Mixing, stopping, 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)

8 November 2023

CMT Forum: Department of Physics, University of Oxford

W. Krauth; Front. Phys. (2021) (review)

G. Tartero, W. Krauth; AJP to appear (2023) (1+12 miniature algorithms)

P. Höllmer, A. C. Maggs, W. Krauth (2023) (Real-life application)

F. H. L. Essler, W. Krauth (2023) (Bethe ansatz)

Z. Lei, W. Krauth, A. C. Maggs; PRE (2019) (Factor fields)

M. Michel, S. Kapfer, W. Krauth; JCP (2014) (Factorization)

Département de Physique

École normale supérieure

Plan of talk

1. Introduction
2. Mixing and Relaxing, stopping
3. Factoring, balance-breaking
4. Moving towards real-life problems
5. Conclusion

Plan of talk

Département de Physique
École normale supérieure

Markov chains

- Sample space Ω (e.g. point particles, water molecules)
- 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:

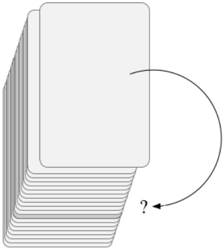
$$t_{\mathrm{mix}}(\epsilon) = \min\{t : d(t) \leq \epsilon\}$$

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

Plan of talk

Département de Physique
École normale supérieure

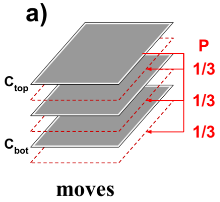
Shuffling of cards 1/6



- $\Omega_N^{\text{shuffle}} = \{\text{Permutations of } \{1, \ldots, N\}\}$
- 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 2/6

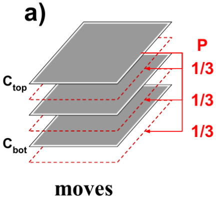


$\mathbf{procedure}$ `top-to-random`
$\mathbf{input}\ \{c_1, \ldots, c_n\}$
$i \leftarrow \mathtt{choice}(\{1, \ldots, n\})$
$\{\hat{c}_1, \ldots, \hat{c}_n\} \leftarrow \{c_2, \ldots, c_i, c_1, c_{i+1}, \ldots, c_n\}$
$\mathbf{output}\ \{\hat{c}_1, \ldots, \hat{c}_n\}$

- 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 3/6



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

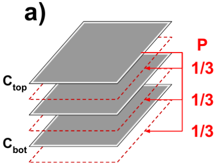
Shuffling of cards 4/6

- $$P_3^{\text{shuffle}} = \frac{1}{3}\begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$
- 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 5/6

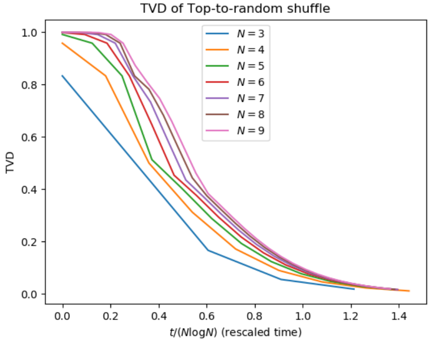


moves

$$
\begin{array}{l}
\textbf{procedure } \texttt{top2random-stop} \\
\textbf{input } \{c_1, \ldots, c_n\} \\
c_{\text{first-n}} \leftarrow c_n \\
\textbf{for } t = 1, 2, \ldots \textbf{ do} \\
\quad \left\{ \begin{array}{l}
\tilde{c}_1 \leftarrow c_1 \\
\{c_1, \ldots, c_n\} \leftarrow \texttt{top2random}(\{c_1, \ldots, c_n\}) \\
\textbf{if } (\tilde{c}_1 = c_{\text{first-n}}) \ \textbf{break}
\end{array} \right. \\
\textbf{output } \{c_1, \ldots, c_n, t\}
\end{array}
$$

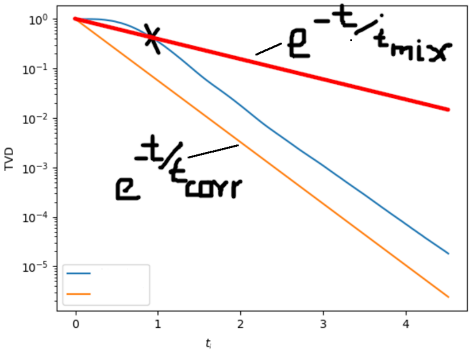
- Expected running time: $n \log n$.

Shuffling of cards 6/6



- Time scale $N \log N$ larger than inverse gap.

Mixing and Relaxation



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

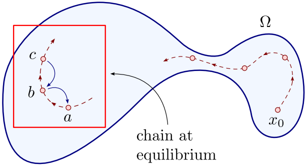
Département de Physique
École normale supérieure

Plan of talk

1. Introduction
2. Mixing and Relaxing, stopping
3. Factoring, balance-breaking
4. Moving towards real-life problems
5. Conclusion

Département de Physique
École normale supérieure

Detailed balance, global balance, lifting



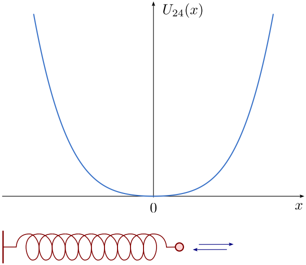
- Reversible transition matrices P satisfy the 'detailed-balance' condition:

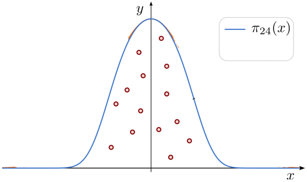
$$\pi_a P_{ab} = \pi_b P_{ba}$$

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

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

Anharmonic potential (1D potential 1/6)





- Anharmonic potential

$$U_{24}(x) = \frac{x^2}{2} + \frac{x^4}{4} \tag{1}$$

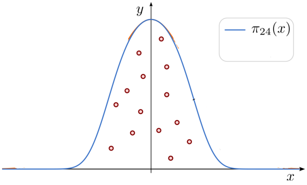
- Boltzmann distribution

$$\pi_{24}(x) = \exp\left[-\beta U_{24}(x)\right] = \pi_2(x)\pi_4(x),$$

with $\beta = 1/(kT)$: inverse temperature

Département de Physique
École normale supérieure

Metropolis algorithm (1D potential 2/6)



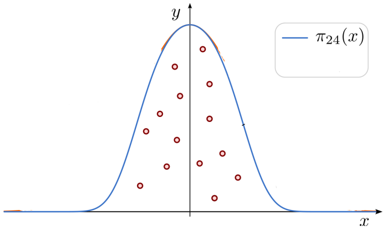
procedure `metropolis`
input x (sample at time t)
$\Delta \leftarrow \texttt{ran}(-\delta, \delta)$
$x' \leftarrow x + \Delta$
$\Upsilon \leftarrow \texttt{ran}(0, 1)$
if $\Upsilon < \min\left[1, \frac{\pi_{24}(x')}{\pi_{24}(x)}\right]$ **:** $x \leftarrow x'$
output x (sample at time $t+1$)

- satisfies detailed balance ...

$$\pi_{24}(x) \min\left[1, \frac{\pi_{24}(x')}{\pi_{24}(x)}\right] = \pi_{24}(x') \min\left[1, \frac{\pi_{24}(x)}{\pi_{24}(x')}\right]$$

- ... famous, but not the end of history.

‘Factorized’ Metropolis (naive) (1D potential 3/6)



procedure `factor-metropolis`
input x
$\Delta \leftarrow \texttt{ran}(-\delta, \delta)$
$x' \leftarrow x + \Delta$
$\Upsilon \leftarrow \texttt{ran}(0, 1)$
if $\Upsilon < \min\left[1, \frac{\pi_2(x')}{\pi_2(x)}\right] \min\left[1, \frac{\pi_4(x')}{\pi_4(x)}\right]$**:**
  $\{ x \leftarrow x'$
output x

———

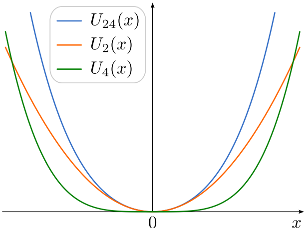
- also satisfies detailed balance...

$$\pi_{24}(x) \min\left[1, \frac{\pi_2(x')}{\pi_2(x)}\right] \min\left[1, \frac{\pi_4(x')}{\pi_4(x)}\right] =$$

$$\pi_{24}(x) \min\left[1, \frac{\pi_2(x')}{\pi_2(x)}\right] \min\left[1, \frac{\pi_4(x')}{\pi_4(x)}\right]$$

- ... but is ‘naive’.

‘Factorized’ Metropolis (patched) (1D potential 4/6)

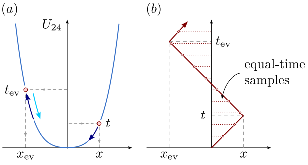


procedure `factor-metropolis(patch)`
input x
$\Delta \leftarrow \texttt{ran}(-\delta, \delta)$
$x' \leftarrow x + \Delta$
$\Upsilon_2 \leftarrow \texttt{ran}(0,1)$; $\Upsilon_4 \leftarrow \texttt{ran}(0,1)$
if $\Upsilon_2 < \min\left[1, \frac{\pi_2(x')}{\pi_2(x)}\right]$ and $\Upsilon_4 < \min\left[1, \frac{\pi_4(x')}{\pi_4(x)}\right]$:
$\{\ x \leftarrow x'$ (move accepted by consensus)
output x

——

- $$\mathbb{P}\left\{\Upsilon_2 < \min\left[1, \frac{\pi_2(x')}{\pi_2(x)}\right] \bigwedge \Upsilon_4 < \min\left[1, \frac{\pi_4(x')}{\pi_4(x)}\right]\right\} = \min\left[1, \frac{\pi_2(x')}{\pi_2(x)}\right] \min\left[1, \frac{\pi_4(x')}{\pi_4(x)}\right]$$
- ... tautological statement, wide-ranging implications

'Lifted' Metropolis (1D potential 5/6)



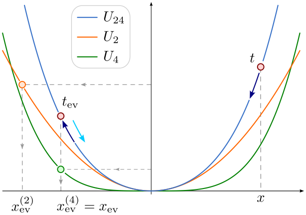
procedure `lifted-metropolis`
input $\{x, \sigma\}$ (lifted sample at time t)
$\Delta \leftarrow \texttt{ran}(0, \delta)$ $(\delta > 0)$
$x' \leftarrow x + \sigma\Delta$ (x' in direction σ from x)
$\Upsilon \leftarrow \texttt{ran}(0, 1)$
if $\Upsilon < \min\left[1, \frac{\pi_{24}(x')}{\pi_{24}(x)}\right]$**:** $x \leftarrow x'$
else: $\sigma \leftarrow -\sigma$
output $\{x, \sigma\}$ (lifted sample at time $t+1$)

——

- breaks detailed balance, satisfies global balance ...
- ... and implements a non-reversible Markov chain.

Peters, de With (2012), Bierkens, Fearnhead, Roberts (2019)

'Lifted factorized' Metropolis (1D potential 6/6)



procedure `factor-zig-zag`
input $\{x, \sigma\}, t$ (lifted sample with $\sigma x \leq 0$)
$x_{\text{ev}}^{(2)} \leftarrow \sigma \sqrt{-2\beta^{-1} \log \texttt{ran}(0,1)}$ (see Eq. (40))
$x_{\text{ev}}^{(4)} \leftarrow \sigma \sqrt[4]{-4\beta^{-1} \log \texttt{ran}(0,1)}$ (see Eq. (41))
$x_{\text{ev}} \leftarrow \sigma \min\left(|x_{\text{ev}}^{(2)}|, |x_{\text{ev}}^{(4)}|\right)$
$t_{\text{ev}} \leftarrow t + |x_{\text{ev}} - x|$
for $t^* = \text{int}(t) + 1, \ldots, \text{int}(t_{\text{ev}})$**:**
$\quad \{$ **print** $x + \sigma(t^* - t)$ (sample of π_{24})
$x \leftarrow x_{\text{ev}};\ \sigma \leftarrow -\sigma;\ t \leftarrow t_{\text{ev}}$ ("zig-zag")
output $\{x, \sigma\}, t$
——

- breaks detailed balance and is 'lifted'...
- ... coordinates independent factor decisions through consensus.

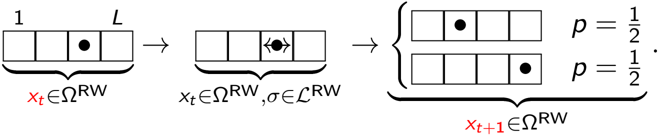
Département de Physique
École normale supérieure

Plan of talk

Département de Physique
École normale supérieure

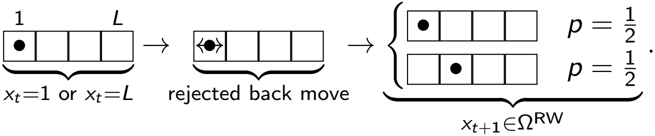
Random walk (RW) on the one-dimensional lattice

- In the bulk:



(2)

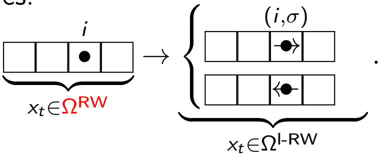
- At the boundary:



(3)

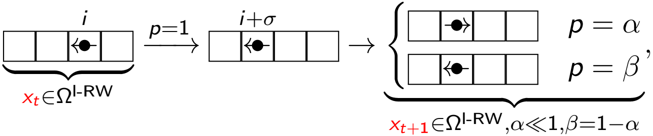
Lifted random walk (l-RW)

- Lifting of samples:



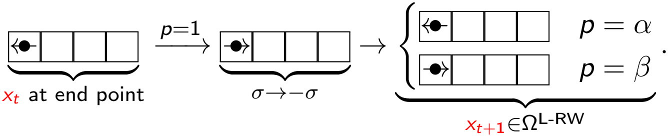
(4)

- In the bulk:



(5)

- At the boundary:



(6)

Diaconis, Holmes, Neal (2000)

Département de Physique
École normale supérieure

Random walk, lifted random walk (examples)

- $1 \equiv \boxed{\bullet}\boxed{}\boxed{}\boxed{} \quad 2 \equiv \boxed{}\boxed{\bullet}\boxed{}\boxed{} \quad 3 \equiv \boxed{}\boxed{}\boxed{\bullet}\boxed{} \quad 4 \equiv \boxed{}\boxed{}\boxed{}\boxed{\bullet}$

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

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

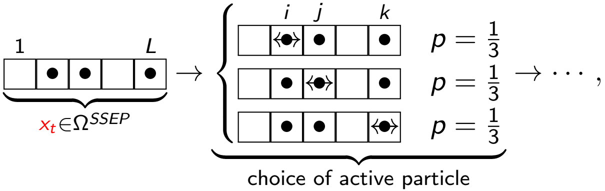
$$1 \equiv \boxed{\bullet\!\!\rightarrow}\boxed{}\boxed{}\boxed{} \quad 3 \equiv \boxed{}\boxed{\bullet\!\!\rightarrow}\boxed{}\boxed{} \quad 5 \equiv \boxed{}\boxed{}\boxed{\bullet\!\!\rightarrow}\boxed{} \quad 7 \equiv \boxed{}\boxed{}\boxed{}\boxed{\bullet\!\!\rightarrow}$$

$$2 \equiv \boxed{\leftarrow\!\!\bullet}\boxed{}\boxed{}\boxed{} \quad 4 \equiv \boxed{}\boxed{\leftarrow\!\!\bullet}\boxed{}\boxed{} \quad 6 \equiv \boxed{}\boxed{}\boxed{\leftarrow\!\!\bullet}\boxed{} \quad 8 \equiv \boxed{}\boxed{}\boxed{}\boxed{\leftarrow\!\!\bullet}$$

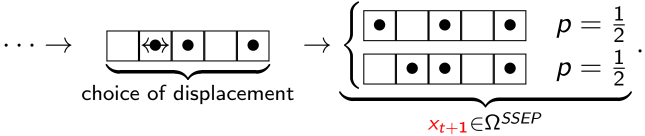
$$P^{\text{lRW}}_{\text{walls}} = \begin{bmatrix} \cdot & \cdot & \beta & \alpha & \cdot & \cdot & \cdot & \cdot \\ \beta & \alpha & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \beta & \alpha & \cdot & \cdot \\ \alpha & \beta & \cdot & \cdot & \cdot & \cdot & \cdot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

Symmetric simple exclusion process (SSEP)

- Move (first part ...)

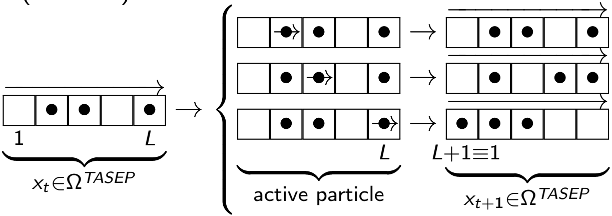


- Move (... second part)

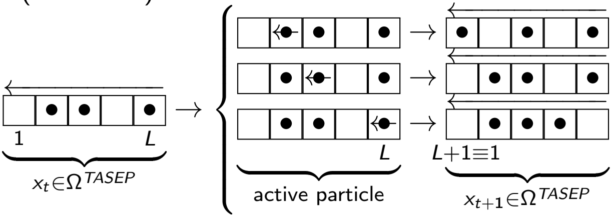


Totally asymmetric simple exclusion process (TASEP)

- Move (forward)



- Move (backward)



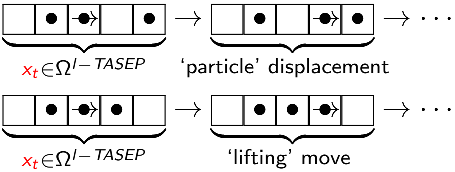
- forward–backward coupling (ad-hoc, or boundary conditions).

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

Département de Physique
École normale supérieure

Lifted TASEP (definition)

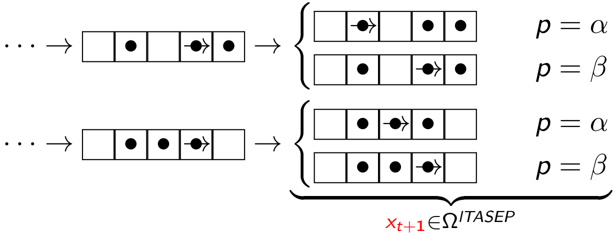
- $\Omega^{l-TASEP} = \Omega^{SSEP} \times \{-1,+1\} \times \{1,\ldots,N\}$, $\mathcal{L} = \emptyset$
- Move (first part ...)



(7)

(8)

- Move (second part ...)



(7b)

(8b)

Département de Physique
École normale supérieure

TASEP (example)

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

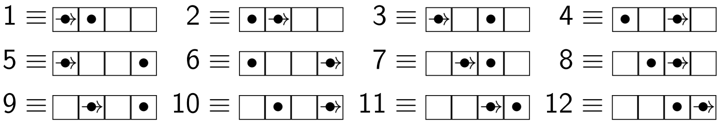
$$
1 \equiv \overrightarrow{\begin{array}{|c|c|c|c|}\hline \bullet & \bullet & & \\ \hline\end{array}} \quad
2 \equiv \overrightarrow{\begin{array}{|c|c|c|c|}\hline \bullet & & \bullet & \\ \hline\end{array}} \quad
3 \equiv \overrightarrow{\begin{array}{|c|c|c|c|}\hline \bullet & & & \bullet \\ \hline\end{array}}
$$

$$
4 \equiv \overrightarrow{\begin{array}{|c|c|c|c|}\hline & \bullet & \bullet & \\ \hline\end{array}} \quad
5 \equiv \overrightarrow{\begin{array}{|c|c|c|c|}\hline & \bullet & & \bullet \\ \hline\end{array}} \quad
6 \equiv \overrightarrow{\begin{array}{|c|c|c|c|}\hline & & \bullet & \bullet \\ \hline\end{array}}
$$

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

Lifted TASEP (example)

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



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

Département de Physique
École normale supérieure

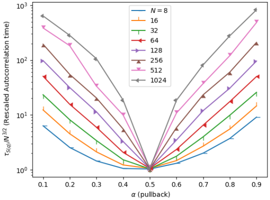
Synopsis of mixing and relaxation times

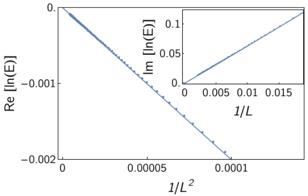
Algorithm	mixing	relaxation (inverse gap)
SSEP	$N^3 \log N$	N^3
TASEP	$N^{5/2}$	$N^{5/2}$
Lifted TASEP (optimal α)	N^2	N^2

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

Parameter-dependence of relaxation times

Algorithm	mixing	relaxation (inverse gap)
Lifted TASEP (optimal α)	N^2	N^2





Département de Physique
École normale supérieure

Conclusion

- A second revolution in Markov-chain Monte Carlo underway.
- Time scales of MCMC much better understood.
- Non-reversible MCMC is what comes after the revolution.
- Exactly solvable non-trivial models of crucial importance.
- Lifted TASEP may serve as paradigm

Département de Physique
École normale supérieure