


# Topological susceptibility in QCD with two flavors and 3-5 colors: A pilot study

Thomas DeGrand<sup>\*</sup>*Department of Physics, University of Colorado, Boulder, Colorado 80309, USA* (Received 29 April 2020; accepted 12 June 2020; published 25 June 2020)

I present a calculation of the topological susceptibility  $\chi_T$  in  $SU(N_c)$  gauge theory with  $N_c = 3-5$  colors and  $N_f = 2$  degenerate flavors of fermions. The results lie on a common curve when expressed in terms of the combination  $N_c m_{PS}^2 t_0$  where  $m_{PS}$  is the pseudoscalar meson mass and  $t_0$  is the flow parameter.  $\chi_T$  approaches its quenched value as the pseudoscalar mass becomes large. The lattice simulations use clover fermions. They are done at a single lattice spacing, roughly matched across  $N_c$ , and over a restricted range of fermion masses.

DOI: [10.1103/PhysRevD.101.114509](https://doi.org/10.1103/PhysRevD.101.114509)

## I. INTRODUCTION, MOTIVATION, AND RESULTS

Real world QCD, with  $N_c = 3$  colors, shares many features with its  $N_c \rightarrow \infty$  limit. Large  $N_c$  expectations [1,2] mostly arise from graph counting, in that only planar diagrams survive in the large  $N_c$  limit. The consequences of these expectations are often applied to nonperturbative observables, like masses or matrix elements, and these applications are qualitatively satisfied by experimental data.

Nonperturbative predictions really need nonperturbative checks, and there is a small lattice literature of simulations of QCD with  $N_c > 3$ . (See [3–7] for a selection of reviews and original work.) Simulation results generally agree with expectations. This paper is another check of large  $N_c$  counting. It is a calculation of the topological susceptibility with two flavors ( $N_f = 2$ ) of degenerate mass fundamental representation fermions and  $N_c = 3, 4$ , and 5 colors.

There are at least three approaches in the literature for studying large  $N_c$  QCD with lattice methods. The largest  $N_c$  values are attained by assuming volume independence and simulating on small spatial volumes (see [7] for a recent review). Fairly large  $N_c$  values (up to  $N_c = 17$ ) have been reached doing quenched simulations [5] on large volumes. Simulations using full QCD, with dynamical fermions, in large volumes are much more costly. However, part of the large  $N_c$  phenomenology is that fermion effects die away at large  $N_c$ . To see their effects die away, it is necessary to include dynamical fermions from the start.

Naive large  $N_c$  counting does not address possible effects depending on the fermion mass  $m_q$ . For many processes,  $N_c$  and  $m_q$  effects approximately factorize:  $Q(N_c, m_q) \sim N_c^p f(m_q)$ . Examples include meson masses ( $m_H$  vs  $m_q$ ), decay constants, and even baryon masses ( $M_B(N_c, m_q, J) = N_c m_0(m_q) + (J(J+1)/N_c)B(m_q) + \dots$  for angular momentum  $J$ ). The qualitative agreement of full QCD ( $N_c = 3$  with dynamical fermions) and quenched QCD (replacing a dynamical fermion by a quenched valence one) is a consequence of this factorization.

But there are (at least) two cases where this factorization should not occur. These cases occur for chiral observables and follow from the scaling of the pseudoscalar decay constant  $f_{PS}$  and condensate  $\Sigma$ :  $f_{PS} \propto N_c^{1/2}$  and  $\Sigma \propto N_c$ . (The behavior of  $f_{PS}$  is directly tested on the lattice; the second relation is only known indirectly: the squared pseudoscalar mass divided by the fermion mass  $m_{PS}^2/m_q$  is seen to be independent of  $N_c$  and this ratio is also proportional to  $\Sigma/f_{PS}^2$ .) The first case involves quantities scaling as  $m_{PS}^2/f_{PS}^2 \propto m_{PS}^2/N_c$  or  $m_q/N_c$ . Examples include higher order corrections to chiral observables,  $O = O_0(1 + C(m_{PS}^2/f_{PS}^2) \log(m_{PS}^2/\Lambda^2) + \dots)$ . These are typically hard to see in simulations because they are subleading corrections. One example, though, has been reported in Ref. [8], the dependence of the gradient flow scale  $t_0$  on  $m_{PS}^2/N_c$  as described by Golterman and Shamir [9].

The second case is the subject of this note: the topological susceptibility  $\chi_T$ . It has very different behavior in the quenched limit and at small fermion mass. In the former case  $\chi_T$  is a constant (call it  $\chi_Q$ ), which is nearly independent of  $N_c$ . In the latter case  $\chi_T \propto m_q \Sigma$  or  $m_{PS}^2 f_{PS}^2$ , so one ought to see scaling as  $\chi_T \propto N_c m_{PS}^2$  at small  $N_c m_{PS}^2$ . In fact, there is an old prediction of a functional form for all

\*thomas.degrand@colorado.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.

mass values, due to Di Vecchia and Veneziano [10] and Leutwyler and Smilga [11],

$$\chi_T = \frac{m_q \Sigma}{N_f} \left[ \frac{\chi_Q}{\chi_Q + m_q \Sigma / N_f} \right] \quad (1)$$

or

$$\frac{1}{\chi_T} = \frac{N_f}{m_q \Sigma} + \frac{1}{\chi_Q}. \quad (2)$$

(The small mass limit of this formula was also derived by Crewther [12].) With  $2m_q \Sigma = f_{PS}^2 m_{PS}^2$  (appropriate to the  $f_{PS} = 93$  MeV convention), Eq. (2) becomes

$$\frac{1}{\chi_T} = \frac{2N_f}{f_{PS}^2 m_{PS}^2} + \frac{1}{\chi_Q}, \quad (3)$$

and with  $f_{PS}(N_c) = \sqrt{N_c} f_0$  the expected scaling behavior of the pseudoscalar decay constant across  $N_c$ , we can write

$$\frac{1}{\chi_T} = \frac{2N_f}{N_c f_0^2 m_{PS}^2} + \frac{1}{\chi_Q}. \quad (4)$$

That is, the inverse topological susceptibility rises linearly from its quenched value with respect to the scaling variable  $1/(N_c m_{PS}^2)$  or  $1/(N_c m_q)$ .

The purpose of this paper is to take a first look at  $\chi_T(m_q, N_c)$ —as the title says, “a pilot study.” This means

- (i)  $N_f = 2$
- (ii)  $N_c = 3, 4, 5$
- (iii) One lattice spacing (loosely speaking), roughly matched across  $N_c$  using a gluonic observable (alternatively, roughly matched in bare 't Hooft coupling  $\lambda = g^2 N_c$ )
- (iv) One simulation volume, a range of intermediate mass fermions, and moderate statistics, so all observations are still tentative

The goal of the paper is to answer a set of physics questions and a set of simulation questions. The physics questions are

- (1) Does  $\chi_T(m_q, N_c)$  actually scale as  $\chi_T(m_q N_c)$  [equivalently  $\chi_T(m_{PS}^2 N_c)$ ], smoothly connected to  $\chi_Q$  at large  $m_{PS}^2 N_c$ ?
- (2) Does  $\chi_T(m_q, N_c)$  follow the Di Vecchia, Veneziano, Leutwyler, Smilga functional form?

The answers are (1) yes, apparently and (2) qualitatively, but not quantitatively, at the lattice spacings studied.

The main simulation question is: it is well known that in ordinary  $N_c = 3$  QCD  $\chi_T$  has a very long simulation autocorrelation time  $\tau$ . How severe an issue is this across  $N_c$ ? The answer is:  $\tau$  grows with  $N_c$ .  $N_c = 3$  or 4 seem to be manageable with the naive approach I took to study the problem, but  $N_c = 5$  already shows clear issues.

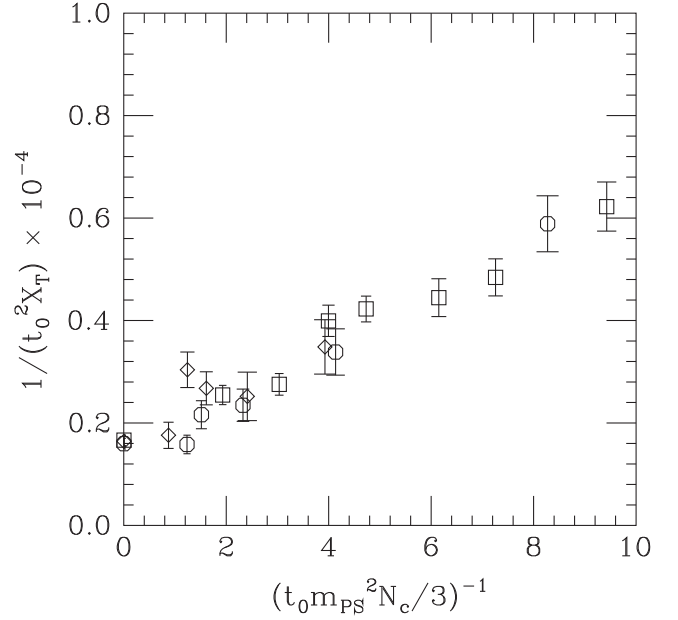


FIG. 1. The inverse topological susceptibility, scaled by  $t_0^2$ , versus  $(t_0 m_{PS}^2 N_c / 3)^{-1}$ . Data are squares for  $N_c = 3$ , octagons for  $N_c = 4$  and diamonds for  $N_c = 5$ .

The result of the simulations described here is displayed in Figs. 1 and 2 (the overall scale is set by the “flow parameter”  $t_0$ ). Monte Carlo results collapse to a common curve, which is a straight line in Fig. 1. In that figure we see that the line extrapolates to the quenched topological susceptibility measured by Ref. [13]. Evidently, the effects

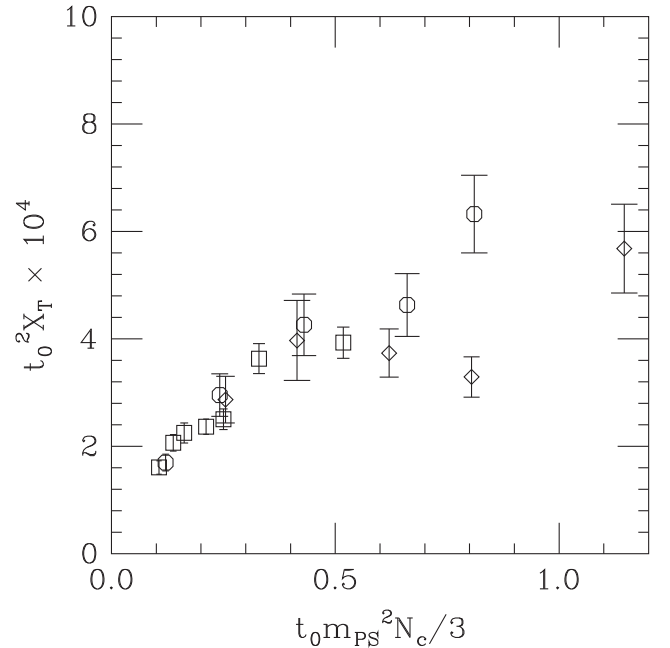


FIG. 2. A more conventional presentation of the data:  $t_0^2 \chi_T$  versus  $t_0 m_{PS}^2 N_c / 3$ .

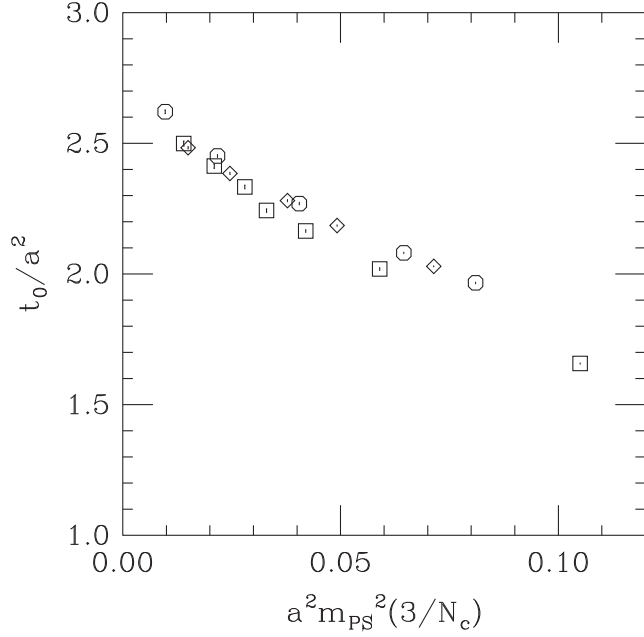


FIG. 3. The flow scale  $t_0(m_{PS})$ , versus  $1/N_c$  times the squared pseudoscalar mass in lattice units,  $(am_{PS})^2/N_c$ , for  $N_c = 3$  (squares), 4 (octagons), and 5 (diamonds).

of dynamical fermions for this observable do not depend separately on  $N_c$  and the fermion mass, but on the combination  $N_c m_q$  or  $N_c m_{PS}^2$ .

In 2001 Dürr [14] presented a similar plot, and comparison to Eq. (4), with  $N_c = 3$  data.

As a contrast, Fig. 3 shows the flow scale  $t_0$  versus  $(am_{PS})^2/N_c$ , the other non-factorizing mass and  $N_c$  dependence. I am just showing it in passing, since it has been discussed before.

The outline of the paper is as follows: Section II describes the calculations: it covers data sets, simulation methodology, and has a discussion of gradient flow based observables. Here is where I describe how I dealt with the long autocorrelations in the data. Section III presents results: first, comparisons of my data to high statistics calculations of the quenched topological susceptibility, then a comparison to previous calculations of the  $N_f = 2$   $SU(3)$  susceptibility. These are checks to make sure that the present calculation seems to be in order. I then discuss my results for the  $N_f = 2$  susceptibility across  $N_c$ . Section IV is a brief summary. A reminder of the derivation of Eq. (2) is given in an Appendix.

## II. TECHNICAL ASPECTS OF THE CALCULATION

### A. Simulation methodology, lattice actions, data sets

The dynamical fermion simulations contained two degenerate flavors of Wilson–clover fermions. The gauge action is the usual plaquette action, with the bare gauge

TABLE I.  $N_f = 2$  dynamical fermion data plotted in the figures. The column labeled by  $N$  gives the number of lattice analyzed for  $t_0$  and  $\chi_T$ .

$\kappa$	$am_q$	$(am_{PS})^2$	$t_0/a^2$	$t_0^2 \chi_T \times 10^4$	$N$
<i>SU(3) <math>\beta = 5.4</math></i>					
0.1250	0.105	0.312(2)	1.657(3)	3.93(29)	500
0.1265	0.059	0.163(2)	2.019(6)	3.63(28)	500
0.1270	0.042	0.116(2)	2.165(6)	2.50(19)	500
0.1272	0.033	0.094(2)	2.243(7)	2.37(14)	500
0.1274	0.028	0.070(2)	2.333(7)	2.25(19)	500
0.1276	0.021	0.057(1)	2.413(8)	2.06(15)	500
0.1278	0.014	0.042(1)	2.500(9)	1.61(12)	500
<i>SU(4) <math>\beta = 10.2</math></i>					
0.1245	0.108	0.309(1)	1.966(4)	6.32(72)	490
0.1252	0.086	0.238(2)	2.081(3)	4.63(58)	490
0.1262	0.054	0.142(1)	2.269(4)	4.26(57)	490
0.1270	0.029	0.074(1)	2.451(5)	2.95(39)	500
0.1275	0.013	0.035(1)	2.621(7)	1.70(16)	500
<i>SU(5) <math>\beta = 16.4</math></i>					
0.1240	0.119	0.339(1)	2.029(2)	5.68(83)	590
0.1252	0.082	0.221(1)	2.185(3)	3.29(38)	590
0.1258	0.063	0.163(1)	2.281(4)	3.74(45)	490
0.1265	0.041	0.104(1)	2.385(4)	3.97(74)	490
0.1270	0.025	0.062(0)	2.483(4)	2.87(44)	490

coupling  $g_0$  parameterized by  $\beta = 2N_c/g_0^2$ . The fermion action uses gauge connections defined as normalized hypercubic (nHYP) smeared links [15–17] (with the arbitrary  $N_c$  implementation of Ref. [6]). The bare quark mass  $m_0^q$  is introduced via the hopping parameter  $\kappa = (2m_0^q a + 8)^{-1}$ . The clover coefficient is fixed to its tree level value,  $c_{SW} = 1$ . The updating scheme is the Hybrid Monte Carlo (HMC) algorithm [18–20] with a multilevel Omelyan integrator [21] and multiple integration time steps [22] with one level of mass preconditioning for the fermions [23].

All lattice volumes are  $16^3 \times 32$  sites. The gauge fields experience periodic boundary conditions; the fermions are periodic in space and antiperiodic in time.

All data sets are 5000 to 6000 trajectories in length. Lattices used for analysis are spaced a minimum of 10 HMC time units apart, so individual bare parameter sets contain 490–600 stored lattices. All data sets (individual  $(\beta, \kappa)$  values) are based on a single stream.

The data sets were collected at approximately equal values of lattice spacing. (The bare gauge coupling is fixed at each  $N_c$  and only  $\kappa$  is varied.) This precludes a discussion of lattice artifacts. However, comparisons across  $N_c$ , or with large  $N_c$  phenomenology, can be done at any value of the lattice spacing.

The data sets are extensions of ones presented in Refs. [6,8] and full spectroscopy is presented in the first of these references. Table I summarizes relevant information for the runs. Across the data sets,  $m_{PS}^2$  the squared

pseudoscalar meson mass is roughly linear in the Axial Ward Identity fermion mass  $m_q$ . The ratio  $(m_{PS}/m_V)^2$  where  $m_V$  is the vector meson mass, spans the range 0.16–0.64.

### B. Gradient flow for length scale

The lattice spacing and the topological charge are measured using the technique of gradient flow [24,25], a smoothing method for gauge fields via diffusion in a fictitious (fifth dimensional) time  $t$ . In continuum language, a smooth gauge field  $B_{t,\mu}$  is constructed through an iterative process

$$\begin{aligned}\partial_t B_{t,\mu} &= D_{t,\mu} B_{t,\mu\nu} \\ B_{t,\mu\nu} &= \partial_\mu B_{t,\nu} - \partial_\nu B_{t,\mu} + [B_{t,\mu}, B_{t,\nu}],\end{aligned}\quad (5)$$

beginning with the original one,

$$B_{0,\mu}(x) = A_\mu(x). \quad (6)$$

A squared length  $t_0$  is defined through the field strength tensor built using smoothed degrees of freedom,  $G_{t,\mu\nu}$ , using the observable

$$\langle E(t) \rangle = \frac{1}{4} \langle G_{t,\mu\nu} G_{t,\mu\nu} \rangle. \quad (7)$$

It is set by fixing the quantity  $t_0^2 \langle E(t_0) \rangle$  to some value  $C(N_c)$

$$t_0^2 \langle E(t_0) \rangle = C(N_c). \quad (8)$$

The choice of  $C(N_c)$  across  $N_c$  is somewhat arbitrary, as it is for  $N_c = 3$ . There is a natural set of choices motivated by the perturbative expansion

$$\begin{aligned}t^2 \langle E \rangle &= \frac{3}{32\pi} (N_c^2 - 1) \alpha(q) [1 + k_1 \alpha + \dots] \\ &= \frac{3}{32\pi} \frac{N_c^2 - 1}{N_c} (4\pi\lambda(q)) [1 + k_1 \alpha + \dots]\end{aligned}\quad (9)$$

where  $\alpha(q)$  is the strong coupling constant at momentum scale  $q \propto 1/\sqrt{t}$  and  $\lambda(q)$  is the corresponding 't Hooft coupling. The large  $N_c$  limit, where matching gluonic observables is achieved by matching the bare 't Hooft couplings, is  $t^2 \langle E \rangle \propto N_c$ . Beyond that, there are many possible choices. In Ref. [8], I tested the leading  $C(N_c) \propto N_c$  behavior by matching  $t_0$  to another gluonic observable, an inflection point on the static potential called  $r_1$ . This choice amounts to fixing the inflection point across  $N_c$ . Most other people adopt a different convention,

$$C(N_c) = C(3) \left( \frac{3N_c^2 - 1}{8N_c} \right), \quad (10)$$

taking  $C(3) = 0.3$  as the usual value used in  $SU(3)$ . This amounts to saying that the ratio  $t_0/r_1$  has a  $1/N_c$  variation away from its  $N_c = 3$  value (or, the large  $N_c$  limit is different from the  $N_c = 3$  ratio), nothing more. I will follow this choice, rather than the one of Ref. [8], because I want to match the quenched results of Ref. [13], and they use the convention of Eq. (10).

The extraction of  $t_0$  from lattice data is standard and is described in Ref. [8]. The gradient flow differential equation is integrated numerically as described by Lüscher [25]. Calculations used the usual ‘‘clover’’ definition of  $E(t)$ . An autocorrelation analysis will be described after the next subsection.

### C. Gradient flow for topological charge—definitions

The topological charge density is defined as

$$q_t(x) = -\frac{1}{32\pi^2} \epsilon_{\mu\nu\rho\sigma} \text{Tr} G_{t,\mu\nu}(x) G_{t,\rho\sigma}(x) \quad (11)$$

and is computed using gauge fields at flow time  $t$  as

$$Q(t) = a^4 \sum_x q_t(x). \quad (12)$$

In a system with periodic boundary conditions, the topological susceptibility is simply

$$\chi_T = \frac{1}{V} \langle Q(t)^2 \rangle. \quad (13)$$

This point actually needs a bit more discussion. Equation (13) implicitly assumes that  $\langle Q(t) \rangle = 0$  when averaged over the measurements taken in the simulation. The observation of  $\langle Q(t) \rangle \neq 0$  is an artifact, indicating that the data has long time autocorrelations. To see if this affects my results, I will compare  $\langle Q^2 \rangle$  to the full correlator  $C(t) = \langle Q(t)^2 \rangle - \langle Q(t) \rangle^2$ .

A second issue is that, at any nonzero lattice spacing,  $\chi_T(t)$  depends on  $t$ . Taking the continuum limit involves measuring  $t_0^2 \chi(t)$  at several lattice spacings and taking the  $a \rightarrow 0$  limit. (This could be done by plotting the data versus  $1/t$  or, reinserting the lattice spacing  $a$ , plotting versus  $a^2/t_0$ ). In principle, this could be done for any  $t$ . The data in this study are all at one lattice spacing, so one has to ask whether the physics hints given by a study at one value of  $a$  are sensitive to the choice of operator [choice of  $t$  for  $Q(t)$ ].

### D. Data analysis

Both  $t_0$  and the topological charge show simulation time autocorrelations. I attempted to estimate the autocorrelation time through the autocorrelation function defined as

$$\rho_A(\tau) = \frac{\Gamma_A(\tau)}{\Gamma_A(0)} \quad (14)$$

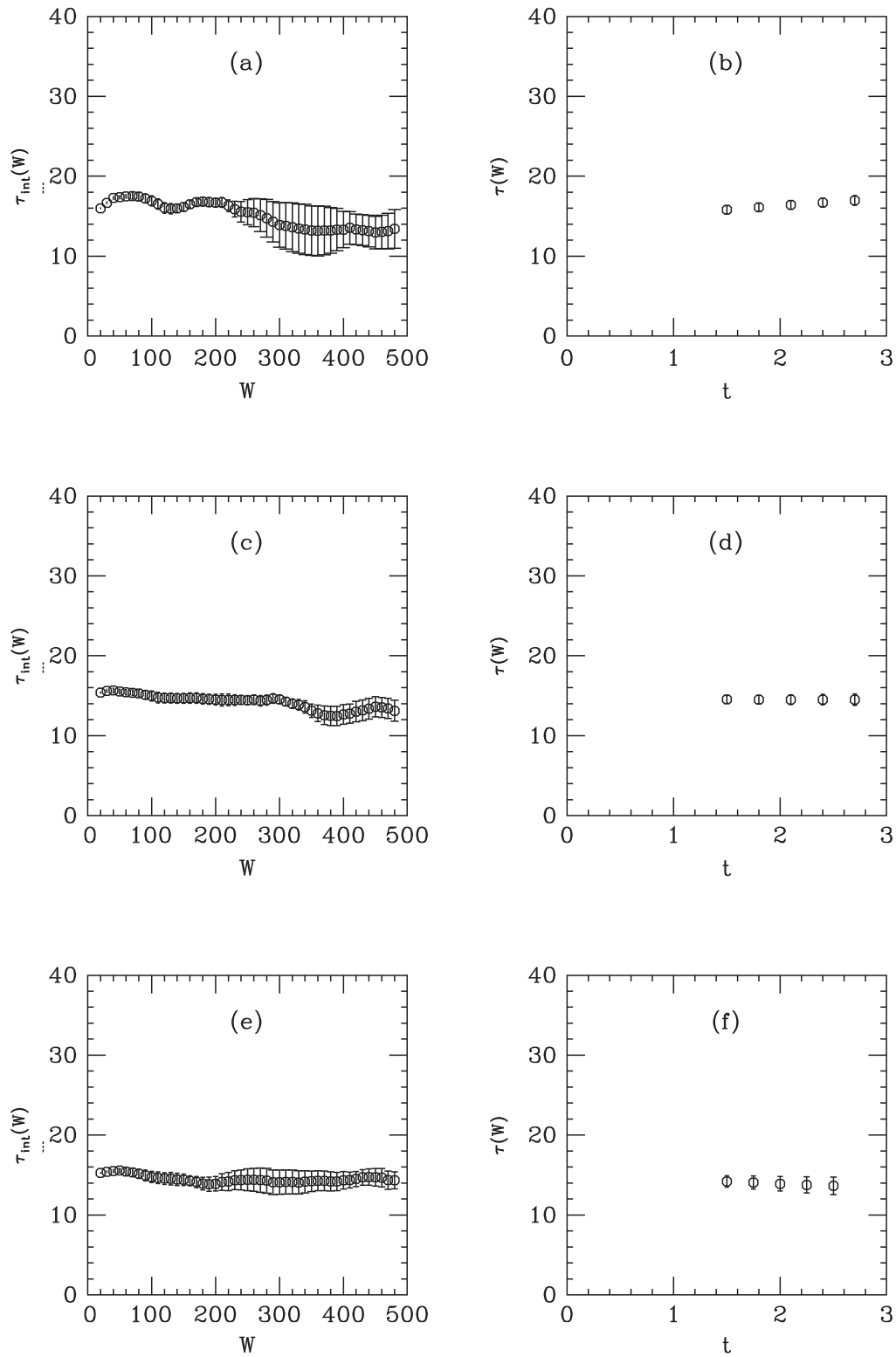


FIG. 4. Integrated autocorrelation time  $\tau_{\text{int}}(W)$  for  $t^2 E(t)$  at fixed  $t$  versus  $W$  and versus  $t$  at fixed  $W = 200$ : a) and b)  $SU(3)$ ,  $\kappa = 0.127$ ; c) and d)  $SU(4)$ ,  $\kappa = 0.1262$ ; e) and f)  $SU(5)$ ,  $\kappa = 0.127$ .



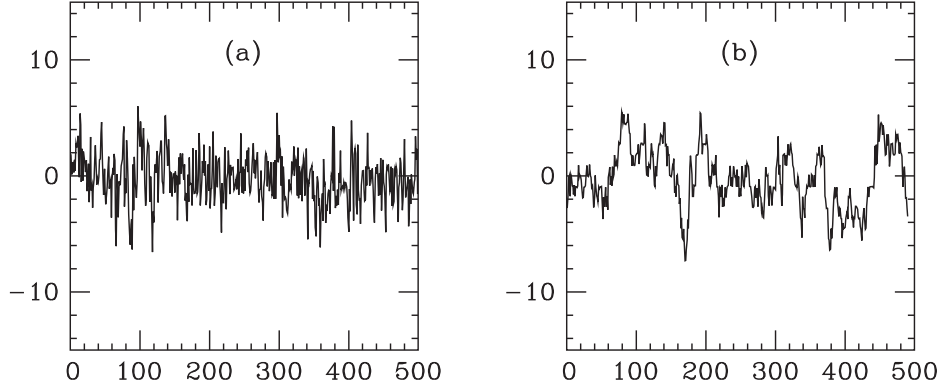


FIG. 5. Time history of the topological charge at  $t = 3$  for (a) an  $SU(3)$  data set ( $\kappa = 0.1274$ ) and (b) an  $SU(5)$  one,  $\kappa = 0.127$ .

(for a generic observable  $A$ ) where

$$\Gamma_A(\tau) = \langle (A(\tau) - \bar{A})(A(0) - \bar{A}) \rangle. \quad (15)$$

The integrated autocorrelation time, up to a window size  $W$ , is

$$\tau_{\text{int}}(W) = \frac{1}{2} + \sum_{\tau=1}^W \rho(\tau). \quad (16)$$

Unless the total length in time of the data set is much larger than the autocorrelation time, it is difficult to estimate an error for  $\tau_{\text{int}}$ . I analyzed my data sets by breaking them into multiple parts, each part being order 1000 trajectories or 100 saved lattices, computing  $\tau$  on each part, and taking an error from the part-to-part fluctuations.

The analysis of  $t_0$  is straightforward. I show a few representative figures, since the data look quite similar across fermion mass and  $N_c$ . Figure 4 shows plots of the integrated autocorrelation time  $\tau_{\text{int}}(W)$  for  $t^2 E(t)$  at  $t = 2.1$  (for  $N_c = 3$  and 4) and  $t = 2$  for  $N_c = 5$  versus  $W$ , and  $\tau_{\text{int}}(W = 200)$  vs flow time  $t$ . The values of  $t_0$  in the table are taken from a jackknife analysis dropping two successive lattices, since these figures indicate that the autocorrelation time is 15-20 trajectories.

Now for the topological charge. The autocorrelation time is large for all  $SU(5)$  data sets. This can be seen by eye from time histories: compare Figs. 5 for an  $SU(3)$  history and an  $SU(5)$  one.

I repeat the calculation of autocorrelation times for  $Q(t)$ . In contrast to the results for  $t^2 E(t)$ , in general  $\tau_{\text{int}}(W)$  is an irregular function of  $W$ . This is already an indicator of long correlations in the data. Results for the same parameter values as in Fig. 4 are shown in Fig. 6.

Fit results come from a jackknife analysis, removing sets of lattices whose length is longer than the estimated integrated autocorrelation time. This would be  $n_J$  successive lattices for  $\tau_{\text{int}} = 10n_J$  molecular dynamics time units. To be explicit: for a given jackknife I compute the

averages  $\langle Q(t) \rangle$ ,  $\langle Q^2(t) \rangle$  and  $C(t) = \langle Q^2(t) \rangle - \langle Q(t) \rangle^2$ ; the uncertainty of each comes from a jackknife. I varied the size of the jackknife beyond the estimate of the integrated autocorrelation time. I estimate the fractional error from loss of statistics as

$$\frac{\Delta(\Delta C(t))}{\Delta C(t)} = \sqrt{\frac{2}{n}} \quad (17)$$

where  $n = N/n_J$ . That gives a rough error bar. The uncertainty in  $C(t)$  increases with  $n_J$  and then either saturates, or at least the growth becomes smaller than what statistics allows one to see. Results for  $t = 3$  are shown in Fig. 7. Other  $t$  values are similar. The results of Fig. 7 suggest the size of the jackknife used to present results. For  $N_c = 3$ , the autocorrelation analysis suggests an autocorrelation time of about 20 trajectories, reasonably constant across  $\kappa$  values and for  $t$  in the range of about 1 to 6, and hence a cut  $n_J = 2$ . Figure 7 encourages setting the jackknife cut at  $n_J = 4$ . For  $N_c = 4$ , the autocorrelation time is about 30 trajectories but jackknife errors do not saturate until  $n_J = 8$ . Finally, for  $N_c = 5$ , the autocorrelation time is about 50 trajectories for the two smallest  $\kappa$  values and 100 for the others, but Fig. 7 instructs us to take  $n_J = 10$  for the three smallest  $\kappa$  values and 20 for the others.

Now for fits to the data. I observe, generally, that at small  $t$ ,  $Q(t)$  has a Gaussian distribution. At large  $t$ , individual configurations “cool,” that is,  $Q$  peaks at equally spaced, roughly integer values. This appears to happen at smaller  $t$  for  $N_c = 5$  than it does for  $N_c = 3$ . I test that the data is Gaussian using the Kolmogorov-Smirnov test [26]. It compares the integrated distributions (the cumulants) of the measured data  $C(x)$  and the theoretical prediction  $P(x)$ . The cumulant of the measured data is  $C(x) = n(x)/N$  where  $n(x)$  is the number of data points with a value smaller than  $x$  and  $N$  the total number of data points. The theoretical prediction for this quantity is found by integrating the distribution:  $P(x) = \int_{-\infty}^x f(y) dy$ . The

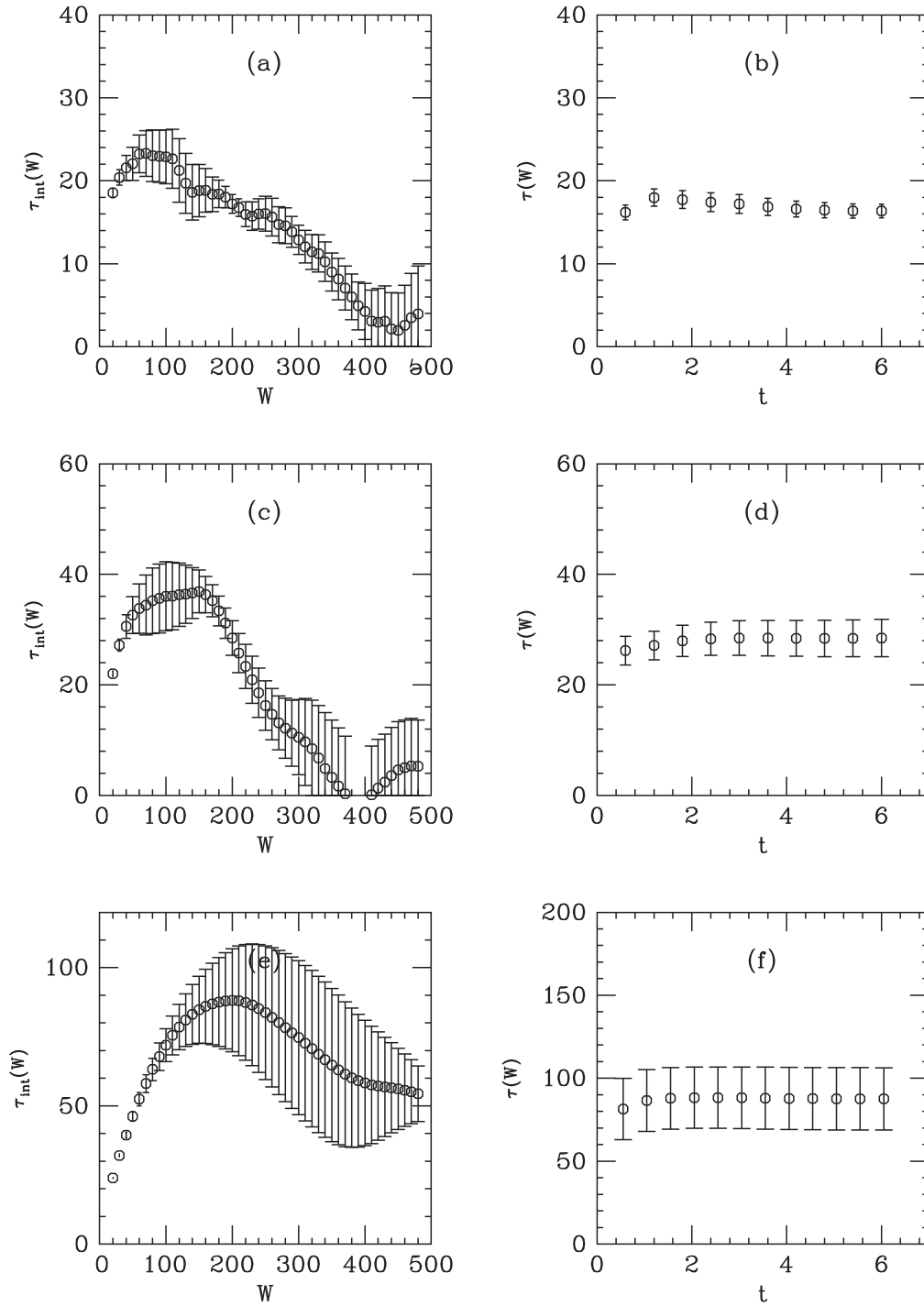


FIG. 6. Integrated autocorrelation time  $\tau_{\text{int}}(W)$  for  $Q(t)$  at fixed  $t$  versus  $W$  and versus  $t$  at fixed  $W = 200$ : (a) and (b)  $SU(3)$ ,  $\kappa = 0.127$ ; (c) and (d)  $SU(4)$ ,  $\kappa = 0.1262$ ; (e) and (f)  $SU(5)$ ,  $\kappa = 0.127$ .

quantity of interest is the largest deviation of  $P$  and  $C$ :  $D = \max_x |P(x) - C(x)|$ . From this the confidence level is given by

$$Q_{KS}(\sqrt{ND}) \quad (18)$$

where

$$Q_{KS}(x) = 2 \sum_{j=1}^{\infty} (-)^{j-1} \exp(-2j^2 x^2). \quad (19)$$

(Note larger  $Q_{KS}$  is better.)

So far, I have not specified a flow time for  $C(t)$ , so I did fits for a range of  $t$  values. Results for  $C(t)$  and  $\langle Q(t) \rangle$  are

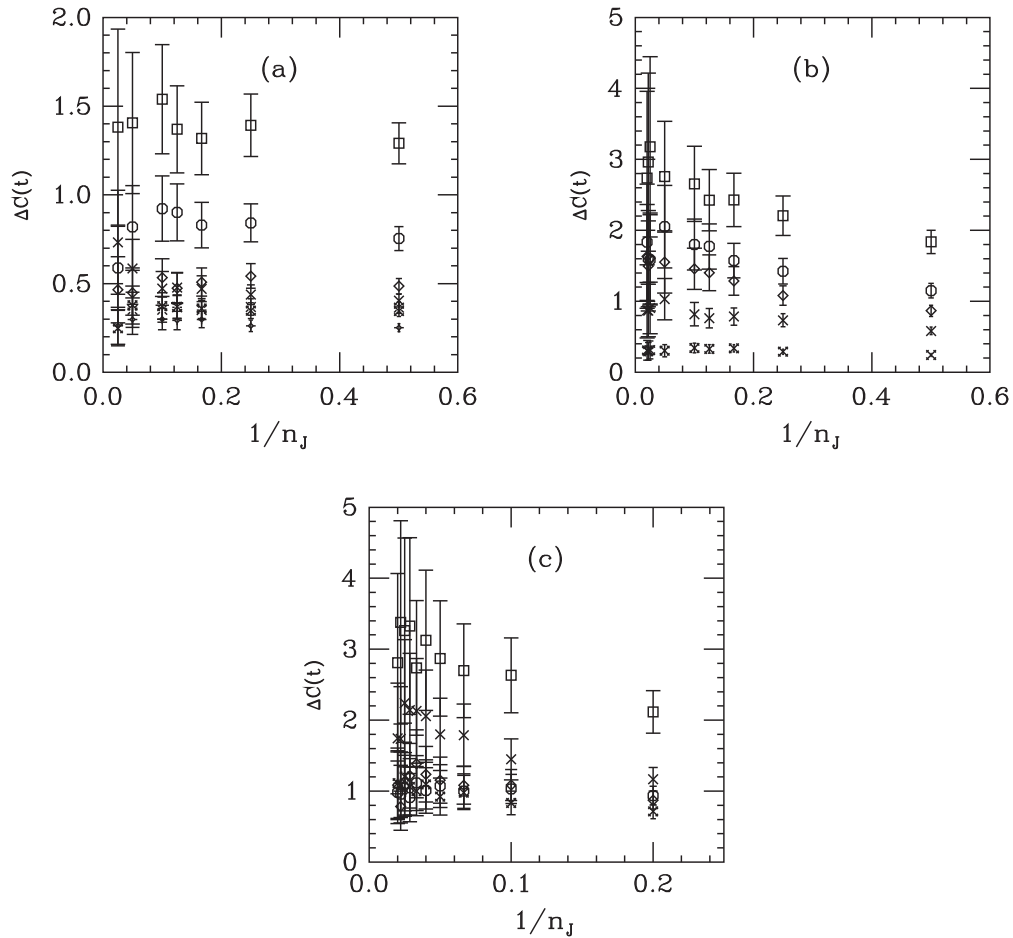


FIG. 7. Uncertainty in  $C(t)$  at  $t = 3$  as a function of the inverse jackknife size for (a)  $SU(3)$ , (b)  $SU(4)$ , and (c)  $SU(5)$ . The different plotting symbols correspond to different  $\kappa$  values, the ordering top to bottom is with decreasing fermion (or pseudoscalar meson) mass.

shown in Figs. 8 and 9. Plots of  $\langle Q^2 \rangle$  are almost identical to those of  $C(t)$ .  $\langle Q(t) \rangle$ , in contrast, is nearly independent of  $t$ .

$Q(t)$  should average to zero. The figures, and the data for  $t = 3$  presented in Table II, show several cases where  $\langle Q \rangle$  sit two standard deviations away from zero. However, even for the most extreme deviations ( $SU(5)$ ,  $\kappa = 0.1265$  and  $0.1265$ ) the difference between  $C(t)$  and  $\langle Q^2(t) \rangle$  is less than the RMS value of the uncertainties of the two determinations.

In all cases, the data is (nearly) Gaussian about its mean. This is checked through a cumulant analysis where the expectation is (the integral of) a Gaussian with  $\langle Q \rangle$  and  $\langle Q^2 \rangle - \langle Q \rangle^2$  taken from Table II. Deviations from Gaussianity occur at long flow time because  $Q$  has cooled to approximate integers, as revealed by steps in the cumulant. It did not seem to be worthwhile to guess a more complicated (Gaussian with steps) distribution for comparison.

A few pictures of cumulants shown in Fig. 10 illustrate the fits.

Figure 8 shows that once  $t$  becomes greater than about 2.5, the value of  $C(t)$  (and  $\langle Q(t)^2 \rangle$ , which is almost

identical) roughly forms a plateau. In a better study, I would fix  $t$  to any convenient value and extrapolate  $\langle Q(t)^2 \rangle$  to  $a = 0$ . For the remainder of this study I will just fix  $t$  to  $t = 3$ . Results are summarized in Table II. All phenomenology in the next section will be done with  $\chi_T = \langle Q(t = 3)^2 \rangle / V$ .

### III. RESULTS

With data sets at one bare gauge coupling per  $N_c$  it is hard to quantify lattice artifacts. I can compare my results to other simulations and ask if they look reasonable. There are two places where this is done.

#### A. Comparison with high precision quenched results

The first one is the quenched limit. The authors of Refs. [27] and [13] published high statistics data for  $t_0^2 \chi_T$  for  $N_c = 2-6$ . I collected a data set much smaller than theirs but comparable to my dynamical sets in size and in lattice spacing, to check against theirs. It is recorded in Table III. My sets are 500 measurements per  $N_c$ , each



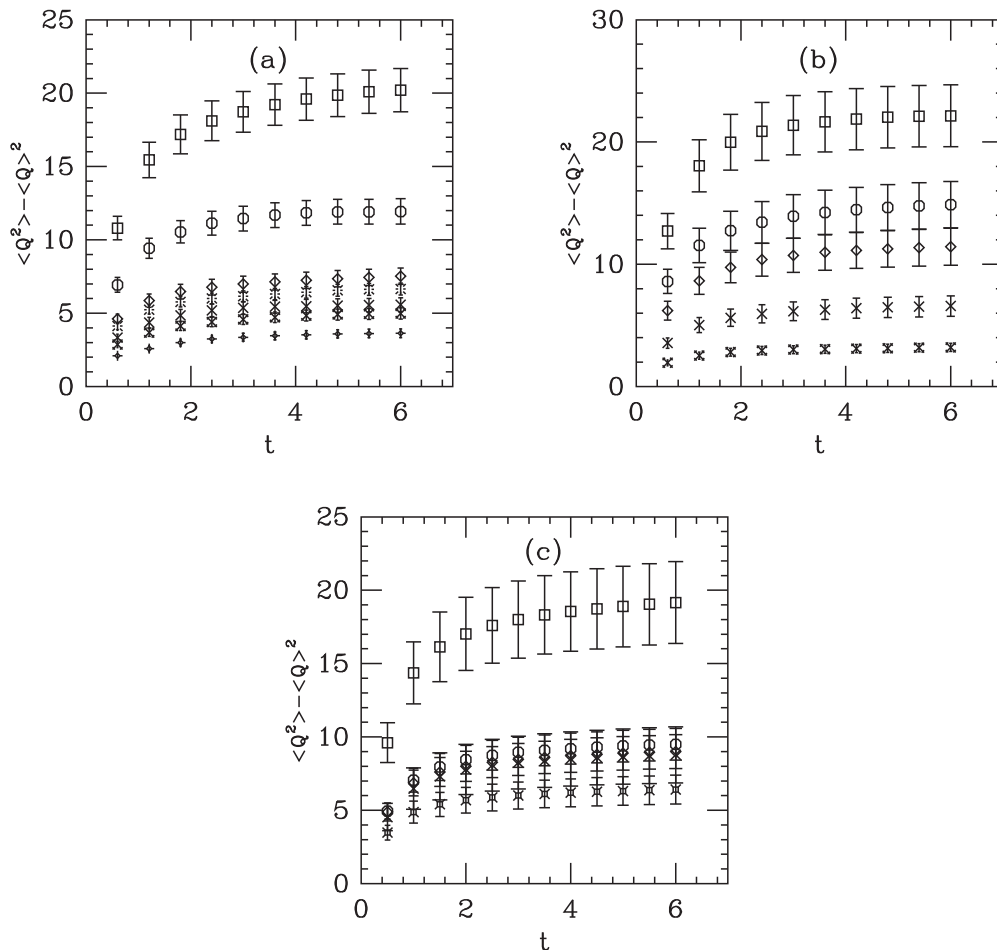


FIG. 8.  $C(t)$  versus  $t$  as a function for (a)  $SU(3)$ , (b)  $SU(4)$ , and (c)  $SU(5)$ . The different plotting symbols correspond to different  $\kappa$  values; the ordering top to bottom is with decreasing fermion (or pseudoscalar meson) mass.

spaced 100 sets of sweeps through the lattice, each sweep consisting of a mix of four Brown—Woch microcanonical over-relaxation steps [28] and a Cabibbo—Marinari heat bath update [29], performed on all  $N_c(N_c - 1)/2$   $SU(2)$  subgroups of the  $SU(N_c)$  link variables.

Figure 11 shows the comparison. Within my large errors, my results are compatible with the high statistics results of Refs. [13,27].

### B. Comparison with high precision $N_f = 2$ $SU(3)$

The next comparison is with high precision  $N_c = 3$ ,  $N_f = 2$  results. I have only been able to find a few recent calculations (most recent studies are for  $N_f > 2$  with physical strange (and beyond) fermion masses). But there are three useful sets.

The first is that of Ref. [30]. I used essentially their techniques: the topological susceptibility is measured from flow. Reference [30] presented data from three small lattice spacings,  $a = 0.075$  fm, 0.065 fm and 0.048 fm (speaking nominally; flow parameters, and hence the lattice spacing  $a$  are computed at each value of bare fermion mass) on very

large lattices. The authors of Ref. [30] provided me with tables of  $t_0^2 \chi_T$  versus  $t_0 m_{PS}^2$ . Most of their data is at smaller pseudoscalar mass than mine.

The next set is that of Ref. [31]. It is a calculation using overlap fermions in a sector of fixed topology. The topological susceptibility is computed from the autocorrelation function of the topological charge density, which in turn is defined in terms of fermionic observables. The lattice spacing is about 0.12 fm. They publish a table of  $\chi_T r_0^4$  versus  $m_{PS} r_0$ , where  $r_0$  is the Sommer parameter [32], an inflection point on the heavy quark potential. I take their value  $r_0 = 0.49$  fm and the value of  $t_0$  quoted in the review by Sommer, Ref. [33],  $\sqrt{t_0} = 0.154$  fm (from Refs. [34,35]) to rescale the data.

The third paper is Ref. [36]. It is a calculation with dynamical domain wall fermions. The topological charge is determined from the eigenmodes of valence overlap fermions. Taking pseudoscalar masses from their Ref. [37], I rescale their numbers (quoted in GeV units but determined from  $r_0$ ). Their data is also shown in Fig. 12.

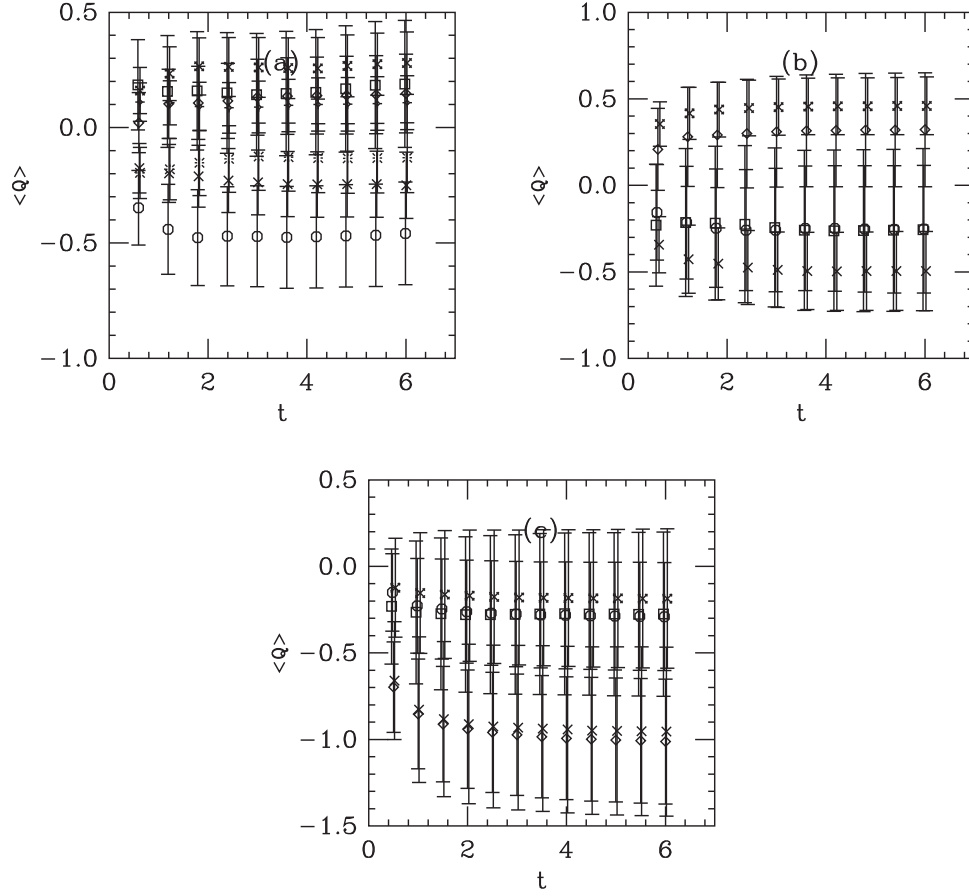


FIG. 9.  $\langle Q(t) \rangle$  versus  $t$  as a function for (a)  $SU(3)$ , (b)  $SU(4)$ , and (c)  $SU(5)$ . The different plotting symbols correspond to different  $\kappa$  values; for numerical values in this cluttered graph, see Table II. The  $x$  axes are slightly displaced for viewing.

The line in the figure is  $t_0^2 \chi_T = (t_0 f_{PS}^2 / 4) m_{PS}^2$  with  $\sqrt{t_0} = 0.154$  fm and  $f_{PS} = 93$  MeV. I show my own data for  $\chi_T(t)$  for two choices of  $t$ , 3 and 1.2.

What points am I trying to make with this busy figure? To begin, at the lattice spacings of these data sets, lattice artifacts are large and are rather different for the two simulations based on flow and the ones based on fermions with exact or nearly exact chiral symmetry. The susceptibility measured by flow is expected to have a lattice artifact  $A$  of the form

$$t_0^2 \chi_T = b t_0 m_{PS}^2 + A \quad (20)$$

where  $A$  scales as  $a^2$ . This is what the authors of Ref. [30] saw. This has been checked in the chiral limit by Münster and Wulkenhaar [38]. In contrast, fermionic zero modes should drive  $\chi_T$  to zero as the fermion mass vanishes. My own fits to the data of Refs. [31] and [36] have intercepts  $A = -0.1(1) \times 10^{-4}$  and  $-0.04(3) \times 10^{-4}$  respectively, while the three sets of Ref. [30] are  $A = 1.23(14) \times 10^{-4}$ ,  $1.02(14) \times 10^{-4}$ , and  $0.02(4) \times 10^{-4}$ . Equation (20) is a good fit to all these data sets. The interesting quantity in Eq. (20) is  $b$ , which should be

TABLE II. Topological charge and related quantities for  $N_f = 2$ , all at flow time  $t = 3$ .

$\kappa$	$\langle Q \rangle$	$\langle Q^2 \rangle - \langle Q \rangle^2$	$\langle Q \rangle^2$
$SU(3) \beta = 5.4$			
0.1250	0.14(27)	18.73(139)	18.75(139)
0.1265	-0.47(22)	11.45(84)	11.68(90)
0.1270	0.13(17)	6.99(54)	7.00(54)
0.1272	-0.13(15)	6.15(37)	6.17(37)
0.1274	-0.24(14)	5.36(44)	5.42(45)
0.1276	0.26(13)	4.58(35)	4.65(35)
0.1278	0.09(11)	3.36(26)	3.37(26)
$SU(4) \beta = 10.2$			
0.1245	-0.24(46)	21.38(242)	21.44(245)
0.1252	-0.26(36)	13.92(177)	14.01(177)
0.1262	0.31(32)	10.73(140)	10.84(146)
0.1270	-0.49(22)	6.16(76)	6.44(86)
0.1275	0.45(16)	3.02(33)	3.24(30)
$SU(5) \beta = 16.4$			
0.1240	-0.28(46)	18.00(263)	18.09(263)
0.1252	-0.28(30)	8.95(103)	9.03(103)
0.1258	-0.97(35)	8.46(109)	9.41(113)
0.1265	-0.93(47)	8.25(180)	9.15(172)
0.1270	-0.18(39)	6.00(92)	6.10(93)

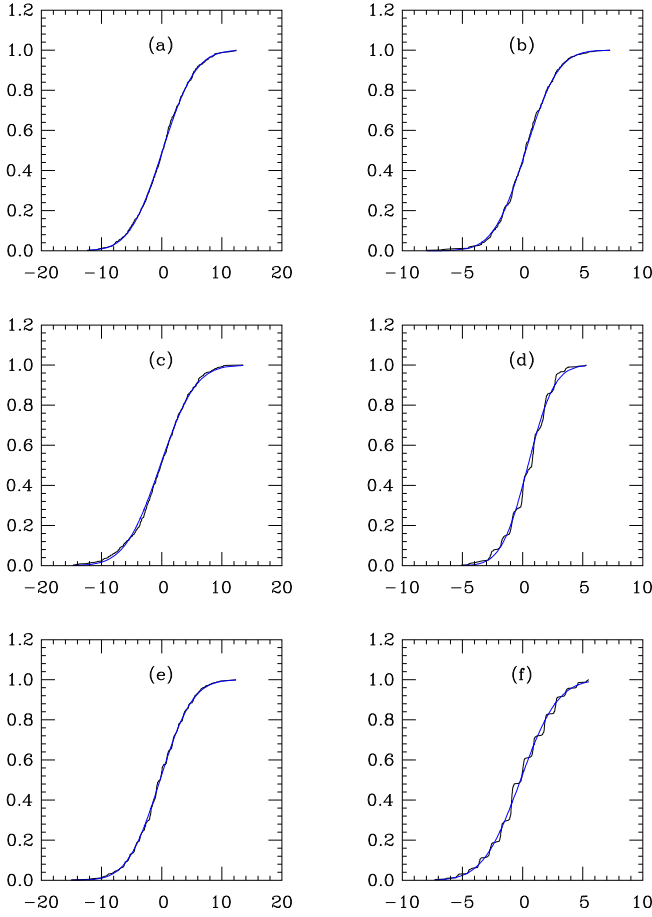


FIG. 10. Cumulants and fits for selected data sets, all at flow time  $t = 3$ .  $D$  is the maximum deviation of the cumulant from the error function and  $Q$  is defined in Eqs. (18). The mean and deviation are taken from Table II.  $SU(3)$ , (a)  $\kappa = 0.125$ — $D = 0.027$ ,  $Q = 0.84$ , (b)  $\kappa = 0.1276$ — $D = 0.028$ ,  $Q = 0.80$ ;  $SU(4)$ , (c)  $\kappa = 0.1245$ — $D = 0.035$ ,  $Q = 0.57$ , (d)  $SU(4)$ ,  $\kappa = 0.1275$ — $D = 0.075$ ,  $Q = 0.007$ ;  $SU(5)$ , (e)  $\kappa = 0.124$ — $D = 0.037$ ,  $Q = 0.38$ , (f)  $\kappa = 0.127$ — $D = 0.066$ ,  $Q = 0.027$ .

$b = t_0 f_{PS}^2/4$  from the leading chiral behavior. This is about  $13.3 \times 10^{-4}$  with  $\sqrt{t_0} = 0.154$  fm and  $f_{PS} = 93$  MeV. The line shows this behavior. A comparison with a ruler shows that the other groups'  $SU(3)$  data is consistent with this value, even though, strictly speaking,  $b$  should have its own lattice artifacts and one would expect agreement only in the continuum limit.

Most of my data is at too large pseudoscalar mass to be expected to be in the linear regime. At best, the lightest three

TABLE III. Quenched data ( $\langle Q \rangle$  etc at  $t = 3$ ) plotted in Fig. 11.

$N_c$	$\beta$	$t_0/a^2$	$\langle Q \rangle$	$\langle Q^2 \rangle - \langle Q \rangle^2$	$\langle Q \rangle^2$
3	5.9	2.255(10)	-0.24(17)	15.43(96)	15.49(95)
4	10.8	2.316(3)	0.37(17)	15.17(97)	15.31(100)
5	17.1	2.267(2)	-0.19(20)	15.49(99)	15.52(99)

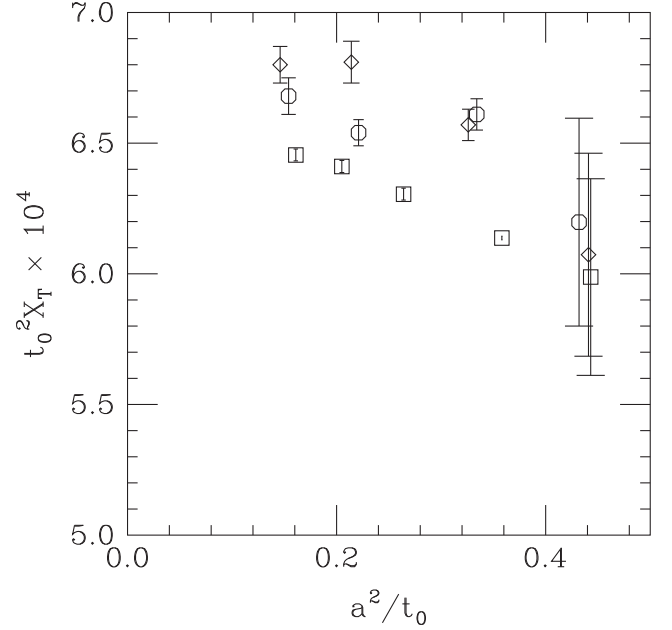


FIG. 11. Quenched  $t_0^2 \chi_T$  from Refs. [13,27] (with the small error bars) and by me (with the large error bars), versus  $a^2/t_0$ . Data are squares for  $N_c = 3$ , octagons for  $N_c = 4$  and diamonds for  $N_c = 5$ .

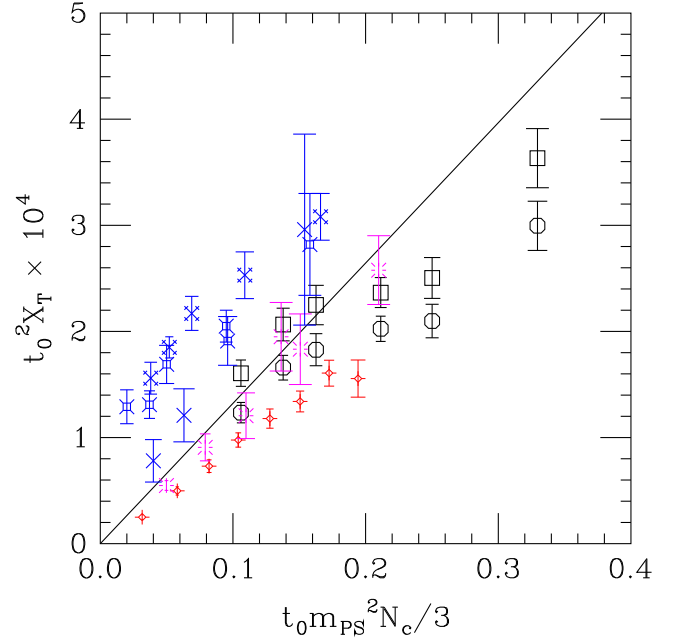


FIG. 12. Comparison of  $N_c = 3$  results for  $t_0^2 \chi_T$  versus  $t_0 m_{PS}^2$ . My results are black squares for  $\chi_T(t = 3)$  and black octagons for  $\chi_T(t = 1.2)$ , while the blue points are data from [30]: fancy crosses, squares, and crosses are data at lattice spacing  $a = 0.075$  fm, 0.065 fm and 0.048 fm, respectively. Red fancy diamonds are from Ref. [36]. Purple bursts are data from Ref. [31]. The line is  $t_0^2 \chi_T = (t_0 f_{PS}^2/4) m_{PS}^2$  with  $\sqrt{t_0} = 0.154$  fm and  $f_{PS} = 93$  MeV.

TABLE IV. Results of fits to Eq. (21).

$N_c$	$1/(t_0^2\chi_Q) \times 10^4$	$C \times 10^4$	$\chi^2/\text{DoF}$
3	0.165(9)	0.048(3)	7.0/6
4	0.146(9)	0.045(5)	9.4/4
5	0.165(10)	0.051(10)	7.5/4
3, 4	0.155(7)	0.048(3)	19.8/11
3, 4, 5	0.158(5)	0.048(3)	29.1/18

points might be light enough. (Note that  $m_{PS}L = 3.26$  and  $3.74$  for the two lightest points; smaller  $m_{PS}$  would require bigger simulation volumes than I used, to avoid finite volume contamination.) This is to be contrasted with the other  $SU(3)$  simulations, where  $t_0m_{PS}^2$  is generally lower than  $0.16$ – $0.19$ . My three lowest points lie in the range  $0.10$ – $0.16$ . Fits to Eq. (20) with more than three points produce  $b$  values which are a factor of two smaller than the expected result, but keeping the lowest three points produces  $(A, b) = 0.36(49) \times 10^{-4}, 12.0(38) \times 10^{-4}$  for the  $t = 3$  susceptibility and  $(A, b) = -0.18(57) \times 10^{-4}, 13.4(45) \times 10^{-4}$  for the  $t = 1.2$  susceptibility. The two choices should have different lattice artifacts, but the important term ( $b$ ) does not seem to be a ridiculous value, nor does it seem to be too dependent on the choice of  $t$  for  $\chi_T$ .

### C. My results across $N_c$

Results across  $N_c$  were displayed in Figs. 1 and 2. Do the data lie on a common curve? I test that by performing a simple linear fit

$$\frac{1}{t_0^2\chi} = \frac{1}{t_0^2\chi_Q} + C \frac{1}{t_0(N_c/3)m_{PS}^2} \quad (21)$$

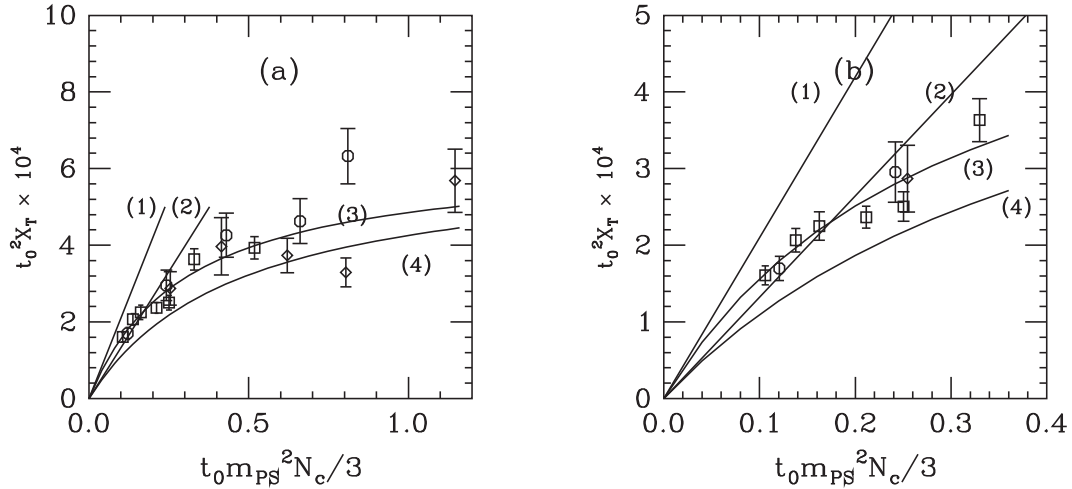


FIG. 13.  $t_0^2\chi_T$  versus  $t_0m_{PS}^2N_c/3$  with several lines. Line (1) is just linear dependence with the slope from the fit to Eq. (21). Line (2) is linear dependence with physical values for  $t_0$  and  $f_{PS}$ . Line (3) is the entire fit function of Eq. (21). Line (4) is the fit function but with physical  $C$ . Panel (b) blows up the small mass region from panel (a).

to individual  $N_c$  values and to various combinations of  $N_c$ . I use my quenched data as inputs to fix the intercept (at  $1/(t_0(N_c/3)m_{PS}^2) = 0$ ). Fit results and the chi-squared per degree of freedom are shown in Table IV. The  $N_c = 3$  and 4 data sets are clearly consistent, and the  $N_c = 5$  topological susceptibility falls on the same curve, although the uncertainty in the slope  $C$  is clearly much greater.

Figure 13 replaces the straight-line presentation with a conventional one of  $t_0^2\chi_T$  versus  $t_0m_{PS}^2N_c/3$ . There are four lines: Line (1) is just linear dependence with the slope from the fit to Eq. (21). Line (2) is linear dependence ( $C = 4/(t_0f_{PS}^2)$ ) with physical ( $SU(3)$ ) values for  $t_0$  and  $f_{PS}$ . Line (3) is the entire fit function of Eq. (21). Line (4) is the fit function but with physical  $C$ . Panel (b) blows up the small mass region of panel (a).

Finally, Fig. 14 shows a third view of curve collapse,  $t_0\chi_T/(m_{PS}^2N_c/3)$  versus  $t_0m_{PS}^2N_c/3$ . This one is a bit dangerous, since  $\chi_T$  from flow does not extrapolate to zero at zero fermion mass: the parameterization blows up there. Overlaid on the data is the expectation of Eq. (4) with physical ( $SU(3)$ ) values for  $t_0$  and  $f_{PS}$ , and  $t_0^2\chi_Q$  taken to be a nominal  $6.25 \times 10^{-4}$ .

What conclusions can be drawn from these figures? First, it is clear that  $\chi_T$  is, broadly speaking, a function of the combination  $m_{PS}^2N_c$ . Second, it is also clear that Eq. (4) with physical ( $SU(3)$ ) values for  $t_0$ ,  $f_{PS}$  and  $t_0^2\chi_Q$  taken from high precision lattice data does not reproduce the data. At this point there are two obvious things to say.

First, this difference could just be due to discretization artifacts at the lattice spacing where the simulations were carried out. A real check requires several lattice spacings and an extrapolation.

Second, the formula Eq. (4) itself could have issues. It is a combination of lowest order chiral perturbation theory

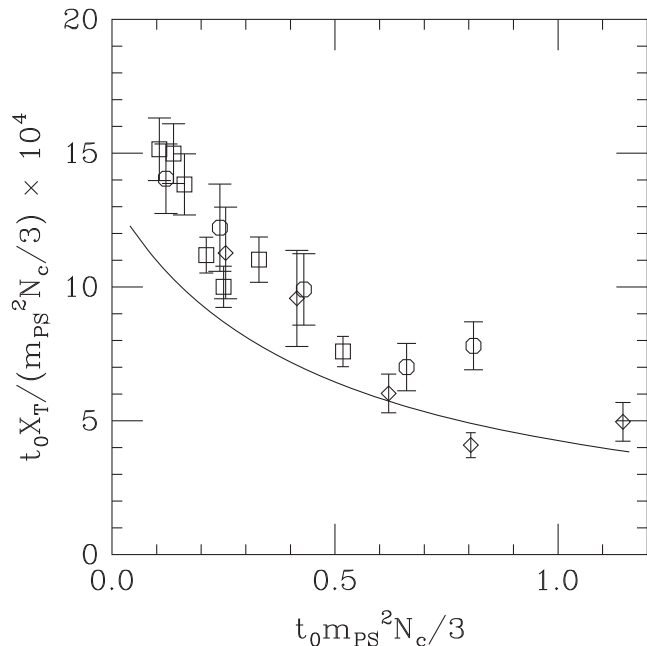


FIG. 14.  $t_0 \chi_T / (m_{PS}^2 N_c / 3)$  versus  $t_0 m_{PS}^2 N_c / 3$  with a line showing the expectation of Eq. (4).

combined with a plausible assumption, that the eta-prime correlator is a bubble sum. Scaling with  $m_{PS}^2 N_c$  is actually scaling with respect to  $m_{PS} / \mu_0^2$  where  $\mu_0$  is the eta-prime mass, combined with scaling of  $\mu_0^2 \propto 1/N_c$  as expected from the Witten-Veneziano relation [39,40]. QCD at intermediate to large sea quark mass does not have to be described by chiral perturbation theory.

#### IV. CONCLUSIONS

This pilot study shows that fermions influence the topological susceptibility through the product  $N_c m_{PS}^2$ . Perhaps it is not a surprising result, but it does illustrate that there are quantities whose  $N_c$  and fermion mass dependence is nonfactorizing.

The other nonfactorizing dependence ( $\propto m_{PS}^2 / N_c$ ) may be more ubiquitous. It appears in all chiral logarithm corrections. High quality data for the topological susceptibility would most likely observe it in the one loop [41] (and beyond) corrections to  $\chi_T$  in the chiral limit. Probably the easiest place to see this generic behavior is in the dependence of  $t_0$  on the pseudoscalar mass, as shown in Fig. 3.

Scaling as  $N_c m_q$  is expected for observables in the epsilon regime (the limit of simulation volume  $V = L^4$  and pseudoscalar mass where  $m_{PS} L \ll 1$  while  $m_H L > 1$  for all other mass scales  $m_H$ ). It appears in predictions for chiral observables such as the finite-volume condensate which involve the scaling combination  $m_q \Sigma V$  (for example,  $\Sigma(V) = m_q \Sigma V f(m_q \Sigma V)$ ). I do not know of any Monte Carlo checks of this scaling.

This is a pilot study: what would it take to produce higher quality data? This presumably means larger volumes, several lattice spacings, and maybe larger  $N_c$ . Larger volumes are needed to push to smaller fermion mass and check for  $m_{PS}^2 N_c$  scaling in a theoretically clean regime. Several lattice spacings are needed, of course, to give a continuum result. Such data sets already exist for  $N_c = 3$ , and the only reason to repeat them is to use them as checks of the methodology for the more interesting larger  $N_c$  cases.

I suspect that such  $N_c = 4$  data sets could be generated with the same techniques as either I or (better) Ref. [30] used, simply consuming more resources. (Neglecting auto-correlation effects, the simulations are dominated by calculation of fermion propagators, involving matrix-times-vector operations; the scaling is roughly  $N_c^2$ .) My experiences with  $N_c = 5$  raise a flag, however. The long autocorrelation time for  $N_c = 5$  compared to lower  $N_c$  values is a clear issue. It is hard to imagine the  $N_c > 5$  will have a shorter autocorrelation time. Of course, I should not say more: I have not tried to do extensive running for  $N_c > 5$  with dynamical fermions at the same lattice spacing as the data presented here. But if I were to keep going with this project, I think I would adopt the open boundary conditions used by Ref. [30] to try to shorten the autocorrelation time.

Another ‘‘pilot area’’ would be to move away from  $N_f = 2$ . For  $N_c = 3$ , this is reasonably well explored by simulations with up, down, and strange quarks, and a recent study by Nogradi and Szikszai [42] covers  $N_f = 2-6$ . These are all tests at low quark mass: what happens as the mass grows? Varying  $N_c$  and  $N_f$  together would allow tests of the Veneziano limit [43,44],  $N_c \rightarrow \infty$  at fixed  $N_f / N_c$ . Is there a universal curve for  $\chi_T(m_{PS}^2)$  across a wide range of  $N_c$  and  $N_f$ , with a scaling variable just  $m_{PS}^2 N_c / N_f$ ?

#### ACKNOWLEDGMENTS

I am grateful to the authors of Ref. [30] for providing me with tables of their data. Some computations were performed on the University of Colorado cluster. I would also like to thank Anna Hasenfratz and Oliver Witzel for comments on the manuscript. My computer code is based on the publicly available package of the MILC collaboration [45]. The version I use was originally developed by Y. Shamir and B. Svetitsky. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award No. DE-SC-0010005. Some of the computations for this work were also carried out with resources provided by the USQCD Collaboration, which is funded by the Office of Science of the U.S. Department of Energy using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359.



## APPENDIX: REVIEW OF THE DERIVATION OF EQ. (2)

Just for completeness, I give a quick derivation of Eq. (2). Since this paper is a lattice calculation, I will assume that we have overlap fermions: their zero modes are chiral and their nonzero ones are not. With nonchiral lattice fermions, this result will be modified by lattice artifacts, but let us neglect them for now. No claim for the originality of this derivation is implied.

Consider the propagator for a bound state of a single flavor of quark, the correlation function of two local pseudoscalar densities,  $\bar{\psi}\gamma_5\psi$ . The first ingredient of the derivation is a plausible assumption for this amplitude, a bubble sum, shown in Fig. 15. The hairpin diagram, the second term in the sum, is

$$H(x, y) = \langle \text{Tr}\gamma_5 \hat{D}(x, x)^{-1} \text{Tr}\gamma_5 \hat{D}(y, y)^{-1} \rangle \quad (\text{A1})$$

where  $\hat{D}(x, y)^{-1}$  is the fermion propagator, the inverse of the Dirac operator. Because only zero modes of the overlap Dirac operator are chiral, the volume integral of the hairpin graph is proportional to the zero mode susceptibility

$$\frac{1}{V} \sum_{x, y} H(x, y) = \frac{\langle Q^2 \rangle}{Vm_q^2} = \frac{\chi}{m_q^2}, \quad (\text{A2})$$

where  $Q$  is just the difference of positive and negative chirality zero modes,  $Q = n_+ - n_-$ .

In quenched QCD, as described by quenched chiral perturbation theory, there is an anomalous coupling of two Goldstone bosons in the flavor singlet channel, parametrized by a coupling with the dimensions of a squared mass. The hairpin graph is analyzed as if each of its quark loops is a propagator for an ordinary pseudoscalar Goldstone meson. That is, the momentum space amplitude for the connected graph (the first term in Fig. 15) is

$$C(q) = f_P \frac{1}{q^2 + m_{PS}^2} f_P \quad (\text{A3})$$

while the hairpin amplitude involving a single flavor is

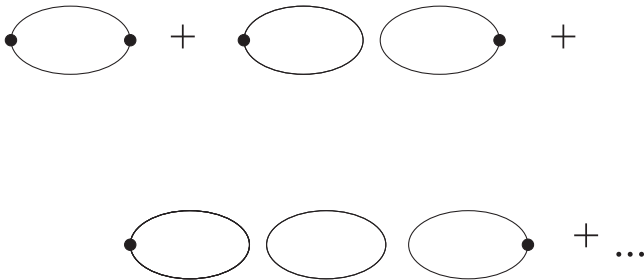


FIG. 15. A set of quark line graphs for the eta-prime meson. The first two terms in the series are the “connected” and “hairpin” graphs.

$$H(q) = f_P \frac{1}{q^2 + m_{PS}^2} \frac{\mu_0^2}{N_f} \frac{1}{q^2 + m_{PS}^2} f_P. \quad (\text{A4})$$

In these expressions,  $f_P = \langle 0 | \bar{\psi}\gamma_5\psi | PS \rangle = \sqrt{2}m_{PS}^2 f_{PS} / (2m_q)$  from the PCAC relation. (Here  $f_{PS} = 93$  MeV.) The quantity  $\mu_0^2$  which couples the fermion loops is the squared mass of the “quenched approximation eta-prime” in the chiral limit. (The factor  $1/N_f$  converts the single-flavor graph into the expectation of the eta-prime mass in  $N_f$ -flavor QCD, since each closed loop has a multiplicity of  $N_f$ , and the wave function (vertex) is scaled by a factor of  $1/\sqrt{N_f}$ .) In full QCD the correlator which gives the mass of the isosinglet meson is the difference  $C(t) - N_f H_{\text{full}}(t)$ , and  $H(t)$  is supposed to represent the first term in a geometric series, the rest of the terms in Fig. 15. This series sums up to

$$\begin{aligned} C(q) - N_f H_{\text{full}}(q) &= C(q) - N_f H(q) + \dots \\ &= f_P \frac{1}{q^2 + m_{PS}^2 + \mu_0^2} f_P, \end{aligned} \quad (\text{A5})$$

shifting the squared mass of the pseudoscalar meson from  $m_{PS}^2$  to  $m_{PS}^2 + \mu_0^2$ .

Computing the quenched susceptibility directly from Eq. (A4) gives

$$\frac{1}{V} \sum_{x, y} H(x, y) = \frac{f_P^2 \mu_0^2}{m_{PS}^4 N_f} = \frac{\mu_0^2 f_{PS}^2}{2N_f m_q^2}. \quad (\text{A6})$$

Equating Eqs. (A2) and (A6), we obtain the Witten-Veneziano [39,40] relation  $\mu_0^2 = 2N_f \chi / f_{PS}^2$ , where  $\chi_Q$  is the quenched zero mode susceptibility.

In full QCD, with dynamical fermions, Eq. (A2) gives the quenched topological susceptibility  $\chi_Q$ . In full QCD, the hairpin is still saturated by zero modes, but Eq. (A5) (evaluated at  $q^2 = 0$ ) says

$$\frac{\chi}{m_q^2} = \frac{f_P^2}{N_f} \left( \frac{1}{m_{PS}^2} - \frac{1}{m_\eta^2} \right). \quad (\text{A7})$$

Substituting for the condensate via  $m_{PS}^2 f_{PS}^2 = 2m_q \Sigma$ , recalling  $m_\eta^2 = \mu_0^2 + m_{PS}^2$ , and using the Witten-Veneziano relation to replace  $\mu_0^2$  by  $\chi_Q$ , we find

$$\chi = \frac{m_q \Sigma}{N_f} \left[ \frac{\chi_Q}{\chi_Q + m_q \Sigma / N_f} \right] \quad (\text{A8})$$

or

$$\frac{1}{\chi} = \frac{N_f}{m_q \Sigma} + \frac{1}{\chi_Q}. \quad (\text{A9})$$

This interpolates between the small- $m_q$  suppression and the quenched result.



- [1] G. 't Hooft, *Nucl. Phys.* **B72**, 461 (1974).
- [2] G. 't Hooft, *Nucl. Phys.* **B75**, 461 (1974).
- [3] B. Lucini and M. Panero, *Phys. Rep.* **526**, 93 (2013).
- [4] T. DeGrand, *Phys. Rev. D* **86**, 034508 (2012).
- [5] G. S. Bali, F. Bursa, L. Castagnini, S. Collins, L. Del Debbio, B. Lucini, and M. Panero, *J. High Energy Phys.* **06** (2013) 071.
- [6] T. DeGrand and Y. Liu, *Phys. Rev. D* **94**, 034506 (2016); **95**, 019902(E) (2017).
- [7] M. G. Perez, *arXiv:2001.10859*.
- [8] T. DeGrand, *Phys. Rev. D* **95**, 114512 (2017).
- [9] O. Bar and M. Golterman, *Phys. Rev. D* **89**, 034505 (2014); **89**, 099905(E) (2014).
- [10] P. Di Vecchia and G. Veneziano, *Nucl. Phys.* **B171**, 253 (1980).
- [11] H. Leutwyler and A. V. Smilga, *Phys. Rev. D* **46**, 5607 (1992).
- [12] R. J. Crewther, *Phys. Lett.* **70B**, 349 (1977).
- [13] M. Cé, M. Garcia Vera, L. Giusti, and S. Schaefer, *Phys. Lett. B* **762**, 232 (2016).
- [14] S. Durr, *Nucl. Phys.* **B611**, 281 (2001).
- [15] A. Hasenfratz and F. Knechtli, *Phys. Rev. D* **64**, 034504 (2001).
- [16] A. Hasenfratz, R. Hoffmann, and S. Schaefer, *J. High Energy Phys.* **05** (2007) 029.
- [17] T. DeGrand, Y. Shamir, and B. Svetitsky, *Phys. Rev. D* **85**, 074506 (2012).
- [18] S. Duane and J. B. Kogut, *Nucl. Phys.* **B275**, 398 (1986).
- [19] S. Duane and J. B. Kogut, *Phys. Rev. Lett.* **55**, 2774 (1985).
- [20] S. A. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, *Phys. Rev. D* **35**, 2531 (1987).
- [21] T. Takaishi and P. de Forcrand, *Phys. Rev. E* **73**, 036706 (2006).
- [22] C. Urbach, K. Jansen, A. Shindler, and U. Wenger, *Comput. Phys. Commun.* **174**, 87 (2006).
- [23] M. Hasenbusch, *Phys. Lett. B* **519**, 177 (2001).
- [24] R. Narayanan and H. Neuberger, *J. High Energy Phys.* **03** (2006) 064; M. Luscher, *Commun. Math. Phys.* **293**, 899 (2010).
- [25] M. Lüscher, *J. High Energy Phys.* **08** (2010) 071; **03** (2014) 92.
- [26] Numerical recipes in *C: The Art of Scientific Computing*, edited by W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2nd ed. (Cambridge University Press, Cambridge, England, 1995).
- [27] M. Cé, C. Consonni, G. P. Engel, and L. Giusti, *Phys. Rev. D* **92**, 074502 (2015).
- [28] F. R. Brown and T. J. Woch, *Phys. Rev. Lett.* **58**, 2394 (1987).
- [29] N. Cabibbo and E. Marinari, *Phys. Lett.* **119B**, 387 (1982).
- [30] M. Bruno, S. Schaefer, and R. Sommer (ALPHA Collaboration), *J. High Energy Phys.* **08** (2014) 150.
- [31] S. Aoki *et al.* (JLQCD and TWQCD Collaborations), *Phys. Lett. B* **665**, 294 (2008).
- [32] R. Sommer, *Nucl. Phys.* **B411**, 839 (1994).
- [33] R. Sommer, *Proc. Sci.*, LATTICE2013 (**2014**) 015 [*arXiv:1401.3270*].
- [34] S. Lottini (ALPHA Collaboration), *Proc. Sci.*, LATTICE2013 (**2014**) 315 [*arXiv:1311.3081*].
- [35] Mattia Bruno and Rainer Sommer (ALPHA Collaboration), *Proc. Sci.*, LATTICE2013 (**2014**) 321 [*arXiv:1311.5585*].
- [36] T. W. Chiu, T.-H. Hsieh, and Y.-Y. Mao (TWQCD Collaboration), *Phys. Lett. B* **702**, 131 (2011).
- [37] T. W. Chiu, T.-H. Hsieh, and Y.-Y. Mao (TWQCD Collaboration), *Phys. Lett. B* **717**, 420 (2012).
- [38] G. Münster and R. Wulkenhaar, *Mod. Phys. Lett. A* **35**, 1950346 (2020).
- [39] E. Witten, *Nucl. Phys.* **B156**, 269 (1979).
- [40] G. Veneziano, *Nucl. Phys.* **B159**, 213 (1979).
- [41] Y. Y. Mao and T.-W. Chiu (TWQCD Collaboration), *Phys. Rev. D* **80**, 034502 (2009).
- [42] D. Negradi and L. Szikszai, *J. High Energy Phys.* **05** (2019) 197.
- [43] G. Veneziano, *Phys. Lett.* **52B**, 220 (1974).
- [44] G. Veneziano, *Nucl. Phys.* **B117**, 519 (1976).
- [45] [https://github.com/milc-qcd/milc\\_qcd/](https://github.com/milc-qcd/milc_qcd/).