Fast non-reversible Markov chains for one-dimensional particle systems

Werner Krauth

Laboratoire de Physique
Ecole normale supérieure
Paris, France

01 July 2022
University of Oxford (UK)
Rudolf Peierls Centre for Theoretical Physics

W. Krauth; Front. Phys. (2021)
S. Kapfer, W. Krauth; PRL (2017)
Z. Lei, W. Krauth, A. C. Maggs; PRE (2019)

A. C. Maggs, W. Krauth; PRE (2022)

Département de Physique
École normale supérieure

Metropolis et al. (1953)

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

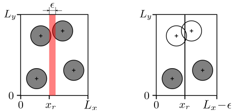
AND

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.

- NB: Equation of state: Pressure as a function of Volume.
- NNB: Pressure: (Rift-Elimination probability)/(rift volume).



Markov chain, transition matrix

- Sample space $\Omega \leftarrow$ disks in a box.
- Markov chain $\leftarrow$ Moves: Sequence of random variables $(X_0 \sim \pi^{\{0\}}, X_1 \sim \pi^{\{1\}}, X_2 \sim \pi^{\{2\}} \dots)$
  X_{t+1} depends on X_t through a transition matrix P.
- A priori probability$\rightarrow$ split matrix: $P_{ij} = \mathcal{A}_{ij}\mathcal{P}_{ij}$ for $i \neq j$
  $\mathcal{A} \Leftrightarrow$ a priori probability; $\mathcal{P} \Leftrightarrow$ filter
  Examples: Metropolis filter, heatbath filter.
- Monte Carlo rejections $\rightarrow P_{ii} \Leftrightarrow$ (filter) rejection probability.
  NB: Modern MCMC algorithms often have no rejections.

NB: Double role of P:

1. For probability distributions: $\pi^{\{t+1\}} = \pi^{\{t\}} P$ (with $\pi^{\{t\}}, \pi^{\{t+1\}}$ non-explicit objects, often even for $t \rightarrow \infty$).
2. For samples: P_{ij}: explicit probability to move from i to j.

Irreducibility

- P irreducible $\Leftrightarrow$ any i can be reached from any j.
- $\pi^{\{0\}}$: Initial probability (explicit, user-supplied). Often concentrated on a single sample $x \in \Omega$.
- P irreducible $\Rightarrow$ unique stationary distribution π with

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

- This is the steady-state version of

$$\pi_i^{\{t\}} = \sum_{j \in \Omega} \pi_j^{\{t-1\}} P_{ji} \quad \forall i \in \Omega.$$

NB: Transition matrix P is stochastic, that is, $\sum_j P_{ij} = 1$.

Département de Physique
École normale supérieure

Probability flows—Global-balance condition

- Global-balance condition:

$$\pi_i = \sum_{j \in \Omega} \underbrace{\overbrace{\pi_j P_{ji}}^{\text{flow } j \to i}}_{\mathcal{F}_{ji}} \qquad \forall i \in \Omega.$$

(NB: This is the steady state of $\pi_i^{\{t+\mathbf{1}\}} = \sum_{j \in \Omega} \pi_j^{\{t\}} P_{ji}$)

- Global-balance condition (second formulation):

$$\pi_i = \overbrace{\sum_{j \in \Omega} \mathcal{F}_{ji}}^{\text{flows entering } i} \qquad \forall i \in \Omega,$$

$$\overbrace{\sum_{k \in \Omega} \mathcal{F}_{ik}}^{\text{flows exiting } i} = \overbrace{\sum_{j \in \Omega} \mathcal{F}_{ji}}^{\text{flows entering } i} \qquad \forall i \in \Omega,$$

(NB: stochasticity condition used $\sum_{k \in \Omega} P_{ik} = 1$).

Département de Physique
École normale supérieure

Reversibility—Detailed-balance condition

- Reversible P satisfies the 'detailed-balance' condition:

$$\underbrace{\overbrace{\pi_i P_{ij}}^{\text{flow } i \to j}}_{\mathcal{F}_{ij}} = \underbrace{\overbrace{\pi_j P_{ji}}^{\text{flow } j \to i}}_{\mathcal{F}_{ji}} \quad \forall i, j \in \Omega.$$

- General P satisfies the 'global-balance' condition

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

- Detailed balance implies global balance.
- Checking detailed balance is easier than checking global balance

Département de Physique
École normale supérieure

Spectrum of reversible transition matrix

- Reversible P:
$$\pi_i P_{ij} = \pi_j P_{ji} \quad \forall i, j \in \Omega.$$
- Reversible P: $A_{ij} = \pi_i^{1/2} P_{ij} \pi_j^{-1/2}$ is symmetric.
- Reversible P:
$$\sum_{j \in \Omega} \underbrace{\pi_i^{1/2} P_{ij} \pi_j^{-1/2}}_{A_{ij}} x_j = \lambda x_i \Leftrightarrow \sum_{j \in \Omega} P_{ij} \left[\pi_j^{-1/2} x_j\right] = \lambda \left[\pi_i^{-1/2} x_i\right].$$
- P and A have same eigenvalues.
- A symmetric: (Spectral theorem): All eigenvalues real, can expand on eigenvectors.
- Irreducible, aperiodic: Single eigenvalue with $\lambda = 1$, all others smaller in absolute value.

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:

$$t_{\mathrm{mix}}(\epsilon) = \min\{t : d(t) \leq \epsilon\} \quad (\epsilon < \frac{1}{2})$$

NB: '$\max_{\pi^{\{\mathbf{0}\}}}$' $\equiv$ 'worst initial distribution $\pi^{\{\mathbf{0}\}}$'

Département de Physique
École normale supérieure

Conductance (bottleneck ratio)

$$\Phi \equiv \min_{S \subset \Omega, \pi_S \le \frac{1}{2}} \frac{\mathcal{F}_{S \to \overline{S}}}{\pi_S} = \min_{S \subset \Omega, \pi_S \le \frac{1}{2}} \frac{\sum_{i \in S, j \notin S} \pi_i P_{ij}}{\pi_S}.$$

- Reversible Markov chains:

$$\frac{1}{\Phi} \le \tau_{\mathrm{corr}} \le \frac{8}{\Phi^2}$$

('$\le$': Sinclair & Jerrum (1986), Lemma (3.3))

- Arbitrary Markov chain (see Chen et al. (1999)):

$$\frac{1}{4\Phi} \le \mathcal{A} \le \frac{20}{\Phi^2},$$

(set time: Expectation of $\max_S (t_S \times \pi_S)$ from equilibrium)

NB: One bottleneck, not many. Lower <u>and</u> upper bound.

Département de Physique
École normale supérieure

Lifting (Chen et al. (1999))

- Markov chain $\Pi \Leftrightarrow$ Lifted Markov chain $\hat{\Pi}$
- $\Omega \ni v$ (sample space) $\Leftrightarrow \hat{\Omega} \ni i$ (lifted sample space)
- P (transition matrix) $\Leftrightarrow \hat{P}$ (lifted transition matrix)
- π_v (stationary probability) $\Leftrightarrow \hat{\pi}_i$
- **Condition 1**: sample space is copied ('lifted'), π preserved

$$\pi_v = \hat{\pi}\left[f^{-1}(v)\right] = \sum_{i \in f^{-1}(v)} \hat{\pi}_i,$$

- **Condition 2**: flows are preserved

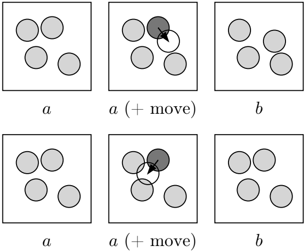
$$\underbrace{\pi_v P_{vu}}_{\text{collapsed flow}} = \sum_{i \in f^{-1}(v), j \in f^{-1}(u)} \overbrace{\hat{\pi}_i \hat{P}_{ij}}^{\text{lifted flow}} .$$

- Π and $\hat{\Pi}$ have the same conductance.

Département de Physique
École normale supérieure

Metropolis algorithm / reversibility

1. The Metropolis et al. (1953) algorithm is reversible.



2. The algorithm used by Metropolis et al. (1953) is non-reversible.

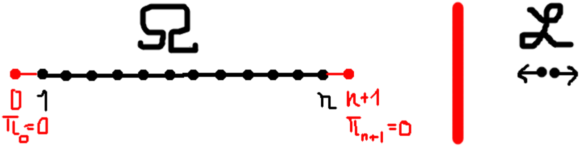
1088 METROPOLIS, ROSENBLUTH, ROSENBLUTH, TELLER, AND TELLER

Our method in this respect is similar to the cell method except that our cells contain several hundred particles instead of one. One would think that such a sample would be quite adequate for describing any one-phase system. We do find, however, that in two-phase systems the surface between the phases makes quite a perturbation. Also, statistical fluctuations may be

configurations with a probability $\exp(-E/kT)$ and weight them evenly.

This we do as follows: We place the N particles in any configuration, for example, in a regular lattice. Then we move each of the particles in succession according to the following prescription:

Metropolis algorithm on path graph (1/3)



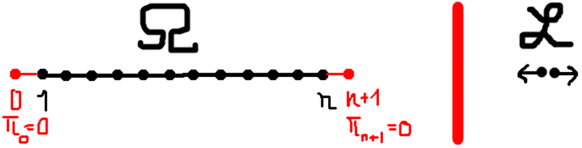
- Sample space = path graph $\Omega = \{1, \ldots, n\}$.
- Phantom vertices and edges.

Metropolis algorithm (NB: $P_{ij} = \mathcal{A}_{ij}\mathcal{P}_{ij}$ for $i \neq j$):

1. Move set $\mathcal{L} = \{+, -\}$.
2. Flat a priori probability $\mathcal{A}$: $\rightarrow \sigma =$ choice($\mathcal{L}$).
3. Metropolis filter: Accept with probability $\min(1, \pi_j/\pi_i)$.
   Reject: Don't move.

Département de Physique
École normale supérieure

Metropolis algorithm on path graph (2/3)



- Detailed balance:

$$\underbrace{\pi_i P_{ij}}_{\mathcal{F}_{ij}} = \underbrace{\pi_j P_{ji}}_{\mathcal{F}_{ji}}$$

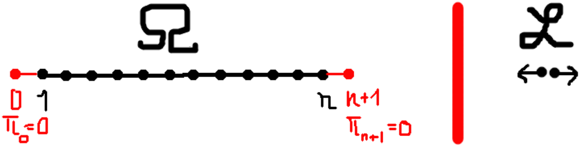
- Metropolis algorithm:

$$\mathcal{F}_{ij} = \frac{1}{2}\min\left(\pi_i, \pi_j\right) \Leftrightarrow P_{ij} = \frac{1}{2}\min\left(1, \pi_j/\pi_i\right)$$

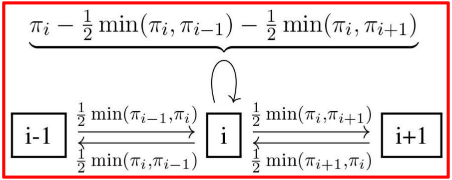
- Metropolis filter (NB: $P_{ij} = \mathcal{A}_{ij}\mathcal{P}_{ij}$):

$$\mathcal{P}_{ij} = \min\left(1, \pi_j/\pi_i\right)$$

Département de Physique
École normale supérieure



- Global balance ($\pi_i = \sum_j \pi_j P_{ji} = \sum_j \mathcal{F}_{ji}$):

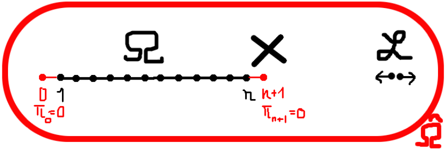


- Crucial role of rejections.

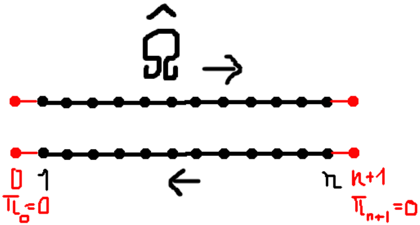
Lifting on the path graph (1/2)

General probability distribution $\pi = (\pi_1, \ldots, \pi_n)$

- 'Lifted' sample space $\hat{\Omega} = \{1, \ldots, n\} \times \{+, -\}$:



- 'Lifted' non-reversible Markov chain $\hat{\Omega} = \Omega \times \{-, +\}$:

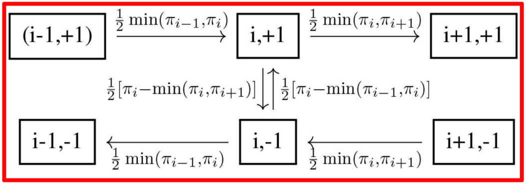


- Diaconis et al. (2000) '

Département de Physique
École normale supérieure

Lifting on the path graph (2/2)

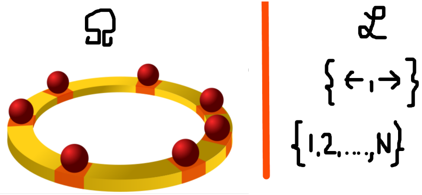
- 'Lifted' non-reversible Markov chain: NB:' only Transport treated



NB: The $\frac{1}{2} \Leftrightarrow \hat{\pi}_{i,\sigma} = \frac{1}{2}\pi_i$

- 'lifted' samples (i,σ) with $\hat{\pi}(i,\sigma) = \frac{1}{2}\pi(i)$.
- Rejections $P_{i,i}$ replaced by lifting moves $(i,\sigma) \to (i,-\sigma)$.

1d hard spheres 1/2



- N spheres, with a sample space Ω, and a move space $\mathcal{L}$.
- $\mathcal{L} = \{-, +\} \times \{1, \ldots, N\}$. Moves sampled from $\mathcal{L}$ at each time step.

Many choices for non-reversible liftings:

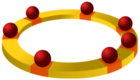
Sequential $\hat{\Omega} = \Omega \times \{1, \ldots, N\}$: Move one disk after the other.

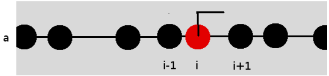
Forward $\hat{\Omega} = \Omega \times \{(-), +\}$: Move only in forward direction.

Particle-lifted forward $\hat{\Omega} = \Omega \times \{1, \ldots, N\} \times \{(-), +\}$: Always move the same disk forward, until it is blocked...

Particle-lifted Forward Metropolis algorithm

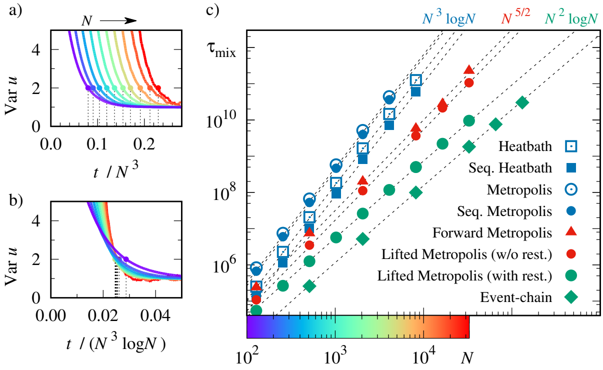
- Move i forward until it is rejected by $i+1$.
- Then move $i+1$ forward until it is rejected, etc.





- $$\mathcal{F}^{\text{lift}}_{(a,i)} = \mathcal{A}^{+}_{i} + \mathcal{R}^{+}_{i-1} = 1.$$
- NB: 1 time step: 1 particle move OR 1 lifting move
- Rejections replaced by liftings $(a, i) \to (a, i+1)$.
- Limit infinitesimal step size: ‘Event-chain Monte Carlo’.

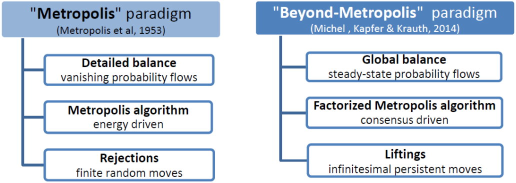
1d hard spheres 2/2



Algorithm	mixing	discrete analogue
Rev. Metropolis	$N^3 \log N$	Symmetric SEP
Forward Metropolis, Lifted (∞)	$N^{5/2}$	TASEP
Event-chain, Lifted (restarts)	$N^2 \log N$	lifted TASEP

- Kapfer—Krauth (2017)

Factorized Metropolis algorithm



- Metropolis algorithm (Metropolis et al (1953))

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

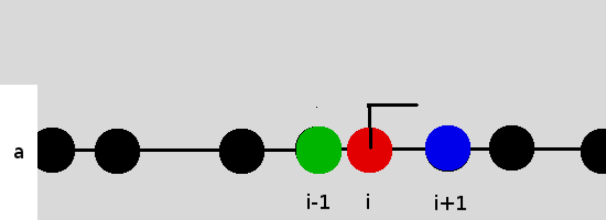
- Factorized Metropolis algorithm (Michel, Kapfer, Krauth (2014) - consensus)

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

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

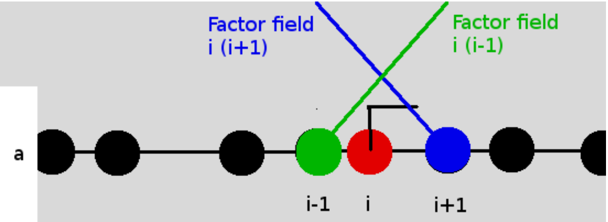
Event-chain algorithm with factor fields (1/4)

- Hard-sphere event-chain algorithm (standard version):



Event-chain algorithm with factor fields (2/4)

- Hard-sphere event-chain algorithm (factor field version):



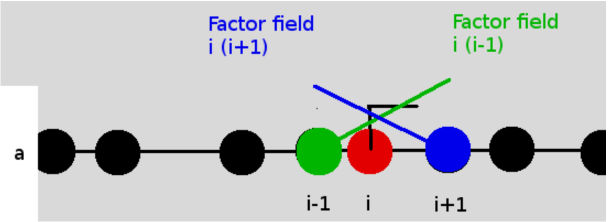
- Adding a constant term to the global energy...

$$U^{\text{fact}} = h \sum_i (x_{i+1} - x_i)$$

- ... will show that it profoundly changes the dynamics.

Event-chain algorithm with factor fields (3/4)

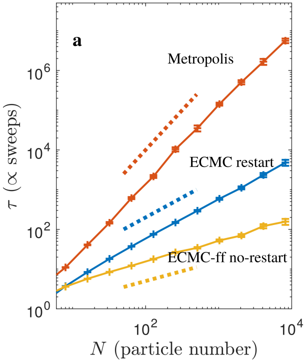
- Hard-sphere event-chain algorithm (variable factor field):



$$U^{\text{fact}} = h \sum_i (x_{i+1} - x_i)$$

Event-chain algorithm with factor fields (4/4)

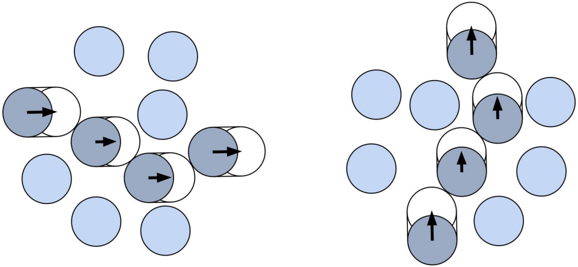
- Scaling of auto-correlation times (optimal factor field):



$$U^{\text{fact}} = h \sum_i (x_{i+1} - x_i)$$

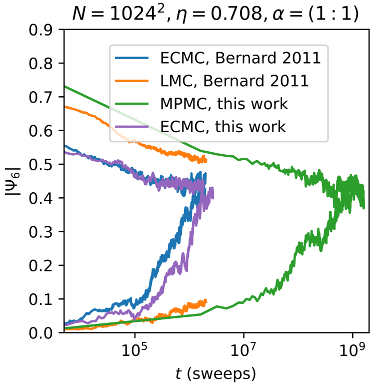
- Algebraic correlations of event steps $u \in \{-1, 1\}$ with event time s: $\langle u(0)u(s) \rangle \sim s^{-2/3}$.
- Lei, Krauth, Maggs (PRE, 2019).

Hard disks: event-chain Monte Carlo (ECMC)



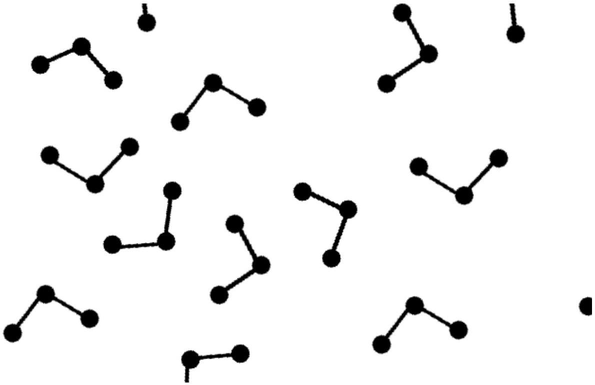
- Bernard, Krauth, Wilson (2009).
- Michel, Kapfer, Krauth (2014) (smooth potentials).
- Many variants.

ECMC and the hard-disk model



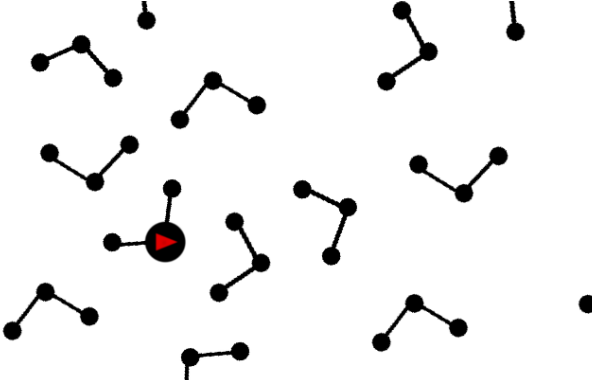
- 10^9 sweeps $\equiv$ 11.4 years (Metropolis, LMC, MPMC)
- 10^6 sweeps $\equiv$ 4.2 days (Event-chain Monte Carlo)

All-Atom Coulomb problem (1/5)



- 3D water model: bond, bending, Lennard-Jones, Coulomb (SPC/Fw).

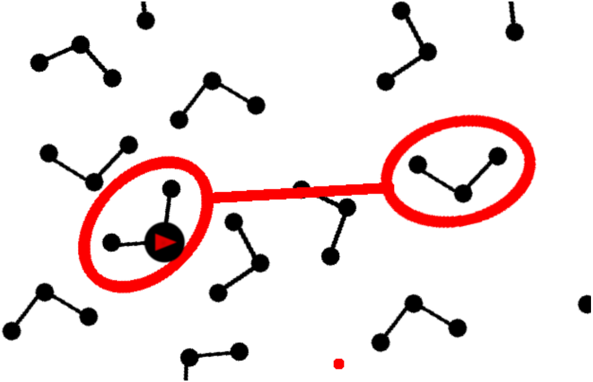
All-Atom Coulomb problem (2/5)



- 3D water model: bond, bending, Lennard-Jones, Coulomb (SPC/Fw).
- Factors and types.

Département de Physique
École normale supérieure

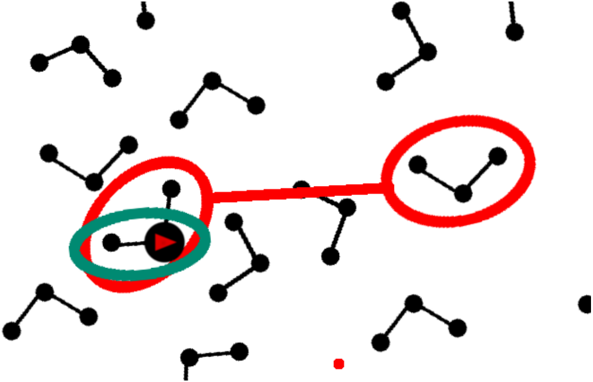
All-Atom Coulomb problem (3/5)



- Factor $M = (I_M, T_M)$: $|I_M| = 6$, two molecules. $T_M =$ 'Coulomb'.

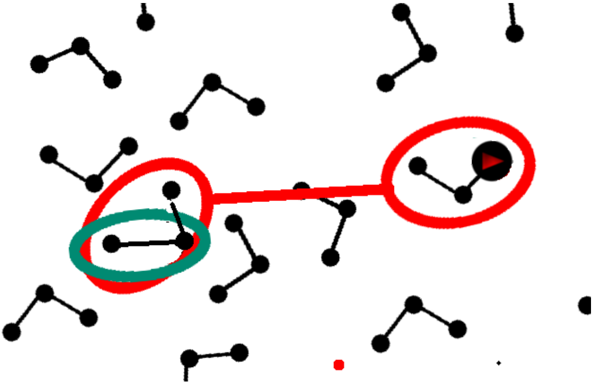
Département de Physique
École normale supérieure

All-Atom Coulomb problem (4/5)



- Water model: bond, bending, Lennard-Jones, Coulomb (SPC/Fw).

All-Atom Coulomb problem (5/5)

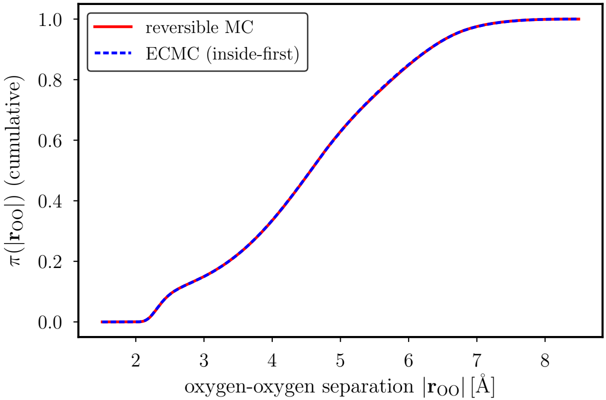


- This is the cell-veto algorithm (Kapfer, Krauth (2016)).
- Thinning, Walker (1977).

Département de Physique
École normale supérieure

ECMC for all-atom water simulations

- ECMC: Event-driven, approximation-free, canonical.
- here oxygen–oxygen distance for 32 water molecules.



See: Faulkner, Qin, Maggs, Krauth (2018).

Conclusions

- Non-reversible lifted Markov chains: From a single particle to the SPC/Fw water model.
- Detailed balance - global balance
- Sampling $\exp(-\beta U)$ without knowing U