# Light scalar meson and decay constant in SU(3) gauge theory with eight dynamical flavors

R. C. Brower, E. Owen, and C. Rebbi
*Department of Physics and Center for Computational Science, Boston University, Boston, Massachusetts 02215, USA*

C. Culver and D. Schaich
*Department of Mathematical Sciences, University of Liverpool, Liverpool L69 7ZL, United Kingdom*

K. K. Cushman
*Department of Physics, Sloane Laboratory, Yale University, New Haven, Connecticut 06520, USA*

G. T. Fleming*
*Department of Physics, Sloane Laboratory, Yale University, New Haven, Connecticut 06520, USA and Theoretical Physics Division, Fermilab, Batavia, Illinois 60510, USA*

A. Gasbarro
*AEC Institute for Theoretical Physics, University of Bern, 3012 Bern, Germany*

A. Hasenfratz and E. T. Neil
*Department of Physics, University of Colorado, Boulder, Colorado 80309, USA*

J. Ingoldby
*Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151, Trieste, Italy*

X. Y. Jin and J. C. Osborn
*Computational Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*

E. Rinaldi
*Interdisciplinary Theoretical and Mathematical Sciences Program (iTHEMS), RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan*

P. Vranas
*Physical and Life Sciences Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA and Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

E. Weinberg
*Department of Physics and Center for Computational Science, Boston University, Boston, Massachusetts 02215, USA and NVIDIA Corporation, Santa Clara, California 95050, USA*

O. Witzel
*Center for Particle Physics Siegen (CPPS), Theoretische Physik 1, Naturwissenschaftlich-Technische Fakultät, Universität Siegen, 57068 Siegen, Germany*

[Lattice Strong Dynamics (LSD) Collaboration]

(Received 30 September 2023; accepted 7 March 2024; published 4 September 2024)

*gfleming@fnal.gov

The SU(3) gauge theory with $N_f = 8$ nearly massless Dirac fermions has long been of theoretical and phenomenological interest due to the near-conformality arising from its proximity to the conformal window. One particularly interesting feature is the emergence of a relatively light, stable flavor-singlet scalar meson $\sigma(J^{PC} = 0^{++})$ in contrast to the $N_f = 2$ theory QCD. In this work, we study the finite-volume dependence of the $\sigma$ meson correlation function computed in lattice gauge theory and determine the $\sigma$ meson mass and decay constant extrapolated to the infinite-volume limit. We also determine the infinite-volume mass and decay constant of the flavor-nonsinglet scalar meson $a_0$.

DOI: 10.1103/PhysRevD.110.054501

## I. INTRODUCTION

SU(3) gauge theory with $N_f$ flavors of massless Dirac fermions has a conformal window for $N_{fc} \leq N_f \leq 16$ [1,2]. See [3] for a review of the early history of constraining $N_{fc}$ and see [4] for the most recent review. While it is not known with certainty whether the massless $N_f = 8$ theory is inside or outside the conformal window [5–7], our collaboration has previously published results [8,9] indicating the massless $N_f = 10$ theory is likely inside the conformal window, and for the rest of this paper we will assume the massless $N_f = 8$ theory is very close to the edge of the conformal window. If it is inside the conformal window, it is most likely a very strongly coupled conformal field theory (CFT) [10]. If it is outside the conformal window, spontaneous chiral symmetry breaking and confinement produce massless Nambu-Goldstone bosons and a spectrum of other hadronic states which may be different relative to QCD due to the proximity of the conformal window.

The continuum SU(3) gauge theory with $N_f = 8$ Dirac fermions with small vectorlike mass terms is not an IR conformal theory. The small mass terms explicitly break chiral symmetry, confinement occurs, and a massive spectrum of hadronic states is generated. Another scenario may be possible at stronger lattice coupling, but we do not consider that here [10].

In our previous papers [11–13], we identified two specific features of the low-energy spectrum which were different from QCD. First, the pion decay constant $F_\pi$ strongly depends on the fermion mass unlike QCD, where $F_\pi$ is nearly constant with a small, linear correction in the fermion mass. Second, the flavor-singlet scalar meson $\sigma(J^{PC} = 0^{++})$ has a light mass $M_\sigma < 1.5 M_\pi$ in the fermion mass region where we compute it, well below the energy threshold for decay to two pions, whereas in QCD it is somewhat heavier $M_\sigma > 1.9 M_\pi$ [14–22], remaining just below decay threshold across in an equivalent fermion mass range. The QCD picture is somewhat consistent with our earlier $N_f = 4$ calculation [12]. However, we also identified several features of the $N_f = 8$ theory which appeared similar to QCD calculations in an equivalent range of fermion masses: The ratios $M_\rho/F_\pi$ and $M_{\text{nucleon}}/F_\pi$ and the decay constants $F_\pi$, $F_\rho$, and $F_{a_1}$ appear consistent with QCD Kawarabayashi-Suzuki-Riazuddin-Fayyazuddin relations [23,24], and the $I = 2$ $\pi\pi$ scattering length $a_{\pi\pi}$ appears to agree with QCD.

In this paper, we focus on larger volume calculations at the various fermion masses (Appendix A) which will allow us to extrapolate our results to the infinite-volume limit, removing one potential source of systematic error (Sec. II E). We also introduce an improved method for analyzing $\sigma$ meson correlation functions with a new subtraction scheme in the rest frame combined with simultaneous analysis in several moving frames (Secs. II B and II C). We rely heavily on the method of Bayesian model averaging [25] (Sec. III). We present new calculations of the flavor-singlet scalar decay constant $F_S$ and the flavor-nonsinglet scalar meson $a_0$ mass and decay constant $F_{a_0}$ (Sec. II D). We also comment briefly on the chiral condensate and its contribution to the Gell-Mann-Oakes-Renner (GMOR) relation and its generalizations (Sec. II F).

As described in our earlier paper [12], we have chosen the bare lattice parameter $\beta = 4.8$ such that the lattice spacing $a$ is as coarse as possible given our current action, so that we can get as close to the chiral limit $am_q \to 0$ as possible with available computing resources. We are working on calculations at $am_q = 0.00056$ which may provide further insights in the near future. If the massless $N_f = 8$ theory is conformal and sufficiently strongly coupled [10], then it is likely a new lattice action that allows for even coarser lattice spacings will be necessary to make further progress.

Phenomenologically, theories that exhibit approximate conformal behavior at strong coupling are anticipated to produce large anomalous dimensions over a wide interval of scales, which can make them attractive as candidate composite Higgs models [26–30]. In particular, the $SU(3) N_f = 8$ theory has been used to build composite Higgs models in [31–33]. The construction of a low-energy effective field theory (EFT) for the lightest composites, to which the rest of the Standard Model can be coupled, is a crucial intermediate step in the creation of these models. In a separate paper [34], we discuss various effective models that can be fit to our data.

## II. STAGGERED TWO-POINT CORRELATION FUNCTION CONSTRUCTION AND MODELING

### A. Staggered two-point correlation function construction

The continuum $N_f = 8$ theory is approximated on a finite lattice by an SU(2) doublet of staggered fermion fields $(\chi_1 \chi_2)$ that carry only an SU(3) color index at each lattice site. Each component of the doublet represents four nondegenerate Dirac fermion "tastes" with spin and taste degrees of freedom spread out over $2^4$ sites of local hypercubes. In the continuum limit where the bare gauge coupling $g_0^2 \to 0$, these tastes become degenerate and equivalent to four Dirac flavors. Hence, the doublet of staggered fields becomes a degenerate $N_f = 8$ theory in this continuum limit. Staggered fermions have a remnant of chiral symmetry that can lead to $(N_f/4)^2$ Nambu-Goldstone bosons when taking the massless chiral limit at finite lattice spacing, assuming the chiral symmetry is spontaneously broken by the gauge interactions. However, to recover the full flavor symmetry, it is essential to take the $g_0^2 \to 0$ continuum limit prior to the $m_q \to 0$ chiral limit.

In general, a staggered meson two-point correlation function where source and sink operators have the same quantum numbers $Q$ is (schematically)

$$C_Q(\vec{p}, |t - t_0|) = \left\langle \sum_{\vec{x}} e^{i\vec{p}\cdot(\vec{x}-\vec{x}_0)} \bar{\chi}(\vec{x}+\vec{\delta}', t)\Gamma_Q(\vec{x}, \vec{\delta}')\tau\chi(\vec{x}, t) \times \bar{\chi}(\vec{x}_0+\vec{\delta}, t_0)\Gamma_Q(\vec{x}_0, \vec{\delta})\tau\chi(\vec{x}, t) \right\rangle, \quad (1)$$

where $\Gamma_Q(\vec{x}, \vec{\delta})$ are phases which refer to the spin-taste structure of the interpolating operators with quantum numbers $Q$ and $\tau$ is either an SU(2) generator for a nonsinglet correlator or a $2 \times 2$ identity matrix for a singlet correlator under the SU(2) staggered doublet symmetry. There are various phase conventions possible; one common choice is [35]. Not shown are gauge matrices required to make the whole thing gauge invariant, e.g., connecting sites $(\vec{x}, t)$ and $(\vec{x}+\vec{\delta}', t)$. Also, translation invariance of the ensemble average $\langle \cdot \rangle$ guarantees the correlation function depends on only the distance $|t - t_0|$ and not the source position $(\vec{x}_0, t_0)$.

In our earlier paper [12], the LSD Collaboration constructed correlation functions with $\vec{p} = 0$ for local and point-split operators. In this study, we focused on constructing singlet and nonsinglet correlation functions of local operators at several different momenta $\vec{p}$ with much higher statistics. On each gauge configuration, we generate a unique set of $N$ random source points $(\vec{x}_0, t_0)_n$ and construct a primitive staggered meson "connected" correlator

$$C(\vec{x}, t) = \frac{1}{N} \sum_n \mathcal{T}_n \mathrm{Tr}_{\mathrm{color}} \left[ G_F(\vec{x}_0, t_0; \vec{x}, t) G_F^\dagger(\vec{x}, t; \vec{x}_0, t_0) \right], \quad (2)$$

where $G_F(\vec{x}_0, t_0; \vec{x}, t)$ is a $3 \times 3$ color matrix of single staggered fermion propagator from the site $(\vec{x}_0, t_0)$ to the site $(\vec{x}, t)$ and $\mathcal{T}_n$ represents the translation of the $n$th source location $(\vec{x}_0, t_0)_n$ to the origin $(\vec{0}, 0)$. Then, we record the value of this averaged primitive correlator for every sink point $(\vec{x}, t)$ in the lattice volume. We refer to this as a connected correlator, because valence fermion lines connect the source and sink points. With postprocessing, we can project this primitive correlator into eight different nonsinglet staggered meson quantum number channels of different momenta $\vec{p}$ using Fourier transform

$$C_Q(\vec{p}, t) = \sum_{\vec{x}} e^{i\vec{p}\cdot\vec{x}} C(\vec{x}, t)\phi_Q(\vec{x}). \quad (3)$$

For example, if we choose the phase $\phi_Q(\vec{x}) = 1$, we get the correlation function for the $\pi_5$ meson, which is the pseudo-Nambu-Goldstone boson.

Also in our earlier work, we explained in detail how we construct a "disconnected" correlator, where valence fermion lines do not connect the source and sink points:

$$D(\vec{p}, |t - t_0|) = \sum_{\vec{x}} \sum_{\vec{x}_0} e^{i(\vec{x}-\vec{x}_0)\cdot\vec{p}} \mathrm{Tr}\left[ G_F(\vec{x}_0, t_0; \vec{x}_0, t_0) \right] \times \mathrm{Tr}\left[ G_F(\vec{x}, t; \vec{x}, t) \right] \quad (4)$$

using a diluted noisy estimator to compute the trace at each site on the lattice for each gauge configuration, which is again recorded as a single value per site in the lattice volume. With postprocessing, we can compute the disconnected correlator for any spatial momentum $\vec{p}$ using FFT and the fast convolution algorithm:

$$\tilde{O}(\vec{p}, \omega) = \sum_{\vec{x}, t} e^{i(\vec{p}\cdot\vec{x}+\omega t)} \mathrm{Tr}\left[ G_F(\vec{x}, t; \vec{x}, t) \right], \quad (5)$$

$$D(\vec{p}, t) = \sum_{\omega} e^{-i\omega t} |\tilde{O}(\vec{p}, \omega)|^2, \quad (6)$$

where the result is automatically invariant under any lattice translation. In an $N_f$-flavor theory, the flavor-singlet scalar correlator for the $\sigma$ meson is then

$$C_\sigma(\vec{p}, t) = \left( \frac{N_f}{4} \right)^2 D(\vec{p}, t) - \frac{N_f}{4} C_{a_{0,1}}(\vec{p}, t) \quad (7)$$

and where $C_{a_{0,1}}(\vec{p}, t)$ is the flavor-nonsinglet scalar meson correlator constructed from Eq. (3) with the appropriate choice of phases. Note this normalization is different from [12], where we dropped an overall factor of $N_f/4$.

Regarding the naming convention of mesons, we note that the continuum SU(8) flavor representation is broken by lattice artifacts to a subgroup SU(2)×taste, a discrete subgroup of SU(4). Meson names will follow the PDG convention for two-flavor mesons: $\pi, a_0, \rho, \ldots$ plus an additional subscript to indicate the representation under the discrete taste group: $\pi_5, a_{0,1}, \rho_i, \ldots$. There is only one scalar meson which is a singlet over the whole flavor group which we name $\sigma$ and no subscript is required. The effects of taste breaking were discussed previously [12] and we will not expand on it, so the taste index does not play a significant role here with one important exception. In the continuum two-flavor theory, the decay $a_0 \to \pi\pi$ is forbidden by isospin symmetry. However, in our staggered $N_f = 8$ theory, the decay $a_{0,1} \to \pi_5\pi_5$ is allowed because the $\pi_5$ and the $a_{0,1}$ are not in the same SU(2) flavor subgroup, as indicated by the different taste indices. It is analogous to the decay of $a_0 \to KK$ in continuum three-flavor theory.

### B. Model for staggered meson correlation functions

We will consider three different types of models for staggered meson two-point correlation functions in this paper. As we are employing Bayesian model averaging, further discussed in Sec. III, we do not have to choose a particular model but rely on the computed model probabilities to distinguish the most likely models for a given correlation function. Within each model type, the number of free parameters in each specific instance of the model will depend upon the number of oscillating and nonoscillating states included.

The model we will use for the staggered meson correlation function in the time domain (model A) is

$$\begin{aligned} C(\vec{p}, t) = c_0\delta_{\vec{p},0} &+ \sum_n \frac{c_n}{2\left(1 - e^{-E_nN_t}\right)\sinh(E_n)} \\ &\times \left[e^{-E_nt} + e^{-E_n(N_t-t)}\right] \\ &+ (-1)^t \sum_j \frac{c'_j}{2\left(1 - e^{-E'_jN_t}\right)\sinh(E'_j)} \\ &\times \left[e^{-E'_jt} + e^{-E'_j(N_t-t)}\right], \end{aligned} \tag{8}$$

where we have chosen to use a particular "relativistic" normalization for the amplitudes. As is typical for the staggered fermions, there are a set of states labeled by $n$ whose contributions do not oscillate in time and another set of states labeled by $j$, with different quantum numbers, that oscillate in time with a factor $(-1)^t$. The energies $E_n$ and $E'_j$ are understood to depend implicitly on the spatial momentum $\vec{p}$. We also allow for the possibility of a $t$-independent contribution to the correlation function, $c_0$, which is generally not present for flavor-nonsinglet correlation functions due to translation invariance of the ensemble average. But, it is the dominant contribution to the flavor-singlet $\sigma$ correlation function and must be treated carefully in order to extract reliable estimates of model parameters. Note that the constant contributes only to the $\vec{p} = 0$ correlator, so one method of dealing with this constant is to work with $\vec{p} \neq 0$ correlators. Given that we are interested in the energy of the $\sigma$ meson in the rest frame, $\lim_{\vec{p}\to 0} E_\sigma(\vec{p}) = M_\sigma$, this approach requires a good understanding of the dispersion relation on the lattice.

To motivate the normalization of amplitudes $c_n$ and $c'_j$ in Eq. (8), we can perform the discrete cosine transform (DCT-I) of the time-domain correlation function into the frequency domain analytically:

$$\tilde{C}(\vec{p}, k) = c_0\delta_{\vec{p},0}\delta_{k,0} + \frac{1}{N_t}\sum_n \frac{c_n}{\hat{E}_n^2 + \hat{\omega}_k^2} + \frac{1}{N_t}\sum_j \frac{c'_j}{\hat{E}_j'^2 + \hat{\omega}_k'^2}, \tag{9}$$

where

$$\hat{E}_n = 2\sinh\frac{E_n}{2}, \qquad \hat{\omega}_k = 2\sin\frac{2\pi k}{2N_t}, \\ \hat{\omega}'_k = 2\sin\left(\frac{\pi}{2} - \frac{2\pi k}{2N_t}\right). \tag{10}$$

Comparing the expression for two different spatial momenta $\vec{p}$, the energies $E_n$ and $E'_j$ will be different, defining some lattice dispersion relation. But the amplitudes $c_n$ and $c'_j$ are momentum independent as normalized and, therefore, frame independent as expected in a Lorentz-invariant theory, hence a "relativistic" normalization.

In our previous work, we considered another method of dealing with the constant $c_0$ which was to analyze the finite difference correlation function for the $\vec{p} = 0$ $\sigma$ meson:

$$\Delta_\sigma(t) = C_\sigma(t+1) - C_\sigma(t). \tag{11}$$

In the model, the cancellation of $c_0$ is exact, but in our lattice calculation there is inherent statistical noise contributing to each time slice, so the cancellation is not exact. In this work, we propose an improved subtraction scheme for $\vec{p} = 0$ correlation functions:

$$\bar{C}(t) = C(t) - \frac{1}{N_t}\sum_{t'=0}^{N_t-1} C(t') \tag{12}$$

for states that have a time-independent part, like the $\sigma$ meson. Given our frequency analysis above, we can see the subtraction is the zero-frequency component of the correlation function $\bar{C}(t) = C(t) - \tilde{C}(0)$. Furthermore, we know explicitly the functional form of the residual constant that comes from the integral of the $t$-dependent part of the correlation function:

$$c_0 - \tilde{C}(0) = -\frac{1}{N_t}\sum_{n=1}^{\infty}\frac{c_n}{\hat{M}_n^2} - \frac{1}{N_t}\sum_{j=1}^{\infty}\frac{c'_j}{4+\hat{M}'^2_j}. \quad (13)$$

Because some of the fit parameters appear in the residual constant, we will include that part in the fit and shift the constant (model B):

$$\begin{aligned}\bar{C}(t) &= \bar{c}_0 + \sum_{n}^{n_{\max}}\frac{c_n}{2(1-e^{-M_nN_t})\sinh(M_n)} \\ &\quad\times\left[e^{-M_nt}+e^{-M_n(N_t-t)}\right] - \frac{c_n}{N_t\hat{M}_n^2} \\ &\quad+\sum_{j}^{j_{\max}}\frac{(-1)^t c'_j}{2(1-e^{-M'_jN_t})\sinh(M'_j)} \\ &\quad\times\left[e^{-M'_jt}+e^{-M'_j(N_t-t)}\right] - \frac{c'_j}{N_t(4+\hat{M}'^2_j)}, \\ \bar{c}_0 &= -\frac{1}{N_t}\sum_{n=n_{\max}+1}^{\infty}\frac{c_n}{\hat{M}_n^2} - \frac{1}{N_t}\sum_{j=j_{\max}+1}^{\infty}\frac{c'_j}{4+\hat{M}'^2_j}. \end{aligned} \quad (14)$$

In counting free parameters, model B will have one more free parameter than model A, and the interpretation of the value of this parameter, $\bar{c}_0$, will depend strongly on the choice of $n_{\max}$ and $j_{\max}$. In particular, we expect $\bar{c}_0 \to 0$ within statistical uncertainties as the number of states included in a particular model instance approaches the limit of available statistics to properly constrain them.

We will also consider a modification of model B (model C) where we constrain $\bar{c}_0 = 0$. It will have the same number of free parameters as model A in an instance where they include the same number of states. In the context of Bayesian model averaging, we expect that model B will have a higher relative probability than model C in instances where $\bar{c}_0$ is statistically nonzero. But with increasing numbers of states eventually model C should become more probable, also indicating the limit in which the power of the available statistics to constrain parameters has been exhausted.

### C. Staggered meson dispersion relation

The functional momentum dependence of energies $E_Q(\vec{p})$ extracted from analysis of two-point correlation functions $C_Q(\vec{p}, t)$ is a complicated, nonperturbative problem, because Lorentz symmetry is broken by the lattice discretization, so the theory is not invariant under boosts. Still, Lorentz symmetry is fully recovered in the continuum limit. Naively, we can expect

$$a^2E_Q^2(\vec{p}) = a^2M_Q^2 + a^2p^2 + O(a^4p^4), \quad (15)$$

where we explicitly show the lattice spacing $a$ in this dimensionless relation and define the spatial momentum components $p_i = 2\pi n_i/(aN_s)$ and $n_i \in \{-N_s/2+1, \ldots, 0, \ldots, N_s/2\}$ and $N_s$ is the number of lattice sites in the spatial directions.

To improve upon this estimate, one would have to understand the dynamics on the lattice of the eigenstates corresponding to these energies. This is a challenging problem, since the eigenstates are not simple single-hadron excitations, in general, but are more likely strongly interacting multihadron states. But the lowest-energy state with given quantum numbers $Q$ may reasonably be expected to behave like a single-hadron state, particularly if its energy is well below the nearest multihadron threshold. In this case, we can approximate the dispersion relation with that of a noninteracting boson on the lattice [36]:

$$\hat{E}_Q^2 = \hat{M}_Q^2 + \hat{p}^2 + O(\hat{p}^4), \quad (16)$$

$$\hat{E}_Q = 2\sinh\frac{aE_Q}{2}, \qquad \hat{M}_Q = 2\sinh\frac{aM_Q}{2}, \qquad \hat{p}_i = 2\sin\frac{ap_i}{2}. \quad (17)$$

In the second equation, we have explicitly put in the lattice spacing dependence $a$. Both lattice dispersion relations correspond to the same continuum relation as $a \to 0$.

In either of these models, Eqs. (15) or (17), the finite size of the lattice along spatial directions $N_s$ directly controls the spacing between the discrete momenta but is not expected to appear explicitly in the finite lattice spacing corrections $O(a^4p^4)$ or $O(\hat{p}^4)$. When we fit our lattice data on two or more volumes at the same value of the bare coupling and mass, we will parametrize our fits so that the same lattice corrections are used on all volumes.

### D. Staggered meson decay constants

The normalization in Eq. (8) was chosen such that $c_n \to |\langle 0|\mathcal{O}|n, \vec{p}=0\rangle|^2$ in the continuum limit with the usual continuum relativistic normalization. Following Eq. (7.5) of [37], we define the pion decay constant

$$\sqrt{2}\hat{F}_{\pi_5}(\hat{E}^2_{\pi_5} - \hat{p}^2) = 2m_q\frac{1}{\sqrt{N_f}}\langle 0|P_5|\pi_5(\vec{p})\rangle \Rightarrow \hat{F}_{\pi_5} = \frac{1}{\sqrt{2}}\frac{m_q\sqrt{|c_{\pi_5}|}}{\hat{E}^2_{\pi_5} - \hat{p}^2}, \quad (18)$$

where $N_f$ in this equation is the number of continuum flavors of a single staggered fermion, i.e., $N_f = 4$. Note we put the hat on the symbol for $\hat{F}_{\pi_5}$ to indicate the form of the lattice dispersion relation used. We could have just as easily used the other form of the lattice dispersion relation, which would lead to a slight different definition of the decay constant. Both definitions should converge to the continuum one in the limit of zero lattice spacing. This definition is slightly different than ones previously used by the LSD Collaboration for the pion decay constant [12,13], but the difference is not statistically significant.

For the isotriplet scalar form factor, there does not seem to be a conventional normalization [38,39] for the decay constant in QCD, as it is an unstable resonance. See the review *"Scalar mesons below 1 GeV"* in [40]. In our $N_f = 8$ theory over the range of fermion masses we have studied, the nonsinglet scalar meson appears to be stable, although close in energy to its decay threshold. We choose to normalize it analogously with the pion decay constant

$$\hat{F}_{a_{0,1}} = \frac{1}{\sqrt{2}} \frac{m_q \sqrt{|c_{a_{0,1}}|}}{\hat{E}^2_{a_{0,1}} - \hat{p}^2}, \tag{19}$$

where $c_{a_{0,1}}$ is the residue of the first pole in the frequency domain representation of the nonsinglet scalar two-point correlation function, Eq. (9).

For the isosinglet scalar decay constant, we use the normalization defined in Eq. (72) of [41]:

$$\hat{F}_S(\hat{E}^2_\sigma - \hat{p}^2) = m_q \langle 0|S(0,0)|\sigma(\vec{p})\rangle, \tag{20}$$

where the scalar current is defined as $S(\vec{x},t) = \sum_{i=1}^{N_f/4} \bar{\chi}_i(\vec{x},t)\chi_i(\vec{x},t)$. The two-point correlation function of this scalar current is defined in Eq. (7), and, in terms of this correlation function, the decay constant is defined

$$\hat{F}_S = \frac{m_q \sqrt{|c_\sigma|}}{\hat{E}^2_\sigma - \hat{p}^2}. \tag{21}$$

In particular, the normalization used in Eq. (7) is essential to correctly normalizing the decay constant.

### E. Finite-volume corrections

In QCD, finite-volume corrections to the pion mass and pion decay constant extracted from a two-point correlation function calculated on a periodic torus of spatial size $L$ can be computed in chiral perturbation theory provided $M_\pi L \gg 1$ and $F_\pi L \gg 1$. See Eq. (6.15) of [42], for example. In $N_f = 8$ over the range of fermion masses for which we have relevant lattice calculations, chiral perturbation theory is unlikely to be a good effective description for two reasons: the strong fermion mass dependence of $F_\pi$ and the stable $\sigma$ meson with $M_\sigma \ll 4\pi F_\pi$. So it is not expected that finite-volume corrections computed in chiral perturbation theory (ChiPT) will exactly match the numerical calculations. Still, it seems likely that whatever low-energy effective theory replaces ChiPT will have much the same structure, as these arise from contributions of virtual pion degrees of freedom that probe the finite volume by wrapping the spatial cycles of the torus, and the pion still is the lightest hadron in the eight-flavor theory. There may be additional contributions from $\sigma$-meson degrees of freedom, but they are expected to be subleading due to the somewhat heavier mass.

We will follow the approach used in [13] and use ChiPT-inspired forms to model our finite-volume corrections:

$$M_Q(L) = M_Q(\infty)\left[1 + \alpha_Q \frac{M_\pi^2}{(4\pi F_\pi)^2} \sum_{n=1}^{\infty} \frac{4\kappa(n)}{\sqrt{n} M_\pi L} \times K_1\left(\sqrt{n} M_\pi L\right)\right], \tag{22}$$

$$F_Q(L) = F_Q(\infty)\left[1 + \beta_Q \frac{M_\pi^2}{(4\pi F_\pi)^2} \sum_{n=1}^{\infty} \frac{4\kappa(n)}{\sqrt{n} M_\pi L} \times K_1\left(\sqrt{n} M_\pi L\right)\right]. \tag{23}$$

The function $\kappa(n)$ counts the number of lattice vectors $\vec{n}$ with integer-valued components of length $\sqrt{n}$, see Table I. In QCD, it is common to expand the sum over modified Bessel functions $K_1$, assuming $M_\pi L \gg 1$, and keep only the leading term, particularly if $M_\pi L \gtrsim 4$ in all, leading to

$$\sum_{n=1}^{\infty} \frac{4k(n)}{\sqrt{n} M_\pi L} K_1(\sqrt{n} M_\pi L) \approx \frac{12\sqrt{2\pi}}{(M_\pi L)^{3/2}} e^{-M_\pi L}. \tag{24}$$

In an earlier paper [13], we also used this approximation for the finite-volume extrapolation of $M_\pi$ and $F_\pi$. We did not observe any significant change in the result if we included more terms in the expansion. In this analysis, we will be conservative and not expand the modified Bessel functions and truncate the sum only after the first eight terms (up to $n = 8$), although we expect it will not make a significant difference relative to keeping just the leading term.

TABLE I. The number of lattice vectors $\vec{n}$ with integer-valued components of length $\sqrt{n}$. Note there are no vectors of length $\sqrt{7}$ and, starting at length 3, there may be multiple inequivalent sets of vectors under the cubic group.

| $n$ | $\vec{n}$ | $\lvert\vec{n}\rvert$ | $\kappa(n)$ |
|---|---|---|---|
| 0 | (0, 0, 0) | 0 | 1 |
| 1 | (1, 0, 0) | 1 | 6 |
| 2 | (1, 1, 0) | $\sqrt{2}$ | 12 |
| 3 | (1, 1, 1) | $\sqrt{3}$ | 8 |
| 4 | (2, 0, 0) | 2 | 6 |
| 5 | (2, 1, 0) | $\sqrt{5}$ | 24 |
| 6 | (2, 1, 1) | $\sqrt{6}$ | 24 |
| 7 | ... | ... | 0 |
| 8 | (2, 2, 0) | $\sqrt{8}$ | 12 |
| 9 | (2, 2, 1) | 3 | 24 |
| 9′ | (3, 0, 0) | 3 | 6 |

Since the infinite-volume extrapolation described in this section implicitly assumes that the pion is a pseudo-Nambu-Goldstone boson, one should use caution when modeling the extrapolated data provided later in this paper, particularly if one wants to explore other finite-volume corrections, e.g., due to a light isosinglet scalar. If one assumes that the massless limit of the theory approaches a conformal fixed point, possible finite-volume corrections were discussed in [43]. In either case, one should use the finite-volume data provided in Supplemental Material [44] when performing further analysis.

### F. The GMOR relation and near-conformality

As a guide to constructing low-energy effective descriptions $N_f = 8$ theory, it would be useful to characterize the extent to which one or a few light states dominates the low-energy dynamics. An important phenomenological tool for characterizing the degree to which the dynamics of the Nambu-Goldstone pions dominates low-energy phenomena in QCD was first described by Gell-Mann, Oakes, and Renner (GMOR) [45]. In their original derivation, they *a priori* assumed pion-pole dominance and derived the GMOR relation as a consequence. Our derivation will not initially assume pole dominance but start with the integral of the axial Ward-Takahashi identity. In our notation, this can be written

$$\sum_{t=0}^{N_t-1} C_{\pi_5}(\vec{0},t) = \frac{1}{m_q}\mathrm{Tr}_{\mathrm{color}}\left[G_F(\vec{0},0;\vec{0},0)\right], \quad (25)$$

which is an exact spectral identity on each gauge configuration, not just in the ensemble average. In the chiral limit $m_q \to 0$ of a theory with spontaneous chiral symmetry breaking, the trace on the right-hand side will approach a constant following the Banks-Casher relation [46,47] and the integrated pion correlation function will diverge due to the massless Nambu-Goldstone pion. Using Eqs. (9) and (18), we can identify the rate of this divergence with parameters in our fit functions:

$$\sum_{t=0}^{N_t-1} C_{\pi_5}(\vec{0},t) \to \frac{c_{\pi_5}}{\hat{M}_{\pi_5}^2} = 2\frac{\hat{F}_{\pi_5}^2\hat{M}_{\pi_5}^2}{m_q^2} \quad \text{as } m_q \to 0. \quad (26)$$

Using the normalization of the isosinglet scalar current in Eq. (20) leads to a generalization of the GMOR relation for general $N_f$:

$$m_q\langle S\rangle = m_q\frac{N_f}{4}\langle\bar{\chi}\chi\rangle \geq \frac{N_f}{2}\hat{F}_{\pi_5}^2\hat{M}_{\pi_5}^2. \quad (27)$$

Now, if we assume spontaneous symmetry breaking and pion pole dominance, the inequality becomes an equality in the limit $m_q \to 0$ and $\langle S\rangle$ approaches a well-defined low-energy constant, which is the usual GMOR relation.

Patella [48] has noted that Eq. (27) should also be true in a mass-deformed CFT with a large mass anomalous dimension $(1 < \gamma^* < 2)$ due to a large contribution to the pion correlation function generated by the running of the mass. They propose examining the GMOR ratio

$$R_G(m_q) \equiv \frac{m_q\langle\bar{\chi}\chi\rangle}{2\hat{F}_{\pi_5}^2\hat{M}_{\pi_5}^2} = \begin{cases} 1, & \text{(near-conformal)} \\ 1 < R_G(0) < \infty, & (\text{CFT}, 1 < \gamma^* < 2) \\ \infty, & (\text{CFT}, 0 < \gamma^* < 1) \end{cases} \quad \text{as } m_q \to 0 \quad (28)$$

for an indication of whether the theory is near-conformal or conformal in the chiral limit. In a near-conformal scenario, it is not clear at what fermion mass $m_q$ one would expect to see the transition from the approximately hyperscaling regime where $R_G(m_q) > 1$ to the spontaneously broken regime where $R_G(m_q) \to 1$ as $m_q \to 0$. Just observing $R_G(m_q) > 1$ at some finite fermion mass is not sufficient to establish IR conformality. In particular, one must follow the correct order of limits: volume to infinity, lattice spacing to zero, and then fermion mass to zero.

## III. BAYESIAN MODEL AVERAGING

### A. General setup

One of the challenges observed in our previous analysis of the light meson spectrum in the $N_f = 8$ theory [12] were large systematic errors due to fit parameters varying significantly over a range of different fits while $\chi^2/\text{d.o.f.}$ did not. We define $\log p(D|M)$ by the usual chi-squared prescription

$$\log p(D|M) \propto -\frac{1}{2}\sum_{t,t'\in T_1}\left(C(t)-f_M(t)\right)\Sigma_{tt'}^{-1}\left(C(t')-f_M(t')\right), \quad (29)$$

where $C(t)$ is correlation function computed from our lattice ensemble $D$, $f_M(t)$ is the function for model $M$ to be fitted by minimizing the log-likelihood, $T_1$ is the subset of times selected for fitting, and $\Sigma_{tt'}$ is the covariance of the correlation function $C(t)$ on the subset $T_1$. Assuming all the quantities are properly estimated from the ensemble, the log-likelihood is expected to sample the chi-squared distribution for degrees of freedom equal to the number of times in $T_1$ minus the number of free parameters in $M$.

Subsequent to our earlier analysis, Jay and Neil proposed [25] a Bayesian model-averaging analysis framework which estimates $\log p(M|D)$, the probability that a model $M$ is a good representation of the data selection $D$. One suggested estimator of the model probability is based on the Akaike information criterion (AIC), provided nuisance model parameters are assigned to account for data subsets not included in the fit. For example, let $M$ be a model with $N_M$ free parameters, and the maximal dataset has $N_T$ times available to be fit. If we perform the fit only on a subset of times $T_1$ of size $N_1$, then the number of data points not included $N_0 = N_T - N_1$ must be assigned nuisance parameters. Thus, for the AIC, the number of relevant parameters is $N_M + N_0$, and the model probability [49] is

$$\log p(M|D) \propto \log p(D|M) - (N_M + N_0). \tag{30}$$

After the model probability has been estimated for the full set of models $\{M\}$ to be considered for the analysis, we normalize this set of probabilities: $\sum_{\{M\}} p(M|D) = 1$. In Appendix B, we provide some details how we perform this sum accurately given the potential for widely varying values of $\log p(M|D)$.

With an reasonable estimate of the model probability, it seems straightforward to construct expectation values and variances of model parameters over the set of possible models considered. For example, the expected value of a model parameter is

$$E(a) = \frac{1}{\Sigma_1} \sum_{\{M|a\in M\}} a_M p(M|D)\Theta\big[p(M|D) - p_{\rm cut}\big]$$
$$\Sigma_1 = \sum_{\{M|a\in M\}} p(M|D)\Theta\big[p(M|D) - p_{\rm cut}\big], \tag{31}$$

where, to make sure the notation is clear, we compute a weighted average over only the subset of models that contain the parameter $\{M|a\in M\}$ and further consider only models where the model weight is greater than some predetermined minimum $p_{\rm cut}$, as enforced by the Heaviside function $\Theta$.

The variance of the model-averaged expectation value has two contributions. The first, and usually dominant, contribution is the weighted average over models of the square of the error estimate $\sigma_{a,M}$ for the parameter $a_M$ in a given model $M$:

$$E(\sigma_a^2) = \frac{1}{\Sigma_1} \sum_{\{M|a\in M\}} \sigma_{a,M}^2 p(M|D)\Theta\big[p(M|D) - p_{\rm cut}\big]. \tag{32}$$

The second, usually subdominant, contribution is the weighted variance of the model estimates of parameters $a_M$, relative to the model-averaged expectation $E(a)$:

$$\mathrm{Var}(a) = \frac{\Sigma_1}{\Sigma_1^2 - \Sigma_2} \sum_{\{M|a\in M\}} \big(a_M - E(a)\big)^2 p(M|D) \times \Theta\big[p(M|D) - p_{\rm cut}\big]$$
$$\Sigma_2 = \sum_{\{M|a\in M\}} p(M|D)^2\Theta\big[p(M|D) - p_{\rm cut}\big]. \tag{33}$$

The final error estimate for the model average of a parameter is to add the two contributions in quadrature:

$$\sigma_a = \sqrt{E(\sigma_a^2) + \mathrm{Var}(a)}. \tag{34}$$

Now we can discuss the motivation behind the probability cut $p_{\rm cut}$. In our experience, the model-averaged $E(a)$ tend to be dominated by a few choices whose $p(M|D) \sim O(1)$. It seems reasonable to expect that $E(\sigma_a^2)$ should be similarly dominated by choices whose $p(M|D) \sim O(1)$ and not $p(M|D) \sim O(p_{\rm cut})$. However, we have observed cases of overfitting for certain models where, as the data selection changes such that $p(M|D)$ decreases, $\sigma_{a,M}^2$ increases at a faster rate, leading to those very unlikely model choices to dominate the model average of the squared error $E(\sigma_a^2)$. $p_{\rm cut}$ can be adjusted to minimize the impact of this scenario.

To understand how this can happen, we recall that uncertainty of a two-point meson correlation function grows exponentially in Euclidean time [50]:

$$\mathrm{Var}[C_Q(\vec{p}, t)] \sim \exp\big[2\big(E_Q(\vec{p}) - M_{\pi_5}\big)t\big]. \tag{35}$$

Now, for a given model function $M$ with its fixed number of exponential terms, there is a certain $t_{\rm min}$ for which $-\chi^2/2 \sim (N_M - t_{\rm max} + t_{\rm min} - 1)/2$, indicating a good fit using the usual chi-squared criteria $\chi^2/\mathrm{d.o.f.} \sim 1$. For fits on the interval $[t, t_{\rm max}]$, $t < t_{\rm min}$, there will be no good fits according to chi-squared, whereas for fits on the interval $[t, t_{\rm max}]$, $t > t_{\rm min}$, $-\chi^2/2$ will approximately increase by $(t - t_{\rm min})/2$ indicating continued goodness of fit. However, as the minimum $t$ increases in a given fit, the number of times not included in the fit also increases: $\Delta N_0 = t - t_{\rm min}$. The net effect of increasing $t > t_{\rm min}$ is to decrease $p(M|D) \propto \exp(-(t - t_{\rm min})/2)$. If $E_Q(\vec{p}) - M_{\pi_5} > 1/4$, we expect that the uncertainties in model parameters will grow faster than the model probability decreases as $t > t_{\rm min}$. Based on these considerations, we have found $p_{\rm cut} = 10^{-3}$ is a reasonable choice for this analysis, and we adopt it throughout. While this analysis was nearing completion, an alternate approach to dealing with these challenges was proposed [51]. It would be interesting to compare these two approaches in future analyses.

In our analysis of $I = 2$ $\pi_5\pi_5$ scattering [13], we implemented Bayesian model averaging. As we had hoped, the systematic uncertainties for $\pi_5$-related observables were greatly reduced in that paper relative to earlier paper [12].

Also, the problem with uncertainties increasing for $t > t_{\min}$ was not apparent, because we were considering primarily $\pi_5$-related observables. We expect this will not be the case for $\sigma$ and $a_{0,1}$-related observables.

### B. Shrinkage estimator of covariance

Suppose one wants to estimate from a multivariate sample a particular element of the covariance matrix; then one usually uses the standard unbiased sample estimator

$$\Sigma_{ij} = \frac{1}{N-1}\sum_{n=1}^{N}\left(x_i^{(n)} - \bar{x}_i\right)\left(x_j^{(n)} - \bar{x}_j\right), \qquad (36)$$

which is derived from the maximum likelihood estimate (MLE) of covariance of a multivariate Gaussian distribution. By the central limit theorem, as $N \to \infty$ the standard estimator approaches the MLE for any distribution. To estimate a full $\mathbb{R}^{K\times K}$ covariance matrix, there are $K(K+1)/2$ independent matrix elements that must be estimated, requiring $N$ independent samples for each one. Furthermore, accurate estimation of the covariance is crucial when using the chi-squared prescription in Eq. (29), since the inverse of the covariance matrix is used and the consequence of poorly estimated small eigenmodes is amplified. Empirically, it has been found that approximately $50K(K+1)/2$ samples are needed in lattice QCD calculations for the standard estimator to be sufficiently accurate for chi-squared fitting [52,53].

If you care about only this particular matrix element, or perhaps one more, then this is the optimal estimator to use. However, if you want to simultaneously estimate three or more elements of the covariance matrix, Stein [54] proved that this was not the optimal estimator in the sense of minimizing the combined mean square error, i.e., $\sum_{ij}(\Sigma_{ij} - \Sigma^*_{ij})^2$, where $\Sigma^*_{ij}$ is the true but unknown covariance. This was so counterintuitive at the time, it was called *Stein's paradox*.

For our purposes, Stein's improved estimator will take the form of the linear shrinkage estimator of covariance:

$$\sigma_{ij}(\lambda) = \lambda\Sigma_{ii}\delta_{ij} + (1-\lambda)\Sigma_{ij}, \qquad \lambda \in [0,1], \qquad (37)$$

and, for a given sample ensemble, there exists some optimal $\lambda^*$ that minimizes the mean squared error (MSE) and $\lambda^* \to 0$ as $N \to \infty$. Since we do not know the true covariance $\Sigma^*$, we must estimate the optimal value. Based on work by Ledoit and Wolf [55], Schäfer and Strimmer [56] gave a fairly straightforward estimator for the optimal value of $\lambda$:

$$\hat{\lambda}^* = \frac{\sum_{i\leq j}\widehat{\mathrm{Var}}(\Sigma_{ij})}{\sum_{i\neq j}\Sigma^2_{ij}}. \qquad (38)$$

In Appendix C, we show a one-pass algorithm to compute the sample estimate of $\widehat{\mathrm{Var}}(\Sigma_{ij})$.

The shrinkage estimator of covariance has been suggested for use in lattice quantum field theory applications for some time [57,58]. Only recently has the shrinkage estimator been actually employed for use in published lattice QCD analyses [59–62]. Recent work by Ledoit and Wolf [63,64] have proposed an improved nonlinear shrinkage estimator. Burda and Jarosz [65] have also developed an improved shrinkage estimator and have developed an open-source PYTHON library called shrinkage to assist in calculations. In this analysis, we have conservatively chosen to use linear shrinkage rather than one of the newer alternatives.

## IV. DETAILED EXAMPLE OF MODEL-AVERAGING ANALYSIS ON A SINGLE ENSEMBLE

We will discuss in detail our analysis of the $96^3 \times 192$, $\beta = 4.8$, $m = 0.00125$ ensemble which is the larger volume companion to the $64^3 \times 128$ ensembles discussed in our previous work [12,13]. It will also serve as a detailed example of how we implemented our model-averaging analysis.

### A. Data selection

In order to compare models fit to different data subsets, we need to first identify the maximal dataset $T$ which could be considered for any model. Although our staggered meson two-point correlation function data are computed from $t = 0$ to $t = N_t - 1 = 191$, the data are first symmetrized: $(C(t) + C(N_t - t))/2 \to C(t)$ and now the largest possible dataset is from $t = 0$ to $t = N_t/2 = 96$. As already mentioned [50], the signal to noise decreases exponentially at large times, so for most correlators, particularly at nonzero momenta $\vec{p} \neq 0$, there is insufficient signal to reasonably include those data points in the analysis, particularly since this will exacerbate the problem of reliable covariance estimation. We will not use data for $t = 0$, 1 given the difficulties of interpreting a staggered correlation function separated by one unit in time in terms of a transfer matrix [37]. We compute the jackknife ratio $C_Q(\vec{p}, t)/C_Q(\vec{p}, 1)$ and choose a minimum value for this quantity for each state $Q$ where there is still good signal to noise for all $\vec{p}$. This defines $t_{\max}$ for each $Q$ and $\vec{p}$.

Figure 1 shows examples of our procedure. On the left, for the $\pi_5$ meson, we see good signal for all momenta to the middle of the lattice, and we also see nice straight lines on the log plot, indicating clear signal of a single decaying exponential. On the right, for the $a_{0,1}$ meson, the situation is somewhat different. There does seem to be pretty good signal to the middle of the lattice, but the nature of the signal changes at large times, with an apparent change of slope and an oscillating signal becoming dominant. We use a rough model to guide our choice of where to draw a horizontal line based on the dispersion relation

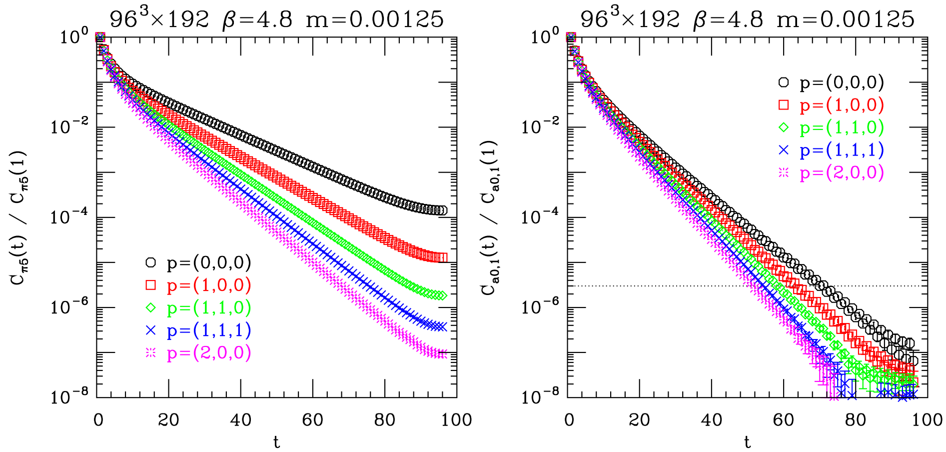



FIG. 1. Two-point correlation functions for $\pi_5$ and $a_{0,1}$ mesons. In the right panel, data points below the horizontal line at $3\times10^{-6}$ were not included in any fits.

$E_Q^2 = M_Q^2 + p^2$ and assuming that a single exponential dominates the correlation function at times $t_c$ where it crosses the line

$$\frac{C_Q(\vec{p},t_c)}{C_Q(\vec{p},1)} = e^{-\sqrt{M_Q^2+p^2}(t_c-1)} = \text{const} \Rightarrow t_c(\vec{p}) \propto \frac{1}{\sqrt{M_{\text{eff}}^2+p^2}}. \tag{39}$$

We compare the computed values to this model, and we see good agreement along the shown cut line. If we lower the cut line, the observed values deviate from the prediction, particularly for $p^2 = 4$, so we conservatively set the cut line at $3\times10^{-6}$. Note that this model will not work well as $t_c \to N_t/2$, since it does not include the additional contribution due to periodic boundary conditions which becomes important in that region. A modified expression involving hyperbolic cosines can be derived, but we did not need it here.

The situation for the $\sigma$ meson correlator is more complicated. In Fig. 2 on the left is the unsubtracted correlator $C_\sigma(\vec{p}=0,t)$. It should be clear that just subtracting some constant value around $c_0 = 262.08\ldots$ in an uncorrelated way, following Eq. (8), would be unsatisfactory because the signal to noise would fall below one in a few time units. The center panel shows $\bar{C}_\sigma(\vec{p}=0,t)$ and, following Eq. (14), the previously large positive constant has been replaced with a 3 orders of magnitude smaller negative constant and greatly enhanced signal to noise. However, we still need to figure out at what time $t_c$ the signal to noise of the exponentially decaying part of the correlator falls below an acceptable level. We cannot judge this from the central panel, since the large time behavior is dominated by the integral of the correlation function. Instead, we compute the ratio $C_\sigma(\vec{p},t)/C_\sigma(\vec{p},1)$ for $\vec{p}\neq0$

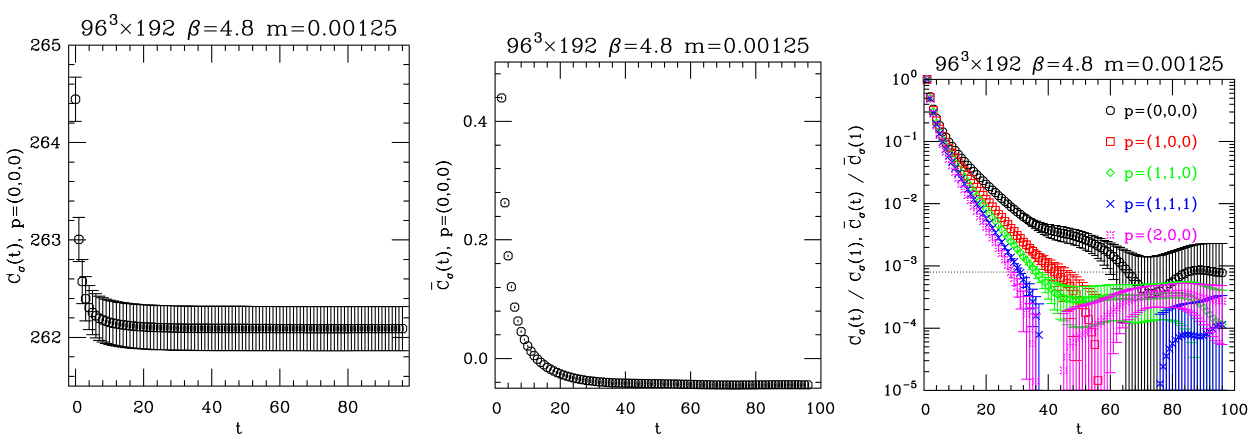



FIG. 2. Two-point correlation functions for $\sigma$ meson. The left panel shows the unsubtracted $\vec{p}=0$ correlator. The center panel shows the $\vec{p}=0$ subtracted correlator. The right panel data points below the horizontal line for $\vec{p}\neq0$ not included in fits. The data from the central panel are included on the right by shifting upward by a sufficiently large constant $\bar{C}(\vec{p}=0,t)+0.045$ so that the result is positive and can be displayed on a log plot. The shifted data cannot be used in the data selection analysis.

TABLE II. Summary of maximum allowed time ranges for fitting in model-averaging procedure for the $96^3 \times 192$, $\beta = 4.8$, $m = 0.00125$ ensemble.

| | $\vec{p} = (0,0,0)$ | $\vec{p} = (1,0,0)$ | $\vec{p} = (1,1,0)$ | $\vec{p} = (1,1,1)$ | $\vec{p} = (2,0,0)$ |
|---|---|---|---|---|---|
| $\pi_5$ | [2, 96] | [2, 96] | [2, 96] | [2, 96] | [2, 96] |
| $a_{0,1}$ | [2, 70] | [2, 63] | [2, 58] | [2, 53] | [2, 50] |
| $\sigma$ | [2, 52] | [2, 41] | [2, 34] | [2, 30] | [2, 27] |

and rely on our crude model Eq. (39) to extrapolate to $\vec{p} = 0$, shown in the right panel. The results in the data selection procedure are summarized in Table II.

### B. Model averaging

As previously discussed, this analysis will use model averaging [25]. In Fig. 3, we show how varying the fitting range $t \in [t_{\min}, t_{\max}]$ affects the relative model probabilities $p(M|D)$. We focus on the $\vec{p} = 0$ mesons, since those states are most susceptible to the presence of $t$-invariant constant contribution to the correlation function. This is true even in the case of the $a_{0,1}$ meson, where the expected constant contribution should vanish in the infinite statistics limit. The $\pi_5$ meson is much less affected by any such constant as can be seen by the preference for model A fits in the model averaging.

### C. Dispersive analysis

Once the model parameters and their errors have been computed for each correlation function computed on a given volume, at a given fermion mass, and at a given spatial momentum $\vec{p}$, the results from various momenta can be used to constrain the values of the parameters in the rest frame using the dispersion relations outlined in Eqs. (15)–(17) for the rest mass $M_Q$ and Eqs. (18)–(21) for the decay constants $\hat{F}_Q$. Parameter estimation is done using least-squared fitting with possible finite lattice spacing corrections included in even powers of $\hat{p}^2$ or $(ap)^2$, as appropriate. Since the number of lattice correction terms needed is unknown *a priori*, we use model averaging to average over the different model choices.

This procedure is probably of marginal benefit for the $\pi_5$ rest mass and decay constant, since those quantities are already very accurately determined directly in the $\vec{p} = 0$ frame and the other momentum frames do not add significant additional information, as shown in Fig. 4. However, these fits also show how the momentum dependence is consistent with the expected dispersion relations up to small lattice artifacts.

For the isosinglet scalar $\sigma$ in Fig. 5 and isotriplet scalar $a_{0,1}$ in Fig. 6, we see similar consistency with the expected dispersion relations. Now, the information from the nonzero momentum frames provides additional significant constraints on the rest mass and decay constant resulting in overall smaller uncertainties than if only the $\vec{p} = 0$ results alone were used. This is particularly important for the $\sigma$ channel, where the correlation function in the $\vec{p} = 0$ frame has a difficult to subtract constant which is not present in nonzero momentum frames.

## V. INFINITE-VOLUME EXTRAPOLATION

We repeat the steps described in detail for one ensemble in Sec. IV for all ensembles in this study. We would like to compare the results of our calculations with various models but those models usually apply to the system only in an infinite volume. We will extrapolate our data to the infinite-volume limit using the model described in Sec. II E.

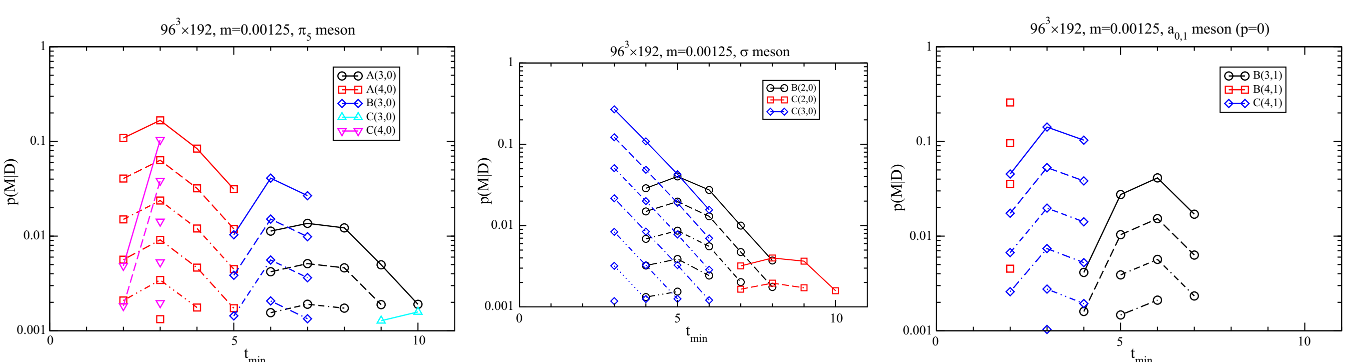



FIG. 3. Relative model probabilities for the $\vec{p} = 0$ $\pi_5$ $\sigma$ and $a_{0,1}$ mesons. The different models are labeled by a letter A, B, C, and integers $(n_{\max}, j_{\max})$, the number of nonoscillating and oscillating states, as described in Sec. II B. The range of time values in each fit $[t_{\min}, t_{\max}]$ are shown in the figures. The uppermost curves correspond to $t_{\max}$ as the maximum value in Table II, and the lower curves correspond to decreasing $t_{\max}$ by one.

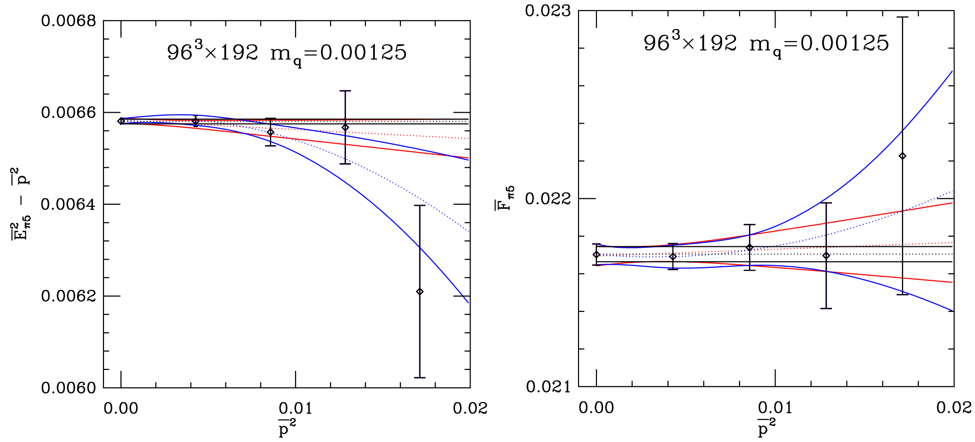



FIG. 4. Momentum dependence of the energy $\hat{E}_{\pi_5}$ and decay constant $\hat{F}_{\pi_5}$. Fits to polynomials in $\hat{p}^2$ up to quadratic order are shown.

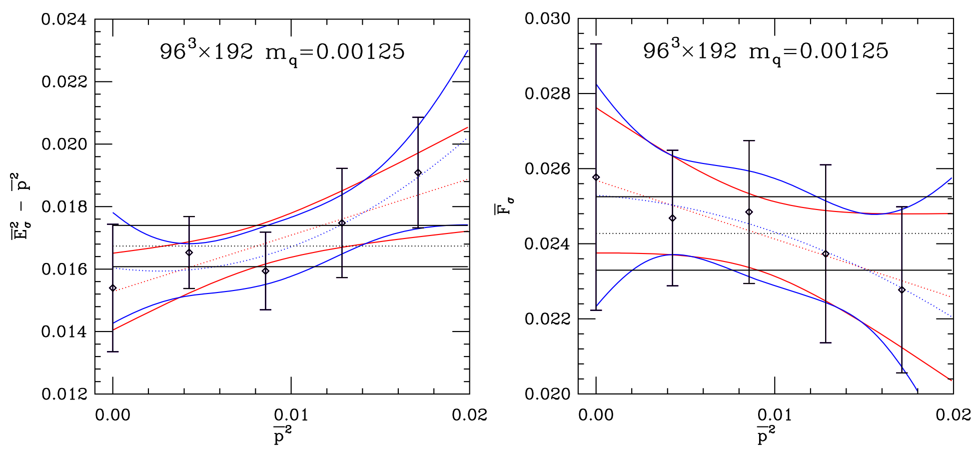



FIG. 5. Momentum dependence of the energy $\hat{E}_{\sigma}$ and decay constant $\hat{F}_{S}$. Fits to polynomials in $\hat{p}^2$ up to quadratic order are shown.

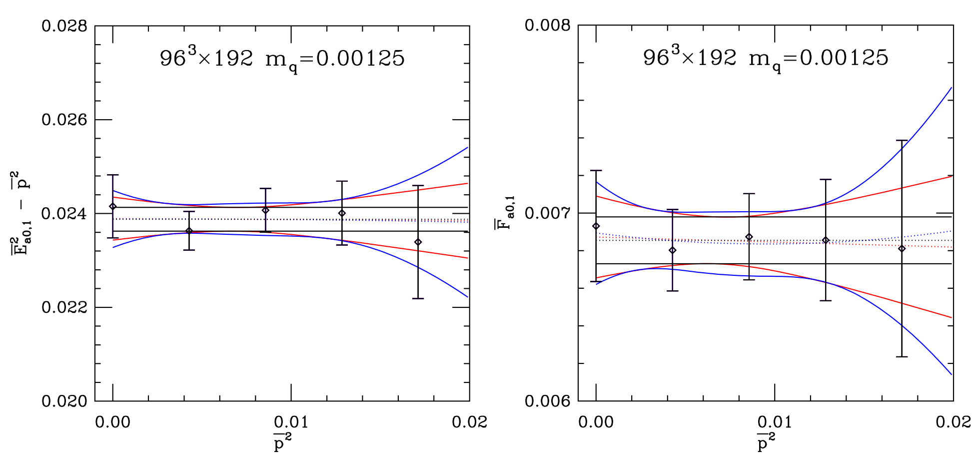



FIG. 6. Momentum dependence of the energy $\hat{E}_{a_{0,1}}$ and decay constant $\hat{F}_{a_{0,1}}$.

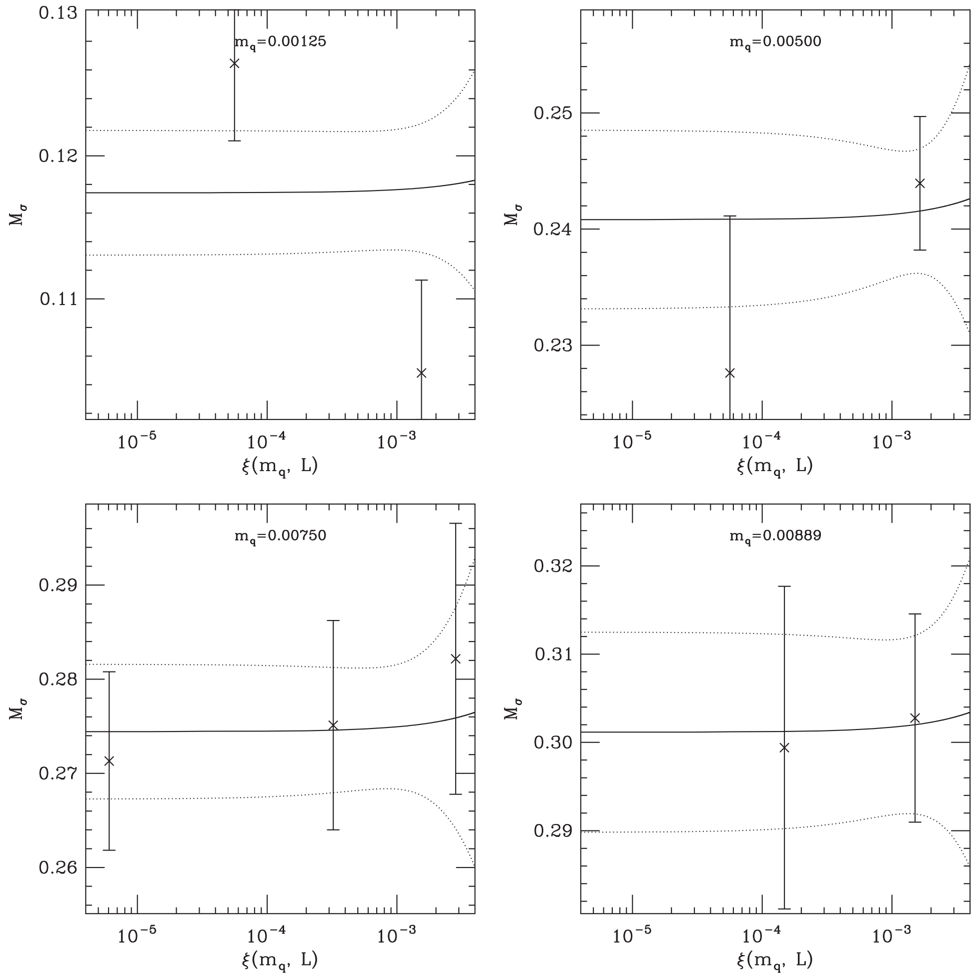



FIG. 7. Infinite-volume extrapolation of the rest masses $M_\sigma(m_q)$. $\alpha_\sigma = 1.9 \pm 16.7$ and $\chi^2/\mathrm{d.o.f.} = 2.05$ with 4 d.o.f.

At each volume and fermion mass, we compute the quantity

$$\xi(m_q, L) \equiv \frac{M_{\pi_5}^2}{(4\pi \hat{F}_{\pi_5})^2} \sum_{n=1}^{8} \frac{4\kappa(n)}{\sqrt{n} M_{\pi_5} L} K_1\left(\sqrt{n} M_{\pi_5} L\right), \qquad (40)$$

where the $m_q$ dependence is implicit in the relevant infinite-volume quantities $M_{\pi_5}$ and $\hat{F}_{\pi_5}$. With this computed quantity, the analysis becomes a simple linear fit.

If we focus just on $M_{\pi_5}$ and $\hat{F}_{\pi_5}$, we know in chiral perturbation theory the quantities $\alpha_{\pi_5}$ and $\beta_{\pi_5}$ defined in Eqs. (22) and (23) appear at a specific order in the chiral expansion and have no implicit fermion mass dependence. We use the same finite-volume model for other masses and decay constants, and we will similarly assume the parameters $\alpha_Q$ and $\beta_Q$ are mass independent as a model choice. This means that $\alpha_Q$ and $\beta_Q$ are determined by a simultaneous fit to the data at all fermion masses and volumes.

The choice of the expansion parameter $\xi(m_q, L)$ being defined in terms of infinite-volume quantities might pose a chicken-and-egg problem when attempting to extrapolate $\pi_5$ data since the infinite-volume values are not known *a priori*. In this case, we start by using the values on the largest volume and then iterate a few times, and the result converges quickly.

An earlier version of the finite-volume extrapolation for $M_{\pi_5}$ and $F_{\pi_5}$ were published previously [13], where it was observed to be a relatively minor correction on our volumes. Our current results are consistent with them, so we focus here on the $\sigma$ channel. The fit of $M_\sigma(m_q, L)$ is shown in Fig. 7, and the fit of $\hat{F}_S(m_q, L)$ is shown in Fig. 8. $\alpha_Q$ and $\beta_Q$ for various channels studied in this work are summarized in Table III. Both from the figures and from the uncertainties on $\alpha_\sigma$ and $\beta_\sigma$ in the table, it is clear that the uncertainties in our $\sigma$ meson observables are still too large to reliably extract the sign and magnitude of these finite-volume corrections. We hope to return to this issue in a future publication.

Studying the other parameters in Table III reveals relationships between parameters which are generated by the strong dynamics and which are qualitatively similar

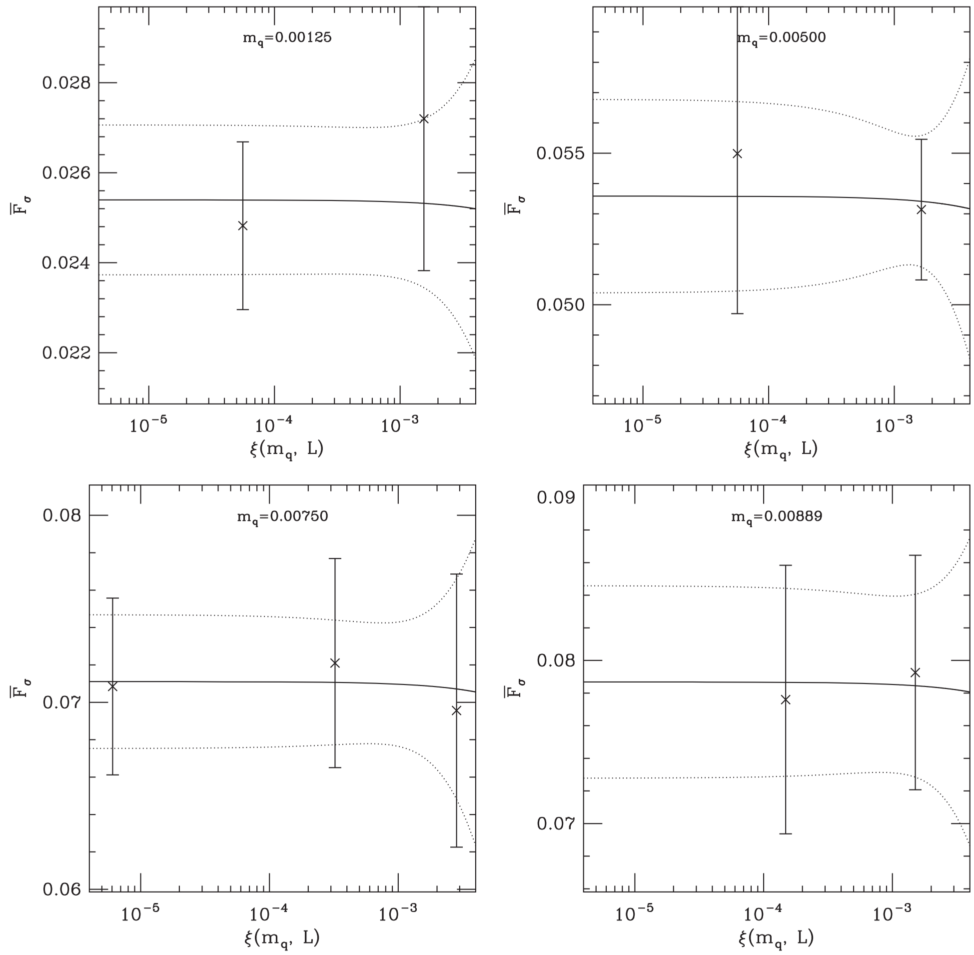



FIG. 8. Infinite-volume extrapolation of the decay constants $\hat{F}_S(m_q)$. $\beta_\sigma = -1.9 \pm 32.0$ and $\chi^2/\mathrm{d.o.f.} = 0.15$ with 4 d.o.f.

to QCD. First, $\mathrm{sgn}(\alpha_Q) = -\mathrm{sgn}(\beta_Q)$ is a well-known feature in QCD. Second, the fact that $\mathrm{sgn}(\alpha_{\pi_5}) = -\mathrm{sgn}(\alpha_{a_{0,1}})$ is also observed in earlier studies [3] and was previously misinterpreted as an indication of "parity doubling" in near-conformal gauge theories, because finite-volume effects would push the masses and decay constants of parity partners $\pi_5$ and $a_{0,1}$ toward degeneracy. We also note that, without a proper infinite-volume extrapolation, if the mass of the $a_{0,1}$ meson were observed to be stable but just below decay threshold, one could wonder whether the state might become unstable in a larger volume. In our calculations, the $a_{0,1}$ meson remains stable even after infinite-volume extrapolation as can be seen in Table IV.

TABLE III. Summary of finite-volume corrections $\alpha_Q$ and $\beta_Q$. All fits have 4 d.o.f. Multiply these parameters by $3/(2\sqrt{2\pi^3}) \approx 0.19$ to compare with [13].

| | $\alpha_Q$ | $\chi^2/\mathrm{d.o.f.}$ | $\beta_Q$ | $\chi^2/\mathrm{d.o.f.}$ |
|---|---|---|---|---|
| $\pi_5$ | 6.53(29) | 3.12 | −9.3(1.3) | 0.65 |
| $\sigma$ | 2(17) | 2.05 | −2(32) | 0.15 |
| $a_{0,1}$ | −27.4(4.6) | 2.88 | 11(11) | 0.29 |

## VI. SYSTEMATIC ERROR ANALYSIS

In our previous $I = 2, \pi\pi$ scattering paper [13], we made a crude estimate of the relative systematic errors affecting our statistical determinations of the $\pi_5$ meson mass $M_{\pi_5}(m_q)$ and decay constant $F_{\pi_5}(m_q)$. Our current statistical-only estimate of uncertainties for quantities like $M_{\pi_5}(m_q)$ and $\hat{F}_{\pi_5}(m_q)$ as shown in Table IV are likely underestimates due to various factors: a small number of independent samples; various modeling choices regarding dispersion relations and finite-volume effects; data quality cuts and model probability cuts in the model-averaging procedure; plus the interplay between the amount of independent data and choices made in the rest of the analysis through the reliability of the shrinkage estimator of covariance.

We would like to estimate how large these effects might be in terms of a single relative systematic error parameter $\rho$ across all the ensembles. We will estimate $\rho$ using a number

TABLE IV. Final summary of infinite-volume ground -state rest masses and decay constants in lattice units. Only statistical uncertainties are shown. Data for $M_{\rho_i}$ copied from [12] for convenience. See Table IX for results with systematic uncertainties included. Supplemental results for all fit parameters are available [44].

| $m_q$ | 0.00125 | 0.00222 | 0.00500 | 0.00750 | 0.00889 |
|---|---|---|---|---|---|
| $M_{\pi_5}$ | 0.081082(32) | 0.10870(12) | 0.165691(73) | 0.205711(33) | 0.22534(13) |
| $\hat{F}_{\pi_5}$ | 0.021677(40) | 0.02794(12) | 0.03982(10) | 0.048314(66) | 0.05262(15) |
| $M_\sigma$ | 0.1174(44) | 0.1545(79) | 0.2408(77) | 0.2744(71) | 0.301(11) |
| $\hat{F}_S$ | 0.0254(17) | 0.0361(37) | 0.0536(32) | 0.0711(36) | 0.0787(59) |
| $M_{a_{0,1}}$ | 0.1536(10) | 0.2070(53) | 0.3119(28) | 0.3773(18) | 0.4193(32) |
| $\hat{F}_{a_{0,1}}$ | 0.00691(14) | 0.00944(43) | 0.1480(27) | 0.01829(24) | 0.02047(42) |
| $M_{\rho_i}$ [12] | 0.1709(65) | 0.2197(37) | 0.3024(63) | 0.36962(77) | 0.4093(21) |

of different observables and then combine those estimates to get an average value for $\rho$. For example, if $\sigma_M$ is the statistical-only estimate of the uncertainty of a given mass $M$, we would like to estimate a relative systematic uncertainty $\rho_M$ such that the total uncertainty is

$$M(m_q) \pm \sqrt{\sigma^2_{M(m_q)} + M(m_q)^2 \rho_M^2}. \tag{41}$$

We assume that the systematic effect is similar across all the different ensembles labeled by different fermion masses $m_q$ so that the parameter $\rho_M$ does not depend on $m_q$.

To estimate $\rho_M$, we do not want to assume any explicit functional dependence for $M(m_q)$, in particular, that we would expect to be valid for small $m_q$ including as $m_q \to 0$. Instead, we imagine that whatever the correct function, it is relatively smooth and slowly varying and can be approximated by a Taylor series expansion around the midpoint $m_0 = 0.00507$ of our range of $m_q$ and $|m_q - m_0| \leq \Delta_m = 0.00382$. We can fit the data to a polynomial

$$M(m_q) \approx \sum_{n=0}^{n_{\max}} a_n (m_q - m_0)^n. \tag{42}$$

Given that we have only five different $m_q$ values, we will compare the $\chi^2$ and AIC values for $n_{\max} = 2$ and $n_{\max} = 3$ and use those comparisons to estimate $\rho_M$. We will also use the ratio test to check for convergence of the series on $m_0 \pm \Delta_m$:

$$\frac{|a_{n+1}\Delta^{n+1}|}{|a_n \Delta^n|} < 1, \quad \forall \ n. \tag{43}$$

Actually, the ratio test requires only the ratio $< 1$ as $n \to \infty$ for convergence, but we will assume convergence if its true term by term up to the largest $n$ we can fit. For this analysis, we will use the data in Table IV.

### A. Fits using statistical-only data

In this section, in Table V we show fits of Eq. (42) to the statistical-only data from Table IV for $n_{\max} = 2, 3$. We then

TABLE V. Basic fits using statistical-only data from Table IV to model function in Eq. (42). Bolded entries indicate observables where model probabilities are higher for $n_{\max} = 2$ than $n_{\max} = 3$.

| Observable | $n_{\max}$ | $\chi^2$ | $\log p(M\|D)$ | $a_0$ | $a_1$ | $a_2$ | $a_3$ |
|---|---|---|---|---|---|---|---|
| $M_{\pi_5}$ | 2 | 1052.0 | −528.0 | 0.166768(60) | 18.563(11) | −1000.9(6.4) | |
| | 3 | 159.5 | −83.8 | 0.167345(63) | 17.472(38) | −950.4(6.6) | 100500(3400) |
| $\hat{F}_{\pi_5}$ | 2 | 76.0 | −41.0 | 0.040144(82) | 3.967(15) | −222.3(8.4) | |
| | 3 | 14.3 | −11.2 | 0.040275(84) | 3.581(51) | −207.4(8.6) | 33600(4200) |
| $M_\sigma$ | 2 | 1.6 | **−3.8** | 0.2382(64) | 22.9(1.3) | −2250(640) | |
| | 3 | 0.8 | **−4.4** | 0.2384(64) | 19.5(4.1) | −2122(657) | $29(33) \times 10^4$ |
| $\hat{F}_S$ | 2 | 0.6 | **−3.3** | 0.0551(28) | 6.93(62) | −200(290) | |
| | 3 | 0.4 | **−4.2** | 0.0550(28) | 6.1(2.0) | −160(300) | $6(16) \times 10^4$ |
| $M_{a_{0,1}}$ | 2 | 14.7 | −10.4 | 0.3097(22) | 33.82(36) | −1820(210) | |
| | 3 | 0.0 | −4.0 | 0.3136(25) | 28.0(1.5) | −1870(210) | $46(12) \times 10^4$ |
| $\hat{F}_{a_{0,1}}$ | 2 | 1.9 | **−3.9** | 0.01483(23) | 1.726(46) | −89(23) | |
| | 3 | 0.0 | **−4.0** | 0.01489(23) | 1.50(17) | −82(23) | $19(14) \times 10^3$ |

TABLE VI. Ratios for convergence testing of fits in Table V. The fit parameter covariance matrix (not shown) was used to compute these uncertainties. Bolded entries indicate observables where the convergence test may fail due to large values or uncertainties.

| Observable | $n_{\max}$ | $\Delta_m\lvert\frac{a_1}{a_0}\rvert$ | $\Delta_m\lvert\frac{a_2}{a_1}\rvert$ | $\Delta_m\lvert\frac{a_3}{a_2}\rvert$ |
|---|---|---|---|---|
| $M_{\pi_5}$ | 2 | 0.42521(36) | 0.2060(14) | |
| | 3 | 0.39884(95) | 0.2078(15) | 0.404(15) |
| $\hat{F}_{\pi_5}$ | 2 | 0.3775(19) | 0.2141(86) | |
| | 3 | 0.3397(56) | 0.2212(97) | 0.618(88) |
| $M_\sigma$ | 2 | 0.368(24) | 0.37(12) | |
| | 3 | 0.312(68) | 0.42(15) | **0.52(66)** |
| $\hat{F}_S$ | 2 | 0.481(54) | 0.11(16) | |
| | 3 | 0.43(14) | 0.10(19) | **1.6(5.9)** |
| $M_{a_{0,1}}$ | 2 | 0.4171(64) | 0.206(26) | |
| | 3 | 0.342(20) | 0.254(35) | **0.94(26)** |
| $\hat{F}_{a_{0,1}}$ | 2 | 0.445(15) | 0.197(54) | |
| | 3 | 0.358(46) | 0.211(64) | **0.86(72)** |

test for convergence by computing the ratios in Eq. (43) and collect the results in Table VI.

If we first look at the model probabilities, we see when the fit is highly constrained, indicated by large $\chi^2$ values, then the fit with $n_{\max} = 3$ is preferred relative to $n_{\max} = 2$. This is the expected behavior, since adding extra fit parameters in a highly constrained fit usually reduces the $\chi^2$ by a sufficient amount to increase the model probability. However, if the fit is poorly constrained, indicated by a small $\chi^2$, adding extra parameters may not increase the model probability. Observables where this occurs are highlighted in Table V, and those observables are probably too noisy to help constrain the systematic error parameter $\rho$.

Looking at the convergence test in Table VI, again we highlight examples where data were too noisy to pass the test with confidence. Again, we will not use those observables to help constrain $\rho$. Note also the strong overlap in the lists of rejected observables from both tables. Finally, we do not expect that the functions will have extremal points in the region where it approximates the data. The zeros of the derivatives are shown in Table VII.

### B. Estimating relative systematic error

To estimate the relative systematic error parameter $\rho$, from Eq. (41) as $\rho$ increases the error bars will increase and the corresponding $\chi^2$ will decrease. What value of $\chi^2$ should we choose to determine $\rho$? *A priori*, two interesting values come to mind: (I) the mean value of the chi-squared distribution for $k$ degrees of freedom, i.e., $\chi^2_k(\rho^{(\mathrm{I})}) = k$; (II) the value of $\chi^2_k$ such that one expects 68% of the time a random sample of the chi-squared distribution

TABLE VII. Zeros of the derivatives of fits in Table V.

| Observable | $n_{\max}$ | $f'(m_q) = 0$ | $f''(m_q) = 0$ |
|---|---|---|---|
| $M_{\pi_5}$ | 2 | 0.013 | |
| | 3 | $0.0070 \pm 0.0069i$ | 0.0070 |
| $\hat{F}_{\pi_5}$ | 2 | 0.013 | |
| | 3 | $0.0059 \pm 0.0056i$ | 0.0059 |
| $M_\sigma$ | 2 | 0.0089 | |
| | 3 | $0.0062 \pm 0.0040i$ | 0.0062 |
| $\hat{F}_S$ | 2 | 0.021 | |
| | 3 | $0.0046 \pm 0.0054i$ | 0.0046 |
| $M_{a_{0,1}}$ | 2 | 0.0131 | |
| | 3 | $0.0052 \pm 0.0043i$ | 0.0052 |
| $\hat{F}_{a_{0,1}}$ | 2 | 0.0135 | |
| | 3 | $0.0053 \pm 0.0050i$ | 0.0053 |

should be less than or equal to that value, i.e., $\chi^2_1(\rho) = 1$, $\chi^2_2(\rho^{(\mathrm{II})}) \approx 2.3$.

*A posteriori*, we noticed that from a model-averaging perspective, the four-parameter cubic polynomial fit has the higher model probability at $\rho = 0$ in cases where the statistical error is small compared to the expected systematic error. In the $\rho \to \infty$ limit, $\chi^2 \to 0$ and the most likely model is the one with the smallest $n_{\max}$. As $\rho$ increases, there is a point where the quadratic and cubic polynomial fits have equal probability. We define

$$\mathrm{AIC}(\rho, n_{\max}) = \frac{1}{2}\chi^2_{5-n_{\max}} + n_{\max} + 1 \qquad (44)$$

and choose a third interesting value of $\rho$: (III) $\mathrm{AIC}(\rho^{(\mathrm{III})}, 2) = \mathrm{AIC}(\rho^{(\mathrm{III})}, 3)$. Note that this does not always have a solution, particularly if the quadratic fit has a lower AIC at $\rho = 0$. *A posteriori*, we can rationalize this choice as the point where the quadratic and cubic descriptions of the data are equally good (or bad) from an information-theoretic perspective.

### C. Summary of systematic error analysis

From Table VIII, we can see there are eight unique $\rho$ values from approximately 0.007 to 0.021. Rather than pick

TABLE VIII. Various estimates of the systematic error parameter $\rho$ as determined by methods described in the text.

| Observable | $n_{\max}$ | $\rho^{(\mathrm{I})}$ | $\rho^{(\mathrm{II})}$ | $\rho^{(\mathrm{III})}$ |
|---|---|---|---|---|
| $M_\pi$ | 2 | 0.0169 | 0.0158 | 0.0162 |
| | 3 | 0.0068 | 0.0068 | 0.0162 |
| $F_\pi$ | 2 | 0.0210 | 0.0196 | 0.0197 |
| | 3 | 0.0098 | 0.0098 | 0.0197 |

TABLE IX. Final summary of infinite-volume ground-state rest masses and decay constants with relative systematic error of $\rho = 0.0165$ included following Eq. (41). Data for $M_{\rho_i}$ derived from [12] for convenience. Results with only statistical errors in Table IV.

| $m_q$ | 0.00125 | 0.00222 | 0.00500 | 0.00750 | 0.00889 |
|---|---|---|---|---|---|
| $M_{\pi_5}$ | 0.0811(13) | 0.1087(18) | 0.1657(27) | 0.2057(34) | 0.2253(37) |
| $\hat{F}_{\pi_5}$ | 0.02168(36) | 0.02794(47) | 0.03982(66) | 0.04831(80) | 0.05262(88) |
| $M_\sigma$ | 0.1174(48) | 0.1545(83) | 0.2408(87) | 0.2744(85) | 0.301(12) |
| $\hat{F}_S$ | 0.0254(17) | 0.0361(37) | 0.0536(33) | 0.0711(38) | 0.0787(60) |
| $M_{a_{0,1}}$ | 0.1536(27) | 0.2070(63) | 0.3119(58) | 0.3773(65) | 0.4193(76) |
| $\hat{F}_{a_{0,1}}$ | 0.00691(17) | 0.00944(45) | 0.1480(36) | 0.01829(39) | 0.02047(54) |
| $M_{\rho_i}$ [12] | 0.1709(71) | 0.2197(52) | 0.3024(80) | 0.3696(61) | 0.4093(71) |

just one, we consider a few summary statistics: the arithmetic mean $\rho_a = 0.0157$, the median $\rho_m = 0.0165$, or the geometric mean $\rho_g = 0.0148$. All give relatively similar values close to the central grouping. We make a conservative choice and choose the largest of the three $\rho = \rho_m = 0.0165$. If we compare this estimate to the previous rough guess of 0.01 quoted in [13], it is nice to see they are not too different and that 0.01 falls within the range of estimated values. A final summary of our results with the relative systematic error included is given in Table IX.

In Fig. 9, we compare the previously computed results for $M_\sigma$ with combined statistical and systematic errors as described in [12] with the new results of Table IX for $M_\sigma$ and $M_{\pi_5}$. The values for $\sqrt{8t_0}/a$ are taken from Table I of [12], and the plot style is similar to Fig. 10 of [12]. The conclusion we draw from this comparison was that the previous analysis method for computing $M_\sigma$ led to systematically lower mass values and that the method used previously to estimate systematic errors was sufficiently conservative as to cover the downward shift of the result.

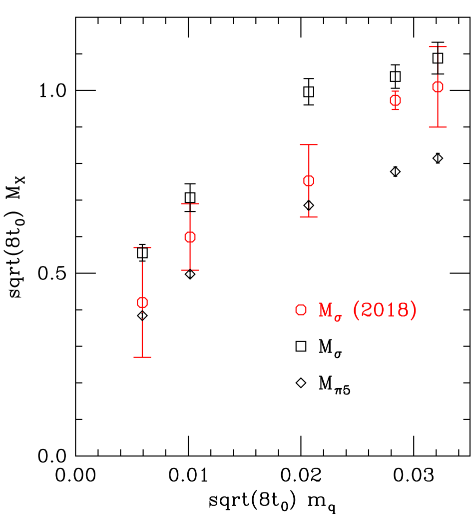



FIG. 9. A comparison of $M_\sigma$ computed previously [12] with the results from Table IX.

## VII. COMMENTS ON CHIRAL EXPANSIONS

While the SU(3) $N_f = 8$ theory with massive Dirac fermions is a potentially interesting theory on its own, being a possible candidate for composite dark matter [66], the theory closer to the chiral limit $m_q \to 0$ might also be relevant for composite Higgs phenomenology. As stated in Sec. I, our results alone are not sufficient to establish with certainty whether the massless $N_f = 8$ theory is inside or outside the conformal window. But, the low-energy content of the two scenarios is quite different: in one case a very strongly coupled conformal field theory and in the other case massless Nambu-Goldstone bosons and possible light flavor-singlet scalar resonance with a mass of the order of $F_\pi$. Specific models will appear quite different depending on the scenario, and when fitted to our data, those models will, in general, be an expansion in some small parameter which vanishes in the chiral limit. We will discuss different specific extrapolations in detail in a companion paper [34].

Here we note that, based on one's *a priori* expectation for the nature of the low-energy theory in the chiral limit, the choice of expansion parameter can lead to very different presentations of the data. One could naively plot results versus the fermion mass $m_q$, or some power of the fermion mass $m_q^{1/(1+\gamma^*)}$, $0 \le \gamma^* \le 2$ motivated by assuming conformal symmetry in the chiral limit, or $\chi \equiv M_{\pi_5}^2/(4\pi\hat{F}_{\pi_5})^2$ by assuming spontaneous chiral symmetry breaking in the chiral limit. In theory like SU(3) $N_f = 2$ (QCD), these choices often do not make any appreciable difference in the presentation of the data. But in this theory, if the chiral limit is conformal, the expansion parameter $\chi$ does not vanish as $m_q \to 0$. Visually, we can see the difference in Fig. 10. Since the value of $\gamma^*$ is a dynamical parameter that can be determined only through a careful extrapolation, we plot three representative values that cover weakly and strongly coupled CFTs plus an intermediate value. Regardless of which value of $\gamma^*$ is chosen, $\chi$ varies significantly over the range of $m_q$ with a fair degree of curvature which makes it difficult to estimate how small $m_q$ must be before the constant term in $\chi$ dominates the leading $m_q$-dependent term. Of course, if $\chi$ vanishes in the chiral limit, then the

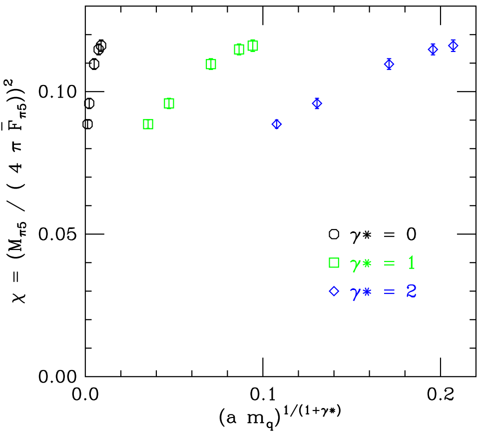



FIG. 10. Chiral expansion parameter $\chi = M_{\pi_5}^2/(4\pi\hat{F}_{\pi_5})^2$ versus other chiral expansion parameters $m_q^{1/(1+\gamma^*)}$. If the theory is conformal, $\chi$ should be nonzero in the chiral limit. If the theory is spontaneously broken, $\chi$ should be zero in the chiral limit.

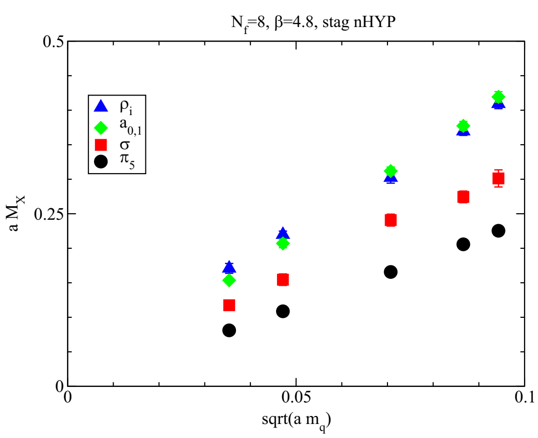
constant term will never dominate. This suggests it will be difficult to distinguish with much certainty given our current results whether or not $\chi$ vanishes in the chiral limit. Calculations at smaller fermion masses are needed.

We can now present the results for the spectrum in two different ways. In Fig. 11, on the left is a presentation appropriate when assuming the theory is conformal in the chiral limit with a mass anomalous dimension $\gamma^* \approx 1$. In units of the lattice spacing $a$, the masses of all the hadrons are expected to extrapolate to zero, since any nonzero hadron mass would break conformal symmetry. On the right is a presentation assuming the chiral symmetry is spontaneously broken in the chiral limit, and the relevant scale of chiral symmetry breaking is set by $4\pi\hat{F}_{\pi_5}$. All the hadron masses except the pion should be nonzero in the chiral limit. Plotted this way, the pion is shown as a simple curve since $M_{\pi_5}/4\pi\hat{F}_{\pi_5} = \sqrt{\chi}$. This also makes it easy to display the energy threshold as a dotted line for decay to two pions. In the current dataset, both the flavor-singlet and nonsinglet scalar mesons appear to be unable to decay to two pions.

Focusing solely on the data presented in this section, it is still far from clear whether or not the chiral parameter $\chi$ vanishes in the chiral limit. On the other hand, $\chi$ varies significantly with the fermion mass, which also suggests we are far from the hyperscaling limit where $\chi$ should be a nonzero constant. Recent numerical studies with improved gauge action [10] suggest that the SU(3) $N_f = 8$ system could be at the opening of the conformal window or at least very close to it. There are indications of an infrared fixed point at much stronger couplings than what is probed by our data in this paper. This is so even if $N_f = 8$ is below the conformal window. Therefore, corrections to scaling in the gauge coupling could be significant. This can explain our inability to distinguish between the conformal and chirally broken scenarios.

## VIII. GMOR RATIO RESULTS

As discussed in Sec. II F, numerical studies of the GMOR ratio can shed light on the low-energy behavior of the $N_f = 8$ theory by measuring how much the ground-state pion pole contributes the pseudoscalar two-point correlation function. A value close to unity indicates pion pole dominance. Table X shows the computed values for the chiral condensate and the integrated pseudoscalar correlation function. Although computed by two different techniques, the results agree extremely well with Eq. (25). Using the largest volume data at each fermion mass for the condensate and the statistical-only data in Table IV, we compute the GMOR ratio in Eq. (28) and propagate the statistical-only errors. We then apply the relative systematic



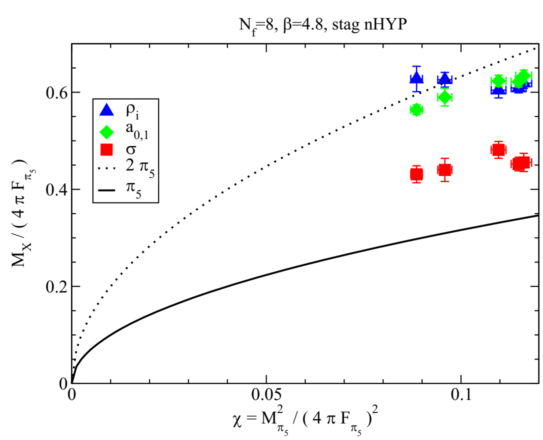




FIG. 11. Two different presentations of the spectrum from Table IX. On the left, in units of the lattice spacing $a$ versus a chiral expansion parameter assuming conformal symmetry and $\gamma^* \approx 1$. On the right, in units of the chiral breaking scale $4\pi\hat{F}_{\pi_5}$ versus a chiral expansion parameter assuming spontaneous chiral symmetry breaking. The dotted line on the right indicates the energy threshold for decays to two pions.

TABLE X. Values for the staggered chiral condensate $\langle\bar{\chi}\chi\rangle$, computed using a noisy estimator, and the integrated pseudoscalar correlation function, computed using a point source. The reader can verify that the columns satisfy Eq. (25). Only statistical errors are shown. The rightmost column shows the GMOR ratio defined in Eq. (28) with errors computed as described in the text.

| $m_q$ | $L$ | $\langle\bar{\chi}\chi\rangle$ | $\sum_t C_{\pi_5}(\vec{0},t)$ | $R_G(m_q)$ |
|---|---|---|---|---|
| 0.00125 | 96 | 0.0121704(53) | 9.732(28) | 2.462(42) |
| | 64 | 0.0121220(70) | 9.641(17) | |
| 0.00222 | 48 | 0.019918(14) | 9.039(25) | 2.397(45) |
| 0.00500 | 48 | 0.040808(18) | 8.164(41) | 2.344(41) |
| | 32 | 0.040521(21) | 8.103(12) | |
| 0.00750 | 48 | 0.058447(12) | 7.808(20) | 2.219(37) |
| | 32 | 0.058445(26) | 7.786(16) | |
| | 24 | 0.058063(73) | 7.752(32) | |
| 0.00889 | 32 | 0.068086(28) | 7.661(37) | 2.152(38) |
| | 24 | 0.067814(46) | 7.615(17) | |

error correction estimated in Sec. VI. The results are shown in the rightmost column in Table X. The lowest pole does not fully dominate the pion correlation function in our fermion mass range, as the result is larger than one. If we assume a mass-deformed CFT is the correct low-energy description, then the lowest pion pole will never dominate the pseudoscalar correlation function at any fermion mass, as it is not a pseudo-Nambu-Goldstone boson, as in Eq. (28).

In the case the $N_f = 8$ theory is outside the conformal window, one of the poles contributing to the pion correlation function would have a pole position at $M_{\pi_5} + M_\sigma$ and a residue proportional to $g^2_{\pi\pi\sigma}$ which we will assume is $\mathcal{O}(\hat{F}^2_{\pi_5})$. For the leading pion pole to dominate, $M^2_{\pi_5}/(M_{\pi_5} + M_\sigma)^2 \ll 1$. In this work, the ratio varies from 0.167(8) to 0.183(9), which in this scenario is interpreted as not small enough to ensure pion pole dominance. A direct calculation of the coupling $g_{\pi\pi\sigma}$ and/or further calculations at lighter fermion masses should shed light on this issue.

We did not perform an infinite volume extrapolation of the condensate data in Table X similar to the ones described in Sec. II E. The systematic effect of this correction might be significant on the scale of the uncertainties shown for the GMOR ratio. But the effect is unlikely to be significant relative to the deviation of the ratio from unity. In the future, if a detailed model is to be fit to these data, the modeler should consider including these neglected corrections.

## IX. DISCUSSION

In this investigation, we have made many methodological improvements with respect to our earlier lattice study of the $N_f = 8$ theory [12]. In particular, we have employed two different methods for dealing with time-independent contributions to the flavor-singlet scalar correlator, first by using the subtraction scheme developed in Sec. II A and then by working with moving frames and applying the dispersion relation described in Sec. II C. We were able to substantially reduce the systematic uncertainties of our fit results using the Bayesian model-averaging approach. Additionally, we used improved “linear” shrinkage estimators for data covariance which we found were more reliable given the amount of statistics. There was an open question in our previous paper whether finite-volume effects could be significant even when $M_{\pi_5}L \gtrsim 5.3$. Now, we can see that the finite-volume effects are mild and do not play a significant role in the final result. We find that $M_\sigma/M_{\pi_5}$ ranges from 1.45 to 1.34 as $M_{\pi_5}/M_{\rho_i}$ increases from 0.47 to 0.55.

We computed a new observable, the scalar decay constant $\hat{F}_S$, which, as we show in a companion paper [34], provides useful independent constraints on various low-energy effective theories. We also computed the flavor-nonsinglet scalar meson mass $M_{a_{0,1}}$ and decay constant $\hat{F}_{a_{0,1}}$. The proximity of the $a_{0,1}$ to the decay threshold suggests that a careful elastic scattering analysis might be warranted in the future if more accurate results are desired.

In this work, we have focused on the SU(3) $N_f = 8$ theory with small, but nonzero, fermion masses and have discussed in general terms in Secs. VII and VIII what hints these results might give us regarding the chiral limit of the theory without appealing to a specific low-energy EFT. In a companion paper [34], we took an alternate approach assuming the results of Table IX as definitive and attempted to match the results to two EFTs, a dilaton EFT and a mass-deformed CFT, which both assume that the underlying gauge theory is near-conformal and strongly influenced by a nearby conformal fixed point. The main difference between the two EFTs is whether or not the theory actually becomes conformal in the chiral limit. The reader should refer to the companion paper for the details, but it should perhaps not be surprising given the general discussion above that the conclusion is the current data does not reach light enough fermion masses to be definitive.

## ACKNOWLEDGMENTS

R. C. B. and C. R. acknowledge U.S. Department of Energy (DOE) Grant No. DE-SC0015845. K. K. Ċ. acknowledges support from the DOE through the Computational Sciences Graduate Fellowship (DOE CSGF) through Grant No. DE-SC0019323 and also from the P. E. O. Scholar award. G. T. F. acknowledges support from DOE Grant No. DE-SC0019061. A. G. is supported by SNSF Grant No. 200021_17576. A. H. and E. T. N. acknowledge support by DOE Grant No. DE-SC0010005. J. I. acknowledges support from ERC Grant No. 101039756. D. S. was supported by United Kingdom Research and Innovation Future Leader Fellowship No. MR/S015418/1 and No. MR/X015157/1 and STFC Grants No. ST/T000988/1 and No. ST/X000699/1. P. V. acknowledges the support of the DOE under Contract No. DE-AC52-07NA27344 (Lawrence Livermore National Laboratory, LLNL). We thank the LLNL Multiprogrammatic and Institutional Computing program for Grand Challenge supercomputing allocations. We also thank Argonne Leadership Computing Facility (ALCF) for allocations through the INCITE program. A. L. C. F. is supported by DOE Contract No. DE-AC02-06CH11357. Computations for this work were carried out in part on facilities of the USQCD Collaboration, which are funded by the Office of Science of the DOE, and on Boston University computers at the MGHPCC, in part funded by the National Science Foundation (Grant No. OCI-1229059). This research utilized the NVIDIA GPU accelerated Summit supercomputer at Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the DOE Office of Science under Contract No. DE-AC05-00OR22725.

## APPENDIX A: ENSEMBLES

A summary of the ensembles used in this paper is shown in Tables XI and XII.

## APPENDIX B: NORMALIZATION OF MODEL PROBABILITIES

When performing an aggressive model-averaging analysis by considering a wide range of models $\{M\}$ and a wide range of data subset selections $T_1$ for each model, the resulting set of $\log p(M|D)$ can vary by several orders of magnitude, making it numerically challenging to perform an accurate calculation of $\sum_{\{M\}} p(M|D)$. In particular, exponentiating each $\log p(M|D)$ and then performing the sum seems like a bad idea. So, we work directly with $\log p(M|D)$ to compute the log of the sum. Let $\ell_n$ be a sorted list of the $\log p(M|D)$: $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_N$. We can construct the partial sums recursively:

$$s_1 = \ell_1, \qquad s_n = s_{n-1} + \log\left(1 + e^{\ell_n - s_{n-1}}\right) \quad (n > 1). \tag{B1}$$

The final sum over model probabilities is $\sum_{\{M\}} p(M|D) = \exp s_N$. The key observation is that sorting the list ensures that two wildly different numbers are not combined at any step with accompanying large loss of precision.

TABLE XI. Ensembles, or Markov chains, used in this study with $0.005 \leq m_q \leq 0.00889$. "Try" assigns a label to each Markov chain and the label "**C**" indicates the combined summary for all chains at a given mass and volume. "Period" indicates how often the correlation functions were computed.

| Volume | Mass | Try | Traj | Period (Traj) | Block (Traj) | $N_{\text{blk}}$ |
|---|---|---|---|---|---|---|
| $24^3 \times 48$ | 0.00889 | 1 | [250, 25000] | 10 | **100** | **247** |
| $32^3 \times 64$ | | 1 | [1040, 7000] | 40 | 80 | 75 |
| | | 2 | [1040, 7000] | 40 | 80 | 75 |
| | | 3 | [1040, 7000] | 40 | 80 | 75 |
| | | 4 | [1040, 7000] | 40 | 80 | 75 |
| | | **C** | | | **80** | **300** |
| $24^3 \times 48$ | 0.0075 | 1 | [350, 10000] | 10 | **90** | **107** |
| $32^3 \times 64$ | | 1 | [255, 1395] | 10 | 100 | |
| | | | [1400, 25160] | 5 | 100 | **249** |
| $48^3 \times 96$ | | 1 | [250, 9990] | 10 | 70 | 139 |
| | | 2 | [250, 9990] | 10 | 70 | 139 |
| | | **C** | | | **70** | **278** |
| $32^3 \times 64$ | 0.005 | 1 | [251, 29641] | 5 | 100 | 293 |
| | | 2 | [20011, 22815] | 2 | 100 | 28 |
| | | 3 | [29001, 31653] | 2 | 100 | 26 |
| | | 4 | [10001, 13293] | 2 | 100 | 32 |
| | | **C** | | | **100** | **379** |
| $48^3 \times 96$ | | 1 | [250, 4200] | 10 | 50 | 79 |
| | | 2 | [250, 3390] | 10 | 50 | 63 |
| | | **C** | | | **50** | **142** |

TABLE XII. Ensembles, or Markov chains, used in this study with $0.00125 \leq m_q \leq 0.00222$. "Try" assigns a label to each Markov chain and the label "**C**" indicates the combined summary for all chains at a given mass and volume. "Period" indicates how often the correlation functions were computed.

| Volume | Mass | Try | Traj. | Period (Traj.) | Block (Traj.) | $N_{\mathrm{blk}}$ |
|---|---|---|---|---|---|---|
| $48^3 \times 96$ | 0.00222 | 1 | [250, 11190] | 2 | 120 | 91 |
| | | 2 | [1000, 9930] | 2 | 120 | 74 |
| | | 3 | [210, 1450] | 10 | 120 | 10 |
| | | 4 | [210, 1410] | 10 | 120 | 10 |
| | | 5 | [210, 1360] | 10 | 120 | 9 |
| | | 6 | [210, 1290] | 10 | 120 | 9 |
| | | 7 | [210, 1350] | 10 | 120 | 9 |
| | | **C** | | | **120** | **212** |
| $64^3 \times 128$ | 0.00125 | r0 | [200, 2060] | 10 | 120 | 15 |
| | | r1 | [200, 1990] | 10 | 120 | 15 |
| | | r2 | [200, 2010] | 10 | 120 | 15 |
| | | r3 | [200, 2070] | 10 | 120 | 15 |
| | | s0 | [8436, 17088] | 12 | 120 | 72 |
| | | s1 | [7644, 17472] | 12 | 120 | 82 |
| | | s2 | [7212, 17412] | 12 | 120 | 86 |
| | | **C** | | | **120** | **300** |
| $96^3 \times 128$ | | 2 | [500, 3144] | 2 | 80 | 34 |
| | | 3 | [500, 3282] | 2 | 80 | 35 |
| | | **C** | | | **80** | **69** |

## APPENDIX C: UNBIASED SAMPLE ESTIMATOR FOR THE VARIANCE OF THE COVARIANCE

Using *Mathematica*'s MomentConvert[] functionality, it is a few lines of code to express the unbiased sample estimator for $\widehat{\mathrm{Var}}(\Sigma_{ij})$ in terms of raw moments:

```
centMom11Est = MomentConvert[CentralMoment[{1, 1}], "SampleEstimator"];
bias = MomentConvert[centMom11Est, {Moment, n}];
MomentConvert[(centMom11Est - bias)^2, {Moment, n}].
```

The result is

$$\frac{n^3}{n-1}\widehat{\mathrm{Var}}(\mathrm{Cov}(x,y)) = (n-1)\mu_{2,2} - 2(n-1)(\mu_{2,1}\mu_{0,1} + \mu_{1,0}\mu_{1,2}) + \mu_{2,0}\mu_{0,2} + (n-2)(\mu_{2,0}\mu_{0,1}^2 + \mu_{1,0}^2\mu_{0,2}) - (n-2)\mu_{1,1}^2 + 2(3n-4)\mu_{1,1}\mu_{1,0}\mu_{0,1} - 2(2n-3)\mu_{1,0}^2\mu_{0,1}^2, \tag{C1}$$

where we use *Mathematica*'s convention for raw moments:

$$\mu_{i,j;\mathcal{S}} = \frac{1}{n}\sum_{(x,y)\in\mathcal{S}} x^i y^j. \tag{C2}$$

Following Pébay [67], we would like to construct a one-pass, parallelizable computation. To explain the notation, $\mathcal{S}$ is a set of $n$ samples that can be partitioned into two subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ of $n_1$ and $n_2$ samples, respectively, so $n_1 + n_2 = n$. The computation can be parallelized by performing computations on the subsets and combining the results. In the special case where $n_1 = n-1$ and $n_2 = 1$, the results simplify and can be used as a one-pass algorithm:

$$\mu_{i,j;\mathcal{S}} = \frac{n_1}{n}\mu_{i,j;\mathcal{S}_1} + \frac{n_2}{n}\mu_{i,j;\mathcal{S}_2} = \mu_{i,j;\mathcal{S}_1} + \frac{n_2}{n}\left(\mu_{i,j;\mathcal{S}_2} - \mu_{i,j;\mathcal{S}_1}\right), \tag{C3}$$

where the first form is symmetric and more useful when $\mathcal{S}_1$ and $\mathcal{S}_2$ are of comparable size and the second form is better suited when $\mathcal{S}_2$ is a single sample $(x, y)$:

$$\mu_{i,j;\mathcal{S}} = \mu_{i,j;\mathcal{S}_1} + \frac{1}{n}(x^i y^j - \mu_{i,j;\mathcal{S}_1}). \tag{C4}$$

[1] W. E. Caswell, Asymptotic behavior of non-Abelian gauge theories to two loop order, Phys. Rev. Lett. **33**, 244 (1974).

[2] T. Banks and A. Zaks, On the phase structure of vector-like gauge theories with massless fermions, Nucl. Phys. **B196**, 189 (1982).

[3] G. T. Fleming, Strong interactions for the LHC, Proc. Sci. LATTICE2008 (**2008**) 021.

[4] V. Drach, Composite electroweak sectors on the lattice, Proc. Sci., LATTICE2019 (**2020**) 242.

[5] K. I. Ishikawa, Y. Iwasaki, Y. Nakayama, and T. Yoshie, IR fixed points in $SU(3)$ gauge theories, Phys. Lett. B **748**, 289 (2015).

[6] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. H. Wong, Is SU(3) gauge theory with 13 massless flavors conformal?, Proc. Sci. LATTICE2018 (**2018**) 198.

[7] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. H. Wong, Case studies of near-conformal $\beta$-functions, Proc. Sci. LATTICE2019 (**2019**) 121.

[8] T. Appelquist *et al.* [Lattice Strong Dynamics (LSD) Collaboration], Approaching conformality with ten flavors, arXiv:1204.6000.

[9] T. Appelquist *et al.* [Lattice Strong Dynamics (LSD) Collaboration], Near-conformal dynamics in a chirally broken system, Phys. Rev. D **103**, 014504 (2021).

[10] A. Hasenfratz, Emergent strongly coupled ultraviolet fixed point in four dimensions with eight Kähler-Dirac fermions, Phys. Rev. D **106**, 014513 (2022).

[11] T. Appelquist *et al.* (LSD Collaboration), Lattice simulations with eight flavors of domain wall fermions in SU(3) gauge theory, Phys. Rev. D **90**, 114502 (2014).

[12] T. Appelquist *et al.* (Lattice Strong Dynamics Collaboration), Nonperturbative investigations of SU(3) gauge theory with eight dynamical flavors, Phys. Rev. D **99**, 014509 (2019).

[13] T. Appelquist *et al.* [Lattice Strong Dynamics (LSD) Collaboration], Goldstone boson scattering with a light composite scalar, Phys. Rev. D **105**, 034505 (2022).

[14] T. Kunihiro, S. Muroya, A. Nakamura, C. Nonaka, M. Sekiguchi, and H. Wada (SCALAR Collaboration), Scalar mesons in lattice QCD, Phys. Rev. D **70**, 034504 (2004).

[15] M. Wakayama, T. Kunihiro, S. Muroya, A. Nakamura, C. Nonaka, M. Sekiguchi, and H. Wada, Lattice QCD study of four-quark components of the isosinglet scalar mesons: Significance of disconnected diagrams, Phys. Rev. D **91**, 094508 (2015).

[16] D. Howarth and J. Giedt, The sigma meson from lattice QCD with two-pion interpolating operators, Int. J. Mod. Phys. C **28**, 1750124 (2017).

[17] R. A. Briceno, J. J. Dudek, R. G. Edwards, and D. J. Wilson, Isoscalar $\pi\pi$ scattering and the $\sigma$ meson resonance from QCD, Phys. Rev. Lett. **118**, 022002 (2017).

[18] R. A. Briceno, J. J. Dudek, and R. D. Young, Scattering processes and resonances from lattice QCD, Rev. Mod. Phys. **90**, 025001 (2018).

[19] R. A. Briceno, J. J. Dudek, R. G. Edwards, and D. J. Wilson, Isoscalar $\pi\pi, K\bar{K}, \eta\eta$ scattering and the $\sigma, f_0, f_2$ mesons from QCD, Phys. Rev. D **97**, 054513 (2018).

[20] D. Guo, A. Alexandru, R. Molina, M. Mai, and M. Döring, Extraction of isoscalar $\pi\pi$ phase-shifts from lattice QCD, Phys. Rev. D **98**, 014507 (2018).

[21] A. Rodas, J. J. Dudek, and R. G. Edwards, The quark mass dependence of $\pi\pi$ scattering in isospin 0, 1 and 2 from lattice QCD, Phys. Rev. D **108**, 034513 (2023).

[22] A. Rodas, J. J. Dudek, and R. G. Edwards, Constraining the quark mass dependence of the lightest resonance in QCD, arXiv:2304.03762.

[23] K. Kawarabayashi and M. Suzuki, Partially conserved axial vector current and the decays of vector mesons, Phys. Rev. Lett. **16**, 255 (1966).

[24] Riazuddin and Fayyazuddin, Algebra of current components and decay widths of rho and K* mesons, Phys. Rev. **147**, 1071 (1966).

[25] W. I. Jay and E. T. Neil, Bayesian model averaging for analysis of lattice field theory results, Phys. Rev. D **103**, 114502 (2021).

[26] G. Panico and A. Wulzer, *The Composite Nambu-Goldstone Higgs* (Springer, New York, 2016), Vol. 913, 10.1007/978-3-319-22617-0.

[27] G. Ferretti, Gauge theories of partial compositeness: Scenarios for Run-II of the LHC, J. High Energy Phys. 06 (2016) 107.

[28] O. Witzel, Review on composite Higgs models, Proc. Sci. LATTICE2018 (**2019**) 006.

[29] G. Cacciapaglia, G. Ferretti, T. Flacke, and H. Serôdio, Light scalars in composite Higgs models, Front. Phys. **7**, 22 (2019).

[30] G. Cacciapaglia, C. Pica, and F. Sannino, Fundamental composite dynamics: A review, Phys. Rep. **877**, 1 (2020).

[31] L. Vecchi, A dangerous irrelevant UV-completion of the composite Higgs, J. High Energy Phys. 02 (2017) 094.

[32] T. Appelquist, J. Ingoldby, and M. Piai, Nearly conformal composite Higgs model, Phys. Rev. Lett. **126**, 191804 (2021).

[33] T. Appelquist, J. Ingoldby, and M. Piai, Composite two-Higgs doublet model from dilaton effective field theory, Nucl. Phys. **B983**, 115930 (2022).

[34] T. Appelquist *et al.* [Lattice Strong Dynamics (LSD) Collaboration], Hidden conformal symmetry from the lattice, Phys. Rev. D **108**, L091505 (2023).

[35] D. Daniel and S. N. Sheard, Perturbative corrections to staggered fermion lattice operators, Nucl. Phys. **B302**, 471 (1988).

[36] J. Engels, F. Karsch, and H. Satz, Finite size effects in euclidean lattice thermodynamics for noninteracting bose and Fermi systems, Nucl. Phys. **B205**, 239 (1982).

[37] G. Kilcup and S. R. Sharpe, A tool kit for staggered fermions, Nucl. Phys. **B283**, 493 (1987).

[38] F. Shi, T. G. Steele, V. Elias, K. B. Sprague, Y. Xue, and A. H. Fariborz, Holder inequalities and isospin splitting of the quark scalar mesons, Nucl. Phys. **A671**, 416 (2000).

[39] K. Maltman, Scalar meson decay constants and the nature of the $a_0(980)$, Nucl. Phys. **A680**, 171 (2000).

[40] P. A. Zyla *et al.* (Particle Data Group), Review of particle physics, Prog. Theor. Exp. Phys. **2020**, 083C01 (2020).

[41] Y. Aoki *et al.* (LatKMI Collaboration), Light flavor-singlet scalars and walking signals in $N_f = 8$ QCD on the lattice, Phys. Rev. D **96**, 014508 (2017).

[42] M. Golterman, Applications of chiral perturbation theory to lattice QCD, in *Les Houches Summer School: Session 93: Modern perspectives in lattice QCD: Quantum Field Theory and High Performance Computing* (Oxford University Press, 2009), arXiv:0912.4042.

[43] T. Appelquist, G. T. Fleming, M. F. Lin, E. T. Neil, and D. A. Schaich, Lattice simulations and infrared conformality, Phys. Rev. D **84**, 054501 (2011).

[44] G. T. Fleming *et al.* [Lattice Strong Dynamics (LSD) Collaboration], Dataset for "light scalar meson and decay constant in SU(3) gauge theory with eight dynamical flavors", 10.5281/zenodo.8007955 (2023).

[45] M. Gell-Mann, R. J. Oakes, and B. Renner, Behavior of current divergences under $SU_3 \times SU_3$, Phys. Rev. **175**, 2195 (1968).

[46] T. Banks and A. Casher, Chiral symmetry breaking in confining theories, Nucl. Phys. **B169**, 103 (1980).

[47] H. Leutwyler and A. V. Smilga, Spectrum of Dirac operator and role of winding number in QCD, Phys. Rev. D **46**, 5607 (1992).

[48] A. Patella, GMOR-like relation in IR-conformal gauge theories, Phys. Rev. D **84**, 125033 (2011).

[49] E. T. Neil and J. W. Sitison, Model averaging approaches to data subset selection, Phys. Rev. E **108**, 045308 (2023).

[50] G. P. Lepage, The analysis of algorithms for lattice field theory, in *Theoretical Advanced Study Institute in Elementary Particle Physics* (World Scientific, Singapore, 1989), 10.1142/9789814540988.

[51] E. T. Neil and J. W. Sitison, Improved information criteria for Bayesian model averaging in lattice field theory, Phys. Rev. D **109**, 014510 (2024).

[52] C. Michael, Fitting correlated data, Phys. Rev. D **49**, 2616 (1994).

[53] J. N. Simone, Improved data covariance estimation techniques applied to lattice QCD, *Proceedings of the the 35th International Symposium on Lattice Field Theory, Granada, Spain* (2017).

[54] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proc. Third Berkeley Symp. on Math. Statist. and Prob. Berkeley Symp. on Math. Statist. and Prob., **1**, 197 (1956).

[55] O. Ledoit and M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Anal. **88**, 365 (2004).

[56] J. Schäfer and K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, Stat. Appl. Genet. Mol. Biol. **4**, 32 (2005).

[57] A. C. Lichtl, Model optimization in the presence of correlations, in *Proceedings of the Fourth International Workshop on Numerical Analysis and Lattice QCD), QCDNA* (Yale University, New Haven, 2007).

[58] S. D. Cohen, G. T. Fleming, and H.-W. Lin, Exponential time series and lattice quantum field theory, in *Exponential Data Fitting and Its Applications*, edited by V. Pereyra and G. Scherer (Bentham Science Publishers, Sharjah, 2010), pp. 71–93, 10.2174/978160805048211001010071.

[59] E. Rinaldi, S. Syritsyn, M. L. Wagman, M. I. Buchoff, C. Schroeder, and J. Wasem, Lattice QCD determination of neutron-antineutron matrix elements with physical quark masses, Phys. Rev. D **99**, 074510 (2019).

[60] S. R. Beane *et al.* (NPLQCD and QCDSF Collaborations), Charged multihadron systems in lattice QCD + QED, Phys. Rev. D **103**, 054504 (2021).

[61] P. Shanahan, M. Wagman, and Y. Zhao, Collins-Soper kernel for TMD evolution from lattice QCD, Phys. Rev. D **102**, 014511 (2020).

[62] A. Bazavov *et al.* (Fermilab Lattice and MILC Collaborations), D-meson semileptonic decays to pseudoscalars from four-flavor lattice QCD, Phys. Rev. D **107**, 094516 (2023).

[63] O. Ledoit, M. Wolf *et al.*, Optimal estimation of a large-dimensional covariance matrix under Stein's loss, Bernoulli **24**, 3791 (2018).

[64] O. Ledoit and M. Wolf, Analytical nonlinear shrinkage of large-dimensional covariance matrices, Ann. Stat. **48**, 3043 (2020).

[65] Z. Burda and A. Jarosz, Cleaning large-dimensional covariance matrices for correlated samples, Phys. Rev. E **105**, 034136 (2022).

[66] T. Appelquist *et al.* [Lattice Strong Dynamics (LSD) Collaboration], Lattice calculation of composite dark matter form factors, Phys. Rev. D **88**, 014502 (2013).

[67] P. Pébay, Formulas for robust, one-pass parallel computation of covariances and arbitrary-order statistical moments, Sandia report, 2008, 10.2172/1028931.