Mixing, stopping, coupling, lifting, and other keys to the second Markov-chain revolution

Werner Krauth
Department of Physics, Ecole normale supérieure, Paris (France)
Rudolf Peierls Centre for Theoretical Physics, University of Oxford (UK)

Cavendish Quantum Colloquium
University of Cambridge, Cambridge (UK)
31 January 2024

S. C. Kapfer, W. Krauth; PRL (2017)
'Irreversible Local Markov Chains with Rapid Convergence towards Equilibrium'

Z. Lei, W. Krauth, A. C. Maggs; PRE (2019)
'Event-chain Monte Carlo with factor fields'

F. Essler, W. Krauth; arxiv:2306.13059
'Lifted TASEP: a Bethe-ansatz integrable paradigm for non-reversible Markov chains'

Département de Physique
École normale supérieure

Metropolis et al (1953) (1/3)

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

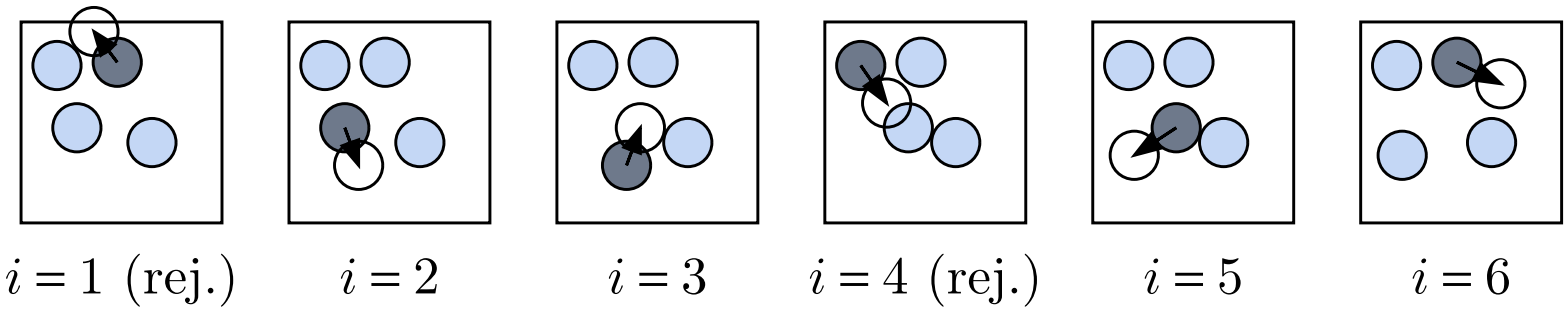
AND

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

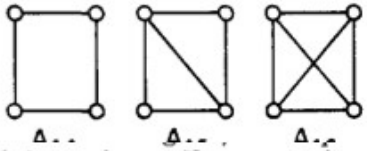
(Received March 6, 1953)

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.

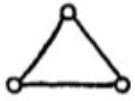
Metropolis et al (1953) (2/3)



Metropolis et al (1953) (3/3)



distinguished by primes. For example, A_{33} is given schematically by the diagram

$A_{3,3}$: 

and mathematically as follows: if we define $f(r_{ij})$ by

$$f(r_{ij})=1 \quad \text{if} \quad r_{ij}<d,$$
$$f(r_{ij})=0 \quad \text{if} \quad r_{ij}>d,$$

then

$$A_{3,3}=\frac{1}{\pi^2 d^4}\int\cdots\int dx_1 dx_2 dx_3 dy_1 dy_2 dy_3 (f_{12}f_{23}f_{31}).$$

The schematics for the remaining integrals are indicated in Fig. 6.

The coefficients $A_{3,3}$, $A_{4,4}$, and $A_{4,5}$ were calculated were put down at random, subject to $f_{12}=f_{23}=f_{34}=f_{15}=1$. The number of trials for which $f_{45}=1$, divided by the total number of trials, is just $A_{5,5}$.

The data on $A_{4,6}$ is quite reliable. We obtained

VI. CONCLUSION

The method of Monte Carlo integrations over configuration space seems to be a feasible approach to statistical mechanical problems which are as yet not analytically soluble. At least for a single-phase system a sample of several hundred particles seems sufficient. In the case of two-dimensional rigid spheres, runs made with 56 particles and with 224 particles agreed within statistical error. For a computing time of a few hours with presently available electronic computers, it seems possible to obtain the pressure for a given volume and temperature to an accuracy of a few percent.

In the case of two-dimensional rigid spheres our results are in agreement with the free volume approximation for $A/A_0<1.8$ and with a five-term virial expansion for $A/A_0>2.5$. There is no indication of a phase transition.

Alder–Wainwright (1962) (1/2)

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)

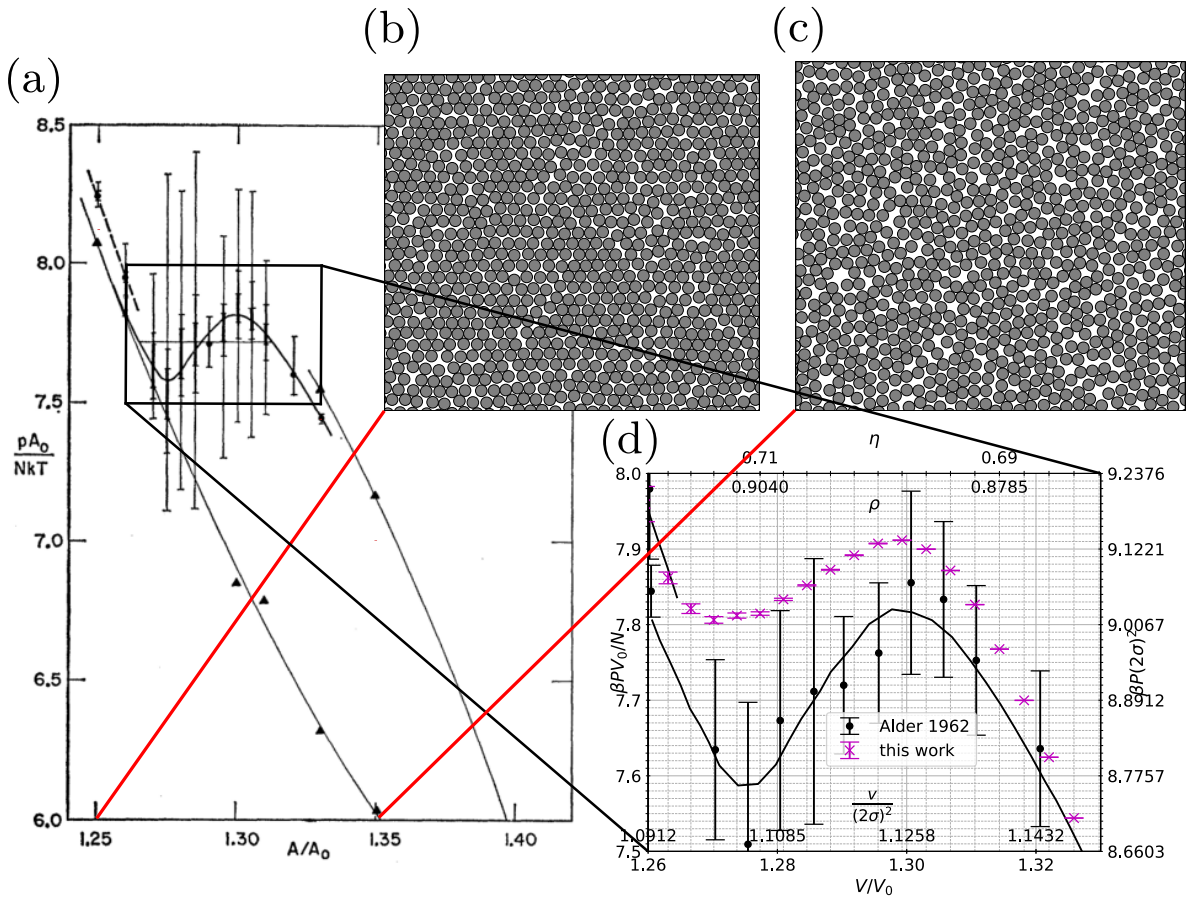
The study of a two-dimensional system consisting of 870 hard-disk particles in the phase-transition region has shown that the isotherm has a van der Waals-like loop. The density change across the transition is about 4% and the corresponding entropy change is small.

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

Alder–Wainwright (1962) (2/2)



- Generic 2D systems cannot crystallize (Peierls, Landau 1930s) but they can turn solid (Alder & Wainwright, 1962).
- Nature of transition disputed for decades.

Département de Physique
École normale supérieure

Kosterlitz–Thouless (1973)

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

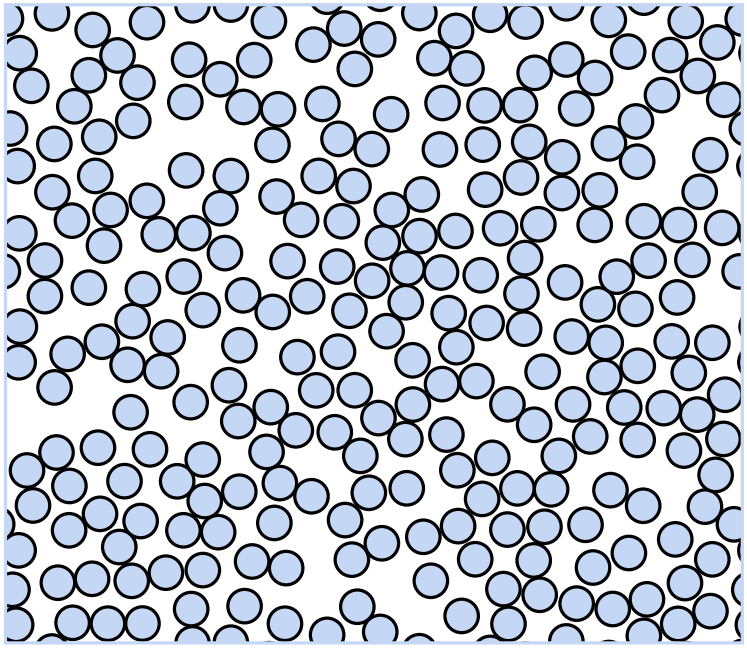
Received 13 November 1972

1. Introduction

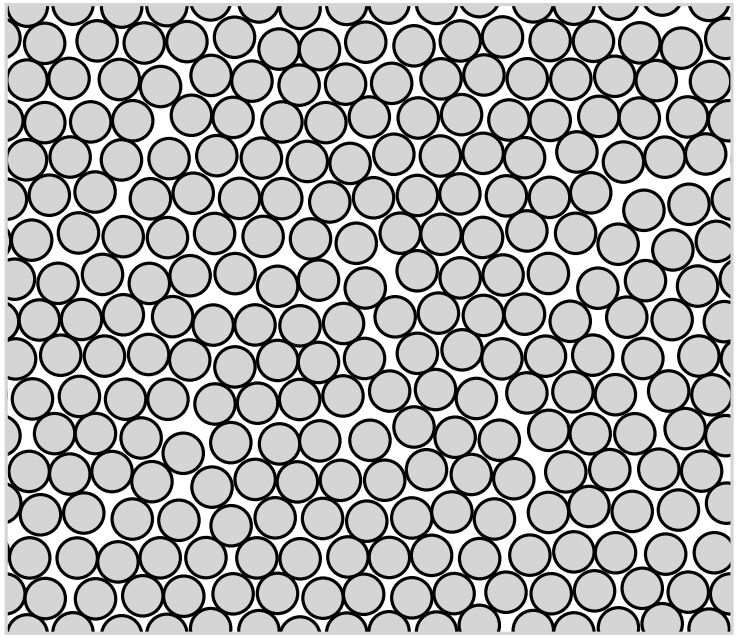
Peierls (1935) has argued that thermal motion of long-wavelength phonons will destroy the long-range order of a two-dimensional solid in the sense that the mean square deviation of an atom from its equilibrium position increases logarithmically with the size of the system, and the Bragg peaks of the 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

Possible phases in two dimensions



density $\eta = 0.48$



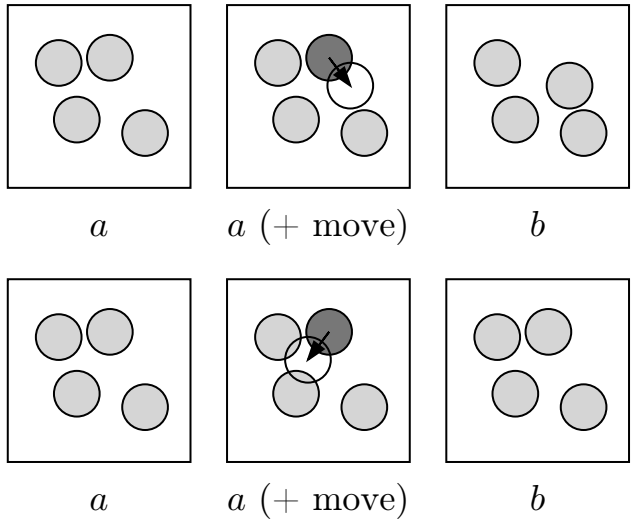
$\eta = 0.72$

Phase	positional order	orientational order
solid	algebraic	long-range
hexatic	short-range	algebraic
liquid	short-range	short-range

Département de Physique
École normale supérieure

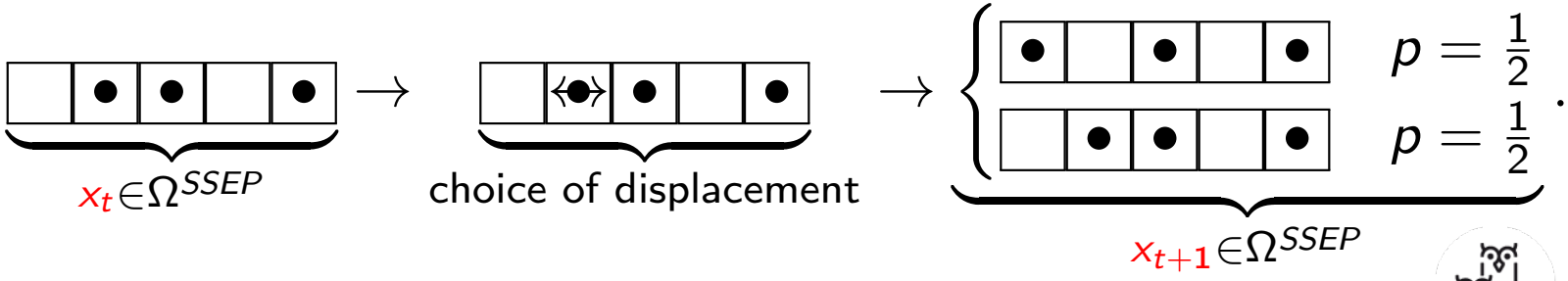
Metropolis algorithm and the SSEP

1. Metropolis (1953) algorithm:



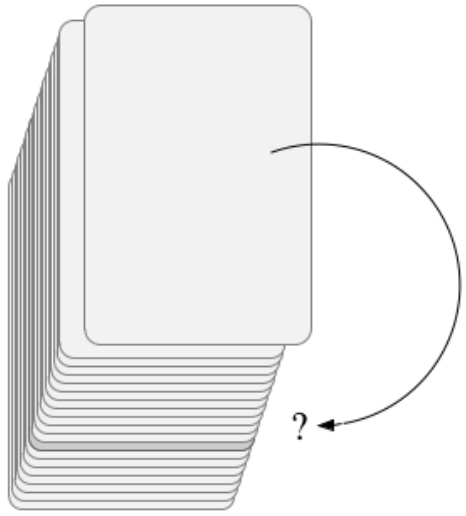
The Metropolis algorithm is reversible.

2. SSEP:



Département de Physique
École normale supérieure

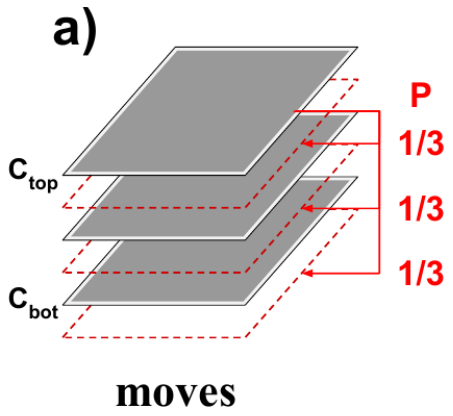
Shuffling of cards 1/5



- $\Omega_n^{\text{shuffle}} = \{\text{Permutations of } \{1, \ldots, n\}\}$
- 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

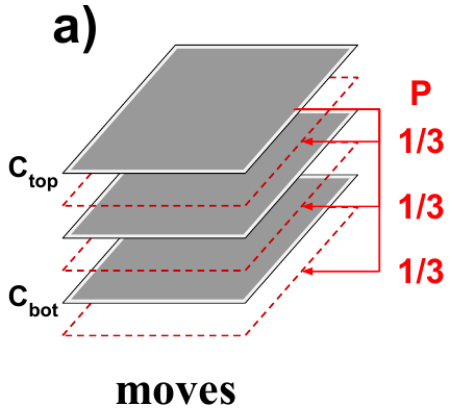


procedure top-to-random
input $\{c_1, \ldots, c_n\}$
$i \leftarrow$ choice$(\{1, \ldots, n\})$
$\{\hat{c}_1, \ldots, \hat{c}_n\} \leftarrow \{c_2, \ldots, c_i, c_1, c_{i+1}, \ldots, c_n\}$
output $\{\hat{c}_1, \ldots, \hat{c}_n\}$

- 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

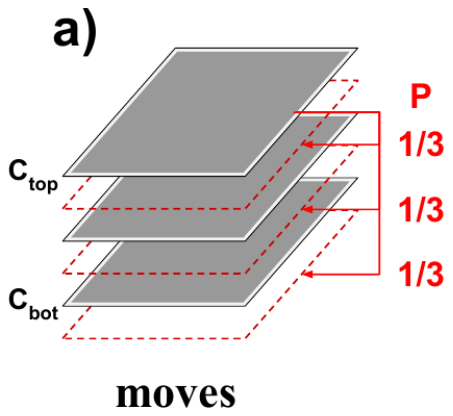
Shuffling of cards 4/5

- $$P_3^{\text{shuffle}} = \frac{1}{3}\begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$
- 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

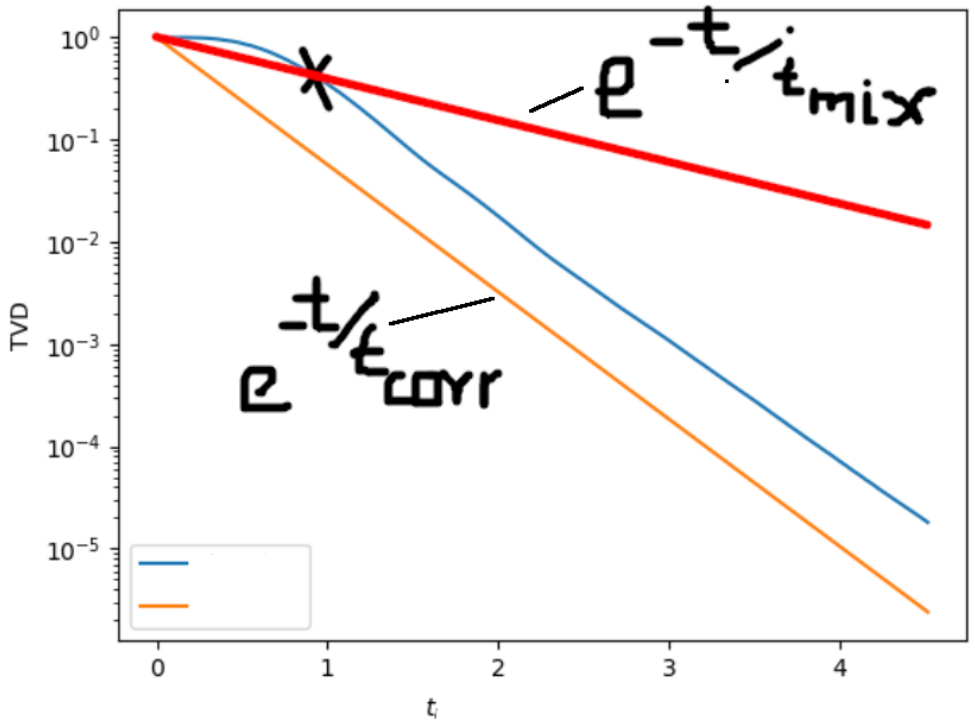


procedure `top2random-stop`
input $\{c_1, \ldots, c_n\}$
$c_{\text{first-n}} \leftarrow c_{\text{n}}$
for $t = 1, 2, \ldots$ **do**
$\tilde{c}_1 \leftarrow c_1$
$\{c_1, \ldots, c_n\} \leftarrow$ `top2random`$(\{c_1, \ldots, c_n\})$
if $(\tilde{c}_1 = c_{\text{first-n}})$ **break**
output $\{c_1, \ldots,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

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

Département de Physique
École normale supérieure

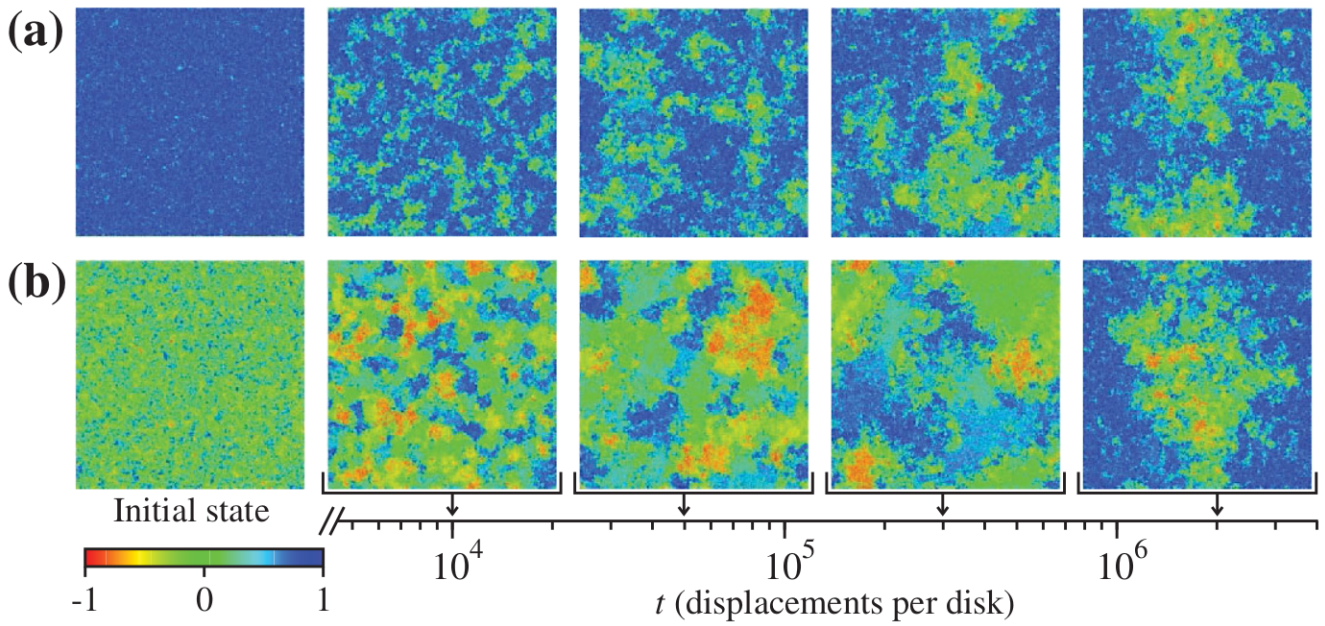
Mixing and Relaxation



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

Département de Physique
École normale supérieure

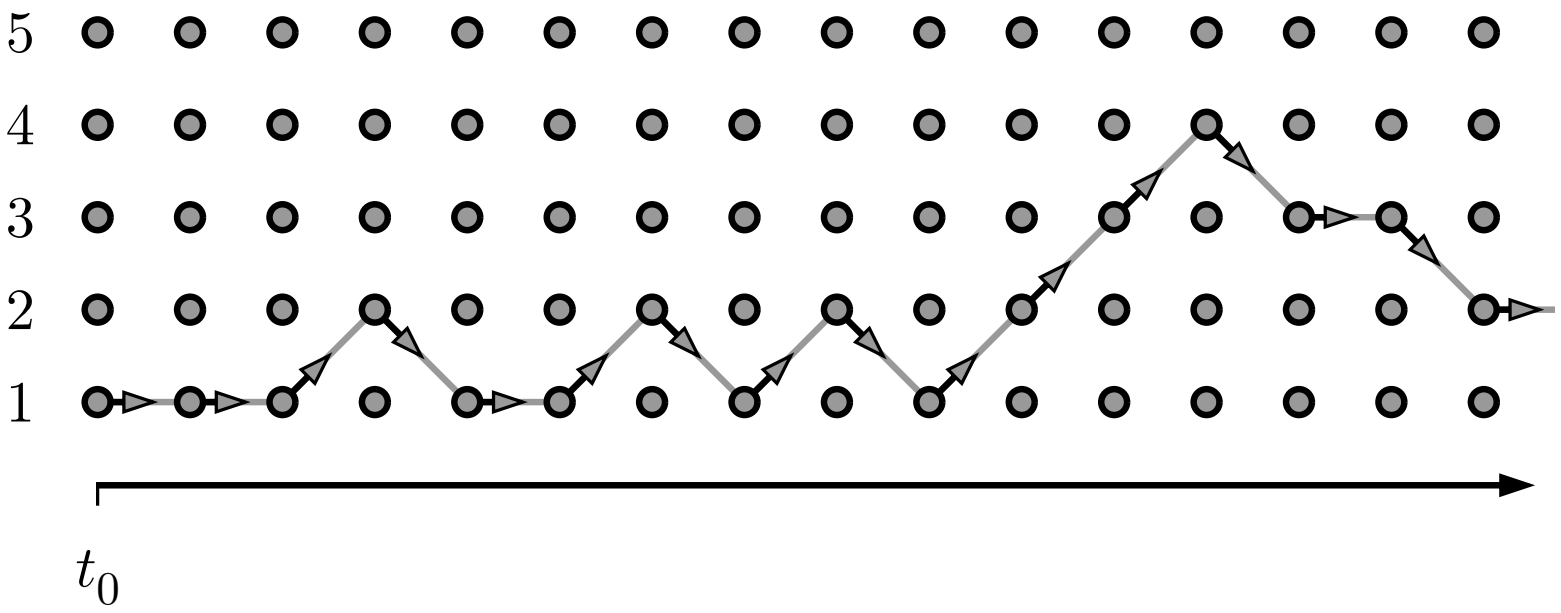
Example II: A non-asymptotic time scale



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

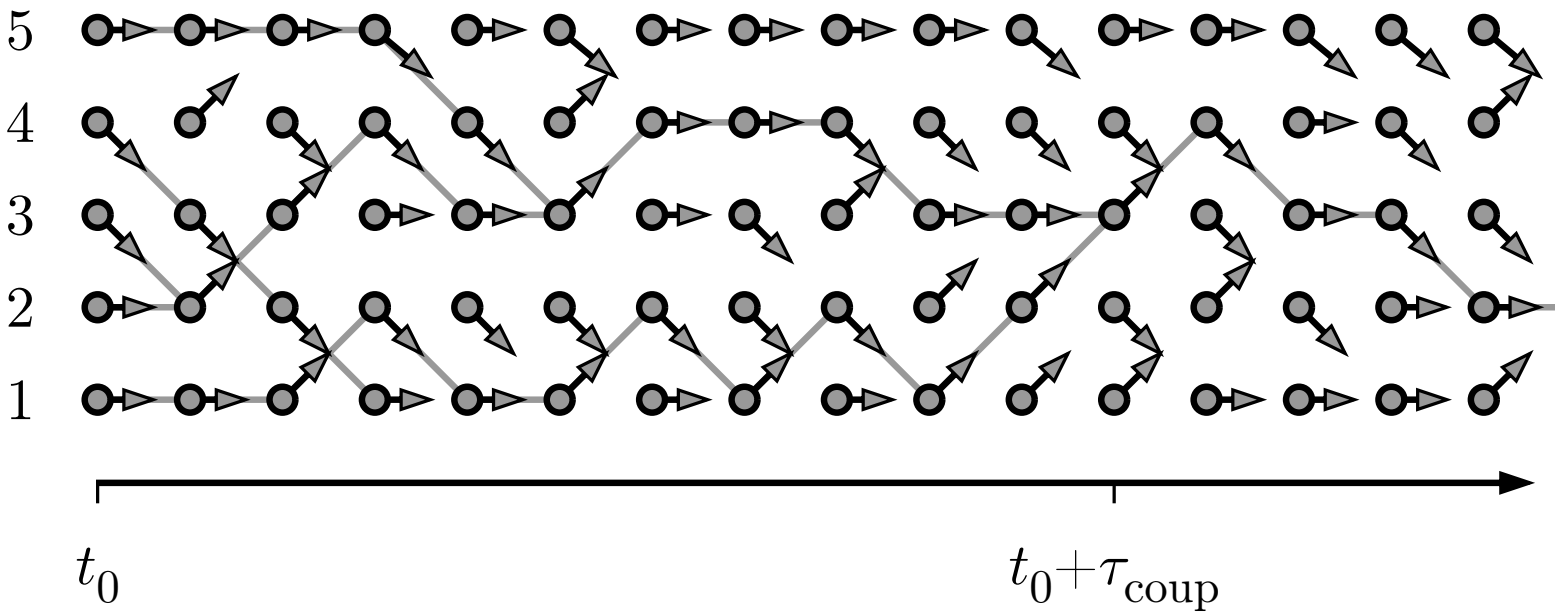
Département de Physique
École normale supérieure

Markov chain



- 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).

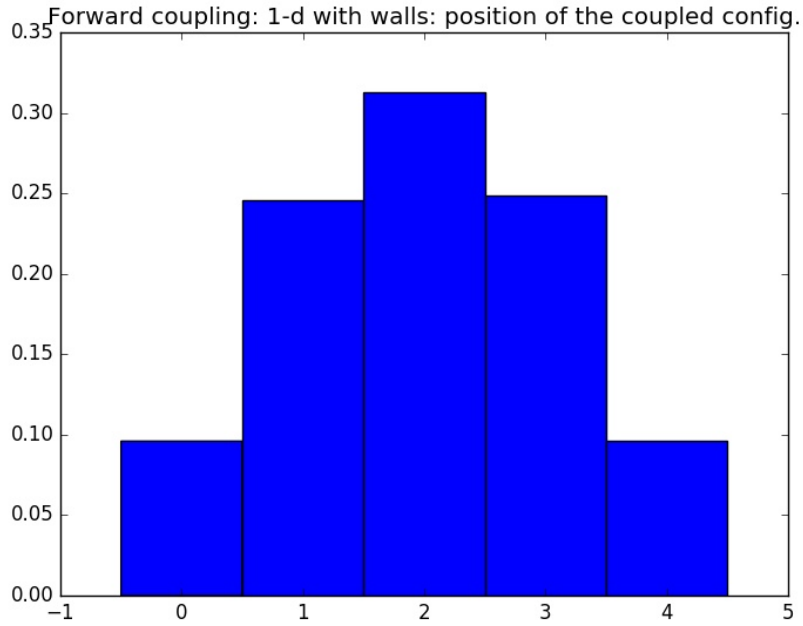
Markov chain (random maps), coupling 2/3

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.

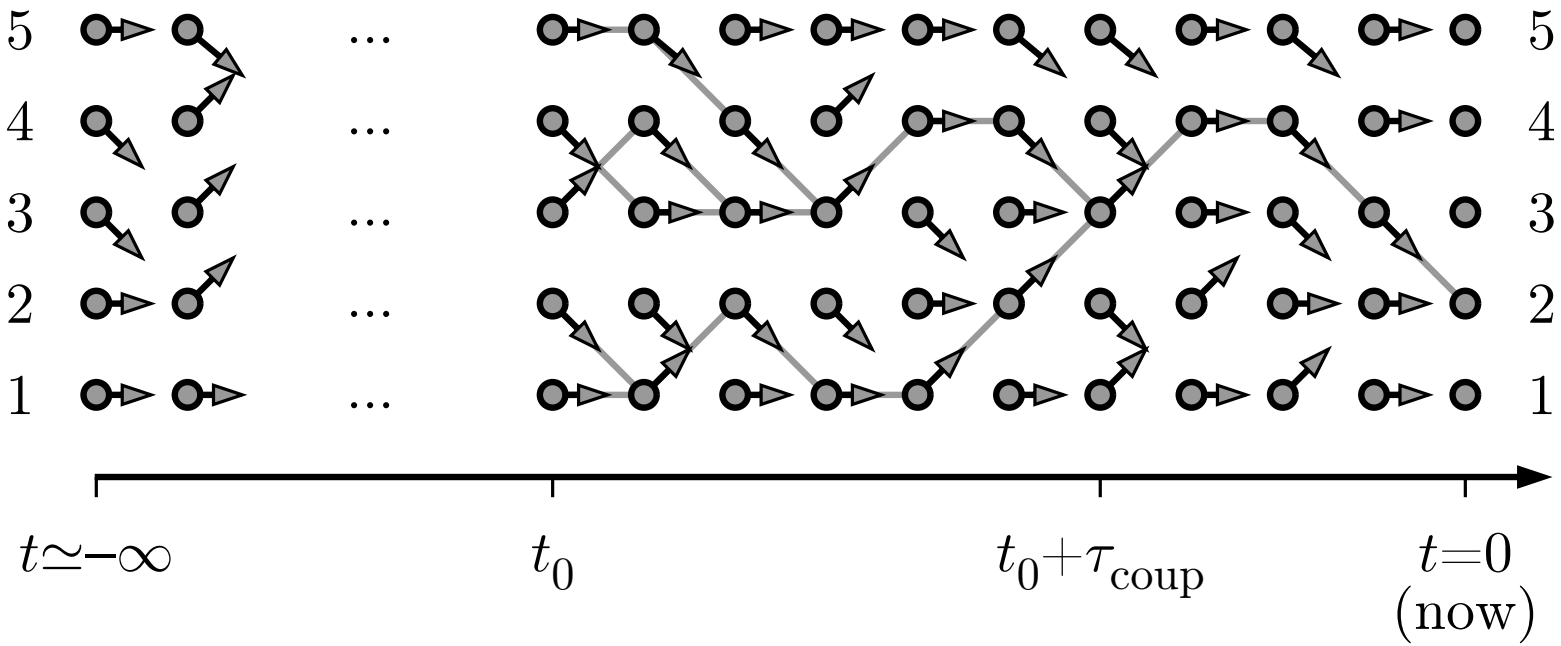
Département de Physique
École normale supérieure

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)

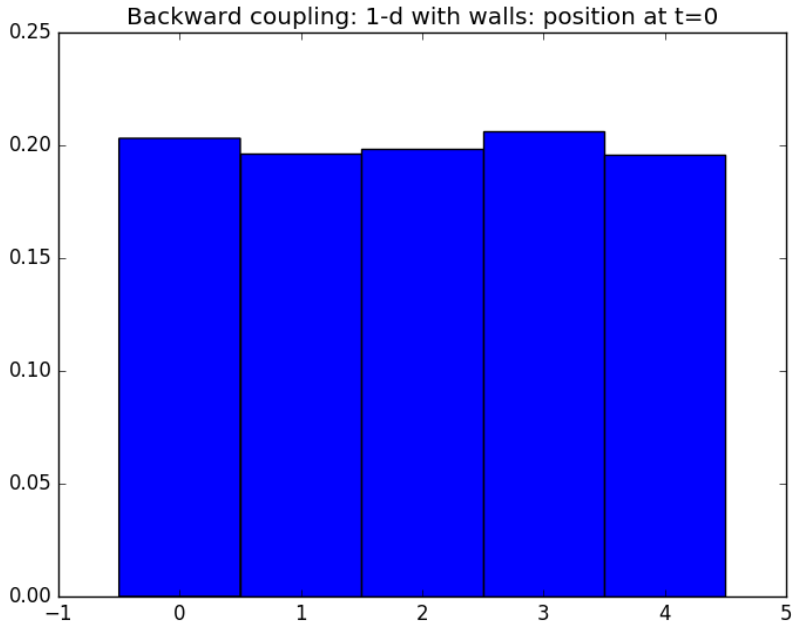
Coupling from the past 2/3

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

Example III: Perfect Monte Carlo samples of hard disks

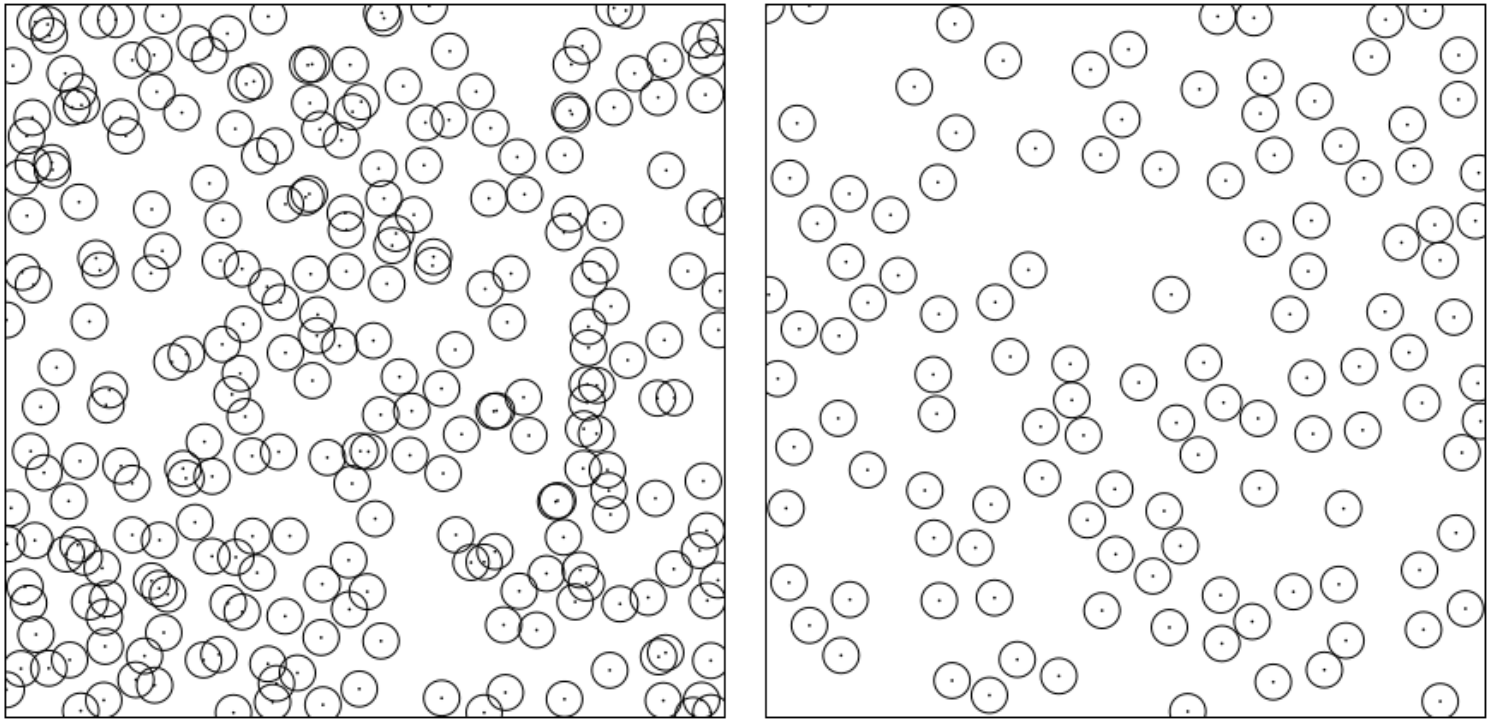
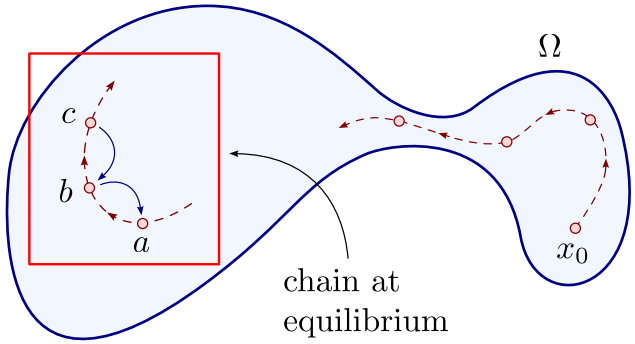


Figure 10: Perfectly random samples of the Strauss point process. In both panels the point

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

Detailed balance, global balance, lifting



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

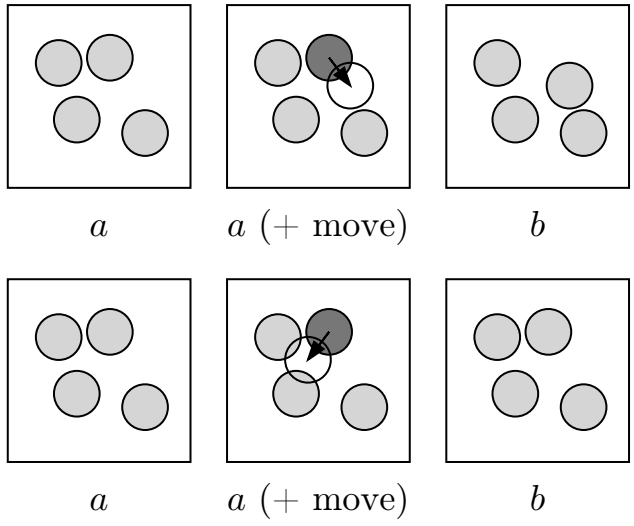
$$\pi_a P_{ab} = \pi_b P_{ba}$$

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

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

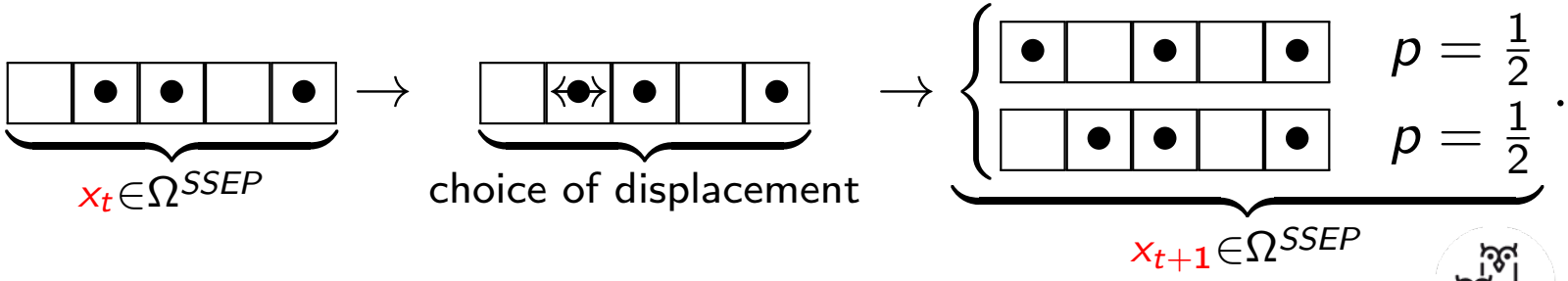
Metropolis algorithm and the SSEP

1. Metropolis (1953) algorithm:



The Metropolis algorithm is reversible.

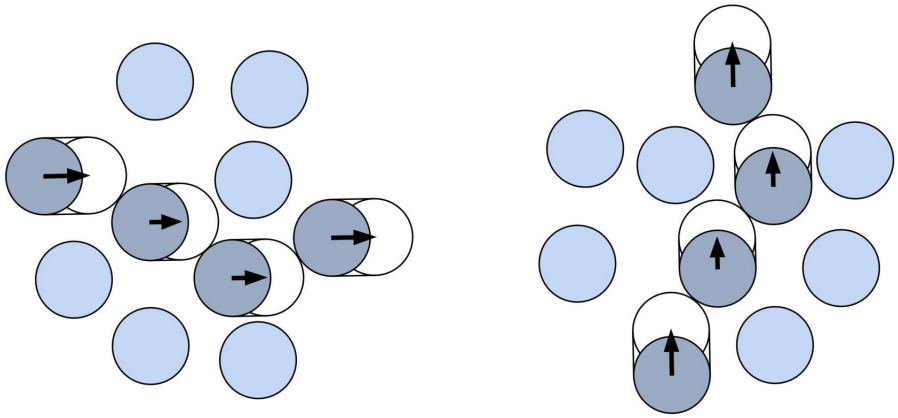
2. SSEP:



Département de Physique
École normale supérieure

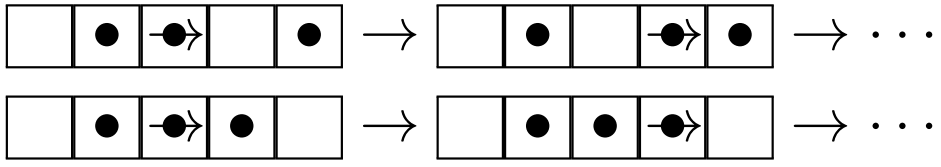
Event-chain Monte Carlo and the lifted TASEP

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



Non-reversible, unchanged stationary distribution, orders of magnitude faster than Metropolis.

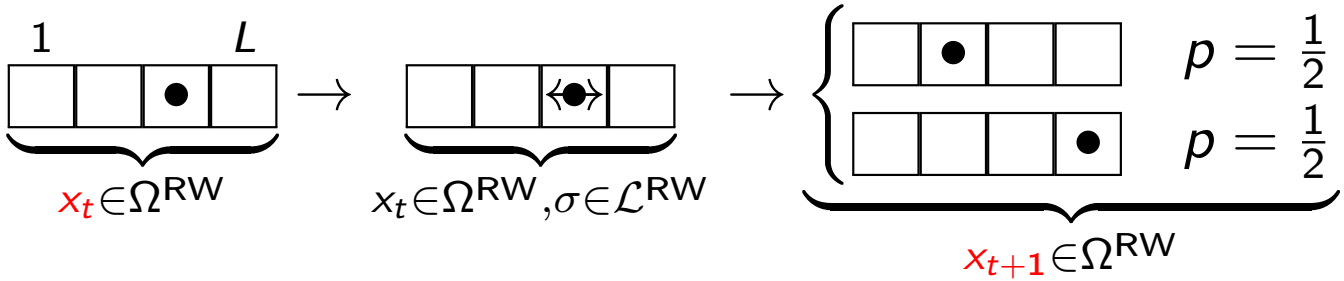
2. Lifted TASEP:



Département de Physique
École normale supérieure

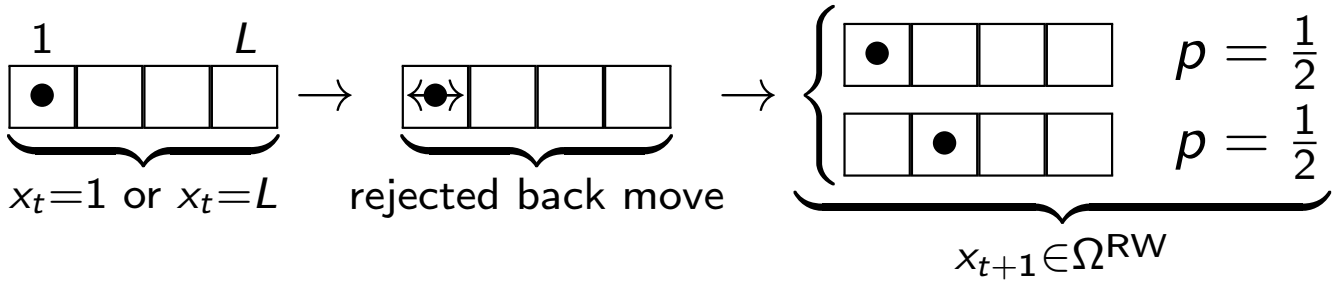
Random walk (RW) on the one-dimensional lattice

- In the bulk:



(1)

- At the boundary:

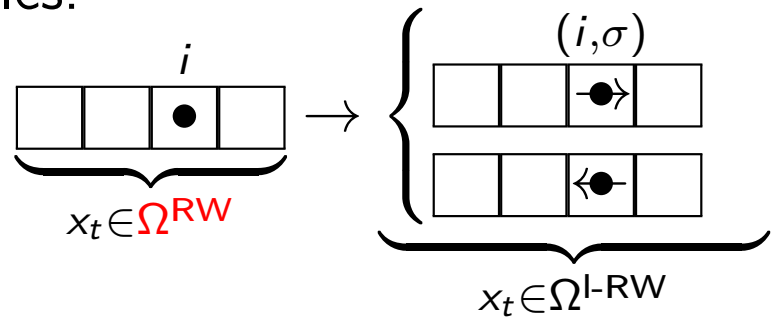


(2)

Département de Physique
École normale supérieure

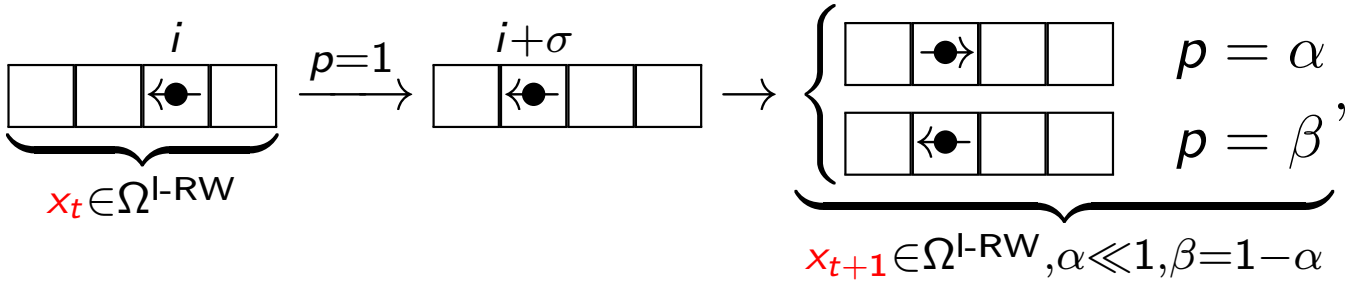
Lifted random walk (l-RW)

- Lifting of samples:



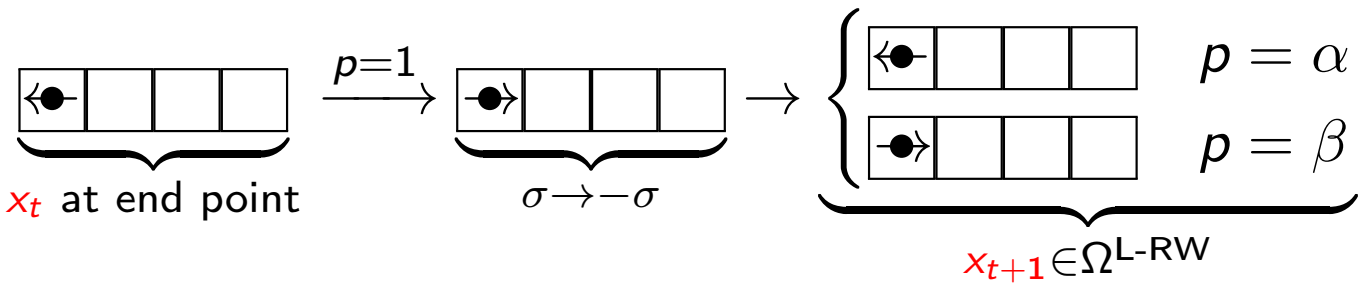
(3)

- In the bulk:



(4)

- At the boundary:

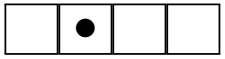
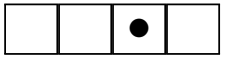


(5)

Diaconis, Holmes, Neal (2000)

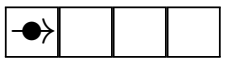
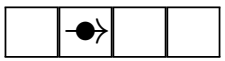
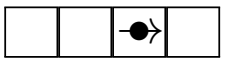
Département de Physique
École normale supérieure

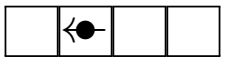
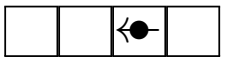
Random walk, lifted random walk (examples)

- $1 \equiv$  $2 \equiv$  $3 \equiv$  $4 \equiv$ 

$$P^{RW}_{\text{walls}} = \frac{1}{2}\begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & 1 \\ \cdot & \cdot & 1 & 1 \end{bmatrix}$$

- Lifted random walk (NB: $\alpha + \beta = 1, \quad \alpha \sim 1/L$)

$1 \equiv$  $3 \equiv$  $5 \equiv$  $7 \equiv$ 

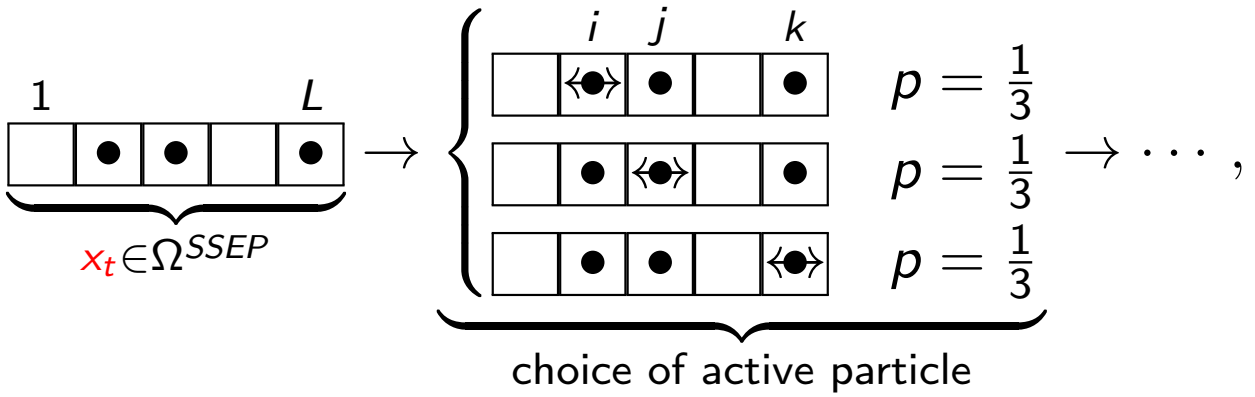
$2 \equiv$  $4 \equiv$  $6 \equiv$  $8 \equiv$ 

$$P^{\text{lRW}}_{\text{walls}} = \begin{bmatrix} \cdot & \cdot & \beta & \alpha & \cdot & \cdot & \cdot & \cdot \\ \beta & \alpha & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \beta & \alpha & \cdot & \cdot \\ \alpha & \beta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \beta & \alpha \\ \cdot & \cdot & \alpha & \beta & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & \beta \\ \cdot & \cdot & \cdot & \cdot & \alpha & \beta & \cdot & \cdot \end{bmatrix},$$

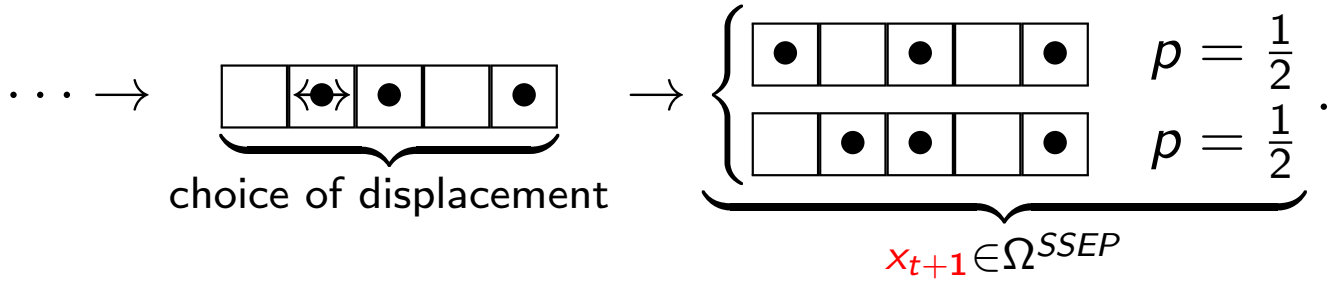
Département de Physique
École normale supérieure

Symmetric simple exclusion process (SSEP)

- Move (first part ...)



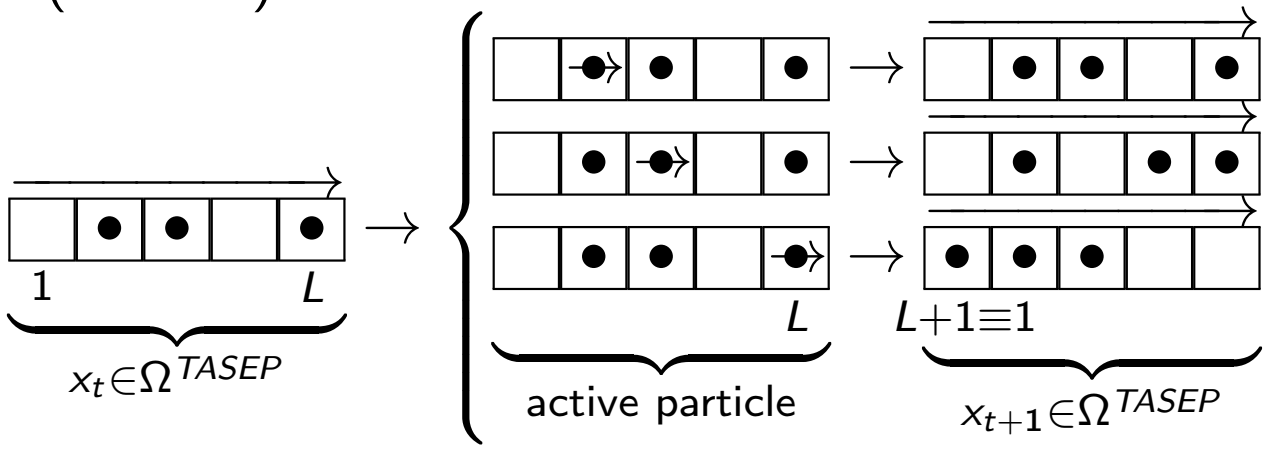
- Move (... second part)



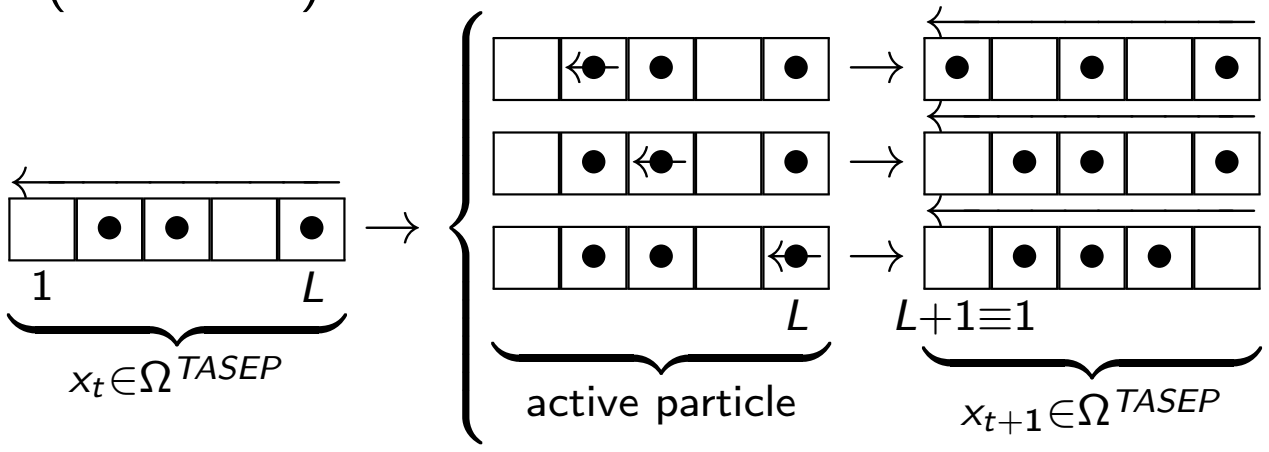
Département de Physique
École normale supérieure

Totally asymmetric simple exclusion process (TASEP)

- Move (forward)



- Move (backward)



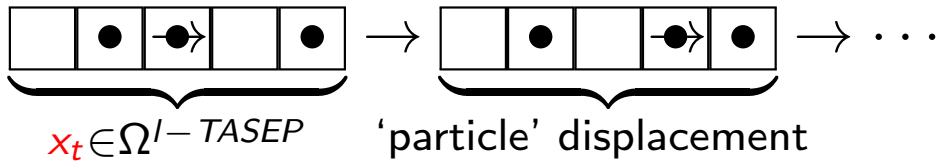
- forward–backward coupling (ad-hoc, or boundary conditions).

NB: Non-reversible, i.e. non-equilibrium, but samples equilibrium Boltzmann distribution.

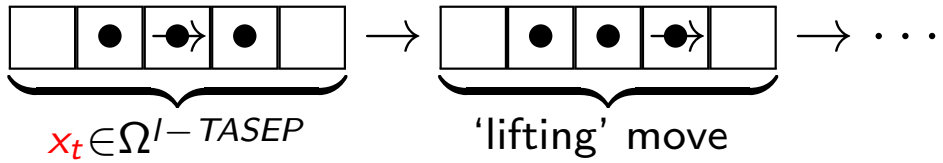
Département de Physique
École normale supérieure

Lifted TASEP (definition)

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

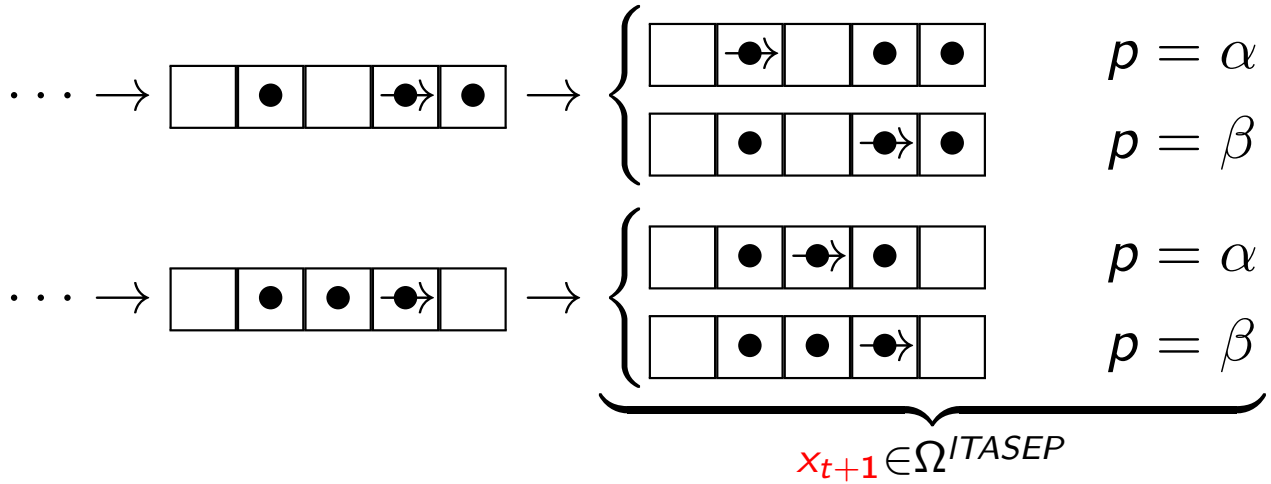


(6)



(7)

- Move (second part ...)

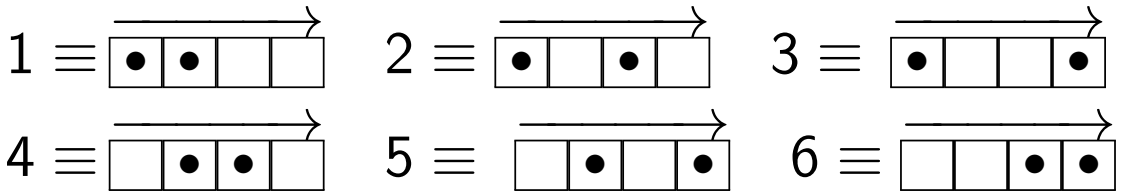


(6b)

(7b)

TASEP (example)

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

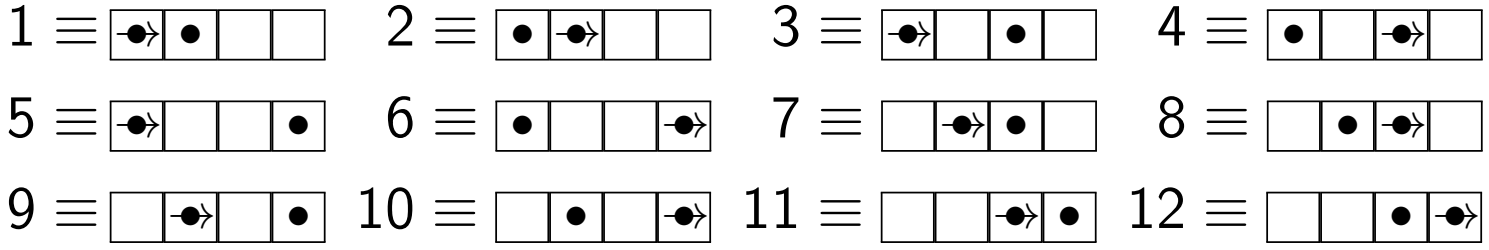


$$P^{TASEP} = \frac{1}{2}\begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & 1 & 1 & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & 1 & \cdot & \cdot & \cdot & 1 \end{bmatrix} .$$

Département de Physique
École normale supérieure

Lifted TASEP (example)

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



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

Integrability, spectrum of $P_{N,L}^{I-T}$. Ex: $(N = 3)$

- Eqs for left eigenvector ψ, complex eigenvalue λ:

$$\lambda\psi_{\{\vec{j},k,l\}} = \beta\psi_{\{\overrightarrow{j-1},k,l\}} + \alpha\psi_{\{j,\overrightarrow{k-1},l\}}, \quad j < k-1$$
$$\lambda\psi_{\{j,\vec{k},l\}} = \beta\psi_{\{j,\overrightarrow{k-1},l\}} + \alpha\psi_{\{j,k,\overrightarrow{l-1}\}}, \quad j < k-1$$
$$\lambda\psi_{\{j,k,\vec{l}\}} = \beta\psi_{\{j,k,\overrightarrow{l-1}\}} + \alpha\psi_{\{\overrightarrow{j-1},k,l\}}, \quad j < k-1$$

  and similar equations for $j = k-1$...

- Bethe ansatz:

$$\psi_{\{\vec{j},k,l\}} = A_{\vec{\bullet}\circ\circ} z_1^j z_2^k z_3^l + B_{\vec{\bullet}\circ\circ} z_1^j z_2^l z_3^k + \ldots + F_{\vec{\bullet}\circ\circ} z_1^l z_2^k z_3^j$$
$$\psi_{\{j,\vec{k},l\}} = A_{\circ\vec{\bullet}\circ} z_1^j z_2^k z_3^l + B_{\circ\vec{\bullet}\circ} z_1^j z_2^l z_3^k + \ldots + F_{\circ\vec{\bullet}\circ} z_1^l z_2^k z_3^j$$
$$\psi_{\{j,k,\vec{l}\}} = A_{\circ\circ\vec{\bullet}} z_1^j z_2^k z_3^l + B_{\circ\circ\vec{\bullet}} z_1^j z_2^l z_3^k + \ldots + F_{\circ\circ\vec{\bullet}} z_1^l z_2^k z_3^j$$

Département de Physique
École normale supérieure

Integrability, Bethe-ansatz equations ($N = 3$)

- Comparing coefficients:

$$\prod_{a=1}^{3}\left(\lambda - \frac{\beta}{z_a}\right) = \frac{\alpha^3}{z_1 z_2 z_3}, \tag{8}$$

- Periodic boundary conditions:

$$z_a^{L-1} = \left(\frac{\lambda - \beta/z_a}{\alpha}\right) \prod_{b\neq a=1}^{3} T(z_a, z_b), \quad a = 1, 2, 3. \tag{9}$$

with

$$T(z_a, z_b) = \frac{\lambda - \alpha - \beta/z_b}{\lambda - \alpha - \beta/z_a}.$$

- Numerical solutions of eqs (8) and (9) agree with complete (complex) spectrum of $P_{3,L}^{l-T}$.
- Essler & Krauth (2023).

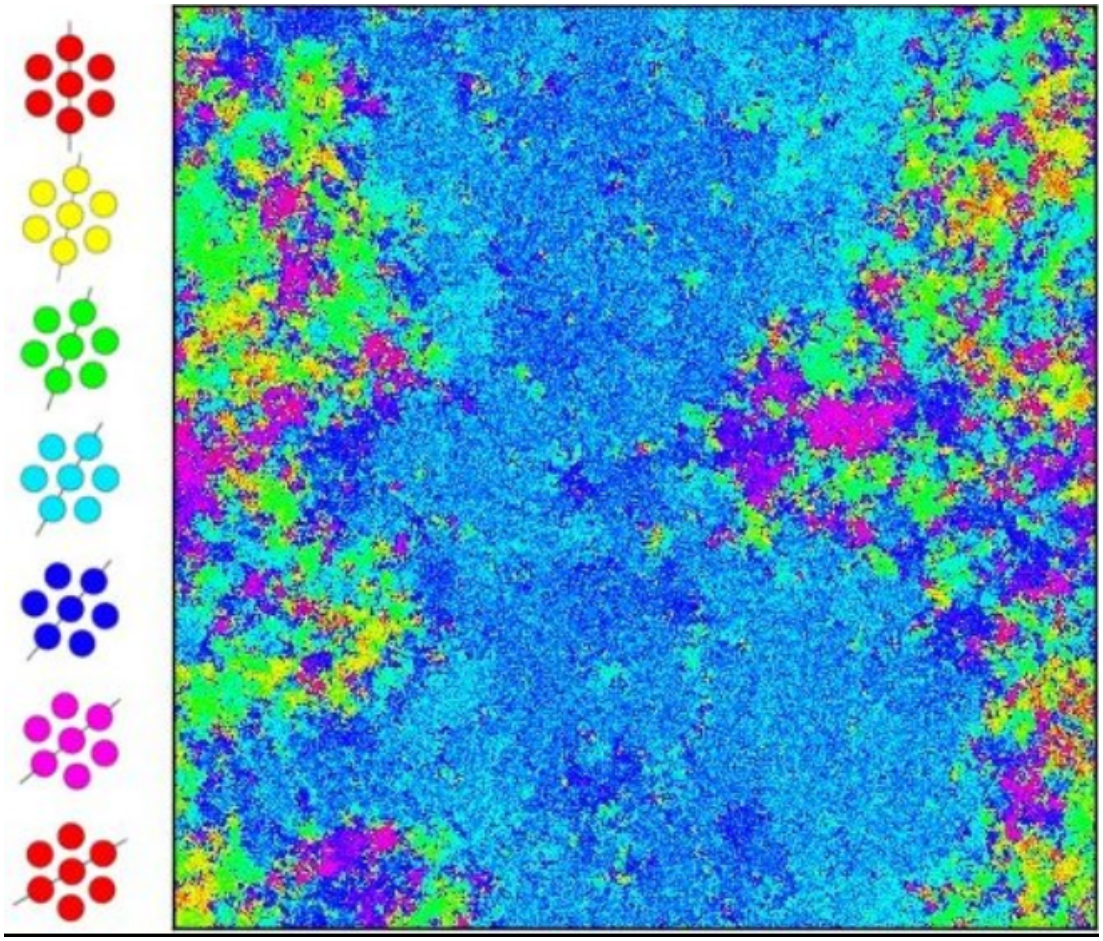
Département de Physique
École normale supérieure

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

- continuous-space versions available (Kapfer & Krauth (2017))
- see Essler & Krauth (2023)

Example IV: Equilibrium non-equilibrium



- Equilibrated sample of 10^6 disks (from Bernard & Krauth 2011, see also Li et al. 2022)

Département de Physique
École normale supérieure

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.

Département de Physique
École normale supérieure