Multiscale Monte Carlo Equilibration

Michael G. Endres
(MIT)

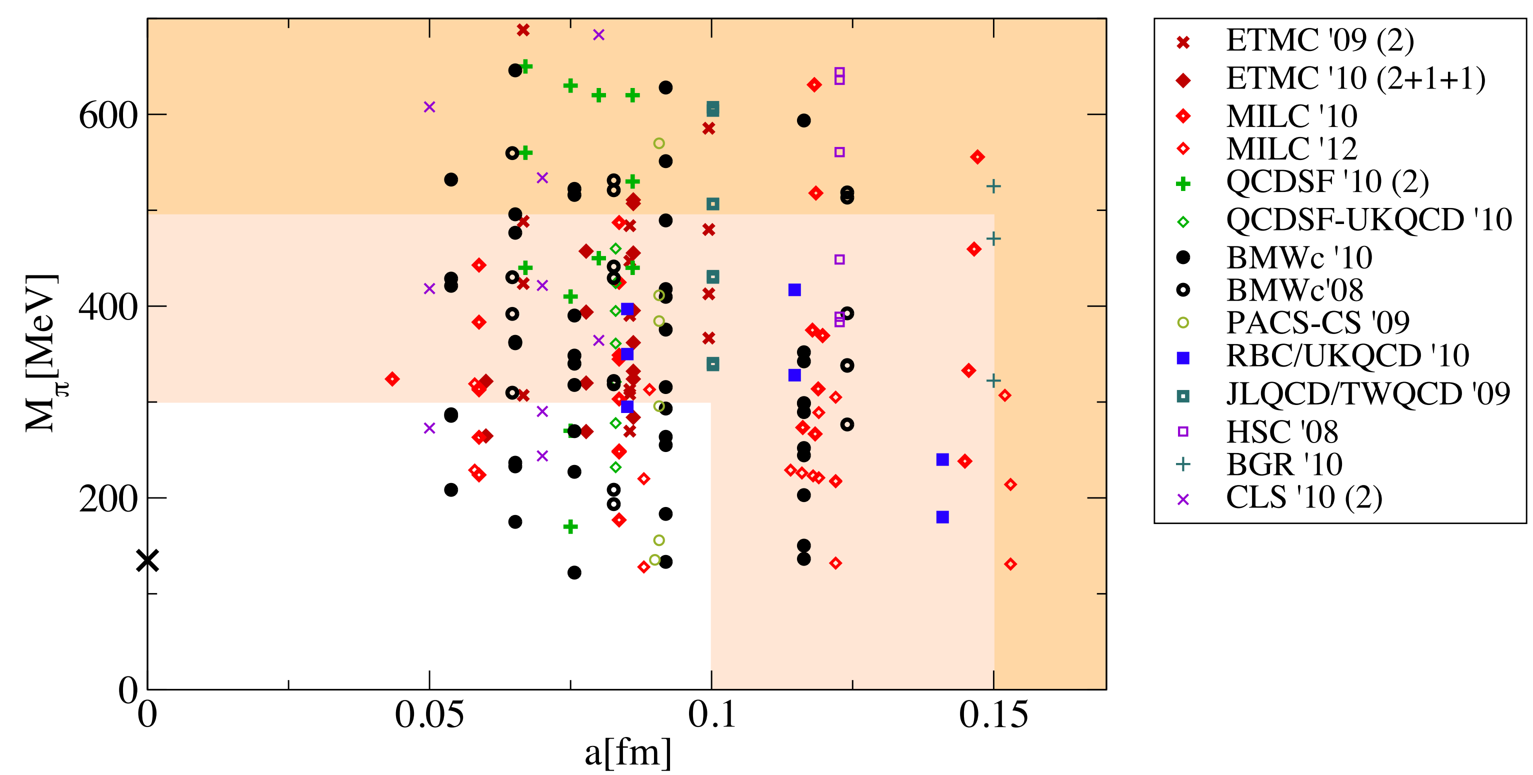
Nuclear Physics from Lattice QCD
INT-UW • Seattle, WA
May 2, 2016

- **M. G. E.**, R. C. Brower, W. Detmold, K. Orginos and A. V. Pochinksy Phys.Rev. D **92** (2015) 114516 [arXiv:1510.04675]
- W. Detmold and **M. G. E.** (coming soon)

Motivation

- Critical slowing down
  - poor sampling of topology when a < 0.05 fm
  - physical pion masses are costly
- Generation of large physical volume lattices

C. Hoelbling (2014) [arXiv:1410.3403]



Markov Chain Monte Carlo

Nonperturbative studies of statistical and quantum systems often rely on Markov Chain Monte Carlo techniques to stochastically approximate path-integrals and observables, which define the system

$$Z = \sum_{s \in \Sigma} e^{-\mathcal{A}(s)} \qquad \langle \mathcal{O} \rangle = \sum_{s \in \Sigma} \mathcal{P}(s) \mathcal{O}(s) \qquad \mathcal{P}(s) = \frac{e^{-\mathcal{A}(s)}}{Z}$$

$$\langle \mathcal{O} \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(s_i) + \mathrm{O}\left(N^{-1/2}\right) \qquad s_i \text{ drawn from } \mathcal{P}(s) = \frac{e^{-\mathcal{A}(s)}}{Z}$$

Markov Chain Monte Carlo

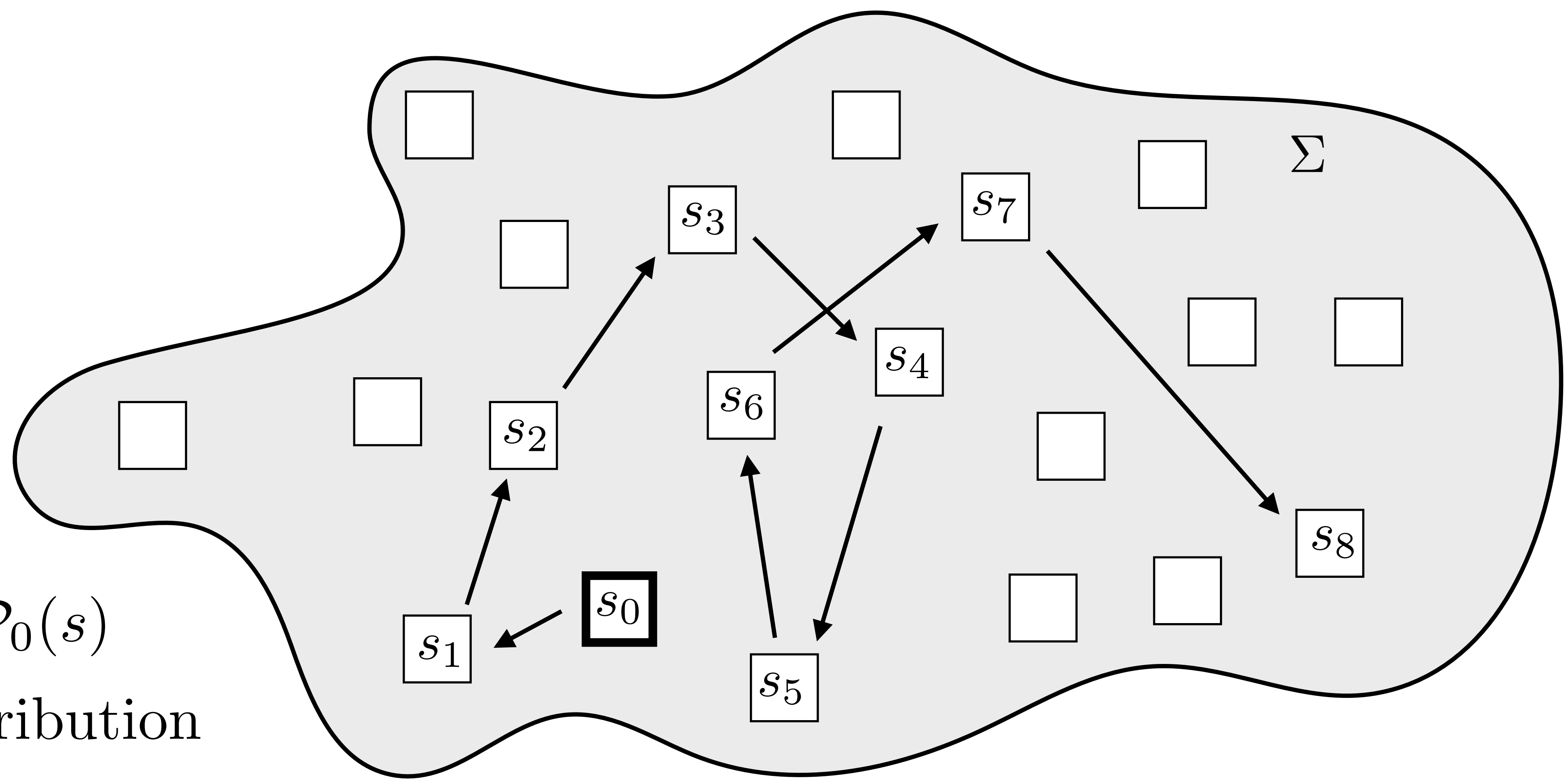
Generation of s_i determined by a transition probability $\mathcal{M}(s', s)$ for $s \to s'$

$$\sum_{s \in \Sigma} \mathcal{M}(s', s)\mathcal{P}(s) = \mathcal{P}(s')$$

(stationary distribution)

$$\sum_{s' \in \Sigma} \mathcal{M}(s', s) = 1$$

(probability conservation)



s_0 drawn from $\mathcal{P}_0(s)$

$\mathcal{P}_0$ a known distribution

Markov Chain Monte Carlo

Generation of s_i determined by a transition probability $\mathcal{M}(s', s)$ for $s \to s'$

$$\sum_{s \in \Sigma} \mathcal{M}(s', s)\mathcal{P}(s) = \mathcal{P}(s') \qquad \sum_{s' \in \Sigma} \mathcal{M}(s', s) = 1$$

Under suitable conditions (e.g., detailed balance, etc.)…

$$\mathcal{M}|\chi_n\rangle = e^{-1/\tau_n}|\chi_n\rangle \qquad \tau_n > 0 \text{ for all } n \geq 0$$

$$\langle\tilde{\chi}_m|\chi_n\rangle = \delta_{mn}$$

$$\langle s|\chi_n\rangle = \chi_n(s)$$

$$\langle\tilde{\chi}_n|s\rangle = \tilde{\chi}_n(s) = \chi_n(s)/\mathcal{P}(s)$$

Markov Chain Monte Carlo

Generation of s_i determined by a transition probability $\mathcal{M}(s', s)$ for $s \to s'$

$$\sum_{s \in \Sigma} \mathcal{M}(s', s) \mathcal{P}(s) = \mathcal{P}(s') \qquad \sum_{s' \in \Sigma} \mathcal{M}(s', s) = 1$$

Under suitable conditions (e.g., detailed balance, etc.)…

$$\mathcal{M}|\chi_n\rangle = e^{-1/\tau_n}|\chi_n\rangle \qquad \tau_n > 0 \text{ for all } n \geq 0$$

$$\langle\tilde{\chi}_m|\chi_n\rangle = \delta_{mn}$$

$$\langle s|\chi_n\rangle = \chi_n(s)$$

$$\langle\tilde{\chi}_n|s\rangle = \tilde{\chi}_n(s) = \chi_n(s)/\mathcal{P}(s)$$

$$\chi_0(s) = \mathcal{P}(s)$$

$$\tilde{\chi}_0(s) = 1$$

$$\tau_0 = \infty$$

Markov Chain Monte Carlo — evolution

Spectral decomposition:

$$\mathcal{M} = \sum_{n\geq 0} e^{-1/\tau_n} |\chi_n\rangle\langle\tilde{\chi}_n| \quad \Rightarrow \quad \mathcal{M}^\tau = \sum_{n\geq 0} e^{-\tau/\tau_n} |\chi_n\rangle\langle\tilde{\chi}_n|$$

Evolution of probability distribution and expectation values:

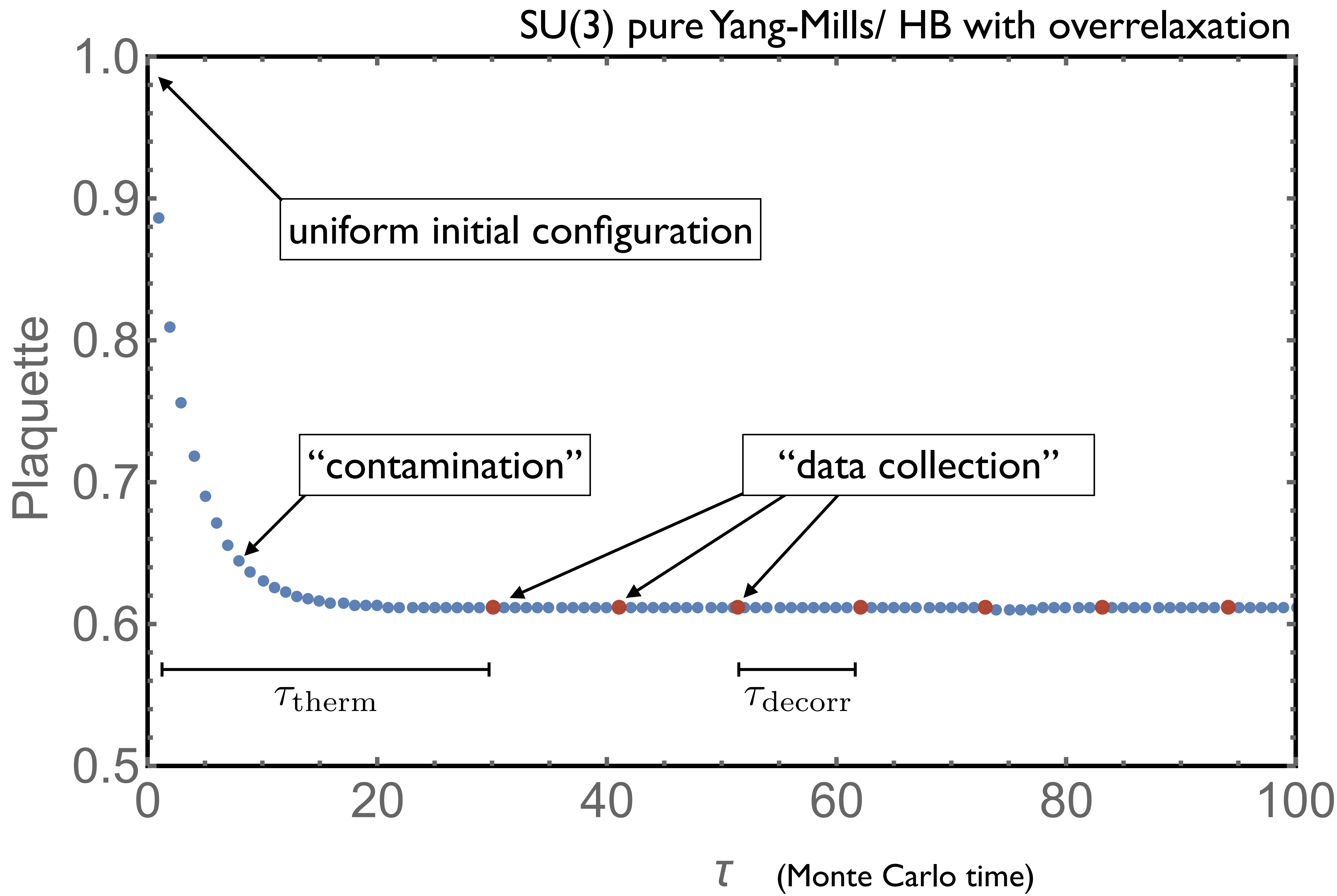
$$\mathcal{P}_\tau(s) = \langle s|\mathcal{M}^\tau|\mathcal{P}_0\rangle \; = \mathcal{P}(s) + \sum_{n>0} \langle s|\chi_n\rangle \langle\tilde{\chi}_n|\mathcal{P}_0\rangle e^{-\tau/\tau_n}$$

$$\langle\mathcal{O}\rangle_\tau = \langle\mathcal{O}|\mathcal{M}^\tau|\mathcal{P}_0\rangle \; = \langle\mathcal{O}\rangle + \sum_{n>0} \langle\mathcal{O}|\chi_n\rangle \langle\tilde{\chi}_n|\mathcal{P}_0\rangle e^{-\tau/\tau_n}$$

evolution time scales

overlap of initial distribution $P_0(s)$ onto mode n

Markov Chain Monte Carlo — equilibration

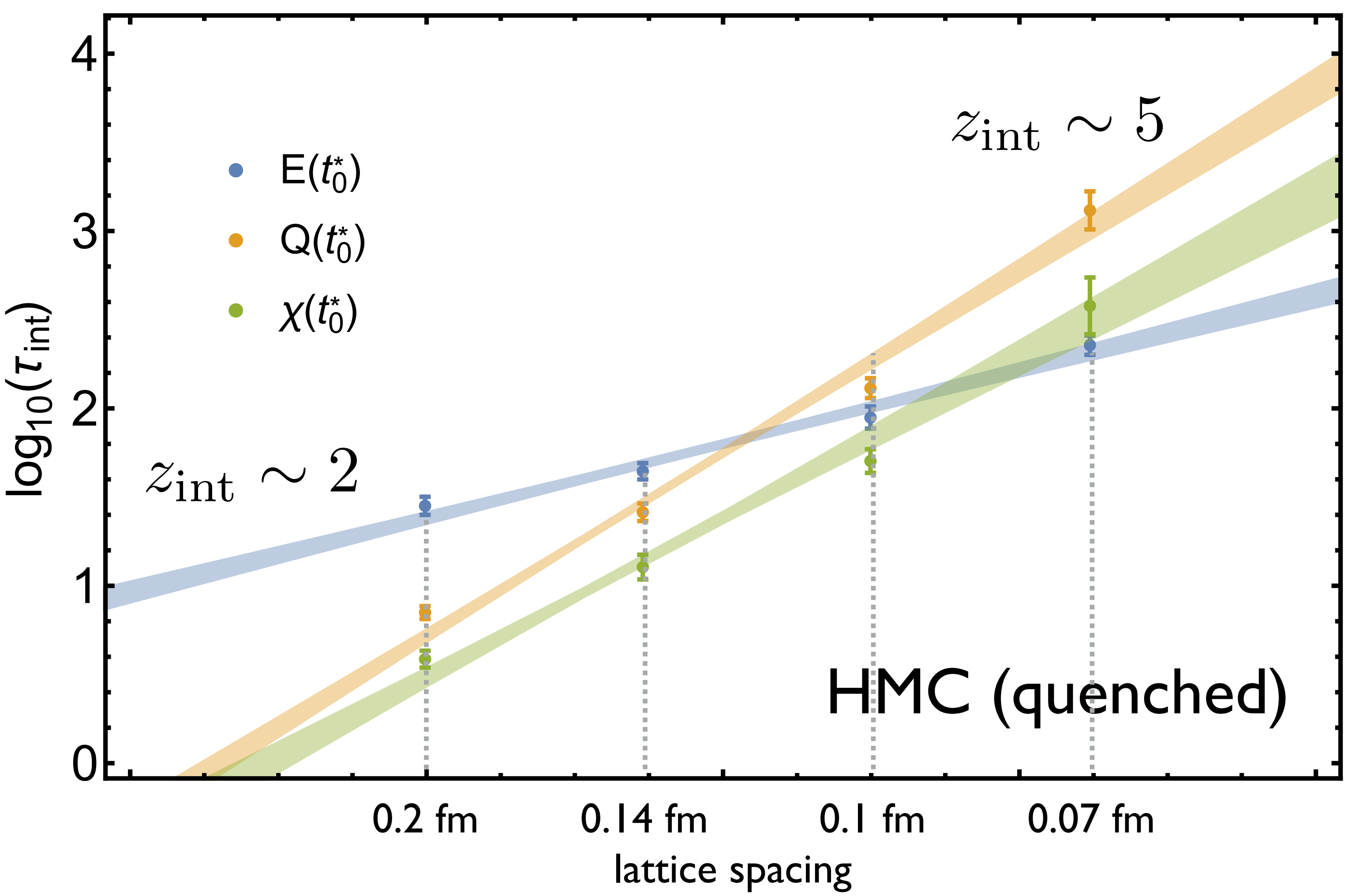


Markov Chain Monte Carlo — critical slowing down

Fine lattices decorrelate slower than coarse lattices

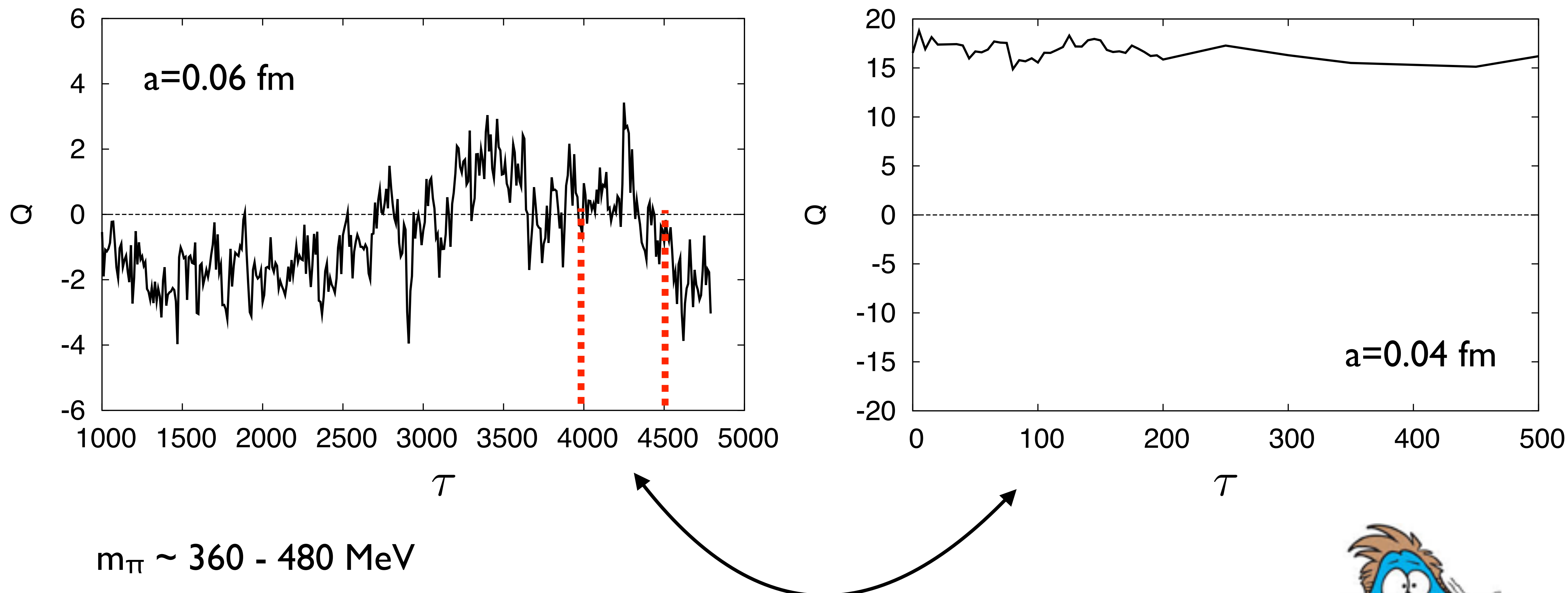
Integrated autocorrelation time

$$\tau_{\rm int}(\mathcal{O}) \sim \left(\frac{1}{a}\right)^{z_{\rm int}(\mathcal{O})} \Rightarrow \quad cost \sim \left(\frac{1}{a}\right)^{D+\max_{\mathcal{O}} z_{\rm int}(\mathcal{O})} \sim \left(\frac{1}{a}\right)^{9}$$



Markov Chain Monte Carlo — topological freezing

On a periodic lattice, topological charge fluctuations become *exponentially* suppressed in 1/a → Incorrect sampling



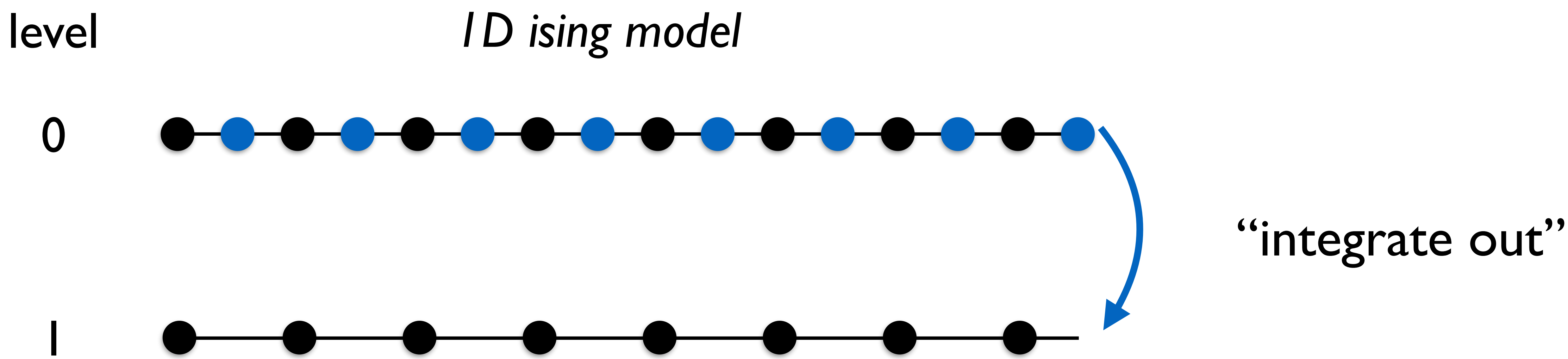
$m_\pi \sim 360$ - 480 MeV

S. Schaefer *et al.*, PoS **LAT2009** (2009) 032

Multiscale Monte Carlo methods

- **Ultimate goal:** an algorithm which allows for efficient updating of modes on multiple scales while retaining detailed balance
    - some progress for simple systems, but remains challenging for gauge theories (QCD in particular)
- **More modest goal of this work:** realization of a multiscale *thermalization* algorithm; the strategy draws upon many ideas:
    - standard Monte Carlo techniques
    - multigrid concepts of *restriction* (coarse-graining) and *prolongation* (refinement)
    - real space renormalization

A multiscale updating algorithm: Ising spin chain



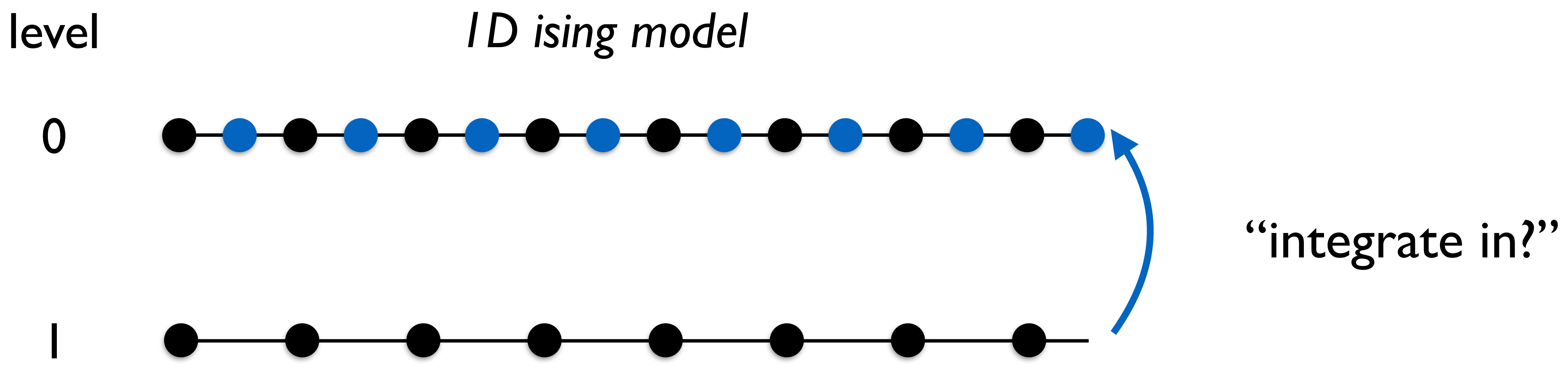
$$Z^{[\ell]} = \sum_{\{S\}} e^{-H^{[\ell]}}$$

$$H^{[0]} = J \sum_i S_{2i}(S_{2i+1} + S_{2i-1})$$

$$H^{[1]} = R(J) \sum_i S_{2i+1} S_{2i-1}$$

$$R(J) = \frac{1}{2} \cosh^{-1}\left(e^{2J}\right)$$

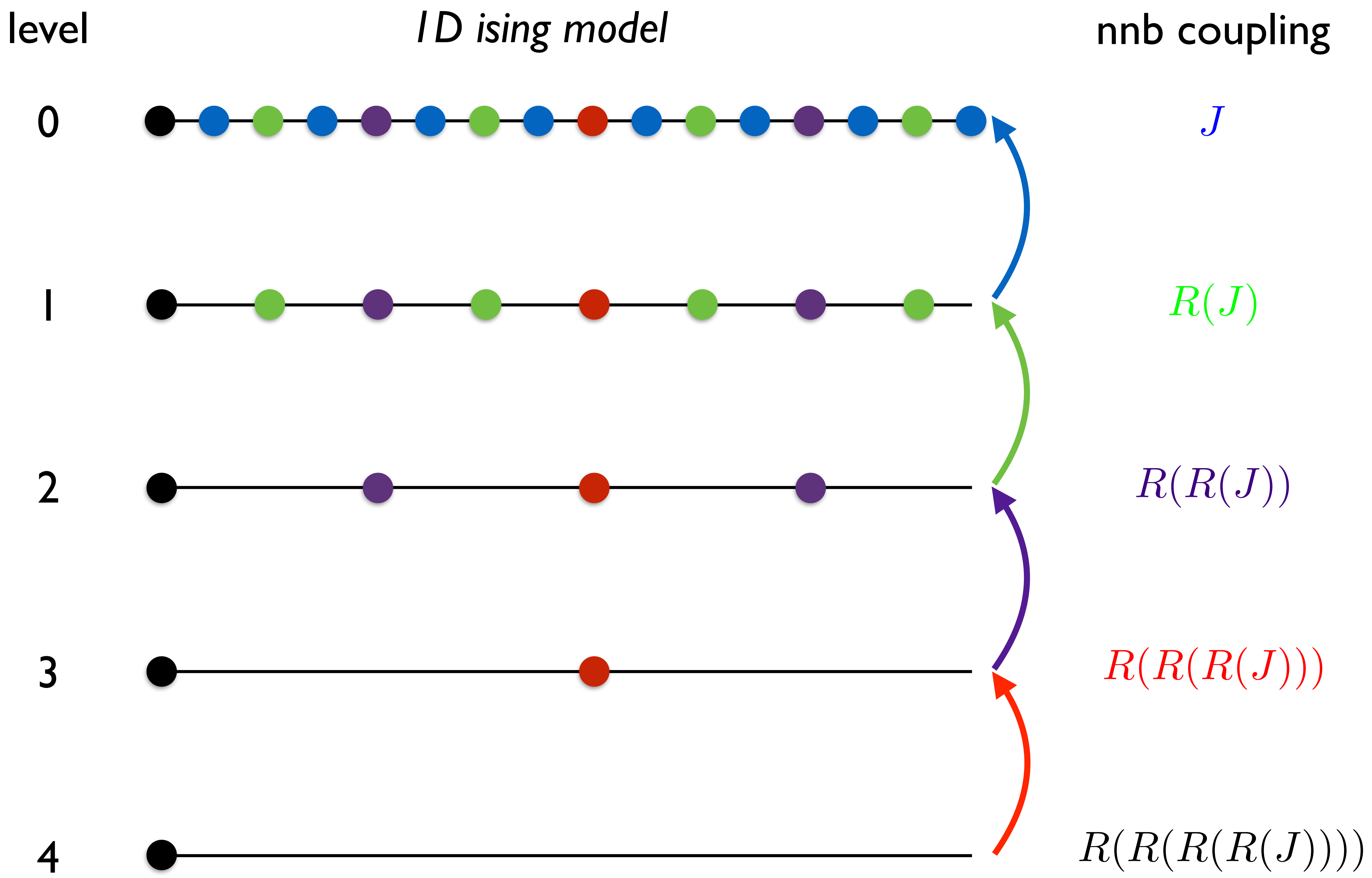
A multiscale updating algorithm: Ising spin chain



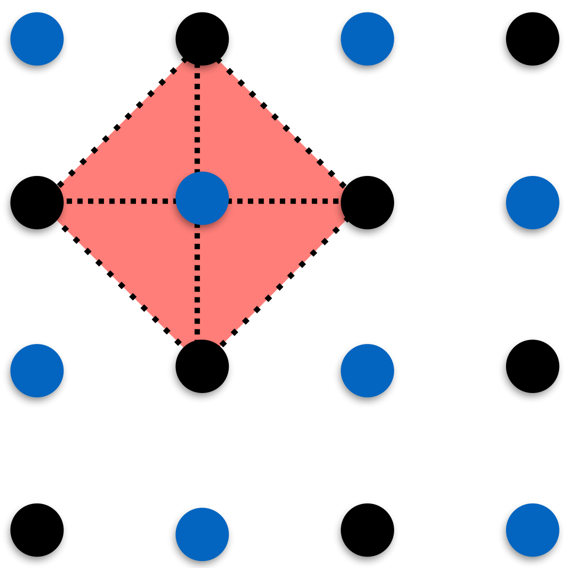
"Integrating in" spins at the fine level (0) requires a single "heat bath" update per (undefined) site:

$$\mathcal{P}(S_{2i}) = \frac{e^{-JS_{2i}(S_{2i+1}+S_{2i-1})}}{\cosh\left(J(S_{2i+1}+S_{2i-1})\right)}$$

A multiscale updating algorithm: Ising spin chain



A multiscale updating algorithm: Ising spin chain



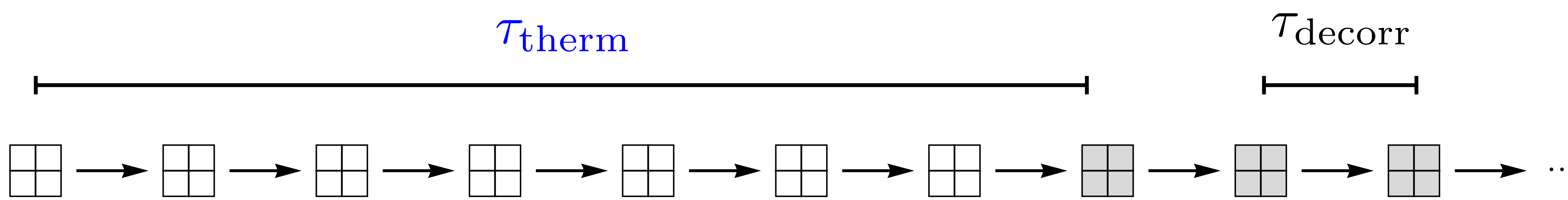
Generalization is nontrivial:

- Higher dimensions:
  - coarse graining induces increasingly complicated interactions
  - “integrating in” at coarse levels cannot be achieved with a single heat bath update
- Gauge theories:
  - continuous variables associated with links
  - nonlocal actions due to fermion determinants

Generalization to more complicated systems

- Generalization is achieved with approximations:
  - truncation of the coarse action; implies inexact RG matching
  - one-to-one refinement prescription based on interpolation, rather than exact prescription
- *Rethermalization* is crucial in order to correct for the errors induced by such approximations
- Effectiveness/use of approach depends on several factors
  - time scales associated with the conventional algorithm
  - refinement prescription
  - RG matching

Time scales



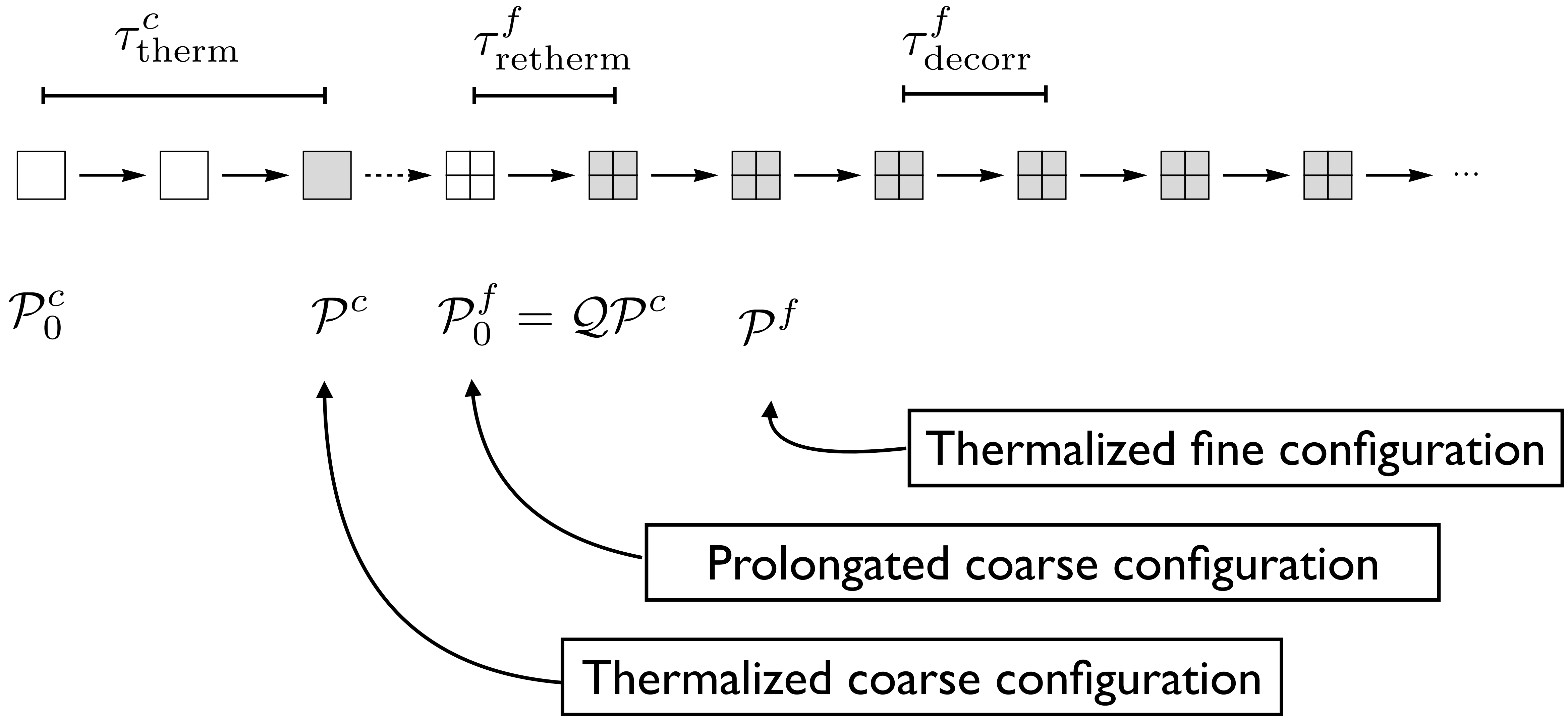
$$\mathcal{P}_\tau(s) = \mathcal{P}(s) + \sum_{n>0} \langle s|\chi_n\rangle \langle\tilde{\chi}_n|\mathcal{P}_0\rangle e^{-\tau/\tau_n}$$

$$\tau_{\text{decorr}} \lesssim 2\tau_1$$

$$\langle\mathcal{O}\rangle_\tau = \langle\mathcal{O}\rangle + \sum_{n>0} \langle\mathcal{O}|\chi_n\rangle \langle\tilde{\chi}_n|\mathcal{P}_0\rangle e^{-\tau/\tau_n}$$

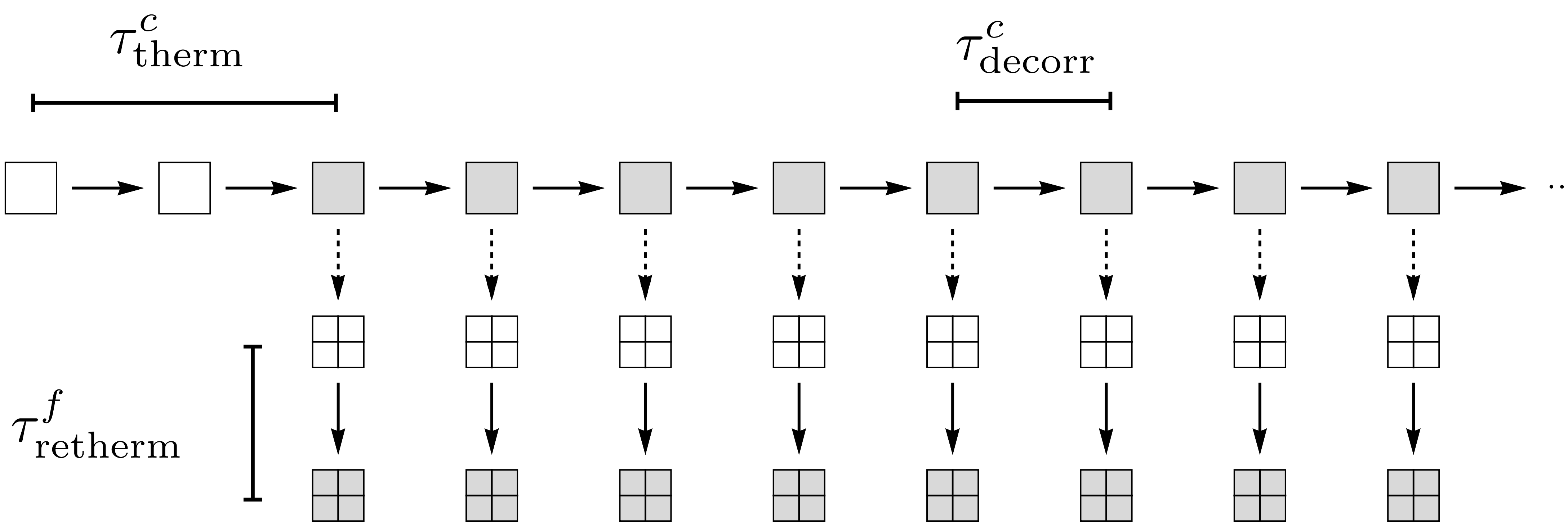
Goal: to find an initial probability distribution for which this overlap vanishes for small n

Time scales



Faster thermalization achieved if: $\tau^c_{\mathrm{therm}} + \tau^f_{\mathrm{retherm}} < \tau^f_{\mathrm{therm}}$

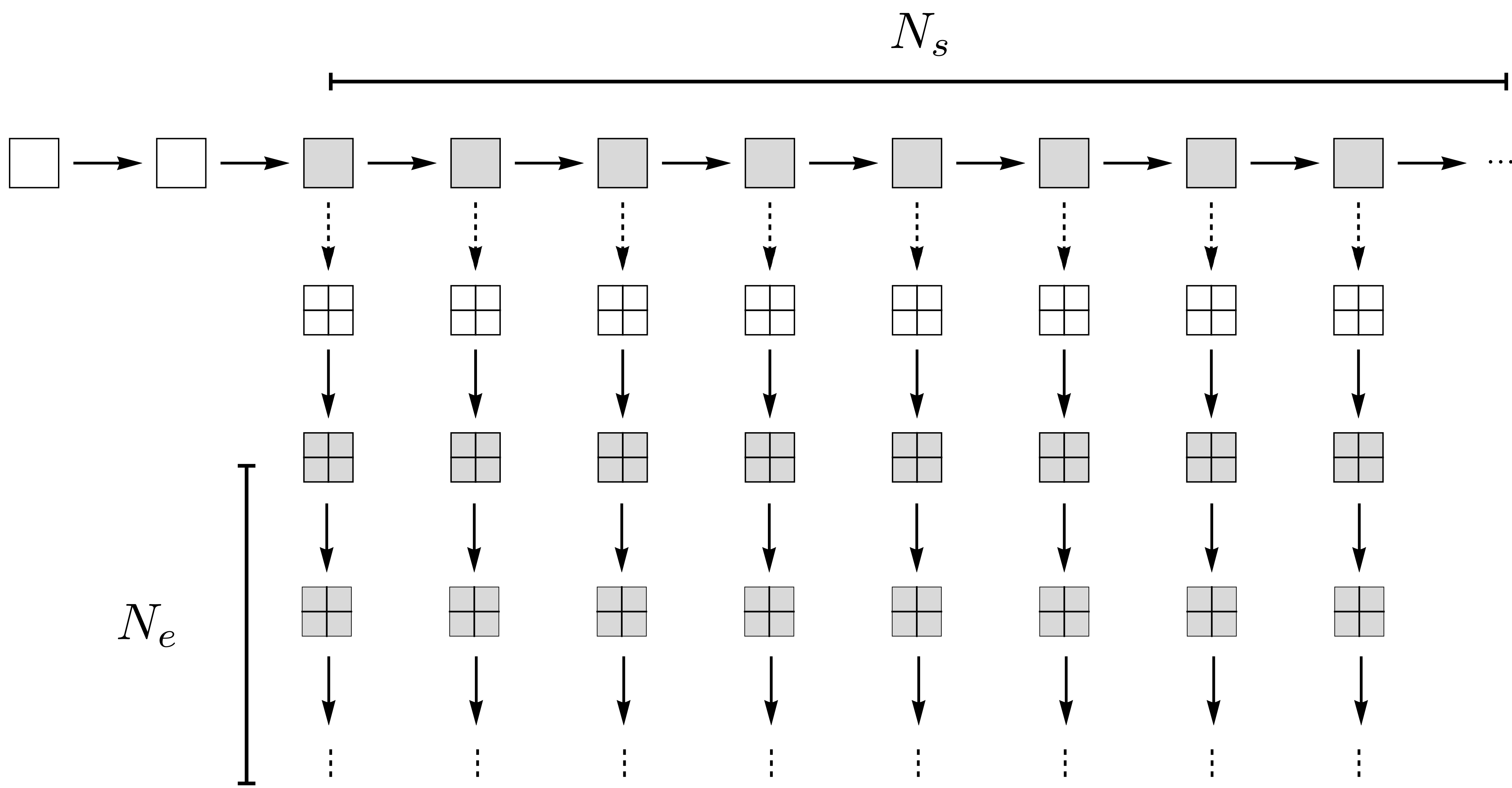
Time scales



Faster ensemble generation achieved if:

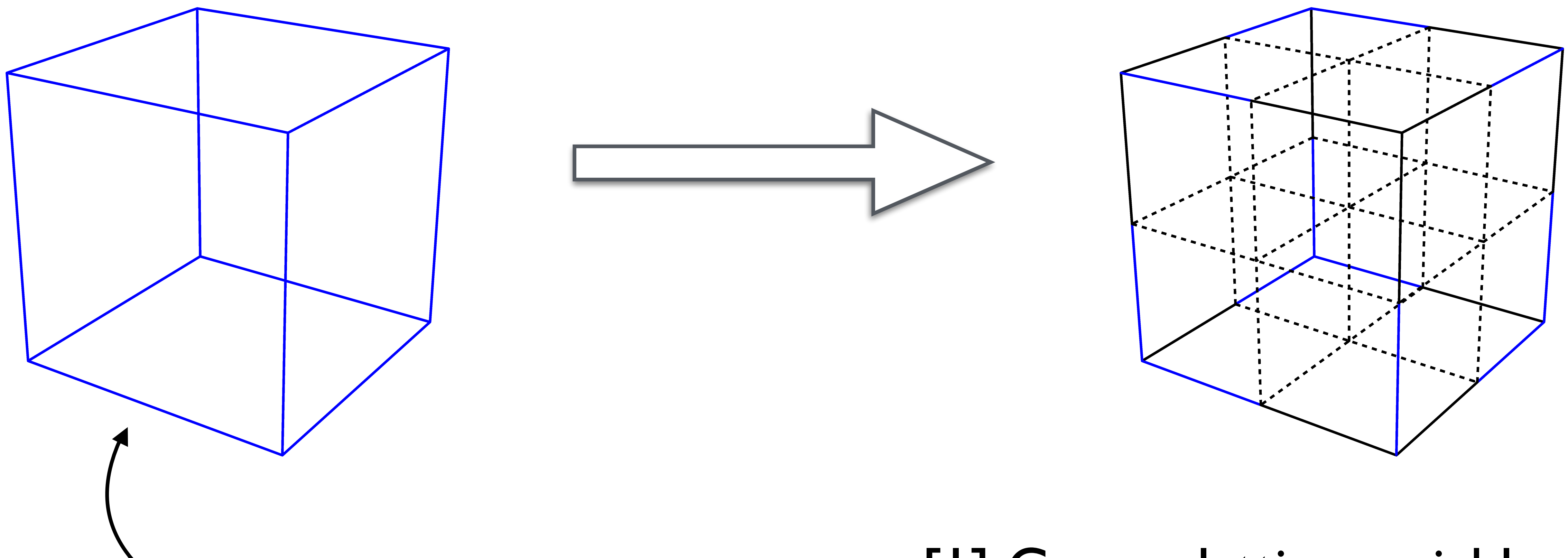
$$\tau^c_{\rm decorr} + \tau^f_{\rm retherm} \leq \tau^f_{\rm decorr}$$

Time scales



— more efficient use of computational resources
— greater statistical power due to *fully decorrelated streams*
— reduced critical slowing down; e.g., well sampled topology

Interpolation of gauge fields (à la ’t Hooft)



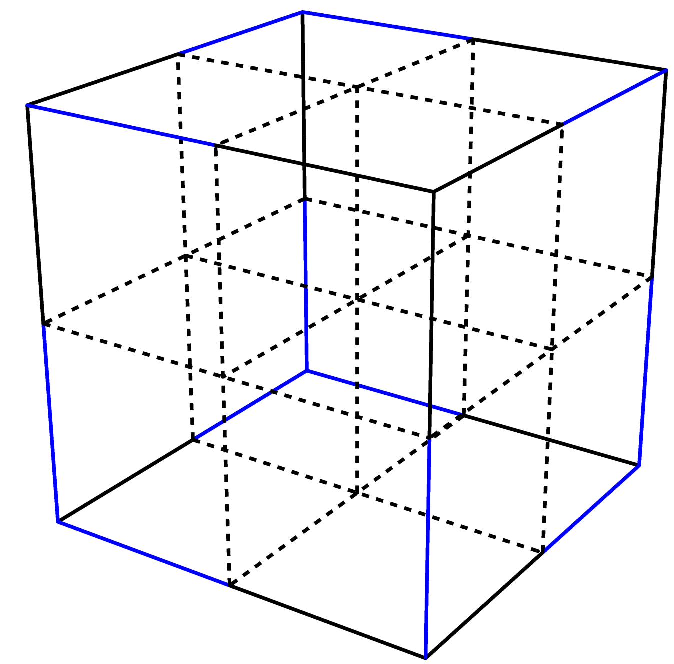
$$U_\mu(x) = e^{iaA_\mu(x)}$$

[1] Coarse lattice variables are transferred to the fine lattice

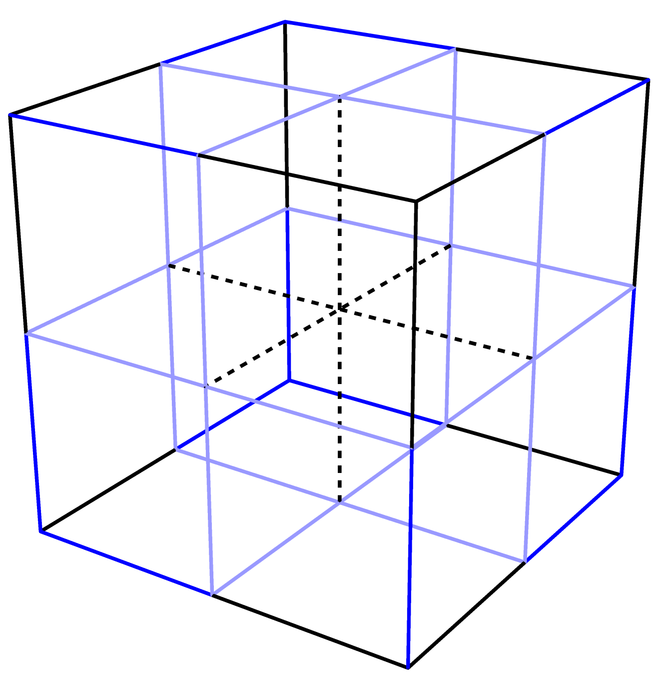
———— set to unity by a gauge choice

- - - - - undefined bond variables (set to unity)

Interpolation of gauge fields (à la ’t Hooft)



[2] Interior links are obtained by first minimization of action defined on 2x2 plaquettes

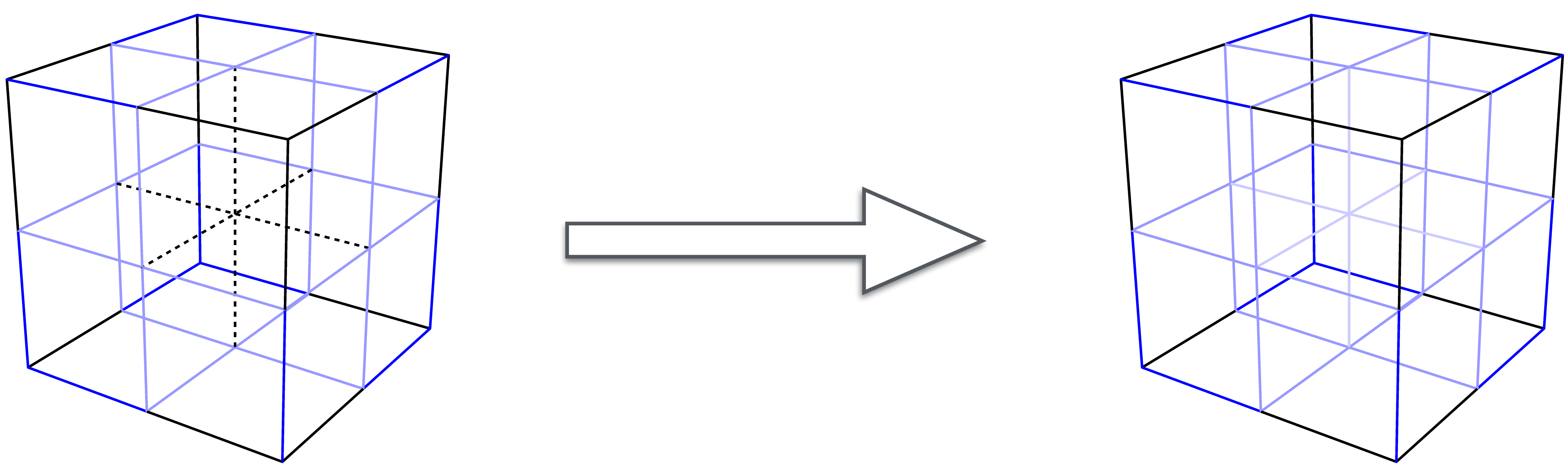


set to unity by a gauge choice

undefined bond variables
(set to unity)

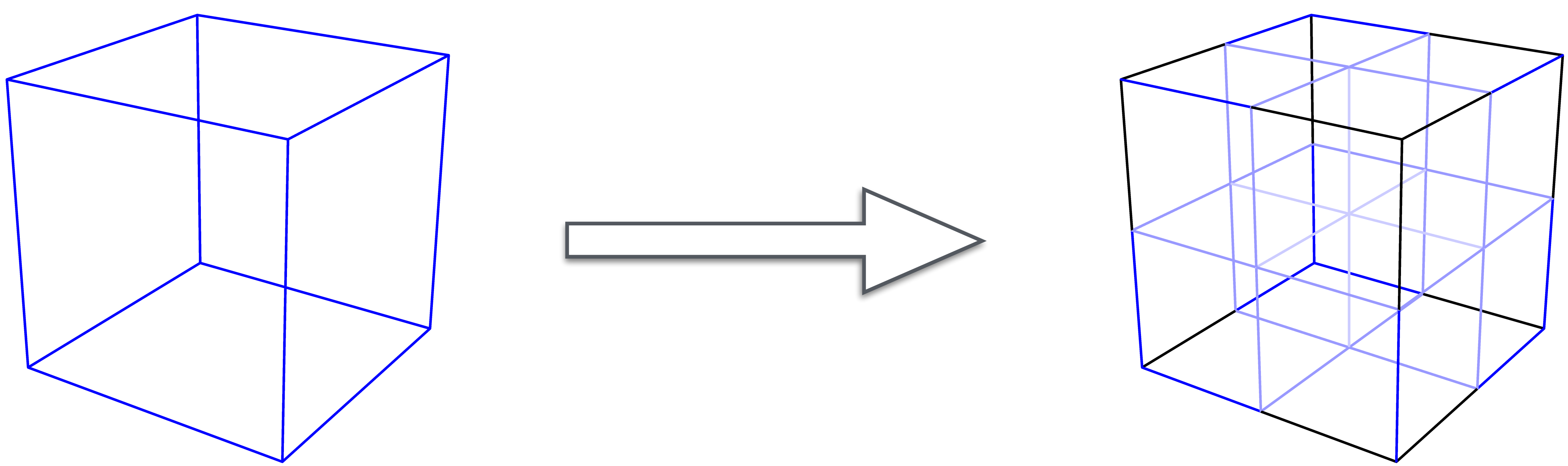
Interpolation of gauge fields (à la ’t Hooft)

[3] Minimization is repeated sequentially for interior cells



Properties of the interpolation

- — Implementation is simple and efficient
- — Can be performed locally
- — Preserves long distance properties of coarse configuration
  - — subset of even dimensional Wilson loops exactly
  - — topological charge at sufficiently fine lattice spacing
  - — discrete rotational invariance
- — Breaks a subset of discrete translational symmetry
  - — rapidly restored upon rethermalization



Numerical studies (three color Yang-Mills theory)

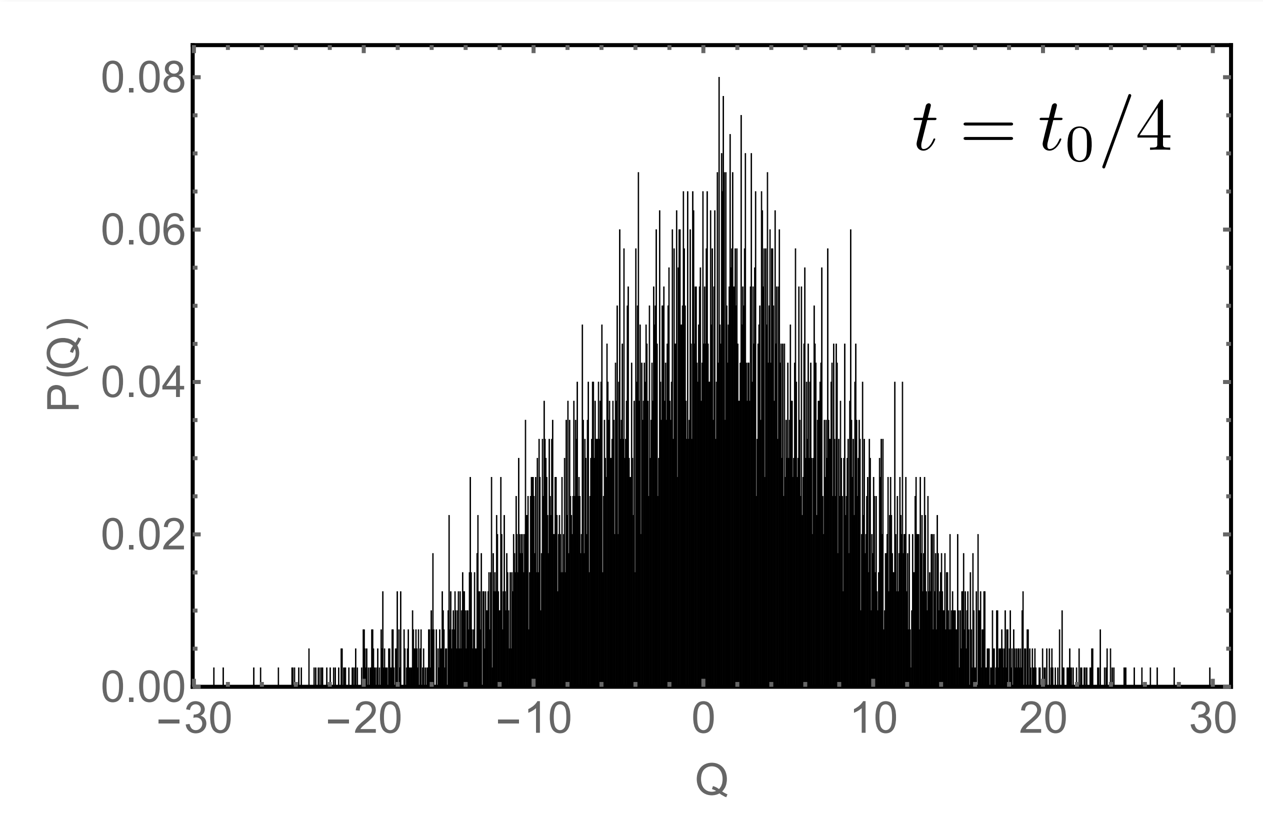
Lattice	β	a [fm]	N
$12^3 \times 24$	5.626	0.1995(20) fm	385
$16^3 \times 36$	5.78	0.1423(5) fm	385
$24^3 \times 48$	5.96	0.0999(4) fm	185
$32^3 \times 72$	6.17	0.0710(3) fm	185

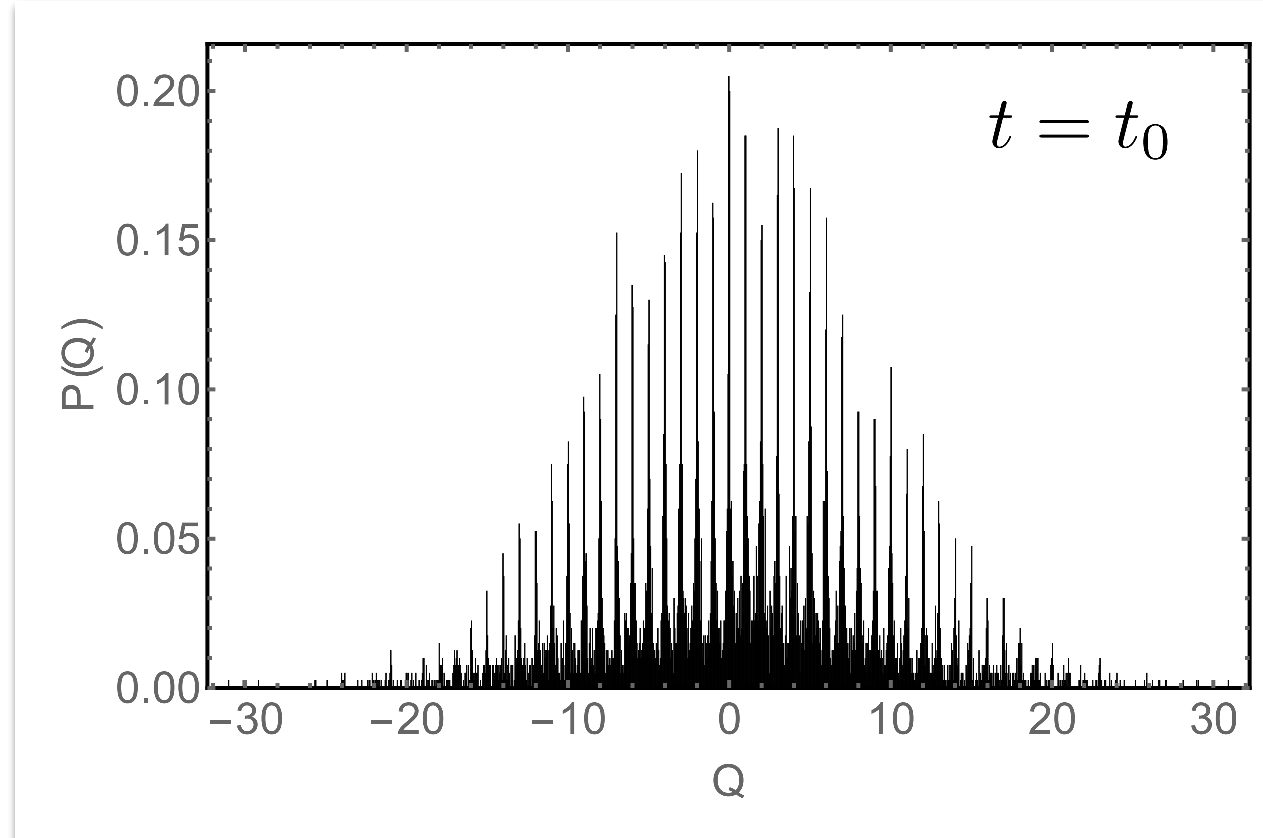
- Two *pairs* of RG matched ensembles (plaquette action)
- All ensembles correspond to a fixed physical volumes ~ 2.3 fm
- Studied long-distance observables such as large Wilson loops and various quantities under "Wilson flow" (diffusion)
    - e.g., powers of the topological charge (Q, χ=Q^2), action density (E)

Wilson flow

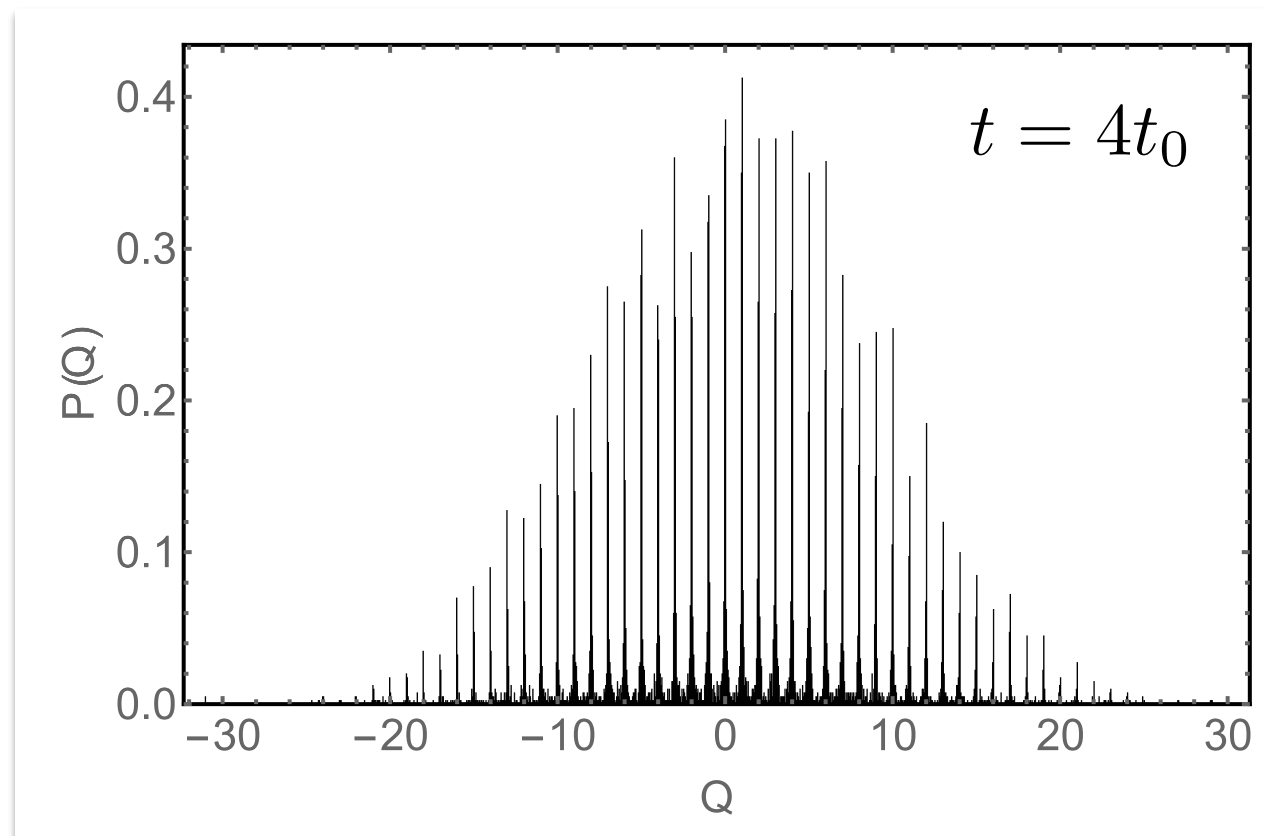
- Diffusion as a function of a fictitious fifth dimensional "flow time" t
- Local observables measures on flowed configurations probe distance scales $\sqrt{8t}$
- Can define a reference scale t_0 via
  - $t_0^2 E(t_0) = 0.3$ corresponding to $\sqrt{8t_0} \sim r_0 \sim 0.5$ fm
- Smoothing properties of Wilson flow are useful for measuring topological charge (e.g., via gluonic definition)
  - short distance fluctuations removed
  - charge approaches near-integer values

Topological charge under Wilson flow





$a \sim 0.1 \text{ fm}$

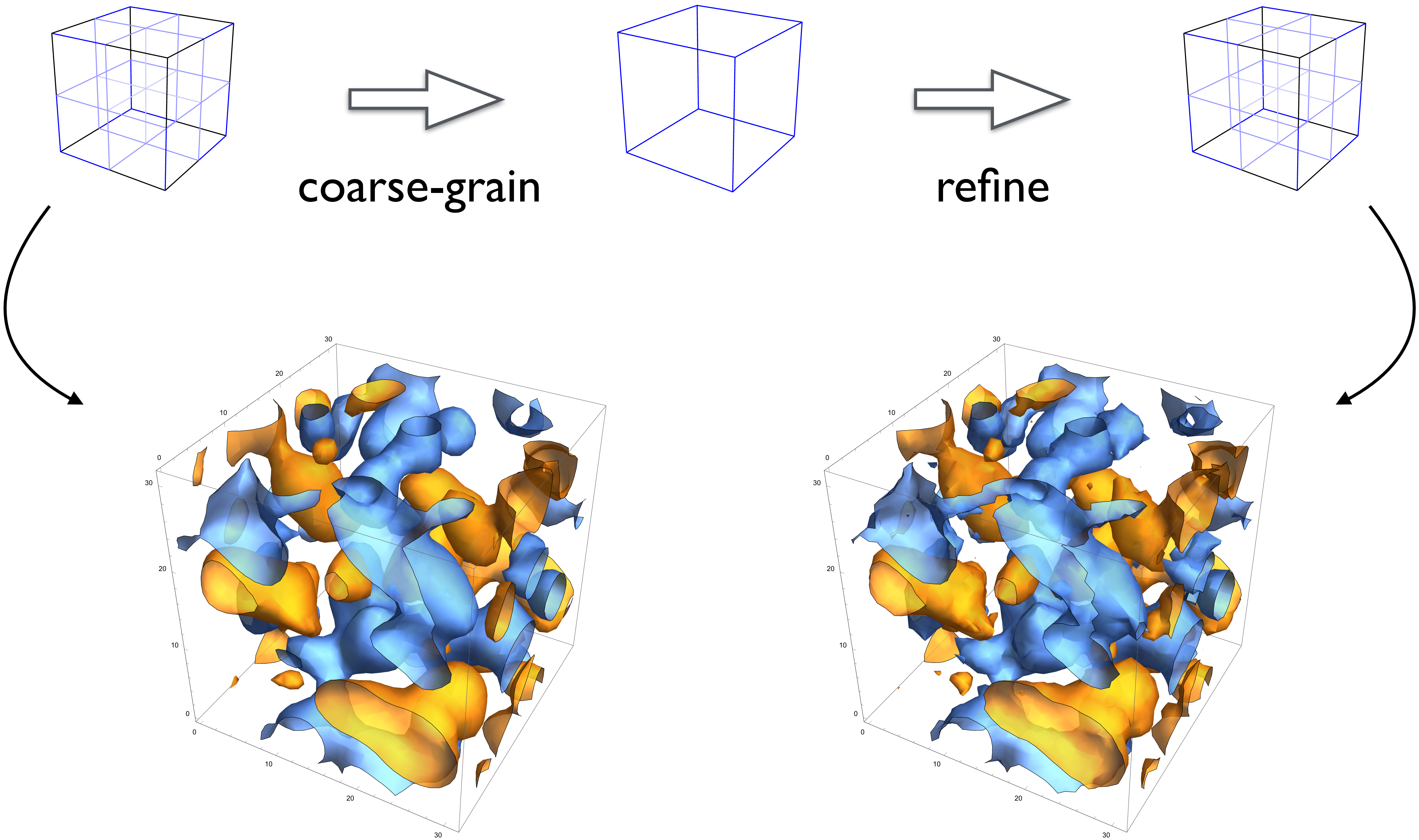


IMPORTANT NOTE!

t : "flow time" — smoothing of fields
τ : "Monte Carlo" — evolution

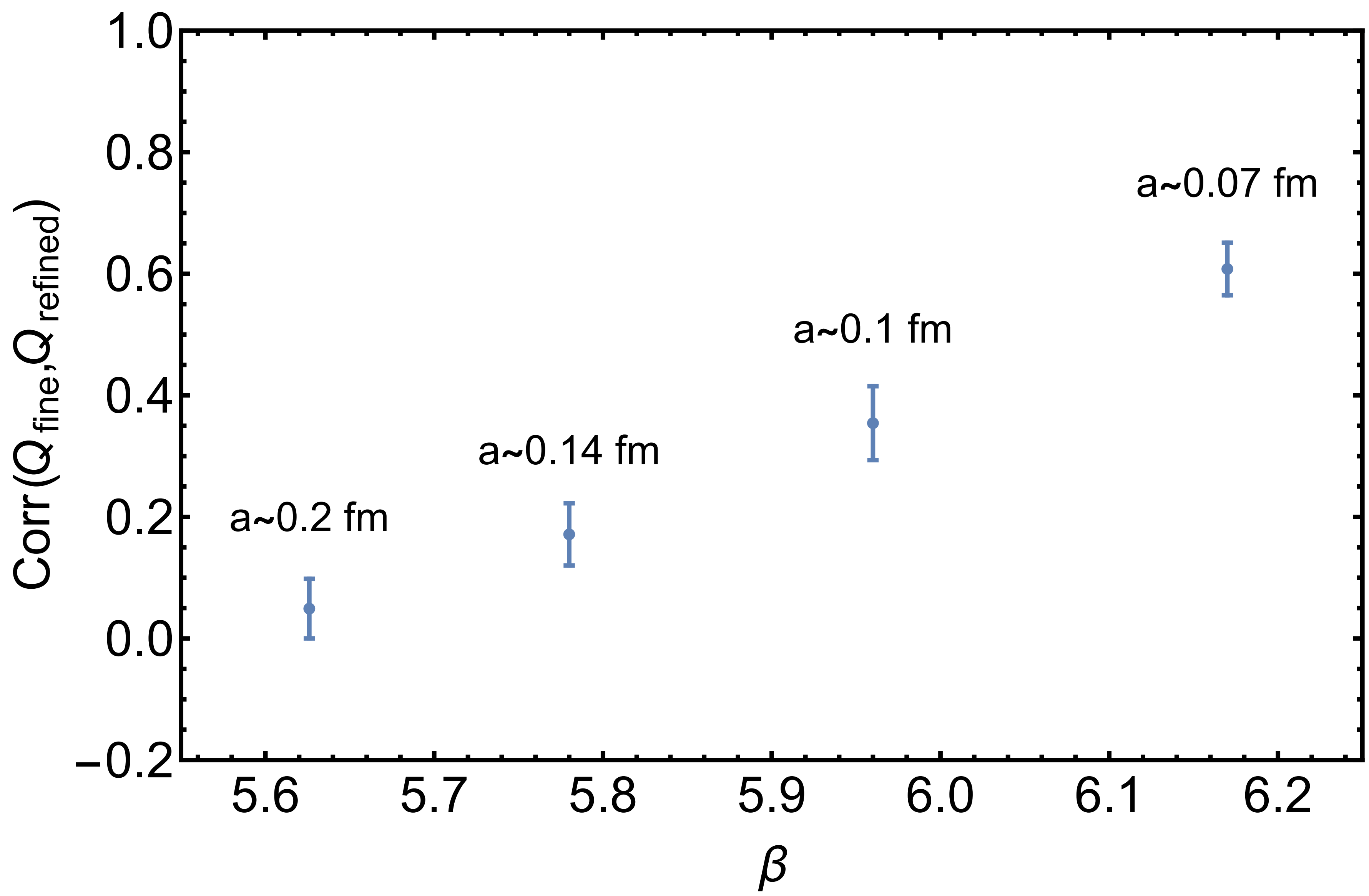
Neither of these correspond to physical time!

Interpolation — topological charge density



(single time slice)

Interpolation — topological charge

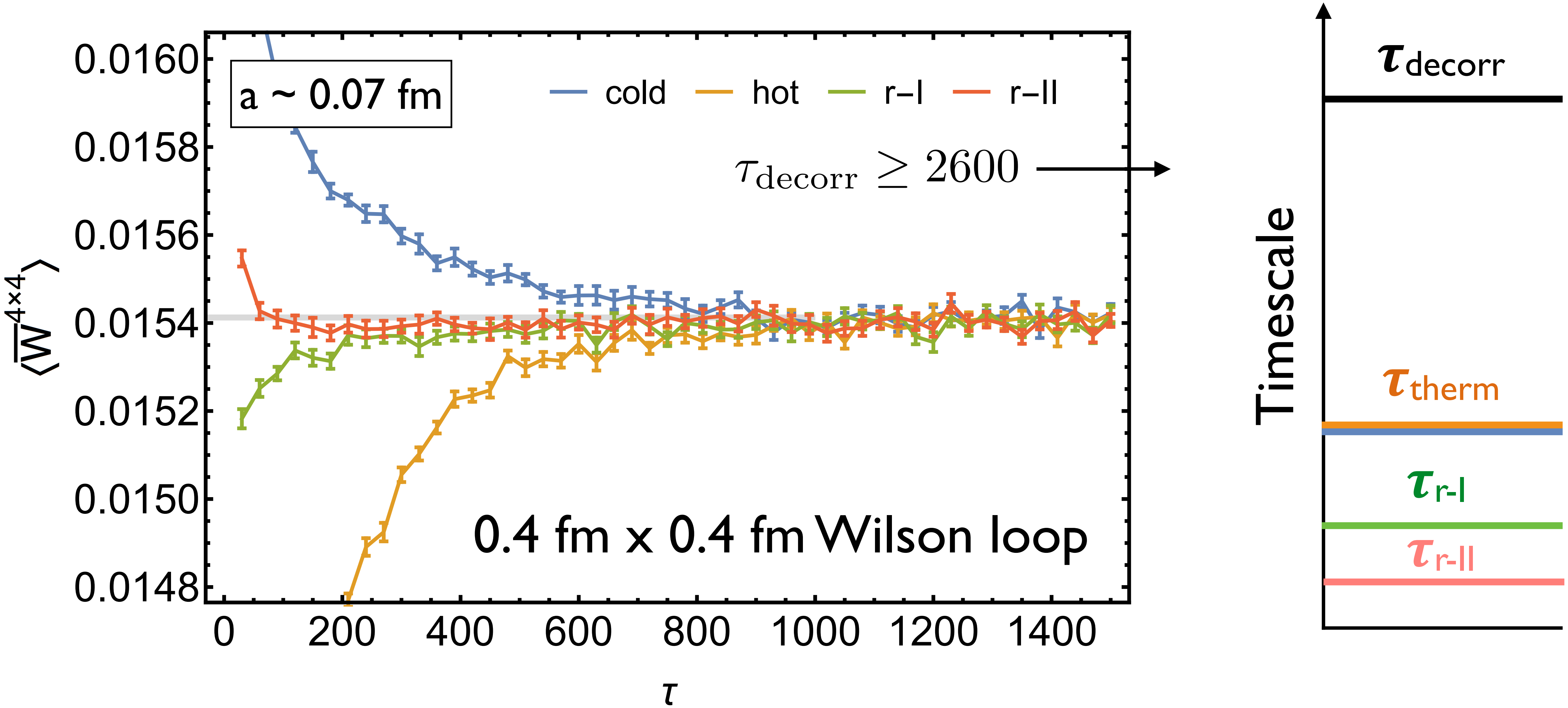


At sufficiently fine lattice spacing, topological charge of the coarse action is preserved configuration by configuration

Thermalization and rethermalization — HMC

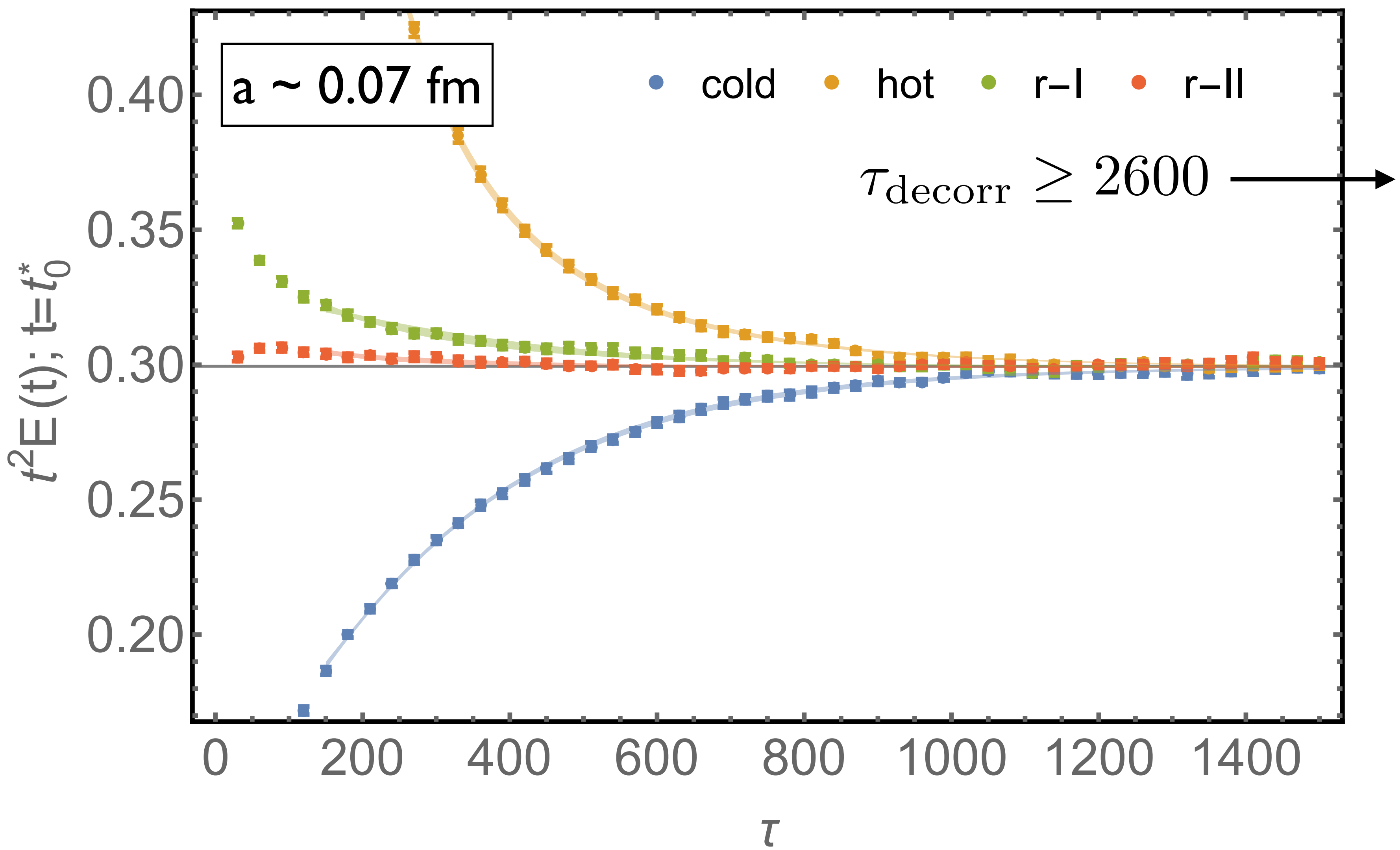
- Ensembles of size N_s=24
- Thermalization times probed by long distance observables measured at various Wilson flow times: $\chi(t)$, $E(t)$
- Thermalization considered for four ensembles:
  - disordered (hot)
  - ordered (cold)
  - restriction followed by prolongation of fine lattices (r-I)
  - prolongation of an RG matched coarse ensemble, generated using a Wilson action (r-II)

(Re)thermalization — Wilson loops



- Long-distance observables *rethermalize* on time scales shorter than
  - thermalization time for hot/cold starts (standard approach)
  - decorrelation time for fine evolution

(Re)thermalization — E(t_0)



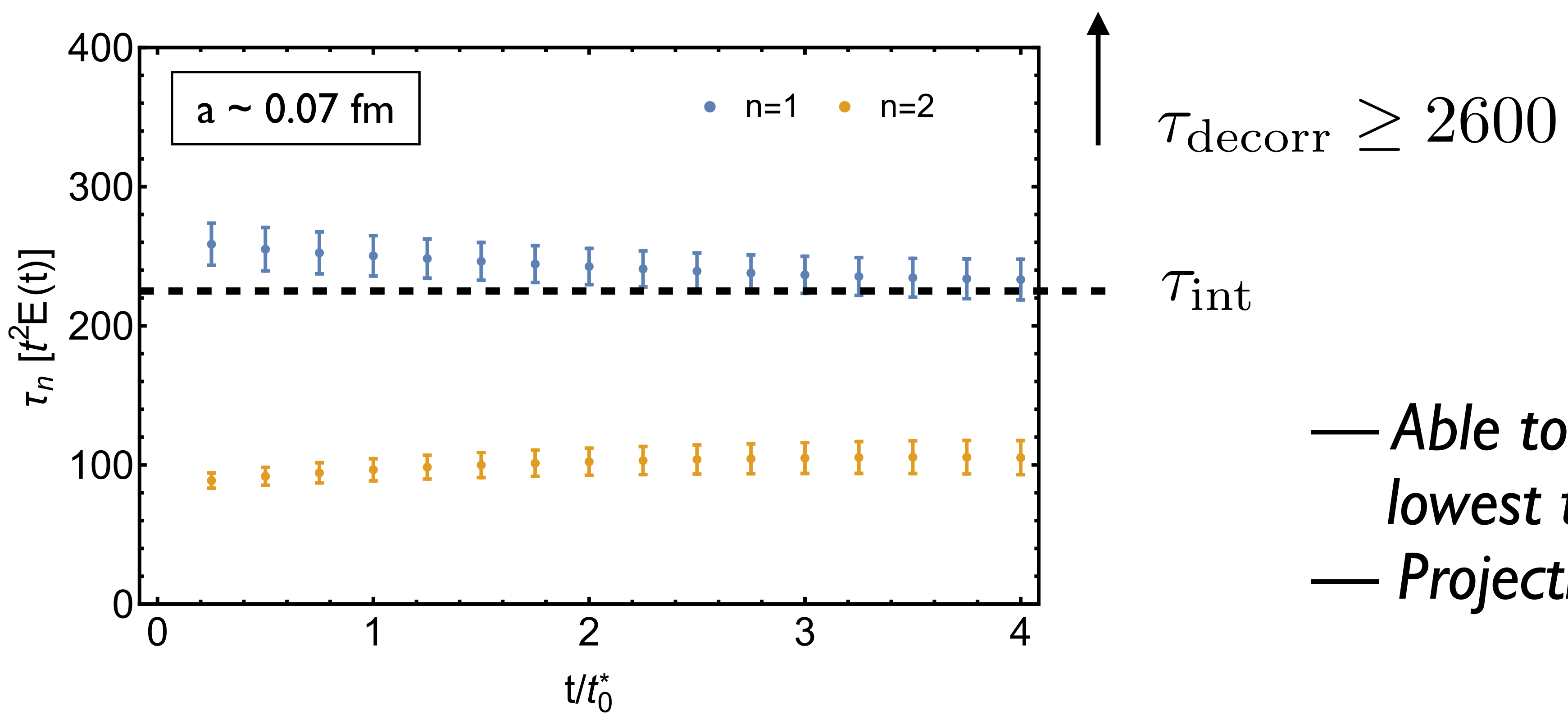
Least-squares fitting:

$$f^\alpha(\tau) = z_0 + z_1^\alpha e^{-\tau/\tau_1} + z_2^\alpha e^{-\tau/\tau_2}$$

$$z_0^\alpha = \langle \mathcal{O}^\alpha \rangle \qquad z_n^\alpha = \langle \mathcal{O}^\alpha | \chi_n \rangle \langle \tilde{\chi}_n | \mathcal{P}_0 \rangle$$

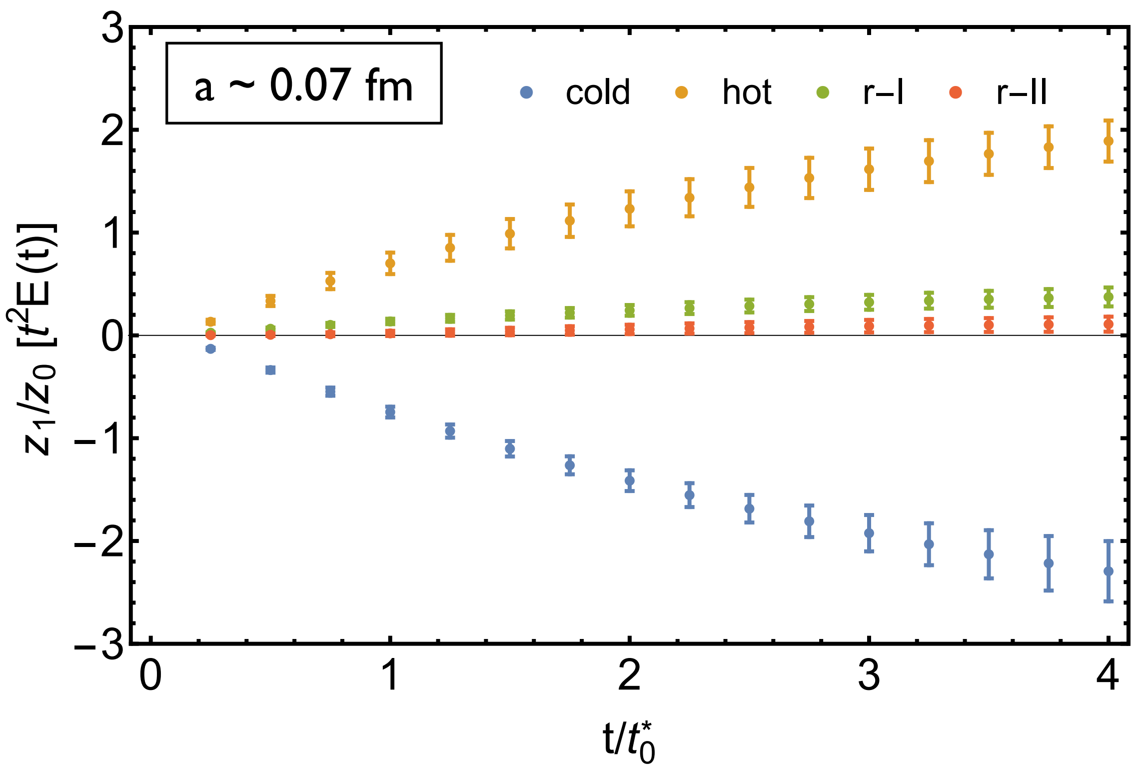
(different observables have common exponents)

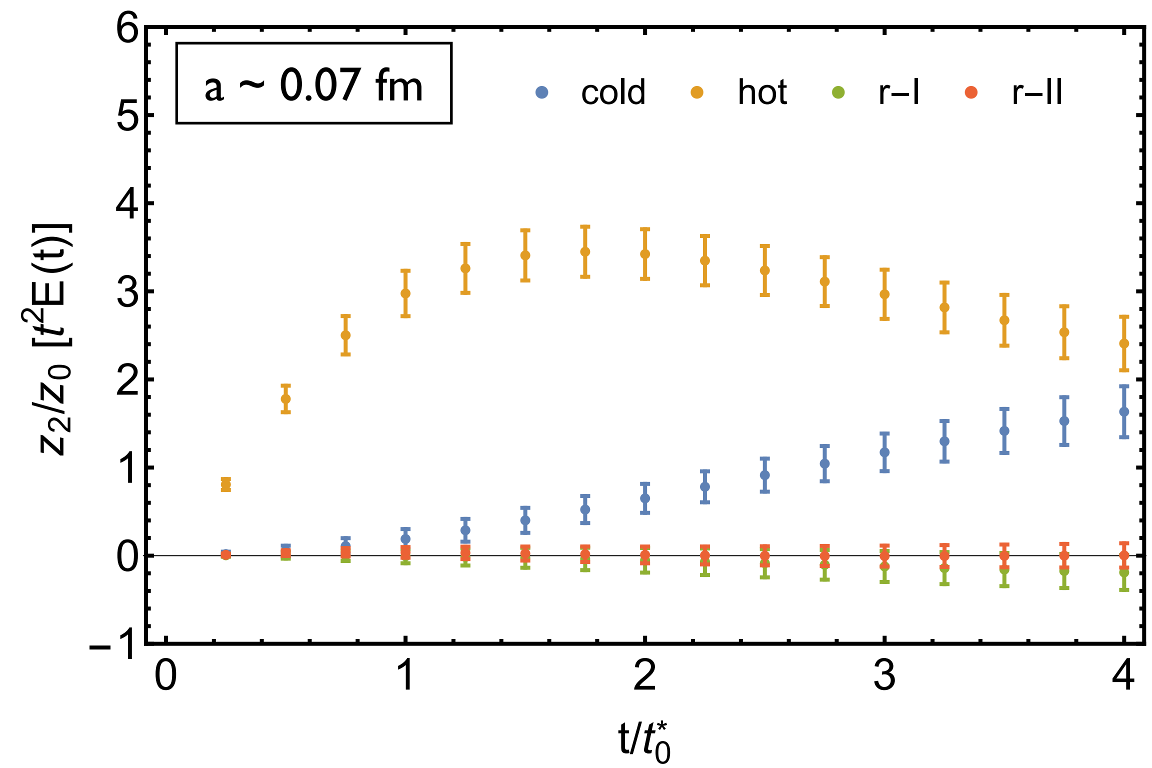
E(t) fit results as a function of flow time t/t_0



— *Able to remove at least the lowest two modes (n=1,2)*

— *Projection is independent of t_0*





Inclusion of fermions

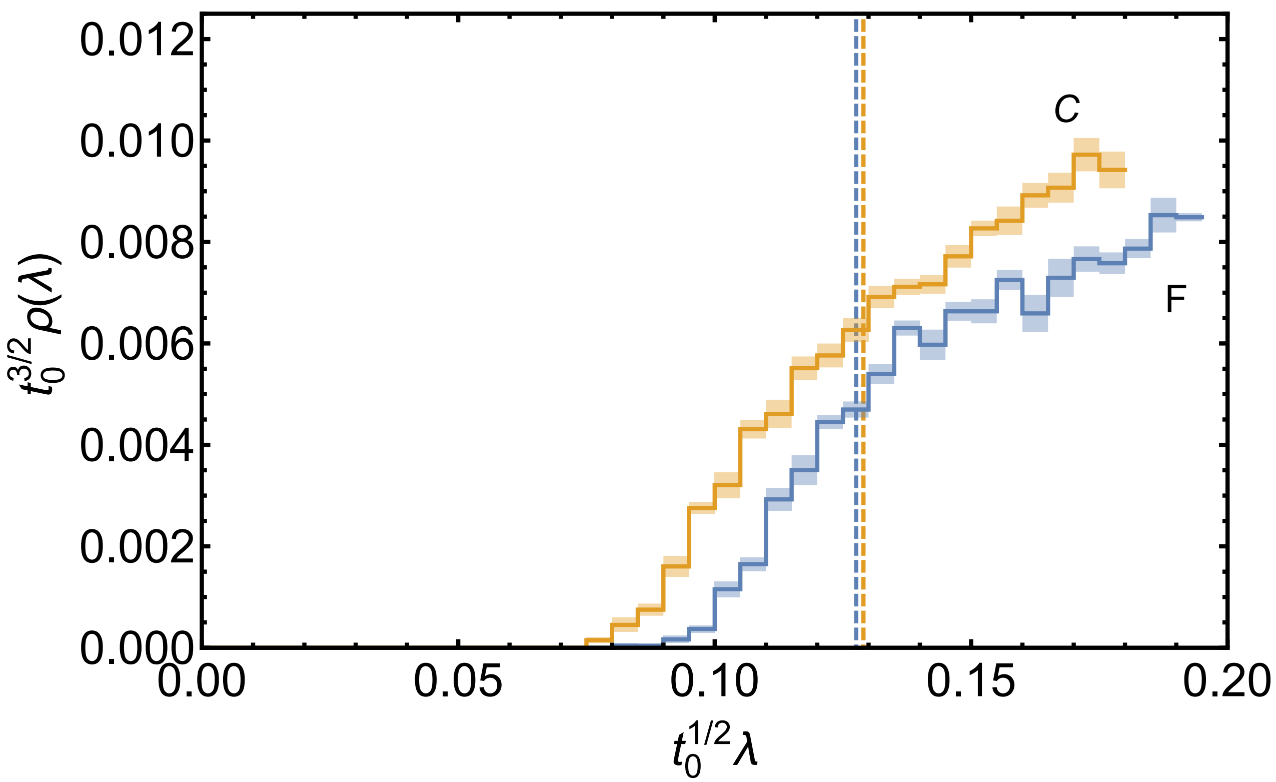
W. Detmold and **M. G. E.** (coming soon)

Dirac determinant

$$Z = \int [dA] \, e^{-S(A)} \Delta(A)$$

- Multiple scales (e.g., t_0, m_ρ, m_π, …)
- RG matching requires tuning of multiple coarse couplings (β, m_f, …)
- New challenges with interpolated configurations: spurious zero modes associated with the Dirac operator
  - initial HMC gauge evolution involves large fermion forces
  - numerical instabilities due to finite step size in evolution

Inclusion of fermions — Dirac spectrum



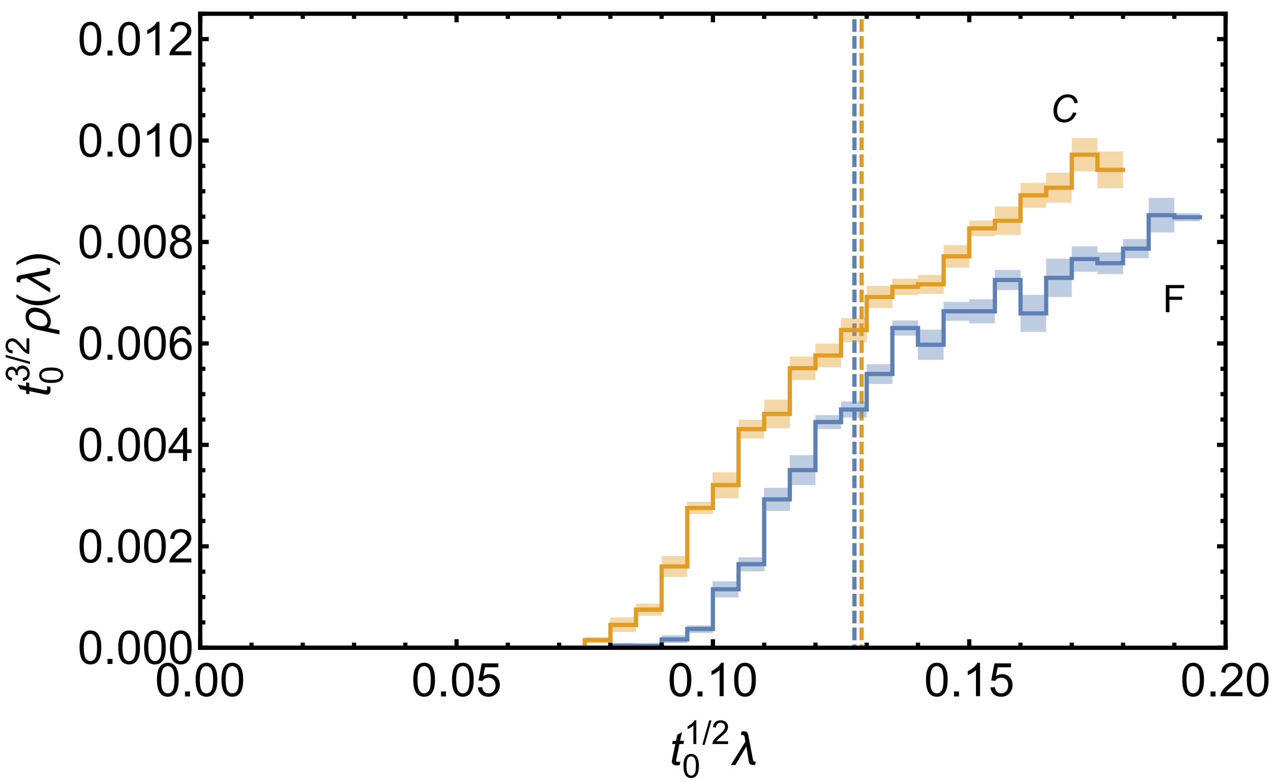
C : coarse ensemble
F : fine ensemble

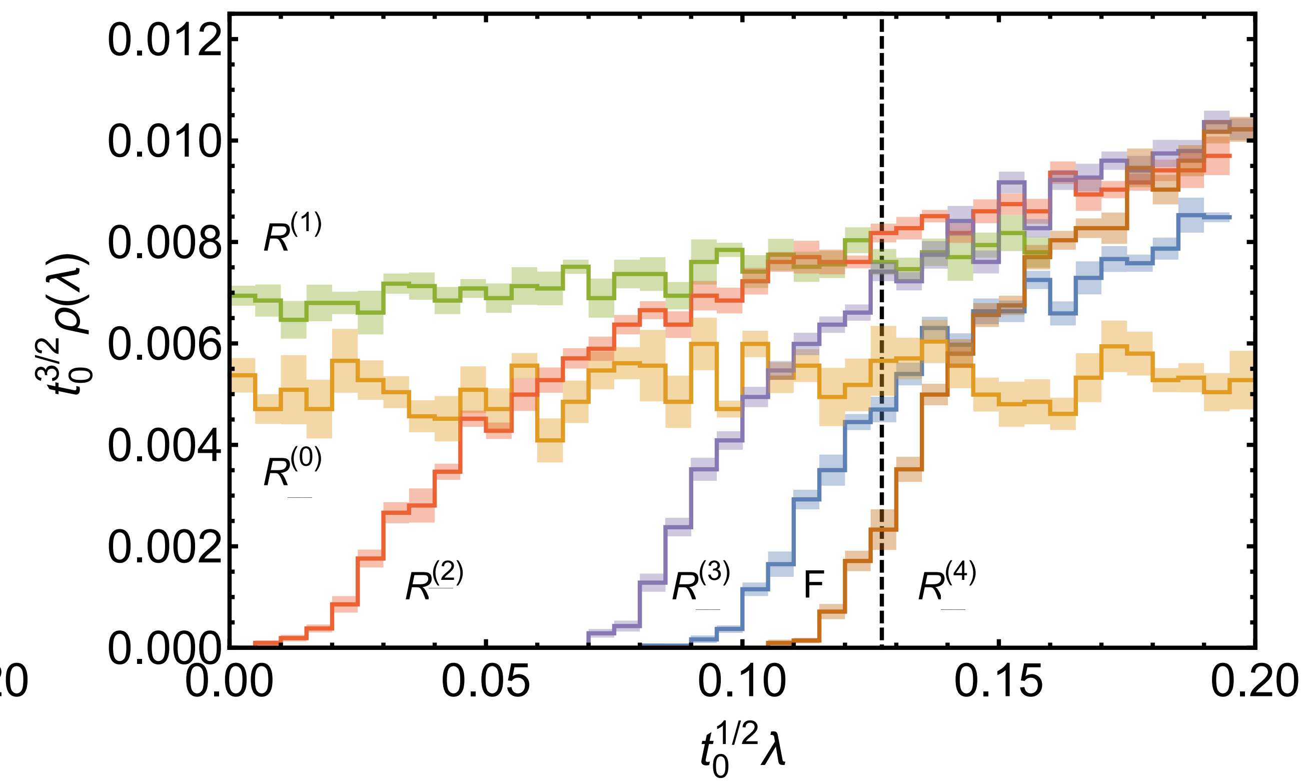
$$\lambda = \text{eigenvalues of } \gamma_\mu D_\mu + m$$

Two color QCD with two fermion flavors (isospin limit)

- Coarse (12^3x36) and fine (24^3x72) lattices
  - nonperturbatively matched actions using t_0 and m_π
- Heavy fermions, corresponding to $m_\pi/m_p \sim 0.85$
- Differences in Dirac spectrum presumably due to lattice artifacts

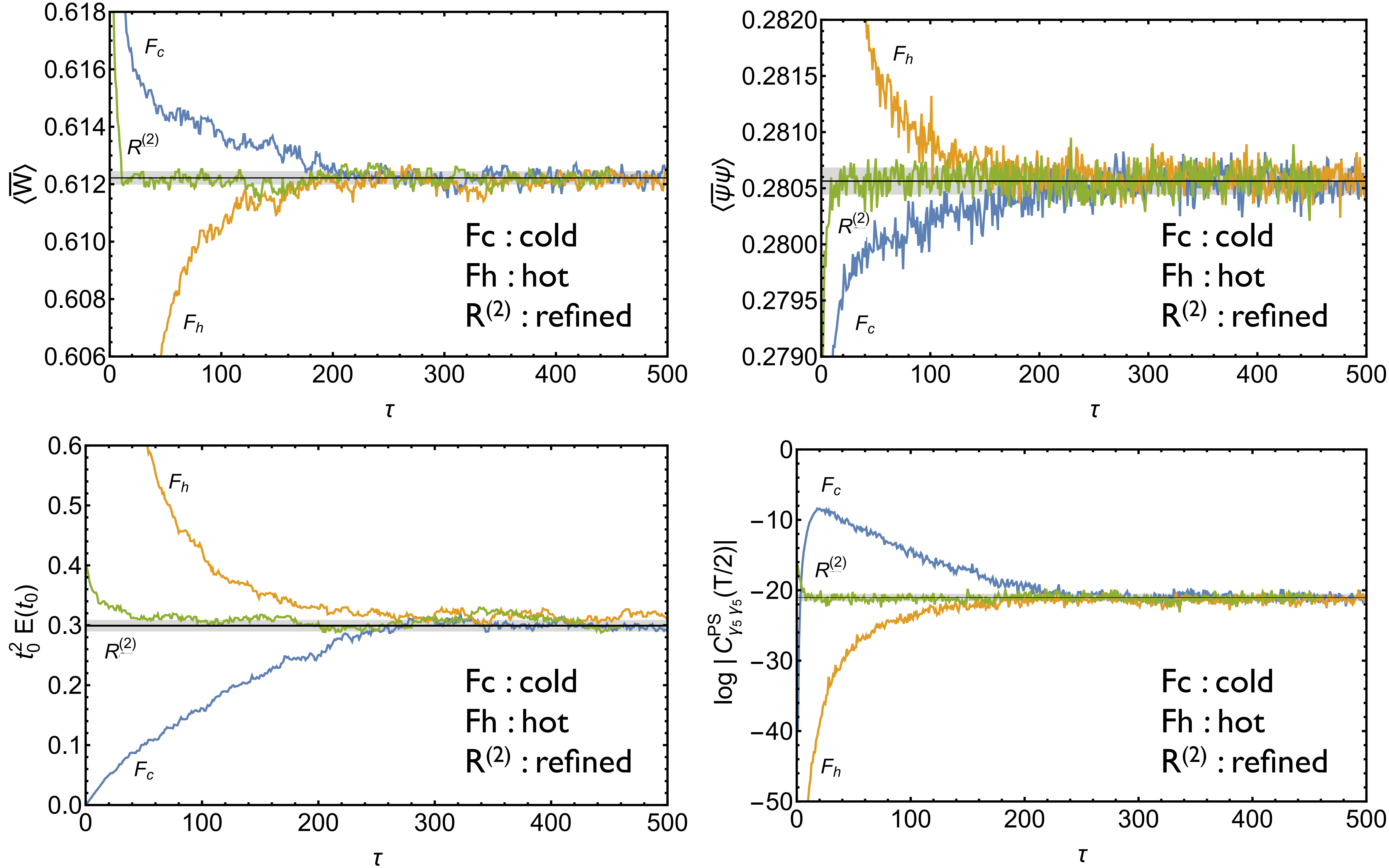
Inclusion of fermions — Dirac spectrum





- Refinement via interpolation: C → $R^{(0)}$
- Quenched evolution of interpolated fields: $R^{(0)} \to R^{(1)} \to R^{(2)} \to \ldots$
  - produces a gap in the Dirac spectrum, eliminates large initial forces
  - only impacts short distance features of ensemble if evolution is short

(Re)thermalization — various observables



Conclusion & Outlook

- Efficient multi-stream generation of uncorrelated gauge configurations
- Significantly reduces the problem of critical slowing down
  - enables numerical simulations at ultra-fine lattice spacings (a<0.05fm) with well-sampled topological charge
  - more efficient simulations expected at physical pion masses
- Alternatively, enables efficient numerical simulations at large volumes
- Method is general:
  - successfully applied to Hybrid Monte Carlo simulations of pure Yang-Mills theory and QCD_2 with dynamical fermions
  - systems beyond QCD (e.g., quantum Monte Carlo?)

Conclusion & Outlook

- Some remaining open issues:
  - efficient tuning methods for matching coarse and fine action
  - reliable methods for determining if ensemble is thermalized
  - removing inherited lattice artifacts in fine topological charge distribution
  - better understanding of spurious zero-modes
  - better methods for handling large initial fermion forces (e.g., evolution on multiple time scales)