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
Simons Center for Computational Physical Chemistry, New York University, USA

26 April 2024
Institute for Advanced Computational Science
Stony Brook University (USA)

This work was supported by a grant from the Simons Foundation

This work was supported by the Alexander von Humboldt Foundation

Département de Physique
École normale supérieure

Plan of talk

1. Introduction (+ Example I)
2. Mixing and Relaxing (+ Example II)
3. Stopping and Coupling (+ Example III)
4. Equilibrium out of Equilibrium (+ Example IV)
5. Conclusion

Département de Physique
École normale supérieure

Markov chains (1/2)

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

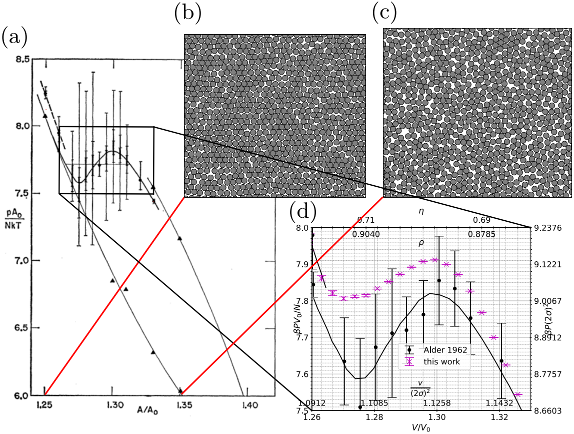
AND

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.

Example I: Equilibrated (?) samples



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

Markov chains (2/2)

- 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:

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

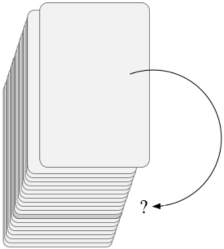
- Usually $\epsilon = 1/4$ is taken, $\epsilon = 1/e$ would be better.

Département de Physique
École normale supérieure

Plan of talk

Département de Physique
École normale supérieure

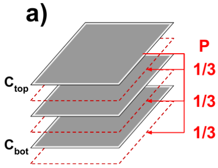
Shuffling of cards 1/5



- $\Omega_N^{\text{shuffle}} = \{\text{Permutations of } \{1, \ldots, N\}\}$
- 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



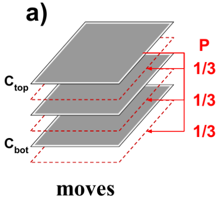
moves

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

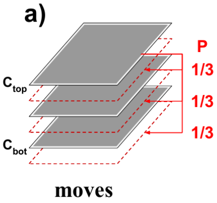


- $\Omega_3^{\text{shuffle}} = \{1 \equiv \{1,2,3\}, 2 \equiv \{1,3,2\}, 3 \equiv \{2,1,3\}, 4 \equiv \{2,3,1\}, 5 \equiv \{3,1,2\}, 6 \equiv \{3,2,1\}\}$.
- $$P = \frac{1}{3}\begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

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

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

Shuffling of cards 5/5

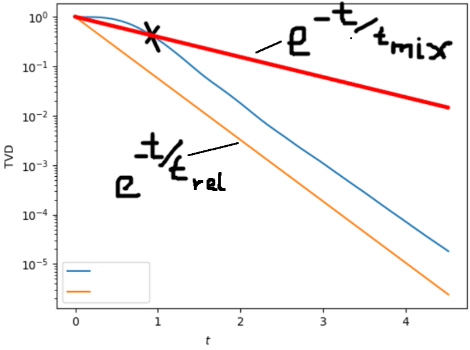



```
procedure top2random-stop
input {c_1, ..., c_n}
c_first-n ← c_n
for t = 1, 2, ... do
  { c̃_1 ← c_1
  { {c_1, ..., c_n} ← top2random({c_1, ..., c_n})
  { if (c̃_1 = c_first-n) break
output {c_1, ..., c_n, t}
```

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

Département de Physique
École normale supérieure

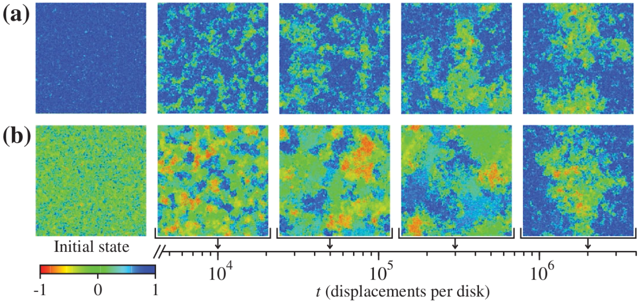
Mixing and Relaxation



- $t_{\text{mix}} = ||\pi^{\{t_{\text{mix}}\}} - \pi||_{\text{TV}} = 1/e$, (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

Example II: A non-asymptotic time scale

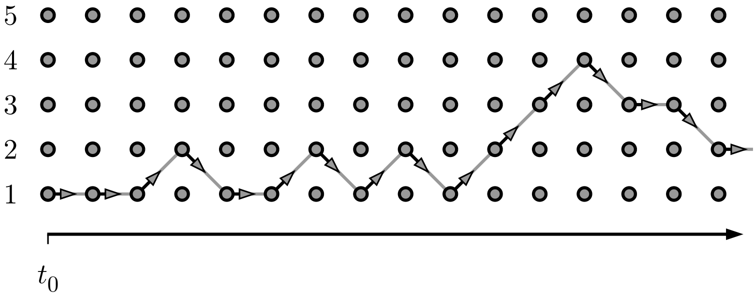


- Coarsening in hard disks (from Bernard & Krauth 2011)...
- ... an example of a non-asymptotic mixing-time scale

Plan of talk

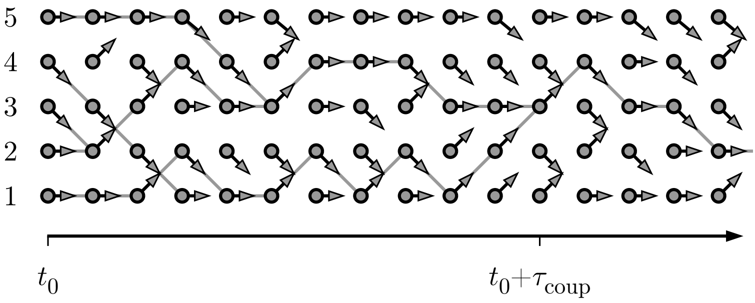
Département de Physique
École normale supérieure

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).

Département de Physique
École normale supérieure

Markov chain (random maps), coupling 2/3

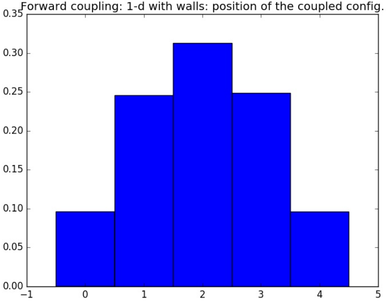
procedure `forward-coupling`
$\mathcal{P} \leftarrow \{1, \ldots, N\}$
$t \leftarrow 0$
while True:
$$\begin{cases} t \leftarrow t + 1 \\ \mathcal{P} \leftarrow \{\min\left[\max(b + \textbf{choice}\{-1, +1\}, 1), N\right] \text{ for } b \in \mathcal{P}\} \\ \textbf{if } |\mathcal{P}| = 1\textbf{: break}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

———

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

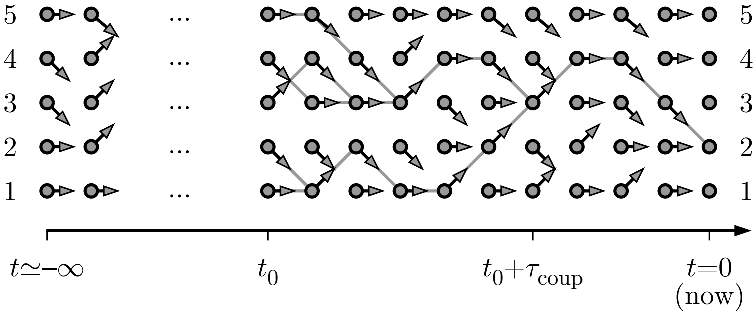
Département de Physique
École normale supérieure

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`

$t_{\text{tot}} \leftarrow 0$

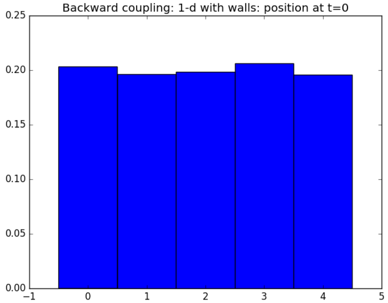
while True:

$$\begin{cases} t_{\text{tot}} \leftarrow t_{\text{tot}} - 1 \\ \mathcal{A}_{t_{\text{tot}}} \leftarrow \texttt{draw-arrows} \text{ (draw arrows at time } t_{\text{tot}}) \\ \mathcal{P} \leftarrow \{1, \ldots, N\} \\ \textbf{for } t = t_{\text{tot}}, t_{\text{tot}} + 1, \ldots, -1\textbf{:} \\ \quad \{ \; \mathcal{P} \leftarrow \{b + \mathcal{A}_t(b) \text{ for } b \in \mathcal{P}\} \\ \textbf{if } |\mathcal{P}| = 1\textbf{: break} \end{cases}$$

output $\mathcal{P}$ ((perfect) sample)

- Propp & Wilson (1997)

Coupling from the past 3/3



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

Example III: Perfect Monte Carlo samples of hard disks

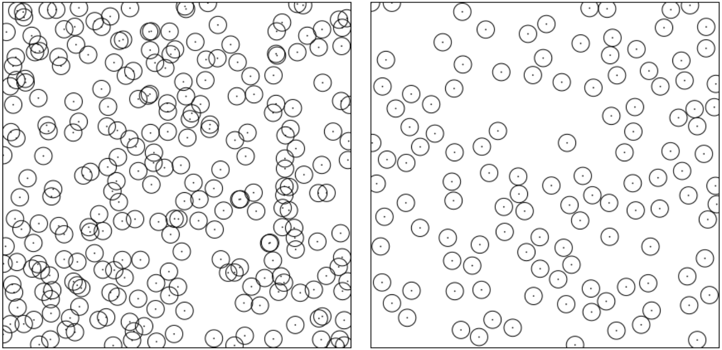
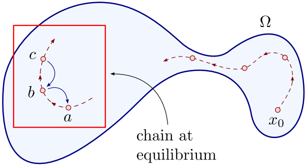


Figure 10: Perfectly random samples of the Strauss point process. In both panels the point

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

Plan of talk

Detailed balance, global balance, lifting



- Reversible transition matrices P satisfy the ‘detailed-balance’ condition:

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

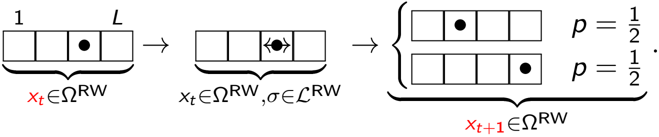
- Non-reversible transition matrices P only satisfy ‘global balance’:

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

Département de Physique
École normale supérieure

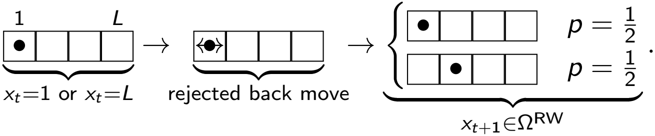
Random walk (RW) on the one-dimensional lattice

- In the bulk:



(1)

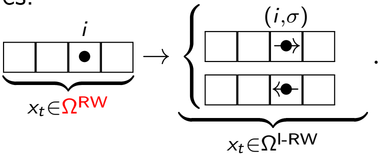
- At the boundary:



(2)

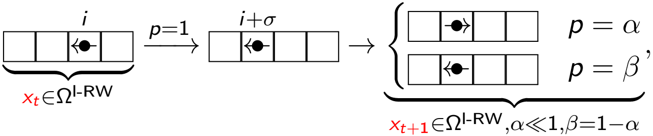
Lifted random walk (l-RW)

- Lifting of samples:



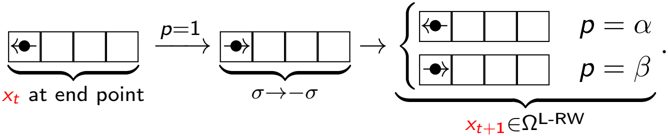
(3)

- In the bulk:



(4)

- At the boundary:



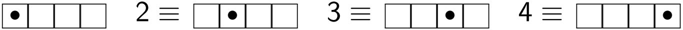
(5)

Diaconis, Holmes, Neal (2000)

Département de Physique
École normale supérieure

Random walk, lifted random walk (examples)

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

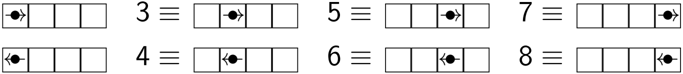


$$P^{RW}_{\text{walls}} = \frac{1}{2}\begin{bmatrix} 1 & 1 & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & 1 \\ \cdot & \cdot & 1 & 1 \end{bmatrix}$$

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

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

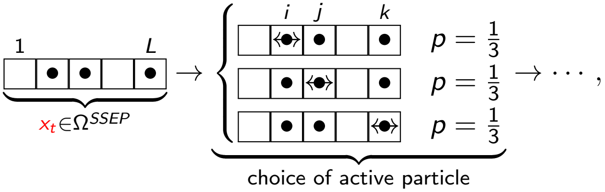
$2 \equiv$ $4 \equiv$ $6 \equiv$ $8 \equiv$



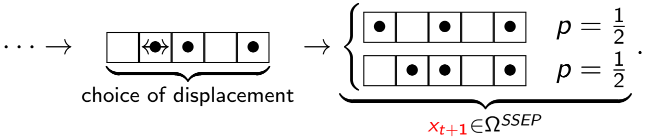
$$P^{\text{lRW}}_{\text{walls}} = \begin{bmatrix} \cdot & \cdot & \beta & \alpha & \cdot & \cdot & \cdot & \cdot \\ \beta & \alpha & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \beta & \alpha & \cdot & \cdot \\ \alpha & \beta & \cdot & \cdot & \cdot & \cdot & \cdot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

Symmetric simple exclusion process (SSEP)

- Move (first part ...)



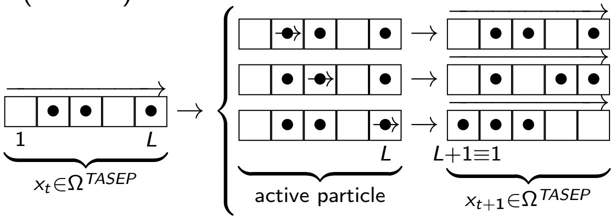
- Move (... second part)



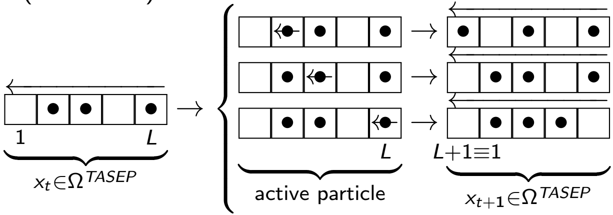
Département de Physique
École normale supérieure

Totally asymmetric simple exclusion process (TASEP)

- Move (forward)



- Move (backward)



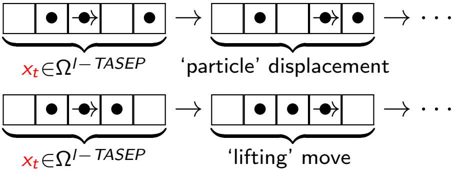
- forward–backward coupling (ad-hoc, or boundary conditions).

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

Département de Physique
École normale supérieure

Lifted TASEP (definition)

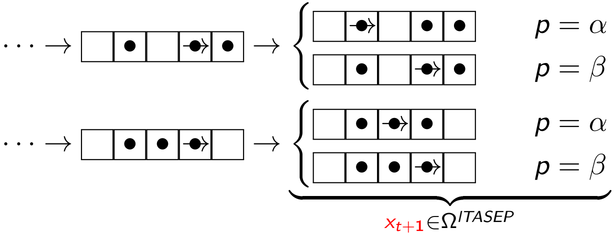
- $\Omega^{l-TASEP} = \Omega^{SSEP} \times \{-1, +1\} \times \{1, \dots, N\}$, $\mathcal{L} = \emptyset$
- Move (first part ...)



(6)

(7)

- Move (second part ...)



(6b)

(7b)

Département de Physique
École normale supérieure

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

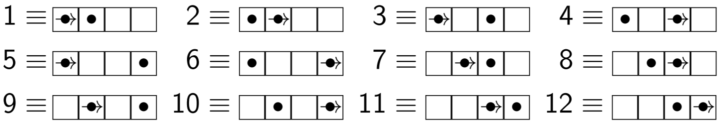
$$1 \equiv \overrightarrow{\boxed{\bullet}\boxed{\bullet}\boxed{}\boxed{}} \quad 2 \equiv \overrightarrow{\boxed{\bullet}\boxed{}\boxed{\bullet}\boxed{}} \quad 3 \equiv \overrightarrow{\boxed{\bullet}\boxed{}\boxed{}\boxed{\bullet}}$$

$$4 \equiv \overrightarrow{\boxed{}\boxed{\bullet}\boxed{\bullet}\boxed{}} \quad 5 \equiv \overrightarrow{\boxed{}\boxed{\bullet}\boxed{}\boxed{\bullet}} \quad 6 \equiv \overrightarrow{\boxed{}\boxed{}\boxed{\bullet}\boxed{\bullet}}$$

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

Lifted TASEP (example)

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



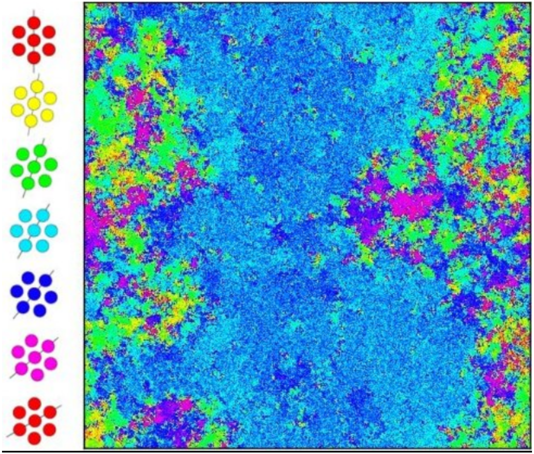
$$P = \begin{bmatrix} \alpha & \beta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \alpha & \beta & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \beta & \alpha & \cdot & \cdot & \cdot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

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

- 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 & 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