MIT Open Access Articles

Multiscale Monte Carlo equilibration: Pure Yang-Mills theory

Massachusetts Institute of Technology

Multiscale Monte Carlo equilibration: Pure Yang-Mills theory

Michael G. Endres[*]

Center for Theoretical Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

Richard C. Brower[†]

Department of Physics, Boston University, Boston, Massachusetts 02215, USA

William Detmold[‡]

Center for Theoretical Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

Kostas Orginos[§]

Department of Physics, College of William and Mary, Williamsburg, Virginia 23187-8795, USA
and Jefferson Laboratory, 12000 Jefferson Avenue, Newport News, Virginia 23606, USA

Andrew V. Pochinsky[‖]

Center for Theoretical Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

We present a multiscale thermalization algorithm for lattice gauge theory, which enables efficient parallel generation of uncorrelated gauge field configurations. The algorithm combines standard Monte Carlo techniques with ideas drawn from real space renormalization group and multigrid methods. We demonstrate the viability of the algorithm for pure Yang-Mills gauge theory for both heat bath and hybrid Monte Carlo evolution, and show that it ameliorates the problem of topological freezing up to controllable lattice spacing artifacts.

I. INTRODUCTION

Numerical simulations of lattice quantum chromodynamics (QCD) and other lattice gauge theories rely on Markov chain Monte Carlo techniques to evaluate the path integral that defines the theory and its correlation functions. Ensembles generated in a Markov process, however, are often highly correlated due to slow modes in the stochastic evolution. Such correlations reduce the effective sample size of the generated ensemble, and thus directly influence the efficiency of such simulations. The problem of slow modes becomes particularly acute in the vicinity of critical points where the continuum limit is defined, resulting in what is commonly known as critical slowing down. Although a variety of algorithmic developments, such as cluster algorithms [1,2] and the worm algorithm [3], have dramatically reduced the problem of critical slowing down for some simple statistical models, they appear to have limited utility for gauge theories such as QCD, where

simulations at lattice spacings $a < 0.05$ fm remain extremely challenging.

In gauge theories, topological quantities[1] (e.g., topological charge and susceptibility) are examples of observables that couple strongly to slow modes of the stochastic evolution. In the continuum and at infinite volume, topological charge is invariant under continuous local deformations of a field configuration. By contrast, on a finite lattice, changes in topology are possible through local updates. Such changes, however, require traversals over large action barriers in configuration space, which in the continuum become infinite and result in the breakup of the configuration space into distinct topological sectors. The likelihood of such tunneling events rapidly diminishes in the approach to the continuum as the height of such topological barriers diverge, resulting in a problem known as topological freezing. This phenomenon was first observed in quenched calculations with improved gauge actions [4] as well as in more recent dynamical simulations [5]. Recently, open boundary conditions (BCs) in time [6] were proposed as a method for enhancing changes in

[*]endres@mit.edu
[†]brower@bu.edu
[‡]wdetmold@mit.edu
[§]kostas@wm.edu
[‖]avp@mit.edu

---

[1]Below, we will specify to a particular definition of topology, but the evolution properties and connection to slow modes are insensitive to these details.

topology by allowing charge to flow in and out through the boundaries. Although offering an improvement in topological tunneling over periodic BCs, open BC simulations still suffer from critical slowing down [7].

It is important to note that for gauge theories, critical slowing down persists even in the absence of topological freezing, because the evolution of long distance (slow) modes can only arise through the application of many local updates at the scale of the lattice spacing. As the lattice spacing is reduced, the number of updates required to move modes at a given physical scale increases. Multiscale evolution algorithms offer the prospect of performing Markov process updates that change modes at different physical scales more efficiently. A number of such approaches have been explored in the literature, primarily for models that are simpler than QCD, and have met with some success (see, e.g., [8–13]). We are unaware, however, of any successful work in this direction relevant to QCD.

In this study we investigate a less ambitious direction, namely a multiscale *thermalization* algorithm, which combines standard heat bath (HB) or hybrid Monte Carlo (HMC) updating methods with the real-space renormalization group (RG) and multigrid concepts of restriction and prolongation between pairs of matched coarse and fine lattices and lattice actions. The algorithm proceeds in four steps:

(1) A coarse action is determined by a RG transformation from the target (fine) lattice action[2];

(2) A set of N_s independent equilibrated coarse configurations are subsequently generated by a conventional Monte Carlo process;

(3) Each coarse configuration is then prolongated (or refined), thereby producing a set of N_s configurations defined on the fine lattice;

(4) The prolongated (fine) ensemble is then equilibrated (or rethermalized) and evolved in parallel using a conventional algorithm to produce an ensemble of N_e decorrelated configurations for each of the N_s independent streams.

This procedure may be generalized to have several levels of refinement proceeding from the coarsest to the finest target ensemble. At each level, the coarse action should follow an RG flow of the underlying gauge dynamics.

Assuming that the computational cost of the coarse evolution and prolongation are negligible compared to the fine evolution, the efficiency of this strategy is determined by the rethermalization time of the prolongated ensemble compared to the decorrelation time for fine evolution. Under the physically reasonable assumption that the distribution of prolongated configurations only differs from that of the target distribution for fine configurations by cutoff artifacts, one might expect the former time scale to be

shorter than the latter. Given this is indeed the case, the scheme will provide an efficient method for initializing field configurations at a fine lattice spacing for subsequent parallel evolution, ultimately yielding decorrelated ensembles of size $N_s \times N_e$.

In many cases, thermalization is considerably more challenging than evolution, and therefore we expect the approach to have significant advantages. Computationally, the parallel nature of the fine evolution of multiple independent streams means that ensembles can be generated more efficiently using fewer computational resources. The trade-off between the parameters N_s and N_e opens possibilities for optimizing the statistical power of subsequent analysis and the use of hardware resources. Furthermore, the strategy can be implemented on a hierarchy of different coarse/fine pairs resulting in rapid thermalization at multiple scales, thus enabling simulations at very fine lattice discretizations.

To test the viability of our strategy, we study a variety of observables that probe long distance scales in pure $SU(3)$ gauge theory. To facilitate our studies, we utilize restriction as a device for preparing coarse ensembles corresponding to a renormalized coarse action. An appealing feature of our restriction and prolongation operations is that they well preserve the topological charge distribution of the ensembles to which they are applied. As a consequence, the ensembles obtained by prolongation will possess properly distributed topology up to lattice artifacts which are inherited from the coarse action. The prolongator in fact satisfies the stronger property of preserving the topological charge for individual configurations at sufficiently small, but presently accessible, lattice spacings. This property is demonstrated numerically by studying the growth in correlations in topology between ensembles before and after restriction and prolongation as a function of the inverse lattice spacing, as shown in Fig. 1 (left). These features are important, since they enable us to achieve thermalized ensembles in time frames which are far shorter than the decorrelation time for fine evolution, providing the latter is controlled by topology.

Finally, a key measure for establishing the success of our approach is the requisite rethermalization time for an ensemble prepared via prolongation to return to equilibrium under standard updating procedures. This time is to be compared with the thermalization time for a typical ordered ("cold") or disordered ("hot") start, as well as the decorrelation time for fine evolution. To address this, we monitor the (re)thermalization times for a variety of observables, including the topological susceptibility and rectangular Wilson loops. In light of the fact that our prolongator preserves topology, we emphasize the study of rethermalization times for nontopological long-distance quantities and demonstrate that they are significantly shorter than the thermalization times and decorrelation times of topological quantities in conventional evolution. In Fig. 1 (right), we

---

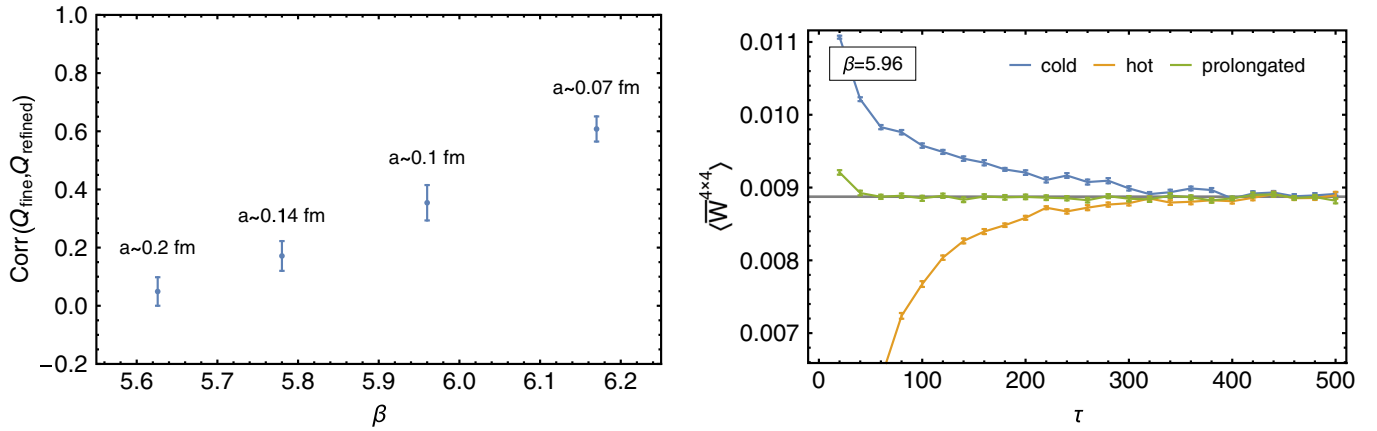[2]In practice, an approximation to the RG transformation is used.

FIG. 1 (color online).   Left: Correlations in topology between an ensemble before (fine) and after restriction and prolongation (refined), as a function of the coupling, β (see Sec. IV D for details). Right: Average Wilson loop (approximately 0.4 fm × 0.4 fm in dimension) as a function of rethermalization time for an initial ensemble obtained via prolongation of a matched coarse ensemble, and corresponding estimates as a function of the thermalization time for cold and hot starts; (re)thermalization time is measured in terms of the number of unit length HMC trajectories, τ (see Sec. IV F and Appendix D for details). The horizontal band indicates a high precision determination of the Wilson loop obtained from a large decorrelated ensemble.

provide an illustrative comparison of the rethermalization time for a representative prolonged ensemble of size $N_s = 24$, and the corresponding thermalization times for hot and cold initial ensembles, as probed by a 0.4 fm × 0.4 fm Wilson loop (see Sec. IV F and Appendix D for details). For this observable and a large range of other quantities that we investigate, we see that the rethermalization time for the prolonged ensemble is dramatically shorter than the thermalization times measured for hot and cold starts. It should be emphasized that our choice of prolongator is designed to preserve a large class of Wilson loops on all scales, and it is likely this feature that enables the rapid thermalization seen in this example.

In the remainder of this paper, we elaborate and expand on the results highlighted above. Before doing so, we first review some basic concepts and known results relating to Markov processes, which provide a theoretical basis for our strategy. Following an overview of notation and definitions, we then introduce the specific choice of restriction and prolongation operations used throughout this work. The latter is carefully chosen so as to retain an imprint of the long distance correlations and topological charge of the coarse configurations. Then, we demonstrate the viability of the multiscale approach through numerical studies of pure Yang-Mills gauge theory in two parts: first, by showing that the proposed prolongation procedure preserves the topological charge on a configuration by configuration basis for sufficiently fine lattice spacing, and second by demonstrating that the rethermalization time required to correct the distribution obtained by prolongation of a coarse ensemble is shorter than the decorrelation time for fine evolution. For the second studies we consider two commonly used algorithms, namely, HB and HMC. The latter case is of greater interest, since it is the algorithm

used in state-of-the-art QCD simulations with dynamical fermions. Finally, we conclude with a detailed discussion of the potential applications and pitfalls of these approaches, as well as an outlook on future directions. Appendixes are devoted to technical details of the simulation and prolongation algorithms as well as technical aspects of the data analysis that are required to extract the (re)thermalization and evolution time scales in this work.

II. PRELIMINARY CONSIDERATIONS

We begin by reviewing the basic aspects of Markov processes and their use in Monte Carlo importance sampling. A sequence of configurations

$$s_1 \to s_2 \to s_3 \to \cdots \to s_\tau, \tag{1}$$

labeled for simplicity by a discrete index τ, is generated in a Markov process described by the transition matrix $\mathcal{M}$. For simplicity, we assume $\mathcal{M}$ acts on a discrete configuration space Σ. The matrix elements $\mathcal{M}(s', s)$ give the transition probabilities for the configuration s to go to the configuration s'. Under adequate conditions (see, e.g., [14]), there exists a stationary distribution $\chi_0(s) = \mathcal{P}(s)$, which is a right eigenstate of the transition matrix, i.e., satisfying $\mathcal{P}(s') = \sum_s \mathcal{M}(s', s)\mathcal{P}(s)$, normalized to $\sum_s \mathcal{P}(s) = 1$ with eigenvalue $\lambda_0 = 1$. The left eigenstate is $\tilde{\chi}_0(s) = 1$ as a consequence of probability conservation: $\sum_{s'} \mathcal{M}(s', s) = 1$. Expectation values of operators $\mathcal{O}(s)$ averaged over the stationary distribution are given by the inner product $\langle \mathcal{O} \rangle = \sum_s \mathcal{O}(s)\mathcal{P}(s)$. Although it is not required for a valid algorithm, let us assume that $\mathcal{M}$ satisfies detailed balance, such that

$$\mathcal{M}(s',s)\mathcal{P}(s) = \mathcal{M}(s,s')\mathcal{P}(s'). \qquad (2)$$

In this case, the spectrum of $\mathcal{M}$ consists of real eigenvalues λ_n, with $|\lambda_n| \equiv e^{-1/\tau_n}$ ordered such that $|\lambda_n| \leq |\lambda_{n-1}|$; note that $|\lambda_n| \leq 1$ for all n. The corresponding right and left eigenvectors are given by $\chi_n(s)$ and $\tilde{\chi}_n(s) = \chi_n(s)/\mathcal{P}(s)$, respectively, and are chosen to be real and mutually orthonormal, satisfying $\sum_s \tilde{\chi}_n(s)\chi_{n'}(s) = \delta_{nn'}$. The spectral decomposition of τ applications of the transition matrix is given by

$$\mathcal{M}^\tau(s',s) = \sum_{n \geq 0} \chi_n(s')\lambda_n^\tau \tilde{\chi}_n(s). \qquad (3)$$

It follows that the τth configuration in Eq. (1) is drawn from the distribution

$$\mathcal{P}_\tau(s') = \mathcal{P}(s') + \sum_{n>0} \chi_n(s')\lambda_n^\tau \left[\sum_s \tilde{\chi}_n(s)\mathcal{P}_1(s)\right], \qquad (4)$$

where $\mathcal{P}_1(s)$ is an initial probability distribution. The expectation of an operator at this point in the Markov process is given by

$$\langle \mathcal{O}\rangle_\tau = \langle \mathcal{O}\rangle + \sum_{n>0}\left[\sum_s \chi_n(s)\mathcal{O}(s)\right]\left[\sum_s \tilde{\chi}_n(s)\mathcal{P}_1(s)\right]\lambda_n^\tau. \qquad (5)$$

At late times in the Markov process, both the distribution and the observables converge to their stationary values exponentially, assuming the existence of a gap in the spectrum of $\mathcal{M}$. The rate of this convergence (i.e., thermalization time) is dominated by the exponential correlation time $\tau_{\exp} \equiv \tau_1$. In addition to the time scale, $\tau_{\exp}$, the rate of thermalization is influenced by the overlap between the initial distribution $\mathcal{P}_1(s)$ and left eigenvectors $\tilde{\chi}_n(s)$, for $n > 0$; in particular, if the overlap vanishes or is exponentially small for $n = 1$, then the relevant thermalization time scale would be governed by the shorter time scale, τ_2. It is observationally established in many examples that near a phase transition, $\tau_{\exp} \sim \xi^z$ where ξ is the largest correlation length of the system (in dimensionless units) and z is a dynamical critical exponent. Local updating processes are diffusive by nature, implying an exponent $z \sim 2$. However, in some cases, such as lattice gauge theories, the scaling can be far worse (e.g., $z \sim 5$ for topological quantities [5]). For critical systems of spatial extent L, $\xi \sim L$, and the scaling becomes $\tau_{\exp} \sim L^z$; this kind of volume scaling is a hallmark property of critical slowing down.

A second time scale (or set of scales) of interest is the integrated autocorrelation time $\tau_{\mathrm{int}}(\mathcal{O})$, which characterizes the correlations in measurements of an observable after thermalization due to the sequential nature of the Markov process. In contrast with $\tau_{\exp}$, this time scale depends not only on the algorithmic details (e.g., the eigenvalues and eigenvectors of $\mathcal{M}$) but also on how well the observable in question couples to the various modes of the stochastic process. Because of the presence of such correlations, the estimated uncertainties on a given quantity $\langle \mathcal{O}\rangle$ are enhanced by a factor $\sqrt{2\tau_{\mathrm{int}}(\mathcal{O})}$ compared to those obtained under the assumption that the ensemble is decorrelated. The integrated autocorrelation time is defined by

$$\tau_{\mathrm{int}}(\mathcal{O}) = \frac{1}{2} + \sum_{\Delta>0}\frac{\Gamma_\Delta(\mathcal{O})}{\Gamma_0(\mathcal{O})}, \qquad (6)$$

where

$$\Gamma_\Delta(\mathcal{O}) = \sum_{s's}\delta\mathcal{O}(s')\mathcal{M}^\Delta(s',s)\delta\mathcal{O}(s)\mathcal{P}(s) \qquad (7)$$

is the lag-Δ autocovariance function, and $\delta\mathcal{O}(s) = \mathcal{O}(s) - \langle\mathcal{O}\rangle$. Using Eq. (3), this expression may be written as

$$\Gamma_\Delta(\mathcal{O}) = \sum_{n>0}a_n(\mathcal{O})\lambda_n^\Delta, \qquad a_n(\mathcal{O}) = \left[\sum_s \delta\mathcal{O}(s)\chi_n(s)\right]^2, \qquad (8)$$

and consequently the integrated correlation time may be expressed as

$$\tau_{\mathrm{int}}(\mathcal{O}) = \left[\sum_{n>0}a_n(\mathcal{O})\right]^{-1}\sum_{n>0}a_n(\mathcal{O})\eta_n,$$
$$\eta_n = \frac{1}{2} + \frac{\lambda_n}{1-\lambda_n}, \qquad (9)$$

where $\eta_n > 0$ for all $n > 0$. Under the assumption that $\mathcal{O}$ is real (or is the real part of an observable), then $a_n(\mathcal{O}) \geq 0$ for all $n > 0$, and one can establish the bound $\tau_{\mathrm{int}}(\mathcal{O}) \leq \hat{\tau}_{\mathrm{int}}$, where

$$\hat{\tau}_{\mathrm{int}} \leq \frac{1}{2} + \frac{|\lambda_1|}{1-|\lambda_1|} \leq \tau_{\exp} + \frac{1}{12}\frac{1}{\tau_{\exp}}. \qquad (10)$$

It follows that integrated autcorrelation times are at worst on the order of $\tau_{\exp}$ when the latter is large. Interestingly, this bound does not preclude the possibility that $\hat{\tau}_{\mathrm{int}} \ll \tau_{\exp}$.

In a standard Markov chain Monte Carlo simulation, represented schematically by Fig. 2(a), there are two relevant time scales associated with the algorithm: the equilibration or thermalization time $\tau_{\mathrm{therm}} \propto \tau_{\exp}$, and the decorrelation time for observables, which is bounded by $2\hat{\tau}_{\mathrm{int}}$. The former will depend to some extent on the initial configuration, drawn from the probability distribution $\mathcal{P}_1(s)$; if the initial configuration is drawn from the
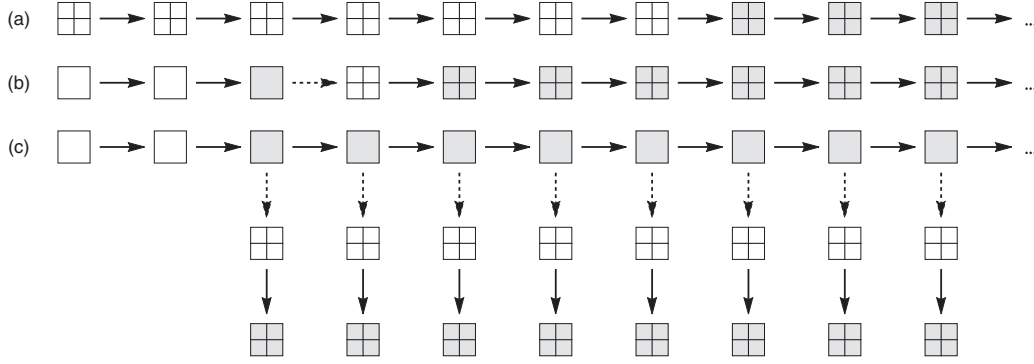
FIG. 2.   Ensemble generation strategies: single fine lattice stream (a), single coarse lattice stream, followed by refinement followed by a single fine lattice stream (b), and a single coarse lattice stream, followed by parallel refinement and rethermalization of refined lattices (c). In all cases, ⊞ represents a fine configuration, □ represents a coarse configuration, unshaded shapes correspond to unthermalized configurations and shaded shapes correspond to thermalized configurations. For each simulation strategy, ensemble averages are performed over shaded (fine) configurations, either generated from a single stream (a,b) or in parallel (c).

stationary distribution $\mathcal{P}(s)$, then the thermalization time will vanish.[3]

Next, let us introduce operators that map probability distributions between fine and coarse configuration spaces. Borrowing the terminology of multigrid, we refer to these as *restriction* operators, $\mathcal{R}$, when mapping from the fine to coarse configuration space and *prolongation* operators, $\mathcal{Q}$, when mapping from the coarse to fine configuration space. To facilitate the discussion, we adorn all coarse and fine quantities with the labels (c) and (f), respectively. For example, fine and coarse configurations are labeled as $s^f \in \Sigma^f$ and $s^c \in \Sigma^c$, where Σ^f and Σ^c represent the fine and coarse configuration spaces, respectively. The restrictor and prolongator can be represented by the matrices $\mathcal{R}(s^c, s^f)$ and $\mathcal{Q}(s^f, s^c)$ which act on fine and coarse configuration spaces, respectively. The restrictor and prolongator should be probability preserving, and therefore must satisfy $\sum_{s^f} \mathcal{Q}(s^f, s^c) = \sum_{s^c} \mathcal{R}(s^c, s^f) = 1$. Such transformations can be one to one, in which case the rectangular matrices $\mathcal{R}$ and $\mathcal{Q}$ have at most one nonzero entry per row and column, or they can be probabilistic. Both restriction and prolongation operations are nonunique, need not satisfy $\mathcal{R}\mathcal{Q} = 1$, and cannot satisfy $\mathcal{Q}\mathcal{R} = 1$ since the rank of $\mathcal{Q}$ and $\mathcal{R}$ is that of $\dim(\Sigma^c)$ and not $\dim(\Sigma^f)$. Explicitly, the restriction operation acting on a fine probability distribution $\mathcal{P}^f$ produces a coarse probability distribution, given by

$$\mathcal{P}^c(s^c) = \sum_{s^f} \mathcal{R}(s^c, s^f)\mathcal{P}^f(s^f), \qquad (11)$$

and can be interpreted as a renormalization group transformation (e.g., decimation or block spin averaging in a

simple implementation). This can be seen by noting the equality of partition functions $\sum_{s^c}\mathcal{P}^c(s^c) = \sum_{s^f}\mathcal{P}^f(s^f)$. On the other hand, the prolongation operation maps a coarse probability distribution to a fine distribution, given by

$$\mathcal{P}^f(s^f) = \sum_{s^c} \mathcal{Q}(s^f, s^c)\mathcal{P}^c(s^c), \qquad (12)$$

and can be interpreted as a kind of inverse RG transformation.

With the concepts of restriction and prolongation in hand, consider a simulation represented schematically by Fig. 2(b), corresponding to the scenario in which $N_s = 1$ and $N_e \gg 1$. Here, evolution is first performed on a coarse lattice using an algorithm represented by the coarse transition matrix $\mathcal{M}^c$ (which implicitly depends on a coarse action) until it is thermalized. Subsequently the lattice is prolongated, and finally rethermalized using an algorithm represented by the fine transition matrix $\mathcal{M}^f$. Note that the subsequent rethermalization is needed to correct the prolongated configuration at the scale of the fine cutoff. In this example, there are now three relevant time scales associated with the algorithm in its entirety: the coarse thermalization time τ^c_{therm}, the rethermalization time τ^f_{retherm}, and the decorrelation time of the fine evolution, bounded by $2\hat{\tau}^f_{\text{int}}$. The procedure represented by Fig. 2(b) will be computationally less costly than that shown in Fig. 2(a) provided $\tau^c_{\text{therm}} + \tau^f_{\text{retherm}} < \tau^f_{\text{therm}}$. Nevertheless, the improvements that can be found here are attenuated by the cost of the generation of a large ensemble since $\tau^f_{\text{therm}}/(N_e 2\hat{\tau}^f_{\text{int}}) \to 0$ as $N_e \to \infty$.

As previously discussed, the rethermalization time of the prolongated configuration is at worst governed by the time scale τ^f_{exp}, which is algorithm dependent, and overlap factors, which depend in part on the initial refined

---

[3]Strictly speaking, it does not make sense to talk about a thermalized configuration, but rather a configuration that is drawn from a thermalized distribution.

distribution and are thereby controllable. In light of Eq. (4) and Eq. (5), we can in principle accelerate the approach to equilibrium of the fine ensemble, by setting to zero the overlap of our refined ensemble with a fixed set of the slowest modes,

$$\sum_{s^f s^c} \tilde{\chi}_n^f(s^f) \mathcal{Q}(s^f, s^c) \mathcal{P}^c(s^c) = 0. \tag{13}$$

Removal of the lowest mode in this fashion, for example, would imply that the rethermalization time is no longer governed by $\tau_{\exp}^f$, but rather by the shorter time scale, τ_2^f. In practice this is difficult to achieve, but by judicious choices, one seeks to approximate this condition for as wide a range of slow modes as possible. Note that this condition depends on the prolongator, on the Markov process used for fine evolution, and implicitly on the renormalized coarse action. All of these factors are therefore important in maximizing the efficiency of our algorithm.

Finally, let us consider a simulation represented schematically by Fig. 2(c), corresponding to the scenario in which $N_s \gg 1$, and $N_e = 1$. In this case, the evolution is first performed on a coarse lattice using an algorithm represented by the transition matrix $\mathcal{M}^c$ until an ensemble of decorrelated configurations are generated. The ensemble of decorrelated coarse configurations are subsequently prolongated, and finally rethermalized using an algorithm represented by the transition matrix $\mathcal{M}^f$. This procedure has three time scales associated with it: the coarse thermalization time τ_{therm}^c, the decorrelation time for coarse evolution, bounded by $2\hat{\tau}_{\text{int}}^c$, and the rethermalization time τ_{retherm}^f. The procedure is computationally less costly than Fig. 2(a) provided that $\tau_{\text{therm}}^c + N_s \hat{2}\tau_{\text{int}}^c + N_s \tau_{\text{retherm}}^f < \tau_{\text{therm}}^f + N_s 2\hat{\tau}_{\text{int}}^f$, where N_s is the size of the target ensemble being generated. For large N_s, the condition reduces to $2\hat{\tau}_{\text{int}}^c + \tau_{\text{retherm}}^f < 2\hat{\tau}_{\text{int}}^f$. Since the decorrelation time for coarse evolution is usually negligible compared to that for fine evolution, the approach will be less computationally costly when the decorrelation time for fine evolution exceeds the rethermalization time for the prolongated ensemble. Note that the computationally most intensive component of this algorithm, namely rethermalization, is embarrassingly parallel, and so each stream can be generated with maximal efficiency on available computing resources.

Our goal for the remainder of this paper is to explore choices of $\mathcal{P}^c$ and $\mathcal{Q}$, given a fine transition matrix $\mathcal{M}^f$, such that Fig. 2(c) becomes a viable simulation strategy in pure gauge theory, and to investigate the time scales associated with the various approaches described above. Provided that the fine evolution is ergodic, the proposed algorithm as a whole will also be ergodic. Given that the prolongated ensemble inherits the long-distance properties

of the coarse ensemble, the relevant parts of the fine configurations space Σ^f are in some sense evenly populated. Since the rethermalization merely corrects the missing short-distance part of the prolongated distribution, our underlying assumption is that all of the fine configuration space can be covered within the rethermalization time.

III. ACTIONS, OBSERVABLES, RESTRICTION, AND PROLONGATION

Here, we provide explicit details pertaining to the implementation of our multiscale thermalization algorithm. We begin by considering a D-dimensional hypercubic lattice, with lattice spacing a and periodic boundary conditions. Let us label the sites of the lattice with D-vectors $\mathbf{n}$, with components $n_\mu = \mathbf{n} \cdot \mathbf{e}_\mu$, where $\mathbf{e}_\mu$ is a unit basis vector in the μ direction and $\mu = 0, \ldots, D - 1$. Note that a D-dimensional hypercubic lattice comprises q-dimensional unit "q-cells," where $q = 0, \ldots, D$. For example, with $D = 4$, the lattice comprises sites ($q = 0$), bonds ($q = 1$), and plaquettes ($q = 2$), and so on. The total number of such cells is given by

$$N_q = N_0 \binom{D}{q}, \tag{14}$$

where N_0 is the total number of lattice sites.

The Wilson action [15] for pure lattice Yang-Mills theory is given by

$$S = \beta \sum_{\mathbf{n}} \sum_{\mu < \nu} [1 - W_{\mu\nu}^{1\times 1}(\mathbf{n})], \tag{15}$$

where $U_\mu(\mathbf{n}) \in SU(N_c)$ are variables associated with the bonds of the lattice, $W_{\mu\nu}^{1\times 1}(\mathbf{n})$ are 1×1 Wilson loops associated with the plaquettes of the lattice, and $\beta = 2N_c/g^2$ is the coupling. Note that if Wilson lines are given by

$$L_\mu^m(\mathbf{n}) = \mathcal{P} \prod_{n=0}^{m-1} U_\mu(\mathbf{n} + n\mathbf{e}_\mu), \tag{16}$$

where $\mathcal{P}$ is the path-ordering symbol, then a rectangular $m \times n$ Wilson loop in the $\mu - \nu$ plane is given by

$$W_{\mu\nu}^{m\times n}(\mathbf{n})$$
$$= \frac{1}{N_c} \Re \text{Tr} L_\mu^m(\mathbf{n}) L_\nu^n(\mathbf{n} + m\mathbf{e}_\mu) L_\mu^m(\mathbf{n} + n\mathbf{e}_\nu)^\dagger L_\nu^n(\mathbf{n})^\dagger, \tag{17}$$

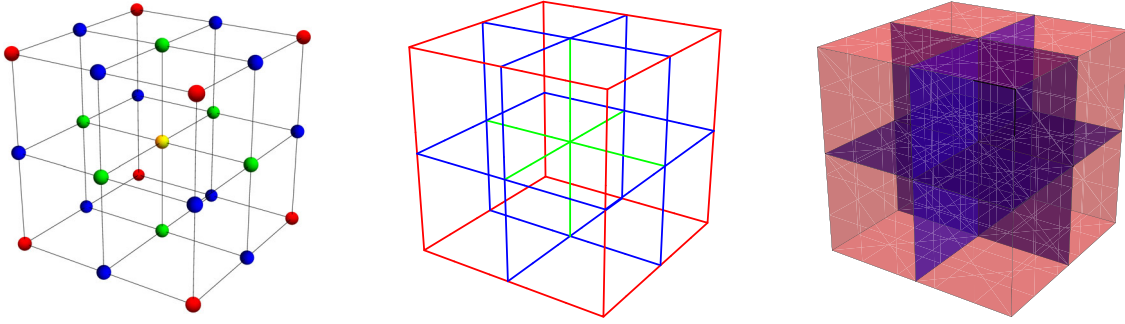and the corresponding space-time averaged Wilson loop is given by

FIG. 3 (color online). Classification of lattice cells, labeled by the integers 0 (red), 1 (blue), 2 (green), and 3 (yellow). Left: $\chi(\mathbf{n})$. Center: $\chi_\mu(\mathbf{n})$. Right: $\chi_{\mu\nu}(\mathbf{n})$.

$$\bar{W}^{m\times n} = \frac{1}{N_2}\sum_{\mathbf{n}}\sum_{\mu<\nu} W_{\mu\nu}^{m\times n}(\mathbf{n}). \tag{18}$$

For the purpose of this study, we consider restriction and prolongation operations that take an ensemble associated with a "fine" lattice with spacing a to an ensemble associated with a "coarse" lattice with spacing $2a$, and back. To facilitate the discussion, we begin by classifying the various q-cells of the fine lattice according to their positions with respect to the 2^D hypercubes which define the coarse lattice. We define the function

$$\chi(\mathbf{n}) = \sum_\mu (n_\mu \bmod 2), \tag{19}$$

which allows us to associate integers $0, \dots, D$ to the sites of the 2^D hypercubes, as shown in Fig. 3(a) for $D = 3$ space-time dimensions. The subset of sites associated with the coarse lattice satisfy $\chi(\mathbf{n}) = 0$ and are consequently given by $\mathbf{n}/2 \in \mathbb{Z}^D$. Note that this convention is but one of 2^D possibilities for the alignment of the coarse lattice with respect to the fine; since the fine lattice theory is invariant under lattice translations, any choice is acceptable without loss of generality. Similarly, we may define the quantities

$$\begin{aligned}
\chi_\mu(\mathbf{n}) &= \chi(\mathbf{n} - n_\mu \mathbf{e}_\mu), \\
\chi_{\mu\nu}(\mathbf{n}) &= \chi(\mathbf{n} - n_\mu \mathbf{e}_\mu - n_\nu \mathbf{e}_\nu), \\
\chi_{\mu\nu\sigma}(\mathbf{n}) &= \chi(\mathbf{n} - n_\mu \mathbf{e}_\mu - n_\nu \mathbf{e}_\nu - n_\sigma \mathbf{e}_\sigma), \\
\chi_{\mu\nu\sigma\rho}(\mathbf{n}) &= \chi(\mathbf{n} - n_\mu \mathbf{e}_\mu - n_\nu \mathbf{e}_\nu - n_\sigma \mathbf{e}_\sigma - n_\rho \mathbf{e}_\rho),
\end{aligned} \tag{20}$$

which associate the integers $0, \dots, D - q$ with the remaining q-cells of the lattice, where $q = 0, \dots, D$. The classification of bonds is shown in Fig. 3(b) and the classification of plaquettes is shown in Fig. 3(c) for $D = 3$ space-time dimensions.

For this study, we consider the simplest restriction (or blocking) procedure, which proceeds by assigning products of fine bond variables U_μ^f to the coarse bond variables U_μ^c,

$$U_\mu^c(\mathbf{n}/2) = U_\mu^f(\mathbf{n})U_\mu^f(\mathbf{n} + \mathbf{e}_\mu), \tag{21}$$

for all values of $\mathbf{n}$ satisfying $\chi(\mathbf{n}) = 0$. Other schemes are also possible; however, the specific choice plays a minor role in the present implementation of our algorithm. In all cases, information is lost in the restriction operation.

Prolongation proceeds in two steps: first, the transfer of the coarse lattice variables to an appropriate subset of bonds on the fine lattice, and second, interpolation of the transferred variables to the remaining undefined bonds of the fine lattice. The prolongator is designed to preserve the long-distance structure of the theory (as it is encoded by the configurations), including correlation lengths and topological charge. As a consequence, if the coarse configurations to which prolongation is applied are thermalized, the resulting fine configurations will also be thermalized except for short-distance defects at the scale of the cutoff. Fine evolution can correct for such short-distance defects, and it is reasonable to expect that the evolution time required to bring the entire prolongated ensemble into thermal equilibrium will be short. Since topology freezing is one of the major issues in present-day simulations that we aim to address, it is advantageous for the restriction and prolongation procedures to preserve the topological charge either with configuration by configuration within an ensemble or in terms of its distribution over the ensemble.[4]

The first step of the prolongation procedure is to associate the coarse bond variables with the fine lattice. Since there are more bonds on the fine lattice than the coarse, there is no unique prescription for doing this. However, a simple choice is to demand that

$$U_\mu^f(\mathbf{n}) = U_\mu^c(\mathbf{n}/2), \qquad U_\mu^f(\mathbf{n} + \mathbf{e}_\mu) = 1, \tag{22}$$

for all $\mathbf{n}$ satisfying $\chi(\mathbf{n}) = 0$. Note that the gauge freedom of the fine action allows us to set one of the two fine bond

---

[4]Note that the topological charge depends on the ultraviolet regulator and a particular definition will be discussed below. It is only in the limit of weak coupling where configurations satisfy an admissibility criterion [16] that the definition becomes unique.
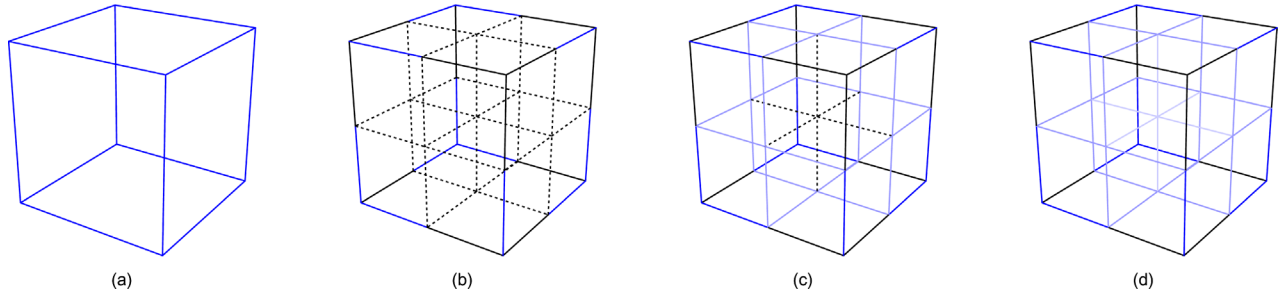
FIG. 4 (color online).   Schematic description of prolongation: (a) coarse bond variables; (b) transfer of coarse bond variables to a fine lattice; (c) interpolation performed on 2×2 plaquettes; (d) interpolation performed on $2 \times 2 \times 2$ cubes; the procedure is continued for the remaining stages (not shown). Black bond variables are set to unity, as allowed by gauge freedom, and dashed bond variables are undefined at intermediate stages of the refinement; dark blue bond variables are transferred from the coarse lattice, whereas the lighter blue bond variables are determined via the interpolation.

variables above to unity; in Fig. 4(b) we show an example of this assignment for a $2 \times 2 \times 2$ cell, having transferred bond variables from the coarse unit cell shown in Fig. 4(a). The remaining bond variables are undefined and may be set to arbitrary values; in this study we initially set them to unity. Note that 2×2 Wilson loops originating from the even sites of the fine lattice [i.e., where $\chi(\mathbf{n}) = 0$] are exactly equal to the plaquettes of the coarse lattice. Furthermore, all even length Wilson loops originating from the even sites are exactly preserved by the map. This is the key to our construction, and it implies that there is a set of long distance loops from which the renormalization group invariant area law can be computed (i.e., using Creutz ratios constructed from even-sided loops).

The second step, interpolation, is designed to remove the most damaging ultraviolet defects induced by the first step. There are a number of ways to carry out the interpolation of gauge fields (see, e.g., [16–18]). Following the approach of 't Hooft [18], we use an interpolation which respects the Lüscher bound for sufficiently smooth configurations [16], and as a consequence, exactly preserves the topological charge for those configurations. The gauge field interpolation is carried out by sequentially minimizing the partial actions,

$$S_d = \beta \sum_{\mathbf{n}} \sum_{\mu<\nu} \delta_{d,\chi_{\mu\nu}(\mathbf{n})} [1 - W_{\mu\nu}^{1\times 1}(\mathbf{n})] \qquad (23)$$

with respect to "active" bond variables which satisfy $\chi_\mu(\mathbf{n}) = d + 1$, for $d = 0, \ldots, D - 2$. The interpolation proceeds starting from low dimensional to high dimensional cells. A useful property of this prescription is that at each stage of the interpolation, the active bond variables in one 2^{d+2} cell are completely decoupled from those in neighboring 2^{d+2} cells. Thus the interpolation can be performed locally at each stage.

At stage $d = 0$, the minimization can be performed analytically, following [18]; however, the analytic forms become complicated for $d > 0$. In this study, we followed a numerically simpler procedure for performing the

minimization that is valid for all stages. Specifically, repeated applications of APE smearing [19,20] of the form

$$U_\mu(\mathbf{n}) \to U_\mu'(\mathbf{n})$$
$$= \mathbb{P}_{SU(N_c)} \left[U_\mu(\mathbf{n}) + c \sum_{\sigma=\pm} \sum_\nu \delta_{d,\chi_{\mu\nu}(\mathbf{n})} T_{\mu\nu}^\sigma(\mathbf{n}) \right] \quad (24)$$

were performed on the active bonds at a given stage, where

$$T_{\mu\nu}^+(\mathbf{n}) = U_\nu(\mathbf{n}) U_\mu(\mathbf{n} + \mathbf{e}_\nu) U_\nu(\mathbf{n} + \mathbf{e}_\mu)^\dagger \qquad (25)$$

and

$$T_{\mu\nu}^-(\mathbf{n}) = U_\nu(\mathbf{n} - \mathbf{e}_\nu)^\dagger U_\mu(\mathbf{n} - \mathbf{e}_\nu) U_\nu(\mathbf{n} - \mathbf{e}_\nu + \mathbf{e}_\mu)^\dagger \quad (26)$$

are forward and backward oriented staple operators, $\mathbb{P}_{SU(N_c)}$ is a projection operator onto $SU(N_c)$, and c is a small parameter to be specified later. The number of times this smearing is applied to the gauge fields will also be specified later.

Before moving on to numerical studies, we define several additional quantities, which will prove useful later on: partially space-time averaged plaquettes, associated with the different plaquettes subsets,

$$\bar{W}_d^{1\times 1} = \frac{1}{N_d^{1\times 1}} \sum_{\mathbf{n}} \sum_{\mu<\nu} \delta_{d,\chi_{\mu\nu}(\mathbf{n})} W_{\mu\nu}^{1\times 1}(\mathbf{n}), \qquad (27)$$

and average displaced 2×2 Wilson loops, given by

$$\bar{W}_d^{2\times 2} = \frac{1}{N_d^{2\times 2}} \sum_{\mathbf{n}} \sum_{\mu<\nu} \delta_{d,\chi(\mathbf{n})} W_{\mu\nu}^{2\times 2}(\mathbf{n}). \qquad (28)$$

The normalization for these quantities are given by

$$N_d^{1\times 1} = 2(D - d)(D - d - 1) \binom{D}{d} \frac{N_s}{2^D} \qquad (29)$$

and

$$N_d^{2\times2} = \binom{D}{2}\binom{D}{d}\frac{N_0}{2^D}, \qquad (30)$$

respectively. Note that

$$\sum_{d=0}^{D} N_d^{1\times1} = \sum_{d=0}^{D} N_d^{2\times2} = \binom{D}{2}N_0, \qquad (31)$$

which is just the total number of plaquettes on the lattice.

IV. SIMULATIONS

A. Target ensembles

In the remainder of this work we consider pure $SU(3)$ gauge theory in $D = 4$ dimensions and make use of four decorrelated target ensembles of size N, generated in a standard way. Physical observables on these ensembles will serve as benchmarks that the multiscale thermalization algorithm should reproduce. The ensembles are described in Table I and have lattice spacings ranging from approximately 0.07 to 0.2 fm, separated by multiples of $\sqrt{2}$. The lattice spacings were determined from empirical formulas relating the Sommer scale in lattice units (r_0/a) to the coupling [21,22], taking $r_0 = 0.5$ fm. The spatial extents of the lattices were chosen to be approximately 2.25–2.40 fm; the temporal extents were chosen to be twice the spatial extents in order to minimize thermal effects. Standard boundary conditions, periodic in all directions, were used throughout. Ensembles were generated with the Cabibbo-Marinari HB algorithm [23] combined with overrelaxation [24]. Each HB sweep was performed on a checkerboard sweep schedule with $N_{\text{hb}} = 1$ attempted updates to each $SU(2)$ subgroup per bond variable via the method of Creutz [25]. Each HB sweep was followed by $N_{\text{ov}} = 10$ overrelaxation sweeps following the same checkerboard sweep schedule. For all ensembles, 1500 HB sweeps were initially performed for thermalization starting from a weak field configuration; subsequent configurations were saved after every 100 sweeps for future use.

TABLE I. Decorrelated target ensembles of size N, generated using HB with 10 overrelaxation sweeps. Lattice spacing is set via the Sommer scale $r_0 = 0.5$ fm, based on the works [21] (coarsest) and [22] (finer). The reference scale t_0 is defined in Sec. IV B.

Lattice	β	a [fm]	N	t_0/a^2
$12^3 \times 24$	5.626	0.1995(20)	385	0.72966(69)
$16^3 \times 36$	5.78	0.1423(5)	385	1.3858(15)
$24^3 \times 48$	5.96	0.0999(4)	185	2.7891(45)
$32^3 \times 72$	6.17	0.0710(3)	185	5.5007(83)

B. Wilson flow

Wilson flow [26–28] was used to define a number of the observables studied in this work. The diffusive nature of the flow allows us to consider a series of observables, which probe different length scales at different flow times, t. Wilson flow was applied to the target ensembles described in Table I using both a fixed step size algorithm [27] and an adaptive step size algorithm [29]. The accuracy of the integration along the flow is controlled by the size of the step in the former case and a tolerance level in the latter case (see [29] for an explicit definition of this tolerance). The adaptive approach is more efficient because the flow has a smoothing effect on the fields. Consequently, the forces that drive the flow become smaller with flow time, thus enabling the use of larger step sizes at later times. We have established the validity of our implementation of the adaptive step Wilson flow by direct comparison with fixed step size Wilson flow for the target ensembles in Table I. For the autocorrelation time and (re)thermalization studies performed later in this work, the adaptive step size algorithm was used, due to its higher efficiency.

For the target ensembles, Wilson flow measurements were performed using a fixed step size of 0.01; results for the quantity $t^2 E(t)$ (for this study, we use the clover-leaf definition) are provided in Fig. 5 as a function of t/t_0. The Wilson flow scale, t_0, is defined by $t_0^2 E(t_0) = 0.3$; values of this scale and corresponding statistical errors were obtained by linearly interpolating the nearest estimates of $t^2 E(t)$ to 0.3. The results from this analysis are provided in Table I; for the 24^3 and 32^3 ensembles, we obtained estimates of t_0 which are consistent with [27]. For the same ensembles, the adaptive step size algorithm was used with a tolerance of 0.01. Measurements along the flow were made in multiples of $t_0^\star/4$, where $t_0^\star$ is introduced in Table II as a nominal value for t_0 [this parameter was chosen to be sufficiently close to t_0, but also a multiple of 0.01 so that we may directly compare estimates of $E(t)$ using both methods]. For the 32^3 ensemble, we found a maximum deviation of about 0.02% in the estimates of $E(t)$ for $t \in [0, 5t_0^\star]$. For 12^3, 16^3, and 24^3 lattices, the deviation was less than 0.001% on the same interval. The good agreement for these ensembles not only validates our implementation, but also indicates that our choice of tolerance level is adequate for the studies we are pursuing. Note that for $t \gtrsim 5t_0$, the flow radius $\sqrt{8t}$ exceeds the size of our lattices, and beyond that, the Wilson flow was used primarily for its smoothing properties in determining the topological charge. For all ensembles, the topological charge was found to be consistent between fixed and adaptive step algorithms over the entire range of flow times on a per configuration basis.

C. Restriction and prolongation

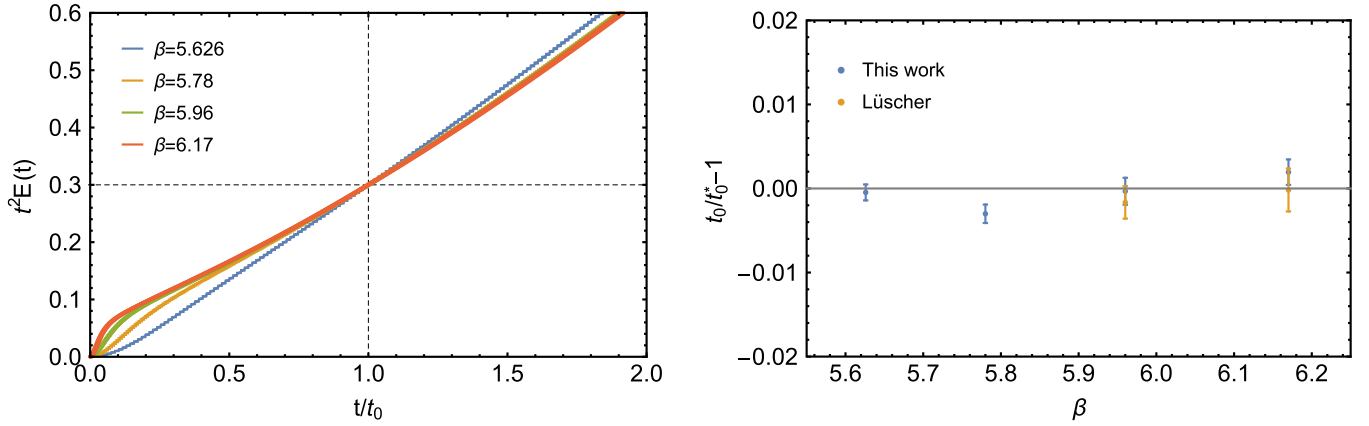As a first test of the prolongation algorithm described in Sec. III, we investigate how the target fine ensembles

FIG. 5 (color online).   Left: $t^2 E(t)$ as a function of t/t_0; determined values of t_0 are provided in Table I. Right: Deviation between t_0 obtained using Wilson flow with a fixed step size (this work) and the nominal values $t_0^\star$ introduced in Table II. Corresponding results obtained in [27] (Lüscher) are shown for comparison.

described in Table I are modified by the application of restriction using Eq. (21), followed immediately by prolongation using Eq. (22) and Eq. (23). The gauge field interpolation was performed by sequentially minimizing the partial actions S_d until at each stage, d, the relative change in the partial action reached 0.001%. For the ensembles considered, the action minimization required by our interpolation procedure was performed using repeated applications of APE smearing, using Eq. (24) with $c = 0.05$. The variation of the average plaquette, $\langle \bar{W}^{1\times 1} \rangle$, and partially averaged plaquette $\langle \bar{W}_d^{1\times 1} \rangle$ are shown in Fig. 6 for each ensemble as a function of the number of smearing applications, beginning with undefined bonds set to unity. Notice from the results that the average plaquette $\langle \bar{W}^{1\times 1} \rangle$, which is proportional to the action, S, up to an overall additive constant, is not a monotonically increasing function of the number of cooling sweeps. This is due to the fact that it is not the total action that is being minimized at each stage of the interpolation, but rather the partial action S_d.

The average partial plaquettes $\langle \bar{W}_d^{1\times 1} \rangle$ and displaced 2×2 Wilson loops $\langle \bar{W}_d^{2\times 2} \rangle$ are shown in Fig. 7 as a function of the target ensemble coupling for each ensemble

after restriction and prolongation. The former demonstrates that although the prolongated configurations retain an imprint of the coarse lattice, the configurations are nonetheless smooth by comparison to configurations from the associated target ensemble. The latter observable provides a measure of the reduced translational symmetry of the restriction/prolongation operators. Note that $\langle \bar{W}_0^{2\times 2} \rangle$ is just the average plaquette measured on the coarse lattice, whereas $\langle \bar{W}_4^{2\times 2} \rangle$ corresponds to fully displaced plaquettes. Clear signals of the reduced translational symmetry is evident with approximately a factor of 4 difference between the two. Later, we explore the rate at which displaced 2×2 Wilson loops converge to the same value as a function of Monte Carlo evolution time, since this provides a measure of how quickly the full translational symmetry of the fine theory is restored.

D. Topological charge

The topological charge, $Q(t)$, and susceptibility, $\chi(t)$ [defined as the square of $Q(t)$ divided by the spacetime volume], is determined as a function of the Wilson flow time using the three-loop improved gluonic definition of the topological charge operator [30]. As a function of flow time, the topological charge is expected to approach integer values. A measure of the deviation of the topological charge from integer values over the ensemble is given by

$$\epsilon(Q) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Q_i - [Q_i])^2}, \qquad (32)$$

where $[Q_i]$ is defined as the nearest integer to Q_i for each decorrelated configuration, labeled by i. Note that a locally uniform distribution for Q_i about integer values (e.g., when the distribution for Q is broad and smooth on scales much larger than unity) yields $\epsilon(Q) = 1/\sqrt{12}$. This measure is

TABLE II. Wilson flow parameters used for comparative study of Q on the fine and refined lattices, autocorrelation time studies, and (re)thermalization time studies; all flows were performed using an adaptive step size [29] with a maximum allowed step of 0.2.

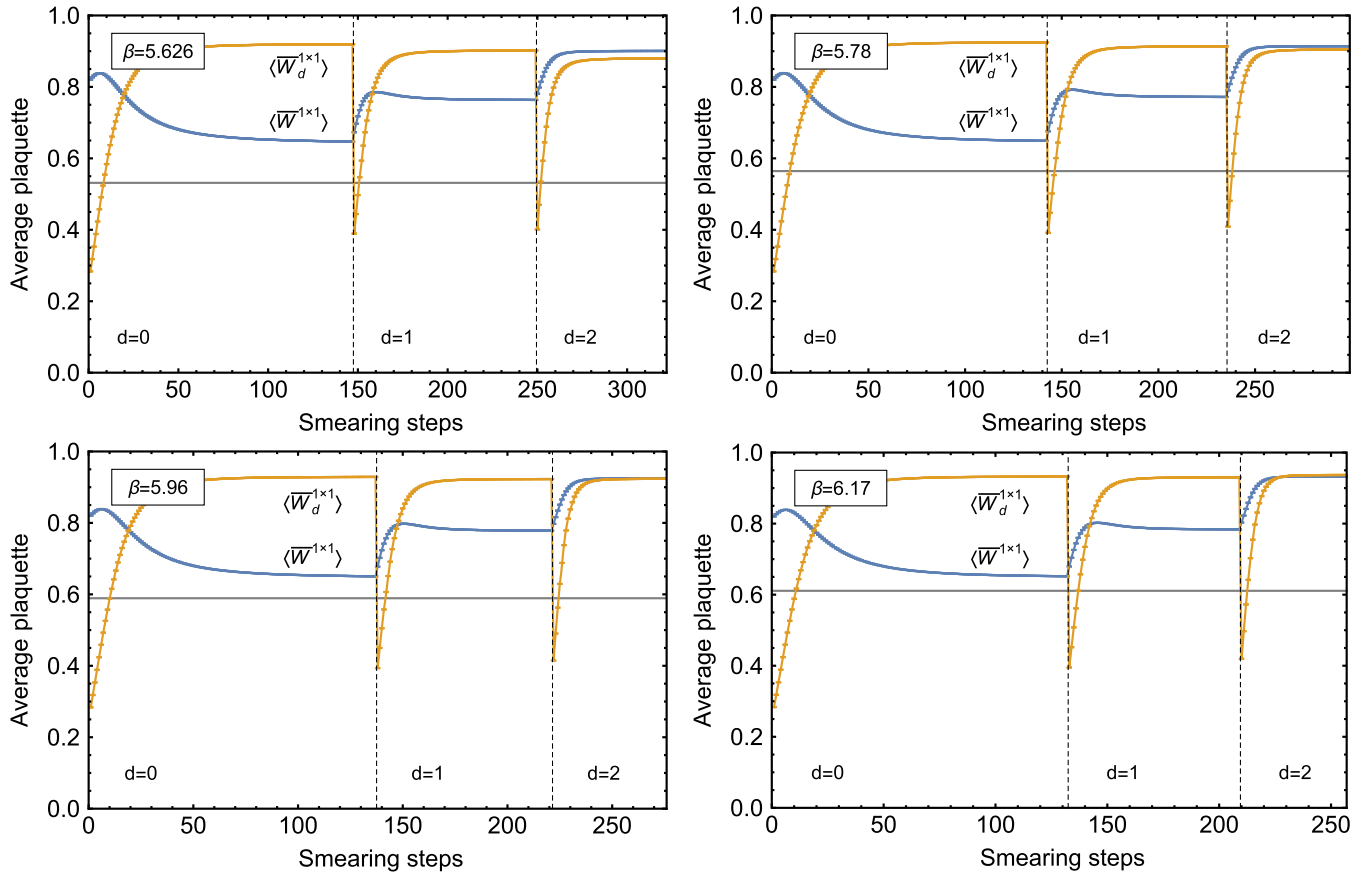| Lattice	$t_0^\star/a^2$	Flow time extent [$t_0^\star/a^2$]	Measurement frequency [$t_0^\star/a^2$]	Tolerance
$12^3 \times 24$	0.73	40	1/4	0.01
$16^3 \times 36$	1.39	25	1/4	0.01
$24^3 \times 48$	2.79	14	1/4	0.01
$32^3 \times 72$	5.49	8	1/4	0.01

FIG. 6 (color online).   Average plaquette as a function of the number of cooling sweeps at each level of interpolation.

expected to approach zero for the gluonic definition of the topological charge at late flow times. Plots of this quantity are provided in Fig. 8 for the target (fine) ensembles and ensembles obtained from restriction and prolongation (refined), as described in the previous section. In units of the nominal scale $t_0^\star$, we find that the topological charge approaches integer values faster as the lattice spacing is decreased. This can be understood in terms of smoothness

of the gauge field configurations. Following [27], we may consider the quantity

$$s_p = N_c[1 - W_{\mu\nu}^{1\times1}(\mathbf{n})], \qquad (33)$$

measured on field configurations at flow time t_0, where p is a plaquette associated with site $\mathbf{n}$ and basis vectors $\mathbf{e}_\mu$ and $\mathbf{e}_\nu$. Given the measure of the configurations' smoothness,
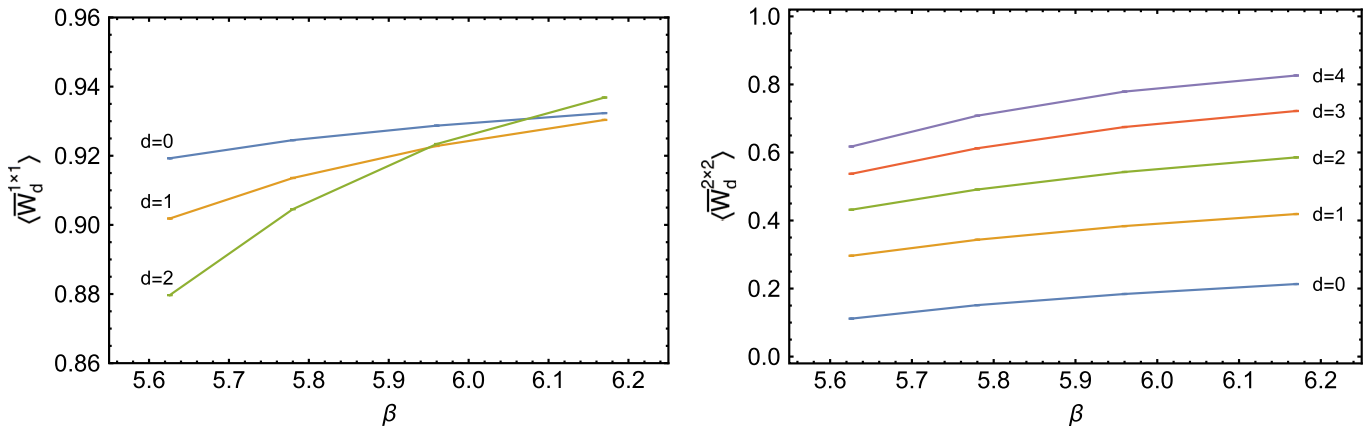

FIG. 7 (color online).   Partially averaged plaquette (left) and average displaced 2×2 Wilson loop (right) as a function of β, measured on the four target ensembles after restriction and prolongation.
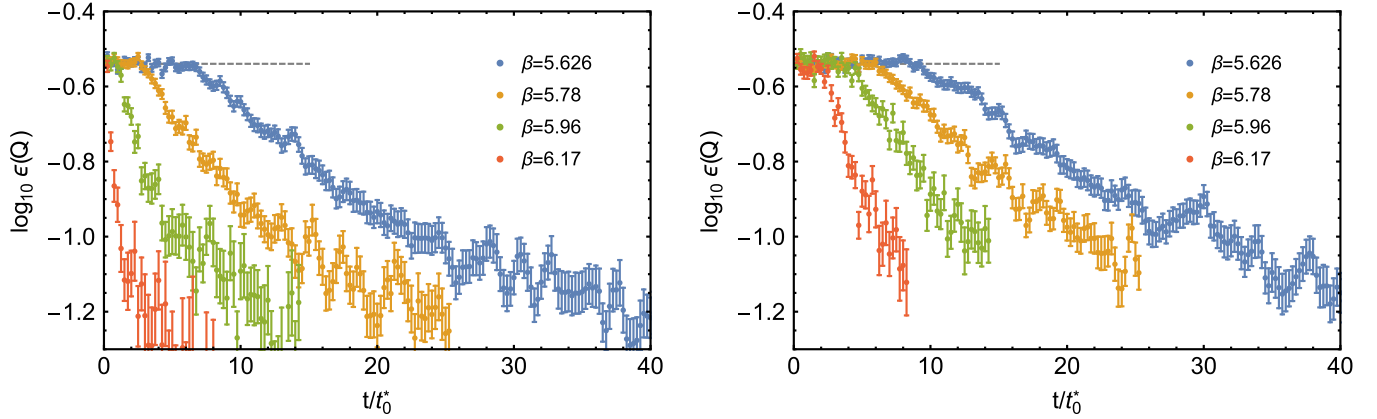
FIG. 8 (color online).   Plots of $\epsilon(Q)$ as a function of $t/t_0^\star$ for fine (left) and refined (right) ensembles; dashed line corresponds to $1/\sqrt{12}$.

$h = \max_p(s_p)$, gauge configurations satisfying the admissibility criterion $h < 0.067$ will fall into distinct topological sectors [27]. Configurations that violate the bound due to lattice artifacts, on the other hand, will not. In this study, we find that 0% of the configurations satisfy the criterion for $a \gtrsim 0.1$ fm, whereas only 9% of the configurations satisfy the criterion at $a \sim 0.07$ fm; these results appear consistent with [27]. Note that according to that study, at $a \sim 0.05$ fm, this percentage increases to about 70%. For equal flow times in lattice units, we find that the fine ensemble is more likely to have configurations that satisfy the admissibility condition than its refined counterpart; this result is counterintuitive, is due to the fact that the refined ensemble is smoother at the scale of the lattice spacing, and thus undergoes diffusion under Wilson flow at a rate slower than the fine ensemble. As a by-product of this, we see in Fig. 8(b) that the topological charge attains integer values at a somewhat slower rate than that in Fig. 8(a).

In Fig. 9, we show scatterplots of the topological charge measured at the longest flow time for each of the target ensembles, and the corresponding topological charge measured after restriction and prolongation. As was evident in Fig. 8, we see that for both fine and refined ensembles, the topological charge takes on integer values at large flow times. Furthermore, for large β we find that the distributions become increasingly skewed, indicating increasing correlation between the topological charge of the original and refined configurations. In Fig. 10 (left) we provide plots of this correlation as a function of flow time for each value of β. We note that the correlations in the topological charge are largely independent of flow time for $t \gtrsim t_0$; this observation holds even at early times, where the topological charge need not take integer values. In Fig. 1 (left), the correlations in the topological charge measured at the latest flow time are plotted for each value of β. Despite the large violations of the admissibility condition at the lattice spacings considered in this work, we nonetheless see a clear increasing trend in the topological charge correlations

between fine and refined lattices. Our expectation is that these correlations will rapidly approach unity as the lattice spacing is further reduced by a factor of $\sqrt{2} - 2$. Finally, in Fig. 10 (right) we show the p-values obtained from a two-sample Kolmogorov-Smirnov test that the topological charge distributions on fine and refined ensembles come from the same underlying distribution. The p-values were obtained as a function of Wilson flow time, and in each case are consistent or exceed 0.05 after flow times of a few $t_0^\star$. Already at $\beta = 6.17$ ($a = 0.07$ fm), it is difficult to distinguish between the distributions, and at even finer lattice spacings, the distributions will be exactly preserved. This is important as it indicates that when we apply prolongation to a coarse ensemble with a well-sampled topological charge distribution, the resulting fine ensemble will continue to have a well sampled topological charge distribution, albeit with lattice artifacts inherited from the coarse level of discretization. These differences can be corrected by the fine evolution, or by improvement of the coarse action.

E. Algorithms and autocorrelations

To establish the viability of our multiscale approach, we must first determine the decorrelation time scales associated with conventional gauge evolution methods and their associated scaling behavior with the lattice spacing. In this work, we focus on two particular algorithms, namely, HB and HMC. The former is described in Sec. IV A; however, for this application we consider the parameter choices $N_{hb} = 100$ and $N_{ov} = 0$. The total number of updates is given by τ. For the later case, we use a potential-kinetic-potential (PQP)-type leapfrog algorithm, with τ trajectories of unit length and the number of leapfrog steps per trajectory tuned to yield an approximate acceptance rate of 70% for each coupling. Full details regarding the tuning of this algorithm are provided in Appendix A. Note that the strategy we take here of keeping the trajectory length fixed, rather than scaling it inversely with the lattice spacing, differs from that of [6,7]. The algorithmic
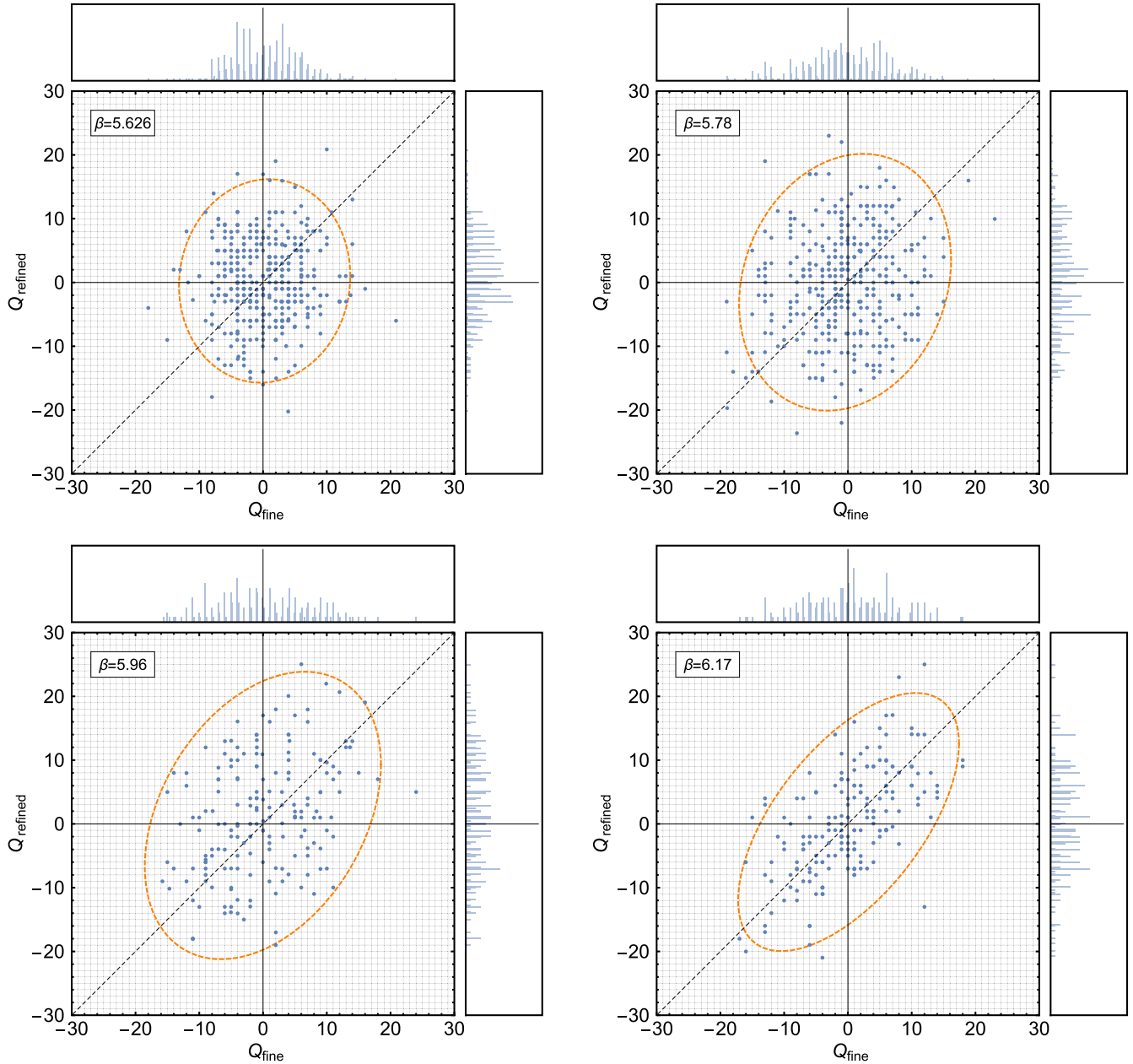
FIG. 9 (color online). Plot of Q_{fine} versus Q_{refined} determined at the maximum flow time extent for each ensemble pair. Dashed ellipses indicates 95% confidence interval centered about the mean. The skewed nature of the ellipses with respect to the diagonal suggests that the refined topological charge distribution is broader than the fine distribution.

implementations we consider are chosen to provide a relatively simple benchmark for comparison with the multiscale approach. More sophisticated and efficient implementations exist and could be used in both the traditional and the multiscale approaches, but require significant tuning and optimization. We do not pursue these directions in this initial quenched investigation.

Integrated autocorrelation times were determined for various observables using the methods described in Appendix B. Errors on the autocorrelation times were estimated using a highly efficient implementation of the jackknife method, with jackknife blocks of size N_J. Obtained errors were consistent with those obtained with analytic approximations described in [31] and based on [32,33]. In Table III, we provide details of the ensembles generated for these estimates, including the ensemble size N, jackknife block size N_J, and measurement frequency of observables $\Delta\tau$. Note that the total number of trajectories per ensemble is given by $N\Delta\tau$. In the same table, we report the autocorrelation times for the topological charge, topological susceptibility, and the quantity $t^2 E(t)$ at flow time $t_0^\star$, all of which are long distance observables. Note that the
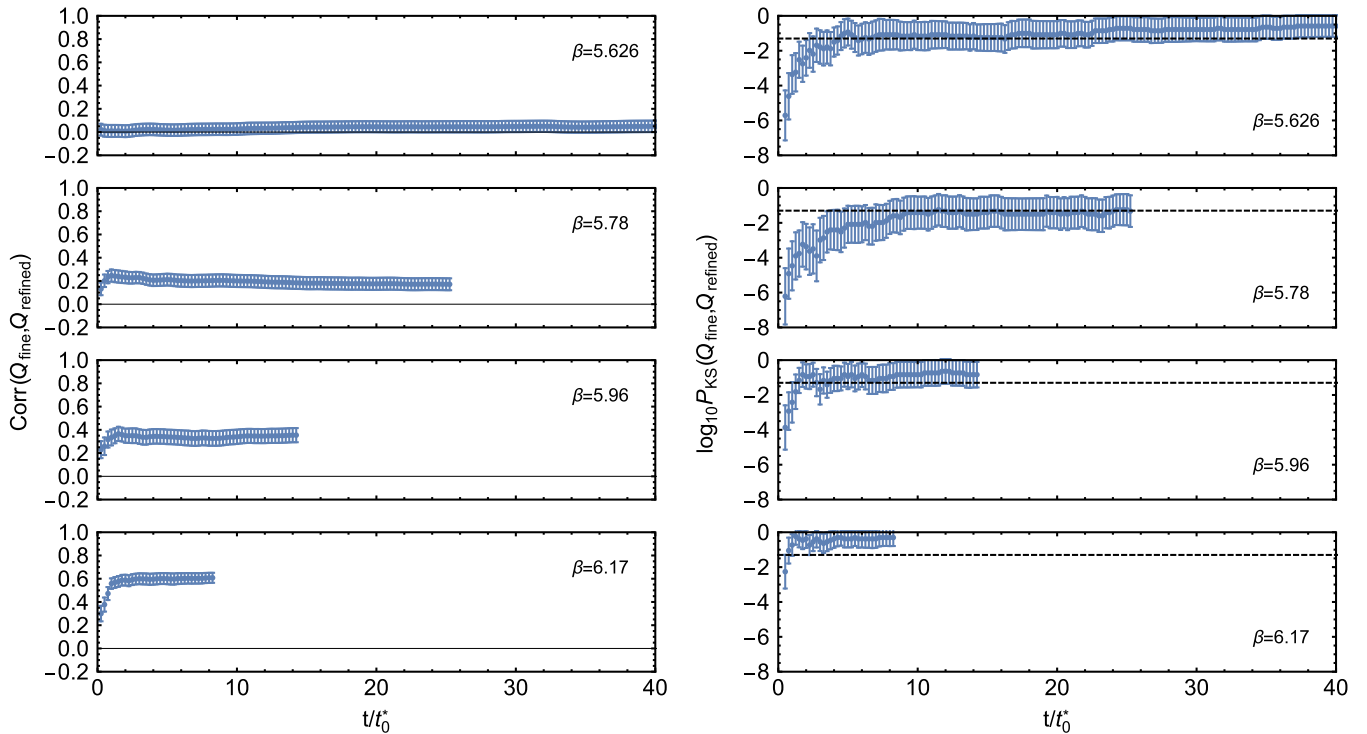
FIG. 10 (color online). Left: Correlation between the topological charge on fine and refined ensembles as a function of flow time for the four values of β considered in this work. Right: P-values obtained from a two-sample Kolmogorov-Smirnov test that the topological charge distributions on fine and refined lattices come from the same underlying distribution, as a function of the Wilson flow time. Dashed line corresponds to 0.05.

choice of $N_J \Delta \tau$ exceeds $2\tau_{\text{int}}$ estimated for each observable, suggesting a self-consistency in our error estimates. For the finer lattice spacings, we find that the integrated autocorrelation time for Wilson loops of all sizes were significantly less than that of the Wilson flow quantities provided in Table III. For 24^3 and 32^3 ensembles, our sampling resolution was insufficient to obtain reliable estimates of the integrated autocorrelation times for Wilson loops, and therefore such estimates for all β are

omitted. The integrated autocorrelation times for each observable were fit to the functional form

$$\tau_{\text{int}} = \text{const} \times \left(\frac{r_0}{a}\right)^{z_{\text{int}}}, \qquad (34)$$

and the fit results are provided in Table IV. Note that at a fixed physical volume, the computational cost to obtain decorrelated measurements of an observable is proportional
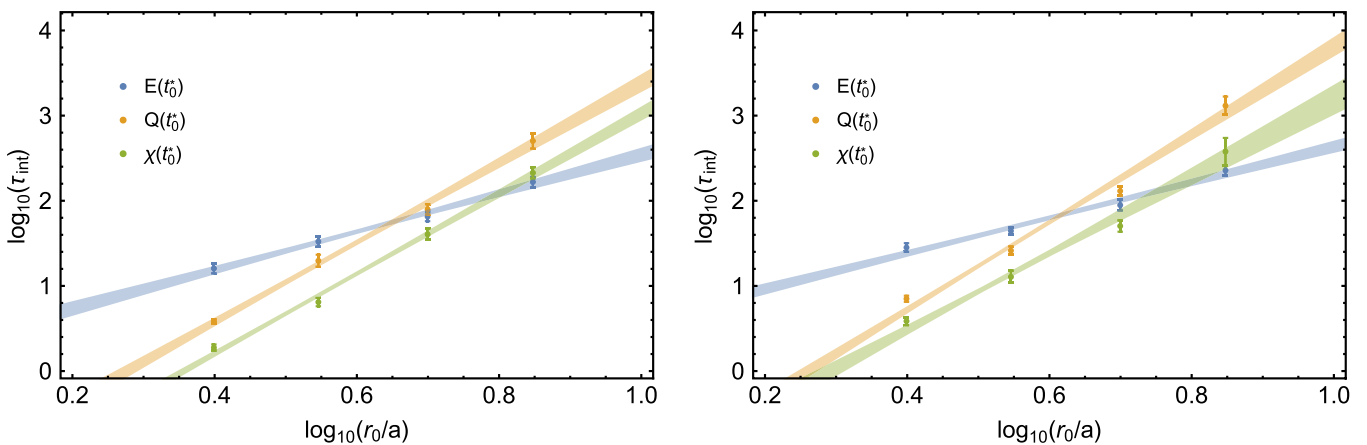

FIG. 11 (color online). Integrated autocorrelation times for Wilson flow observables as a function of r_0/a using HB (left) and HMC (right) algorithms.

TABLE III. Integrated autocorrelation times for various observables obtained from N measurements performed on every $\Delta\tau$th update.

Algorithm	Lattice	N	$\Delta\tau$	N_J	$\tau_{\text{int}}(E(t_0^\star))$	$\tau_{\text{int}}(Q(t_0^\star))$	$\tau_{\text{int}}(\chi(t_0^\star))$
HB	$12^3 \times 24$	5000	1	50	16.0(2.1)	3.8(0.3)	1.9(0.2)
	$16^3 \times 36$	12000	1	100	33.2(4.4)	19.7(3.2)	6.5(0.7)
	$24^3 \times 48$	8000	4	100	65.2(8.1)	79.4(11.1)	40.3(6.0)
	$32^3 \times 72$	9000	8	300	166.2(22.7)	504.4(103.9)	212.6(32.2)
HMC	$12^3 \times 24$	9000	1	50	28.2(3.0)	7.1(0.6)	3.9(0.4)
	$16^3 \times 36$	24000	1	100	44.2(4.7)	26.0(2.9)	12.7(2.0)
	$24^3 \times 48$	12000	4	100	88.8(12.7)	130.0(16.7)	50.5(7.8)
	$32^3 \times 72$	9000	8	300	227.3(28.5)	1307.5(322.8)	378.4(139.4)

to $\tau_{\text{int}} \times (r_0/a)^D$ for HB, and $\tau_{\text{int}} \times (r_0/a)^{D+1}$ for HMC, where $D = 4$ powers of r_0/a arise from trivial scaling of the number of lattice sites. The latter has an additional power of r_0/a due to the fact that the number of steps per trajectory needed to attain constant acceptance probability is roughly inversely proportional to the lattice spacing (see Appendix A for details). Fig. 11 shows the autocorrelation times as a function of the inverse lattice spacing, exhibiting the expected critical slowing down as the continuum limit is approached at large β. Results are shown for both HB and HMC evolution.

F. Thermalization and rethermalization

With autocorrelation times for conventional gauge evolution at hand, we are finally in a position to assess the utility of the multiscale strategy proposed in Sec. I. To make the analysis quantitative, we proceed by studying the equilibration properties of both conventionally prepared initial configurations (thermalization) and initial configurations prepared by prolongation (rethermalization). For these studies, we consider ensembles of size $N_s = 24$ under Markov evolution using the same action and algorithms as in the previous section, and the two gauge couplings, $\beta = 5.96$ and $\beta = 6.17$, corresponding to $a \sim 0.1$ fm and $a \sim 0.07$ fm, respectively. Thermalization studies were performed using initial configurations drawn from an ordered delta-function distribution (i.e., a cold start) and from a random distribution (i.e., hot start). Rethermalization studies were performed using initial configurations which had been prepared in two ways, in each case utilizing a subset of the decorrelated target ensembles described in Table I. In the first case (r-I), configurations were prepared by restriction and prolongation of $\beta = 5.96$ and $\beta = 6.17$ configurations, similar to the analysis of topological charge correlations in Sec. IV D. In

TABLE IV. Integrated autocorrelation time fit results for z_{int} for various observables.

Algorithm	$z_{\text{int}}(E(t_0^\star))$	$z_{\text{int}}(Q(t_0^\star))$	$z_{\text{int}}(\chi(t_0^\star))$
HB	2.3(2)	4.7(2)	4.7(2)
HMC	2.1(2)	5.2(2)	4.5(4)

the second case (r-II), initial configurations were prepared via prolongation of RG-matched coarse ensembles generated using the Wilson action. The matching was performed via the Sommer scale, with coarse couplings corresponding to $\beta = 5.626$ ($a \sim 0.2$ fm) and $\beta = 5.78$ ($a \sim 0.14$ fm).

Each (re)thermalization study was performed using both HB and HMC algorithms. In the latter case, acceptance probabilities were found to be exponentially small at the start of (re)thermalization. For both lattice spacings, we therefore initially evolved the ensembles for 24 trajectories without an accept/reject step in order to achieve reasonable acceptance probabilities; beyond that, evolution was performed with an accept/reject step. In all cases, the warm-up period necessary to achieve reasonable acceptance probabilities was significantly shorter than the (re)thermalization time.

We begin by considering the (re)thermalization properties of the average displaced 2×2 Wilson loops, defined in Eq. (28). Recall from Fig. 7 (right) that initially the displaced Wilson loops vary widely with the degree of displacement, d. For example, at $\beta = 6.17$, the average value ranges from approximately 0.2 at $d = 0$ to 0.8 at $d = 4$. As illustrated in Fig. 12 (top, center) for HB and Fig. 13 (top, center) for HMC, despite this wide initial variation, the displaced plaquettes converge to a single value for all d after several Monte Carlo updates. This holds true for both r-I (top) and r-II (center) ensembles, and indicates that the translational symmetry of the fine ensemble is restored rapidly as a function of the rethermalization time. In the same figures, we show the total Wilson loop (bottom) obtained by an appropriately weighted average over the five displaced loops. Here we see that the r-I ensemble "overshoots" the thermalized average, whereas the r-II ensemble converges more rapidly, and without overshooting. Although further investigation is needed to better understand these differences, it is encouraging to see that the case r-II converges so well, given that is the case where coarse ensembles had been generated (as they would be in practical applications) rather than produced artificially by restriction.

In Fig. 14 and Fig. 15, we show (re)thermalization curves for the topological susceptibility, $(t_0^\star)^2\chi(t_0^\star)$, and Wilson flow quantity $t^2E(t)$, for the flow times $t = t_0^\star/4, t_0^\star$ and
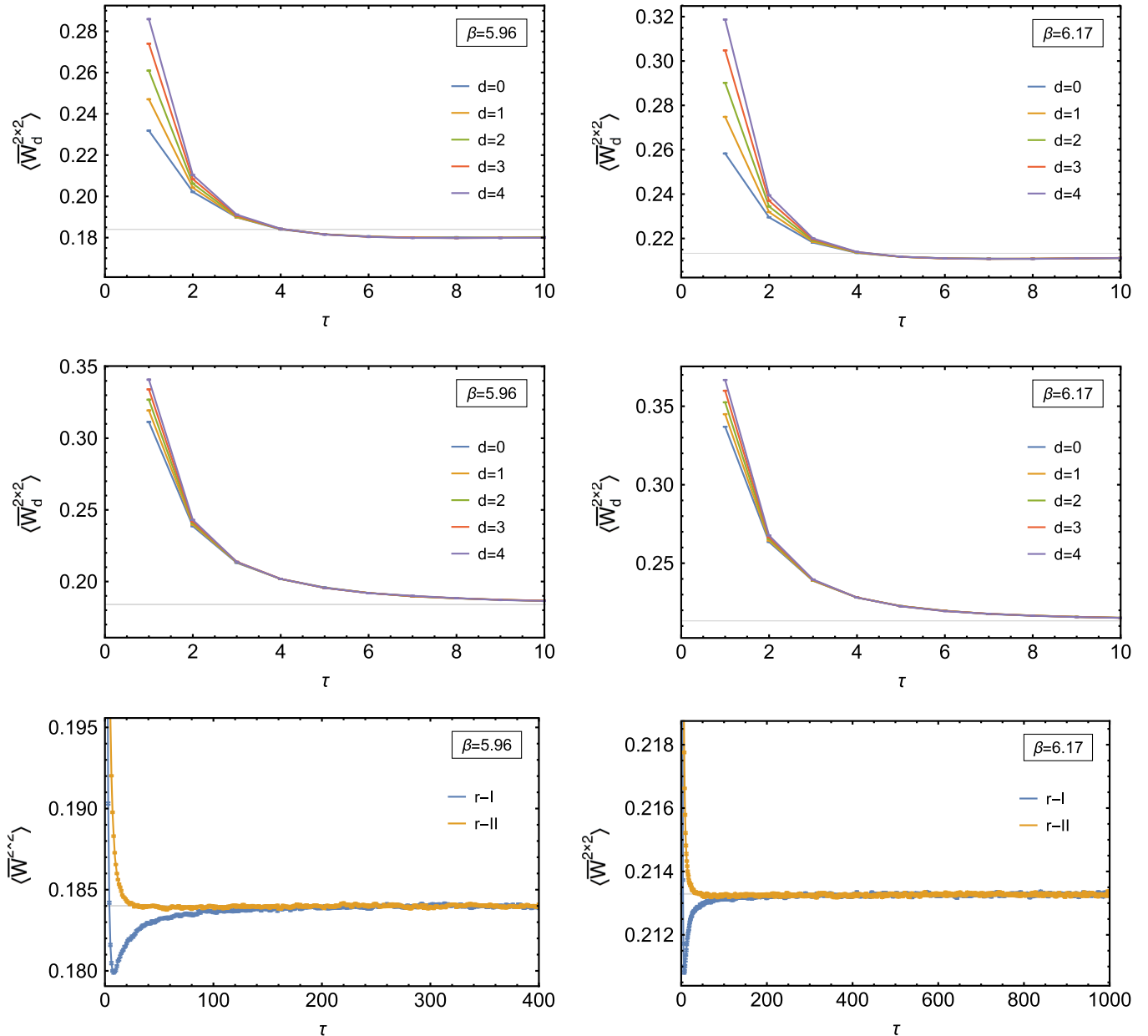
FIG. 12 (color online). Top: HB rethermalization of average displaced 2×2 Wilson loops (r-I). Center: HB rethermalization of average displaced 2×2 Wilson loops (r-II). Bottom: HB rethermalization of 2×2 Wilson loops.

$4t_0^\star$. Recall from the previous section that the topological charge distributions for these ensembles are well preserved upon prolongation. Consequently, observables derived from topology, such as the susceptibility, are by construction thermalized up to lattice artifacts. From studies of the topological susceptibility, this indeed appears to be the case. Furthermore, for nontopological quantities measured on ensembles obtained by prolongation, we find that rethermalization times appear significantly shorter than the thermalization times of either hot and cold starts.

A quantitative comparison of the (re)thermalization times for each observable requires determination of both the exponents and the overlap factors for each (re) thermalization curve. For each observable, we therefore perform a combined multiexponential fit to all four (hot, cold, r-I, and r-II) (re)thermalization curves, as a function of (re)thermalization time. We include in these fits estimates of the observable obtained from the much larger thermalized target ensemble (therm), effectively corresponding to $\tau = \infty$. We considered fits of the functional form

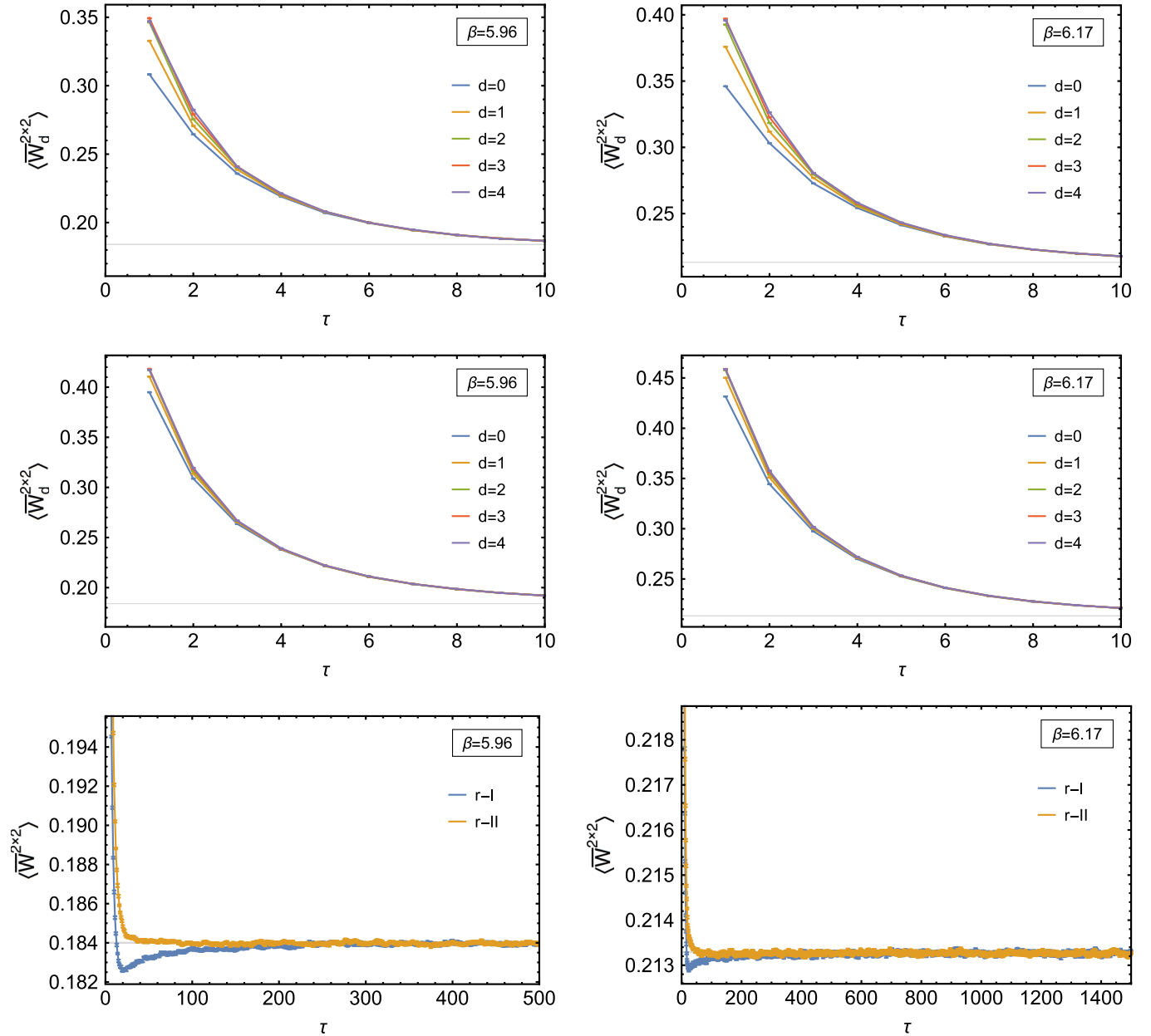$$f^\alpha(\tau) = \sum_{n=0}^{2} z_n^\alpha e^{-E_n \tau}, \qquad (35)$$

FIG. 13 (color online).   Top: HMC rethermalization of average displaced 2×2 Wilson loops (r-I). Center: HMC rethermalization of average displaced 2×2 Wilson loops (r-II). Bottom: HMC rethermalization of 2×2 Wilson loops.

where α labels the ensemble, and impose the constraints $E_0 = 0$, $z_0^\alpha = z_0$ for all α and $z_n^{\text{therm}} = 0$ for $2 \geq n > 0$. The least squares fits were performed using the variable projection method [34]. A brief description of how the constraints were imposed in this approach is provided in Appendix C. Reliable correlated multiexponential fits to the data were difficult to achieve due to the small ensemble sizes; consequently uncorrelated fits to data were performed, with errors estimated via a bootstrap analysis. Our aim is to obtain estimates of the relevant evolution time scales rather than precise values, so this simplified analysis is sufficient. With larger ensembles and more

frequent measurements, coupled fits to multiple observables could be performed, potentially constraining higher states in the evolution. The leading exponents $\tau_n = 1/E_n$ ($n = 1$, 2) determined from each fitted observable are provided in Table V at flow time t_0. Note that generally $\tau_1 \neq \tau_{\text{exp}}$, since the observable under study may not couple strongly enough to the slowest mode, and furthermore, the statistics may be insufficient to resolve the effects of that mode. Nonetheless, we expect the bound $\tau_{\text{int}} \leq \tau_1$ to hold for each observable that is considered. A comparison of these time scales can be made from the data provided in Table III and Table V, and it suggests that this is indeed the case.
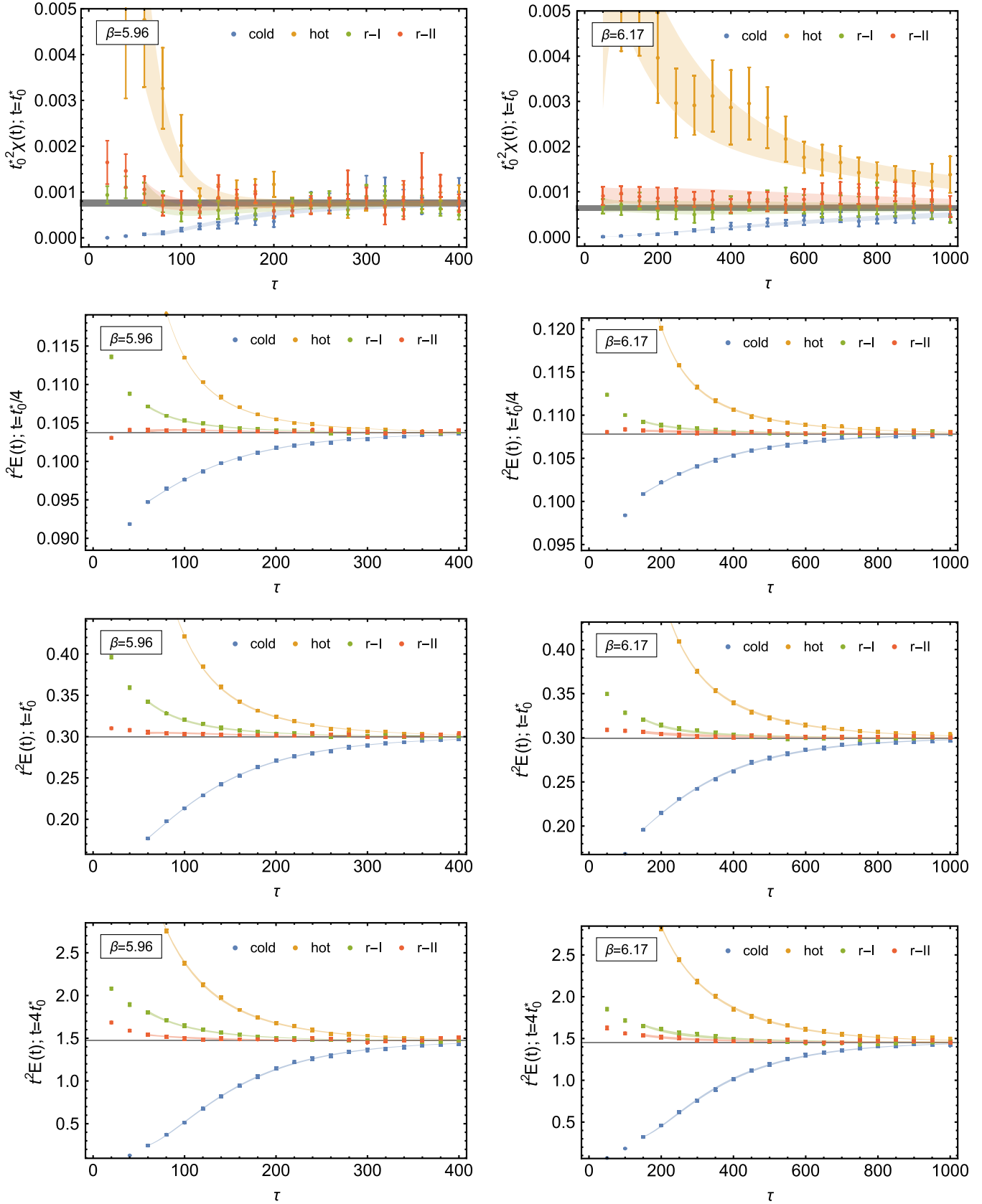
FIG. 14 (color online). HB (re)thermalization of topological susceptibility and $E(t)$ as a function of the number of sweeps, for various values of t.
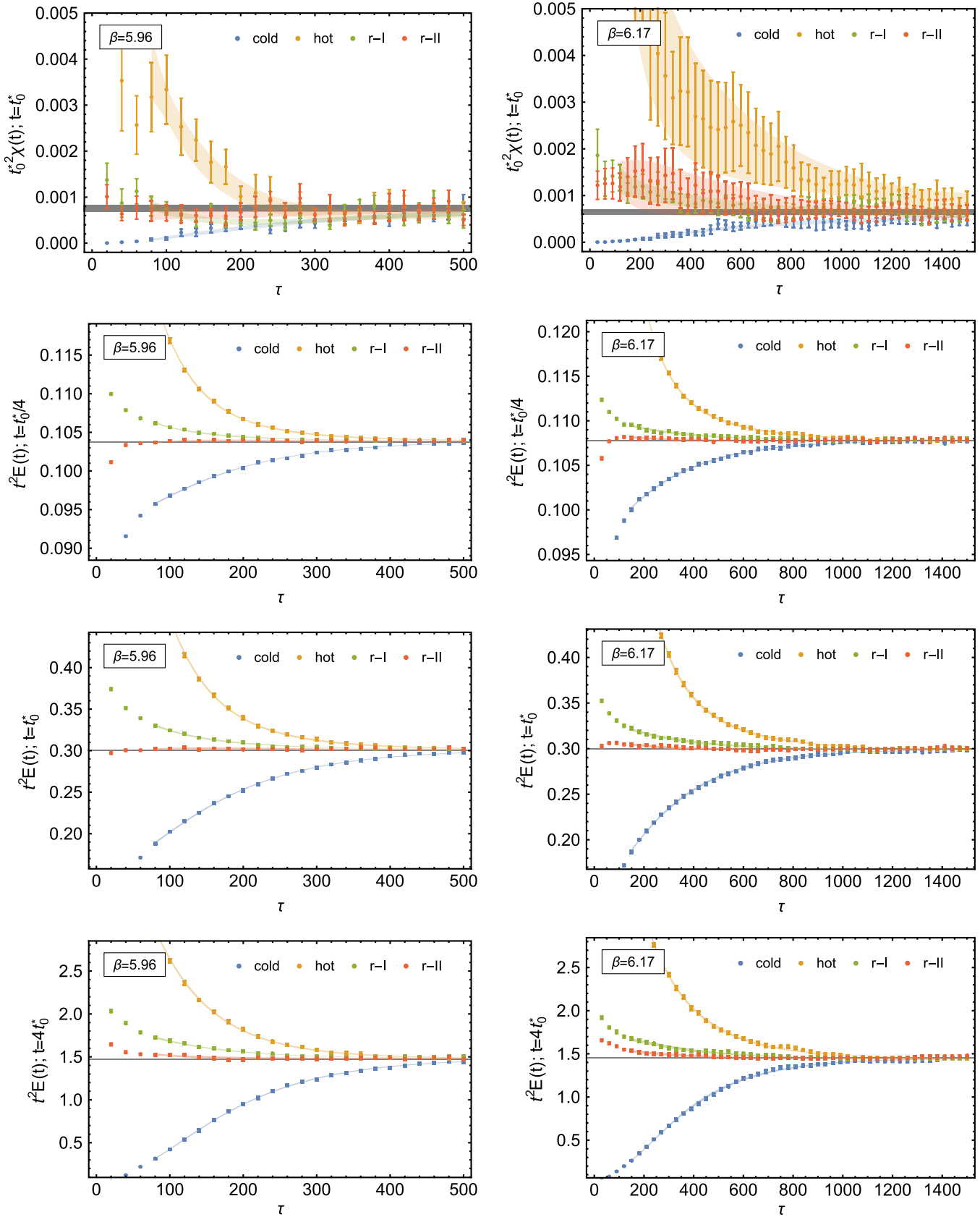
FIG. 15 (color online).   HMC (re)thermalization of topological susceptibility and $E(t)$ as a function of the number of trajectories, for various values of t.

TABLE V.   Fit results for (re)thermalization curves.

Algorithm	Lattice	τ range	$\tau_1(E(t_0^\star))$	$\tau_2(E(t_0^\star))$	τ range	$\tau_1(\chi(t_0^\star))$	$\tau_2(\chi(t_0^\star))$
HB	$24^3 \times 48$	60–400	89.1(2.3)	32.5(1.2)	60–400	95.9(49.9)	24.9(8.9)
	$32^3 \times 72$	150–1000	219.3(7.6)	72.2(4.5)	50–1000	623.6(84.9)	140.4(149.6)
HMC	$24^3 \times 48$	80–500	115.6(4.0)	53.2(5.4)	80–500	187.5(117.3)	86.9(33.0)
	$32^3 \times 72$	150–1500	250.3(10.8)	96.5(6.2)	120–1500	511.2(227.9)	83.8(52.1)

Finally, in Fig. 16 we plot extracted values for τ_n and z_n^α as a function of the Wilson flow time for the observable $t^2E(t)$, for $32^3 \times 72$ ensembles (re)thermalized via HMC. For this illustrative case, the fitted exponents are insensitive to the flow time, with $\tau_1/\tau_2 \sim 2.5$. This stability suggests that the fits are picking out the true exponents governing the evolution dynamics. The overlap factors, on the other hand, need not be independent of the flow time. For flow times $t/t_0^\star > 1$ we find $z_1/z_0 \gtrsim \mathcal{O}(1)$ and $z_2/z_0 \gtrsim \mathcal{O}(1)$ for hot and cold ensembles. The ensemble r-I has significantly reduced overlaps by comparison, with z_2/z_0 consistent with zero over the full range of flow times. The ensemble r-II exhibits the most impressive behavior, with both z_1/z_0 and z_2/z_0 consistent with zero over the full flow time range. The result suggests a lower bound on the rethermalization time scale, given by τ_3. It would be particularly interesting

in this example to determine with higher precision the number of low lying states that have been eliminated, thus further constraining this bound.

From analysis of the autocorrelation times in the preceding section, and the (re)thermalization time scales determined here, we may draw several conclusions. First, the rethermalization times for prolonged ensembles are significantly shorter than the thermalization times for hot and cold starts. This result implies that the simulation strategy advocated in Fig. 2(b) is more efficient than that of Fig. 2(a). Second, the rethermalization times for nontopological long-distance observables are significantly shorter than the decorrelation time scale for fine evolution, which is bounded from below by twice the integrated autocorrelation time for topological charge. An explicit comparison of these time scales can be made from Table III and Table V.
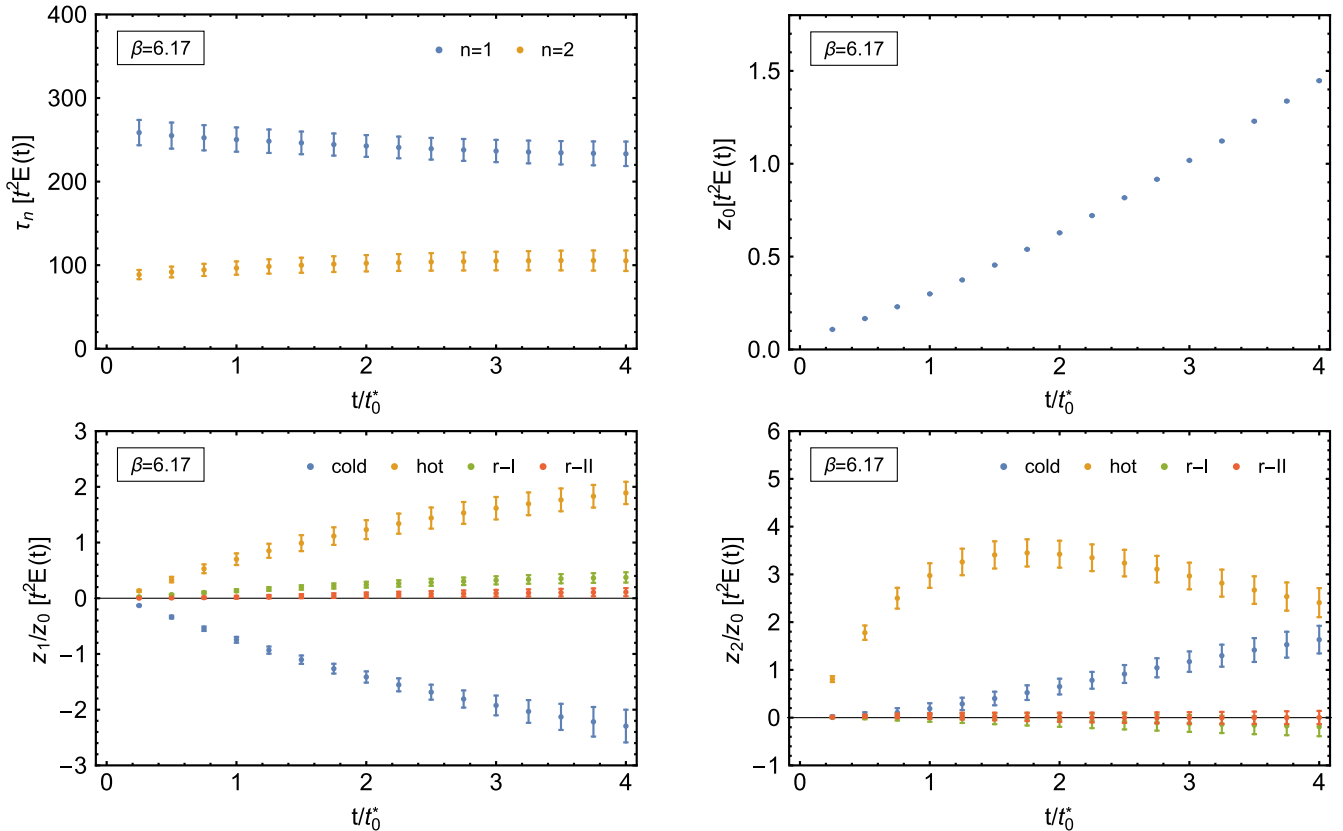

FIG. 16 (color online).   Multiexponential fit results (decay constants and overlap factors) for HMC (re)thermalization of $E(t)$ as a function of the number of trajectories, for various values of t.

For example, in the case r-II (HMC, $a \sim 0.07$ fm), where τ_3 (which was undetermined from fits) appears to be the dominant rethermalization time scale, rethermalization to quarter-percent levels can be achieved in the time $\tau_{\text{retherm}} \sim 6\tau_3 < 6\tau_2 \sim 600$. The decorrelation time for the topological charge provides a lower bound $2\hat{\tau}_{\text{int}} > 2\tau(Q(t_0^*)) \sim 2600$, and therefore the efficiency of the algorithm, as described by Fig. 2(c), is conservatively estimated to be greater than $2\hat{\tau}_{\text{int}}/\tau_{\text{retherm}} \sim 4$.[5] The result provides compelling evidence that an ensemble generation strategy along the lines of Fig. 2(c) is not only viable but also a superior alternative to approaches presently available. Note that this assessment becomes particularly conclusive in the regime of ultrafine lattice spacings, where proper sampling of the topological charge is presently impractical by conventional means due to topological freezing. As mentioned in Sec. I, the two strategies described in Fig. 2(b) and Fig. 2(c) represent only the extremes in a range of algorithms, defined by different choices of N_s and N_e. In general, the optimal choice for these parameters depends on the time scales observed for a particular target action and the computational facilities that are available.

V. CONCLUSION

We have proposed a multiscale equilibration strategy for Yang-Mills gauge theories, which can be used to rapidly initialize large numbers of Monte Carlo streams, thereby increasing the efficiency of simulations. This algorithm shares many features with multigrid solvers [35–38] which have been used to dramatically decrease the computational cost of matrix inversion, a large component of lattice QCD calculations. The effectiveness of our multilevel strategy for equilibration was demonstrated for the case of pure $SU(3)$ gauge theory using both heat bath and hybrid Monte Carlo updating procedures. In both cases, the time scales governing the rethermalization of prolonged ensembles were measured to be considerably shorter than the decorrelation times for conventional evolution, as estimated from the autocorrelations in the topological charge. Furthermore, the ratio of these quantities decreases parametrically as the continuum limit is approached. As a consequence, prolongation of a RG-matched coarse ensemble followed by rethermalization provides a new way to reduce critical slowing down in lattice gauge theory simulations. In particular, the poor sampling of topology at fine lattice spacing is ameliorated by evolving multiple fine-action streams derived from a coarse-action ensemble with well-sampled topology. Although unexplored in this study, multiple levels of refinement would offer additional speed-ups in thermalization.

---

[5]This accounting does not include any additional reduction in computational cost attributed to reduced communication overhead from having multiple streams.

The successful application of our strategy requires a nonperturbative real space renormalization group procedure to match the physical scales at the coarse and fine levels of evolution. Generally the matching need not be precise to realize improvements since the subsequent evolution (rethermalization) eliminates any effects of the mistuning of the coarse action. However, the precision with which the tuning is carried out will influence the rethermalization times of the prolongated ensembles. A RG transformation of the fine lattice action induces many operators in the coarse action, and these should, in principle, be included in the coarse evolution. In this study, we have ignored all but the local plaquette coupling in the coarse action, yet still attain impressively short rethermalization times. Numerical methods have been developed for nonperturbatively determining the induced couplings along a RG flow for simple systems (see, e.g., [39]), and their use in tuning would likely result in a further reduction in rethermalization times.

From a practical standpoint, the utility of our approach can be realized in several ways. The method can be directly applied toward generation of very large physical volume ensembles in cases where the matching is already known (e.g., from previous small volume studies). In this way one can avoid the long initialization time for large lattice streams. Our expectation is that for a given target lattice spacing, rethermalization times for prolongated ensembles will be insensitive to volume, and therefore the efficiency of the algorithm will be unaffected by the volume scaling. Alternatively the method can be used to start random ensembles on coarse lattices to initialize a large ensemble of fully independent streams. These streams will start with different topological charges so that together they will sample the topological charge distribution dictated by the coarse lattice. In principle, the topological charge distribution could be reweighted if the continuum distribution is determined by some other means.

A potential weakness of our multilevel approach is that its effective use requires prior knowledge of the RG matching of actions, which in turn requires simulations at the fine level. Nevertheless, there are several ways in which to proceed. One can exploit the fact that the tuning need not be exact at the coarse level and perform "sloppy matching" studies using small volume ensembles with poorly sampled topology; in this case matching would proceed by considering long distance observables which are relatively insensitive to topology. Another possibility is to use a finite volume scheme such as the Schrödinger functional [40] to define the gauge coupling. By requiring a fixed coupling constant for a coarse and fine ensemble, one can then obtain a matching of the bare gauge couplings that result from calculations at two lattice spacings with the desired ratio. Available results that could be used to obtain matched coarse and fine ensembles, including dynamical fermions, can be found in a recent review [41]. Finally, in

the ultrafine limit, one can make contact with perturbative calculations in order to carry out the matching.

The multilevel methods described here naturally extend to simulations of full QCD and, more generally, gauge theories with matter fields. The presence of fermions has no impact on the details of the prolongation. The only additional ingredient is that the RG matching must be performed for more than one physical scale (e.g., Λ_{QCD}, the pion mass and the kaon mass for $2 + 1$ flavor QCD), and therefore requires tuning of multiple parameters in the gauge and Dirac actions. In cases such as this where multiple physical scales are present, rapid thermalization with multiple levels of refinement may be particularly advantageous. Perfect action constructions [42] may be useful in this regard.

Although the multiscale algorithm presented here provides an efficient means for thermalization, it is important to draw a distinction between it and a more ambitious multigrid Monte Carlo dynamics. The latter implements a fully recursive evolution, including multiple scale evolution, while maintaining exact detailed balance on the finest level. While there has been some success in constructing such multiscale methods for simpler field theories [8–13], it is an open challenge to construct such an algorithm, particularly for QCD, due to the presence of a nonlocal fermion determinant in the measure.

ACKNOWLEDGMENTS

APPENDIX A: HMC TUNING

Acceptance probabilities for each HMC step size $\delta\tau$ and β were estimated from 100 independent trajectories, starting from thermalized configurations. Estimated acceptance rates are displayed in Fig. 17 (left). The estimated acceptance rates were then fit to the expected functional form

$$P_{\rm acc} = {\rm erfc}\left(\frac{1}{2}\Omega^{1/2}\delta\tau^2\right); \qquad (A1)$$

the extracted fit values for $\Omega^{1/2}$ are provided in Table VI in units of the scale t_0. Fitting the extracted values of $\Omega^{1/2}$ as a function of scale parameter t_0, we find $\Omega^{1/2}(t_0) = 1096(14)t_0$. For each of the HMC studies in this work, the trajectory length is fixed to unity, and the acceptance rate is chosen to be approximately 70%. The nominal number of steps per trajectory, $\delta\tau^\star$, used in these studies, is provided in Table VI. As a function of the lattice spacing, the requisite number of steps per trajectory length to achieve such an acceptance rate exhibits a power-law behavior, as demonstrated in Fig. 17 (right). Fitting the data, we find $1/\delta\tau \propto (r_0/a)^{1.012(17)}$. In light of the fact that the HMC algorithm is not renormalizable [44], the scaling behavior observed in this study may be regarded as empirical in nature.
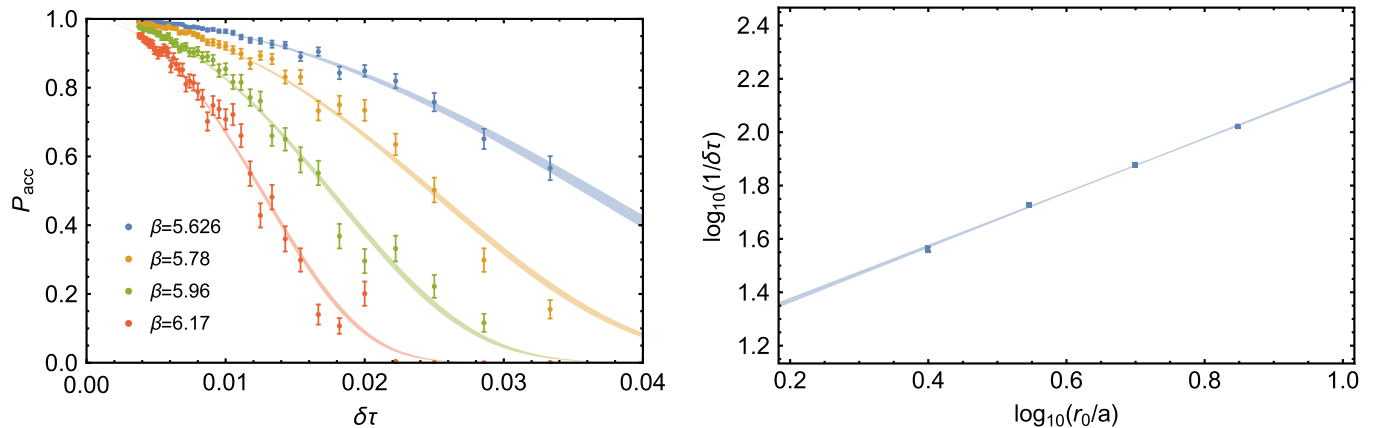

FIG. 17 (color online). Left: HMC acceptance probabilities as a function of step size, $\delta\tau$, for a fixed trajectory of unit length. Right: Number of steps per trajectory ($1/\delta\tau$) as a function of lattice spacing, for a fixed acceptance probability of 70%.

TABLE VI. Fit parameters for $\Omega^{1/2}$ in units of t_0, estimated HMC step size $\delta\tau$, and nominal HMC step size $\delta\tau^\star$ required to achieve a 70% acceptance rate, for a trajectory of unit length.

Lattice	β	$\Omega^{1/2}/t_0$	$1/\delta\tau$	$1/\delta\tau^\star$
$12^3 \times 24$	5.626	1044(39)	36.5(0.7)	37
$16^3 \times 36$	5.78	1138(28)	53.3(0.5)	54
$24^3 \times 48$	5.96	1084(27)	75.5(0.9)	75
$32^3 \times 72$	6.17	1098(23)	105.1(1.1)	105

APPENDIX B: AUTOCORRELATIONS

Consider a set of data, X, comprising arbitrarily spaced measurements x_τ, labeled by $\tau \in [1, N]$. The autocorrelation function at lag-time Δ is defined by

$$\rho_\Delta(X) = \frac{\Gamma_\Delta(X)}{\Gamma_0(X)}, \tag{B1}$$

where $\Gamma_\Delta(X)$ is the corresponding autocovariance function for X. To estimate $\Gamma_\Delta(X)$, first consider the set δX comprising the elements $\delta x_\tau = x_\tau - \bar{x}$ for all τ, where

$$\bar{x} = \frac{1}{N} \sum_{\tau=1}^{N} x_\tau, \tag{B2}$$

and the function

$$\hat{\Gamma}_\Delta(\delta X) = \frac{1}{N} \sum_{\tau=1}^{N-\Delta} \delta x_\tau \delta x_{\tau+\Delta}, \tag{B3}$$

defined on this set. If X consists of uniformly sampled data, then the autocovariance function is simply given by

$$\Gamma_\Delta(X) = \frac{N}{N-\Delta} \hat{\Gamma}_\Delta(\delta X). \tag{B4}$$

Note that the computational cost of evaluating $\hat{\Gamma}_\Delta(\delta X)$ is $\mathcal{O}(N \log N)$ using standard discrete Fourier transform methods, whereas the computational cost for evaluating $\Gamma_\Delta(X)$ for arbitrarily spaced samples is generally $\mathcal{O}(N^2)$.

Assuming X consists of uniformly spaced data, the errors on $\Gamma_\Delta(X)$ can be computed efficiently via the following jackknife procedure. First partition the measurements x_τ into N/N_J consecutive blocks of size N_J, where it is assumed that $N \bmod N_J$ equals zero. Labeling the partitions by the integers $j \in [1, N/N_J]$, define the jackknife ensemble X^j comprising the $N - N_J$ elements of X, with elements x_τ on the interval $\tau \in ((j-1)N_J, jN_J]$ omitted. Furthermore, define the set δX^j which comprises elements

$$\delta x_\tau^j = \begin{cases} 0 & (j-1)N_J < \tau \leq jN_J \\ x_\tau - \bar{x}^j & \text{otherwise} \end{cases}. \tag{B5}$$

Note that although X^j comprises $N - N_J$ elements, δX^j comprises N elements, of which N_J vanish. The autocorrelation function on the jth jackknife ensemble is given by

$$\Gamma_\Delta(X^j) = \frac{N}{g_\Delta(X^j)} \hat{\Gamma}_\Delta(\delta X^j), \tag{B6}$$

where the piecewise function $g_\Delta(X^j)$ quantifies the degeneracy of distances Δ for the ensemble X^j. If the set of integers m_α for $\alpha \in [1, 5]$ label the quantities $\{N_J, (j-1)N_J, jN_J, N - jN_J, N - (j-1)N_J\}$ in ascending order, then

$$g_\Delta(X^j) = \begin{cases} N - N_J - 2\Delta & 0 \leq \Delta < m_1 \\ N - N_J - m_1 - \Delta & m_1 \leq \Delta < m_2 \\ N - N_J - m_1 - m_2 & m_2 \leq \Delta < m_3 \\ N - N_J - m_1 - m_2 + m_3 - \Delta & m_3 \leq \Delta < m_4 \\ N - N_J - m_1 - m_2 + m_3 - m_4 & m_4 \leq \Delta < m_5 \\ N - \Delta & m_5 \leq \Delta < N \end{cases}. \tag{B7}$$

Once the jackknife estimates $\rho_\Delta(X^j)$ are obtained, the standard error is determined by

$$\delta\rho_\Delta^2(X) = \frac{N}{N-N_J} \sum_{j=1}^{N/N_J} [\rho_\Delta(X^j) - \rho_\Delta(X)]^2. \tag{B8}$$

The integrated correlation time is given by

$$\tau_{\text{int}}(X) = \frac{1}{2} + \sum_{\Delta=1}^{\Delta_{\max}} \rho_\Delta(X), \tag{B9}$$

where the cutoff lag-time $\Delta_{\max}$ is defined as the minimum time at which

$$\rho_\Delta(X) > \sqrt{\delta\rho_\Delta^2(X)} \tag{B10}$$

following [31]. Once $\Delta_{\max}$ is selected, the errors on τ_{int} may be determined from the jackknife estimates $\tau_{\text{int}}(X^j)$ via

$$\delta\tau_{\text{int}}^2(X) = \frac{N}{N-N_J} \sum_{j=1}^{N/N_J} [\tau_{\text{int}}(X^j) - \tau_{\text{int}}(X)]^2. \tag{B11}$$

APPENDIX C: VARIABLE PROJECTION WITH CONSTRAINTS

Assume we have a set of measurements $y^\alpha(\tau)$ and covariance matrix $\Gamma^{\alpha\beta}(\sigma, \tau)$. Given the fit function

$$f^\alpha(\tau) = \sum_n z_n^\alpha e^{-E_n \tau}, \tag{C1}$$

and the constraint $\phi z = 0$, we may construct

$$\chi^2(z, E) = \sum_{\alpha\beta} \sum_{\sigma\tau} [y^\alpha(\sigma) - f^\alpha(\sigma)] V^{\alpha\beta}(\sigma, \tau)[y^\alpha(\tau) - f^\alpha(\tau)] + \sum_{\alpha\beta} \sum_{mn} \xi_m^\alpha \phi_{mn}^{\alpha\beta} z_n^\beta, \tag{C2}$$

where $V = \Gamma^{-1}$, and ξ_n^α are Lagrange multipliers for each constraint. Note that ϕ is generally a rectangular matrix.

Next, we may express Eq. (C2) as

$$\chi^2(z, E) = \chi^2(0,0) - 2\sum_\alpha \sum_n z_n^\alpha u_n^\alpha(E)$$
$$+ \sum_{\alpha\beta} \sum_{mn} z_m^\alpha W_{mn}^{\alpha\beta}(E) z_n^\beta + \sum_{\alpha\beta} \sum_{mn} \xi_m^\alpha \phi_{mn}^{\alpha\beta} z_n^\beta,$$

$$(C3)$$

where

$$u_n^\alpha(E) = \sum_\beta \sum_{\sigma\tau} e^{-E_n\sigma} V^{\alpha\beta}(\sigma\tau) y^\beta(\tau), \qquad (C4)$$

and

$$W_{mn}^{\alpha\beta}(E) = \sum_{\sigma\tau} e^{-E_m\sigma} V^{\alpha\beta}(\sigma, \tau) e^{-E_n\tau}. \qquad (C5)$$

Minimizing this function with respect to z yields

$$z(E) = W^{-1}(E)u(E)$$
$$- W^{-1}(E)\phi^\intercal[\phi W^{-1}(E)\phi^\intercal]^{-1}\phi W^{-1}(E)u(E), \quad (C6)$$

where indices α and n have been suppressed. One can then construct a reduced χ^2, which is only a function of E and given by

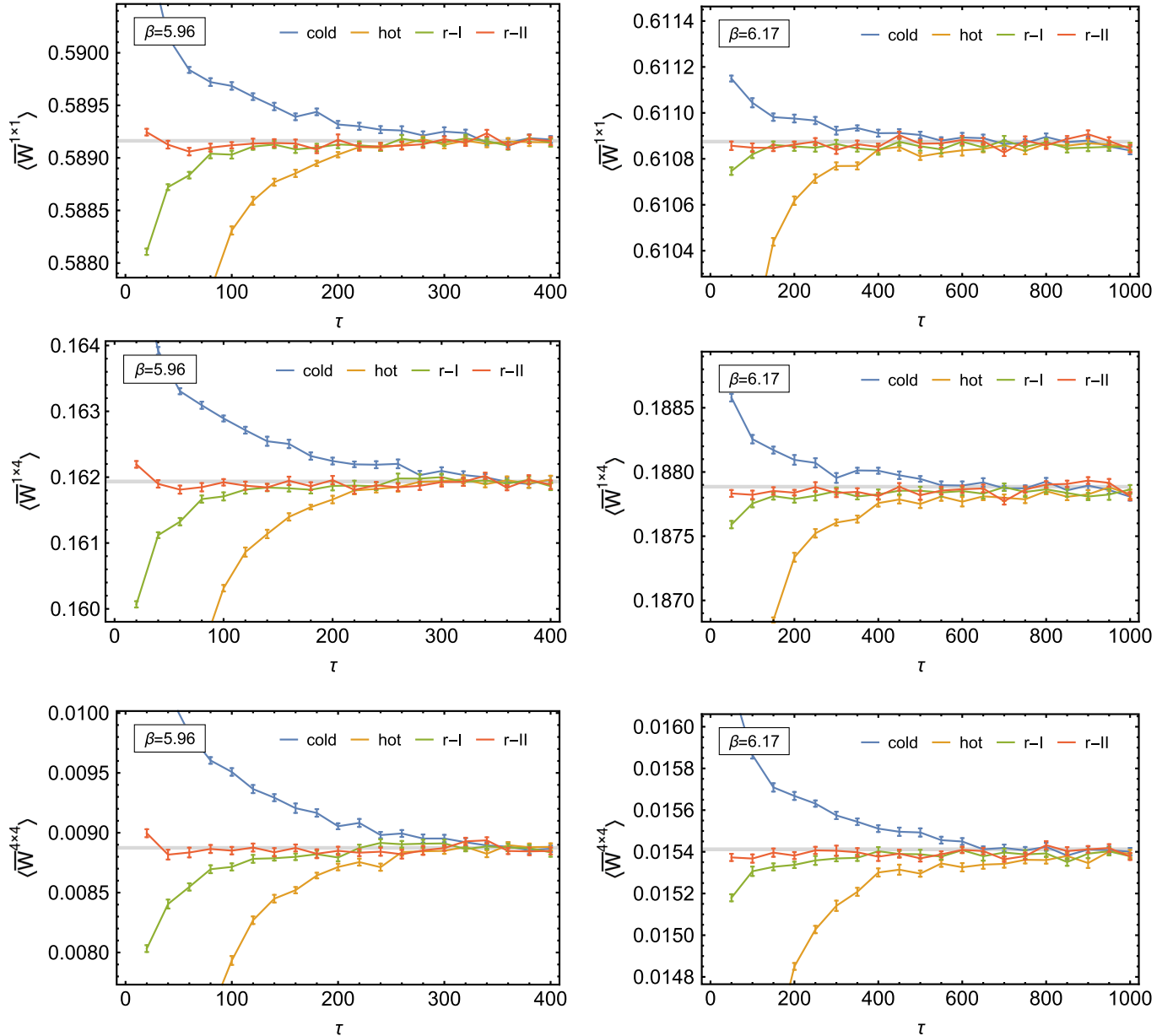$$\chi_r^2(E) = \chi^2(z(E), E). \qquad (C7)$$

FIG. 18 (color online).   HB (re)thermalization of space-time Wilson loops W as a function of the number of sweeps.

Numerical minimization of $\chi_r^2(E)$ proceeds with standard methods, yielding E_{min}; the corresponding overlaps $z(E_{min})$ can be reconstructed from Eq. (C6).

APPENDIX D: THERMALIZATION AND RETHERMALIZATION OF WILSON LOOPS

(Re)thermalization curves for Wilson loops of various shapes are provided in Fig. 18 for HB and Fig. 19 for HMC. As with the Wilson flow quantity $t^2E(t)$, we find that the

rethermalization time for the r-II ensemble is significantly shorter than that of r-I. Furthermore, both rethermalization times are shorter than their hot and cold counterparts. The (re)thermalization times for Wilson loops generally appear to be significantly shorter than that of $t^2E(t)$, even for physically large Wilson loops of size 0.5 fm. This is likely attributed to the fact that our prolongator preserves all even size Wilson loops originating at sites **n** satisfying $\chi(\mathbf{n}) = 0$.
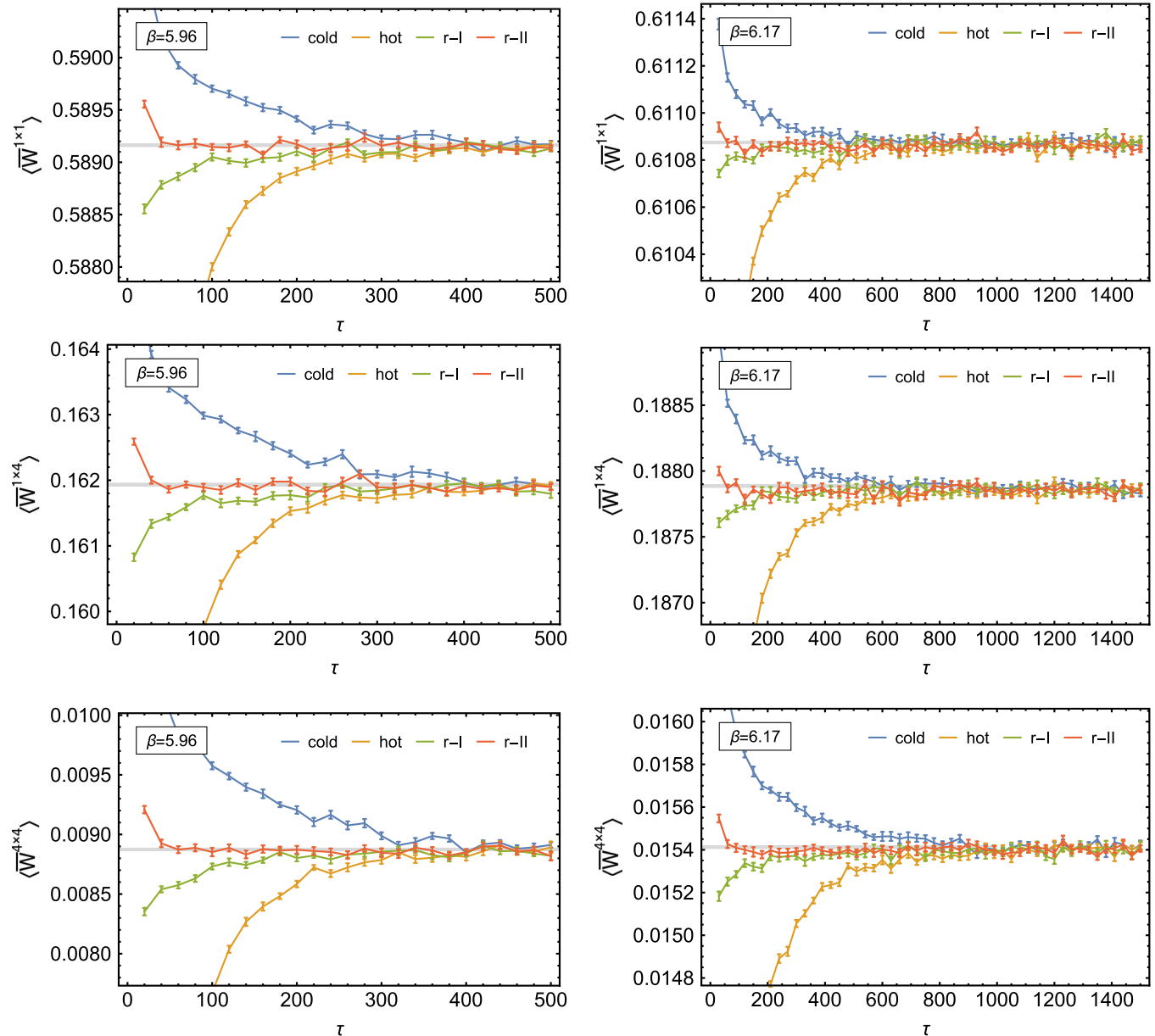

FIG. 19 (color online).   HMC (re)thermalization of space-time Wilson loops W as a function of the number of trajectories.

[1] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[2] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).

[3] N. Prokof'ev and B. Svistunov, Phys. Rev. Lett. **87**, 160601 (2001).

[4] Y. Aoki, T. Blum, N. H. Christ, C. Dawson, T. Izubuchi, R. D. Mawhinney, J. Noaki, S. Ohta, K. Orginos, A. Soni, and N. Yamada, Phys. Rev. D **73**, 094507 (2006).

[5] S. Schaefer, R. Sommer, and F. Virotta (ALPHA Collaboration), Nucl. Phys. **B845**, 93 (2011).

[6] M. Lüscher and S. Schaefer, J. High Energy Phys. 07 (2011) 036.

[7] G. McGlynn and R. D. Mawhinney, Phys. Rev. D **90**, 074502 (2014).

[8] J. Goodman and A. D. Sokal, Phys. Rev. Lett. **56**, 1015 (1986).

[9] R. G. Edwards, J. Goodman, and A. D. Sokal, Nucl. Phys. **B354**, 289 (1991).

[10] R. G. Edwards, S. J. Ferreira, J. Goodman, and A. D. Sokal, Nucl. Phys. **B380**, 621 (1992).

[11] M. Grabenstein and B. Mikeska, Nucl. Phys. B, Proc. Suppl. **34**, 765 (1994).

[12] W. Janke and T. Sauer, Nucl. Phys. B, Proc. Suppl. **34**, 771 (1994).

[13] M. Grabenstein and K. Pinn, Phys. Rev. D **50**, 6998 (1994).

[14] M. Lüscher, arXiv:1002.4232.

[15] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).

[16] M. Lüscher, Commun. Math. Phys. **85**, 39 (1982).

[17] A. Phillips and D. Stone, Commun. Math. Phys. **103**, 599 (1986).

[18] G. 't Hooft, Phys. Lett. B **349**, 491 (1995).

[19] M. Falcioni, M. Paciello, G. Parisi, and B. Taglienti, Nucl. Phys. **B251**, 624 (1985).

[20] M. Albanese *et al.*, Phys. Lett. B **192**, 163 (1987).

[21] R. Edwards, U. Heller, and T. Klassen, Nucl. Phys. **B517**, 377 (1998).

[22] M. Guagnelli, R. Sommer, and H. Wittig, Nucl. Phys. **B535**, 389 (1998).

[23] N. Cabibbo and E. Marinari, Phys. Lett. **119B**, 387 (1982).

[24] F. R. Brown and T. J. Woch, Phys. Rev. Lett. **58**, 2394 (1987).

[25] M. Creutz, Phys. Rev. Lett. **43**, 553 (1979); **43**, 890(E) (1979).

[26] R. Narayanan and H. Neuberger, J. High Energy Phys. 03 (2006) 064.

[27] M. Lüscher, J. High Energy Phys. 08 (2010) 071.

[28] R. Lohmayer and H. Neuberger, *Proc. Sci.*, LATTICE2011 (2011) 249 [arXiv:1110.3522].

[29] P. Fritzsch and A. Ramos, J. High Energy Phys. 10 (2013) 008.

[30] P. de Forcrand, M. G. Pérez, and I.-O. Stamatescu, Nucl. Phys. **B499**, 409 (1997).

[31] M. Lüscher, Comput. Phys. Commun. **165**, 199 (2005).

[32] N. Madras and A. D. Sokal, J. Stat. Phys. **50**, 109 (1988).

[33] U. Wolff (ALPHA Collaboration), Comput. Phys. Commun. **156**, 143 (2004); **176**, 383(E) (2007)].

[34] G. Golub and V. Pereyra, Inverse Probl. **19**, R1 (2003).

[35] R. Babich, J. Brannick, R. C. Brower, M. A. Clark, T. A. Manteuffel, S. F. McCormick, J. C. Osborn, and C. Rebbi, Phys. Rev. Lett. **105**, 201602 (2010).

[36] R. Babich, J. Brannick, R. C. Brower, M. A. Clark, S. D. Cohen, J. C. Osborn, and C. Rebbi, *Proc. Sci.*, LAT2009 (2009) 031 [arXiv:0912.2186].

[37] A. Frommer, K. Kahl, S. Krieg, B. Leder, and M. Rottmann, SIAM J. Sci. Comput. **36**, A1581 (2014).

[38] J. Brannick, A. Frommer, K. Kahl, B. Leder, M. Rottmann, and A. Strebel, Numer. Math. 1 (2015).

[39] M. Hasenbusch, K. Pinn, and C. Wieczerkowski, Phys. Lett. B **338**, 308 (1994).

[40] M. Lüscher, S. Sint, R. Sommer, P. Weisz, and U. Wolff, Nucl. Phys. **B491**, 323 (1997).

[41] R. Sommer and U. Wolff, Nucl. Part. Phys. Proc. **261–262**, 155 (2015).

[42] P. Hasenfratz and F. Niedermayer, Nucl. Phys. **B414**, 785 (1994).

[43] R. G. Edwards and B. Joo, Nucl. Phys. B, Proc. Suppl. **140**, 832 (2005).

[44] M. Lüscher and S. Schaefer, J. High Energy Phys. 04 (2011) 104.