Topological susceptibility of the 2D O(3) model under gradient flow

Wolfgang Bietenholz,[1,2,*] Philippe de Forcrand,[3] Urs Gerber,[2] Héctor Mejía-Díaz,[1] and Ilya O. Sandoval[1]

[1]*Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, A.P. 70-543, C.P. 04510 Ciudad de México, Mexico*

[2]*Albert Einstein Center for Fundamental Physics, Institute for Theoretical Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland*

[3]*Institut für Theoretische Physik, ETH Zürich, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland*

(Received 21 September 2018; published 5 December 2018)

The 2D O(3) model is widely used as a toy model for ferromagnetism and for quantum chromodynamics. With the latter it shares—among other basic aspects—the property that the continuum functional integral splits into topological sectors. Topology can also be defined in its lattice regularized version, but semiclassical arguments suggest that the topological susceptibility χ_t does not scale towards a finite continuum limit. Previous numerical studies confirmed that the quantity $\chi_t\xi^2$ diverges at large correlation length ξ. Here we investigate the question whether or not this divergence persists when the configurations are smoothened by the gradient flow (GF). The GF destroys part of the topological windings; on fine lattices this strongly reduces χ_t. However, even when the flow time is so long that the GF impact range—or smoothing radius—attains $\xi/2$, we still do not observe evidence of continuum scaling.

DOI: 10.1103/PhysRevD.98.114501

I. INTRODUCTION

We are going to deal with the 2D O(3) model, a nonlinear σ model, which is also known as the Heisenberg model, or $\mathbb{C}P(1)$ model. It is a highly popular toy model both in solid state physics—where it describes ferromagnets—and in particle physics, where it shares fundamental features with quantum chromodynamics (QCD). In particular, it is asymptotically free [1], it has a dynamically generated mass gap (which was computed analytically [2] and numerically, e.g., in Ref. [3]), and in the continuum formulation its configurations are divided into topological sectors, due to $\Pi_2[S^2] = \mathbb{Z}$.

On a square lattice of unit spacing, the standard lattice action reads

$$S[\vec{e}] = \beta \sum_{\langle xy \rangle} (1 - \vec{e}_x \cdot \vec{e}_y), \qquad \vec{e}_x \in S^2, \qquad x \in \mathbb{Z}^2, \qquad (1.1)$$

where $\langle xy \rangle$ are nearest-neighbor lattice sites. The naïve continuum limit leads to

$$S[\vec{e}] = \frac{\beta}{2} \int d^2x \partial_\mu \vec{e}(x) \cdot \partial_\mu \vec{e}(x), \qquad (1.2)$$

but at the quantum level it is far from obvious if the continuum limit is well defined. This is a long-standing issue, which arises in the context of topology. In particular, the crucial question is whether or not the topological susceptibility χ_t exhibits a continuum scaling behavior, i.e., whether or not the term $\chi_t\xi^2$, which is supposed to be a scaling term, is finite in the continuum limit $\xi \to \infty$ (where ξ is the correlation length in lattice units).

Numerous studies have addressed this question before, including Refs. [4–10]. After a period of confusion, the consensus seemed to be that $\chi_t\xi^2$ diverges at large ξ. This conclusion is consistent with studies with different lattice actions in the same universality class [11–13].

The interpretation of this observation is delicate: since the functional integral includes all field fluctuations, the notion of topology is generally nontrivial. For asymptotically free theories, one usually refers to the weak-coupling regime, where one assumes smooth configurations to dominate. For them topological sectors are well defined, along with a topological density q_x. The topological susceptibility can be assembled as $\chi_t = \sum_x \langle q_0 q_x \rangle$, where solely the contact term at lattice site $x = 0$ causes the divergence [10,13]. This implies that it is an UV effect, in agreement with the picture of more and more abundant tiny (with respect to ξ) topological windings as we approach the continuum limit [6]. However, that is not easily compatible with the assumption of dominant smooth configurations, and it is questionable if the perturbative expansion applies anywhere, even in the UV regime.

We have referred to the usual procedure which invokes the correlation length ξ to set a scale. It grows exponentially

*Corresponding author.
wolbi@nucleares.unam.mx

when we increase the inverse coupling β, which defines the standard continuum limit, say at fixed physical size L/ξ. However, if this leads to a divergence of $\chi_t\xi^2$, then there is no way to renormalize it by subtracting counterterms. Alternatively, one might consider taking χ_t as a reference quantity to fix the scale and keep $\langle Q^2\rangle = L^2\chi_t$ constant when β grows. Then L increases so slowly that the (standard) size L/ξ shrinks to zero; this is what it takes to sufficiently damp the topological fluctuations. Therefore the continuum limit is ambiguous: there is no way to obtain scaling for all observables, which are supposed to be physical. This situation is strange and unusual, although 3D U(1) lattice gauge theory has an analogous feature [14], where the string tension takes the rôle of χ_t.

Here we stay with the standard formulation of the continuum limit. A conceivable way out is that the divergence of $\chi_t\xi^2$ is not truly physical, but it can be overcome by systematic smoothing, which suppresses these tiny topological windings. Although smoothing techniques have been applied before, e.g., in Ref. [8], this question has not yet been addressed with the systematic method of the gradient flow (GF). Unlike *ad hoc* approaches, this smoothing procedure is justified based on the renormalization group [15–17]. Here we explore the fate of the term $\chi_t\xi^2$ after applying the GF.

The GF has a characteristic radius, which we call impact range (cf. Sec. II. 4); within this range the configurations become much smoother, so one might well expect most tiny topological windings to be eliminated, and—possibly—the divergence of χ_t as well. On the other hand, semiclassical topological windings fulfil the equations of motion, so they are not destroyed by the GF. They are just a measure-zero subset of all configurations, and the tricky question remains what happens to the frequent tiny winding in the presence of fluctuations. Hence it is hard to predict whether or not the divergence of $\chi_t\xi^2$ will even survive the GF, which motivates our numerical study.

There are other models with topological sectors, including 4D SU(N) gauge theories and QCD, which suffer from the same problem, if one uses a naïve lattice formulation of the topological charge density. However, in QCD this problem is less severe: solutions without GF are known [18], and the GF provides another solution [19]. In the absence of quarks, Ref. [20] discusses in detail how the GF cures χ_t in 4D SU(3) gauge theory. This raises the question if this remedy could also cure the divergence in the 2D O(3) model, which we are going to investigate.

Section II describes the numerical tools used in this study. Section III comments on the semiclassical picture, and Sec. IV presents our results for the topological scaling behavior at the quantum level, based on extensive Monte Carlo simulations. We discuss the outcome in Sec. V. The Appendix compares various numerical implementations of the GF. Preliminary results (with short GF times) have been published in two proceedings contributions [21].

II. NUMERICAL TECHNIQUES

A. Algorithm

Our simulations were performed with the Wolff cluster algorithm [22]. It adapts the concept of the Swendsen-Wang algorithm [23] from the Ising model to the O(N) models, where it is highly efficient. In our study we employed both the single-cluster as well as the multicluster variant.

B. Topological charge

Regarding the topological charge of a lattice configuration, we applied the geometric formulation, which was introduced in Ref. [4]. For periodic boundary conditions—which we assume throughout this article—it assigns an integer charge $Q[\vec{e}] \in \mathbb{Z}$ to each configuration (except for a subset of measure zero); this formulation is reviewed e.g., in Ref. [13], Sec. III.[1] In the absence of a θ term, symmetry implies $\langle Q\rangle = 0$,[2] and hence the topological susceptibility takes the simple form

$$\chi_t = \frac{1}{V}\langle Q^2\rangle. \tag{2.1}$$

C. Correlation length

The natural scale of the system is set by the correlation length ξ, which corresponds to the inverse energy gap. It is obtained by a two-parameter fit to the correlation function between spin averages in layers at fixed instances in Euclidean time, say x_2 and $x_2 + r$, with $0 \le r < L$,

$$C(r) = \langle \vec{s}_{x_2} \cdot \vec{s}_{x_2+r}\rangle \propto \cosh\left(\frac{r - L/2}{\xi}\right), \qquad \vec{s}_{x_2} = \frac{1}{L}\sum_{x_1=1}^{L} \vec{e}_x, \tag{2.2}$$

where we refer to a periodic $L \times L$ lattice with sites $x = (x_1, x_2)$. This proportionality relation holds when $|r - L/2|$ is sufficiently small. In practice we follow the recipe of Ref. [25] to determine ξ by a fit in the range $L/3 \le r \le 2L/3$ (varying this range leads to minor modifications of the fitting result for ξ). Our results agree with the literature, in particular with ξ values given in Refs. [3,25,26].

In each volume $V = L \times L$ we tune β such that $L/\xi \simeq 6$. Hence increasing the lattice volume corresponds to a system of fixed physical size, which approaches the continuum limit. The corresponding values of β and ξ

[1] Alternative definitions of the topological charge of a lattice configuration were suggested in Refs. [7,12,24].

[2] This can be seen from the topological charge density in the continuum, $q(x) = \epsilon_{\mu\nu}\vec{e}(x) \cdot (\partial_\mu \vec{e}(x) \times \partial_\nu \vec{e}(x))/8\pi$: a global spin flip $\vec{e}(x) \to -\vec{e}(x)$, which is the $Z(2)$ subgroup of the global O(3) symmetry, changes the sign of q.

TABLE I. Overview of the parameters in our study: we consider eight volumes $V = L \times L$; in each one β is tuned such that $L/\xi \simeq 6$, and the GF time unit amounts to $t_0 = L^2/5760$. When we apply the GF, the correlation length ξ does not change significantly up to flow time $10t_0$. Before the GF, at $t = 0$, ξ agrees fairly well with the second moment correlation length ξ_2, which can be measured more precisely. Our numerical measurements are based on $\mathcal{S}_\xi$ and $\mathcal{S}_{\xi_2}$ independent simulations for ξ and ξ_2, respectively, where each simulation generated 10^5 configurations.

					ξ		ξ_2
L	β	$\mathcal{S}_\xi$	$\mathcal{S}_{\xi_2}$	t_0	$t = 0$	$t = 10t_0$	$t = 0$
24	1.263	3	10	0.1	4.03(9)	4.02(5)	3.96(1)
36	1.37	4	5	0.225	5.97(10)	5.96(7)	6.01(1)
54	1.458	5	5	0.506	8.95(9)	8.96(7)	8.93(4)
80	1.535	4	5	1.111	12.99(17)	13.05(11)	13.24(4)
120	1.607	3	5	2.5	20.14(18)	19.87(13)	19.77(11)
180	1.677	3	5	5.625	31.09(36)	30.39(20)	30.01(18)
270	1.743	3	5	12.656	44.80(30)	45.32(8)	44.97(24)
404	1.807	2	5	28.336	68.34(52)	67.56(19)	67.66(31)

are given in Table I (the column for ξ at $t = 0$ contains the results before application of the GF).

These results are based on sets of $\mathcal{S}_\xi$ independent measurements (see Table I), each of which involves 10^5 configurations. If we insert the errors as obtained from the fits to Eq. (2.2), the independent results are not fully consistent: requiring a unique value at each L, we obtain $\chi^2/\text{d.o.f.} \simeq 4.0$, which shows that the errors are underestimated.[3] Therefore we amplify the errors by a factor of 2, which leads to consistency, in particular to $\chi^2/\text{d.o.f.} \simeq 1.0$. These extended errors are inserted into the Gaussian propagation to obtain the error of the average values given in Table I.

Despite the sizable statistics, the uncertainty in ξ is non-negligible; for comparison, the relative errors on χ_t are much smaller—see Sec. IV. Therefore we also measured the second moment correlation length ξ_2. It is obtained from the Fourier transform of the spin-spin correlation function $\langle \vec{e}_x \cdot \vec{e}_y \rangle$ at zero momentum (χ_m), and at the lowest nonzero momentum ($\mathcal{F}$),

$$\chi_m = \frac{1}{V}\sum_{x,y}\langle \vec{e}_x \cdot \vec{e}_y \rangle, \quad \mathcal{F} = \frac{1}{V}\sum_{x,y}\langle \vec{e}_x \cdot \vec{e}_y \rangle \cos\left(\frac{2\pi(x_1 - y_1)}{L}\right),$$
$$\xi_2 = \frac{1}{2\sin(\pi/L)}\sqrt{\frac{\chi_m}{\mathcal{F}} - 1}, \quad (V = L \times L), \tag{2.3}$$

where χ_m is the magnetic susceptibility (at magnetization zero). ξ_2 can be measured more precisely than ξ (cf. Table I), since it does not require any fit. In this case we have performed $\mathcal{S}_{\xi_2}$ independent measurements, each one based on 10^5 configurations (again the errors are somewhat enhanced for compatibility of the individual results). Strictly speaking, this is not the physical scale, but it is known to coincide with ξ to high accuracy: in the large-L limit the discrepancy is below 1‰ [27], and at $L/\xi_2 \simeq 4$, $L \geq 70$ it is still below 1% [13].[4] (A systematic comparison in other models is given in Ref. [28].)

D. Gradient flow

The GF in the O(N) models has been formulated in Refs. [29]. In the continuum, the spin components $e(x)^i$ are altered by integrating the differential equation

$$\partial_t e(t,x)^i = P^{ij}(t,x)\Delta e(t,x)^j,$$
$$P^{ij}(t,x) = \delta^{ij} - e(t,x)^i e(t,x)^j, \tag{2.4}$$

where Δ is the Laplace operator and t is the GF time (of dimension [length]^2), which starts at $t = 0$, i.e., $\vec{e}(0,x) = \vec{e}(x)$ and $t \geq 0$. The GF preserves the spin norm, which corresponds to the condition $\vec{e} \cdot \partial_t \vec{e} = 0$.

The concept of the GF is based on the heat kernel $K(t,x)$ [15–17]:

$$K(t,x) = \frac{e^{-x^2/4t}}{(4\pi t)^{d/2}}, \tag{2.5}$$

which allows us to estimate its impact range, or smoothing radius, $\bar{x}(t)$ as (in d dimensions)

$$\bar{x}(t) = \left(\int d^d x \; x^2 K(t,x)\right)^{1/2} = \sqrt{2dt}. \tag{2.6}$$

On the lattice we deal with the spin field $\vec{e}(t)_x$, and we insert the standard lattice Laplacian:

[3] Possible reasons are the fixed fitting range, and the use of the same configurations to measure the correlation function over all distances, although we only include a fixed x_2 and $r = 0 \ldots L - 1$ (i.e., one-to-all but no all-to-all correlations).

[4] This was observed for the "constraint lattice action"; cf. Sec. III.

$$\Delta\vec{e}(t,x) \to \sum_{\mu=1}^{d}[\vec{e}(t)_{x+\hat{\mu}} + \vec{e}(t)_{x-\hat{\mu}}] - 2d\vec{e}(t)_x, \qquad |\hat{\mu}| = 1. \tag{2.7}$$

For the numerical integration of Eq. (2.4) also the GF time t has to be discretized. Here we apply the Runge-Kutta method; see e.g., Ref. [30]. We first compute the gradients to all spin components at all lattice sites, and then we rotate all spins *simultaneously* (afterwards the normalization $|\vec{e}_x| = 1$ is readjusted at each site). In small and moderate volumes we used the four-point Runge-Kutta method, with time step $dt = 10^{-4}$.[5]

In this project, the GF integration took most of the computation time. In order to handle lattice sizes up to $L = 404$, it was mandatory to implement an adaptive step size. We applied the Dormand-Prince algorithm [31], which gradually increases dt, if the Runge-Kutta four-point and five-point gradients agree to high accuracy. At long flow times this method provides a gain in computing time by several orders of magnitude: once a configuration is quite smooth, dt can be enhanced drastically without causing significant artifacts. This is discussed in the Appendix.

In order to compare results in different volumes, and therefore at different couplings, we need a GF time unit t_0, which has to be determined by referring to a dimensional observable. Such a time unit allows for the matching of results from different couplings and volumes and therefore for a controlled continuum extrapolation (which is not obvious for *ad hoc* smoothing techniques). In QCD, t_0 is usually fixed by the condition $\langle E\rangle t_0^2 = 0.3$ [15] [or $\langle E\rangle t_0^2 = 0.1$ for SU(2) Yang-Mills theory [32]], where the density $E = -\mathrm{Tr}[G_{\mu\nu}G_{\mu\nu}]/2$ serves as an observable, which is easily measurable ($G_{\mu\nu}$ is a lattice field strength tensor).

In Refs. [21] we have used the corresponding density in the 2D O(3) model, $\langle E\rangle = \langle S\rangle/\beta V$. However, this turned out to be impractical: for increasing GF time t the (dimensionless) term $\langle E\rangle t$ rises from 0 to some maximum and decreases again. The value of this maximum decreases as we enlarge L, so in order to capture all volumes under consideration, we had to take a small reference value like $\langle E\rangle t_0^{\mathrm{short}} = 0.08$ (for instance, the value 0.1 is never attained at $L = 404$). Thus we obtained short time units $t_0^{\mathrm{short}} \lesssim 0.1$, and up to $6t_0^{\mathrm{short}}$ the impact range attained at most 1.6 lattice spacings.

In order to probe much larger impact ranges, of $\mathcal{O}(\xi)$, we now refer directly to ξ as our reference observable to fix t_0. We define it such that $10t_0$—the longest GF time in our study—corresponds to an impact range of about $\xi/2$:

[5]We checked that the results coincide within the errors with those obtained at $dt = 10^{-5}$. On the other hand, when we increase the step size to $dt = 10^{-3}$ we noticed (in a few cases) non-negligible artifacts; they typically emerge at an early stage of the GF.

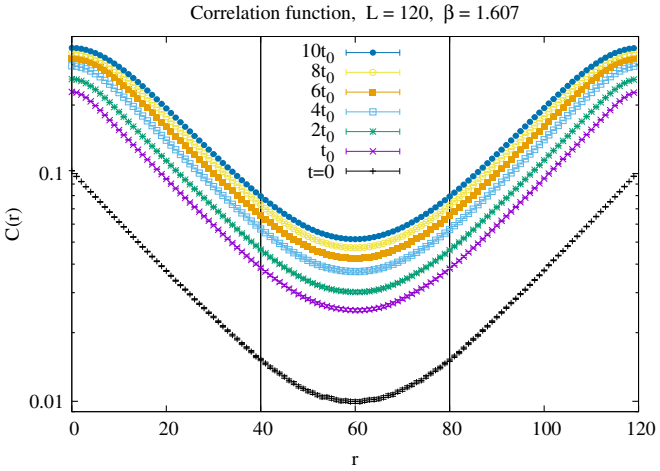

FIG. 1. The correlation function $C(r) = \langle \vec{s}_{x_2} \cdot \vec{s}_{x_2+r}\rangle$, measured before and after the GF, at $L = 120$ (as an example). At fixed separation r, $C(r)$ rises under GF, but the value of ξ—obtained from a fit to relation (2.2) in the interval $40 \le r \le 80$—remains practically constant.

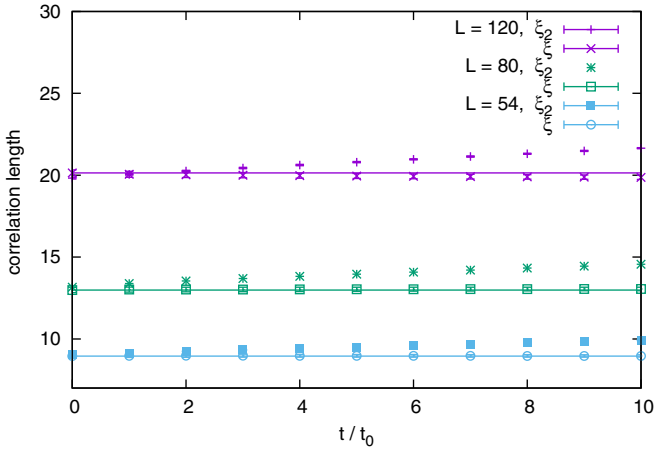

FIG. 2. A comparison of the correlation length ξ, and the second moment correlation length ξ_2, under GF up to flow time $10t_0$. We see that they initially agree well, but as the GF proceeds, ξ_2 increases, whereas ξ remains stable [this is visualized with horizontal lines at the values of $\xi(t = 0)$]. We show three volumes as examples for this generic effect.

$$t_0 = \frac{1}{5760}L^2 \simeq \frac{1}{160}\xi^2 \to \bar{x}(10t_0) \simeq \frac{1}{2}\xi. \tag{2.8}$$

Table I contains the GF time unit t_0, as well as the correlation length measured at $10t_0$; we see that it hardly changes compared to $t = 0$.[6]

[6]In the framework of finite temperature gauge theory, Ref. [33] discusses the question how long the GF time can be, before destroying physical information. Our results for $\xi(t)$ show that—in our case—we are on the safe side, at least up to $10t_0$.

An example for the GF time evolution of the correlation function $C(r)$ of Eq. (2.2) is illustrated in Fig. 1: at a fixed distance r it increases under GF, but the value of ξ remains virtually unaffected. This is consistent with the fact that $\bar{x}$ is still small compared to L, $\bar{x}(10t_0) \simeq L/12$, so it does not reach out to the interval, where we performed fits to relation (2.2). The GF does, however, have the expected effect of suppressing the statistical errors in ξ (they are amplified with the factor of 2, as at $t = 0$; cf. Sec. II. 3).

On the other hand, after applying the GF the second moment correlation length ξ_2 increases above its value at $t = 0$, as illustrated in Fig. 2. This property is generic[7]; it implies that $\xi_2(t > 0)$ cannot be used to set an (approximate) scale. Instead our results for $\xi(t)$ justify the use of the scale $\xi_2(0)$ even after the GF, all the way up to $t/t_0 = 10$.

III. THE SEMICLASSICAL PICTURE

Reference [4] was the first study to show that the numerical results for the topological susceptibility χ_t, based on Monte Carlo simulations of the standard action (1.1), do not seem compatible with continuum scaling, i.e., with the scaling towards a finite continuum limit, which is naïvely expected. In particular, the dimensionless term $\chi_t\xi^2$ seems to diverge in the continuum limit. Small topological windings, which may occur in lattice configurations with low action, were blamed for this effect; it was suspected that their dominant rôle on fine lattices prevents continuum scaling [4,5].

Reference [6] provided a comprehensive semiclassical explanation for this behavior. It generally considered 2D $\mathbb{C}\mathrm{P}(N-1)$ models,[8] where the continuum instanton action (the minimal action within the topological sector $|Q| = 1$) amounts to

$$S_{\mathrm{inst}} = \beta\epsilon_{\mathrm{inst}}, \qquad \epsilon_{\mathrm{inst}} = 2\pi N. \tag{3.1}$$

On the lattice, a single topological winding ($Q = \pm 1$) with minimal action is denoted as a dislocation. Its action was numerically obtained as [6]

$$S_{\mathrm{disloc}} = \beta\epsilon_{\mathrm{disloc}}, \qquad \epsilon_{\mathrm{disloc}} \simeq 6.69 \cdot N/2. \tag{3.2}$$

At the quantum level, the fate of a model depends on the balance between action and entropy. In this case, a perturbative calculation of the β function suggests that the fate of this model depends on $\epsilon_{\mathrm{disloc}}$ as follows [6]:

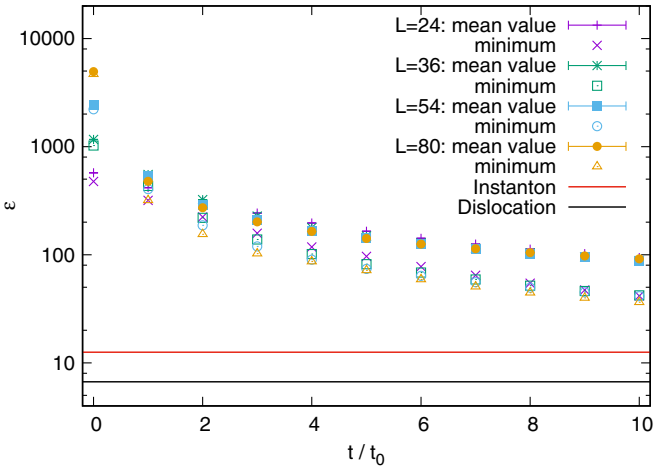

FIG. 3. The quantity $\langle\epsilon\rangle = \langle S\rangle/\beta$ measured in volumes $V = L \times L$, at $L/\xi \approx 6$. In each volume, and at each flow time $t/t_0 = 0, 1 \ldots 10$, we used 50 000 configurations with topological charge $|Q| = 1$. We also show the minima of ϵ within this set. Even after a long GF, up to flow time $10t_0$, these minima are still more than a factor of 5 above the dislocation value $\epsilon_{\mathrm{disloc}} \simeq 6.69$. We infer that dislocations and their vicinities hardly contribute to the statistics.

$$\epsilon_{\mathrm{disloc}} \begin{cases} > 4\pi & \text{continuum scaling,} \\ < 4\pi & \text{divergence in the continuum limit.} \end{cases} \tag{3.3}$$

This implies that continuum scaling of χ_t is safe at $N \geq 4$. At $N = 3$ it can still be arranged for by adding nonstandard terms to the lattice action [34]. $N = 2$, however, is a peculiar case, where ϵ_{inst} coincides with the bound derived from the β function. In this case, which corresponds to the O(3) model, the semiclassical picture predicts the term $\chi_t\xi^2$ to diverge in the continuum limit.

This semiclassical argument is not rigorous, of course; there is no compelling reason for it to be conclusive at the full quantum level. Still, a variety of numerical studies ultimately suggested that this prediction is confirmed; cf. Sec. I.

Reference [11] applied a sophisticated lattice action, a (truncated) *classically perfect action*, which was constructed by means of classical block spin renormalization group transformations. It involves couplings over several lattice spacings, which exclude dislocations with $\epsilon_{\mathrm{disloc}} < 4\pi$, but $\chi_t\xi^2$ still was found to diverge logarithmically in the continuum limit.

Very different are topological lattice actions, in particular the *constraint action*, where all configurations have action 0, if the relative angles between all nearest-neighbor spins are below some bound δ.[9] In this case, the dislocations are extremely degenerate, with $\epsilon_{\mathrm{disloc}} = 0$. Hence one might expect a very bad divergence of $\chi_t\xi^2$ in the continuum limit,

[7]Note that the entire configurations contribute to the terms χ_m and $\mathcal{F}$, in contrast to the fits, which determine ξ within a limited range. Hence the short-distance deformation of the correlation function (see Fig. 1) is likely to cause the distortion of ξ_2.

[8]All the 2D $\mathbb{C}\mathrm{P}(N-1)$ models, $N = 2, 3, 4\ldots$, have topological sectors (labeled by $Q \in \mathbb{Z}$), in contrast to the 2D O(N) models with $N \neq 3$.

[9]If at least one nearest-neighbor angle exceeds δ, then the action is infinite; i.e., such configurations are excluded from the functional integral.

TABLE II. The topological susceptibility χ_t of Eq. (2.1) on $L \times L$ lattices, with the values of β and t_0 given in Table I, based on $\mathcal{S}_\chi$ measurements with 10^5 configurations each.

		χ_t (in units of 10^{-3})				
L	$\mathcal{S}_\chi$	$t=0$	t_0	$2t_0$	$5t_0$	$10t_0$
24	5	7.54(1)	5.80(1)	4.85(1)	3.516(7)	2.681(6)
36	5	4.736(9)	2.926(6)	2.319(5)	1.677(3)	1.356(3)
54	5	2.982(7)	1.388(3)	1.103(2)	0.856(2)	0.743(1)
80	5	1.87(1)	0.662(2)	0.552(1)	0.466(1)	0.423(1)
120	5	1.150(6)	0.321(3)	0.287(2)	0.255(2)	0.235(2)
180	3	0.691(2)	0.1614(4)	0.1491(4)	0.1360(4)	0.1266(4)
270	3	0.422(1)	0.0843(2)	0.0765(2)	0.0705(2)	0.0662(2)
404	2	0.2538(8)	0.0414(1)	0.0392(1)	0.0362(1)	0.0342(1)

which is attained in this case by $\delta \to 0$. It turned out, however, that the divergence is still compatible with a logarithmic dependence on ξ [13].

Here this question will be revisited under application of the GF.[10] Before doing so, however, we begin with an observation about the relevance of the semiclassical picture. To this end, it is sufficient to consider modest lattice volumes, of sizes $L = 24 \ldots 80$, with the β values of Table I. In each volume we selected 50 000 configurations with topological charge $|Q| = 1$.

Figure 3 refers to the quantity $\epsilon = S/\beta$ [S being the lattice action (1.1)]: it shows the mean value $\langle \epsilon \rangle$, as well as the minimum obtained in each volume. At GF time $t = 0$ even the minima (in this set of configurations) are orders of magnitude above the instanton and dislocation values. This suggests that—although configurations with ϵ down to ϵ_{disloc} exist—their contribution to a typical expectation value is negligible in our settings.[11]

When we apply the GF, as described in Sec. II, the configurations become smoother and the action decreases, so one might suspect that now the semiclassical configurations (or at least their vicinity) become relevant. Figure 3 shows that this is *not* the case: even when we run the GF up to $10t_0$, the averages and minima (within a set of 50 000 configurations, at any instant t) are still more than a factor of 5 times larger than ϵ_{disloc}.

Figure 3 further shows that this observation hardly depends on the volume. It raises the question how relevant the semiclassical consideration really is, since it does not refer to the statistically significant contributions (unless presumably in tiny physical volumes). Nevertheless, our goal is a direct verification of its prediction; this is the question to be addressed in the next section.

IV. TOPOLOGY UNDER THE GRADIENT FLOW

Based on $\mathcal{S}_\chi$ sets of 10^5 configurations in each volume (see Table II), we finally measured the topological susceptibility χ_t, given in Eq. (2.1). Unlike the case of ξ, the results for χ_t from these $\mathcal{S}_\chi$ independent simulations are consistent within our estimated errors: in this case we obtain a ratio $\chi^2/\text{d.o.f.} \simeq 0.90$ (the cluster algorithm allows us to avoid topological autocorrelations).

Our results for χ_t are listed in Table II. They are averaged over all simulations, and each of their standard errors enters the Gaussian composition of the final error. The evolution under GF is illustrated in Fig. 4, which shows the dimensionless term $\langle Q^2 \rangle = \chi_t V$. In large lattice volumes, i.e., on fine lattices, we see a rapid decrease of $\langle Q^2 \rangle$ when the GF starts, in particular from $t = 0$ to t_0 (in the Appendix we will see that most of this effect happens even within a first small fraction of t_0). At a later stage $\langle Q^2 \rangle$ still keeps decreasing but at an ever slower rate.

At last we arrive at the discussion of the "scaling term" $\chi_t \xi^2$. Regarding the correlation length, we rely on the

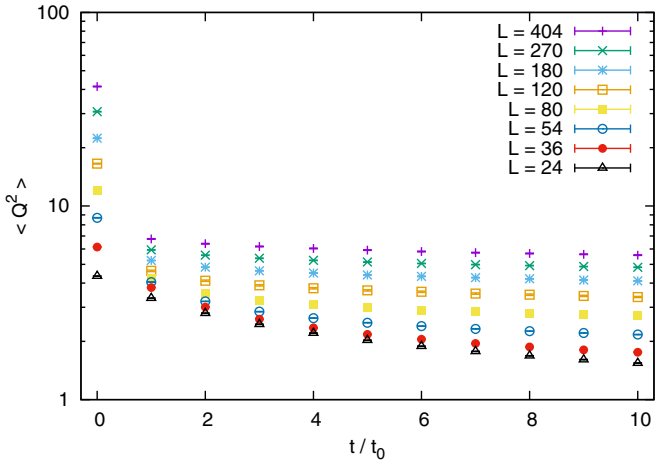

FIG. 4. GF time evolution of the expectation value $\langle Q^2 \rangle = \chi_t V$.

[10] According to Eq. (2.8) we deal with an impact range which is adjusted to $L/12$; it attains 33.7 lattice spacings in our largest volume. This strongly differs from Refs. [21] (see Sec. II. 4) and also from Ref. [11], where the coupling range of the "perfect lattice action" was fixed to a couple of lattice spacings, while ξ increased up to 58.

[11] Actually such configurations have the highest probability $p[\vec{e}] \propto \exp(-S[\vec{e}])$, but configurations with a significantly larger action have a much higher degeneracy, such that they overwhelmingly dominate the functional integral.

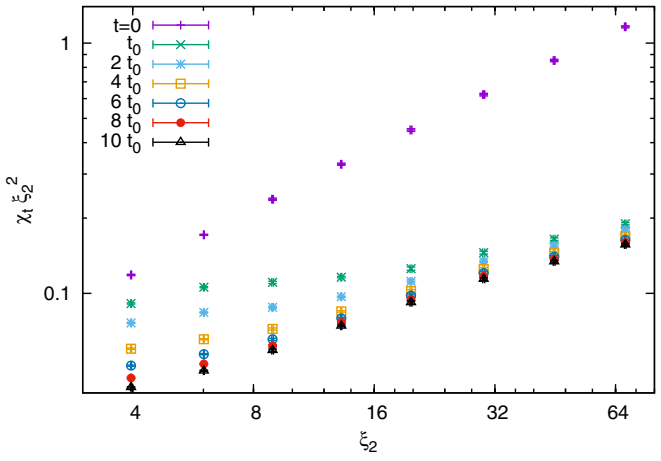

FIG. 5. Illustration of the nonscaling of the term $\chi_t\xi^2$. The scale used in this plot is the second moment correlation length ξ_2.

property that the flow times are not excessively long, so that physical aspects are not affected, and in particular the long-range scale ξ should not change, as we argued in Sec. II. 4. In fact, our results in Table I confirm that the modifications of ξ are minor: in each volume, $\xi(0)$ and $\xi(10t_0)$ agree within less than 1.7σ. [It is also noteworthy that the sign of $\xi(0) - \xi(10t_0)$ differs in our results from different lattice volumes, which further shows the absence of a systematic effect of the GF on ξ up to $10t_0$.]

Trusting the stability of ξ, we replace it by $\xi_2(0)$, for which we have precise results—see Table I—and use it at any flow time $t \in [0, 10t_0]$; cf. Sec. II. 3. This yields the scaling plot in Fig. 5.

It is an unambiguous observation that—after any fixed multiple of the flow time unit t_0 that we considered—the quantity $\chi_t\xi^2$ keeps growing as we increase the correlation length; we cannot observe convergence towards a finite continuum limit. This trend is most obvious in our largest lattice volumes and at long flow times. At relatively short GF, in particular at flow time t_0, the term, which is supposed to scale, looks almost stable up to $L = 80$, $\xi \approx 13$, but even closer to the continuum limit it turns into the (qualitative) behavior observed at long flow times.

As a first hypothesis, we assume the asymptotic behavior at large ξ to be logarithmic.[12] This can be expressed by the ansatz

$$\chi_t\xi^2 = a_1 \ln(a_2\xi + a_3), \qquad a_i = \text{const}, \qquad (4.1)$$

which was successful in fits to results obtained with topological lattice actions [13]. As an alternative, we consider another three-parameter ansatz, which describes a power law,

$$\chi_t\xi^2 = b_1\xi^{b_2} + b_3, \qquad b_i = \text{const}, \qquad (4.2)$$

as in Refs. [13,21]. That behavior corresponds to the semiclassical picture of Ref. [6] (for the case of 4D Yang-Mills gauge theory, this property is worked out explicitly in Ref. [35]).

We first consider the data before the GF. In this case, we perform fits over the entire range $L = 24\ldots404$, so there are 5 "degrees of freedom," and we obtain at $t = 0$

$$\begin{aligned} a_1 &= 1.3(1), \qquad & a_2 &= 0.021(3), \\ a_3 &= 1.014(2), \qquad & \chi^2/\text{d.o.f.} &= 3.76, \\ b_1 &= 0.0522(5), \qquad & b_2 &= 0.741(2), \\ b_3 &= -0.026(1), \qquad & \chi^2/\text{d.o.f.} &= 0.04. \end{aligned}$$

The power-law fit has a tiny value of χ^2/d.o.f. (which appears accidental), but this quantity is somewhat large for the logarithmic fit.[13] However, even there the uncertainties of the fitting parameters are moderate. The observation that the constants a_1, a_2, b_1, and b_2 are all larger than 0 (far beyond the errors) confirms that the data before GF are incompatible with continuum scaling.

Figure 6 shows the constants $a_i(t)$ and $b_i(t)$ obtained from the fits to the functions (4.1) and (4.2) at GF times $t = t_0, 2t_0\ldots10t_0$. The lower plot also shows the ratio χ^2/d.o.f., as a measure of the quality of the fits. All fits were performed in the range $L = 54\ldots404$; hence they capture six data points, corresponding to six lattice volumes. In all these cases, i.e., after the GF, the fits to the logarithmic ansatz (4.1) are superior, as we see from the lower plot in Fig. 6 (this behavior agrees with Refs. [21]).

The essential observation, however, is based on the upper plot: it shows that the constants a_1, a_2, b_1, and b_2 keep on being larger than zero during the GF; zero values are well beyond the errors.[14] Therefore, even after the GF our data are incompatible with a scaling of $\chi_t\xi^2$ towards a finite continuum limit.

V. CONCLUSIONS

There is a variety of models with topological sectors, and some of them are plagued by problems with the continuum scaling of χ_t. This is not the case in the simple 1D O(2) model, where—for a multitude of lattice actions—$\chi_t\xi$ exhibits a straight convergence to its continuum value of $1/2\pi^2$ [13,36].

[12]Here and in the following we refer to ξ, which is conceptually correct, although in practice it is replaced by ξ_2, as we explained before.

[13]The fits to our preliminary data that we considered in Refs. [21] had a similar quality for both functions. On the other hand, in Ref. [13] we observed superiority of the logarithmic function for data obtained with the constraint lattice action.

[14]In light of Fig. 6, one might question the behavior of a_2 and b_1 at short $t \leq t_0$. However, in Refs. [21] we arrived at the same conclusion also for such short GF times.

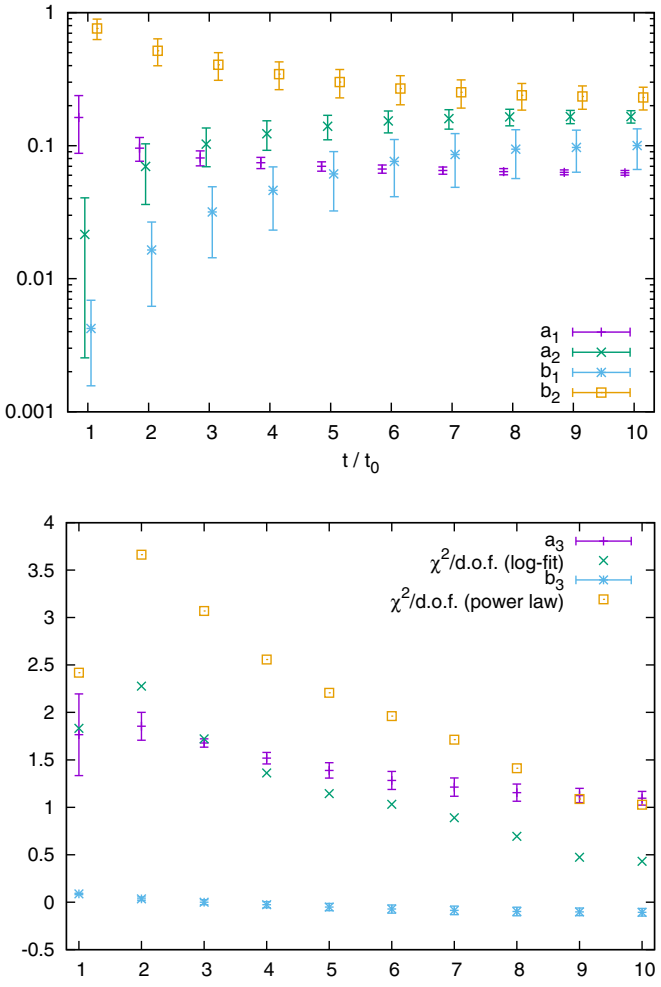

FIG. 6. The constants $a_i(t)$ and $b_i(t)$ of the fits to the logarithmic function (4.1) and to the power law (4.2), respectively, along with the χ^2/d.o.f. ratios. The fits were performed at any flow time $t/t_0 = 1\ldots10$, in the range $L = 54\ldots404$ (d.o.f. $= 3$). In all cases the logarithmic fits have a better quality, in contrast to the case $t = 0$. The constants a_1, a_2, b_1, and b_2 are positive and incompatible with 0 at any GF time up to $10t_0$. Hence our data are incompatible with continuum scaling.

In naïve lattice formulations of 4D Yang-Mills gauge theory, as well as QCD, this problem appears, but there are various ways to overcome it; see Ref. [37] for pure SU(3) gauge theory and the aforementioned Refs. [18,19] for QCD.

Regarding the 2D $\mathbb{C}\mathrm{P}(N-1)$ models, the numerical results confirm the semiclassical picture of Ref. [6] that we sketched in Sec. 3: no problem occurs at $N \geq 4$, and at $N = 3$ there is a divergence, but it can be avoided by nonstandard lattice actions; see e.g., Refs. [34,38].

There remains the case $N = 2$, which is peculiar indeed: in this model, which is equivalent to the 2D O(3) model, no way around the divergent continuum limit of $\chi_t\xi^2$ is known; we have seen that not even the GF, which is a safe remedy in other models, helps in this specific case.

This does not mean that *all* topological terms in the 2D O(3) model are ill defined. Even without GF, there is evidence for the opposite to hold for the following quantities:

(i) The correlation function of the topological charge density q_x, $\langle q_x q_y\rangle$, is well defined (i.e., finite in the continuum limit) at all separations $x - y$, expect for $x = y$. That point alone causes the divergence of $\chi_t = \sum_y \langle q_x q_y\rangle$ [10,13], and the situation is similar in QCD [18].[15]

(ii) The kurtosis $c_4 = (3\langle Q^2\rangle^2 - \langle Q^4\rangle)/V$ is a characteristic of the distribution of the topological charges (it vanishes if this distribution is Gaussian). In the continuum limit, the ratio c_4/χ_t converges to a value close to -1 [40] (which is the value of a dilute instanton gas).

(iii) If we add a θ term, $S[\vec{e}]_\theta = S[\vec{e}] - \mathrm{i}\theta Q[\vec{e}]$, with $-\pi < \theta \leq \pi$, we obtain an expectation value $\langle Q\rangle$, which does not need to vanish anymore. Therefore we now have to refer to the general expression for χ_t:

$$\langle Q\rangle = -\mathrm{i}\partial_\theta \ln Z(\theta),$$
$$\chi_t = \frac{1}{V}(\langle Q^2\rangle - \langle Q\rangle^2) = -\frac{1}{V}\partial_\theta^2 \ln Z(\theta). \qquad (5.1)$$

The expectation value $\langle Q\rangle$ is well defined at any vacuum angle θ, but the function $\langle Q\rangle(\theta)$ has an infinite slope at $\theta = 0$. This is the picture elaborated in Ref. [41], without GF, which is sketched schematically in Fig. 7. It implies that θ remains finite under renormalization[16] and that $\chi_t(\theta)$ does exhibit continuum scaling at any $\theta \neq 0$.

We have seen that the picture of Ref. [41]—in particular the infinite slope at $\theta = 0$—seems to (qualitatively) persist under the GF. This extends our previous observation [21] to much longer flow times.

Reference [41] did not consider this behavior unnatural, although $\chi_t(\theta = 0)$ is supposed to be an observable. This scenario requires the free energy $F(\theta) = -\beta^{-1} \ln Z(\theta)$ to take an unusual—though conceivable—form, where $F(\theta)$ and $F'(\theta)$ have removable singularities at $\theta = 0$, which give rise to a divergence of $F''(0)$ [a prototype for such a function is $F(\theta) \propto \theta^2 \ln\theta$].

Here we present our numerical results, which support this scenario. We leave it to the reader to decide whether he or she considers this property as fatal for the topology of the 2D O(3) model.

[15] In a fixed topological sector, the correlation $\langle q_x q_y\rangle$ at large separation can be employed for an indirect measurement of χ_t [39].

[16] Reference [41] concludes that each value of $\theta \in [0, \pi]$ represents a different continuum theory.

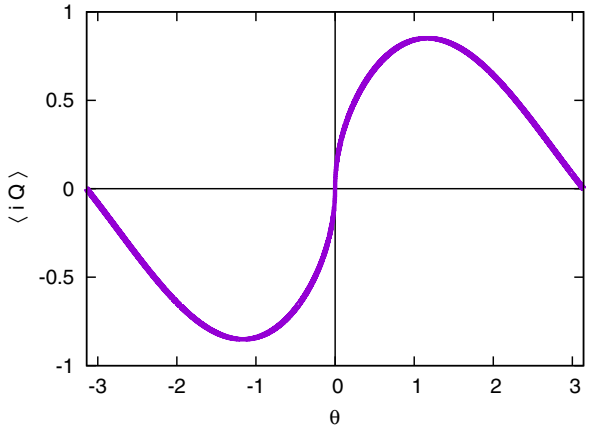

FIG. 7. A schematic illustration of the expectation value $i\langle Q\rangle$ as a function of the vacuum angle θ, in the continuum limit. The peculiarity of the 2D O(3) model is that its slope—which is proportional to χ_t—seems to diverge at $\theta = 0$. This picture corresponds to Refs. [41,42]; our results suggest that its qualitative features persist under the GF.

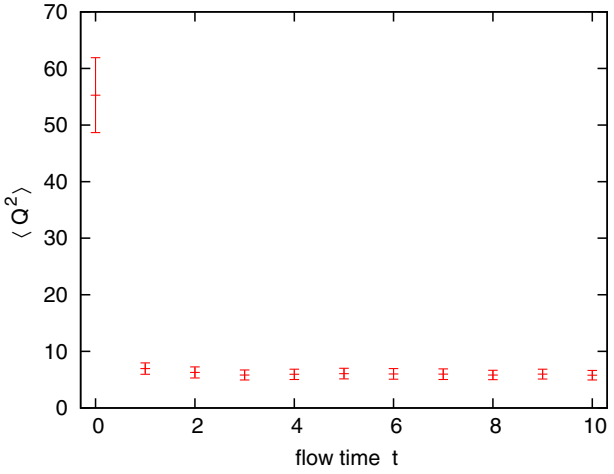

FIG. 8. The mean value $\langle Q^2\rangle(t)$, obtained from 100 configurations at $L = 404$, at flow times $t = 0\ldots 10t$. The results coincide for all three GF implementations under consideration.

ACKNOWLEDGMENTS

We are indebted to Martin Lüscher for suggesting this project and for helpful advice on its realization. We further thank Uwe-Jens Wiese for instructive discussions. The computations were performed on the cluster of ICN/UNAM; we thank Luciano Díaz and Eduardo Murrieta for technical support. This work was supported by DGAPA-UNAM through Grant No. IN107915 and through the program PASPA-DGAPA, by the Albert Einstein Center for Theoretical Physics and by the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC Grant Agreement No. 339220. P. d. F. thanks the CERN Theory Division for its hospitality.

APPENDIX: NUMERICAL INTEGRATION OF THE GRADIENT FLOW

This Appendix compares various implementations of the GF based on the Runge-Kutta method; for a pedagogical description of this method we recommend Ref. [30]. In particular we are going to address the performance of the Dormand-Prince adaptive step size algorithm [31].

That algorithm allowed us to handle lattices up to size $L = 404$ with a high statistics of 2×10^5 configurations; see Tables I and II. In the smaller volumes we could run the GF at a fixed step size of $dt = 10^{-4}$, and extensive tests demonstrated the consistency with the Dormand-Prince algorithm, up to $L = 270$. This Appendix is going to concentrate on $L = 404$, where fixed $dt = 10^{-4}$ production runs are prohibitively expensive. Instead we refer to a sample of 100 test configurations, which were generated at $\beta = 1.807$ (the value used in our study), well thermalized and independent.

Our tests have further shown that the most delicate part of the GF is the very beginning. This is expected: possible artifacts due to the finite step size dt are most likely before the configurations become smooth. In this Appendix we consider flow time $t = 0$ to $10 \simeq 0.35t_0$. This interval is of primary interest: we will see that most of the reduction of $\langle Q^2\rangle$ that we observe up to $10t_0$ (see Fig. 4) happens in the very first flow period.

Strictly speaking, the application of the Dormand-Prince algorithm requires two parameters: the initial time step dt_0 and a "tolerance parameter" ε. If the gradients computed by the Runge-Kutta method with four points and with five points[17] coincide within this tolerance, i.e., the norm of their difference is below ε, then dt will be increased in the subsequent step—in the opposite case it will be decreased.[18]

Regarding the initial time step dt_0, we ran numerous tests with $dt_0 = 10^{-3}$ and $dt_0 = 10^{-4}$: when everything else was kept fixed, we never found any difference which could be significant at our level of precision. After just a few time steps one obtains results, which are practically indistinguishable. Since this choice hardly affects the computation time, we used $dt_0 = 10^{-4}$ in our production runs and also in the tests to be presented in this Appendix. Hence our discussion focuses on the tolerance parameter ε.

We are going to compare three numerical implementations of the GF:

(i) fixed step size $dt = 10^{-4}$,
(ii) Dormand-Prince adaptive step size with $dt_0 = 10^{-4}$ and $\varepsilon = 10^{-6}$ (as used in our production runs), and

[17]Referring specifically to these two gradients is motivated by the fact that some ingredients of their computation are identical.

[18]This is untypical, since the configurations become gradually smoother under the GF, but it does occasionally happen; i.e., the increase of dt is not strictly monotonic.

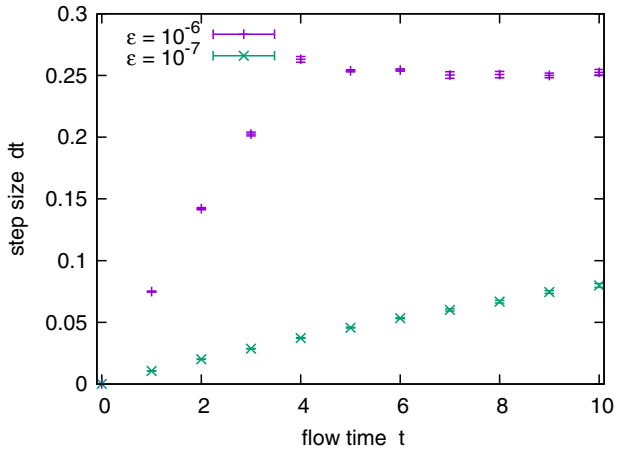

FIG. 9. The step size evolution of the Dormand-Prince algorithm with an initial time step $dt_0 = 10^{-4}$ and with tolerance parameter $\varepsilon = 10^{-6}$ or 10^{-7}.

(iii) the same Dormand-Prince algorithm with $\varepsilon = 10^{-7}$.

First we consider the topological charges of these 100 test configurations. We checked for possible deviations when we apply these GF implementations, but they fully agree at any $t = 1, 2, 3 \ldots 10$. Figure 8 shows the value of the $\langle Q^2 \rangle$ obtained from this sample. It confirms that most of the destruction of topological windings happens very early, at $t < 1 \simeq 0.035 t_0$. This corresponds to an impact range below two lattice spacings; hence it matches the picture of a quick destruction of numerous tiny dislocations (compared to the correlation length $\xi \simeq 68$). The topological windings that persist can either be large or small with a structure, which resists the GF for a longer flow time. We saw that these remaining windings still make the topological susceptibility diverge.

Figure 9 illustrates how dt increases when we apply the Dormand-Prince algorithm, with tolerance parameter $\varepsilon = 10^{-6}$ or $\varepsilon = 10^{-7}$. The difference between these two scenarios is significant: in particular, at $\varepsilon = 10^{-6}$ the step size soon attains a remarkable magnitude of $dt \approx 0.25$; at $t \approx 4$ the configurations are already sufficiently smooth to allow for this value. (That case also confirms that, in exceptional cases, the algorithm can temporarily decrease dt; cf. footnote 18.)

Since we did not observe any significant difference in the results, the use of this value of ε is highly motivated. It provides a gain in computation time by several orders of magnitude: this gain can be estimated by assuming the GF to take computation time $\propto 1/dt$, although adaptive step size algorithms require some additional operations.

[1] A. M. Polyakov, Phys. Lett. **59B**, 79 (1975).
[2] P. Hasenfratz, M. Maggiore, and F. Niedermayer, Phys. Lett. B **245**, 522 (1990).
[3] J.-K. Kim, Phys. Rev. D **50**, 4663 (1994).
[4] B. Berg and M. Lüscher, Nucl. Phys. **B190**, 412 (1981).
[5] B. Berg, Phys. Lett. **104B**, 475 (1981).
[6] M. Lüscher, Nucl. Phys. **B200**, 61 (1982).
[7] A. Di Giacomo, F. Farchioni, A. Papa, and E. Vicari, Phys. Lett. B **276**, 148 (1992); Phys. Rev. D **46**, 4630 (1992).
[8] C. Michael and P. S. Spencer, Phys. Rev. D **50**, 7570 (1994).
[9] F. Farchioni and A. Papa, Nucl. Phys. **B431**, 686 (1994); B. Allés and M. Beccaria, Phys. Rev. D **52**, 6481 (1995).
[10] J. Balog and M. Niedermaier, Nucl. Phys. **B500**, 421 (1997); Phys. Rev. Lett. **78**, 4151 (1997).
[11] M. Blatter, R. Burkhalter, P. Hasenfratz, and F. Niedermayer, Phys. Rev. D **53**, 923 (1996).
[12] M. D'Elia, F. Farchioni, and A. Papa, Phys. Rev. D **55**, 2274 (1997).
[13] W. Bietenholz, U. Gerber, M. Pepe, and U.-J. Wiese, J. High Energy Phys. 12 (2010) 020.
[14] M. Göpfert and G. Mack, Commun. Math. Phys. **82**, 545 (1982).
[15] M. Lüscher, J. High Energy Phys. 08 (2010) 071; Proc. Sci., LATTICE2010 (**2010**) 015.
[16] M. Lüscher and P. Weisz, J. High Energy Phys. 02 (2011) 051.
[17] H. Makino, O. Morikawa, and H. Suzuki, Prog. Theor. Exp. Phys. **2018**, 053B02 (2018); Y. Abe and M. Fukuma, Prog. Theor. Exp. Phys. **2018**, 083B02 (2018).
[18] L. Giusti, G. C. Rossi, and M. Testa, Phys. Lett. B **587**, 157 (2004); M. Lüscher, Phys. Lett. B **593**, 296 (2004).
[19] M. Bruno, S. Schaefer, and R. Sommer, J. High Energy Phys. 08 (2014) 150.
[20] M. Cè, C. Consonni, G. P. Engel, and L. Giusti, Phys. Rev. D **92**, 074502 (2015).
[21] I. O. Sandoval, W. Bietenholz, P. de Forcrand, U. Gerber, and H. Mejía-Díaz, J. Phys. Conf. Ser. **912**, 012024 (2017); H. Mejía-Díaz, W. Bietenholz, K. Cichy, Philippe de Forcrand, A. Dromard, U. Gerber, and I. O. Sandoval, Eur. Phys. J. Web Conf. **175**, 11024 (2018).
[22] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[23] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
[24] G. Martinelli, R. Petronzio, and M. A. Virasoro, Nucl. Phys. **B205**, 355 (1982); B. Berg and C. Panagiotakopoulos, Nucl. Phys. **B251**, 353 (1985).
[25] U. Wolff, Nucl. Phys. **B334**, 581 (1990).
[26] J. Apostolakis, C. F. Baillie, and G. C. Fox, Phys. Rev. D **43**, 2687 (1991).
[27] M. Campostrini, A. Pelissetto, P. Rossi, and E. Vicari, Phys. Lett. B **402**, 141 (1997).
[28] M. Caselle and A. Nada, Phys. Rev. D **96**, 074503 (2017).

[29] K. Kikuchi and T. Onogi, J. High Energy Phys. 11 (2014) 094; H. Makino and H. Suzuki, Prog. Theor. Exp. Phys. **2015**, 033B08 (2015).
[30] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes, The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge, England, 2007).
[31] J. R. Dormand and P. J. Prince, J. Comput. Appl. Math. **6**, 19 (1980).
[32] T. Hirakida, E. Itou, and H. Kouno, arXiv:1805.07106.
[33] A. M. Eller and G. D. Moore, Phys. Rev. D **97**, 114507 (2018).
[34] D. Petcher and M. Lüscher, Nucl. Phys. **B225**, 53 (1983).
[35] U.-J. Wiese, Nucl. Phys. B, Proc. Suppl. **17**, 639 (1990).
[36] W. Bietenholz, R. Brower, S. Chandrasekharan, and U. J. Wiese, Phys. Lett. B **407**, 283 (1997).
[37] M. Lüscher and F. Palombi, J. High Energy Phys. 09 (2010) 110.
[38] R. Burkhalter, M. Imachi, Y. Shinno, and H. Yoneyama, Prog. Theor. Phys. **106**, 613 (2001).
[39] S. Aoki, H. Fukaya, S. Hashimoto, and T. Onogi, Phys. Rev. D **76**, 054508 (2007); I. Bautista, W. Bietenholz, A. Dromard, U. Gerber, L. Gonglach, C. P. Hofmann, H. Mejía-Díaz, and M. Wagner, Phys. Rev. D **92**, 114510 (2015).
[40] W. Bietenholz, K. Cichy, P. de Forcrand, A. Dromard, and U. Gerber, Proc. Sci., LATTICE2016 (**2016**) 321.
[41] M. Bögli, F. Niedermayer, M. Pepe, and U.-J. Wiese, J. High Energy Phys. 04 (2012) 117.
[42] A. Zamolodchikov (private communication).