# Recent developments in sampling methods

Manon Michel

CNRS, Laboratoire de mathématiques Blaise Pascal, Université Clermont-Auvergne

April, 4th 2022, IXXI, Lyon
Machine Learning and sampling methods for climate and physics

Particles: S. Kapfer (Erlangen), W. Krauth (ENS)

PDMP: A. Monemvassitis, A. Guillin (UCA)

Polymer: T. A. Kampmann, J. Kierfeld (Dortmund)

Bayesian inference: A. Durmus (ENS Saclay)

Complexity: Y. Deng, X. Tan (Hefei)

Normalizing flows: T. Guyon, V. Souveton, A. Guillin (UCA), G. Lavaux (IAP), J. Jasche (SU)

# Outline

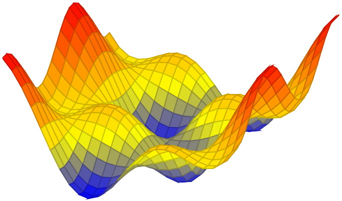
Sampling and the Monte carlo method

Upgrading the dynamics

Reducing the computational complexity

Producing non-local moves

## In inference



### Optimization

$$x \longrightarrow \boxed{f(x,\theta)} \longrightarrow y$$

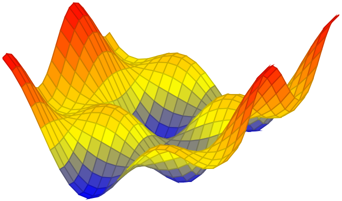
Find best $\theta_{\text{min}}$ minimizing some score function/maximizing the likelihood.

### Bayesian approach: from deterministic to probabilistic approach

Look at the full probability distribution
$P(\theta|(x,y)) \propto P((x,y)|\theta) \cdot P_{\text{prior}}(\theta)$

- Full information, uncertainty quantification
- Model flexibility (hierarchical, $P(\theta) = P(y|\theta)P(\theta|\gamma)P(\gamma)$)
- Analogy with energy landscapes in statistical physics

# In inference



## Optimization

$$x \longrightarrow \boxed{f(x,\theta)} \longrightarrow y$$

Find best $\theta_{\text{min}}$ minimizing some score function/maximizing the likelihood.

## Bayesian approach: from deterministic to probabilistic approach

Look at the full probability distribution
$P(\theta|(x,y)) \propto P((x,y)|\theta) \cdot P_{\text{prior}}(\theta)$

- Full information, uncertainty quantification
- Model flexibility (hierarchical, $P(\theta) = P(y|\theta)P(\theta|\gamma)P(\gamma)$)
- Analogy with energy landscapes in statistical physics

$\rightarrow$ Description by high-dimensional integrals!

## Sampling by Markov-chain Monte Carlo

### Goal

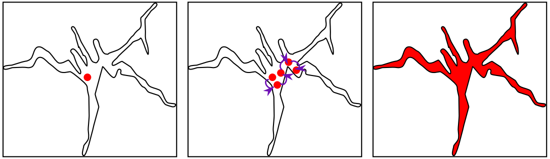
| high-dimensional integral | | Average over random $x_i$ | | Generate $x \sim \pi(x)$ |
|---|---|---|---|---|
| $\langle\theta\rangle = \int_\Omega \pi(\mathrm{d}x)\theta(x)$ | $\Longleftrightarrow$ | $\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N}\theta(x_i)$ | $\Longleftrightarrow$ | $\mathrm{rand}(0,1) \to \pi$ |
| $\pi(\mathrm{d}x) \propto \exp(-\beta E(x))\mathrm{d}x$ | | | | |

# Sampling by Markov-chain Monte Carlo

## Goal

| high-dimensional integral<br>$\langle\theta\rangle = \int_\Omega \pi(\mathrm{d}x)\theta(x)$<br>$\pi(\mathrm{d}x) \propto \exp(-\beta E(x))\mathrm{d}x$ | $\Longleftrightarrow$ | Average over random $x_i$<br>$\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N}\theta(x_i)$ | $\Longleftrightarrow$ | Generate $x \sim \pi(x)$<br>$\mathrm{rand}(0,1) \to \pi$ |
|---|---|---|---|---|

## Markov process $K(\cdot)$



## Sampling by Markov-chain Monte Carlo

### Goal

high-dimensional integral
$\langle\theta\rangle = \int_\Omega \pi(\mathrm{d}x)\theta(x)$
$\pi(\mathrm{d}x) \propto \exp(-\beta E(x))\mathrm{d}x$

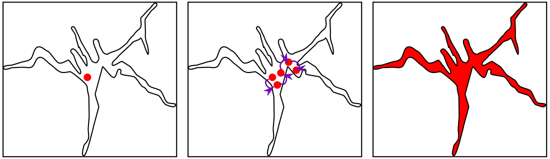
$\Longleftrightarrow$

Average over random $x_i$
$\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N}\theta(x_i)$

$\Longleftrightarrow$

Generate $x \sim \pi(x)$
$\mathrm{rand}(0,1) \to \pi$

### Markov process $K(\cdot)$



Master equation

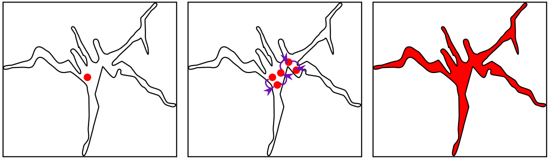
$$\frac{\mathrm{d}P(\mathrm{d}x,t)}{\mathrm{d}t} = \int_\Omega (P(\mathrm{d}x',t)K(x',\mathrm{d}x) - P(\mathrm{d}x,t)K(x,\mathrm{d}x'))$$

# Sampling by Markov-chain Monte Carlo

## Goal

high-dimensional integral
$\langle\theta\rangle = \int_\Omega \pi(\mathrm{d}x)\theta(x)$
$\pi(\mathrm{d}x) \propto \exp(-\beta E(x))\mathrm{d}x$
$\Longleftrightarrow$
Average over random $x_i$
$\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N}\theta(x_i)$
$\Longleftrightarrow$
Generate $x \sim \pi(x)$
$\mathrm{rand}(0,1) \to \pi$

## Markov process $K(\cdot)$



Master equation

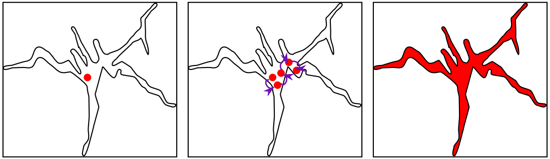
$$\frac{\mathrm{d}P(\mathrm{d}x,t)}{\mathrm{d}t} = \int_\Omega \left(P(\mathrm{d}x',t)K(x',\mathrm{d}x) - P(\mathrm{d}x,t)K(x,\mathrm{d}x')\right)$$

$$\frac{\mathrm{d}\pi(\mathrm{d}x)}{\mathrm{d}t} = \underbrace{0 = \int_\Omega \left(\pi(\mathrm{d}x')K(x',\mathrm{d}x) - \pi(\mathrm{d}x)K(x,\mathrm{d}x')\right)}_{\textbf{Global balance}}$$

And $\pi$ unique by **ergodicity**.

# Sampling by Markov-chain Monte Carlo

## Markov process



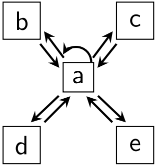
Detailed balance $\pi(\mathrm{d}x')K(x',\mathrm{d}x)=\pi(\mathrm{d}x)K(x,\mathrm{d}x')$

Master equation

$$\frac{\mathrm{d}P(\mathrm{d}x,t)}{\mathrm{d}t}=\int_\Omega \left(P(\mathrm{d}x',t)K(x',\mathrm{d}x)-P(\mathrm{d}x,t)K(x,\mathrm{d}x')\right)$$

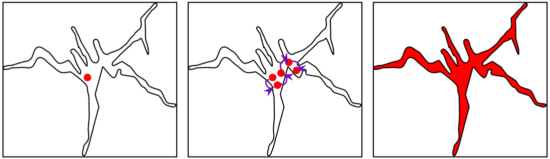
$$\frac{\mathrm{d}\pi(\mathrm{d}x)}{\mathrm{d}t}=\underbrace{0=\int_\Omega \left(\pi(\mathrm{d}x')K(x',\mathrm{d}x)-\pi(\mathrm{d}x)K(x,\mathrm{d}x')\right)}_{\textbf{Global balance}}$$

And $\pi$ unique by **ergodicity**.



# Sampling by Markov-chain Monte Carlo

## Markov process



Master equation

$$\frac{\mathrm{d}P(\mathrm{d}x,t)}{\mathrm{d}t} = \int_\Omega (P(\mathrm{d}x',t)K(x',\mathrm{d}x) - P(\mathrm{d}x,t)K(x,\mathrm{d}x'))$$

$$\frac{\mathrm{d}\pi(\mathrm{d}x)}{\mathrm{d}t} = \underbrace{0 = \int_\Omega (\pi(\mathrm{d}x')K(x',\mathrm{d}x) - \pi(\mathrm{d}x)K(x,\mathrm{d}x'))}_{\textbf{Global balance}}$$

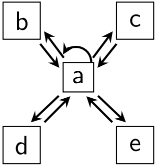
And $\pi$ unique by **ergodicity**.

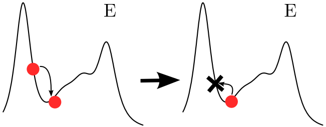
Detailed balance $\pi(\mathrm{d}x')K(x',\mathrm{d}x) = \pi(\mathrm{d}x)K(x,\mathrm{d}x')$

$K(x,\mathrm{d}x') = q(x,x')a(x,x')\mathrm{d}x'$
$+(1 - \int_\Omega q(x,y)a(x,y)\mathrm{d}y)\delta_{x=x'}$

$a(x,x') = \min\left(1, \frac{q(x',x)}{q(x,x')}\exp(-\beta\Delta E_{xx'})\right)$

Hastings-Metropolis algorithm
(Metropolis et al (1953), Hastings (1977))




Rejection

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

**Equation of State Calculations by Fast Computing Machines**

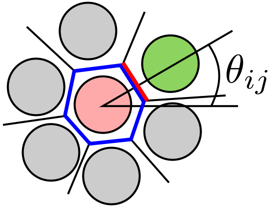
NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
*Los Alamos Scientific Laboratory, Los Alamos, New Mexico*

AND

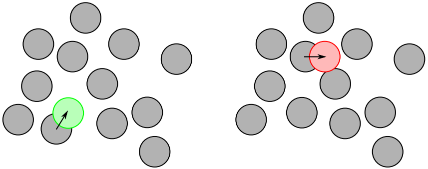
EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(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 algorithm



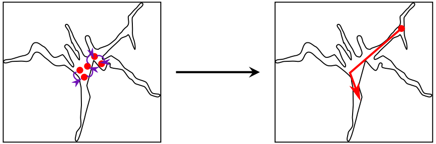
## Diffusive dynamics

- Correlated sample: $\sigma^2(\bar{\Theta}) \propto \tau(\Theta)$
  $C_\Theta(t) = \frac{\langle \Theta(t'+t)\Theta(t')\rangle - \langle \Theta^2 \rangle}{\langle \Theta^2 \rangle - \langle \Theta \rangle^2}$
- Around 2nd order phase transition $\tau \propto \xi^z \propto L^z$
  $C_\Theta(t) \sim \exp(-t/\tau)$

## Challenges

$$K(x,\mathrm{d}x') = q(x,x')a(x,x')\mathrm{d}x' + \left(1 - \int_\Omega q(x,y)a(x,y)\mathrm{d}y\right)\delta_{x=x'}$$

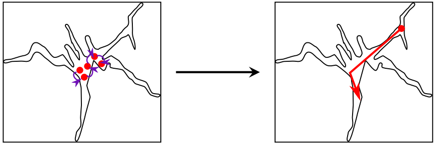
Efficient dynamics over the state space?



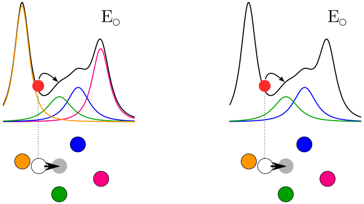
## Challenges

$$K(x,\mathrm{d}x') = q(x,x')a(x,x')\mathrm{d}x' + \left(1 - \int_\Omega q(x,y)a(x,y)\mathrm{d}y\right)\delta_{x=x'}$$

### Efficient dynamics over the state space?



### Computational complexity of each move?

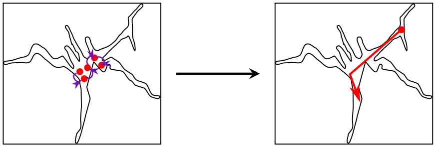


Computation of $N$ terms → Only computation of a few terms?

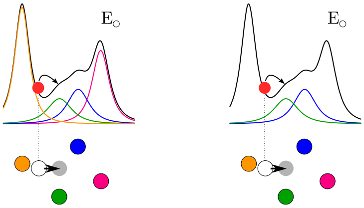
## Challenges

$$K(x, \mathrm{d}x') = q(x, x')a(x, x')\mathrm{d}x' + \left(1 - \int_{\Omega} q(x, y)a(x, y)\mathrm{d}y\right)\delta_{x=x'}$$

### Efficient dynamics over the state space?

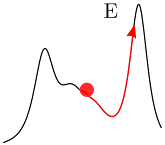


### Computational complexity of each move?




Computation of $N$ terms → Only computation of a few terms?

### High energy barrier and non-local moves?



# Outline

Sampling and the Monte carlo method

Upgrading the dynamics

Reducing the computational complexity

Producing non-local moves

# Outline

Sampling and the Monte carlo method

Upgrading the dynamics

- Non-reversibility, Event-chain Monte Carlo
- Event-chain Monte Carlo
- Piecewise deterministic Markov processes
- Invariance through interplay of transport and direction changes
- Replacing time reversibility by potential symmetries

Reducing the computational complexity

Producing non-local moves

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

**Equation of State Calculations by Fast Computing Machines**

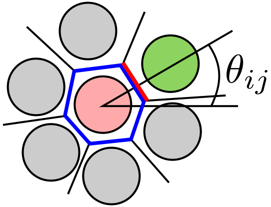
NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
*Los Alamos Scientific Laboratory, Los Alamos, New Mexico*

AND

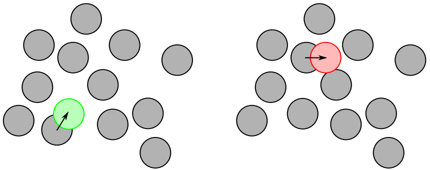
EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(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.

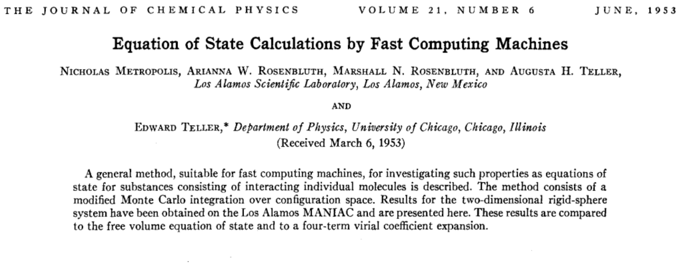



## Metropolis algorithm



## Diffusive dynamics

- Correlated sample: $\sigma^2(\bar{\Theta}) \propto \tau(\Theta)$
  $C_\Theta(t) = \frac{\langle \Theta(t'+t)\Theta(t')\rangle - \langle \Theta^2 \rangle}{\langle \Theta^2 \rangle - \langle \Theta \rangle^2}$
- Around 2nd order phase transition $\tau \propto \xi^z \propto L^z$
  $C_\Theta(t) \sim \exp(-t/\tau)$



THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

**Equation of State Calculations by Fast Computing Machines**

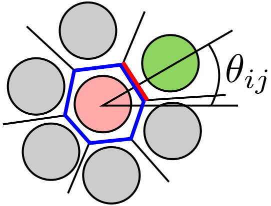
NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
*Los Alamos Scientific Laboratory, Los Alamos, New Mexico*

AND

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

(Received March 6, 1953)

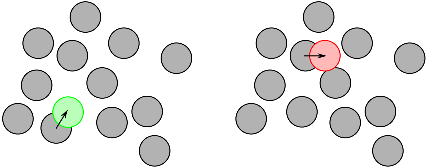
A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional 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.

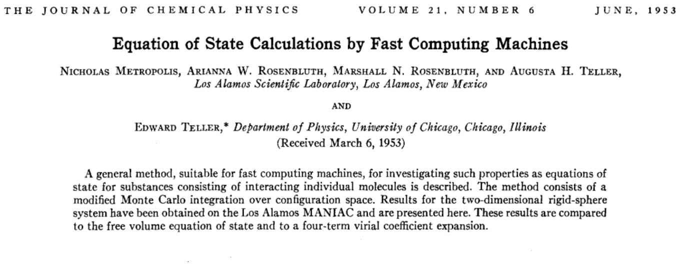



## How to produce collective moves?

- Continuous state space. No discrete symmetry as for spin lattices to easily build global $q$ (Cluster algorithms).
- With detailed balance in hard-core particle systems: symmetric proposal probabilities $q$ are necessary for the scheme to be rejection-free.

## Metropolis algorithm





THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

**Equation of State Calculations by Fast Computing Machines**

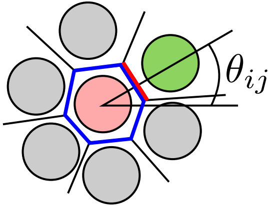
NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
*Los Alamos Scientific Laboratory, Los Alamos, New Mexico*

AND

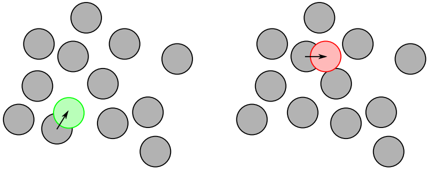
EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(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 algorithm



## How to produce collective moves?

- Continuous state space. No discrete symmetry as for spin lattices to easily build global $q$ (Cluster algorithms).
- With detailed balance in hard-core particle systems: symmetric proposal probabilities $q$ are necessary for the scheme to be rejection-free.
- Break DB: Non-reversibility?

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

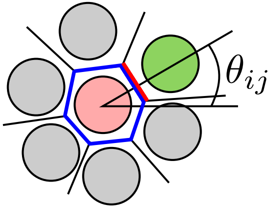
**Equation of State Calculations by Fast Computing Machines**

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller,
*Los Alamos Scientific Laboratory, Los Alamos, New Mexico*

and

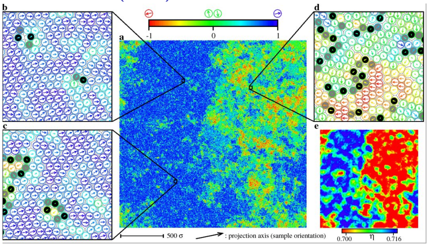
Edward Teller,* *Department of Physics, University of Chicago, Chicago, Illinois*
(Received March 6, 1953)

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.

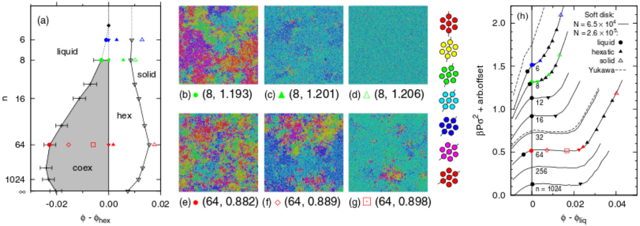



## Event-chain Monte Carlo

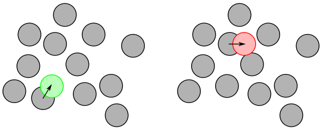
Bernard et al (2009)




Michel et al (2014), Kapfer et al (2015)

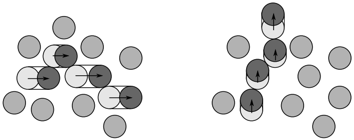


## Metropolis algorithm
(Metropolis et al. (1953))

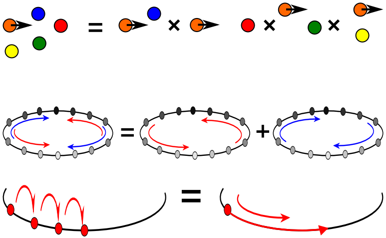


- Acceptance through Metropolis filter. $\min(1, \prod_i \exp(-\beta \Delta E_i)) = \exp(-\beta[\sum_i \Delta E_i]_+)$ **Rejections**
- Moves are:
  - Randomly proposed
  - Local
  - Finite
- Detailed balance

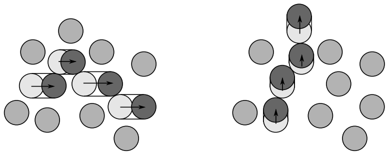
## Event-chain Monte Carlo
(Bernard et al (2009), Michel et al. (2014)))



- Direction change set by **factorized Metropolis** filter. $\prod_i \min(1, \exp(-\beta \Delta E_i)) = \exp(-\sum_i \beta[\Delta E_i]_+)$ **Rejection free**
- Moves are:
  - **Set** by additional variable
  - **Persistent** on global scale
  - **Infinitesimal**
- **Global** balance

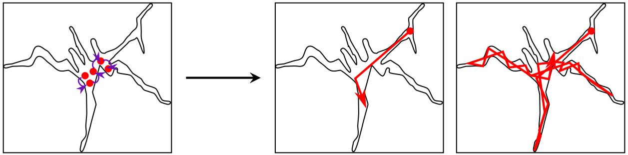


## Event-chain Monte Carlo
(Bernard et al (2009), Michel et al. (2014)))



- Direction change set by **factorized Metropolis** filter. $\prod_i \min(1, \exp(-\beta\Delta E_i)) = \exp(-\sum_i \beta[\Delta E_i]_+)$ **Rejection free**
- Moves are:
  - **Set** by additional variable
  - **Persistent** on global scale
  - **Infinitesimal**
- **Global** balance

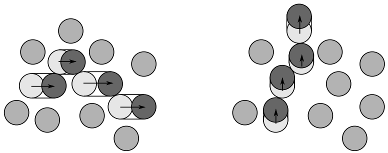
## ECMC for general potential?



| Reversibility | Non-reversibility |
| --- | --- |
| $\pi(\mathrm{d}x')K(x',\mathrm{d}x) = \pi(\mathrm{d}x)K(x,\mathrm{d}x')$ | $\int_{x'} \pi(\mathrm{d}x')K(x',\mathrm{d}x) = \int_{x'} \pi(\mathrm{d}x)K(x,\mathrm{d}x')$ |
| $\rightarrow$ Rejection, | $\rightarrow$ Direction change, |
| i.e. acceptance prob. $(a)$ | i.e. proposal prob. $(q)$ |

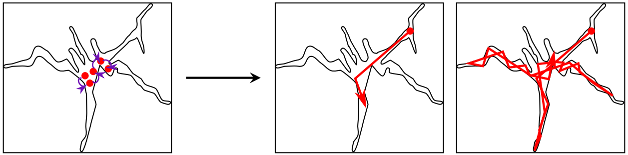
- How to upgrade to non-reversibility in general case? How to ensure global balance and ergodicity through only direction changes set by $q$?

## Event-chain Monte Carlo
(Bernard et al (2009), Michel et al. (2014)))



- Direction change set by **factorized Metropolis** filter. $\prod_i \min(1, \exp(-\beta\Delta E_i)) = \exp(-\sum_i \beta[\Delta E_i]_+)$ **Rejection free**
- Moves are:
  - **Set** by additional variable
  - **Persistent** on global scale
  - **Infinitesimal**
- **Global** balance

## ECMC for general potential?

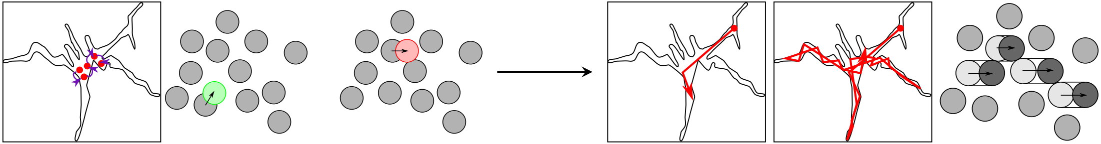


| Reversibility | Non-reversibility |
|---|---|
| $\pi(\mathrm{d}x')K(x', \mathrm{d}x) = \pi(\mathrm{d}x)K(x, \mathrm{d}x')$ → Rejection, i.e. acceptance prob. $(a)$ | $\int_{x'} \pi(\mathrm{d}x')K(x', \mathrm{d}x) = \int_{x'} \pi(\mathrm{d}x)K(x, \mathrm{d}x')$ → Direction change, i.e. proposal prob. $(q)$ |

- How to upgrade to non-reversibility in general case? How to ensure global balance and ergodicity through only direction changes set by $q$?
- Global symmetry hunt
- Piecewise deterministic Markov process

## General upgrading of the dynamics?

Sampling $x \sim \pi$ ($\propto \exp(-E(x))$, $E : \Omega \to \mathbb{R}$ the potential)
through Markov kernel $K(x, \mathrm{d}x') = q(x, x')a(x, x')\mathrm{d}x' + (1 - \int_\Omega q(x, y)a(x, y)\mathrm{d}y)\delta_{x=x'}$

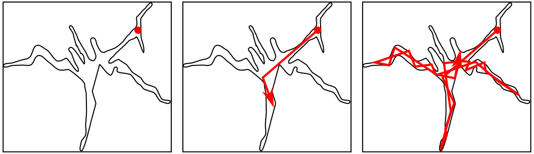


| Reversibility | Non-reversibility |
|---|---|
| $\pi(\mathrm{d}x')K(x', \mathrm{d}x) = \pi(\mathrm{d}x)K(x, \mathrm{d}x')$ | $\int_{x'} \pi(\mathrm{d}x')K(x', \mathrm{d}x) = \int_{x'} \pi(\mathrm{d}x)K(x, \mathrm{d}x')$ |
| $\to$ Rejection, i.e. acceptance prob. ($a$) | $\to$ Direction change, i.e. proposal prob. ($q$) |
| **Local symmetry constraint** | **Global symmetry?** |

State space extension $\Omega \to \Omega \times \mathcal{D}$ to set the proposal probabilities

$\pi(x) \to \tilde{\pi}(x, e) = \pi(x) \times \mu(e)$, $e \sim$ *direction* (**Careful!**)

## Piecewise Deterministic Markov Process



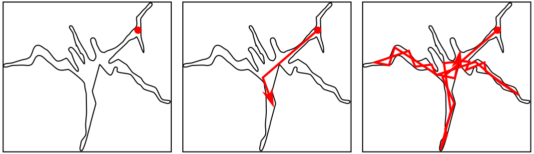
Goal: Global symmetry, no state space partition

No rejection, only direction changes.

No back-and-forth along a fixed trajectory.

No line partition

# Piecewise Deterministic Markov Process



## Goal: Global symmetry, no state space partition

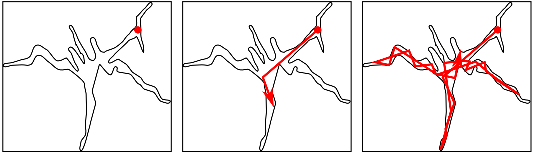
No rejection, only direction changes.

No back-and-forth along a fixed trajectory.

No line partition

$\rightarrow$ **Piecewise deterministic Markov process**

## Piecewise Deterministic Markov Process



Goal: Global symmetry, no state space partition

No rejection, only direction changes.

No back-and-forth along a fixed trajectory.

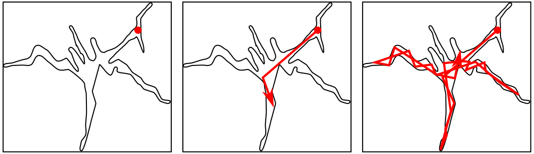
No line partition

$\rightarrow$ **Piecewise deterministic Markov process**

### PDMP characterizing elements (Davis (1993), in MCMC: Bouchard-Côté et al (2018), Bierkens et al (2019))

- Differential flow $(\phi_t)_{t\geq 0}$
- Jump rate $\lambda(x, e) + \bar{\lambda}$
- Markov kernel $Q$ (*repel kernel*)

# Piecewise Deterministic Markov Process



## Goal: Global symmetry, no state space partition

No rejection, only direction changes.
No back-and-forth along a fixed trajectory.
No line partition
$\rightarrow$ **Piecewise deterministic Markov process**

## PDMP characterizing elements (Davis (1993), in MCMC: Bouchard-Côté et al (2018), Bierkens et al (2019))

- Differential flow $(\phi_t)_{t\geq 0}$
- Jump rate $\lambda(x,e)+\bar{\lambda}$
- Markov kernel $Q$ (*repel kernel*)

## Infinitesimal generator $\mathcal{A}f = \lim_{t\to 0}\frac{P_t f - f}{t}$, $D_\phi f(x,e) = \lim_{t\to 0}\frac{f(\phi_t(x,e))-f(x,e)}{t}$

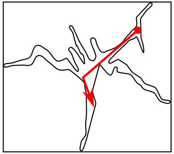
$$\mathcal{A}f = \underbrace{D_\phi f(x,e)}_{\text{Transport}} + \underbrace{\lambda(x,e)\int_{\mathcal{D}}(f(x,e')-f(x,e))Q((x,e),\mathrm{d}e')}_{\text{Events - Direction changes}} + \underbrace{\bar{\lambda}\int_{\mathcal{D}}(f(x,e')-f(x,e))\mu(\mathrm{d}e')}_{\text{Refreshment}}$$

## Invariance: Transport compensated by the direction changes

Infinitesimal generator $\mathcal{A}f = \lim_{t\to 0} \frac{P_t f - f}{t}$

$$\mathcal{A}f = \underbrace{D_\phi f(x,e)}_{\text{Transport}} + \underbrace{\lambda(x,e)\int_{\mathcal{D}}(f(x,e') - f(x,e))Q((x,e),\mathrm{d}e')}_{\text{Events - Direction changes}} + \underbrace{\bar\lambda\int_{\mathcal{D}}(f(x,e') - f(x,e))\mu(\mathrm{d}e')}_{\text{Refreshment}}$$

Conditions for $\tilde\pi = \pi \times \mu$ invariant: $\int_{\Omega\times\mathcal{D}} \mathcal{A}f\,\mathrm{d}\pi\mathrm{d}\mu = 0$

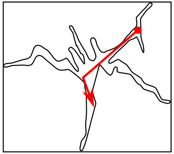


$$\int_{\Omega\times\mathcal{D}} D_\phi f(x,e)\pi(\mathrm{d}x)\mu(\mathrm{d}e)$$
$$= \int_{\Omega\times\mathcal{D}}\int_{\mathcal{D}} \lambda(x,e)(f(x,e') - f(x,e))Q((x,e),\mathrm{d}e')\pi(\mathrm{d}x)\mu(\mathrm{d}e)$$

## Invariance: Transport compensated by the direction changes

Infinitesimal generator $\mathcal{A}f = \lim_{t\to 0} \frac{P_t f - f}{t}$

$$\mathcal{A}f = \underbrace{D_\phi f(x,e)}_{\text{Transport}} + \underbrace{\lambda(x,e)\int_{\mathcal{D}}(f(x,e') - f(x,e))Q((x,e),\mathrm{d}e')}_{\text{Events - Direction changes}} + \underbrace{\bar{\lambda}\int_{\mathcal{D}}(f(x,e') - f(x,e))\mu(\mathrm{d}e')}_{\text{Refreshment}}$$

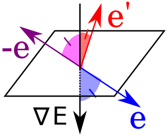


Conditions for $\tilde{\pi} = \pi \times \mu$ invariant: $\int_{\Omega\times\mathcal{D}} \mathcal{A}f \,\mathrm{d}\pi\mathrm{d}\mu = 0$

$$\int_{\Omega\times\mathcal{D}} D_\phi f(x,e)\pi(\mathrm{d}x)\mu(\mathrm{d}e)$$
$$= \int_{\Omega\times\mathcal{D}}\int_{\mathcal{D}} \lambda(x,e)(f(x,e') - f(x,e))Q((x,e),\mathrm{d}e')\pi(\mathrm{d}x)\mu(\mathrm{d}e)$$

With a flow along $e$, by integration by part, $(\pi(x) \propto \exp(-E(x)))$

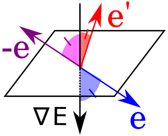
$$\underbrace{\int_{\mathcal{D}} \langle \boldsymbol{\nabla} \mathbf{E}(\mathbf{x}), -\mathbf{e}\rangle_+ f(x,e)\boldsymbol{\mu}(\mathbf{de})}_{\text{brought by transport}} = \underbrace{\int_{\mathcal{D}}\int_{\mathcal{D}} \langle \boldsymbol{\nabla} \mathbf{E}(\mathbf{x}), \mathbf{e}\rangle_+ f(x,e')\mathbf{Q}(\mathbf{e}\to\mathbf{e}')\boldsymbol{\mu}(\mathbf{de})}_{\text{redistributed by direction change}}$$



# Event-chain Monte Carlo/PDMP-sampling in a few words

With a flow along $e$, by integration by part, $(\pi(x) \propto \exp(-E(x)))$

$$\underbrace{\int_{\mathcal{D}} \langle \nabla \mathbf{E}(\mathbf{x}), -\mathbf{e}\rangle_+ f(x,e)\boldsymbol{\mu}(\mathbf{de})}_{\text{brought by transport}} = \underbrace{\int_{\mathcal{D}}\int_{\mathcal{D}} \langle \nabla \mathbf{E}(\mathbf{x}), \mathbf{e}\rangle_+ f(x,e')\mathbf{Q}(\mathbf{e}\to\mathbf{e}')\boldsymbol{\mu}(\mathbf{de})}_{\text{redistributed by direction change}}$$




## Main idea

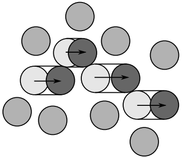
Find some symmetries on the way the energy change in order to get some balance

$$\sum_{\Delta} \langle \nabla_\Delta E, e\rangle = 0 \to \sum_{\substack{\Delta \\ \langle \nabla_\Delta E, e\rangle > 0}} \langle \nabla_\Delta E, e\rangle = \sum_{\substack{\Delta \\ \langle \nabla_\Delta E, e\rangle < 0}} -\langle \nabla_\Delta E, e\rangle$$

$$\to \sum_{\boldsymbol{\Delta}} \langle \nabla_{\boldsymbol{\Delta}} \mathbf{E}, \mathbf{e}\rangle_+ = \sum_{\boldsymbol{\Delta}} \langle \nabla_{\boldsymbol{\Delta}} \mathbf{E}, -\mathbf{e}\rangle_+$$

## Chasing down symmetries

### Pairwise interactions



Exploitation of mirror symmetry through factorization $\boldsymbol{\nabla}_{x_i} E_{ij}(x) = -\boldsymbol{\nabla}_{x_j} E_{ij}(x)$

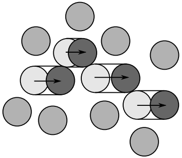
(i.e. $\mathrm{div} E_{ij} = 0$)

**Deterministic** kernel $Q$

Michel et al (2014)

## Chasing down symmetries

### Pairwise interactions



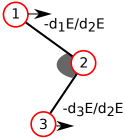
Exploitation of mirror symmetry through factorization $\boldsymbol{\nabla}_{x_i} E_{ij}(x) = -\boldsymbol{\nabla}_{x_j} E_{ij}(x)$

(i.e. $\mathrm{div} E_{ij} = 0$)

**Deterministic** kernel $Q$

Michel et al (2014)

### $n$-body interactions




Exploitation of translational invariance div $\mathbf{E} = \mathbf{0}$

$\rightarrow \sum_{i_k} \langle \nabla x_{i_k} E_{i_1 \cdots _n}, v\rangle = 0$

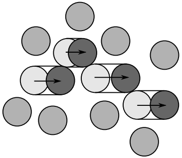
$$\rightarrow \sum_{i_k} \langle \nabla x_{i_k} E_{i_1 \dots i_n}, v\rangle_+ = \sum_{i_k} \langle \nabla x_{i_k} E_{i_1 \dots i_n}, -v\rangle_+$$

**Non-deterministic** kernel $Q$

Harland et al (2017)

# Chasing down symmetries

## Pairwise interactions



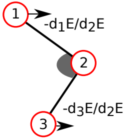
Exploitation of mirror symmetry through factorization
$\boldsymbol{\nabla}_{x_i} E_{ij}(x) = -\boldsymbol{\nabla}_{x_j} E_{ij}(x)$

(i.e. $\mathrm{div} E_{ij} = 0$)

**Deterministic** kernel $Q$

Michel et al (2014)

## $n$-body interactions




Exploitation of translational invariance div $\mathbf{E} = \mathbf{0}$

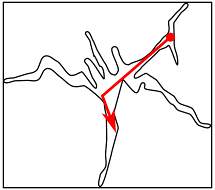
$\to \sum_{i_k} \langle \nabla x_{i_k} E_{i_1 \cdots_n}, v \rangle = 0$

$$\to \sum_{i_k} \langle \nabla x_{i_k} E_{i_1 \dots i_n}, v \rangle_+ = \sum_{i_k} \langle \nabla x_{i_k} E_{i_1 \dots i_n}, -v \rangle_+$$

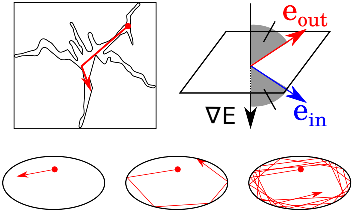
**Non-deterministic** kernel $Q$

Harland et al (2017)

## In the general case?



## General case: Exploiting rotational invariance




### Deterministic kernel $Q$

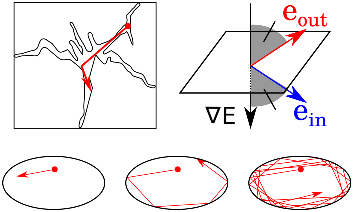
No a priori symmetry, but if reflection or flip:

$\nabla E \cdot e_{in} = -\nabla E \cdot e_{out}$

$Q(e_{\text{in}} \to e_{\text{out}}) = \delta(e_{\text{out}} - R_{\boldsymbol{\nabla} E(x)}(e_{\text{int}}))$

(Peters et al (2012), Michel et al (2014), Bouchard-Côté et al (2018), Bierkens et al (2019))

# General case: Exploiting rotational invariance



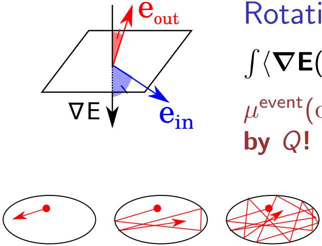

## Deterministic kernel $Q$

No a priori symmetry, but if reflection or flip:

$\nabla E \cdot e_{in} = -\nabla E \cdot e_{out}$

$Q(e_{\text{in}} \to e_{\text{out}}) = \delta(e_{\text{out}} - R_{\boldsymbol{\nabla} E(x)}(e_{\text{int}}))$

(Peters et al (2012), Michel et al (2014), Bouchard-Côté et al (2018), Bierkens et al (2019))




## Rotational invariance around $\boldsymbol{\nabla} E$:

$\int \langle \boldsymbol{\nabla} \mathbf{E}(\mathbf{x}), \mathbf{e} \rangle \boldsymbol{\mu}(\mathrm{d}\mathbf{e}) = \mathbf{0} \to \int \langle \boldsymbol{\nabla} \mathbf{E}(\mathbf{x}), \mathbf{e} \rangle_+ \boldsymbol{\mu}(\mathrm{d}\mathbf{e}) = \int \langle \boldsymbol{\nabla} \mathbf{E}(\mathbf{x}), -\mathbf{e} \rangle_+ \boldsymbol{\mu}(\mathrm{d}\mathbf{e})$

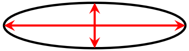
$\mu^{\text{event}}(\mathrm{d}e) = \langle \boldsymbol{\nabla} E(x), -e \rangle_+ \mu(\mathrm{d}e) / \int \langle \boldsymbol{\nabla} E(x), -e \rangle_+ \mu(\mathrm{d}e)$ **should be conserved by $Q$!**

- Independent pick of new directions $Q(e_{\text{in}} \to e_{\text{out}}) \propto \langle \boldsymbol{\nabla} E(x), -e_{\text{out}} \rangle_+$
- Non-reversible in $E$

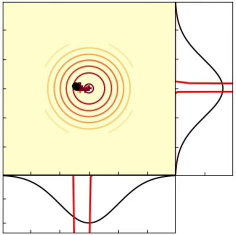
(Michel et al (2020))

# Illustration - Anisotropic Gaussian

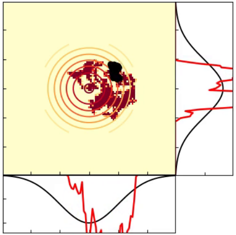
Gaussian distribution $E = \sum_i x_i^2/(2\sigma_i^2)$, $\sigma_i \in [1, 1000]$ - 400 **dimensions**
(section of the dimensions with the largest variances)



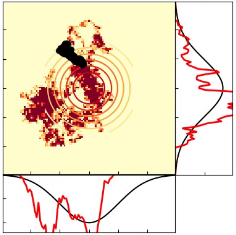
Metropolis
Reversibility and Local symmetry



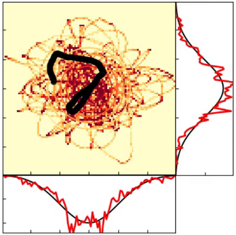
HMC
Reversibility and kinetic energy



Reflexion
Irreversibility and Local Symmetry



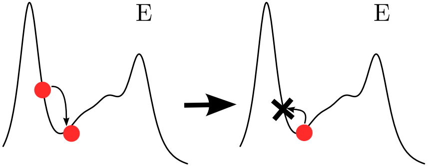
Direct pick
Irreversibility and Global Symmetry



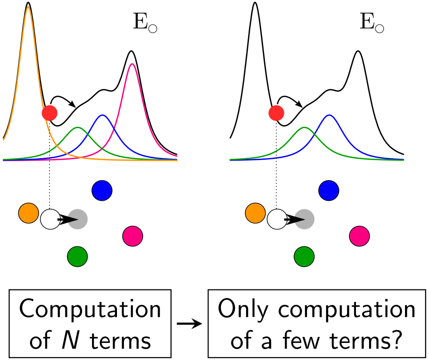
# Outline

# What about Complexity?

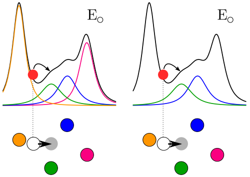
Metropolis algorithm:
$p(i \to j) = \min(1, \exp(-\beta \Delta E))$




N interaction terms
$\Delta E = \sum_{i=1}^{N} \Delta E_i$



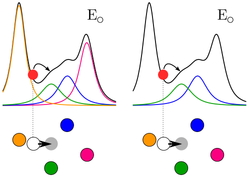
# Complexity reduction for irreversible MC algorithms




## Factorized transitions: superposition of Poisson process (PP)

Direction changes ruled by a Poisson process of rate $\lambda = \sum_{i=1}^{N} \lambda_i$, $\lambda_i = \max(0, dE_i)$.

# Complexity reduction for irreversible MC algorithms



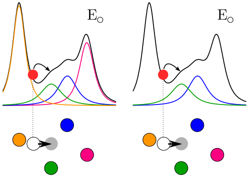

## Factorized transitions: superposition of Poisson process (PP)

Direction changes ruled by a Poisson process of rate $\lambda = \sum_{i=1}^{N} \lambda_i$,
$\lambda_i = \max(0, dE_i)$.

## Complexity reduction by *thinning* (Lewis and Schedler (1979))

Consider the bound $\lambda^{\mathrm{Bound}} \geq \lambda$
Superposition of PP: $\lambda^{\mathrm{Bound}} = \lambda + \lambda^{\mathrm{Fake}}$

# Complexity reduction for irreversible MC algorithms

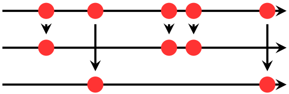


## Factorized transitions: superposition of Poisson process (PP)

Direction changes ruled by a Poisson process of rate $\lambda = \sum_{i=1}^{N} \lambda_i$,
$\lambda_i = \max(0, dE_i)$.

## Complexity reduction by *thinning* (Lewis and Schedler (1979))

Consider the bound $\lambda^{\mathrm{Bound}} \geq \lambda$
Superposition of PP: $\lambda^{\mathrm{Bound}} = \lambda + \lambda^{\mathrm{Fake}}$



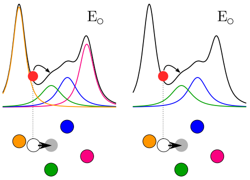
Bound, $\lambda^{\mathrm{Bound}}$
Fake, $\lambda^{\mathrm{Fake}}$
True, $\lambda$

True event with probability $\lambda/\lambda^{\mathrm{Bound}}$

# Complexity reduction for irreversible MC algorithms

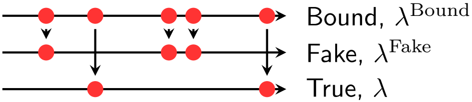
## Factorized transitions: superposition of Poisson process (PP)

Direction changes ruled by a Poisson process of rate $\lambda = \sum_{i=1}^{N} \lambda_i$,
$\lambda_i = \max(0, dE_i)$.



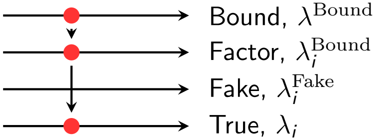

## Complexity reduction by *thinning* (Lewis and Schedler (1979))

Consider the bound $\lambda^{\mathrm{Bound}} \geq \lambda$
Superposition of PP: $\lambda^{\mathrm{Bound}} = \lambda + \lambda^{\mathrm{Fake}}$




True event with probability $\lambda / \lambda^{\mathrm{Bound}}$

Writing now $\lambda^{\mathrm{Bound}} = \sum_i \lambda_i^{\mathrm{Bound}}, \lambda_i^{\mathrm{Bound}} \geq \lambda_i, \forall i$
And $\lambda_i^{\mathrm{Bound}} = \lambda_i + \lambda_i^{\mathrm{Fake}}$

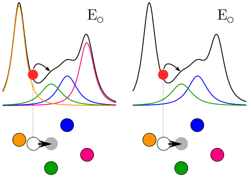



Pick a potentially rejecting term $i$ with $\lambda_i^{\mathrm{Bound}} / \lambda^{\mathrm{Bound}}$ and resample a true event with $\lambda_i / \lambda_i^{\mathrm{Bound}}$.

# Complexity reduction for irreversible MC algorithms

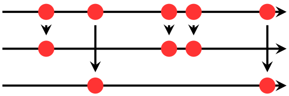
## Factorized transitions: superposition of Poisson process (PP)

Direction changes ruled by a Poisson process of rate $\lambda = \sum_{i=1}^{N} \lambda_i$, $\lambda_i = \max(0, dE_i)$.




## Complexity reduction by *thinning* (Lewis and Schedler (1979))

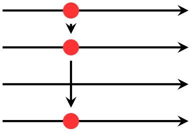
Consider the bound $\lambda^{\mathrm{Bound}} \geq \lambda$
Superposition of PP: $\lambda^{\mathrm{Bound}} = \lambda + \lambda^{\mathrm{Fake}}$



Bound, $\lambda^{\mathrm{Bound}}$
Fake, $\lambda^{\mathrm{Fake}}$
True, $\lambda$

True event with probability $\lambda/\lambda^{\mathrm{Bound}}$
Logistic regression: Bouchard-Côté et al (2018); Soft spheres: Kapfer et al (2016)

Writing now $\lambda^{\mathrm{Bound}} = \sum_i \lambda_i^{\mathrm{Bound}}, \lambda_i^{\mathrm{Bound}} \geq \lambda_i, \forall i$
And $\lambda_i^{\mathrm{Bound}} = \lambda_i + \lambda_i^{\mathrm{Fake}}$



Bound, $\lambda^{\mathrm{Bound}}$
Factor, $\lambda_i^{\mathrm{Bound}}$
Fake, $\lambda_i^{\mathrm{Fake}}$
True, $\lambda_i$

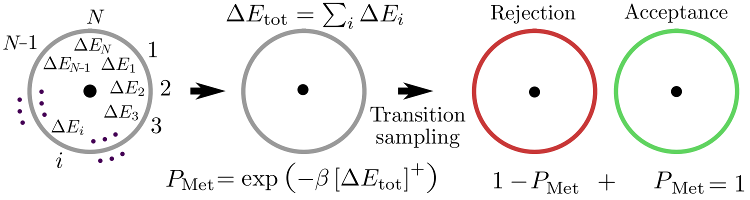
Pick a potentially rejecting term $i$ with $\lambda_i^{\mathrm{Bound}}/\lambda^{\mathrm{Bound}}$ and resample a true event with $\lambda_i/\lambda_i^{\mathrm{Bound}}$.

## Clock Monte Carlo method (Michel et al (2019))

Metropolis filter

$P_{\mathrm{rej}} = 1 - P_{\mathrm{Met}}$

One-step Bernoulli process

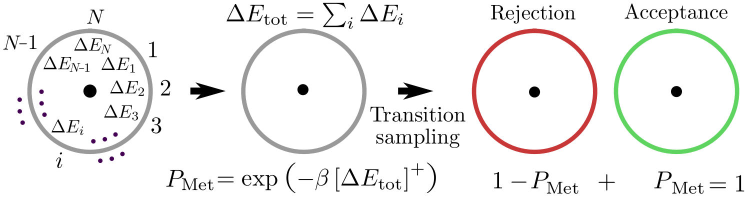


## Clock Monte Carlo method (Michel et al (2019))

### Metropolis filter

$P_{\mathrm{rej}} = 1 - P_{\mathrm{Met}}$

### One-step Bernoulli process




### Factorized filter and its consensus rule

sampling rejection $\leftrightarrow$ sampling first factor rejecting

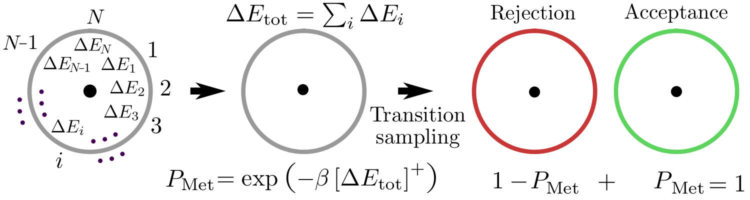
$P_{\mathrm{rej}} = 1 - P_{\mathrm{fac}} = \sum_i (1 - p_i) \prod_{j<i} p_j$

## Clock Monte Carlo method (Michel et al (2019))

### Metropolis filter

$P_{\text{rej}} = 1 - P_{\text{Met}}$

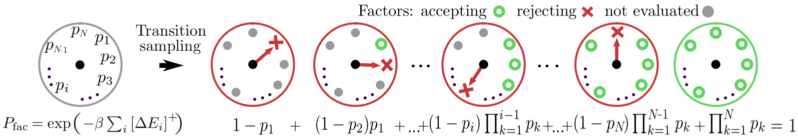
### One-step Bernoulli process




### Factorized filter and its consensus rule

sampling rejection $\leftrightarrow$ sampling first factor rejecting

$P_{\text{rej}} = 1 - P_{\text{fac}} = \sum_i (1 - p_i) \prod_{j<i} p_j$

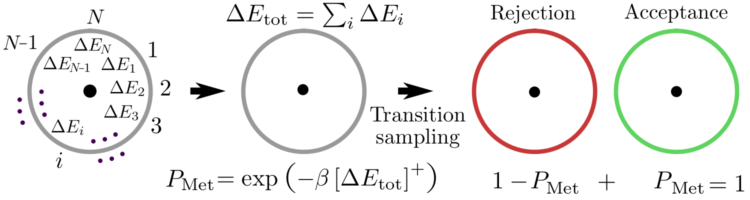


# Clock Monte Carlo method (Michel et al (2019))

## Metropolis filter

$P_{\text{rej}} = 1 - P_{\text{Met}}$

## One-step Bernoulli process

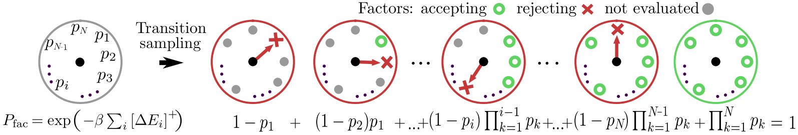



## Factorized filter and its consensus rule

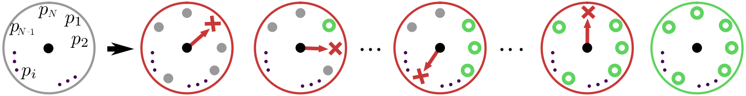
sampling rejection ↔ sampling first factor rejecting

$P_{\text{rej}} = 1 - P_{\text{fac}} = \sum_i (1 - p_i) \prod_{j<i} p_j$

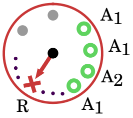
N-step Bernoulli process!



# Clock Monte Carlo method



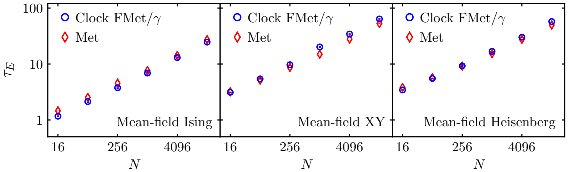

## Complexity reduction




- Consider a bound Bernoulli process $p_B = \prod p_{B,i}$, with $\forall i, p_i \geq p_{B,i}$
- $P_{\mathrm{rej}}(i) = p_i^R \prod_{j<i} (p_j^{A_1} + p_j^{A_2})$
- Sampling a clock is replaced by the sampling of a random path of successive events $(A_1)$ or $(A_2)$ until a true rejection $(R)$ is sampled or the path is of length $N$.
  - Given configuration-independent bounds, successive bound rejections sampled in O(1).
  - Complexity $\mathcal{C}$ = number of attempted bound rejections $\sim \mathrm{O}(\ln p_B / \ln p_{Fac})$, $\sim 1$ if $p_B$ and $p_{\mathrm{Fac}}$ scale with N similarly.
  - Long-range cluster algorithms (Luijten et al (1995), Fukui et al (2008)): effectively uses a factorized filter.

# Performance analysis

Integrated autocorrelation times

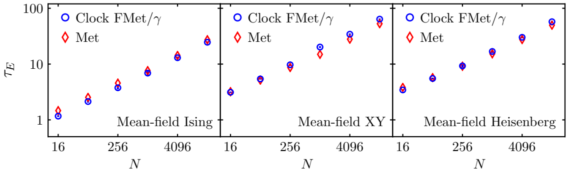



## Overall acceleration $\mathcal{A} \sim \mathrm{O}(N/\mathcal{C}\gamma)$

- A complexity speedup $\mathrm{O}(N/\mathcal{C})$;
- But a smaller acceptance rate slowing-down, $\gamma = p_{\mathrm{Metro}}/p_{\mathrm{Fact}}$

## Performance analysis

Integrated autocorrelation times




Overall acceleration $\mathcal{A} \sim \mathrm{O}(N/\mathcal{C}\gamma)$

- A complexity speedup $\mathrm{O}(N/\mathcal{C})$;
- But a smaller acceptance rate slowing-down, $\gamma = p_{\mathrm{Metro}}/p_{\mathrm{Fact}}$

The energy extensivity nature directly controls the performance!

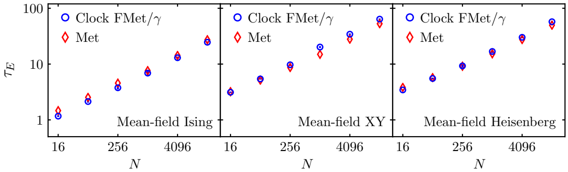
- $\ln\gamma \propto \sum_i |\Delta E_i| - |\sum_i \Delta E_i|$
- $\mathcal{C} \sim \ln p_B / \ln p_{\mathrm{Fac}} \sim \sum_i \max |\Delta E_i| / \sum_i |\Delta E_i|$

## Performance analysis

Overall acceleration $\mathcal{A} \sim \mathrm{O}(N/\mathcal{C}\gamma)$

- A complexity speedup $\mathrm{O}(N/\mathcal{C})$;
- But a smaller acceptance rate slowing-down, $\gamma = p_{\mathrm{Metro}}/p_{\mathrm{Fact}}$

Integrated autocorrelation times




The energy extensivity nature directly controls the performance!

- $\ln \gamma \propto \sum_i |\Delta E_i| - |\sum_i \Delta E_i|$
- $\mathcal{C} \sim \ln p_B / \ln p_{\mathrm{Fac}} \sim \sum_i \max |\Delta E_i| / \sum_i |\Delta E_i|$

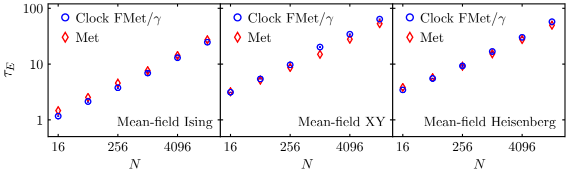
- Strict extensivity:
  - $\sum_i \max |\Delta E_i| \sim \mathrm{O}(1)$
  - $\mathcal{A} \sim \mathrm{O}(N)$

## Performance analysis

Overall acceleration $\mathcal{A} \sim \mathrm{O}(N/\mathcal{C}\gamma)$

- A complexity speedup $\mathrm{O}(N/\mathcal{C})$;
- But a smaller acceptance rate slowing-down, $\gamma = p_{\mathrm{Metro}}/p_{\mathrm{Fact}}$

Integrated autocorrelation times



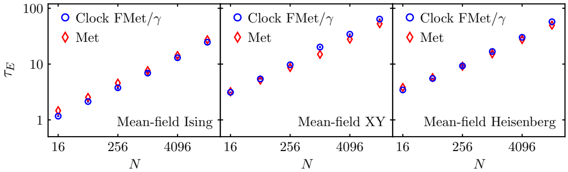

The energy extensivity nature directly controls the performance!

- $\ln \gamma \propto \sum_i |\Delta E_i| - |\sum_i \Delta E_i|$
- $\mathcal{C} \sim \ln p_B / \ln p_{\mathrm{Fac}} \sim \sum_i \max |\Delta E_i| / \sum_i |\Delta E_i|$

- Strict extensivity:
  - $\sum_i \max |\Delta E_i| \sim \mathrm{O}(1)$
  - $\mathcal{A} \sim \mathrm{O}(N)$
- Sub-extensivity:
  - $\sum_i \max |\Delta E_i| \sim \mathrm{O}(N^\alpha)$
  - $\mathcal{A} \sim \mathrm{O}(N^\kappa), 0 \leq \kappa < 1$
  - Box $\sim N/N^\omega$

## Performance analysis

Integrated autocorrelation times




### Overall acceleration $\mathcal{A} \sim \mathrm{O}(N/\mathcal{C}\gamma)$

- A complexity speedup $\mathrm{O}(N/\mathcal{C})$;
- But a smaller acceptance rate slowing-down, $\gamma = p_{\mathrm{Metro}}/p_{\mathrm{Fact}}$

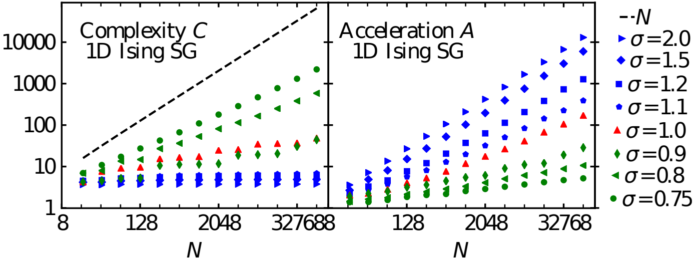
### The energy extensivity nature directly controls the performance!

- $\ln \gamma \propto \sum_i |\Delta E_i| - |\sum_i \Delta E_i|$
- $\mathcal{C} \sim \ln p_B / \ln p_{\mathrm{Fac}} \sim \sum_i \max |\Delta E_i| / \sum_i |\Delta E_i|$

- Strict extensivity:
  - $\sum_i \max |\Delta E_i| \sim \mathrm{O}(1)$
  - $\mathcal{A} \sim \mathrm{O}(N)$
- Sub-extensivity:
  - $\sum_i \max |\Delta E_i| \sim \mathrm{O}(N^\alpha)$
  - $\mathcal{A} \sim \mathrm{O}(N^\kappa), 0 \leq \kappa < 1$
  - Box $\sim N/N^\omega$
- Marginal extensivity:
  - $\sum_i \max |\Delta E_i| \sim \mathrm{O}(\ln N)$
  - $\mathrm{O}(N/(\ln N)^2) \leq \mathcal{A} \leq \mathrm{O}(N/\ln N)$
  - Box $\sim \ln N$ can be necessary

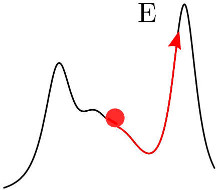
## Applications: 1D long-range Ising spin glass

- $\mathcal{H} = -c(N)\sum_{i<j} \frac{s_{ij}}{r_{ij}^{\sigma}} S_i S_j$
- $s_{ij} = \pm 1$
- $c(N)^{-2} = \sum_{j>1}\langle J_{1j}^2 \rangle$
- $\beta = 1$




- $\sigma > 1$: Strict extensivity:
  - $\sum_i \max|\Delta E_i| \sim \mathrm{O}(1)$
  - Box $= 2$
  - $\mathcal{A} \sim \mathrm{O}(N)$
- $\sigma < 1$: Sub-extensivity:
  - $\sum_i \max|\Delta E_i| \sim \mathrm{O}(N^{1-\sigma})$
  - Box $\sim N^{2(1-\sigma)}$
  - $\mathcal{A} \sim \mathrm{O}(N^{\kappa})$
- $\sigma = 1$: Marginal extensivity:
  - $\sum_i \max|\Delta E_i| \sim \mathrm{O}(\ln N)$
  - Box $= \ln N$
  - $\mathcal{A} \sim \mathrm{O}(N/(\ln N)^2)$

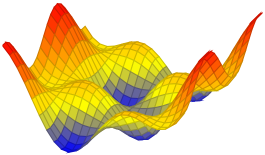
# Outline




## How to propose moves which can overpass high barriers?

- Parallel tempering
- Population Monte Carlo
- Overrelaxation method
- Umbrella sampling
- Adaptive Monte Carlo
- and much more

## Generative models: Normalizing flows



- Learn an invertible mapping $x \sim \pi \leftrightarrow z \sim \nu$ (typically $\nu$ Gaussian)

| | | | | |
|---|---|---|---|---|
| high-dimensional integral<br>$\langle\theta\rangle = \int_\Omega \pi(\mathrm{d}x)\theta(x)$<br>$\pi(\mathrm{d}x) \propto \exp(-\beta E(x))\mathrm{d}x$ | $\Longleftrightarrow$ | Average over random $x_i$<br>$\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N}\theta(x_i)$ | $\Longleftrightarrow$ | Generate $x \sim \pi(x)$<br>$\mathrm{rand}(0,1) \rightarrow \pi$ |

## Normalizing flows (Rezende et al (2015), Kingma et al (2016))

### Invertible mapping $f$

- $f(x, \theta) = z$
- $P_x(z) = \pi(f^{-1}(z))|\det J_{f^{-1}}|$
- $P_z(x) = \nu(f(x))|\det J_f|$
- Typically Kullback-Leibler divergence

Find invertible transformation with computable Jacobian?

# Key points

## Boltzmann generators

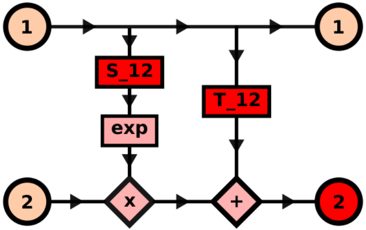
- Real NVP architecture (invertibility, computable Jacobian) (Dinh et al (2017))
- Use of Boltzmann distribution (Noé et al (2019))
  - $\pi$ is known, used during training
  - Samples obtained from latent space ($z \to x$) can be unbiased by importance sampling

# Invertible block

- Real NVP architecture (invertibility, computable Jacobian) (Dinh et al (2017))

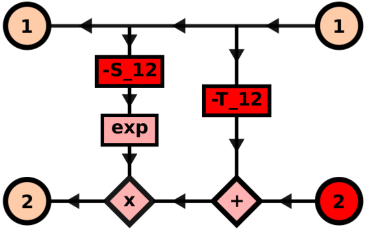
Coordinates divided into two parts, $S_{12}$ and $T_{12}$ are (non-invertible) networks.

**Forward**




$$\begin{cases} \vec{y}_1 = \vec{x}_1 \\ \vec{y}_2 = \vec{x}_2 \odot \exp(S_{12}(\vec{x}_1;\vec{\theta})) + T_{12}(\vec{x}_1;\vec{\theta}) \end{cases}$$

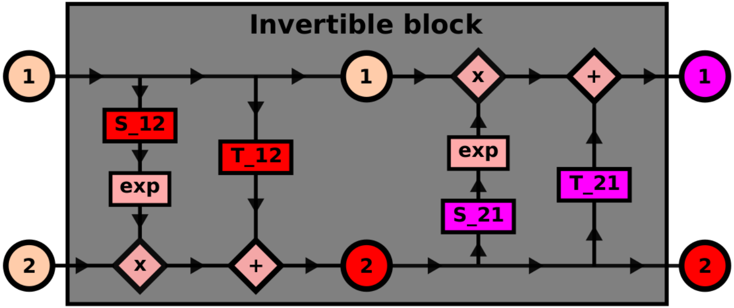
**Inverse**




$$\begin{cases} \vec{x}_1 = \vec{y}_1 \\ \vec{x}_2 = \left(y_2 - T_{12}(\vec{y}_1;\vec{\theta})\right) \odot \exp(-S_{12}(\vec{y}_1;\vec{\theta})) \end{cases}$$

## Invertible block

- Real NVP architecture (invertibility, computable Jacobian) (Dinh et al (2017))



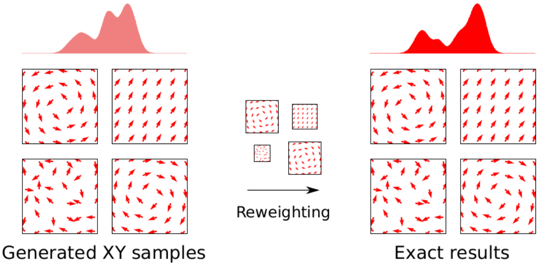

(Plot from T. Guyon)

## Importance sampling

- Use of Boltzmann distribution (Noé et al (2019))

Importance sampling :

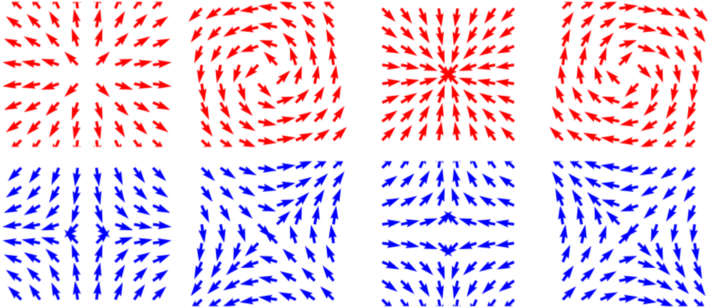
$$\mathbb{E}_{\vec{x}\sim\mu_X}\left[\mathcal{O}(\vec{x})\right] = \mathbb{E}_{\vec{x}\sim q_X}\left[\frac{\mu_X(\vec{x})}{q_X(\vec{x})}\mathcal{O}(\vec{x})\right]$$




Non-normalized weights :

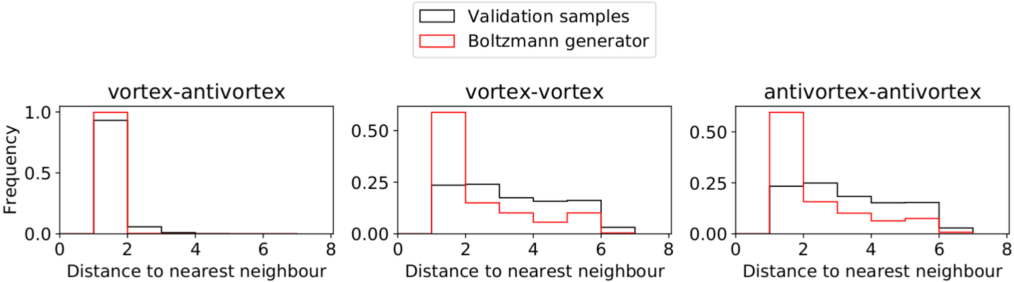
$$\mathbb{E}_{\vec{x}\sim\mu_X}\left[\mathcal{O}(\vec{x})\right] \underset{N_{samples}\to\infty}{\to} \frac{\sum_{samples} w(\vec{x})\mathcal{O}(\vec{x})}{\sum_{samples} w(\vec{x})} \quad \text{with} \quad w(\vec{x}) = \exp\left(-\frac{E(\vec{x})}{T} + \frac{1}{2}\|F_{xz}(\vec{x})\|^2 - \log R_{xz}(\vec{x})\right)$$

## Applications to 2d XY spins (Ongoing work with T. Guyon, A. Guillin)



(Plot from T. Guyon)

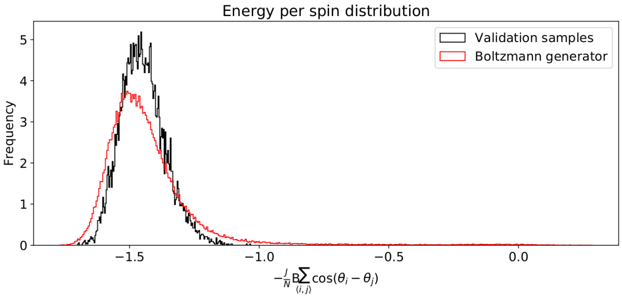
## Applications to 2d XY spins (Ongoing work with T. Guyon, A. Guillin)

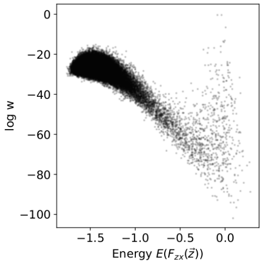



The network has learned the attraction behaviour between vortices and anti-vortices.

The repulsion is underestimated.

## Applications to 2d XY spins (Ongoing work with T. Guyon, A. Guillin)




### Challenges

- Fine tuning to avoid divergence at the importance sampling step
- Could not discuss: Angles? Mode collapse?

# Conclusion

- Upgrading dynamics by non-reversibility obtained by exploiting global symmetries (discrete or continuous).
- Trade-off between efficient exploration and ergodicity? Quantitative theoretical analysis? General implementation?
- Complexity reduction in standard MCMC scheme by factorizing interaction terms.
- Dealing with energy extensivity and strong frustration? Limit for computational complexity reduction?
- Non-local moves by normalizing flows
- Ergodic training set? Fine tuning? Mode collapse? Hard-core potentials?

Particles: S. Kapfer, W. Krauth

PDMP: A. Monemvassitis, A. Guillin

Polymer: T. A. Kampmann, J. Kierfeld

Bayesian inference: A. Durmus

Complexity: Y. Deng, X. Tan

Normalizing flows: T. Guyon, V. Souveton, A. Guillin, G. Lavaux, J. Jasche

# Conclusion

- Upgrading dynamics by non-reversibility obtained by exploiting global symmetries (discrete or continuous).
- Trade-off between efficient exploration and ergodicity? Quantitative theoretical analysis? General implementation?
- Complexity reduction in standard MCMC scheme by factorizing interaction terms.
- Dealing with energy extensivity and strong frustration? Limit for computational complexity reduction?
- Non-local moves by normalizing flows
- Ergodic training set? Fine tuning? Mode collapse? Hard-core potentials?

Particles: S. Kapfer, W. Krauth

PDMP: A. Monemvassitis, A. Guillin

Polymer: T. A. Kampmann, J. Kierfeld

Bayesian inference: A. Durmus

Complexity: Y. Deng, X. Tan

Normalizing flows: T. Guyon, V. Souveton, A. Guillin, G. Lavaux, J. Jasche

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR SuSa

**Thank you for your attention!**

## (Not complete) References


- DB MCMC
  - N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
  - A. Jaster, *Physica A* **264**, 134 (1999).
- Discrete lifted MCMC/HMC
  - P. Diaconis, S. Holmes, and R. M. Neal, *Ann. Appl. Probab.* **10**, 726 (2000)
  - S. Duane, A.D. Kennedy, B. J. Pendleton and D. Roweth, *Physics letters B*, **195**(2), 216-222 (1987).
  - R. M. Neal. *MCMC using Hamiltonian dynamics*. Handbook of Markov Chain Monte Carlo, 2(11), 2011.
  - S.Turitsyn, M. Chertkov and M. Vucelja, *Physica D* **240**, 410 (2011)
- ECMC/PDMP
  - E. P. Bernard, W. Krauth, D. B. Wilson *Phys. Rev. E* **80** 056704 (2009)
  - E. A. J. F. Peters and G. de With, *Phys. Rev. E*, **85**:026703 (2012)
  - M. Michel, S. C. Kapfer and W. Krauth, *J. Chem. Phys.*, **140**, 054116 (2014)
  - J. Harland, M. Michel, T. A. Kampmann and J. Kierfeld, *Phys. Rev. E*, **117** (3), 30001 (2017).
  - M. Michel, A. Durmus and S. Sénécal, *JCGS*, **29**(4): 689–702 (2020)
  - A. Monemvassitis, A. Guillin, M. Michel, *prepint* (2022)
  - M. Davis, *Markov Models & Optimization*, Volume 49. CRC Press. (1993)
  - A. Bouchard-Côté, S. Vollmer and A. Doucet.*JASA*, 113(522): 855-867 (2018)
  - J. Bierkens, P. Fearnhead, G. Roberts. *Ann. Statist.* **47**(3): 1288-1320 (2019)
- Complexity reduction
  - E. Luijten and H. W. J. Bloete, Int. J. Mod. Phys. C 06, 359 (1995)
  - S. Kapfer and W. Krauth, *Phys. Rev. E*, **94**, 031302(R) (2016)
  - M. Michel, X. Tan and Y. Deng, *Phys. Rev. E*, **99**, 010105, (2019)
- Normalizing flows
  - D. Jimenez Rezende and S. Mohamed. *preprint* arXiv:1505.05770, 2015.
  - D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling. *NeurIPS*, (2016).
  - L. Dinh, J. Sohl-Dickstein, and S. Bengio. *ICLR*, 2(017).
  - F. Noé, S. Olsson, J. Khler, and H. Wu. *Science*, **365**(6457), (2019).