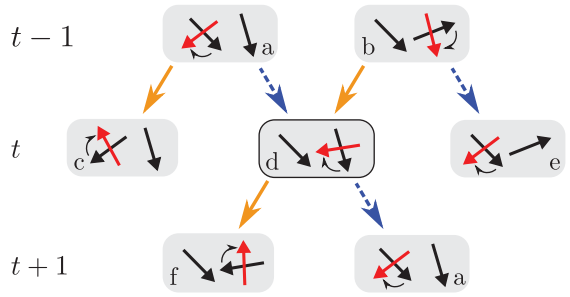


Fig. 3: (Colour on-line) Lifting approach of ECMC. Physical moves $b \to d$, $d \to f$ and $a \to c$ are by the same infinitesimal angle $\Delta\phi$ in clockwise direction, all others are lifting moves that preserve the physical configuration. Note that $\pi_b = \pi_c$ because of eq. (1).

configuration d must be equal to the flow out of it:

$$\underbrace{\pi_a p(a \to d)}_{\mathcal{P}(a\to d)} + \underbrace{\pi_b p(b \to d)}_{\mathcal{P}(b\to d)} = \underbrace{\pi_d p(d \to f)}_{\mathcal{P}(d\to f)} + \underbrace{\pi_d p(d \to a)}_{\mathcal{P}(d\to a)}. \tag{4}$$

Here, $\mathcal{P}(a \to d)$ represents the probability flow from a to d, etc. For ECMC, the probabilities p in eq. (4) coincide with the acceptance probabilities: All configurations carry a lifting variable that specifies the spin that may move and the move itself, $\Delta\phi$.

The statistical weight π_a is trivially equal to π_d because they differ only by a lifting move. Furthermore, π_c equals π_b, as the two configurations differ only by a global rotation. Writing $\Delta E = E_b - E_d$, we thus find

$$\begin{aligned}\mathcal{P}(b \to d) &= \pi_b p^{\text{frac}}_{\text{acc}}(b \to d) = \pi_d p^{\text{frac}}_{\text{acc}}(d \to b) \\ &= \pi_d \min(1, \exp(-\beta\Delta E)).\end{aligned} \tag{5}$$

Note in this equation that $\pi_b p^{\text{frac}}_{\text{acc}}(b \to d) = \pi_d p^{\text{frac}}_{\text{acc}}(d \to b)$, because the factorized transition probabilities satisfy detailed balance. Likewise, the change in energy in going from $a \to c$ is also ΔE and $p(a \to d) = 1 - p(a \to c)$. Therefore, the flow $\mathcal{P}(a \to d)$ satisfies

$$\begin{aligned}\mathcal{P}(a \to d) &= \pi_a(1 - \min(1, \exp(-\beta\Delta E)) \\ &= \pi_d(1 - \min(1, \exp(-\beta\Delta E)).\end{aligned} \tag{6}$$

It follows that the flow into d, namely the sum of $P(a \to d)$ and of $P(b \to d)$, equals π_d. As for the flow out of d, it trivially equals π_d because of the conservation of probabilities. It follows that the global balance of eq. (4) is satisfied. The factorization property and the infinitesimal limit guarantee that the argument carries over to general N (see [15]).

ECMC violates the detailed-balance condition $\mathcal{P}(b \to d) = \mathcal{P}(d \to b)$: A move $d \to b$ would be anti-clockwise, yet all moves within ECMC are, by the initial choice of $\Delta\phi$, clockwise. Also, $\mathcal{P}(a \to d) = 0$, as $E_d > E_f$ and all physical moves from d to f are accepted. Furthermore, for ECMC to be valid, the pair energy must be symmetric (so that $\pi_b = \pi_c$ in fig. 3). Modified XY models, as described in ref. [27], can also be treated, but more general pair energies require special considerations [28].

ECMC with infinitesimal moves requires a scaling of physical time: In one unit of time, as $\Delta\phi$ goes to zero, an infinite number of physical moves take place, but the number of lifting moves remains finite. In an event-driven approach [15,16], the algorithmic complexity can be made to scale with the number of liftings: The lifting variable being set to $\overset{\curvearrowright}{k}$, the angle ϕ_k now rotates clockwise until the "event", *i.e.* a lifting move, is produced through a rejection by a neighbor l. The lifting variable is updated to $\overset{\curvearrowright}{l}$, ϕ_l rotates clockwise, etc. Effectively, one undergoes an infinite number of Monte Carlo steps, giving a continuous trajectory.

The angle ϕ_k corresponding to the next event is easily sampled: We continue to consider a single pair (k, l) of spins, with the lifting variable $\overset{\curvearrowright}{k}$. The i-th infinitesimal update of ϕ_k is noted as the move $i - 1 \to i$ and the weight of the configuration $(\phi_i = \phi_k + i\mathrm{d}\phi, \phi_l)$, π_i. The probability $p_{\text{event}}(0 \to n)$ to accept n subsequent physical moves and then to reject the $(n+1)$-st physical move is

$$\begin{aligned}p_{\text{event}}(0 \to n) &= p_{\text{acc}}(0 \to 1) \cdots p_{\text{acc}}(n-1 \to n) \\ &\quad \times [1 - p_{\text{acc}}(n \to n+1)].\end{aligned} \tag{7}$$

The jth term in this expression is $\min(1, \pi_j/\pi_{j-1})$. Supposing for a moment that π_j is monotonously decreasing with j, this gives

$$\begin{aligned}p_{\text{event}}(0 \to n) &= \frac{\pi_{n-1}}{\pi_0}\left(1 - \frac{\pi_n}{\pi_{n-1}}\right) \\ &= \frac{-1}{\pi_0}\left.\frac{\partial\pi}{\partial\phi_k}\right|_{\phi_k=\phi_n} \mathrm{d}\phi.\end{aligned} \tag{8}$$

This probability is normalized, writing ϕ_{event} the value of ϕ_k at which the event happens:

$$\begin{aligned}&-\frac{1}{\pi_0}\int_0^\infty \left.\frac{\partial\pi}{\partial\phi_k}\right|_{\phi_k=\phi_{\text{event}}} \mathrm{d}\phi_{\text{event}} \\ &= \frac{1}{\pi_0}\int_0^{\pi_0} \mathrm{d}\pi_{\text{event}} = 1.\end{aligned} \tag{9}$$

This integral is sampled by [13]

$$\begin{aligned}\pi_{\text{event}} &= \text{ran}(0, \pi_0) \\ \pi_{\text{event}}/\pi_0 &= \text{ran}(0, 1),\end{aligned} \tag{10}$$

which is equivalent to the following sampling of the energy increase:

$$\Delta E(\phi_{\text{event}}) = -\left[\log\ \text{ran}(0,1)\right]/\beta. \tag{11}$$

Sampling π uniformly between 0 and the present value, π_0 (equivalently, ΔE from its exponential distribution) thus yields the event time, ϕ_{event} (see fig. 4).

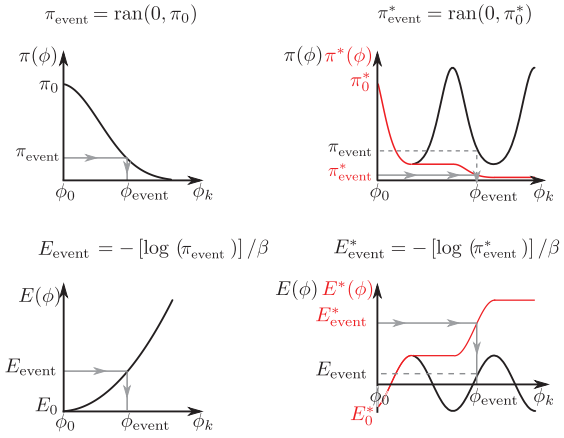


Fig. 4: (Colour on-line) Event-driven implementation of ECMC for a pair of spins (k,l). From a starting point $\phi_k = \phi_0$ of weight π_0 and energy E_0, ϕ_k is updated by infinitesimal moves until $\phi_k = \phi_{\rm event}$. Left: monotonously decreasing distribution π: the lifting event is sampled as $\pi_{\rm event} = {\rm ran}(0,\pi_0)$. Right: general distribution π; $E^*_{\rm event} - E^*(0) = [-\log\ {\rm ran}(0,1)]/\beta$.

For a non-monotonous probability distribution, all negative energy increments correspond to an acceptance probability 1, and disappear from eq. (7). The sampling of the energy increase in eq. (11) turns into the sampling of only the positive energy changes. As shown in fig. 4, this can be expressed as a function E^*, constructed only from the positive increments of the energy E [16].

For a system of more than one pair of spins, the event times $\phi_{\rm event}$ for each neighbor of the lifted spin k can be computed independently in view of the factorized probability of eq. (3), and k turns clockwise up to the earliest event (that involves, say, another spin l). The lifting variable is then set to $\overset{\frown}{l}$.

It follows from eq. (7) that all configurations encountered between two events sample the Boltzmann distribution. Any uniform subset of these configurations can be used for averaging observables. A practical choice consists in outputting spin configurations at regular intervals independent of the occurrence of events.

For the models considered here, we found that the efficiency was not increased by halting and restarting the simulation after fixed displacements. In contrast, switches between moves along the different coordinate axes assure ergodicity in multi-dimensional Hamiltonians as they appear in particles systems [14], but also the related Heisenberg model.

Simulations for the two-dimensional XY model at the critical point. – In the two-dimensional XY model, we consider the susceptibility χ

$$\chi = \frac{\|\sum \boldsymbol{S}_k\|^2}{N}, \tag{12}$$

and estimate the convergence properties by the susceptibility autocorrelation function

$$C_\chi(t) = \frac{\langle\chi(t'+t)\chi(t')\rangle - \langle\chi\rangle^2}{\langle\chi^2\rangle - \langle\chi\rangle^2} \tag{13}$$

at the critical point $\beta = 1.1199$ (see [6]). We suppose that χ is a slow variable of this model. We measure time in sweeps: For ECMC, one sweep corresponds to $\sim N$ lifting events while for LMC, one sweep corresponds to N attempted moves. For SFC, a sweep denotes $\sim N$ spins added to clusters. The complexity of one sweep is $O(N)$ in the three algorithms and the CPU times used per sweep are roughly comparable.

In fig. 5, we show the autocorrelation function for the XY model at its critical point, obtained from very long single runs of the algorithms. For LMC and SFC, the decay of the susceptibility autocorrelation function can be described by a single time scale, while for ECMC, it is well described by two time scales:

$$C_\chi(t) \simeq \begin{cases} \exp(-t/\tau^{\rm LMC}) \quad ({\rm LMC}), \\ \exp(-t/\tau^{\rm SFC}) \quad ({\rm SFC}), \\ A_0\exp(-t/\tau_0^{\rm ECMC}) \\ +\,A_1\exp(-t/\tau_1^{\rm ECMC}) \quad ({\rm ECMC}). \end{cases} \tag{14}$$

For ECMC, this correlation function rapidly decays to $C_\chi \sim 0.1$ on a timescale $\tau_0^{\rm ECMC}$ of about 5 sweeps. A slow mode $\tau_1^{\rm ECMC}$ then sets in. It presents a $z = 2$ scaling ($\tau_1^{\rm ECMC} \sim L^2$, with $N = L^2$). As shown on the right panel of fig. 5, $\tau_1^{\rm ECMC}$ is an order of magnitude smaller than $\tau^{\rm LMC}$. Together with the initial rapid decrease, this makes ECMC about one hundred times faster than LMC. However, the dynamical scaling exponent seems to be unchanged with respect to LMC: After a fast initial decay, on a time scale $\tau_0^{\rm ECMC}$ that appears independent of system size ($z \sim 0$), the relaxation crosses over to a slower decay, characterized by a constant $\tau_1^{\rm ECMC}$ that scales as $z \sim 2$, just as $\tau^{\rm LMC}$. In the absence of a clear explanation of this cross-over, we notice that an analogous phenomenon appears in ECMC simulations for particle systems, where the initial ballistic behavior gives rise to a slow diffusive decay, with what appears to be the dynamical critical exponent of LMC [29].

Three-dimensional XY spin glass model. – We now study ECMC for the three-dimensional XY spin glass model, where the nearest-neighbor coupling constants J_{ij} are drawn from a Gaussian normal distribution of zero mean and unit variance. The algorithm can be formulated as for the ferromagnetic model, and the spins continue to always turn clockwise. We will find evidence that the relaxation dynamics of ECMC differs from the one of LMC. Following [11], we consider the chiral overlap between two independent systems, (1) and (2), with identical coupling constants

$$p_\kappa = \frac{1}{N}\sum_{p=1}^{N} \kappa^{(1)}_{p\perp\mu}\kappa^{(2)}_{p\perp\mu}, \tag{15}$$

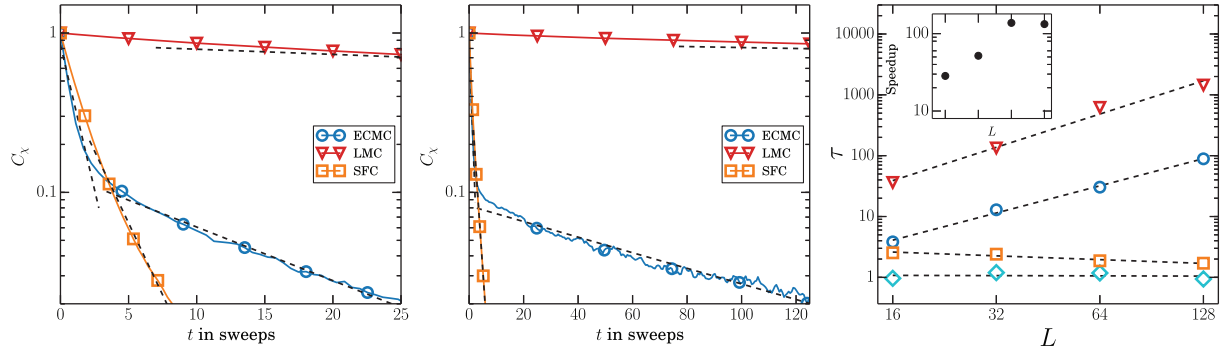


Fig. 5: (Colour on-line) Autocorrelation function $C_\chi(t)$ for the two-dimensional XY model at the critical point $\beta = 1.1199$ for LMC (red, triangle), ECMC (blue, circle), and SFC (yellow, square). Exponential fits (black dotted line) are as in eq. (14). Left: $N = 32^2$. Middle: $N = 128^2$. Right: scaling of the autocorrelation time τ with the system size. Both LMC (red triangles) and the slow scale of ECMC (dark blue circles) are compatible with a dynamical scaling exponent $z \sim 2$. Both the fast scale of ECMC (light blue diamonds) and SFC (yellow squares) are compatible with $z \sim 0$. Inset: speedup of ECMC with respect to LMC *vs.* L.

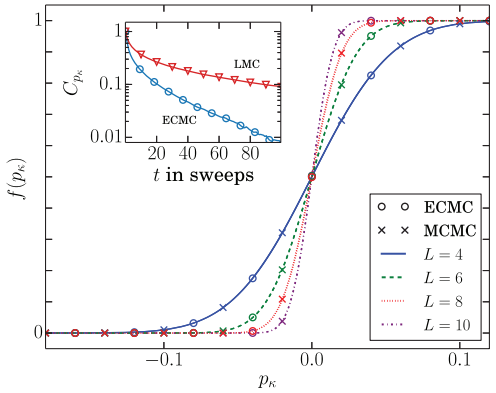


Fig. 6: (Colour on-line) Cumulative distribution of the chiral overlap p_κ for the three-dimensional XY spin glass model for $N = 4^3, 6^3, 8^3, 10^3$ at $\beta = 1.5$, in the high-temperature phase (single samples). Inset: autocorrelation function $C_{p_\kappa}(t)$ for $N = 6^3$ from LMC (red, triangle) and ECMC (blue, circle).

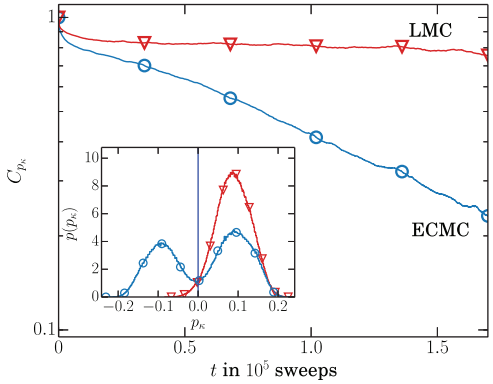


Fig. 7: (Colour on-line) Chiral overlap autocorrelation function from ECMC (blue, circle) and LMC (red, triangle) at $\beta = 3.636$ for a given sample at $N = 6^3$. Inset: distributions of p_κ after 10^6 sweeps for the two algorithms in the same sample. Note the nearly symmetric distribution for ECMC.

with $\kappa^{(i)}_{p\perp\mu}$ being the chirality at a plaquette p, perpendicular to the axis μ, defined as

$$\kappa^{(i)}_{p\perp\mu} = \frac{1}{2\sqrt{2}} \sum_{(i,j)\in p} \mathrm{sgn}(J_{ij}) \sin(\phi_i - \phi_j). \qquad (16)$$

The sum $\sum_{(i,j)\in p}$ is taken over the four bonds encircling the plaquette p clockwise. By construction, p_κ is a symmetric function about zero. As shown in fig. 6, ECMC and LMC agree very well at high temperature. The autocorrelation function of the chiral overlap for LMC and ECMC, shown in fig. 6, gives at high temperature a size-independent speedup by a factor ~ 5 of ECMC.

The phase diagram of the three-dimensional XY spin glass model at low temperature (with the possible existence of separate spin-glass and chiral-glass phases) is still being debated. We consider $\beta = 3.636$, which may be the locus of the spin glass transition [12], or below it, near the transition [30,31]. At this temperature, ECMC exhibits a striking advantage over LMC in one third of samples of size $N = 6^3$, where it explores the configuration space more easily, without using parallel tempering [32]. A typical example of a symmetric chiral overlap distribution profile after 10^6 sweeps (symmetric for ECMC, but not for LMC) is shown in fig. 7, together with the corresponding autocorrelation function. For larger systems, the speedup of ECMC with respect to LMC seems to increase, but already for 10^3 systems, ECMC no longer equilibrates at $\beta = 3.636$.

Conclusion. – In conclusion, we have applied in this work the recent event-chain algorithm to classical spin models, and obtained a considerable algorithmic speedup with respect to the local Monte Carlo algorithm for the two-dimensional XY model at its critical point. The new method appears very general, as we also obtained a clear acceleration for the three-dimensional XY spin glass

model at low temperature. It will be interesting to see how well the event-chain algorithm couples with the traditional acceleration methods, as for example the parallel tempering method, or the overrelaxation approaches that have been much used for spin glasses.

* * *

We thank K. Hukushima for helpful discussions. This work was granted access to the HPC resources of MesoPSL financed by the Region Ile de France and the project Equip@Meso (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche.

REFERENCES

[1] Seiler E., Stamatescu I. O., Patrascioiu A. and Linke V., *Nucl. Phys. B*, **305** (1988) 623.
[2] Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H. and Teller E., *J. Chem. Phys.*, **21** (1953) 1087.
[3] Wolff U., *Phys. Rev. Lett.*, **62** (1989) 361.
[4] Hasenbusch M., Marcu M. and Pinn K., *Physica A*, **208** (1994) 124.
[5] Janke W. and Nather K., *Phys. Rev. B*, **48** (1993) 7419.
[6] Hasenbusch M., *J. Phys. A: Math. Gen.*, **38** (2005) 5869.
[7] Komura Y. and Okabe Y., *J. Phys. Soc. Jpn.*, **81** (2012) 113001.
[8] Kosterlitz J. M. and Thouless D. J., *J. Phys. C*, **6** (1973) 1181; Kosterlitz J. M., *J. Phys. C*, **7** (1974) 1046.
[9] José J. V., Kadanoff L. P., Kirkpatrick S. and Nelson D. R., *Phys. Rev. B*, **16** (1977) 1217.
[10] Amit D. J., Goldschmidt Y. Y. and Grinstein G., *J. Phys. A: Math. Gen.*, **13** (1980) 585.
[11] Kawamura H. and Li M. S., *Phys. Rev. Lett.*, **87** (2001) 187204.
[12] Obuchi T. and Kawamura H., *Phys. Rev. B*, **87** (2013) 174438.
[13] Krauth W., *Statistical Mechanics: Algorithms and Computations* (Oxford University Press) 2006.
[14] Bernard E. P., Krauth W. and Wilson D. B., *Phys. Rev. E*, **80** (2009) 056704.
[15] Michel M., Kapfer S. C. and Krauth W., *J. Chem. Phys.*, **140** (2014) 054116.
[16] Peters E. A. J. F. and de With G., *Phys. Rev. E*, **85** (2012) 026703.
[17] Bernard E. P. and Krauth W., *Phys. Rev. Lett.*, **107** (2011) 155704.
[18] Isobe M. and Krauth W., *J. Chem. Phys.*, **143** (2015) 084509.
[19] Kapfer S. C. and Krauth W., *Phys. Rev. Lett.*, **114** (2015) 035702.
[20] Kampmann T. A., Boltz H. H. and Kierfeld J., *J. Comput. Phys.*, **281** (2015) 864.
[21] Diaconis P., Holmes P. S. and Neal R. M., *Ann. Appl. Probab.*, **10** (2000) 726.
[22] Turitsyn K. S., Chertkov M. and Vucelja M., *Physica D*, **240** (2011) 410.
[23] Suwa H. and Todo S., *Phys. Rev. Lett.*, **105** (2011) 120603.
[24] Fernandes H. C. M. and Weigel M., *Comput. Phys. Commun.*, **182** (2011) 1856.
[25] Sakai Y. and Hukushima K., *J. Phys. Soc. Jpn.*, **82** (2013) 064003.
[26] Ichiki A. and Ohzeki M., *Phys. Rev. E*, **88** (2013) 020101.
[27] Domany E., Schick M. and Swendsen R. H., *Phys. Rev. Lett.*, **52** (1984) 1535.
[28] Michel M. and Krauth W., unpublished.
[29] Kapfer S. C. and Krauth W., *J. Phys. Conf. Ser.*, **454** (2013) 012031.
[30] Pixley J. H. and Young A. P., *Phys. Rev. B*, **78** (2008) 014419.
[31] Romá F. and Domínguez D., *Phys. Rev. B*, **89** (2014) 024408.
[32] Hukushima K. and Nemoto K., *J. Phys. Soc. Jpn.*, **65** (1996) 1604.

Publication 3: Event-chain algorithm for the Heisenberg model: Evidence for $z \simeq 1$ dynamic scaling

Yoshihiko Nishikawa, Manon Michel, Werner Krauth and Koji Hukushima, *PRE*, **92**, 063306 (2015).

This article studies the implementation of the irreversible factorized Metropolis algorithm in systems of three-dimensional Heisenberg spins. It exhibits the important reduction of the critical slowing down obtained in this system and provides evidence for the square-root reduction of the dynamical critical scaling exponent.

Event-chain algorithm for the Heisenberg model: Evidence for $z \simeq 1$ dynamic scaling

Yoshihiko Nishikawa,[1,*] Manon Michel,[2,†] Werner Krauth,[2,‡] and Koji Hukushima[1,3,§]

[1]*Department of Basic Science, University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo 153-8902, Japan*

[2]*Laboratoire de Physique Statistique, Ecole Normale Supérieure, PSL Research University, UPMC, Université Paris Diderot, CNRS, 24 Rue Lhomond, 75005 Paris, France*

[3]*Center for Materials Research by Information Integration, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan*

(Received 5 October 2015; published 14 December 2015)

We apply the event-chain Monte Carlo algorithm to the three-dimensional ferromagnetic Heisenberg model. The algorithm is rejection-free and also realizes an irreversible Markov chain that satisfies global balance. The autocorrelation functions of the magnetic susceptibility and the energy indicate a dynamical critical exponent $z \approx 1$ at the critical temperature, while that of the magnetization does not measure the performance of the algorithm. We show that the event-chain Monte Carlo algorithm substantially reduces the dynamical critical exponent from the conventional value of $z \simeq 2$.

DOI: 10.1103/PhysRevE.92.063306 PACS number(s): 02.70.Tt, 75.10.Hk, 75.10.Nr, 05.10.−a

I. INTRODUCTION

Ever since the advent of the local Metropolis Monte Carlo algorithm (LMC) [1], Monte Carlo simulations of systems with many degrees of freedom have played an important role in statistical physics. Near phase transitions, the LMC is severely hampered by dynamical arrest phenomena such as critical slowing down for second-order transitions, nucleation and coarsening for first-order transitions, and glassy behavior in disordered systems. A number of specialized algorithms then allow one to speed up the sampling of configuration space, namely, the Swendsen-Wang [2] and the Wolff [3] cluster algorithms, the multicanonical method [4], and the exchange Monte Carlo method [5] based on extended ensembles.

The above algorithms respect detailed balance, a sufficient condition for the convergence towards the equilibrium Boltzmann distribution. Recently, algorithms breaking detailed balance but satisfying the necessary global-balance condition have been discussed [6–9]. Among them, the event-chain Monte Carlo algorithm (ECMC) [9] has proven useful in hard-sphere [10,11] and more general particle systems [12,13], allowing one to equilibrate systems larger than previously possible [11,14]. It has also been applied to continuous spin systems [15]. The ECMC uses a factorized Metropolis filter [12] and relies on an additional lifting variable to augment configuration space [16]. It is rejection-free and realizes an irreversible Markov chain. So far, however, the speedup realized by the ECMC with respect to the LMC has always represented a constant factor in the thermodynamic limit, although larger gains are theoretically possible [16,17].

In this paper we apply the ECMC to the three-dimensional ferromagnetic Heisenberg model, defined by the energy

$$E(\{\boldsymbol{S}_i\}) = \sum_{\langle i,j \rangle} E_{ij} = -J \sum_{\langle i,j \rangle} \boldsymbol{S}_i \cdot \boldsymbol{S}_j, \tag{1}$$

where J is the unit of the energy, $\boldsymbol{S}_i$ is a three-component unit vector, and the sum runs over all neighboring pairs of the $N = L^3$ sites of a simple cubic lattice of linear extension L. In our simulations, we consider the critical inverse temperature $\beta_c = J/T_c = 0.6930$ [18]. To describe the dynamics of the system, we compute the autocorrelation functions of the energy, the system magnetization $\boldsymbol{M} = \sum_k \boldsymbol{S}_k$, and the magnetic susceptibility

$$\chi = \frac{|\boldsymbol{M}|^2}{N}.$$

Both the energy and the susceptibility are invariant under global rotations of the spins $\boldsymbol{S}_k$ around a common axis, whereas the magnetization follows the rotation. We will argue that the energy and the susceptibility are slow variables, that is, their slowest time constant describes the correlation (mixing) time of the underlying Markov chain. Under this hypothesis, we will present evidence that the ECMC for the three-dimensional Heisenberg model reduces the dynamical critical exponent from the LMC value of $z \simeq 2$ to $z \simeq 1$. This considerable reduction of mixing times with respect to the LMC may be optimal within the lifting approach [17]. The observed reduction is all the more surprising as in the closely related XY model [15], where the spins are two-dimensional unit vectors, the ECMC realizes speedups by two orders of magnitude with respect to the LMC, but does not seem to lower the dynamical critical exponent.

II. THE ECMC ALGORITHM FOR THE HEISENBERG MODEL

In the LMC, finite local moves are proposed randomly and a move from a configuration a to a configuration b is accepted with the notorious Metropolis filter

$$p^{\mathrm{Met}}(a \to b) = \min[1, \exp(-\beta \Delta E)], \tag{2}$$

where $\Delta E = E_b - E_a$ is the change of the system energy. For Heisenberg spins, as for any system with pairwise interactions,

*nishikawa@huku.c.u-tokyo.ac.jp
†manon.michel@ens.fr
‡werner.krauth@ens.fr
§hukusima@phys.c.u-tokyo.ac.jp

 ©2015 American Physical Society

we may write Eq. (2) as

$$p^{\mathrm{Met}}(a \to b) = \min\left[1, \prod_{\langle i,j\rangle} \exp(-\beta \Delta E_{ij})\right],$$

with the pair energies $\Delta E_{ij} = E^b_{ij} - E^a_{ij}$ [see Eq. (1)]. If the move $b \to a$ is proposed with the same probability as $a \to b$, the detailed-balance condition

$$\pi_a p^{\mathrm{Met}}(a \to b) = \pi_b p^{\mathrm{Met}}(b \to a) \tag{3}$$

is satisfied with the Boltzmann weights $\pi_a = \exp(-\beta E_a)$ and $\pi_b = \exp(-\beta E_b)$. The LMC dynamics is diffusive and detailed balance is enforced through the rejections in the Metropolis filter of Eq. (2).

In contrast to the LMC, the ECMC produces persistent infinitesimal moves that nevertheless amount to finite displacements. Specifically, in the Heisenberg model, it augments the physical space of spin configurations by a lifting variable $(k, \boldsymbol{v})$ that defines the considered infinitesimal counterclockwise rotation of spin k about the axis $\boldsymbol{v}$. This rotation is accepted according to a consensus based on all individual pair energies, namely, the factorized Metropolis filter [12]

$$p^{\mathrm{fact}}(a \to b) = \prod_{\langle i,j\rangle} \min[1, \exp(-\beta \Delta E_{ij})]. \tag{4}$$

For infinitesimal rotations and by virtue of the factorized Metropolis filter, this physical move can only be rejected by a single neighboring spin l and the lifting variable will then be moved as $(k, \boldsymbol{v}) \to (l, \boldsymbol{v})$, keeping the sense of rotation, but passing it on to the spin responsible for the rejection. In the augmented space, the rejections are thus supplanted by events, namely, the lifting moves for arrested physical states. As there are no clockwise moves, obviously $p^{\mathrm{fact}}(b \to a)$ is zero if $p^{\mathrm{fact}}(a \to b) > 0$, so the detailed balance condition of Eq. (3) is broken. Nevertheless, it is easy to show for infinitesimal moves that the more general global balance condition

$$\sum_a \pi_a p^{\mathrm{fact}}(a \to b) = \sum_c \pi_b p^{\mathrm{fact}}(b \to c) \tag{5}$$

is satisfied [12,15], with the stationary Boltzmann weights. Equation (5) describes equality between the global probability flow into the configuration b (on the left-hand side) and the flow out of it (on the right-hand side). In contrast to Eq. (3), $a, b, c, \ldots$ now comprise the lifting variable and there are no rejections in this augmented space (see Ref. [15]).

Practically, while the spin $\boldsymbol{S}_k$ rotates around $\boldsymbol{v}$, the azimuthal angle $\phi_{\boldsymbol{v},k}$ increases from its initial value ϕ_0 until one of its neighbors l triggers a lifting $(k, \boldsymbol{v}) \to (l, \boldsymbol{v})$ at $\phi_{\boldsymbol{v},k} = \phi_{l,\mathrm{event}}$. One no longer samples the acceptation of each infinitesimal rotation of $\boldsymbol{S}_k$, but directly samples the event angle $\phi_{l,\mathrm{event}}$. It is sampled with a single random number in the event-driven approach [12,13]. Precisely, $\phi_{l,\mathrm{event}}$ is given by the sampling of the positive pair energy increase

$$\Delta E_l = -[\ln \mathcal{R}(0,1)]/\beta, \tag{6}$$

where $\mathcal{R}(0,1)$ is a uniform random number between 0 and 1.

For a fixed rotation axis $\boldsymbol{v}$, the ECMC for the Heisenberg model reduces to the one of the XY model: With $(\phi_{\boldsymbol{v},k}, \theta_{\boldsymbol{v},k})$ the spherical coordinates of a spin k in a system where the z axis is aligned with $\boldsymbol{v}$, the pair energy E_{kl} between spins k and l is

$$E_{kl} = -J' \cos(\phi_{\boldsymbol{v},k} - \phi_{\boldsymbol{v},l}) + K, \tag{7}$$

with

$$J' = J \sin\theta_{\boldsymbol{v},k} \sin\theta_{\boldsymbol{v},l},$$
$$K = -J \cos\theta_{\boldsymbol{v},k} \cos\theta_{\boldsymbol{v},l}.$$

Both J' and K depend only on the polar angles $\theta_{\boldsymbol{v}}$ and remain unchanged along the event chain. The azimuthal-angle dependence in Eq. (7) is proportional to $\cos(\phi_{\boldsymbol{v},k} - \phi_{\boldsymbol{v},l})$, as in the XY model. The positive pair energy increase of Eq. (6) then becomes

$$\Delta E_l = -J' \int_{\phi_0}^{\phi_{l,\mathrm{event}}} \max\left(0, \frac{d\cos(\phi_{\boldsymbol{v},k} - \phi_{\boldsymbol{v},l})}{d\phi_{\boldsymbol{v},k}}\right) d\phi_{\boldsymbol{v},k}. \tag{8}$$

To solve Eq. (8) for $\phi_{l,\mathrm{event}}$, one first slices off any full rotations (these n rotations by 2π yield an energy increase of $2nJ'$), leaving a value ΔE_l^f,

$$E^*_{\mathrm{init}} + \Delta E_l^f = -J' \cos(\phi_{l,\mathrm{event}} - \phi_{\boldsymbol{v},l} - 2n\pi), \tag{9}$$

where

$$E^*_{\mathrm{init}} = \begin{cases} E_{kl} & \text{if the initial pair energy derivative is positive} \\ -J' & \text{otherwise.} \end{cases}$$

The true lifting event corresponds to the earliest of the independent event times sampled for all the neighbors of the spin k. In the ECMC, the Monte Carlo time is continuous and proportional to the total displacement of the spins.

The ECMC creates then chains of successive and consistent finite displacements of different spins. The choice of the length ℓ of a chain, defined as the cumulative rotation angles about $\boldsymbol{v}$ of the chain, is free. For the XY model of planar rotators, $\boldsymbol{v}$ is uniquely defined as the axis perpendicular to the sense of rotation. For this reason, the ECMC around this axis is irreducible and the chain length ℓ in this model is best taken equal to the simulation time [15]. For the Heisenberg model, spin rotations must be about at least two axes in order to reach the entire configuration space. The resampling of the rotation axis is performed after the chain length ℓ is reached. All configurations of the chain sample the equilibrium distribution and any uniform subset of them yields valid observable averages. Observables may be integrated during the continuous evolution or, e.g., retrieved at regular intervals independent of the lifting events.

We have checked the correctness of the ECMC and obtained perfect agreement for the mean energy, the specific heat, and the susceptibility with the heat-bath algorithm [19,20] modified with the exchange Monte Carlo method (or parallel tempering) [5] (see Fig. 1).

III. DYNAMICAL SCALING EXPONENT

At the critical temperature T_c, the correlation length ξ of a model undergoing a second-order phase transition equals the system size L and the autocorrelation time of slow variables τ diverges as $\tau \sim L^z$, where z is the dynamical critical exponent. We define time in terms of sweeps: One ECMC sweep corresponds, on average, to N lifting events and one

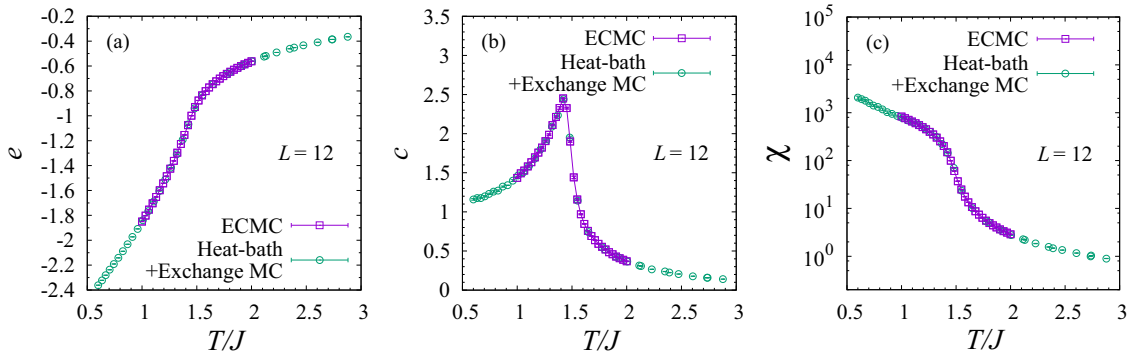


FIG. 1. (Color online) Temperature dependence of (a) the energy density $e = E/N$, (b) the specific heat c, and (c) the magnetic susceptibility χ of the three-dimensional Heisenberg model with $L = 12$. A chain length $\ell = N\pi/10$ is used.

LMC sweep to N attempted moves. For both algorithms, the complexity of one sweep is $O(N)$ and the CPU times used per sweep are roughly the same. Time autocorrelation functions are defined by

$$C_O(t) = \frac{\langle O(t'+t)O(t')\rangle - \langle O(t')\rangle^2}{\langle O^2(t')\rangle - \langle O(t')\rangle^2}, \tag{10}$$

where the angular brackets indicate the thermal average and t' is set sufficiently large for equilibration. The dynamical critical exponent of the LMC for the three-dimensional Heisenberg model was estimated from the autocorrelation function of the magnetization $\boldsymbol{M}$ as $z = 1.96(6)$ [21]. The overrelaxation algorithm [22,23] seems to give $z \simeq 1.10$ [21], which was obtained from the autocorrelation function of the magnetization, and the Wolff algorithm is believed to yield a value close to zero: $z \gtrsim 0$, a value obtained from the susceptibility autocorrelation function [24].

To evaluate the correlation time and the dynamical critical exponent for the ECMC, one must pay attention to the irreversible nature of the underlying Markov chain. During one event chain, spins all rotate in the same sense and the system undergoes global rotations with taking into account the thermal fluctuation. This results in fast oscillations of the magnetization $\boldsymbol{M}$ and a quick decay of its autocorrelation function that is insensitive to the system size (see Fig. 2) and even to the temperature. However, this effect is also visible for a trivial algorithm, which simply performs global rotations (see the inset of Fig. 2). The trivial algorithm satisfies global balance, but its correlation time is infinite, as it does not relax the energy. A similar effect appears in the ECMC for particle systems [9], which likewise is not characterized by the mean net displacement of particles. To characterize the speed of the ECMC, we consider the energy density and the susceptibility that we conjecture to be slow variables at the critical temperature. Both χ and e are insensitive to global rotations and do not oscillate.

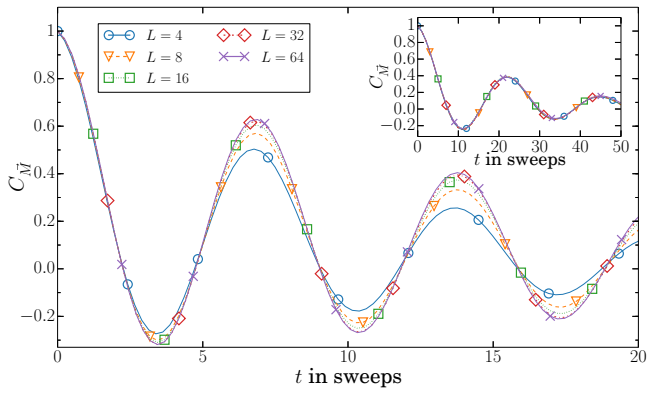


FIG. 2. (Color online) Autocorrelation function of magnetization $C_M(t)$ at the critical temperature for various system sizes. The inset shows the spin autocorrelation function of a trivial algorithm that only performs global rotations in spin space along the two axes.

As shown in Fig. 3, the autocorrelation functions both of the energy density and of the susceptibility are well approximated as a single exponential decay

$$C_\chi(t) = \exp(-t/\tau) \tag{11}$$

on essentially the same time scales. Furthermore, the finite-size behavior of the autocorrelation times indicates $z \simeq 1$ dynamical scaling. This z value is significantly less than for the LMC and very similar to the one obtained for overrelaxation methods, although the $z \simeq 0$ value of the cluster algorithm is not reached.

IV. DISCUSSION AND SUMMARY

The earliest application of lifting [16], the motion of a particle on a one-dimensional N-site lattice with periodic boundary conditions, already featured the decrease of the dynamical scaling exponent from $z = 2$ to $z = 1$ (the reduction of the mixing time from proportional to N^2 to proportional to N). To reach such reductions, the Markov chain must be irreversible. It was pointed out that the square-root decrease of the critical exponent was the optimal improvement [17]. The concepts of factorized Metropolis filters and of infinitesimal moves brought irreversible lifting algorithms to general N-body systems, although only finite speed-ups were realized in the $N \to \infty$ limit. The three-dimensional Heisenberg model however seems to be an ECMC application with a lowered critical dynamical exponent. Our observation relies on the hypotheses that the energy and the susceptibility are indeed slow variables and that the observed decay of the autocorrelation function continues for larger times. However, in Fig. 3, a crossover from $z = 1$ back to $z = 2$ as it was

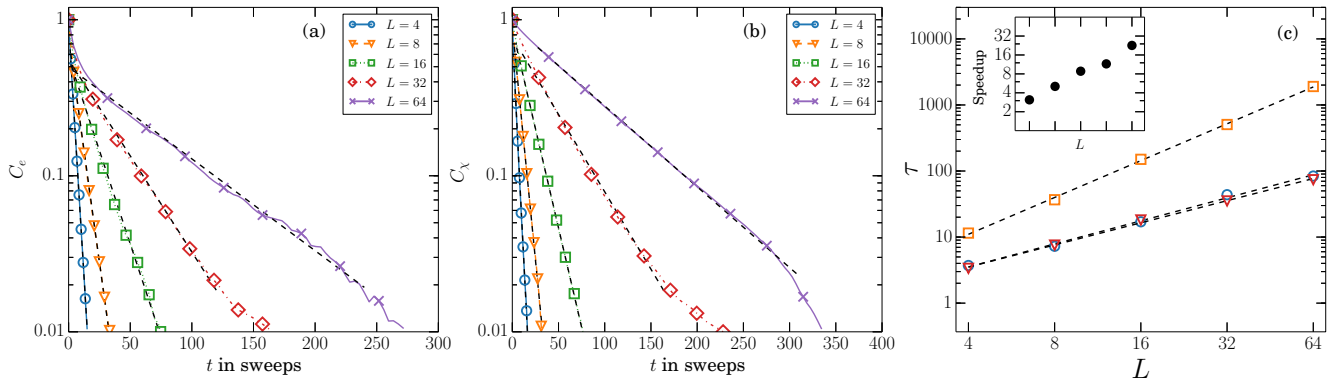


FIG. 3. (Color online) Autocorrelation functions and time constants of the ECMC for the three-dimensional Heisenberg model at its critical point $\beta = 0.693$: (a) energy density autocorrelation function C_e for system sizes $4^3, 8^3, \ldots, 64^3$, (b) susceptibility autocorrelation function C_χ for the ECMC for the three-dimensional Heisenberg system sizes $4^3, 8^3, \ldots, 64^3$, and (c) scaling of the autocorrelation time τ_χ (τ_e) of the susceptibility χ (energy density e) with system size L for the ECMC [blue circles (red triangles)] and of the autocorrelation time of the susceptibility for the LMC (yellow squares). Error bars are smaller than the markers size. The inset shows the speedup for the susceptibility χ in comparison to the LMC for system sizes $4^3, 8^3, \ldots, 64^3$.

observed in the XY model after approximately five sweeps [15] appears unlikely to arise after hundreds of sweeps. The dynamical critical exponent $z \approx 1$ represents a maximal improvement with respect to the $z \approx 2$ of the LMC, supposing again that the theorems of Ref. [17] apply to infinitesimal Markov chains.

In summary, we have successfully applied the ECMC to the Heisenberg model in three dimensions. The ECMC shows considerable promise for spin models and the numerical data presented in this paper allow us to formulate the exciting conjecture that the dynamical critical exponent for the Heisenberg model is $z \simeq 1$. The ECMC is also applicable to frustrated magnets and spin glasses, which involve antiferromagnetic interactions and/or quenched disorder. Our preliminary study indicates that the ECMC algorithm is also useful for a Heisenberg spin glass model. The ECMC can be easily combined with other algorithms such as the exchange Monte Carlo method and the overrelaxation algorithm in the usual manner. This may allow the investigation of the three-dimensional Heisenberg spin glass model in the low-temperature region. Large-scale simulations in this direction are currently in progress. It would be very interesting to understand why the ECMC is so much more successful in the Heisenberg model than both in hard and soft disks and in the XY model.

ACKNOWLEDGMENTS

Y.N. and K.H. thank S. Hoshino and M. J. Miyama for useful discussions and J. Takahashi and Y. Sakai for carefully reading the manuscript. This research was supported by Grants-in-Aid for Scientific Research from the JSPS, Japan (Grants No. 25120010 and No. 25610102) and JSPS Core-to-Core program "Nonequilibrium dynamics of soft matter and information." This work was granted access to the HPC resources of MesoPSL financed by the Region Ile de France and the project Equip@Meso (Reference No. ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche.

[1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[2] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
[3] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[4] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[5] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[6] H. Suwa and S. Todo, Phys. Rev. Lett. **105**, 120603 (2010).
[7] K. S. Turitsyn, M. Chertkov, and M. Vucelja, Physica D **240**, 410 (2011).
[8] H. C. M. Fernandes and M. Weigel, Comput. Phys. Commun. **182**, 1856 (2011).
[9] E. P. Bernard, W. Krauth, and D. B. Wilson, Phys. Rev. E **80**, 056704 (2009).
[10] E. P. Bernard and W. Krauth, Phys. Rev. Lett. **107**, 155704 (2011).
[11] M. Isobe and W. Krauth, J. Chem. Phys. **143**, 084509 (2015).
[12] M. Michel, S. C. Kapfer, and W. Krauth, J. Chem. Phys. **140**, 054116 (2014).
[13] E. A. J. F. Peters and G. de With, Phys. Rev. E **85**, 026703 (2012).
[14] S. C. Kapfer and W. Krauth, Phys. Rev. Lett. **114**, 035702 (2015).
[15] M. Michel, J. Mayer, and W. Krauth, Europhys. Lett. **112**, 20003 (2015).
[16] P. Diaconis, S. Holmes, and R. M. Neal, Ann. Appl. Probab. **10**, 726 (2000).
[17] F. Chen, L. Lovász, and I. Pak, *Proceedings of the 17th Annual ACM Symposium on Theory of Computing* (ACM, New York, 1999), p. 275.
[18] K. Chen, A. M. Ferrenberg, and D. P. Landau, Phys. Rev. B **48**, 3249 (1993).

[19] Y. Miyatake, M. Yamamoto, J. J. Kim, M. Toyonaga, and O. Nagai, J. Phys. C **19**, 2539 (1986).
[20] J. A. Olive, A. P. Young, and D. Sherrington, Phys. Rev. B **34**, 6341 (1986).
[21] P. Peczak and D. P. Landau, Phys. Rev. B **47**, 14260 (1993).
[22] M. Creutz, Phys. Rev. D **36**, 515 (1987).
[23] F. R. Brown and T. J. Woch, Phys. Rev. Lett. **58**, 2394 (1987).
[24] C. Holm and W. Janke, Phys. Rev. B **48**, 936 (1993).

Résumé

Cette thèse porte sur le développement et l'application en physique statistique d'un nouveau paradigme pour les méthodes sans rejet de Monte-Carlo par chaînes de Markov irréversibles, grâce à la mise en œuvre du filtre factorisé de Metropolis et du concept de lifting.

Les deux premiers chapitres présentent la méthode de Monte-Carlo et ses différentes applications à des problèmes de physique statistique. Une des principales limites de ces méthodes se rencontre dans le voisinage des transitions de phase, où des phénomènes de ralentissement dynamique entravent fortement la thermalisation des systèmes.

Le troisième chapitre présente la nouvelle classe des algorithmes de Metropolis factorisés et irréversibles. Se fondant sur le concept de lifting des chaînes de Markov, le filtre factorisé de Metropolis permet de décomposer un potentiel multidimensionnel en plusieurs autres unidimensionnels. De là, il est possible de définir un algorithme sans rejet de Monte-Carlo par chaînes de Markov irréversibles. Le quatrième chapitre examine les performances de ce nouvel algorithme dans une grande variété de systèmes. Des accélérations du temps de thermalisation sont observées dans des systèmes bidimensionnels de particules molles, des systèmes bidimensionnels de spins XY ferromagnétiques et des systèmes tridimensionnels de verres de spins XY. Finalement, une réduction importante du ralentissement critique est exposée pour un système tridimensionnel de spins Heisenberg ferromagnétiques.

Mots-clefs : Méthode de Monte-Carlo - Chaînes de Markov irréversibles- Systèmes désordonnés - Systèmes bidimensionnels - Systèmes de spins classiques - Verre de spins.

Abstract

This thesis deals with the development and application in statistical physics of a general framework for irreversible and rejection-free Markov-chain Monte Carlo methods, through the implementation of the factorized Metropolis filter and the lifting concept.

The first two chapters present the Markov-chain Monte Carlo method and its different implementations in statistical physics. One of the main limitations of Markov-chain Monte Carlo methods arises around phase transitions, where phenomena of dynamical slowing down greatly impede the thermalization of the system.

The third chapter introduces the new class of irreversible factorized Metropolis algorithms. Building on the concept of lifting of Markov chains, the factorized Metropolis filter allows to decompose a multidimensional potential into several unidimensional ones. From there, it is possible to define a rejection-free and completely irreversible Markov-chain Monte Carlo algorithm. The fourth chapter reviews the performance of the irreversible factorized algorithm in a wide variety of systems. Clear accelerations of the thermalization time are observed in bidimensional soft-particle systems, bidimensional ferromagnetic XY spin systems and three-dimensional XY spin glasses. Finally, an important reduction of the critical slowing down is exhibited in three-dimensional ferromagnetic Heisenberg spin systems.

Keywords: Monte Carlo method - Irreversible Markov chains - Disordered systems - Two-dimensional systems - Classical spin systems - Spin glass.

Résumé

Cette thèse porte sur le développement et l'application en physique statistique d'un nouveau paradigme pour les méthodes sans rejet de Monte-Carlo par chaînes de Markov irréversibles, grâce à la mise en œuvre du filtre factorisé de Metropolis et du concept de lifting.

Les deux premiers chapitres présentent la méthode de Monte-Carlo et ses différentes applications à des problèmes de physique statistique. Une des principales limites de ces méthodes se rencontre dans le voisinage des transitions de phase, où des phénomènes de ralentissement dynamique entravent fortement la thermalisation des systèmes.

Le troisième chapitre présente la nouvelle classe des algorithmes de Metropolis factorisés et irréversibles. Se fondant sur le concept de lifting des chaînes de Markov, le filtre factorisé de Metropolis permet de décomposer un potentiel multidimensionnel en plusieurs autres unidimensionnels. De là, il est possible de définir un algorithme sans rejet de Monte-Carlo par chaînes de Markov irréversibles. Le quatrième chapitre examine les performances de ce nouvel algorithme dans une grande variété de systèmes. Des accélérations du temps de thermalisation sont observées dans des systèmes bidimensionnels de particules molles, des systèmes bidimensionnels de spins XY ferromagnétiques et des systèmes tridimensionnels de verres de spins XY. Finalement, une réduction importante du ralentissement critique est exposée pour un système tridimensionnel de spins Heisenberg ferromagnétiques.

Mots Clés

Méthode de Monte-Carlo, Chaînes de Markov irréversibles, Systèmes désordonnés, Systèmes bidimensionnels, Systèmes de spin classiques, Verre de spins

Abstract

This thesis deals with the development and application in statistical physics of a general framework for irreversible and rejection-free Markov-chain Monte Carlo methods, through the implementation of the factorized Metropolis filter and the lifting concept.

The first two chapters present the Markov-chain Monte Carlo method and its different implementations in statistical physics. One of the main limitations of Markov-chain Monte Carlo methods arises around phase transitions, where phenomena of dynamical slowing down greatly impede the thermalization of the system.

The third chapter introduces the new class of irreversible factorized Metropolis algorithms. Building on the concept of lifting of Markov chains, the factorized Metropolis filter allows to decompose a multidimensional potential into several unidimensional ones. From there, it is possible to define a rejection-free and completely irreversible Markov-chain Monte Carlo algorithm. The fourth chapter reviews the performance of the irreversible factorized algorithm in a wide variety of systems. Clear accelerations of the thermalization time are observed in bidimensional soft-particle systems, bidimensional ferromagnetic XY spin systems and three-dimensional XY spin glasses. Finally, an important reduction of the critical slowing down is exhibited in three-dimensional ferromagnetic Heisenberg spin systems.

Keywords

Monte Carlo method, Irreversible Markov chains, Disordered systems, Two-dimensional systems, Classical spin systems, Spin glass