

Multilevel Monte Carlo for path integrals in quantum mechanics

Karl Jansen[†], **Eike Hermann Müller**[★], Robert Scheichl[‡]

[†]DESY Zeuthen/Berlin, [★]University of Bath, [‡]Heidelberg University

Numerical Challenges in Lattice QCD, Mon 15th Aug 2022

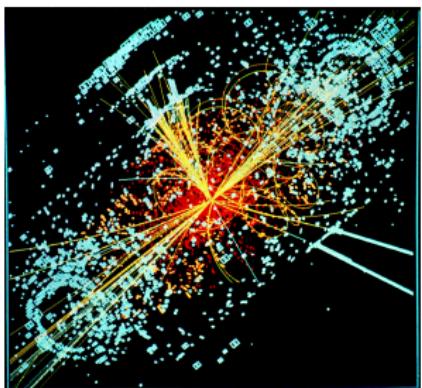
Phys. Rev. D 102, 114512 (2020) [arxiv:2008.03090](https://arxiv.org/abs/2008.03090)



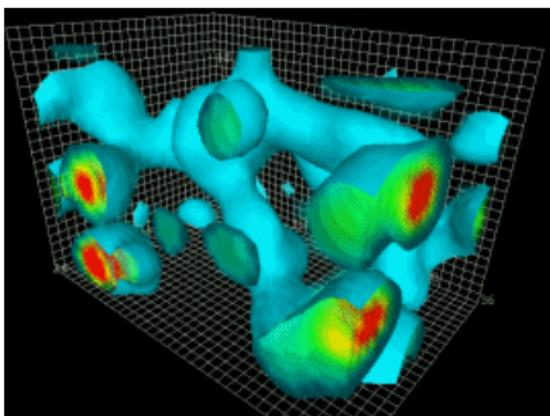
Overview

- 1 Motivation
- 2 Multilevel Monte Carlo for path integrals
 - Path integral formulation of quantum mechanics
 - Computational complexity of Markov chain MC
 - Hierarchical sampling
 - Multilevel Monte Carlo
- 3 Quantum mechanical model problem
 - Topological oscillator
 - Results
- 4 Towards QCD: the 2d Schwinger model
- 5 Conclusion and future work

Computational Quantum Field Theory



CERN particle collision



Quantum Field Theory (QFT) simulations on D -dimensional* space-time lattice

- Discretise on lattice with spacing $a \Rightarrow$ **discretisation error**
- Monte Carlo average over configurations \Rightarrow **statistical error**

Very expensive as $a \rightarrow 0$ (continuum limit)

* $D = 4$ for lattice QCD

Cost analysis and multilevel hierarchy

Total cost for given lattice spacing a and tolerance ϵ_{stat}

- $a^\alpha \propto$ discretisation error (typically $\alpha = 1, 2$) \Rightarrow finer lattices
- $N_{\text{samples}}^{-1/2} \propto$ statistical error $< \epsilon_{\text{stat}}$ \Rightarrow more samples
- integrated autocorrelation time $\tau_{\text{int}} \propto a^{-z}$ ($z = 5$ in some cases)

$$\text{Cost}_{\text{StMC}} = O(\epsilon_{\text{stat}}^{-2} \cdot a^{-D-z})$$

Cost analysis and multilevel hierarchy

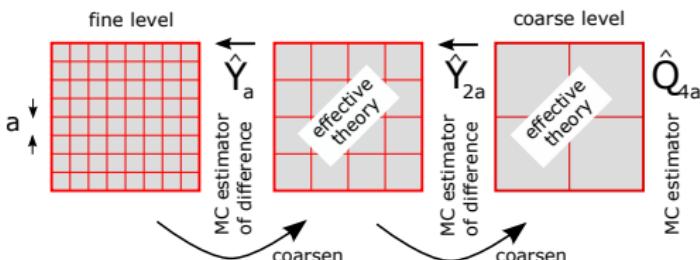
Total cost for given lattice spacing a and tolerance ϵ_{stat}

- $a^\alpha \propto$ discretisation error (typically $\alpha = 1, 2$) \Rightarrow finer lattices
- $N_{\text{samples}}^{-1/2} \propto$ statistical error $< \epsilon_{\text{stat}}$ \Rightarrow more samples
- integrated autocorrelation time $\tau_{\text{int}} \propto a^{-z}$ ($z = 5$ in some cases)

$$\text{Cost}_{\text{StMC}} = O(\epsilon_{\text{stat}}^{-2} \cdot a^{-D-z}) \quad \Rightarrow \quad \text{Cost}_{\text{MLMC}} = O(\epsilon_{\text{stat}}^{-2} \cdot a^{-D+\alpha} + a^{-D})$$

Key idea: use **hierarchy of coarse level theories** to

- ① accelerate sampling (reduce autocorrelations)
- ② exploit multilevel variance reduction [Giles ('08), Dodwell et al. ('15)]



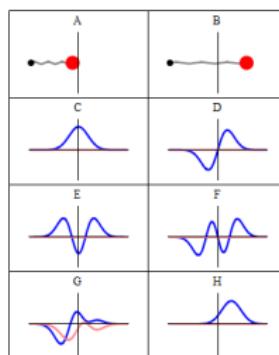
Quantum mechanics ($D = 1$ dimensions)

Particle moving in potential $V(x)$

Lagrangian $\mathcal{L} \Rightarrow$ action S

$$\begin{aligned} S(x(t)) &= \int_0^T \mathcal{L}(x(t)) dt, \quad \mathcal{L} = T - V \\ &= \int_0^T \left\{ \frac{1}{2} \dot{x}(t)^2 - V(x(t)) \right\} dt \end{aligned}$$

stationary $S \Rightarrow$ Euler-Lagrange \Rightarrow Newton



$$V(x) = \frac{1}{2} kx^2, -\frac{\partial V}{\partial x} = -kx$$

$$\frac{\delta S}{\delta x(t)} \Big|_{x_{\text{class}}(t)} = 0 \Rightarrow \frac{\partial \mathcal{L}}{\partial x} = \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}} \Rightarrow \ddot{x}_{\text{class}}(t) = -\frac{\partial V(x)}{\partial x} \Big|_{x_{\text{class}}(t)}$$

Quantum mechanics: “smear out” paths, consider all $x(t)$ with

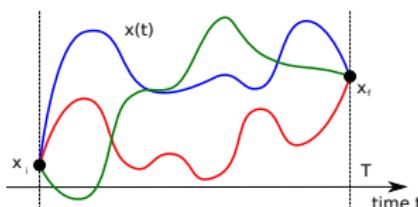
$$\frac{\delta S}{\delta x(t)} \approx 0, \quad \text{penalise deviations from } x_{\text{class}}(t)$$

Path integral formulation of quantum mechanics

Feynman's path integral (in Euclidean time)

observable Q (QoI, quantity-of-interest)

$$\mathbb{E}[Q] = \frac{\int Q(x(t)) e^{-\frac{1}{\hbar} S(x(t))} \mathcal{D}x(t)}{\int e^{-\frac{1}{\hbar} S(x(t))} \mathcal{D}x(t)}$$



Very elegant approach to quantisation

- Integrand highly peaked around classical path $x_{\text{class}}(t)$ as $\hbar \rightarrow 0.$ [†]
- Readily extended to quantum field theory ($D > 1$)
- Resembles statistical field theory with temperature $T \propto \hbar$
- Non-perturbative predictions for strongly coupled theories (lattice QCD)

Only minor problem: **integral is infinite dimensional!**

[†]will work in units where $\hbar = 1$ from now on

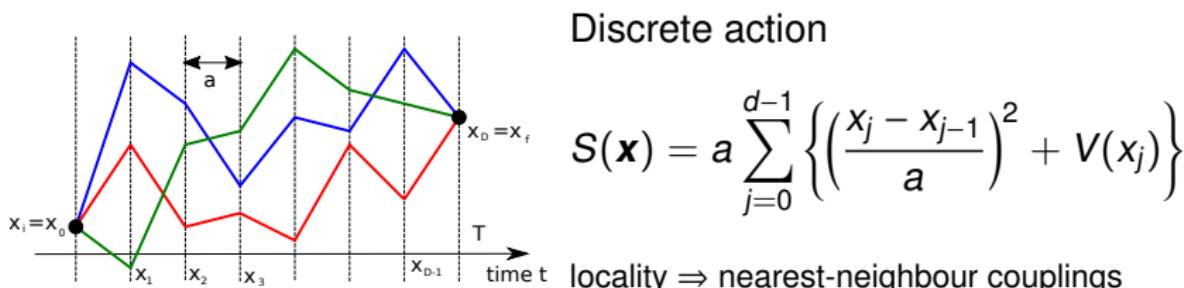
Discrete path integral

Discretisation[‡] → d -dimensional integral:

$$x(t) \rightarrow \mathbf{x} = (x_0, x_1, x_2, \dots, x_{d-1}) \in \mathbb{R}^d, x_j \approx x(t_j = a \cdot j)$$

lattice spacing $a = T/d$ ($\rightarrow 0$ in continuum limit)

$$\mathbb{E}[Q] = \frac{\int Q(\mathbf{x}) e^{-S(\mathbf{x})} dx_0 dx_1 \dots dx_{d-1}}{\int e^{-S(\mathbf{x})} dx_0 dx_1 \dots dx_{d-1}} + \text{discretisation error}$$



[‡]also regularises UV divergences in QFT

Markov chain Monte Carlo

Monte Carlo evaluation + importance sampling:

Draw $\mathbf{x}^{(k)} \sim \pi^*$, $\pi^*(\mathbf{x}) = Z^{-1} e^{-S(\mathbf{x})}$ (Z = normalisation constant)

$$\mathbb{E}[Q] = \frac{1}{N} \sum_{k=1}^N Q[\mathbf{x}^{(k)}] + \text{discretisation error} + \text{sampling error}$$

Sample from π^* : Markov chain $\mathbf{x}^{(0)} \rightarrow \mathbf{x}^{(1)} \rightarrow \mathbf{x}^{(2)} \rightarrow \dots \rightarrow \mathbf{x}^{(N)}$

Metropolis-Hastings: $\mathbf{x}^{(t)} \rightarrow \mathbf{x}^{(t+1)}$, only depends on ratio

$$\frac{\pi^*(\mathbf{x}^{(t+1)})}{\pi^*(\mathbf{x}^{(t)})} = \exp \left[- (S(\mathbf{x}^{(t+1)}) - S(\mathbf{x}^{(t)})) \right]$$

Computational cost

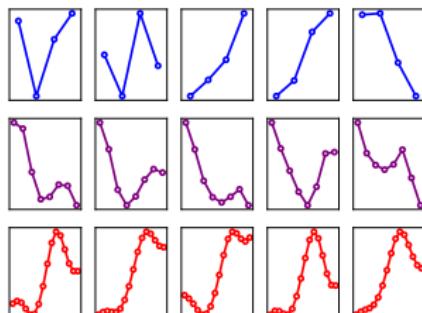
How expensive is a simulation?

- **Discretisation error** (bias)

$$\Delta_{\text{disc}} \leq C_0 a^\alpha \leq \epsilon_{\text{disc}}, \alpha = 1, 2$$

- **Statistical error**

$$\Delta_{\text{stat}} \propto N_{\text{samples}}^{-1/2} \leq \epsilon_{\text{stat}}$$



Total cost

- ① Cost per path: $\text{Cost}_{\text{path}} \propto d \propto a^{-1}$
- ② Number of **independent** samples $N_{\text{samples}} \propto \epsilon_{\text{stat}}^{-2}$
- ③ Integrated autocorrelation time: $\tau_{\text{int}} \propto a^{-z}$

$$\Rightarrow \quad \text{Cost}_{\text{StMC}} = \text{Cost}_{\text{path}} \cdot N_{\text{samples}} \cdot \tau_{\text{int}} \propto \epsilon_{\text{disc}}^{-(1+z)/\alpha} \cdot \epsilon_{\text{stat}}^{-2}$$

Hierarchy of coarse lattices

Growth of **autocorrelations** can be a serious issue
 (phase transition = “freezing” of topological charge, $z \gg 1 \Rightarrow$ later)

⇒ **Hierarchical sampling**

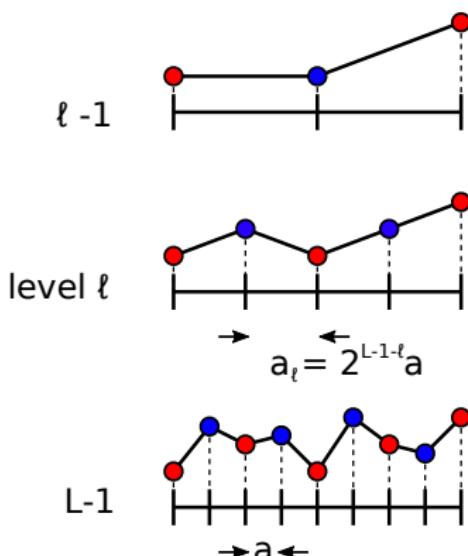
Hierarchy of L (coarser) lattices \mathcal{T}_ℓ

$$\pi_\ell(\mathbf{x}) = Z_\ell^{-1} \exp [-S_\ell(\mathbf{x})]$$

Action S_ℓ on level $\ell = 0, 1, \dots, L - 1$

Finest level ($\ell = L - 1$):

- action $S_{L-1} = S$
- probability density $\pi_{L-1} = \pi^*$



Effective theories

Coarse theories S_ℓ

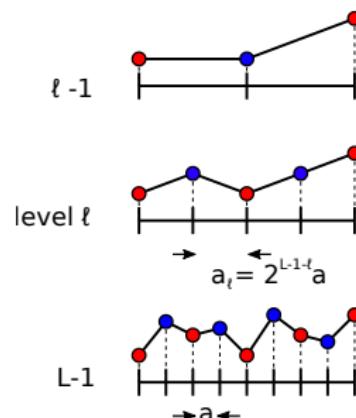
$$\pi_\ell(\mathbf{x}) = Z_\ell^{-1} \exp [-S_\ell(\mathbf{x})]$$

Even/odd decomposition: $\mathbf{x} = [\tilde{\mathbf{x}}, \mathbf{x}']$

Integrate out **odd modes** in S_ℓ

⇒ **effective theory** $S_{\ell-1}$

(Renormalisation Group transformation)



$$S_{\ell-1}(\mathbf{x}') = -\log \left\{ \int e^{-S_\ell([\tilde{\mathbf{x}}, \mathbf{x}'])} d\tilde{\mathbf{x}} \right\} + C_\ell$$

more explicitly:

$$S_{\ell-1}(x_0, x_2, x_4, \dots) = -\log \left\{ \int e^{-S_\ell(x_0, x_1, x_2, x_3, \dots)} dx_1 dx_3 \dots \right\} + C_\ell$$

Effective theories

Coarse theories S_ℓ

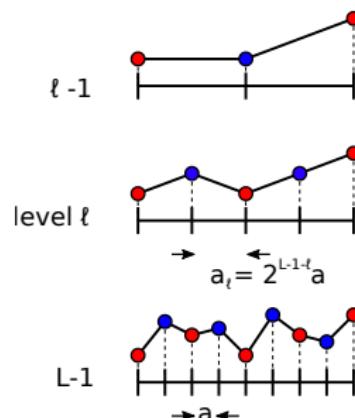
$$\pi_\ell(\mathbf{x}) = Z_\ell^{-1} \exp [-S_\ell(\mathbf{x})]$$

Even/odd decomposition: $\mathbf{x} = [\tilde{\mathbf{x}}, \mathbf{x}']$

Integrate out **odd modes** in S_ℓ

⇒ **effective theory** $S_{\ell-1}$

(Renormalisation Group transformation)



$$S_{\ell-1}(\mathbf{x}') \approx -\log \left\{ \int e^{-S_\ell([\tilde{\mathbf{x}}, \mathbf{x}'])} d\tilde{\mathbf{x}} \right\} + C_\ell$$

more explicitly:

$$S_{\ell-1}(x_0, x_2, x_4, \dots) \approx -\log \left\{ \int e^{-S_\ell(x_0, x_1, x_2, x_3, \dots)} dx_1 dx_3 \dots \right\} + C_\ell$$

(approximation sufficient for us!)

Hierarchical sampling

Idea [Christen & Fox (2005)]

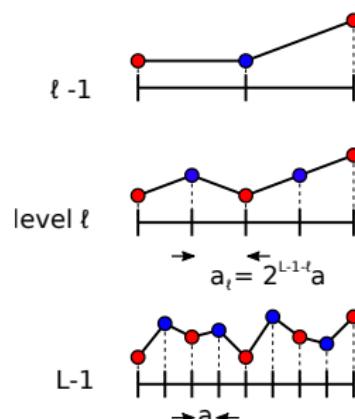
- Use coarse-level sample $\mathbf{y}_{\ell-1}$ to generate fine-level proposal
 $\mathbf{y}_\ell = [\tilde{\mathbf{y}}_\ell, \mathbf{y}_{\ell-1}]$
- Accept/reject with fine-level theory $\Rightarrow \mathbf{x}^{(t+1)} \sim \pi_\ell$

Fill-in odd points

\Rightarrow conditional probability distribution $\tilde{\pi}_\ell$ for
 $\tilde{\mathbf{x}}$, given even points \mathbf{x}' :

$$\tilde{\pi}_\ell(\tilde{\mathbf{x}}|\mathbf{x}') = \tilde{Z}_\ell(\mathbf{x}')^{-1} \exp \left[-\tilde{S}_\ell([\tilde{\mathbf{x}}, \mathbf{x}']) \right]$$

with $\tilde{S}_\ell \approx S_\ell$



also works in higher dimensions \Rightarrow later

Two level Metropolis Hastings step

Algorithm 1 Two-level Metropolis Hastings step on level ℓ

Input: current $\mathbf{x}_\ell^{(t)} \sim \pi_\ell$, proposal distribution $q_{\ell-1}$. Output: new $\mathbf{x}_\ell^{(t+1)} \sim \pi_\ell$

- 1: Let $\mathbf{x}_\ell^{(t)} = [\tilde{\mathbf{x}}_\ell^{(t)}, \mathbf{x}_{\ell-1}^{(t)}]$ and pick $\mathbf{x}_{\ell-1}^{(t+1)} = \mathbf{y}_{\ell-1}$ from $q_{\ell-1}(\cdot | \mathbf{x}_{\ell-1}^{(t)})$.
- 2: **if** $\mathbf{x}_{\ell-1}^{(t+1)} = \mathbf{x}_{\ell-1}^{(t)}$ (coarse level proposal rejected) **then**
- 3: Set $\mathbf{x}_\ell^{(t+1)} \leftarrow \mathbf{x}_\ell^{(t)}$
- 4: **else**
- 5: Pick $\tilde{\mathbf{y}}_\ell$ from $\tilde{\pi}_\ell(\cdot | \mathbf{y}_{\ell-1})$ and let $\mathbf{y}_\ell = [\tilde{\mathbf{y}}_\ell, \mathbf{y}_{\ell-1}]$.
- 6: Compute $\frac{\pi_\ell(\mathbf{y}_\ell)}{\pi_\ell(\mathbf{x}_\ell^{(t)})} \cdot \underbrace{\frac{\tilde{\pi}_\ell(\tilde{\mathbf{x}}_\ell^{(t)} | \mathbf{x}_{\ell-1}^{(t)})}{\tilde{\pi}_\ell(\tilde{\mathbf{y}}_\ell | \mathbf{y}_{\ell-1})}}_{\text{proposal distribution}} \cdot \frac{\pi_{\ell-1}(\mathbf{x}_{\ell-1}^{(t)})}{\pi_{\ell-1}(\mathbf{y}_{\ell-1})} = e^{-\Delta S_\ell}$
- 7: Accept proposal \mathbf{y}_ℓ and set $\mathbf{x}_\ell^{(t+1)} \leftarrow \mathbf{y}_\ell$ with probability $\min\{1, e^{-\Delta S_\ell}\}$;
set $\mathbf{x}_\ell^{(t+1)} \leftarrow \mathbf{x}_\ell^{(t)}$ if proposal is rejected.
- 8: **end if**

$$q_{\ell-1}(\cdot | \mathbf{x}_{\ell-1}^{(t)}) = \pi_{\ell-1}(\cdot) \Rightarrow [\text{Alg. 2, Dodwell (2015)}]$$

Hierarchical sampler

Alg. 1 implicitly defines a proposal distribution $q_\ell^{(\text{TL})}(\cdot | \mathbf{x}_\ell^{(t)})$

⇒ apply Alg. 1 recursively, using standard Metropolis-Hastings method proposal (e.g. HMC) on coarsest level ($\ell = 1$)

Algorithm 2 Hierarchical delayed acceptance sampler

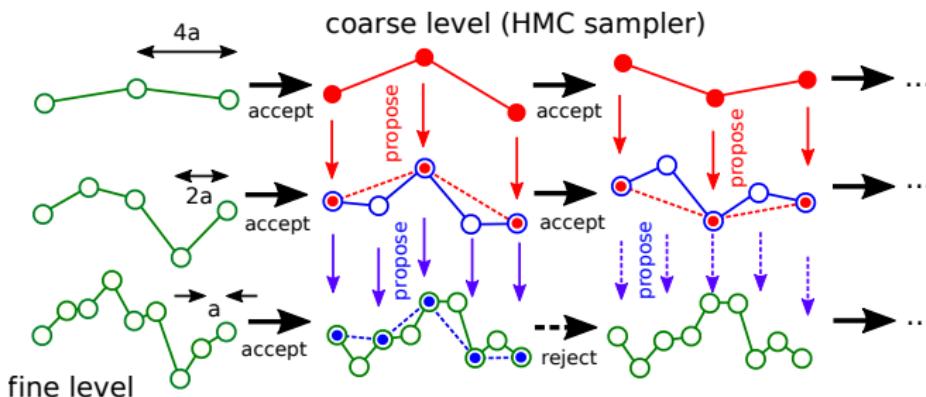
Input: current $\mathbf{x}_\ell^{(t)} \sim \pi_\ell$, Output: new $\mathbf{x}_\ell^{(t+1)} \sim \pi_\ell$

Generate $\mathbf{x}_\ell^{(t+1)}$ using Alg. 1 with level ℓ , current sample $\mathbf{x}_\ell^{(t)} \sim \pi_\ell$ and proposal distribution

$$q_{\ell-1}(\cdot | \mathbf{x}_{\ell-1}^{(t)}) = \begin{cases} q_0^{(\text{MH})}(\cdot | \mathbf{x}_0^{(t)}) & \text{for } \ell = 1 \\ q_{\ell-1}^{(\text{TL})}(\cdot | \mathbf{x}_{\ell-1}^{(t)}) & \text{for } \ell = 2, 3, \dots, L-1 \end{cases}$$

[picking from $q_{\ell-1}$ in line 1 of Alg. 1 induces recursive call of Alg. 1 on level $\ell-1$]

Hierarchical sampler



Hierarchical sampling, as described in Alg. 2, for $L = 3$ levels.

Hierarchical sampling

Hierarchical sampling tames growth of **autocorrelations** by generating non-local updates

$$\text{Cost}_{\text{StMC}} = O(\epsilon_{\text{stat}}^{-2} \cdot a^{-1-z}) = O(\epsilon_{\text{stat}}^{-2} \cdot \epsilon_{\text{disc}}^{-(1+z)/\alpha})$$

↓

$$\text{Cost}_{\text{HSMC}} = O(\epsilon_{\text{stat}}^{-2} \cdot a^{-1}) = O(\epsilon_{\text{stat}}^{-2} \cdot \epsilon_{\text{disc}}^{-1/\alpha})$$

(Recall that $\epsilon_{\text{disc}} \propto a^\alpha$)

Multilevel Monte Carlo

Telescoping sum ($\mathbb{E}[\cdot]$ linear, $\widehat{\cdot}$ unbiased \Rightarrow no additional bias)

$$\begin{aligned}
 \mathbb{E}[Q] &\approx \mathbb{E}[Q_{L-1}] = \mathbb{E}[Q_{L-1} - Q_{L-2}] + \mathbb{E}[Q_{L-2}] \\
 &= \dots \\
 &= \mathbb{E}[Q_{L-1} - Q_{L-2}] + \mathbb{E}[Q_{L-2} - Q_{L-3}] \\
 &\quad + \dots + \mathbb{E}[Q_1 - Q_0] + \mathbb{E}[Q_0] \\
 &= \sum_{\ell=0}^{L-1} \mathbb{E}[Y_\ell] \approx \sum_{\ell=0}^{L-1} \widehat{Y}_\ell, \quad \widehat{Y}_\ell := \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} Y_\ell^{(j)},
 \end{aligned}$$

where

$$Y_\ell := \begin{cases} Q_0 & \text{for } \ell = 0 \\ Q_\ell - Q_{\ell-1} & \text{for } \ell = 1, 2, \dots, L-1, \end{cases}$$

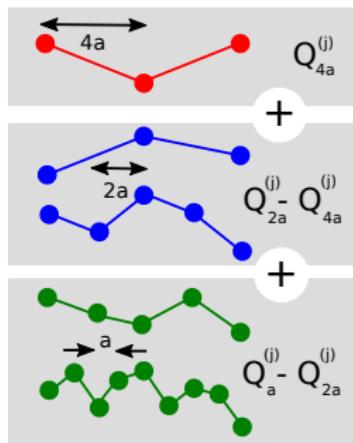
N_ℓ = # samples on level ℓ

Multilevel Monte Carlo

Why does this help?

Differences $Y_\ell = Q_\ell - Q_{\ell-1}$ have smaller variance \Rightarrow estimate Y_ℓ with fewer samples N_ℓ (assuming correct coupling of paths on levels ℓ , $\ell-1$)

$$\begin{aligned} \text{Var}[Y_{L-1}] &\ll \dots \ll \text{Var}[Y_1] \ll \text{Var}[Y_0] \\ \Rightarrow N_{L-1} &\ll \dots \ll N_1 \ll N_0 \end{aligned}$$



Shift cost to coarser levels, where action evaluation is cheap

$$N_\ell^{\text{eff}} = \epsilon_{\text{stat}}^{-2} \left(\sum_{\ell=0}^{L-1} \sqrt{\text{Var}[Y_\ell]} C_\ell^{\text{eff}} \right) \sqrt{\frac{\text{Var}[Y_\ell]}{C_\ell^{\text{eff}}}}, \quad C_\ell^{\text{eff}} = \text{cost of generating independent sample on level } \ell$$

[Giles (2008), Dodwell et al. (2015), Dodwell et al. (2019)]

Multilevel Monte Carlo

Algorithm 3 Multilevel Monte Carlo.

Input: Number of levels L , number of samples per level N_ℓ^{eff} and sub-sampling rates t_ℓ for $\ell = 0, \dots, L - 1$, Output: MLMC estimate $\widehat{Q}_{L,\{N_\ell^{\text{eff}}\}}^{\text{MLMC}}$ for QoI.

```

1: for level  $\ell = 0, \dots, L - 1$  do
2:   for  $j = 1, \dots, N_\ell^{\text{eff}}$  do
3:     if  $\ell = 0$  then
4:       Create new  $\mathbf{x}_0^{(t+t_0)}$  from  $\mathbf{x}_0^{(t)}$  with standard Metropolis-Hastings.
5:       Compute  $Y_0^{(j)} = Q_0(\mathbf{x}_0^{(t+t_0)})$ 
6:     else
7:       Create new  $\mathbf{x}_\ell^{(t+1)}$  from  $\mathbf{x}_\ell^{(t)}$  with Alg. 1 and  $q_{\ell-1}(\cdot | \mathbf{x}_\ell^{(t)}) = \pi_{\ell-1}$  ;
        In practice, use  $t_{\ell-1}$  steps of Alg. 2 to compute approximately
        independent sample  $\mathbf{z}_{\ell-1}^{(t+t_{\ell-1})}$  on level  $\ell - 1$ .
8:       Compute  $Y_\ell^{(j)} = Q_\ell(\mathbf{x}_\ell^{(t+1)}) - Q_{\ell-1}(\mathbf{z}_{\ell-1}^{(t+t_{\ell-1})})$ .
9:     end if
10:    end for
11:  end for
12: Compute  $\widehat{Q}_{L,\{N_\ell^{\text{eff}}\}}^{\text{MLMC}} = \sum_{\ell=0}^{L-1} \widehat{Y}_{\ell,N_\ell^{\text{eff}}}$  with  $\widehat{Y}_{\ell,N_\ell^{\text{eff}}} = \frac{1}{N_\ell^{\text{eff}}} \sum_{j=1}^{N_\ell^{\text{eff}}} Y_\ell^{(j)}$ ,

```

Multilevel Monte Carlo cost

Variance decay

$$\text{Var}[Y_\ell] \leq 2^{-\beta} \cdot \text{Var}[Y_{\ell-1}] \leq \dots \leq 2^{-\beta(\ell-1)} \cdot \text{Var}[Y_1]$$

Cost of Multilevel Monte Carlo (assuming exact sampling from π_ℓ)

$$\text{Cost}_{\text{StMC}} = O(\epsilon_{\text{stat}}^{-2} \cdot \epsilon_{\text{disc}}^{-(1+z)/\alpha})$$



$$\text{Cost}_{\text{HSMC}} = O(\epsilon_{\text{stat}}^{-2} \cdot \epsilon_{\text{disc}}^{-1/\alpha})$$



$$\text{Cost}_{\text{MLMC}} = \begin{cases} O(\epsilon_{\text{stat}}^{-2} + \epsilon_{\text{disc}}^{-1}) & \text{for } \beta > 1 \\ O(\epsilon_{\text{stat}}^{-2} \cdot |\log_2 \epsilon_{\text{disc}}|^2 + \epsilon_{\text{disc}}^{-1}) & \text{for } \beta = 1 \end{cases}$$

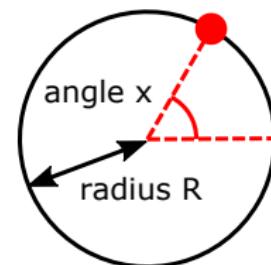
Quantum mechanical model problem

Topological oscillator

Free particle on a ring, $I_0 = m_0 R^2$ (=moment of inertia)

$$S(x(t)) = \frac{I_0}{2} \int_0^T \dot{x}^2 dt \quad \text{with } x \in [-\pi, \pi)$$

$$S(\mathbf{x}) = \frac{I_0}{a} \sum_{j=0}^{d-1} (1 - \cos(x_j - x_{j-1})).$$



Coarse level theories:

⇒ **perturbative matching** (= approximate RG transformation)

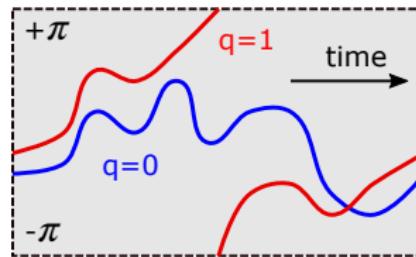
$$I_0 = I_0^{(L-1)} \rightarrow I_0^{(L-2)} \rightarrow I_0^{(L-3)} \rightarrow \cdots \rightarrow I_0^{(\ell)}$$

Topological charge

Topological charge $q = \#$ complete revolutions in time T

$$q(x(t)) = \frac{1}{2\pi} \int_0^T \dot{x} dt \in \mathbb{Z}$$

$$\approx q(\mathbf{x}) = \frac{1}{2\pi} \sum_{j=0}^{d-1} \{x_j - x_{j-1} \bmod [-\pi, \pi)\}$$



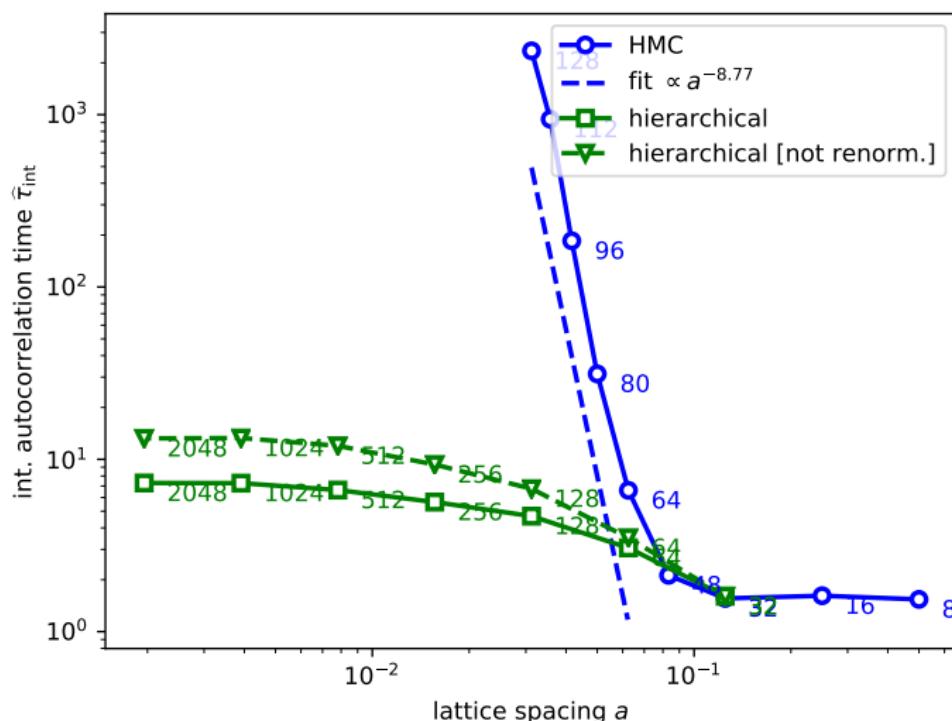
QoI = **topological susceptibility**

$$Q(\mathbf{x}) = \chi_t(\mathbf{x}) = \frac{q(\mathbf{x})^2}{T}, \quad \mathbb{E}[\chi_t] \rightarrow \frac{1}{4\pi^2 l_0} \quad \text{as } a \rightarrow 0, T \rightarrow \infty$$

integrated autocorrelation time $\tau_{\text{int}} \propto a^{-z}$, $z \gg 1$

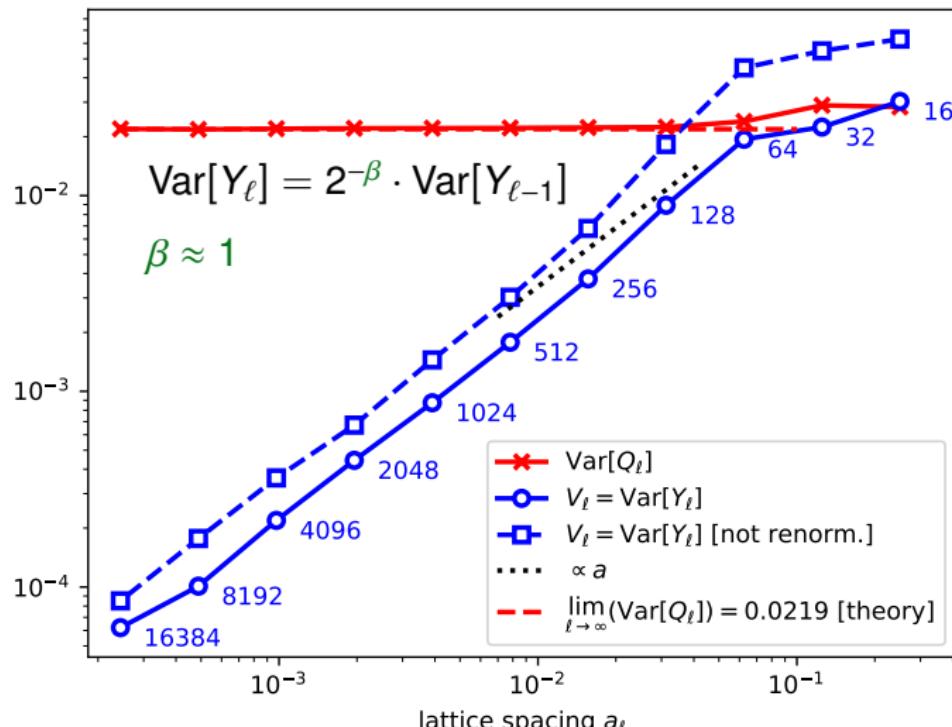
Integrated autocorrelation time

τ_{int} ($d_0 = 32$ for hierarchical sampler)



Variance decay

Decay of $\text{Var}[Y_\ell]$



Multilevel Monte Carlo setup

Compared algorithms

1 Case 1:

- a HMC: Standard MC with single-level HMC sampler
- b HSMC: Standard MC with hierarchical sampler (Alg. 2)
- c MLMC: Multilevel MC, generate coarse level samples with (subsampled) Alg. 2; subsampling rates $t_\ell = \lceil 2\tau_{\text{int},\ell} \rceil$

2 Case 2

- a StMC: Standard MC with cluster algorithm [Wolff (1989)]
- b MLMC: Multilevel MC, generate coarse level samples with (subsampled) cluster algorithm; subsampling rates $t_\ell = \lceil 2\tau_{\text{int},\ell} \rceil$

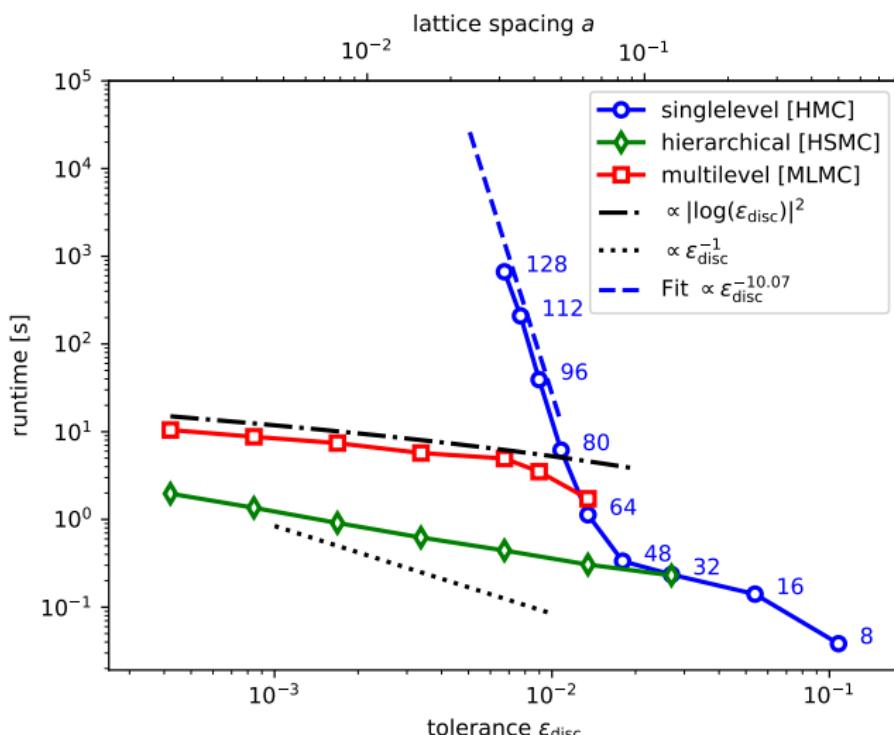
Reported runtimes do not include burn-in (but burn-in can also be accelerated with our methods, see also [Endres et al. (2015)])

All results obtained with C++ code available at

<https://bitbucket.org/em459/mlmcpathintegral>

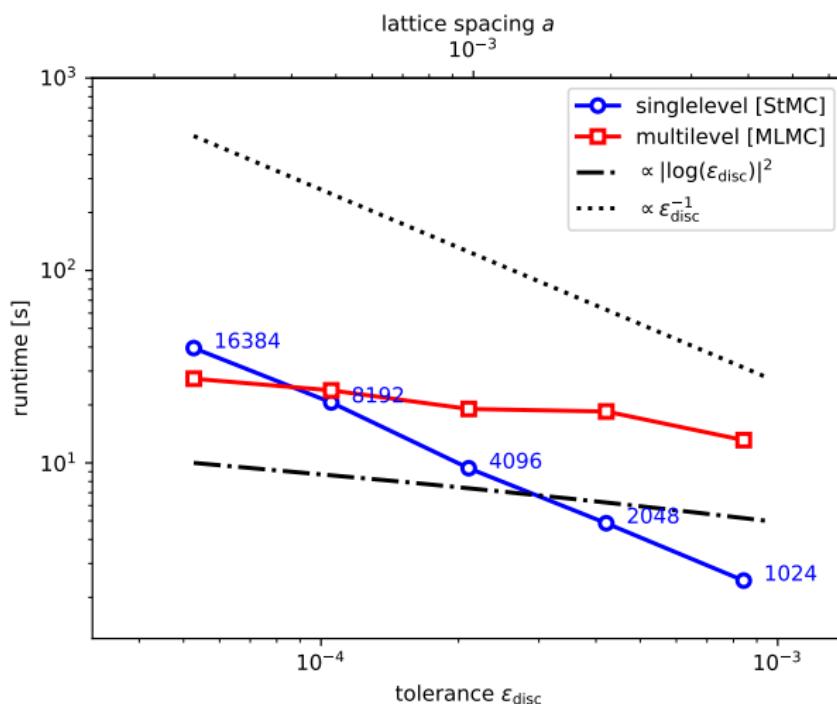
Cost comparison

Runtime for topological oscillator with $\epsilon_{\text{stat}} = 10^{-2}$



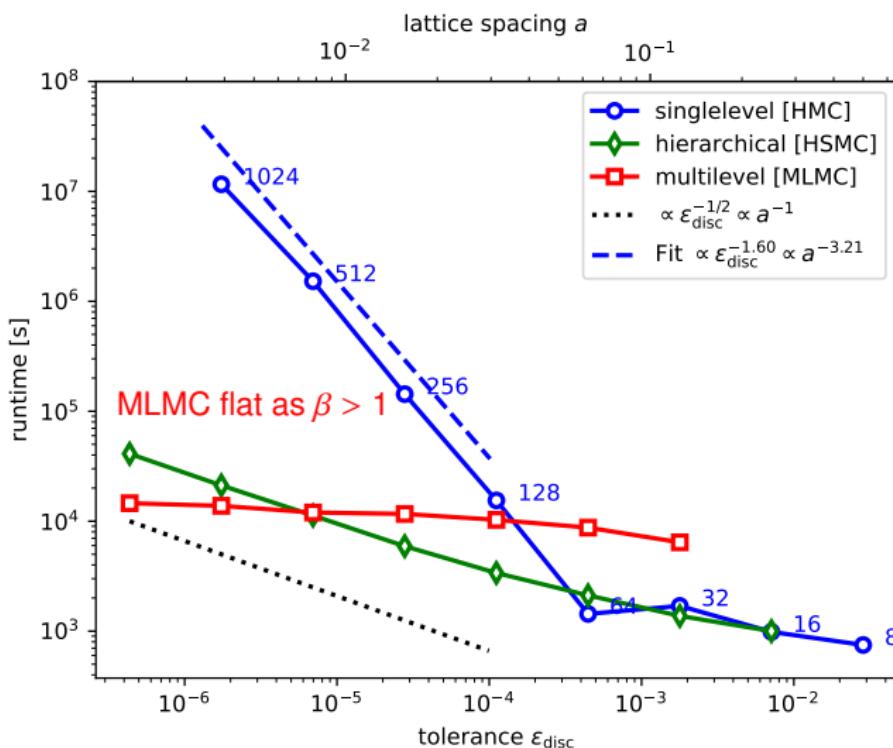
Cost comparison

Runtime, topological oscillator, cluster algorithm with $\epsilon_{\text{stat}} = 10^{-3}$



Cost comparison

Runtime for double well potential with $\epsilon_{\text{stat}} = 10^{-4}$



Towards QCD

2d Schwinger model = toy model for quantum electrodynamics

$$\pi(\textcolor{blue}{A}_\mu) = Z^{-1} \exp[-S(\textcolor{blue}{A}_\mu)] \quad \text{gauge potential } A_\mu(x) \in \mathbb{R} \text{ for all } x \in \Omega, \mu \in \{0, 1\}$$

Action

$$S_{\text{cont}}(\textcolor{blue}{A}_\mu) = \int d^2x \frac{1}{4} \sum_{\mu\nu} \textcolor{red}{F}_{\mu\nu}^2 \quad \text{field strength } F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$$

Invariance under local gauge transformation $G(x) = e^{i\Lambda(x)} \in U(1)$
 (unitary group = rotations in complex plane)

$$A_\mu(x) \mapsto \textcolor{blue}{A}_\mu(x) - \frac{i}{g} G(x) \partial_\mu G^{-1}(x), \quad F_{\mu\nu} \mapsto \textcolor{red}{F}_{\mu\nu}$$

Lattice discretisation

Lattice action design principle: preservation of local gauge invariance[§]

$$S_{\text{lat}}[U] = \beta \sum_p \left(1 - \frac{1}{2} (U_p + U_p^\dagger) \right) \quad \beta = \frac{1}{a^2 g^2}$$

Link $U_{n,\mu} \in U(1)$, $n \in \mathbb{Z} \times \mathbb{Z}$

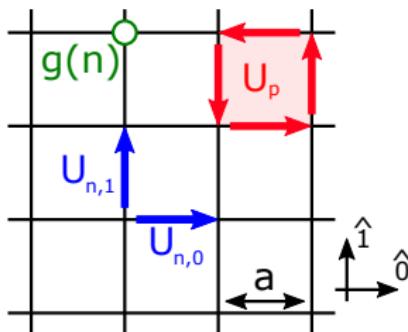
$$A_\mu(x) \rightarrow U_{n,\mu} \approx e^{igaA_\mu(na)}$$

Plaquette

$$U_p = U_{n,0} U_{n+\hat{\alpha},1} U_{n+\hat{\alpha},0}^\dagger U_{n,1}^\dagger$$

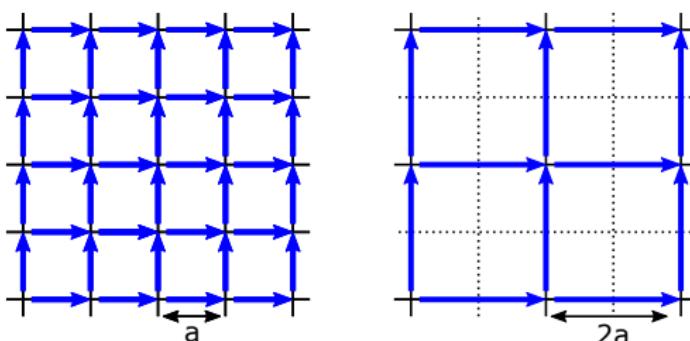
Local gauge transformation $g(n) \in U(1)$

$$U_{n,\mu} \mapsto g(n) U_{n,\mu} g^\dagger(n + \mu),$$



[§]ensures that conservation laws survive discretisation

Lattice hierarchy



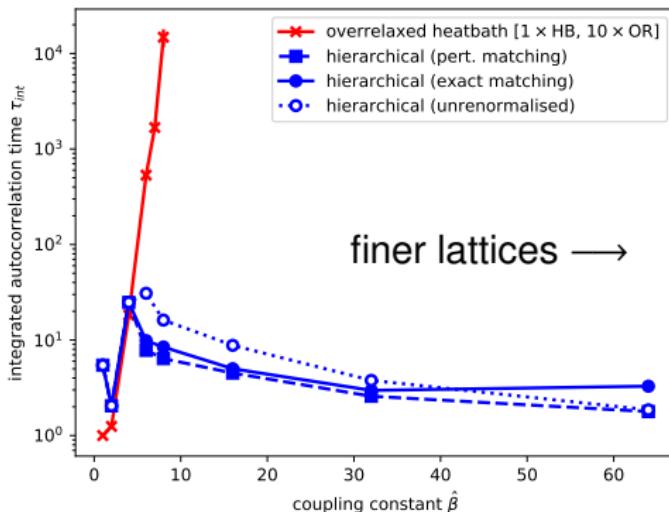
Gauge invariant coarsening

$$U_{n,\mu}^{(\text{coarse})} := U_{n,\mu} U_{n+\hat{\mu},\mu}$$

Fill-in of additional fine links: draw from from 4-dimensional distribution in each coarse cell

Preliminary results

Integrated autocorrelation time τ_{int} on 16×16 lattice, $\beta \propto a^{-2}$
 QoI = topological susceptibility χ_t

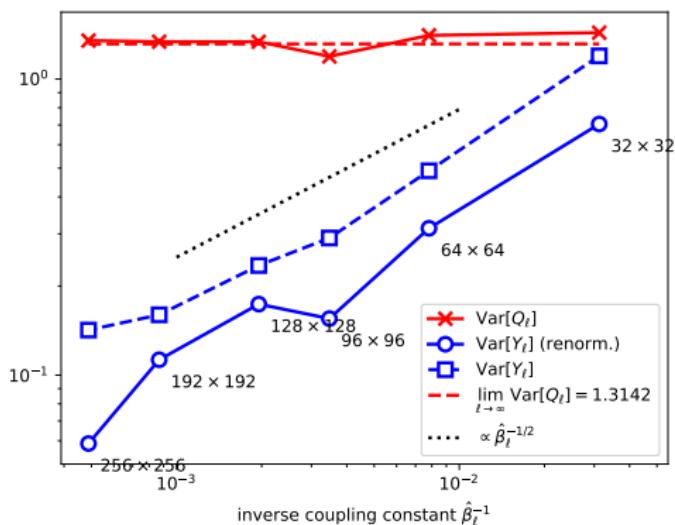


Comparison **hierarchical sampler** ↔ **Gibbs sampler (heatbath + OR)**

[†]scaled by physical volume

Preliminary results

Variance decay constant physical volume



... appears to be consistent with $\text{Var}[Y_\ell] \propto \beta^{-1/2} \propto a$
 bias (discretisation error) $\propto a^2$

Expected performance gains in D dimensions

Total cost for given tolerances ϵ_{disc} , ϵ_{stat}

- discretisation error $< \epsilon_{\text{disc}}$ (typically $\alpha = 1, 2$)
- statistical error $< \epsilon_{\text{stat}}$
- integrated autocorrelation time $\tau_{\text{int}} \propto a^{-z}$ ($z = 5$ in some cases)

$$\text{Cost}_{\text{StMC}} = O(\epsilon_{\text{stat}}^{-2} \cdot \epsilon_{\text{disc}}^{-(D+z)/\alpha})$$

↓ (hierarchical sampling)

$$\text{Cost}_{\text{hierarchical}} = O(\epsilon_{\text{stat}}^{-2} \cdot \epsilon_{\text{disc}}^{-D/\alpha})$$

↓ (multilevel Monte Carlo)

$$\text{Cost}_{\text{MLMC}} = O(\epsilon_{\text{stat}}^{-2} \cdot \epsilon_{\text{disc}}^{-D/\alpha+1} + \epsilon_{\text{disc}}^{-D/\alpha}) \quad \text{for } D > \alpha$$

Continuum limit ($\epsilon_{\text{disc}} \rightarrow 0$): $\frac{\text{Cost}_{\text{hierarchical}}}{\text{Cost}_{\text{MLMC}}} \approx \epsilon_{\text{stat}}^{-2}$

Conclusion

Summary and Outlook

- Cost of evaluating path integrals grows rapidly as $a \rightarrow 0$
- Hierarchical sampling eliminates autocorrelations
- Multilevel Monte Carlo leads to further speedups for $a \ll 1$
 $C_{\text{hierarchical}}/C_{\text{MLMC}} \approx \epsilon_{\text{stat}}^{-2}$ for ϵ_{stat} fixed, $\epsilon_{\text{disc}} \rightarrow 0$
- Coarse level matching improves performance by $2 \times - 3 \times$

Future work

- **We really want to apply this to QFT** (esp. Lattice QCD)
- Expected performance gains from MLMC higher for $D > 1$
- “Coarsen” in other categories (\Rightarrow multiindex MC)
 - Size of physical box
 - Increase dynamical quark mass on coarse levels
- Improved coarse level theories from renormalisation group transformations / lattice perturbation theory / machine learning
- Combine with neural MCMC [Nicoli et al. arXiv:2007.07115]
 \Rightarrow Pan Kessel, Christopher Anders (BIFOLD & TU Berlin)

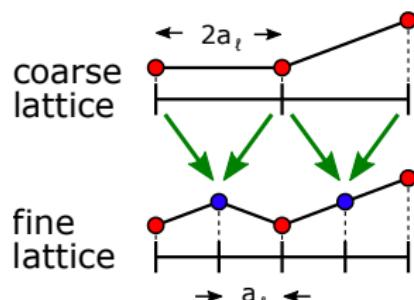
Higher dimensions

Does hierarchical sampling only work in 1d?

Recall two-level Metropolis-Hastings step (Alg. 1) in 1d:

- Fill-in fine unknowns \bullet by drawing from $\tilde{\pi}_\ell(\cdot|\bullet)$ and set $\bullet = [\bullet, \bullet]$
- Accept with ratio

$$\frac{\pi_\ell(\bullet_{\text{new}})}{\pi_\ell(\bullet_{\text{old}})} \cdot \underbrace{\frac{\tilde{\pi}_\ell(\bullet_{\text{old}}|\bullet_{\text{old}})}{\tilde{\pi}_\ell(\bullet_{\text{new}}|\bullet_{\text{new}})} \cdot \frac{\pi_{\ell-1}(\bullet_{\text{old}})}{\pi_{\ell-1}(\bullet_{\text{new}})}}_{\text{proposal distribution}}$$



Higher dimensions

2 dimensions \Rightarrow two-stage fill-in process with rotated lattice

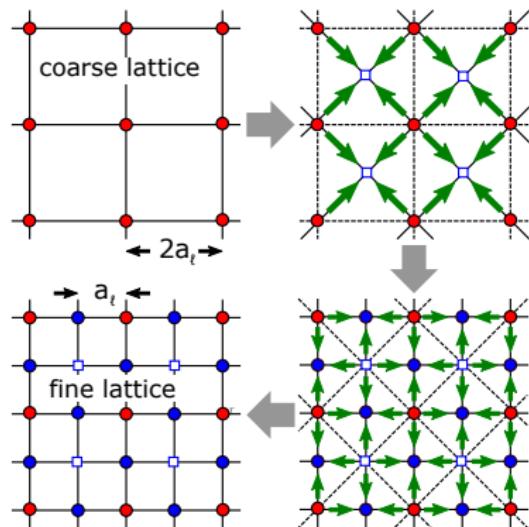
see also [Schmidt (1983), Faas & Hilhorst (1986)] for Ising model

- Fill in $\bullet \rightarrow \square$ with

$$\tilde{\pi}_\ell^{(1)} \sim \exp[-\tilde{S}_\ell^{(1)}]$$

- Fill in $\bullet, \square \rightarrow \bullet$ with

$$\tilde{\pi}_\ell^{(2)} \sim \exp[-\tilde{S}_\ell^{(2)}]$$



proposal-actions

$$\tilde{S}_\ell^{(1)} \approx S[\text{lat. spacing} = \sqrt{2}a_\ell],$$

$$\tilde{S}_\ell^{(2)} \approx S[\text{lat. spacing} = a_\ell]$$

Higher dimensions

Two-level Metropolis-Hastings step (Alg. 1)

- Fill-in fine unknowns \square by drawing from $\tilde{\pi}_\ell^{(1)}(\cdot | \bullet)$
- Fill-in fine unknowns \bullet by drawing from $\tilde{\pi}_\ell^{(2)}(\cdot | \bullet, \square)$ and set
 $\bullet = [\bullet, \square, \bullet]$
- Accept with ratio

$$\frac{\pi_\ell(\bullet_{\text{new}})}{\pi_\ell(\bullet_{\text{old}})} \cdot \underbrace{\frac{\tilde{\pi}_\ell^{(2)}(\bullet_{\text{old}} | \bullet_{\text{old}}, \square_{\text{old}})}{\tilde{\pi}_\ell^{(2)}(\bullet_{\text{new}} | \bullet_{\text{new}}, \square_{\text{new}})}}_{\text{proposal distribution}} \cdot \frac{\tilde{\pi}_\ell^{(1)}(\square_{\text{old}} | \bullet_{\text{old}})}{\tilde{\pi}_\ell^{(1)}(\square_{\text{new}} | \bullet_{\text{new}})} \cdot \frac{\pi_{\ell-1}(\bullet_{\text{old}})}{\pi_{\ell-1}(\bullet_{\text{new}})}$$

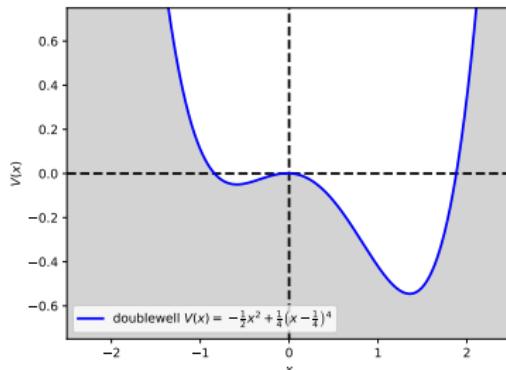
Double-well potential

Particle moving in potential

$$V(x) = -\frac{1}{2}x^2 + \frac{1}{4}\left(x - \frac{1}{4}\right)^4$$

QoI = mean square displacement

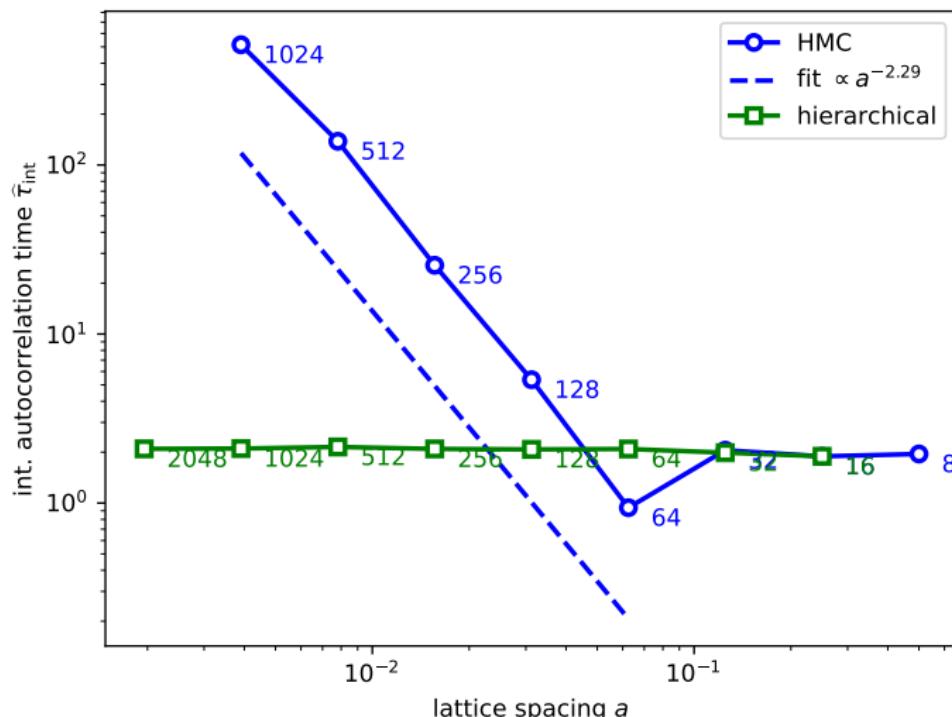
$$Q(\mathbf{x}) = \frac{1}{d} \sum_{j=0}^{d-1} x_j^2$$



Boundary conditions (all numerical experiments): $x(0) = x(T)$
do not consider finite “volume” errors due to $T < \infty$

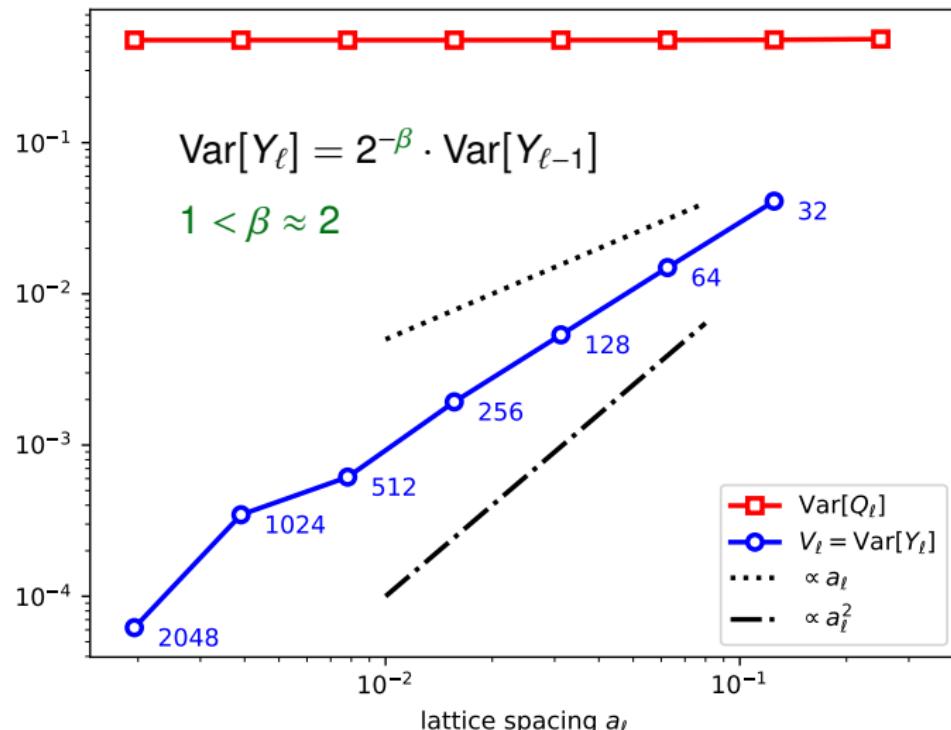
Integrated autocorrelation time

τ_{int} , double well potential ($d_0 = 16$ for hierarchical sampler)



Variance decay

Decay of $\text{Var}[Y_\ell]$, double well potential



Effect of coarse-level matching

Improved coarse level action \Rightarrow better results

- $t_{\text{MLMC}}^{(0)}$: no matching $I_0^{(\ell)} = I_0$ for all $\ell = 0, 1, \dots, L - 1$
- t_{MLMC} : with matching $I_0 = I_0^{(L-1)} \rightarrow I_0^{(L-2)} \rightarrow \dots \rightarrow I_0^{(1)} \rightarrow I_0^{(0)}$

| d | a | ϵ_{disc} | $t_{\text{MLMC}}^{(0)}$ | t_{MLMC} | speedup |
|------|--------|--------------------------|-------------------------|-------------------|---------|
| 64 | 0.0625 | 0.01348 | 4.07 | 1.72 | 2.4x |
| 128 | 0.0312 | 0.00674 | 10.68 | 4.95 | 2.2x |
| 256 | 0.0156 | 0.00337 | 16.47 | 5.72 | 2.9x |
| 512 | 0.0078 | 0.00168 | 19.27 | 7.42 | 2.6x |
| 1024 | 0.0039 | 0.00084 | 21.92 | 8.76 | 2.5x |
| 2048 | 0.0020 | 0.00042 | 25.64 | 10.47 | 2.5x |

Expected performance gains in D dimensions

Multilevel MC literature: **bound total RMS error $< \epsilon$**
balance disc/stat errors $\Rightarrow \epsilon_{\text{disc}} = \epsilon_{\text{stat}} = \epsilon / \sqrt{2}$

$$\text{Cost}_{\text{StMC}} = O(\epsilon^{-2-(D+z)/\alpha})$$

\downarrow (hierarchical sampling)

$$\text{Cost}_{\text{hierarchical}} = O(\epsilon^{-2-D/\alpha})$$

\downarrow (multilevel Monte Carlo)

$$\text{Cost}_{\text{MLMC}} = O(\epsilon^{-1-D/\alpha}) \quad \text{for } D > \alpha$$