Non-reversible Markov chains: From statistical mechanics to chemical physics

Werner Krauth
Department of Physics, Ecole normale supérieure Paris, France
Rudolf Peierls Centre for Theoretical Physics, University of Oxford, UK
Simons Center for Computational Physical Chemistry, New York University, USA

21 May 2024
Simons Center for Computational Physical Chemistry
New York University (USA)

This work was supported by a grant from the Simons Foundation
This work was supported by the Alexander von Humboldt Foundation

Tutorial session: Wednesday 29 May 2024, 9am—4pm (*sic*)

Département de Physique
École normale supérieure

Markov chains (1/4)

URNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUN

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

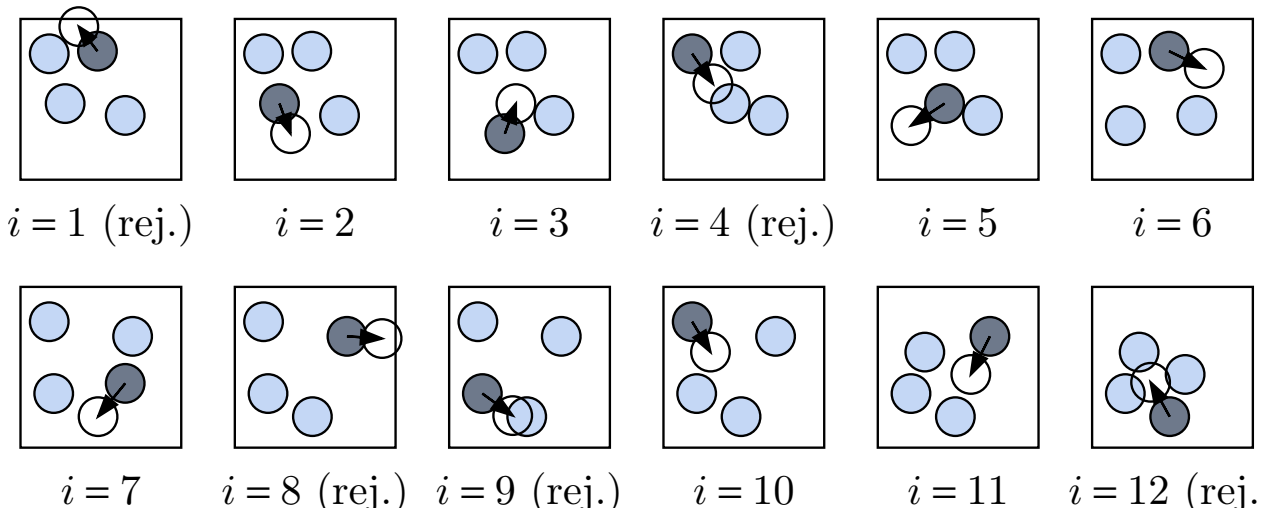
AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

Markov chains (2/4)



- Metropolis algorithm (1953, mod.)

Département de Physique
École normale supérieure

Molecular dynamics

PHYSICAL REVIEW VOLUME 127, NUMBER 2 JULY 15, 1962

Phase Transition in Elastic Disks*

B. J. Alder and T. E. Wainwright
University of California, Lawrence Radiation Laboratory, Livermore, California

(Received October 30, 1961)

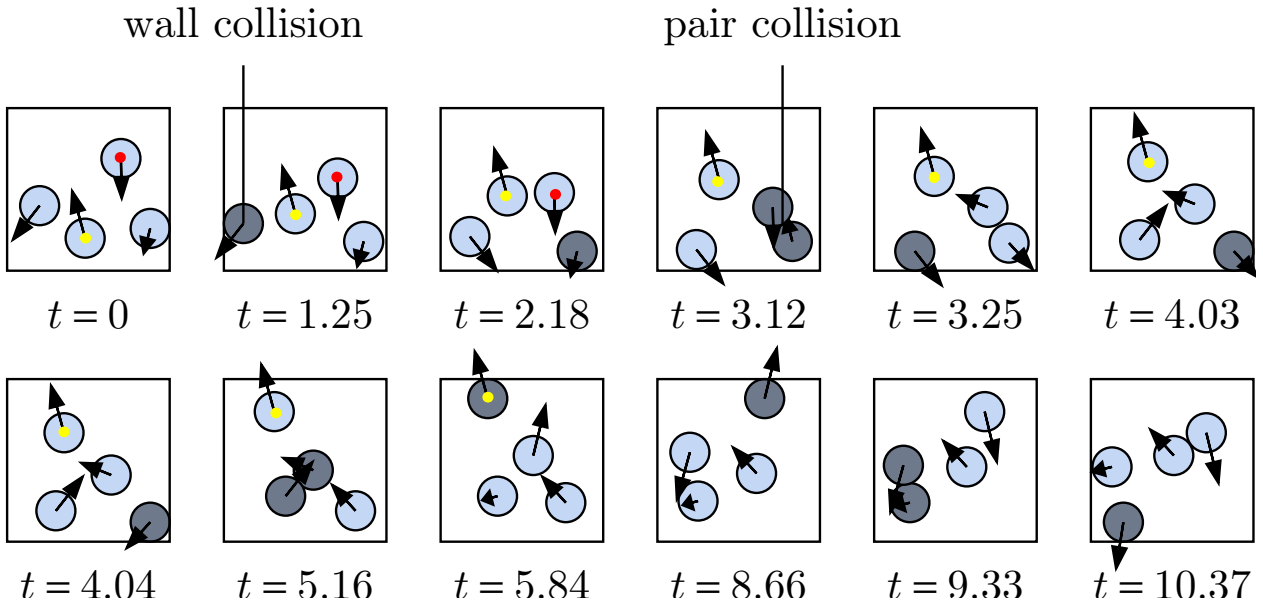
The study of a two-dimensional system consisting of 870 hard-disk particles in the phase-transition region has shown that the isotherm has a van der Waals-like loop. The density change across the transition is about 4% and the corresponding entropy change is small.

A STUDY has been made of a two-dimensional system consisting of 870 hard-disk particles. Simultaneous motions of the particles have been calculated by means of an electronic computer as described previously.[1] The disks were again placed in a periodically repeated rectangular array. The computer program

interchanges it was not possible to average the two branches.

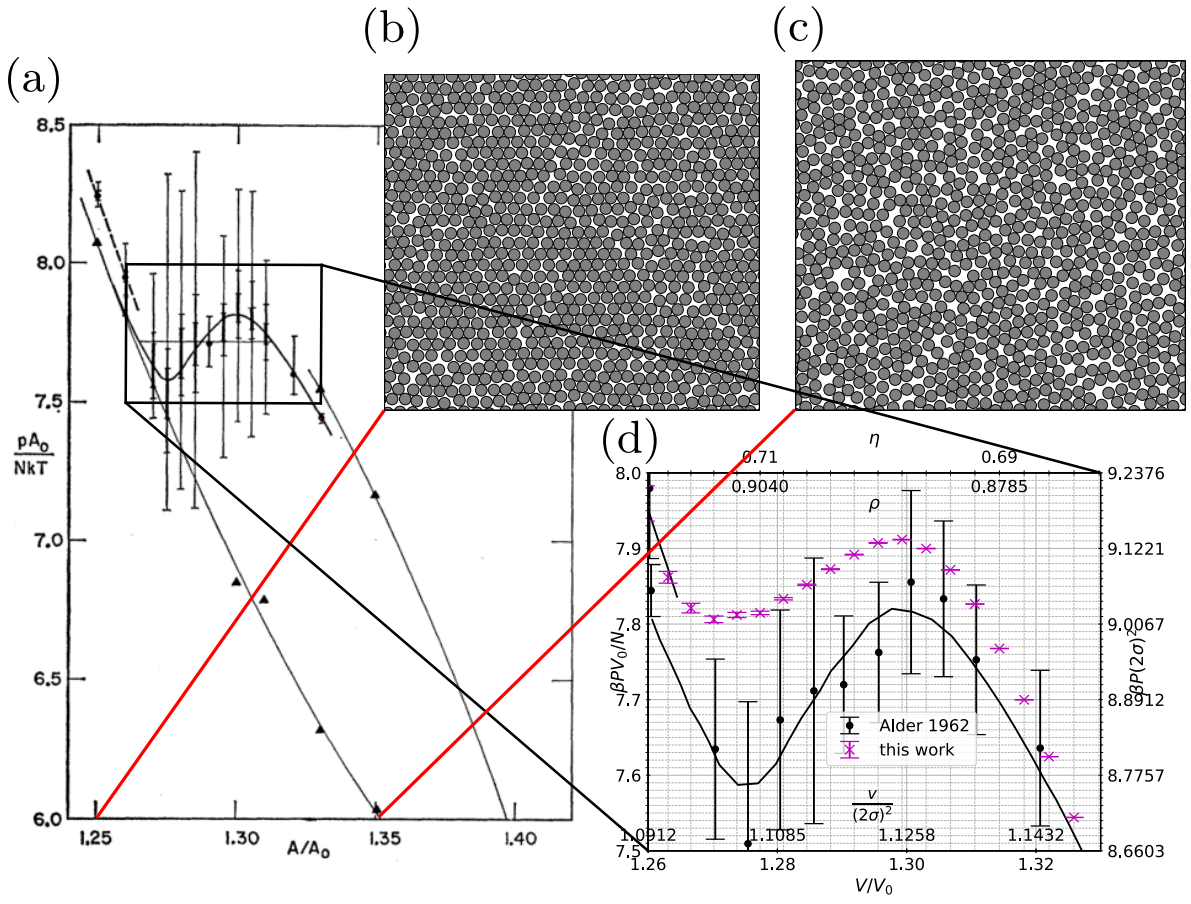
Two-dimensional systems were then studied, since the number of particles required to form clusters of particles of one phase of any given diameter is less than in three dimensions. Thus, an 870 hard-disk system is

Molecular dynamics



- Alder–Wainwright (1957)

Samples



- Alder & Wainwright 1962, see Li et al. (2022)

From ‘numerical work’ to Kosterlitz–Thouless

Ordering, metastability and phase transitions in two-dimensional systems

J M Kosterlitz and D J Thouless
Department of Mathematical Physics, University of Birmingham, Birmingham B15 2TT, UK

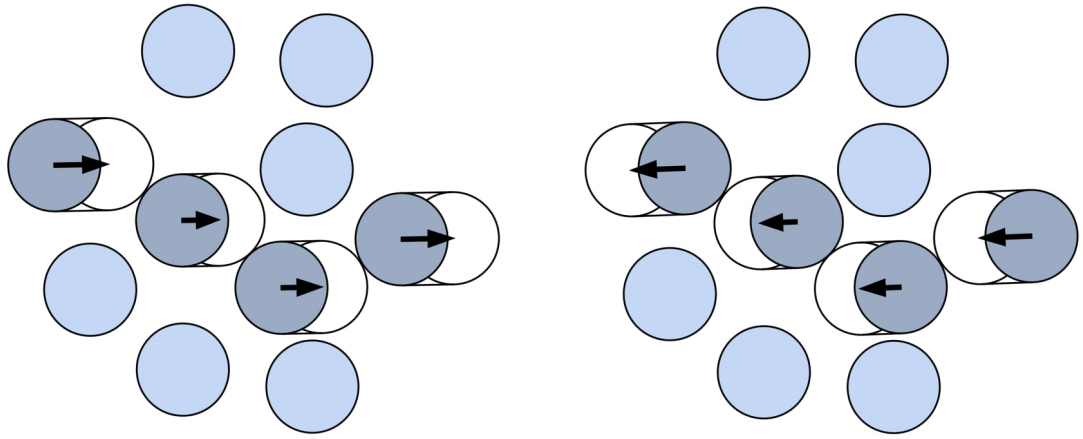
Received 13 November 1972

1. Introduction

Peierls (1935) has argued that thermal motion of long-wavelength phonons will destroy the long-range order of a two-dimensional solid in the sense that the mean square deviation of an atom from its equilibrium position increases logarithmically with the size of the system, and the Bragg peaks of the diffraction pattern formed by the system are broad instead of sharp. The absence of long-range order of this simple form has been shown by Mermin (1968) using rigorous inequalities. Similar arguments can be used to show that there is no spontaneous magnetization in a two-dimensional magnet with spins with more than one degree of freedom (Mermin and Wagner 1966) and that the expectation value of the superfluid order parameter in a two-dimensional Bose fluid is zero (Hohenberg 1967).

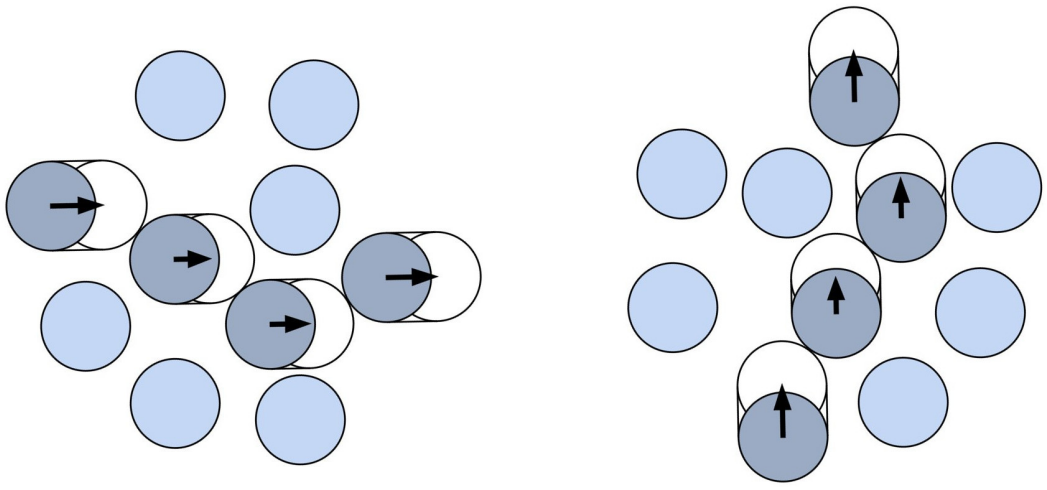
On the other hand there is inconclusive evidence from the numerical work on a two-dimensional system of hard discs by Alder and Wainwright (1962) of a phase transition between a gaseous and solid state. Stanley and Kaplan (1966) found that high-temperature series expansions for two-dimensional spin models indicated a phase

Event-chain Monte Carlo (reversible)



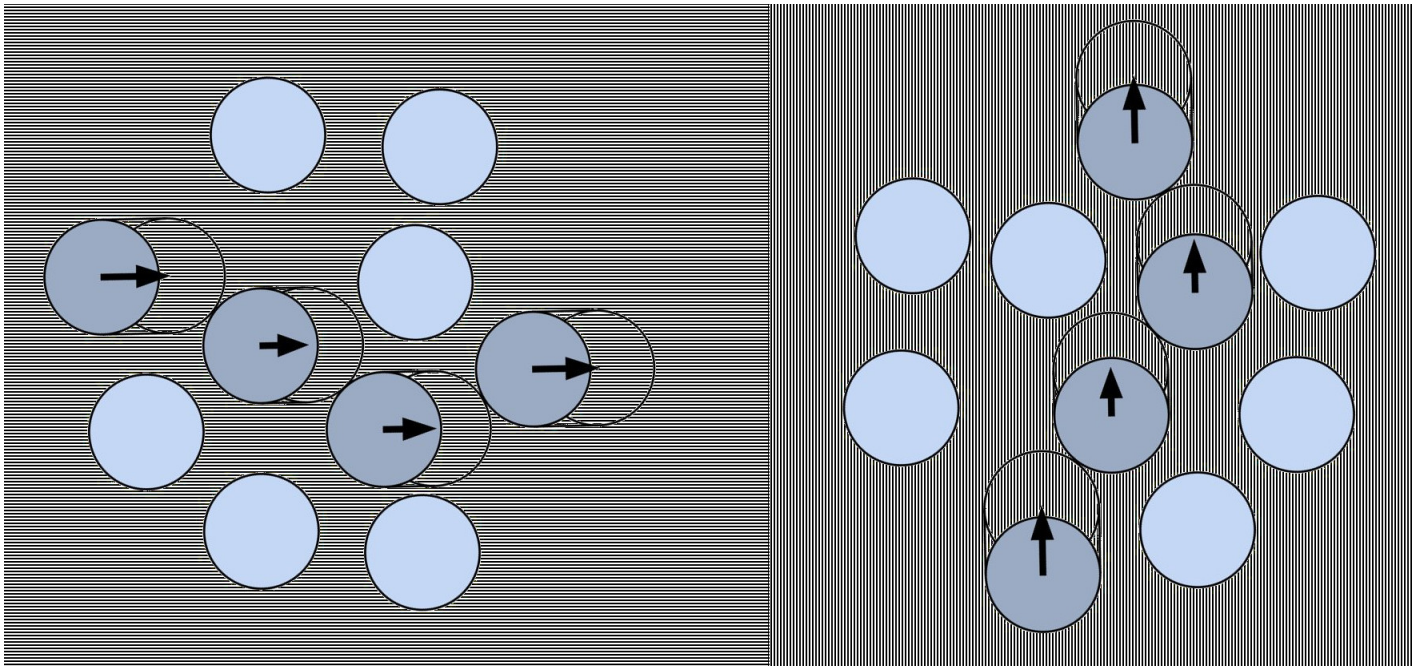
- total length of arrows $= \ell$, fixed (sic!)

Event-chain Monte Carlo (non-reversible)



- Box of size $L \times L$, periodic boundary conditions, N disks.
- Event-chain Monte Carlo: Bernard, Krauth, Wilson (2009).

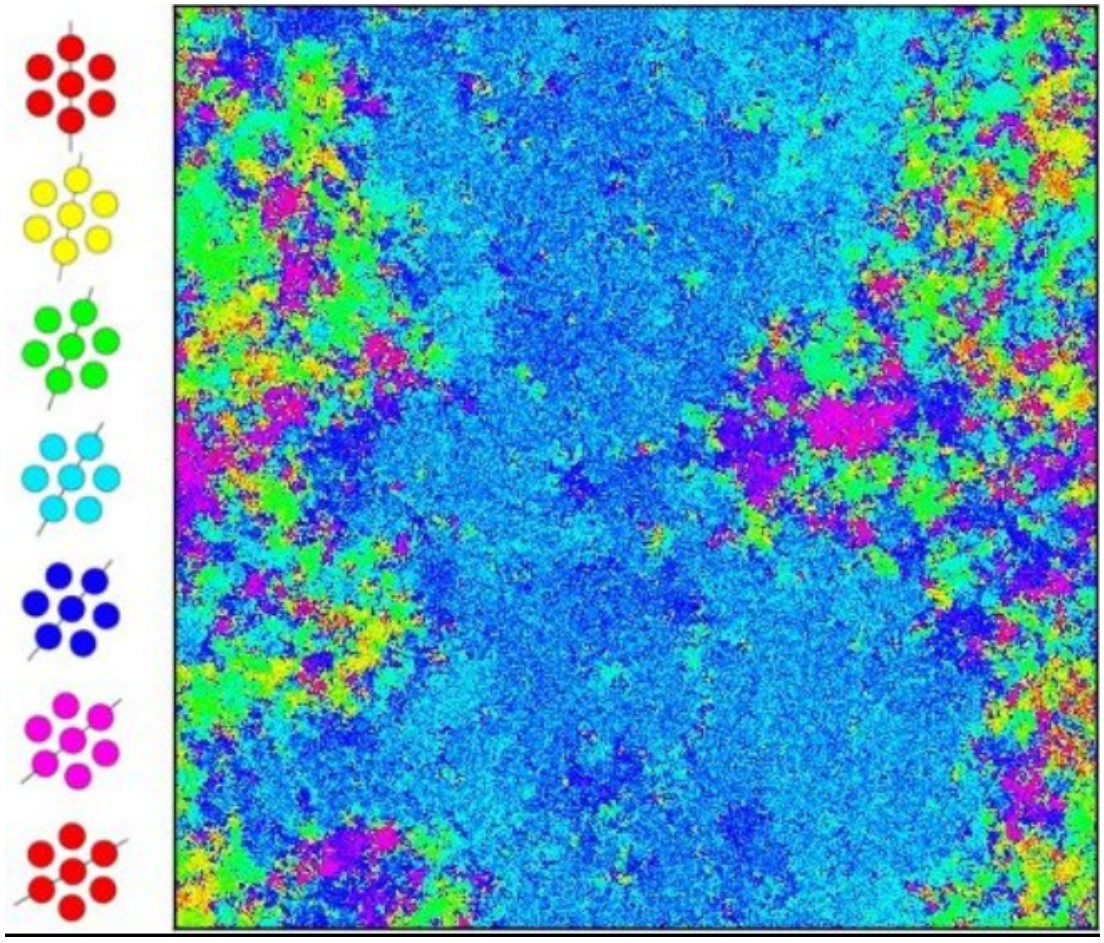
Event-chain Monte Carlo and molecular dynamics



- Box of size $L \times L$, periodic boundary conditions, N disks.
- Event-chain Monte Carlo: Bernard, Krauth, Wilson (2009).
- Molecular dynamics 'on a grating'.
- Hamiltonian Monte Carlo 'on a grating'.

Département de Physique
École normale supérieure

Equilibrium samples from non-equilibrium MCMC



- 10^6 disks (Bernard & Krauth 2011, Li et al. 2022)

Département de Physique
École normale supérieure

Markov chains (3/4)

- Sample space Ω (e.g. hard disks, water molecules, quarks, ...)
- Markov chain $\leftarrow$ Sequence of random variables
  $(X_0 \sim \pi^{\{0\}}, X_1 \sim \pi^{\{1\}}, X_2 \sim \pi^{\{2\}} \dots)$
  X_{t+1} depends only on X_t, t is a 'time'
- Transition matrix P:
  - P_{ij}: conditional probability to move from sample i to sample j.
  - $\pi^{\{t+1\}} = \pi^{\{t\}} P$: Evolve probability distribution at time t to probability distribution at time $t+1$ (with $\pi^{\{t\}}, t > 0$ often non-explicit,even for $t \to \infty$).
- Move set $\mathcal{L}$: ... from which moves are sampled.
- Equilibrium distribution π: Satisfies global balance:

$$\pi_i = \sum_{j \in \Omega} \pi_j P_{ji} \quad \forall i \in \Omega.$$

  NB: P irreducible $\implies$ π unique.
- Aperiodicity: Absence of cycles. P irreducible and aperiodic:

$$\pi^{\{t\}} \to \pi \quad \text{for } t \to \infty$$

Département de Physique
École normale supérieure

Total variation distance, mixing time

- Total variation distance:

$$||\pi^{\{t\}} - \pi||_{\mathrm{TV}} = \max_{A \subset \Omega} |\pi^{\{t\}}(A) - \pi(A)| = \frac{1}{2} \sum_{i \in \Omega} |\pi_i^{\{t\}} - \pi_i|.$$

- Distance:

$$d(t) = \max_{\pi^{\{0\}}} ||\pi^{\{t\}}(\pi^{\{0\}}) - \pi||_{\mathrm{TV}}$$

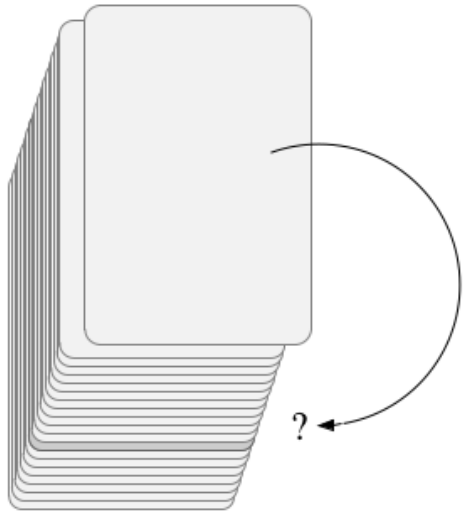
- Mixing time:

$$t_{\mathrm{mix}}(\epsilon) = \min\{t : d(t) \leq \epsilon\}$$

- Usually $\epsilon = 1/4$ is taken, $\epsilon = 1/e$ would be better.

Département de Physique
École normale supérieure

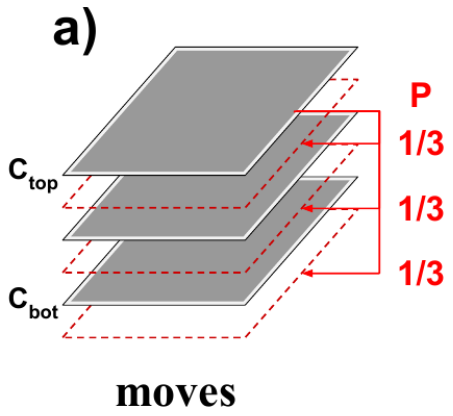
Shuffling of cards 1/5



- $\Omega_N^{\text{shuffle}} = \{\text{Permutations of } \{1, \ldots, N\}\}$
- For $N = 3$:
  $\Omega_3^{\text{shuffle}} = \{1 \equiv \{1,2,3\}, 2 \equiv \{1,3,2\}, 3 \equiv \{2,1,3\}, 4 \equiv \{2,3,1\}, 5 \equiv \{3,1,2\}, 6 \equiv \{3,2,1\}\}$.
- $\pi^{t=0} = \delta((1,,\ldots,N))$

Département de Physique
École normale supérieure

Shuffling of cards 2/5

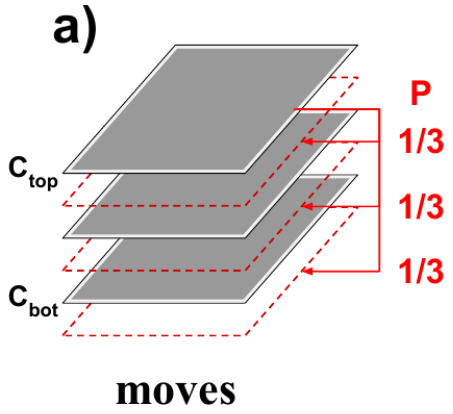


procedure top-to-random
input $\{c_1, \ldots, c_n\}$
$i \leftarrow$ choice$(\{1, \ldots, n\})$
$\{\hat{c}_1, \ldots, \hat{c}_n\} \leftarrow \{c_2, \ldots, c_i, c_1, c_{i+1}, \ldots, c_n\}$
output $\{\hat{c}_1, \ldots, \hat{c}_n\}$

- Insert upper card (c_1) after card i and before card $i+1$
- NB: if $i = 1$, put it back on top.

Département de Physique
École normale supérieure

Shuffling of cards 3/5



- $\Omega_3^{\text{shuffle}} = \{1 \equiv \{1,2,3\}, 2 \equiv \{1,3,2\}, 3 \equiv \{2,1,3\}, 4 \equiv \{2,3,1\}, 5 \equiv \{3,1,2\}, 6 \equiv \{3,2,1\}\}$.
-

$$P = \frac{1}{3}\begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

Département de Physique
École normale supérieure

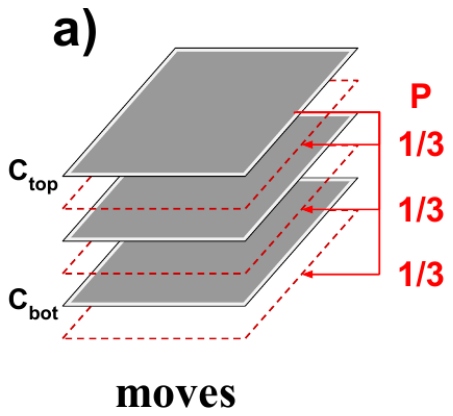
Shuffling of cards 4/5

- $$P_3^{\text{shuffle}} = \frac{1}{3}\begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$
- Eigenvalues of P_N^{shuffle}: $0, \frac{1}{N}, \frac{2}{N}, \ldots, 1 - \frac{2}{N}, 1$
- Degeneracies:

$$N = 2 : [1, 0, 1]$$
$$N = 3 : [2, 3, 0, 1]$$
$$N = 4 : [9, 8, 6, 0, 1]$$
$$N = 5 : [44, 45, 20, 10, 0, 1]$$
$$N = 6 : [265, 264, 135, 40, 15, 0, 1]$$
$$N = 7 : [1854, 1855, 924, 315, 70, 21, 0, 1]$$

Département de Physique
École normale supérieure

Shuffling of cards 5/5

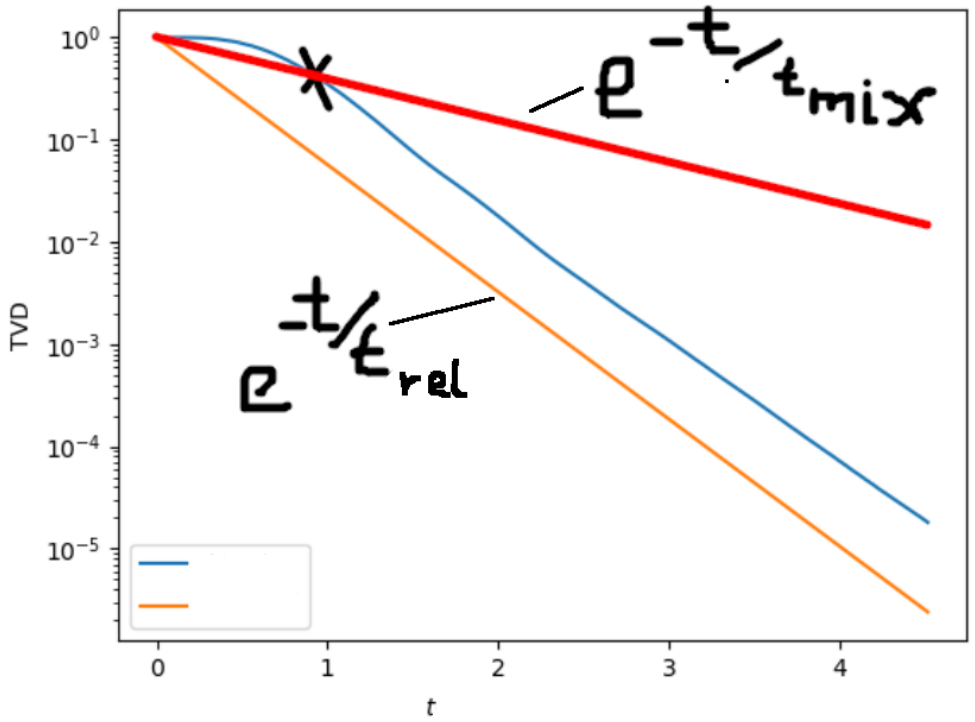


procedure `top2random-stop`
input $\{c_1, \ldots, c_n\}$
$c_{\text{first-n}} \leftarrow c_{\text{n}}$
for $t = 1, 2, \ldots$ **do**
$\tilde{c}_1 \leftarrow c_1$
$\{c_1, \ldots, c_n\} \leftarrow$ `top2random`$(\{c_1, \ldots, c_n\})$
if $(\tilde{c}_1 = c_{\text{first-n}})$ **break**
output $\{c_1, \ldots, c_n, t\}$

- Expected running time: $n \log n$.
- Time scale $n \log n$ larger than inverse gap $n/2$.

Département de Physique
École normale supérieure

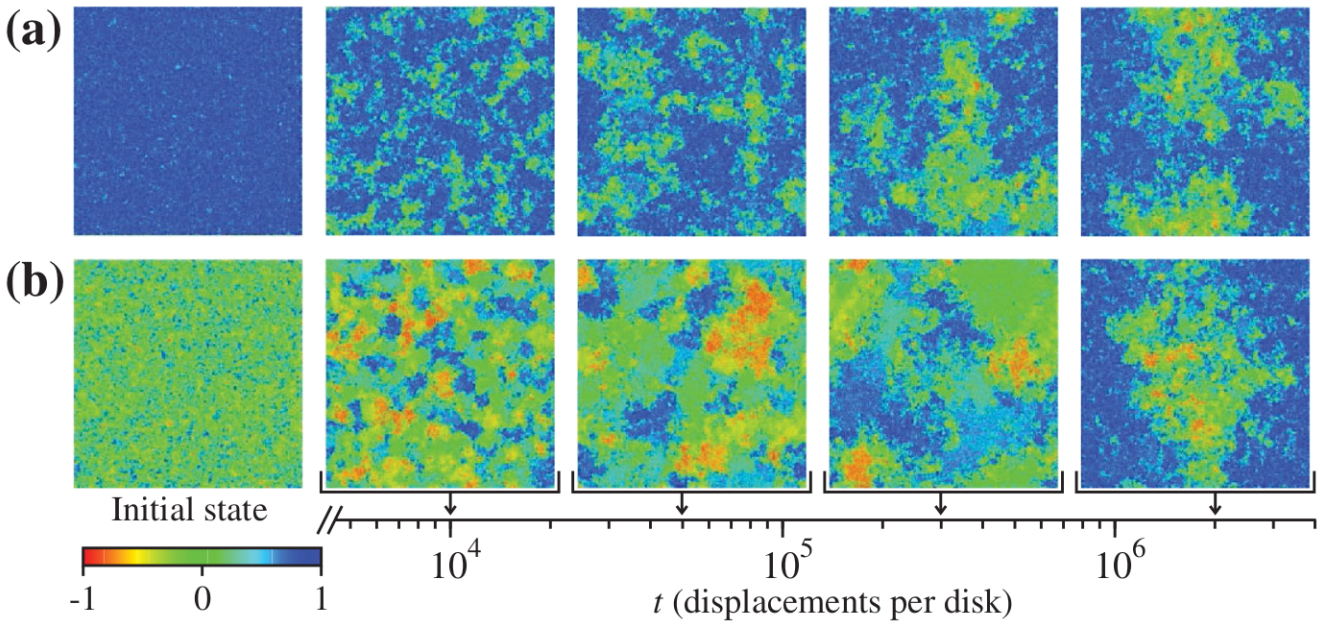
Mixing and Relaxation



- $t_{\text{mix}} = ||\pi^{\{t_{\text{mix}}\}} - \pi||_{\text{TV}} = 1/e$, (non-asymptotic time scale).
- $t_{\text{rel}} =$ inverse gap, (asymptotic time scale).
- $t_{\text{mix}} \gg t_{\text{rel}}$ leads to cutoff phenomenon.
- Aldous–Diaconis (1986)
- Diaconis–Fill-Pitman (1992)

Département de Physique
École normale supérieure

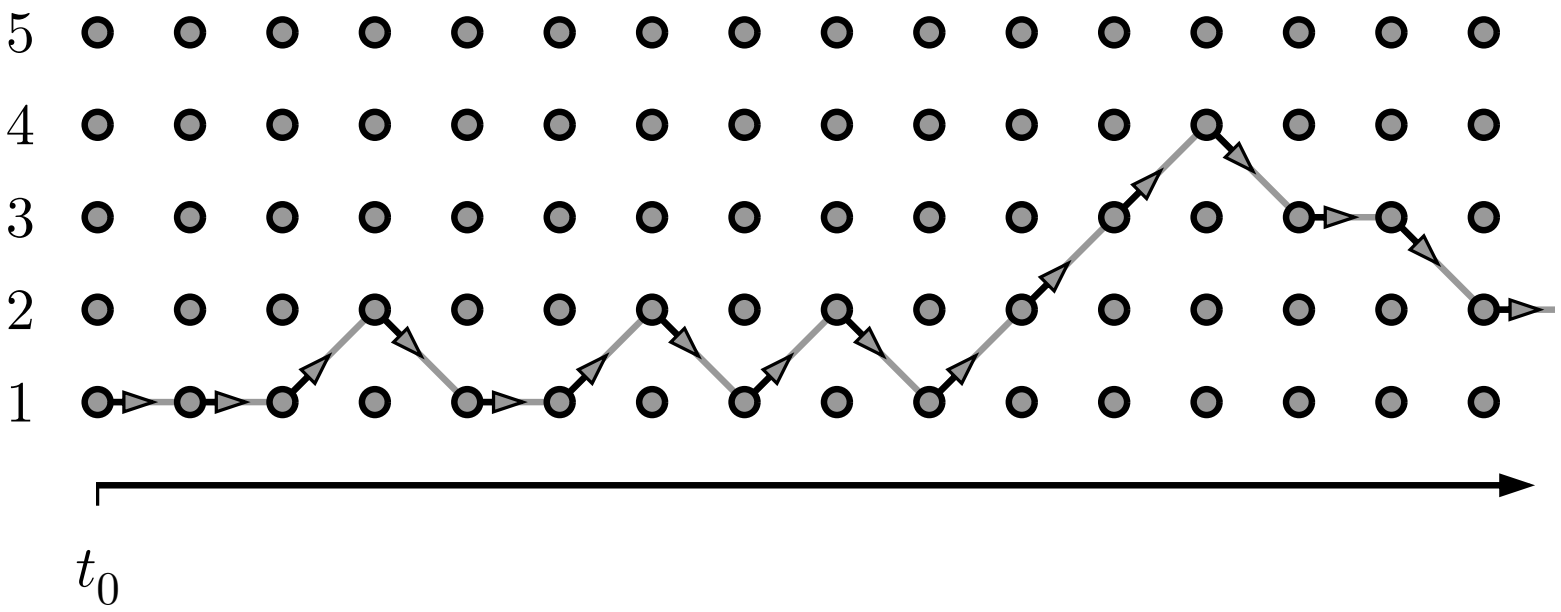
Coarsening: A non-asymptotic time scale



- Coarsening in hard disks (from Bernard & Krauth 2011)...
- ... an example of a non-asymptotic mixing-time scale

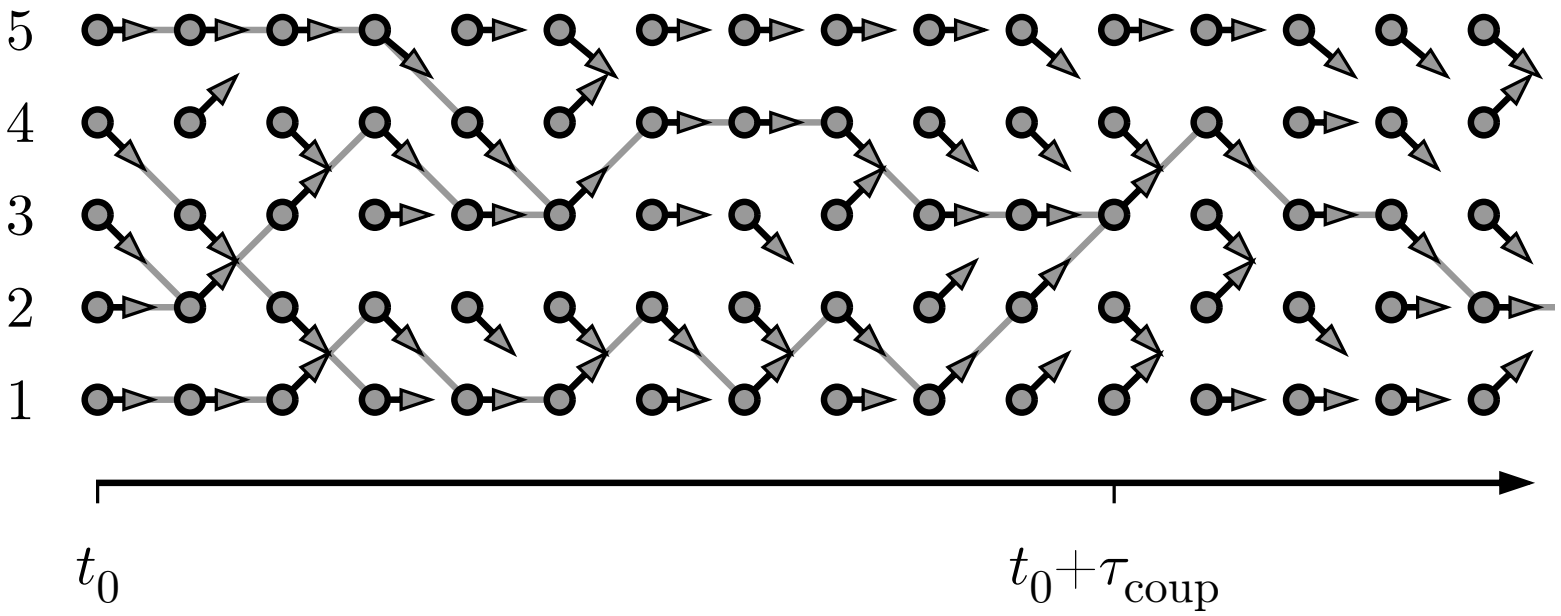
Département de Physique
École normale supérieure

Markov chains 4/4



- Configuration c_t, move δ_t.
- Set $t_0 = 0$.

Markov chain (random maps), coupling 1/3



- Each configuration has its move at each time step.
- Coupling (Doeblin, 1930s).

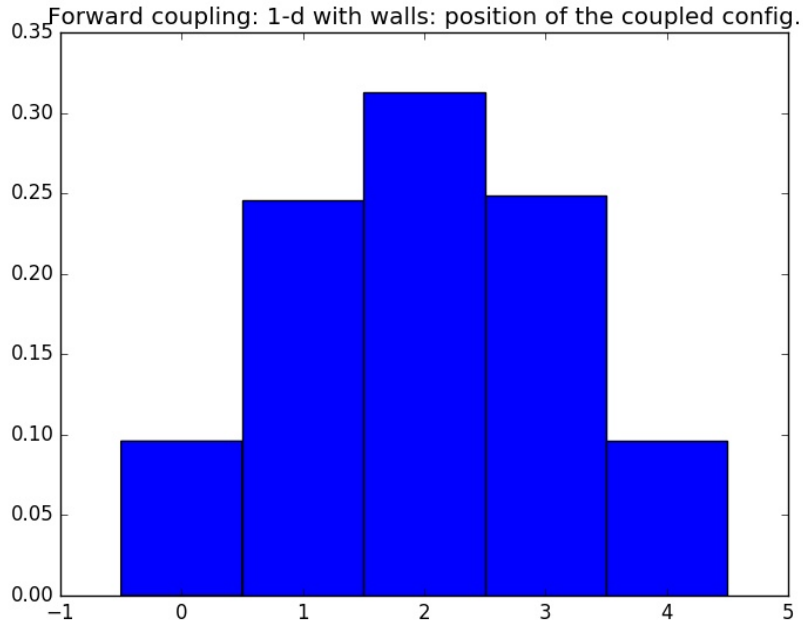
Markov chain (random maps), coupling 2/3

procedure `forward-coupling`
$\mathcal{P} \leftarrow \{1, \ldots, N\}$
$t \leftarrow 0$
while True:
$$\begin{cases} t \leftarrow t+1 \\ \mathcal{P} \leftarrow \{\min\left[\max(b + \textbf{choice}\{-1,+1\}, 1), N\right] \text{ for } b \in \mathcal{P}\} \\ \textbf{if } |\mathcal{P}| = 1\textbf{: break} \end{cases}$$
output $\mathcal{P}, t$ (position, time of coupling)

- Position of coupling not uniform.
- Coupling time larger than mixing time.

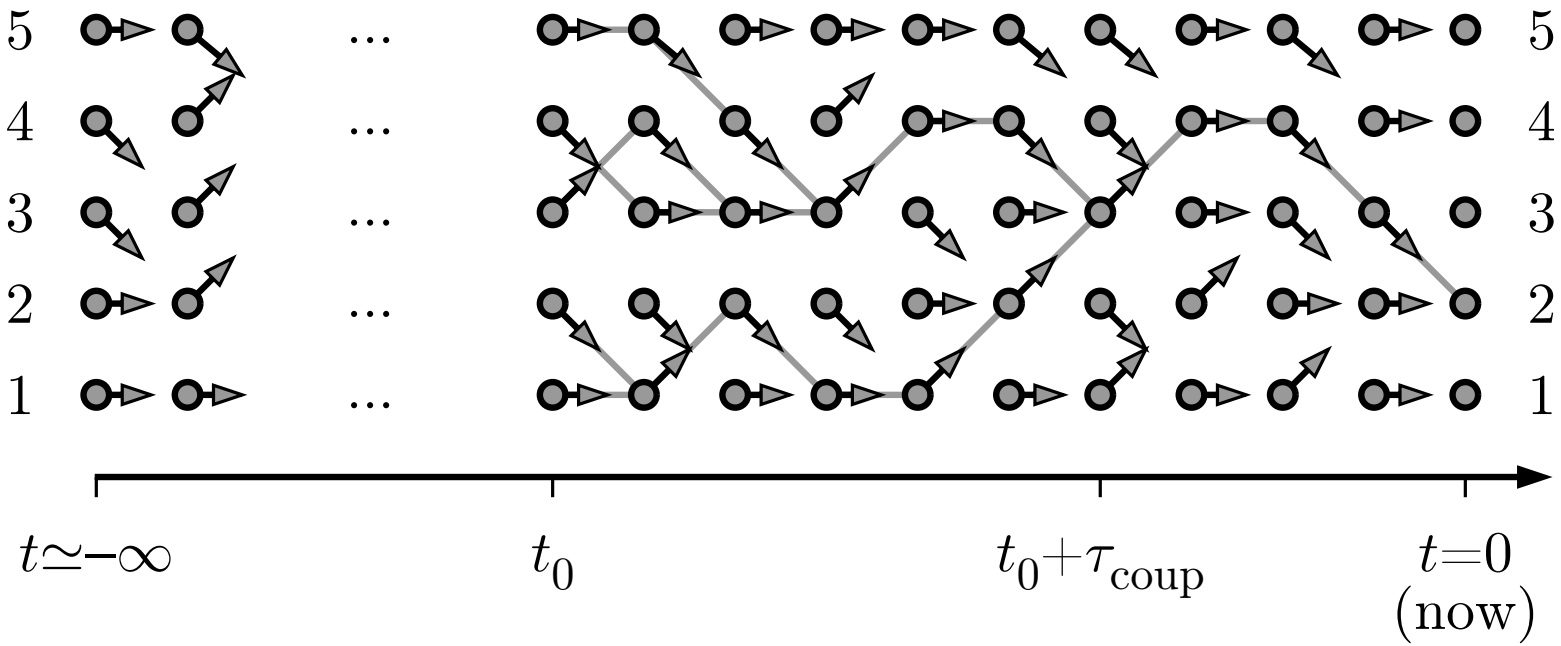
Département de Physique
École normale supérieure

Markov chain (random maps), coupling 3/3



- Histogram of coupling position.

Coupling from the past 1/3



- Starting an MCMC simulation at $t = -\infty$
- Propp & Wilson (1997)

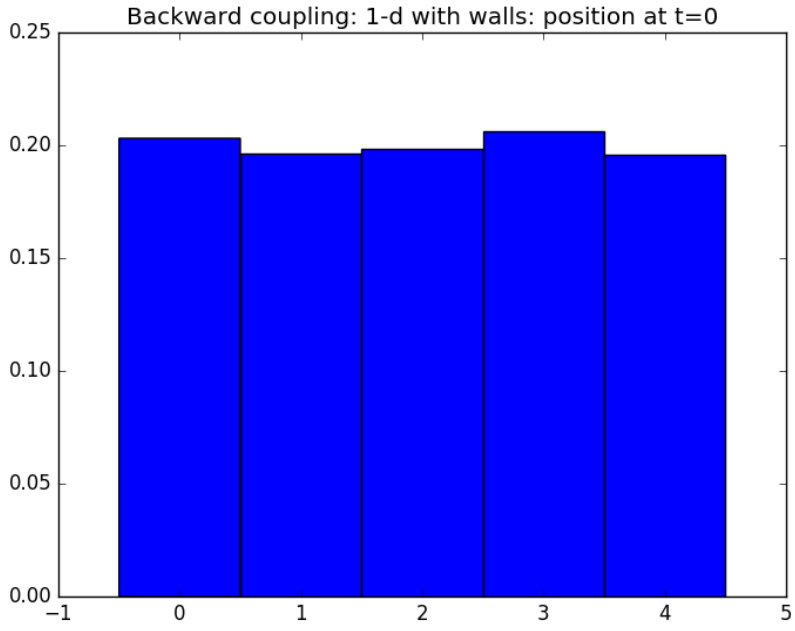
Coupling from the past 2/3

procedure `coupling-from-past`
$t_{\text{tot}} \leftarrow 0$
while True:
$\left\{\begin{array}{l} t_{\text{tot}} \leftarrow t_{\text{tot}} - 1 \\ \mathcal{A}_{t_{\text{tot}}} \leftarrow \texttt{draw-arrows}\ (\text{draw arrows at time } t_{\text{tot}}) \\ \mathcal{P} \leftarrow \{1, \ldots, N\} \\ \textbf{for } t = t_{\text{tot}}, t_{\text{tot}}+1, \ldots, -1\textbf{:} \\ \quad \{\ \mathcal{P} \leftarrow \{b + \mathcal{A}_t(b) \text{ for } b \in \mathcal{P}\} \\ \textbf{if } |\mathcal{P}| = 1\textbf{: break} \end{array}\right.$
output $\mathcal{P}$ ((perfect) sample)

- Propp & Wilson (1997)

Département de Physique
École normale supérieure

Coupling from the past 3/3



- Propp & Wilson (1997)
- see `CouplingFromThePast.py` on my website

Perfect Monte Carlo samples of hard disks

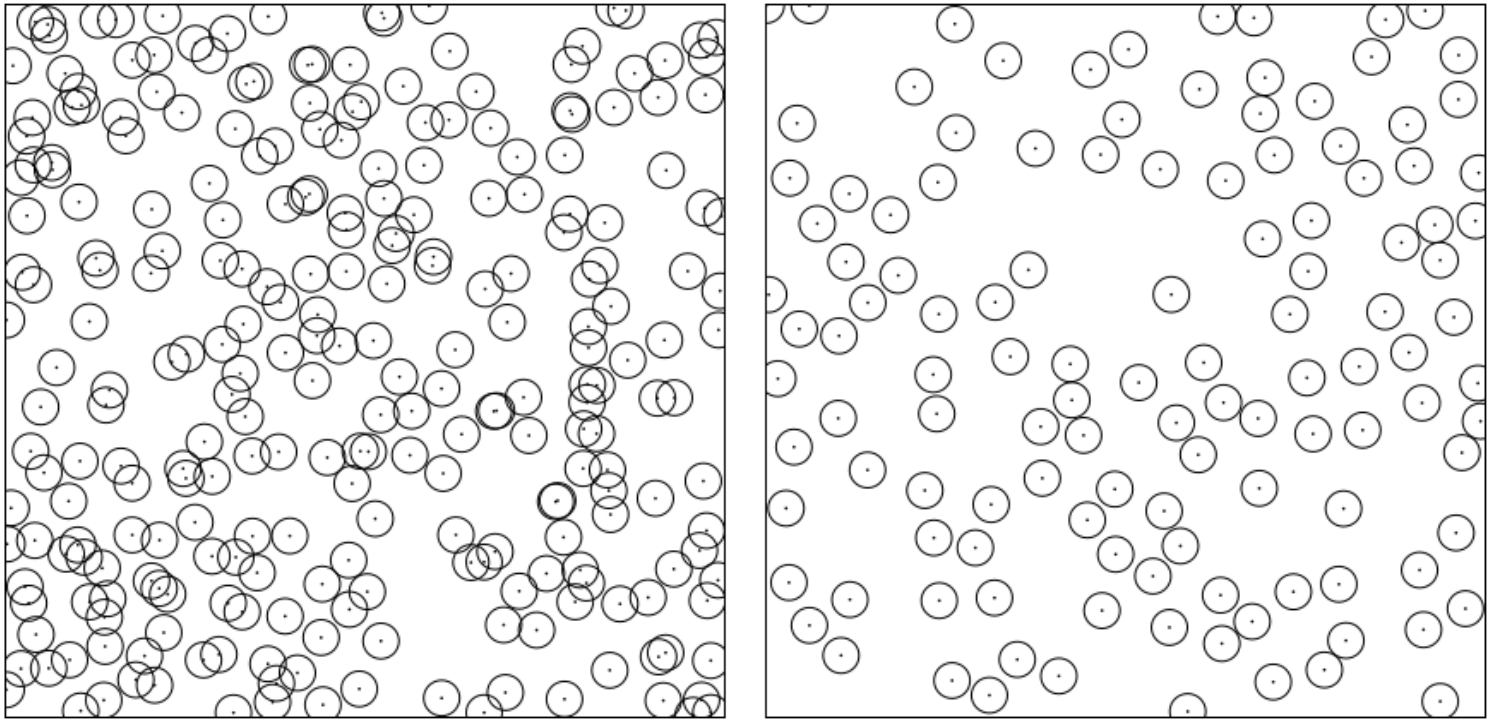
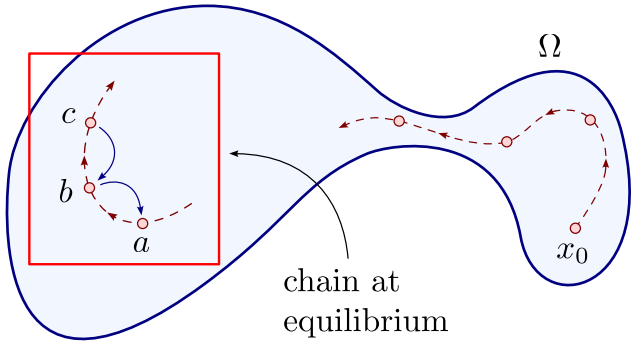


Figure 10: Perfectly random samples of the Strauss point process. In both panels the point

- Perfect sample of hard disks (right) from Wilson (2000)

Detailed balance, global balance, lifting



- Reversible transition matrices P satisfy the 'detailed-balance' condition:

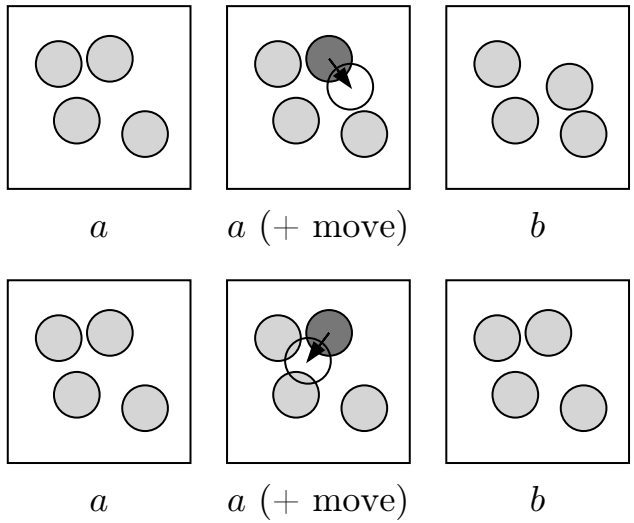
$$\pi_a P_{ab} = \pi_b P_{ba}$$

- Non-reversible transition matrices P only satisfy 'global balance':

$$\pi_a = \sum_{b \in \Omega} \pi_b P_{ba}$$

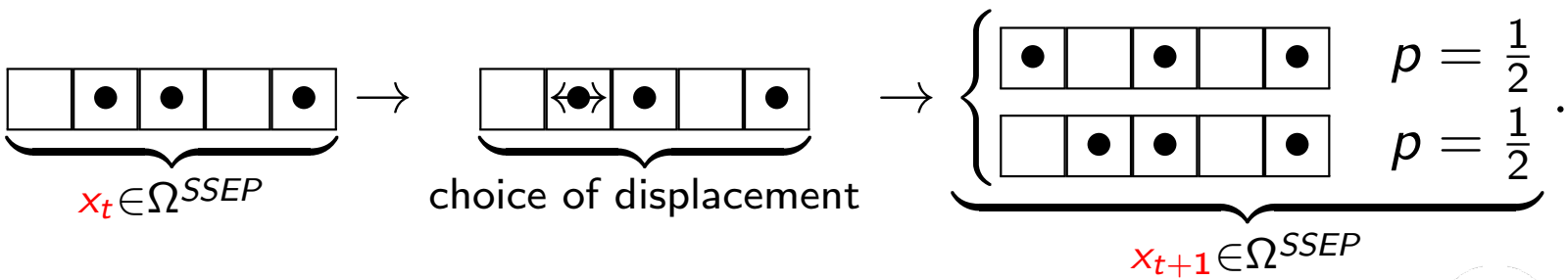
The Metropolis algorithm and the SSEP

1. Metropolis (1953) algorithm:



The Metropolis algorithm is reversible.

2. SSEP:

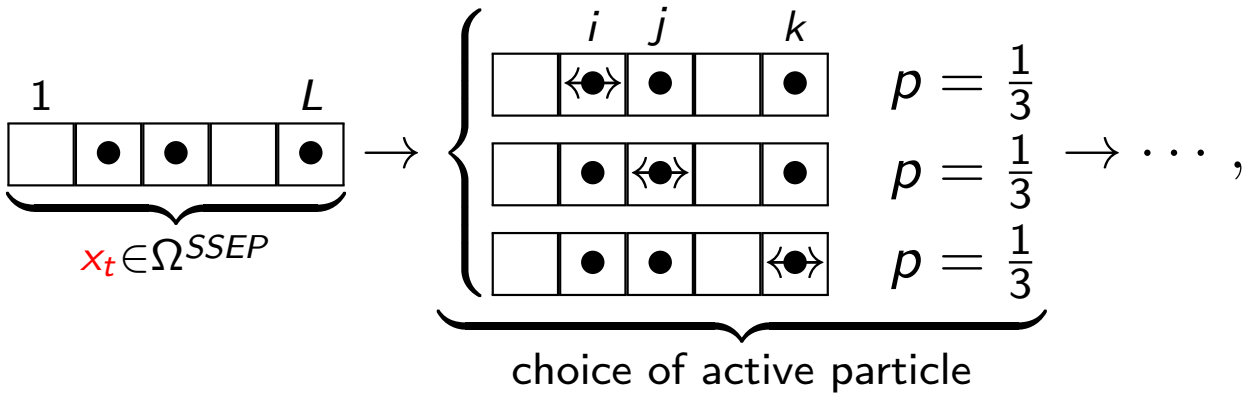


$1D$ lattice reduction of hard-sphere Metropolis algorithm.

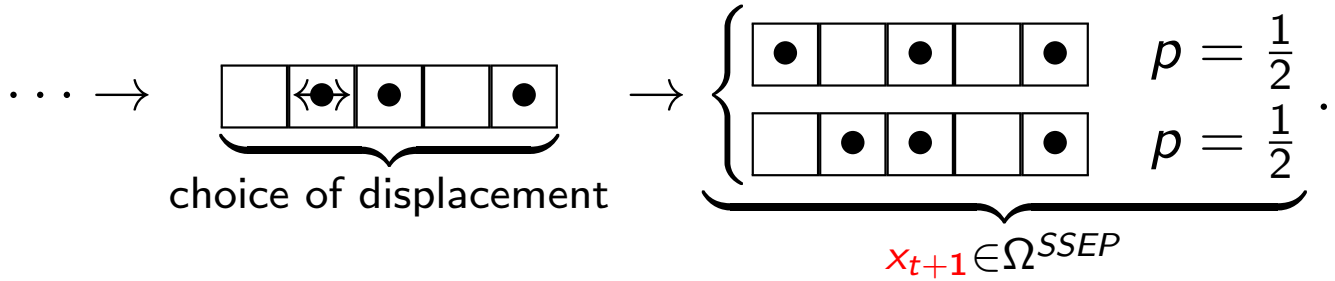
Département de Physique
École normale supérieure

Symmetric simple exclusion process (SSEP)

- Move (first part ...)



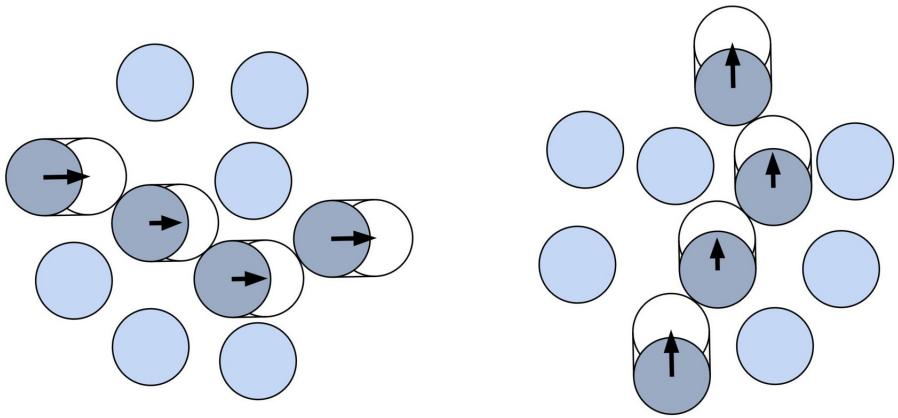
- Move (... second part)



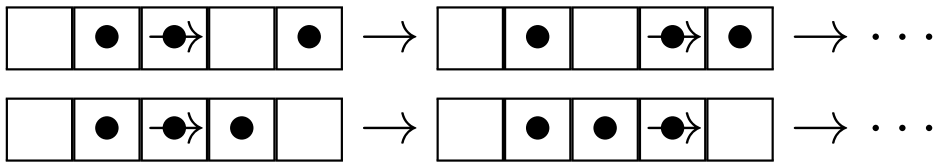
Département de Physique
École normale supérieure

Event-chain Monte Carlo and the lifted TASEP

1. Event-chain Monte Carlo algorithm (Bernard, Krauth, Wilson 2009):



2. Lifted TASEP:

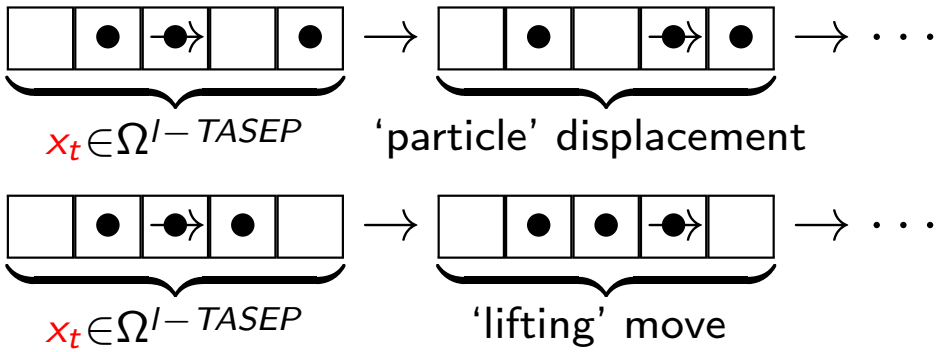


One-dimensional lattice reduction of event-chain Monte Carlo.

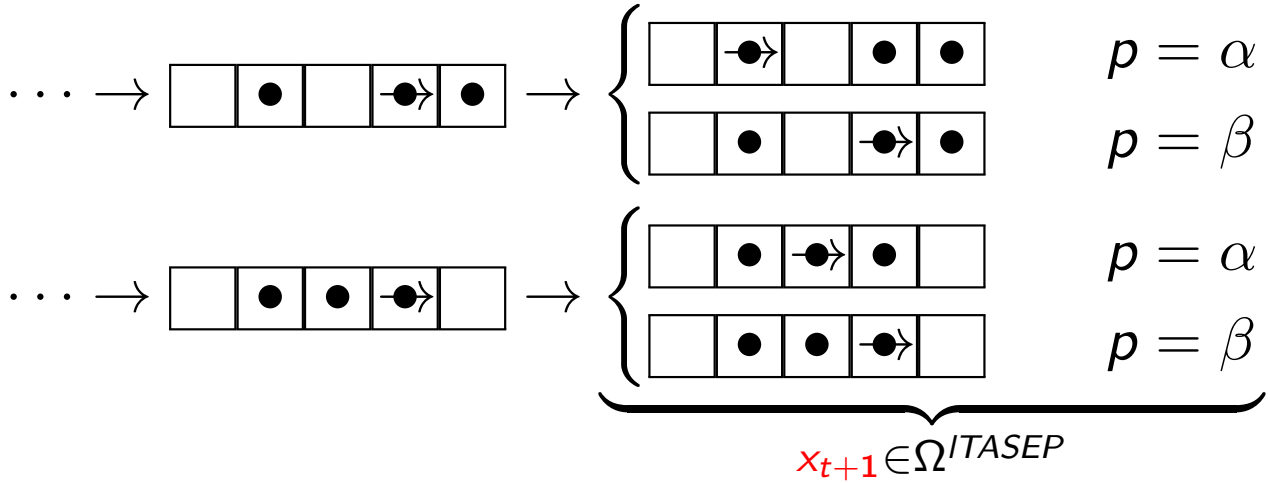
Département de Physique
École normale supérieure

Lifted TASEP (definition)

- $\Omega^{l-TASEP} = \Omega^{SSEP} \times \{-1, +1\} \times \{1, \ldots, N\}$, $\mathcal{L} = \emptyset$
- Move (first part ...)



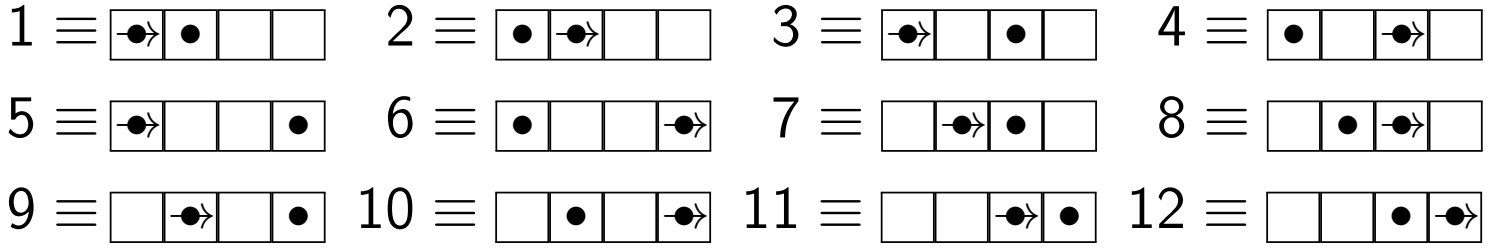
- Move (second part ...)



Département de Physique
École normale supérieure

Lifted TASEP (example)

NB: Consider only the forward-moving sector (pbc):



$$P = \begin{bmatrix}
\alpha & \beta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \alpha & \beta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \beta & \alpha & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \alpha & \beta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \beta & \alpha & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \beta & \alpha & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & \beta & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & \beta & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \beta & \alpha \\
\beta & \alpha & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & \beta \\
\cdot & \cdot & \beta & \alpha & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot
\end{bmatrix}$$

Synopsis of mixing and relaxation times

Algorithm	mixing	relaxation (inverse gap)
SSEP	$N^3 \log N$	N^3
TASEP	$N^{5/2}$	$N^{5/2}$
Lifted TASEP	N^2	N^2 ($N^{3/2}$)

- Bethe ansatz: Essler & Krauth (2023)

Département de Physique
École normale supérieure

Factorized Metropolis algorithm (pair potential)

- Metropolis filter:

$$p^{\mathrm{Met}}(\mathrm{x} \to \mathrm{x}') = \min\left[1, \exp\left(-\beta \sum_{i<j} \Delta U_{i,j}\right)\right]$$

$$p^{\mathrm{Met}}(\mathrm{x} \to \mathrm{x}') = \min\left[1, \prod_{i<j} \exp\left(-\beta \Delta U_{i,j}\right)\right]$$

- Factorized Metropolis filter (Michel, Kapfer, Krauth 2014):

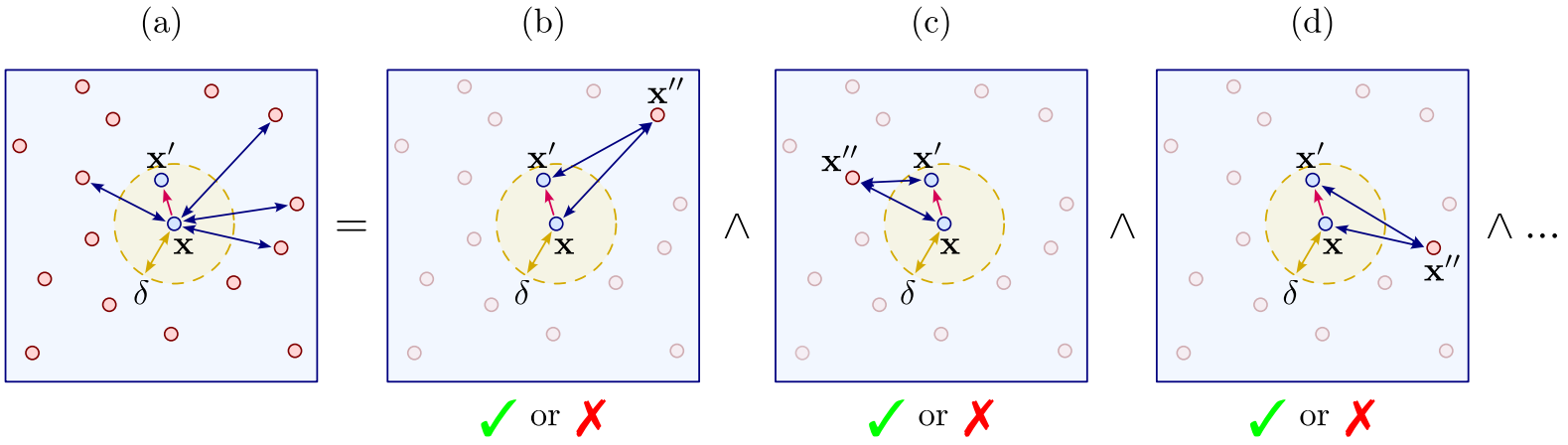
$$p^{\mathrm{Fact}}(\mathrm{x} \to \mathrm{x}') = \prod_{i<j} \min\left[1, \exp\left(-\beta \Delta U_{i,j}\right)\right].$$

$$X^{\mathrm{Fact}}(\mathrm{x} \to \mathrm{x}') = X_{1,2} \wedge X_{1,3} \wedge \cdots \wedge X_{N-1,N}$$

- i.e.: Accept move if all pairs (i, j) accept it.
- Consensus-based, not gradient-based.

Département de Physique
École normale supérieure

Factorized Metropolis algorithm



- Factorized Metropolis filter

$$p^{\mathrm{Fact}}(\mathrm{x} \to \mathrm{x}') = \prod_{i<j} \min\left[1, \exp\left(-\beta \Delta U_{i,j}\right)\right].$$

- Conjunction of Boolean random variables

$$X^{\mathrm{Fact}}(\mathrm{x} \to \mathrm{x}') = X_{1,2} \wedge X_{1,3} \wedge \cdots \wedge X_{N-1,N}$$

Département de Physique
École normale supérieure

Factorized Metropolis algorithm

procedure `factorized-metropolis`
input $\mathbf{X}$ (configuration at time t)
$\mathbf{x} \leftarrow$ **choice**$(\mathbf{X})$ (random particle)
$\mathbf{x}' \leftarrow \mathbf{x} + \Delta\mathbf{x}$ (with $|\Delta\mathbf{x}| < \delta$)
for $\mathbf{x}'' \in \mathbf{X} \setminus \{\mathbf{x}\}$**:**
  $\Upsilon \leftarrow$ `ran`$(0,1)$
  if $\Upsilon > \exp\left[-\beta\left(U_{\mathbf{x}''\mathbf{x}'} - U_{\mathbf{x}''\mathbf{x}}\right)\right]$**: goto** 1
$\mathbf{X} \leftarrow \{\mathbf{x}'\} \cup \mathbf{X} \setminus \{\mathbf{x}\}$
1 **output** $\mathbf{X}$ (configuration at time $t+1$)

- Factorized Metropolis filter

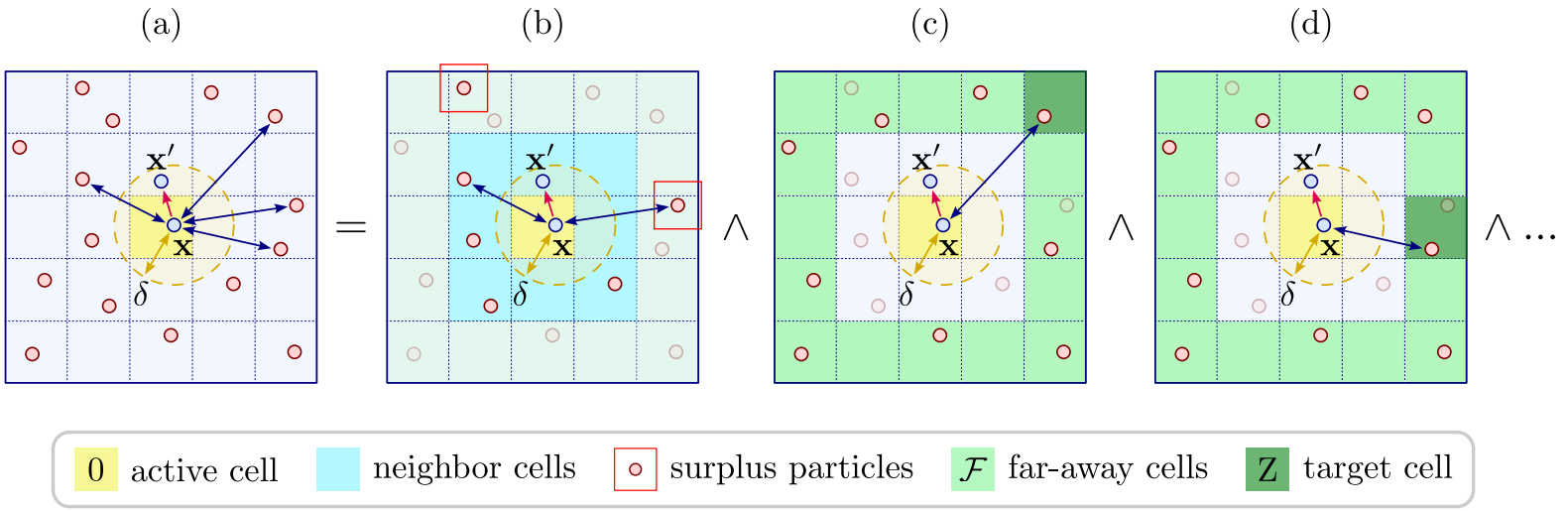
$$p^{\mathrm{Fact}}(\mathrm{x} \to \mathrm{x}') = \prod_{i<j} \min\left[1, \exp\left(-\beta \Delta U_{i,j}\right)\right].$$

- Conjunction of Boolean random variables

$$X^{\mathrm{Fact}}(\mathrm{x} \to \mathrm{x}') = X_{1,2} \wedge X_{1,3} \wedge \cdots \wedge X_{N-1,N}$$

Département de Physique
École normale supérieure

All-atom simulation with non-reversible Markov chains

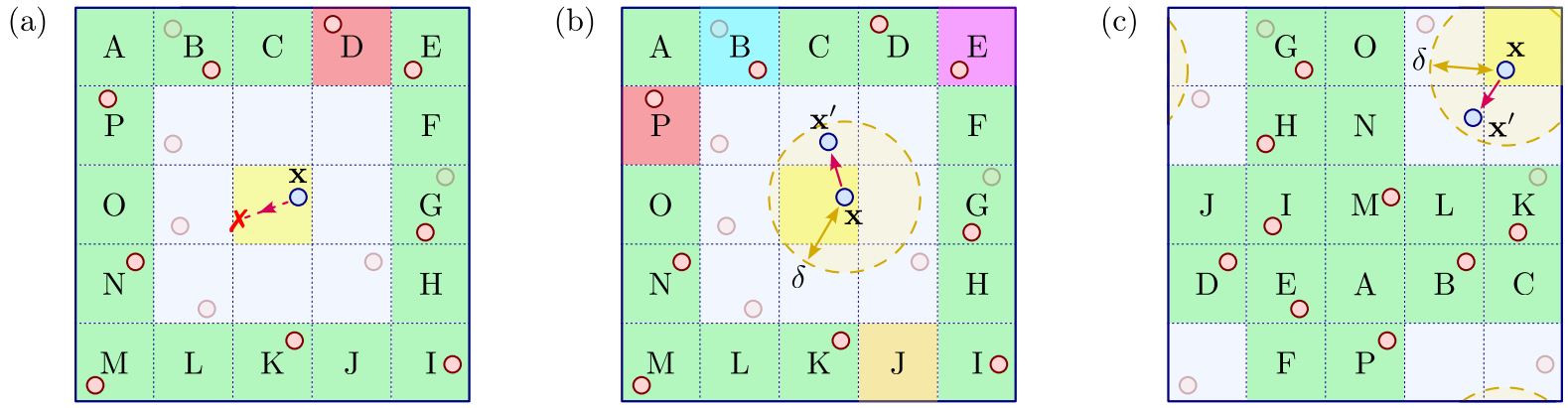


- Veto probability (for particle in cell z)

$$q_{x''} = q_Z \frac{\left[1 - \exp\left(-\beta \Delta U_{x,x''}\right)\right]}{q_Z}$$

- Two-pebble veto.

All-atom simulation with non-reversible Markov chains



$\mathcal{S}_{\text{veto}} = \{\text{D}\}$

$\pi(\mathcal{S}_{\text{veto}}) = q_{\text{D}}\mathrm{d}t$

$\mathcal{S}_{\text{veto}} = \{\text{B, E, J, P}\}$

$\pi(\mathcal{S}_{\text{veto}}) = \prod_{\text{Z}\in\mathcal{S}_{\text{veto}}} q_{\text{Z}} \prod_{\text{Z}\notin\mathcal{S}_{\text{veto}}} (1 - q_{\text{Z}})$

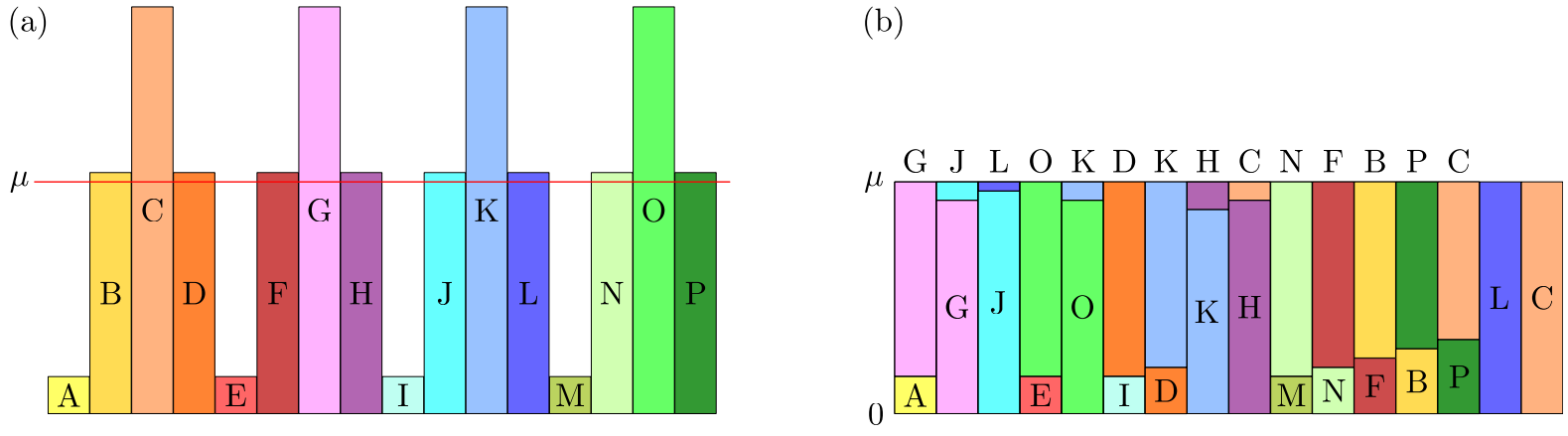
- Veto probability (for particle in cell z)

$$q_{\text{x}''} = q_{\text{Z}} \frac{\left[1 - \exp\left(-\beta \Delta U_{\text{x},\text{x}''}\right)\right]}{q_{\text{Z}}}$$

- $q_{\text{Z}} = 0.0001$: Check cell Z only once in $10,000$ times.

Département de Physique
École normale supérieure

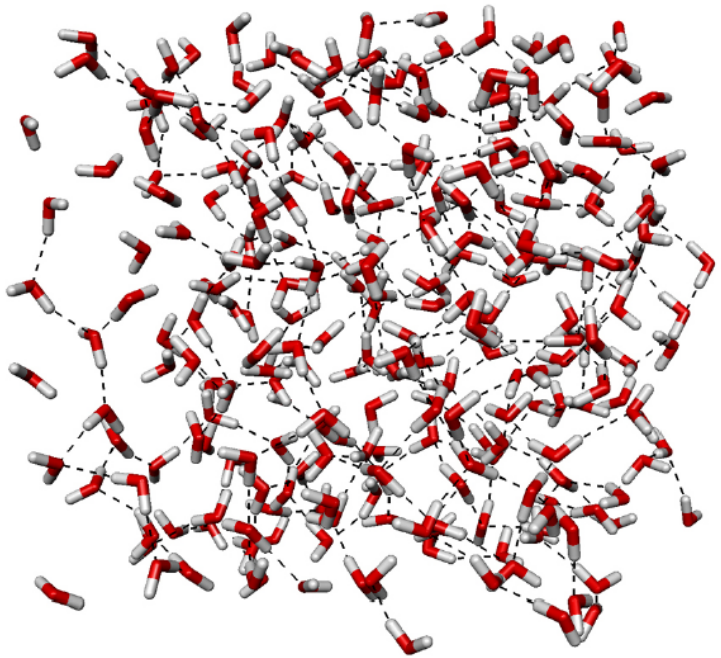
All-atom simulation with non-reversible Markov chains



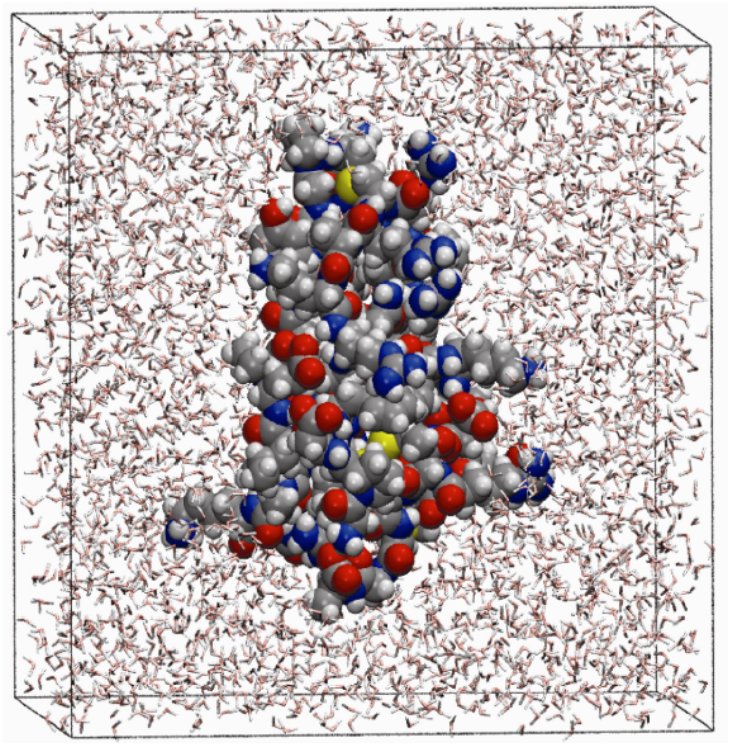
- Walker's algorithm (1977): $\mathcal{O}(1)$ sampling from a constant distribution
- Cell-veto algorithm (Kapfer, Krauth, 2017), samples $\exp(-\beta U)$ without evaluating U.

Département de Physique
École normale supérieure

All-atom simulation with non-reversible Markov chains



(a)



(b)

(a): Native sampling for long-range interactions without cutoffs, space discretizations, time-step errors, thermostat (Höllmer, Maggs, Krauth (2023)).

(b): Yet to be done...

Département de Physique
École normale supérieure

Conclusion & Outlook

Conclusion:

- A second revolution in Markov-chain Monte Carlo underway.
- Time scales of MCMC much better understood.
- Coupling: a way to perfect simulations.
- Non-reversible MCMC is what comes after the revolution.
- Lifting: a practical method to create non-reversible algorithms.

Outlook:

- Sampling $\exp(-\beta U)$ without evaluating U.
- 'Natively cutoff-free' MCMC (Coulomb, LJ) in $\mathcal{O}(1)$.
- Applications in chemical physics.

Département de Physique
École normale supérieure

Thanks to:

- E. P. Bernard, W. Krauth, D. B. Wilson, PRE (2009) *Event-chain algorithms for hard-sphere systems*
- E. P. Bernard, W. Krauth, PRL (2011) *Two-Step Melting in Two Dimensions: First-Order Liquid-Hexatic Transition*
- M. Michel, S. C. Kapfer, W. Krauth, JCP (2014) *Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps*
- S. C. Kapfer, W. Krauth, PRE (2016) *Cell-veto Monte Carlo algorithm for long-range systems*
- S. C. Kapfer, W. Krauth, PRL (2017) *Irreversible Local Markov Chains with Rapid Convergence towards Equilibrium*
- P. Hoellmer, A. C. Maggs, W. Krauth, *Molecular simulation from modern statistics: Continuous-time, continuous-space, exact arXiv:2305.02979*
- F. H. L. Essler, W. Krauth, *Lifted TASEP: a Bethe ansatz integrable paradigm for non-reversible Markov chains arXiv:2306.13059*
- G. Tartero, W. Krauth, *arXiv;2405.XXX Markov-chain sampling for long-range systems without evaluating energies*

Département de Physique
École normale supérieure